DR. LAURENCE FLORI (Orcid ID : 0000-0002-7529-8521)

DR. EMMANUELLE JOUSSELIN (Orcid ID : 0000-0001-9548-202X)

DR. MATHIEU GAUTIER (Orcid ID : 0000-0001-7257-5880)

Article type : Original Article

A genomic map of climate adaptation in Mediterranean cattle breeds

Running Title: Climate Adaptation in Mediterranean cattle

Authors and Affiliations

Laurence FLORI[1§], Katayoun MOAZAMI-GOUDARZI[2], Véronique ALARY[1,3†], Abdelillah ARABA[4†], Ismaïl BOUJENANE[4†], Nadjet BOUSHABA[5†], François CASABIANCA[6†], Sara CASU[7†], Roberta CIAMPOLINI[8†], Armelle COEUR D'ACIER[9†], Corinne COQUELLE [10†], Juan-Vicente DELGADO[11†], Ahmed EL-BELTAGI[12†], Georgia HADJIPAVLOU[13†], Emmanuelle JOUSSELIN[9†], Vincenzo LANDI[14†], Anne LAUVIE[1†], Philippe LECOMTE[1,15†], Christina LIGDA[16†], Caroline MARINTHE[10†], Amparo MARTINEZ[14†], Salvatore MASTRANGELO[17†], Dalal MENNI[4†], Charles-Henri MOULIN[1†], Mona-Abdelzaher OSMAN[12†], Olivier PINEAU[18†], Baldassare PORTOLANO[17†], Clementina RODELLAR[19†], Nahdira SAIDI-MEHTAR[5†], Tiziana SECHI[7†], Guilhem SEMPERE[20,21†], Sophie THEVENON[20,21†], Dimitrios TSIOKOS[22†], Denis LALOË[2‡§] and Mathieu GAUTIER[9,23‡]

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.15004

[†]authors listed alphabetically

[‡]DL and MG should be considered joint senior author

[1]SELMET, INRA, CIRAD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

[2]GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

[3]CIRAD, UMR SELMET, ICARDA, Rabat, Morocco

[4]Institut Agronomique et Vétérinaire Hassan II, Département de Productions et de Biotechnologies Animales, 10101 Rabat, Morocco

[5]Université d'Oran "Mohamed Boudiaf", Département de Génétique Moléculaire Appliquée, 31000 Oran, Algeria

[6]LRDE, INRA, 20250 Corté, France

[7]Agris-Sardegna Servizio Ricerca per la Zootecnica, Olmedo (SS), Italy

[8]Dipartimento di Scienze Veterinarie, Università di Pisa, 56100 Pisa, Italy

[9]CBGP, INRA, CIRAD, IRD, Montpellier Supagro, Univ. Montpellier, Montferrier-sur-Lez, France

[10] Corsica vaccaghji, Corté, France

[11] Department of Genetics, University of Cordoba. C.U. Rabanales 14014-Cordoba, Spain.

[12]APRI, Animal Breeding and Genetics, Cairo, Egypt

[13]Agricultural Research Institute, P.O.Box 22016, 1516 Lefkosia, Cyprus

[14]Animal Breeding Consulting SL, Laboratorio de Genetica Molecular Aplicada, 14071 Cordoba, Spain

[15]CIRAD, UMR SELMET, F34398 Montpellier, France

[16]HAO-Demeter, Veterinary Research Institute, 57001 Thessaloniki, Greece

[17]Dipartimento Scienze Agrarie, Alimentari e Forestali, Università degli Studi di Palermo, 90128 Palermo, Italy

[18]Centre de recherche de la Tour du Valat, 13200 Arles, France

[19]LAGENBIO, Facultad de Veterinaria, Instituto Agroalimentario de Aragón-IA2, Universidad de Zaragoza-CITA, 50013 Zaragoza, Spain

[20]INTERTRYP, Univ. Montpellier, CIRAD, IRD, Montpellier, France

[21]CIRAD, UMR INTERTRYP, F-34398 Montpellier, France

[22]HAO-Demeter, Research Institute for Animal Science, 58100 Pella, Greece

[23]IBC, Institut de Biologie Computationnelle, 34095 Montpellier, France

[§]Corresponding authors

Emails: laurence.flori@inra.fr

denis.laloe@inra.fr

Abstract

Domestic species such as cattle (*Bos taurus taurus* and *B. t. indicus*) represent attractive biological models to characterize the genetic basis of short term evolutionary response to climate pressure induced by their post-domestication history. Here, using newly generated dense SNP genotyping data, we assessed the structuring of genetic diversity of 21 autochtonous cattle breeds from the whole Mediterranean basin and performed genome-wide association analyses with covariables discriminating the different Mediterranean climate sub-types. This provided insights into both the demographic and adaptive histories of Mediterranean cattle. In particular, a detailed functional annotation of genes surrounding variants associated with climate variations highlighted several biological functions involved in Mediterranean climate adaptation such as thermotolerance, UV protection, pathogen resistance or metabolism with strong candidate genes identified (e.g. *NDUFB3*, *FBN1*, *METTL3*, *LEF1*, *ANTXR2* and *TCF7*). Accordingly, our results suggest that main selective pressures affecting cattle in Mediterranean area may have been related to variation in heat and UV exposure, in food resources availability and in exposure to pathogens, such as anthrax bacteria (*Bacillus anthracis*). Furthermore, the observed contribution of the three main bovine ancestries (indicine, European and African taurine) in these different populations suggested that adaptation to local climate conditions may have either relied, on standing genomic variation of taurine origin or adaptive introgression from indicine origin, depending on the local breed origins. Taken together, our results highlight the genetic uniqueness of local Mediterranean cattle breeds and strongly support conservation of these populations.

Keywords

Cattle, Mediterranean, Genetics, SNP, Local adaptation, Climate

Introduction

Domestic species represent attractive biological models to characterize the genetic basis of short term evolutionary response to abiotic selective pressure such as climate. Indeed, during their human mediated migration from restricted domestication centers, they colonized a wide range of new environments and differentiated into a variety of populations or breeds shaped by the combined action of natural and artificial selection, and genetic drift. For instance, Cattle (*Bos taurus*) have been domesticated ca. 10,000 years before present (YBP) in two independent domestication centers, located in the Fertile Crescent for taurine (*Bos t. taurus*) and in the Indus valley for the zebu (*Bos t. indicus* (Loftus *et al.*, 1994). These cattle population then migrated with farmers on the different continents leading to more than 800 different cattle breeds now recognized worldwide (Felius *et al.*, 2014; Felius *et al.*, 2011). Standing at a crossroad of several early cattle migration routes, the Mediterranean basin appeared as a central area for European and African population exchanges. Indeed, the Near East Neolithic farmer populations followed two main routes through Europe migrating from 9,000 YBP to Central Europe via the so-called Danubian route (i.e., through Anatolia, Thrace and the Balkans), and to Southern Europe via the so-called Mediterranean route (i.e., using a maritime route by the Mediterranean sea) reaching Italy, France and Spain, 6,000 to 6,500 YBP (Cherry, 1981; Gautier *et al.*, 2010b; Payne & Hodges, 1997; Porter, 1991). Conversely, in Africa, taurine cattle were introduced through Egypt along the Mediterranean littoral, 6,500 YBP, although migrations from Southern Europe via Mediterranean islands also

possibly occurred (Payne & Hodges, 1997). Finally, if zebu cattle were first introduced in Africa via the Suez route, 3,500 YBP, their two major introductions occurred later via maritime routes to the Horn of Africa with Muslim expansion during the 7th century AD and following the rinderpest epidemics of 19th century (Epstein & Mason, 1984; Payne & Hodges, 1997). Zebu then spread gradually to the entire African continent, including Mediterranean regions and probably in Southern Europe, as suggested by the indicine ancestry detected in some Italian and Greek breeds (Cymbron *et al.*, 2005; Gautier *et al.*, 2010a). During this complex colonization history, and over the course of less than 1,500 generations, cattle became adapted to a variety of local climatic conditions across the Mediterranean basin. (Gautier *et al.*, 2007). Indeed, although the Mediterranean climate is mainly characterized by dry summers and mild and moist winters, a wide range of climate sub-types still persists from "dry-summer temperate" in Southern Europe to “dry-summer subtropical" in North Africa (Peel *et al.*, 2007).

Capitalizing on the genomic resources available for cattle, and in particular in SNP genotyping assays (Matukumalli *et al.*, 2009), it is now possible to scan the genome for regions involved in adaptation to local environment. To that end, Genome-Environment Association (GEA) analyses have facilitated identification of genetic variants associated with population-specific environmental covariables (Coop *et al.*, 2010; De Villemereuil & Gaggiotti, 2015; Frichot *et al.*, 2013; Gautier, 2015; Joost *et al.*, 2007). When combined with a functional annotation of candidate genes or variants, such a population genomics approach has proved to be efficient at highlighting candidate genes or physiological pathways affected by climate variation in both animal (e.g. (Fumagalli *et al.*, 2011; Gao *et al.*, 2017; Hancock *et al.*, 2011; Hancock *et al.*, 2008; Lv *et al.*, 2014)) and plant (e.g. (Abebe *et al.*, 2015; Fournier-Level *et al.*, 2011; Frachon *et al.*, 2018)) species.

Here, using a newly developed genotyping data set, we aimed at providing insights into both the demographic and adaptive histories of Mediterranean cattle, with a focus on climate adaptation. More precisely, we analyzed the structure of the genetic diversity across 21 local cattle breeds inhabiting eight different countries on both sides of the Mediterranean sea (Spain, France, Italy, Greece, Cyprus, Egypt, Algeria and Morocco) and infer their relationships with other breeds. We then scanned the bovine genome to identify footprints of adaptation to the Mediterranean climate using the methods implemented in the BayPass software (Gautier, 2015) and climatic variables associated with the cattle sampling locations. A detailed functional annotation of the identified candidate genes using system biology tools (Flori *et al.*, 2009; Flori *et al.*, 2014; Gautier *et al.*, 2009) uncovered the main physiological pathways mobilized during cattle adaptation to the Mediterranean climate.

Materials and Methods

Animal sampling and genotyping

For the purpose of this study, 344 individuals belonging to 16 different breeds were sampled for DNA extraction in seven Mediterranean countries (Figure 1A and Table 1). For all but CYP individuals (see Table 1 for acronyms details), the sampled tissues corresponded to blood collected in EDTA vacutainer tubes from which DNA was extracted using standard kits (Wizard Genomic DNA purification kit, Promega or Isolate II Blood DNA kit, Bioline). For the 32 NAN individuals, whole genome amplification was further performed using the GenomePlex WGA kit (Sigma-Aldrich). For the CYP individuals, nasal swabs were collected

using the Performagene PG-100 swab kit and DNA was extracted with the PG-AC extraction kit (DNA Genotek). All procedures of animal sample collection (blood sampling and collection of nasal swab samples) followed recommendations of the directive 2010/63/EU. In particular, blood sampling was performed by veterinarians, during prophylactic campaigns, following the standard procedures and relevant national guidelines of animal care. DNA samples were genotyped on the Illumina BovineSNP50 chip (v2) at Labogena platform (INRA, Jouy-en-Josas) using standard procedures (www.illumina.com). Genotypes were then integrated in the WIDDE database (Sempéré *et al.*, 2015). Using WIDDE utilities, these newly generated data were combined with existing publicly available ones (Decker *et al.*, 2014; Gautier *et al.*, 2009; Gautier *et al.*, 2010b; Matukumalli *et al.*, 2009) to generate two data sets referred to as the WORLD-Set and the MED-Set (Figure 1A and Table 1). The WORLD-Set was generated to provide a global view of the relationship of Mediterranean cattle with other breeds and consisted of 1,704 animals belonging to 62 different breeds representative of the bovine worldwide diversity. The MED-Set consisted of the 640 individuals from the WORLD-Set that belonged to 21 breeds (from eight countries) representative of the different Mediterranean climates. Note that four breeds included in the WORLD-Set and inhabiting Mediterranean countries were discarded from the Med-Set, i) the Turkish TUR (n=8) and ANA (n=31 individuals from four different Anatolian populations) populations due to a lack of information on their origins; ii) the Spanish AST (n=13) and NAV (n=30) populations because Asturian cattle are bred in a region distant from the Mediterranean sea and because the animals originated from the different Casta Navarra fighting bulls lines are raised in various Spanish localities; and iii) the Algerian BIS breed because this population has been subjected to significant admixture with Northern

European breeds (Boushaba *et al.*, 2018) making it unrepresentative of its geographic region of origin (see Results section).

For both the WORLD-Set and MED-Set, the minimal SNP genotyping call rates was set to 95% (i.e., animals genotyped for less than 95% of SNPs were discarded) and the minimal individual genotyping call rate was set to 90% over all populations and 75% within each population (i.e., SNPs genotyped for less than 75% of the animals from at least one population were discarded). SNPs with a MAF<0.01 or departing from Hardy-Weinberg equilibrium expectation (exact test p-value<0.001) were also discarded. A total of 33,301 and 39,921 SNPs were finally available in the WORLD-Set and MED-Set, respectively.

Analyses of cattle breed genetic diversity and structure based on the WORLD-Set

Principal component analysis (PCA) based on the SNP genotypes (Figure 2A) was performed with the *smartpca* software (Patterson *et al.*, 2006) and results were visualized with the R package *ade4* (Chessel, 2004). Unsupervised genotype-based hierarchical clustering of the individual animal samples (Table S1) was carried out using the maximum-likelihood method implemented in ADMIXTURE 1.06 (Alexander *et al.*, 2009) and visualized results with custom functions in the R environment (http://wwwq.r-project.org; Table S1 and Figures 2B, S1 and S2). Allele Sharing Distances (ASD) were computed for each pair of individuals using all available SNP information (see (Gautier *et al.*, 2010b) for details) and a neighbor-joining tree (Figure S3) was computed based on the resulting distance matrix using the R package APE (Paradis *et al.*, 2004).

In addition within-population F_{IS} and between populations F_{ST} were estimated (Table 1 and Table S2) using custom R functions implementing the estimator derived by Weir and Cockerham (Weir & Cockerham, 1984). A hierarchical clustering of the population pairwise F_{ST} matrix using the average agglomeration method (Figure S4) was performed with the R package *fastcluster* (Müllner, 2013).

Inference of the genetic history of the breeds included in the MED-Set were estimated using *TreeMix* 1.13 (Pickrell & Pritchard, 2012) together with one breed representative of indicine (NEL), African taurine (NDA) and European taurine (ANG) ancestry taken in the WORLD-Set. First, the maximum likelihood tree was inferred, considering NEL as an outgroup (Figure S5 A) and the residuals from the fit of the model to the data were visualized (Figure S5 B). Then, in order to identify aspects of ancestry not captured by the tree, we sequentially added one to 20 migration events to the tree. For all analyses performed with *TreeMix* a window size of 400 consecutive SNPs (ca. 20 Mb) was used to estimate the standard error of the elements of the (scaled) sample covariance matrix among population allele frequencies. The percentage of the variance in breed relatedness explained by each tree model (Figure S5 C) was calculated according to Pickrell and Pritchard (Pickrell & Pritchard, 2012). In addition, formal tests for admixture were performed using the three-population test (Patterson *et al.*, 2012) and the *threepop* program (Pickrell & Pritchard, 2012) available in the *TreeMix* package (Table S3).

Whole Genome-scan for adaptive differentiation on the MED-Set.

Whole genome-scans for adaptive differentiation and association with population-specific co-variables were performed with BayPass 2.1 (Gautier, 2015). The underlying models explicitly account for the covariance structure among the population allele frequencies, summarized by a scaled covariance matrix Ω, which makes the approach particularly robust to complex demographic histories (Gautier, 2015). The model also included hyperparameters (a_π and b_π) for the Beta distribution of across population allele frequencies which allows to account for the SNP chip ascertainment bias (Gautier, 2015). Identification of overly differentiated SNPs was based on the *XtX* statistics (Gautier, 2015; Gunther & Coop, 2013) estimated under the core model of BayPass together with the matrix Ω. To calibrate the *XtX*s, a pseudo-observed data set (POD) containing 100,000 SNPs simulated under the inference model with hyperparameters equal to those estimated on the real data set was generated and further analyzed under the same conditions following the procedure described in (Gautier, 2015). In particular, we ensured that the posterior estimate of Ω obtained with the POD was similar to that obtained on the real data since the FMD distance proposed by Förstner and Moonen (2003) between the two matrices was found equal to 0.21 and the correlation between their elements was close to one (ρ=0.99992). Similarly, the posterior means of the two hyperparameters a_π and b_π for the Beta distribution of across population allele frequencies obtained on the POD ($\widehat{a_\pi}$ =1.188; $\widehat{b_\pi}$ = 1.191) were almost equal to the ones obtained in the original data ($\widehat{a_\pi}$ = 1.189; $\widehat{b_\pi}$ = 1.190). Taken together, these sanity checks indicated that the POD faithfully mimics the real data set, allowing us to define a 0.1% significance threshold on the *XtX* statistics (*XtX* = 38.6) used to identify genomic regions harboring footprints of selection.

To that end, the UMD3.1/Btau6 bovine genome assembly (Liu *et al.*, 2009) was split into 4,984 consecutive 1-Mb windows with 500kb overlap. Windows which included at least two significant SNPs (at the 0.1% *XtX* POD threshold) were deemed significant. After merging overlapping significant windows, we identified for each resulting region the strongest candidate gene as the one located less than 15 kb from the *XtX* peak. In some instances, some functional candidate genes contiguous to this positional candidate gene were also identified based on their known biological role.

Genome-Environment Association analyses and identification of candidate genes associated with climatic covariables

Using GPS coordinates associated to each breed of MED-Set, 35 climatic covariables (from Bio01 to Bio35) were extracted from the Climond database (https://www.climond.org/; (Kriticos *et al.*, 2012); Tables S5 and S6). This database contains bioclimatic variables with original measurable information on annual, weekly, and seasonal temperatures, soil moisture, radiation, and precipitation, that we used as a proxy to describe the climate (Ficetola *et al.*, 2017). These variables are a summary of climatic conditions between 1961 and 1990 in the form of rasters at about 19 km spatial resolution. The annual average of the Temperature Humidity Index (THI), that may be viewed as a measure of discomfort experienced by an animal in warm weather, was calculated for each breed using formula 5 described in Bohmanova *et al.* (Bohmanova *et al.*, 2007). The geographic patterns of Bio01, Bio07, Bio12, Bio20 and Bio28 were represented in Figure 1B, C, D, E and F using *raster* package (Hijmans & Van Etten, 2012).

A PCA including the 35 climatic covariates (from Bio01 to Bio35) was performed to reduce dimensionality and estimate correlations between the 35 climatic variables using the R packages *ade4* (Chessel, 2004). This analysis revealed a mainly one-dimensional structure, with first four components summarizing 90% of the information from the original set of variables (57.22%, 16.56%, 10.23%, and 6.13% for the axes 1 to 4, respectively). The PCA coordinates of each breed, from the first three axes, are synthesized in Figure S6A by means of a colorplot projected on the geographic map using the R packages *rworldmap* (South, 2011), *sp* (Pebesma & Bivand, 2005), *raster* (Hijmans & Van Etten, 2012) and *adegenet* (Jombart, 2008). This map associates each breed with the climate conditions they are exposed to, revealing an expected South to North Mediterranean climate range (Figure S6B and Table S5). Hence the first PC was found to be mainly built by a temperature *vs.* rain opposition closely linked to the latitude while the close location of variables associated to minimal values of radiation revealed that North and South Mediterranean climates differ by radiation levels in winter. The second PC was not related to the geography, and opposed the temperature range to maximal values of radiation. Based on this characterization, ten different covariables were finally retained for GEA studies corresponding to i) the breed coordinates on PC1 to PC4, ii) Bio01 (annual mean temperature), Bio07 (temperature annual range), Bio12 (annual precipitation), Bio20 (annual mean radiation) and Bio28 (annual mean moisture index) original climate covariables and iii) the THI (Tables S4).

Genome-wide GEA were carried out with the BayPass software (Gautier, 2015) using the default options of the AUX model parameterized with the scaled covariance matrix $\hat{\Omega}$ estimated above (Figure S7). This model explicitly accounts for multiple tests by integrating over (and estimating) the unknown proportion of SNPs actually associated with a given

covariable. The support of the association of each SNP with each covariable was evaluated by a Bayes Factor (BF).

In order to identify candidate genes associated with climatic covariables, all the genotyped SNPs were first annotated using as a gene set reference a list of 14,627 RefSeqGenes anchored on the UMD3.1/Btau6 bovine genome assembly (Liu *et al.*, 2009) that was uploaded from the UCSC RefGene database (http://hgdownload.cse.ucsc.edu/goldenPath/bosTau6/database/). A SNP was then considered as representative of one of these RefSeq genes if localized within its boundary positions extended by 15kb upstream and downstream to account for persistence of Linkage Disequilibrium across populations (Gautier *et al.*, 2007). Indeed, as shown in Figure S8, r^2 measured across the MED-Set populations remained >0.2 on average for SNPs separated by less than 20 kb over the whole genome. Among the 39,921 SNPs from the MED-Set included in the analysis, 14,038 SNPs mapped to 9,023 different RefSeqGenes (see details in Table S6), which corresponds to 8,929 gene symbols. On average, each SNP mapped within 1.2 RefSeq genes (from 1 to 29, median=1) while each RefSeq gene was represented by 3.7 SNP (from 1 to 32, median=1). Genes which included at least one SNP displaying a *BF*>20 (decisive evidence for association according to the Jeffreys' rule (Jeffreys, 1961)) with a given covariable were considered as candidate genes (Table 2 and Table S6).

Functional annotation of candidate genes

Candidate genes were functionally annotated and submitted to gene network analyses using the *Ingenuity Pathway Analysis* software (IPA, Ingenuity®Systems,). Among 8,929 gene symbols represented in the analysis, 8,312 were mapped to the Ingenuity Pathway Knowledge Base (IPKB) and corresponded to the reference dataset. Among the 55 genes

associated with at least one climatic covariable, 49 were mapped in IPKB and were ready for IPA analysis. The top significant functions and diseases (p-value<0.05) were obtained by comparing functions associated with the genes under selection against functions associated with all genes in our reference set using the right-tailed Fisher's exact test. For each network containing at most 140 molecules, a score S was computed based on a right-tailed Fisher exact test for the overrepresentation of associated genes (S=-log(p-value)). A network was considered significant if S>3.

Results

Structuring of genetic diversity among the Mediterranean cattle breeds

To characterize the structuring of genetic diversity among the Mediterranean cattle breeds, we first analyzed the WORLD-Set dataset that consisted of genotyping data (33,301 SNPs) on 1,704 individuals from 62 different populations including the 21 breeds of our Mediterranean specific dataset (MED-Set data set) and 41 other breeds representative of the world-wide cattle diversity (Figure 1, Table 1 and the Material and methods section). As expected, the first factorial plan of PCA on individual genotyping data (Figure 2A) recovered the previously described triangle-like 2-Dimensional global structuring of the worldwide cattle genetic diversity with European taurine (EUT), African taurine (AFT) and zebuine (ZEB) populations at the three apexes (Bradley *et al.*, 1998; Gautier *et al.*, 2009; Gautier *et al.*, 2010b; MacHugh *et al.*, 1997). In agreement with their geographical origin, most individuals belonging to the Mediterranean breeds clustered between individuals from EUT and AFT

origins, with a positioning closely related to their latitude of origin. Indeed, individuals from European Mediterranean countries (e.g., the French Raço di Biòu or RDB and the Spanish Casta Navarra or NAV and Negra Andaluza or NAN populations) were closer to the EUT apex while individuals from African Mediterranean countries tended toward the AFT apex suggesting substantial AFT ancestry (e.g., the Moroccan Oulmès Zaër or OUL, Brune de l'Atlas or BAT and Tidili or TID and the Algerian Chelifienne or CHF populations) with the noticeable exception of the Algerian Biskra (BIS) breed that was located within the EUT populations. More strikingly, individuals belonging to the Turkish grey (TUR), Anatolian (ANA), Cyprus (CYP), Greek Brachykeratiki (GRE), the five Italian Maremmana (MAM), Modicana (MOD), Sardo-Modicana (SAM), Romagnola (RMG) and Chianina (CHI) and to a lesser extent the Corsican COR populations shifted from the EUT-AFR side toward the ZEB apex thereby suggesting zebu admixture in these populations. Similarly, for North and North-East African populations, individuals from the Algerian Cheurfa (CHE) and Guelmoise (GUE) and the Egyptian Baladi (BAL) populations shifted toward the ZEB apex while staying closer to the center of the triangle thus suggesting a substantial ZEB ancestry. Model based hierarchical clustering approaches with varying number K of predefined genetic clusters (Figure 2B and Table S1 for K=3 and Figure S1 for K=2 up to K=40), as implemented in the ADMIXTURE software, essentially confirmed the pattern of the structuring of genetic diversity revealed by PCA, in particular when K=3, if one interprets the clusters as overall EUT, AFT and ZEB ancestries (respectively represented in red, blue and green in Figure 2B). Note however that the clusters underlying the EUT and AFT ancestries might also be assumed to represent the extremes of a North-South gradient for the taurine cattle genetic diversity thus warning against over-interpreting the high contribution of the blue cluster

observed for European breeds in Figure 2B (e.g., about 10% in the Alpine Abondance and up to 20% in the Spanish NAV individuals) as evidence for AFT introgression.

As already suggested by the above results and previous studies (e.g., Gautier et al., 2010), the NJ tree based on the ASD (Figure S4) showed that all the individuals unambiguously clustered according to their population origin. $F_{IS,}$ as estimated from the Bovine SNP50 assay known to be enriched for SNP segregating in European populations (Gautier *et al.*, 2010b; Matukumalli *et al.*, 2009), were small (<0.038 in absolute value) and were all in the range of other European breeds except for the NAN (F_{IS}=0.111) and GRE (F_{IS}=0.116) presumably related to higher inbreeding levels in the sampled individuals (Table 1). Among the 21 Mediterranean populations of the MED-Set, we estimated an overall F_{ST} equal to 0.101 (to compare with the value of 0.158 found among the 62 worldwide populations of the WORLD-Set). The matrix of all population pairwise F_{ST} (Table S2 and Figure S4) was also consistent with PCA and ADMIXTURE results carried out on individual genotyping data (Figure 2). Within the 21 Mediterranean populations, pairwise F_{ST} ranged from 0.00 between the two Algerian CHE and GUE populations to 0.273 between the Spanish Mallorquina (MAL) and the Cyprus CYP (to compare with maximal value of 0.471 found between the Lagune or LAG and Nelore or NEL populations with WORLD-Set).

The *TreeMix* inference model (Pickrell & Pritchard, 2012), provided further insights into the post-domestication history of Mediterranean cattle breeds that led to the observed structuring of genetic diversity. The phylogenetic tree of the relationships between the 21 cattle breeds of the MED-Set together with the Angus (ANG), N'Dama (NDA) and NEL breeds (representative of EUT, AFT and ZEB, respectively) largely recapitulates the known relationships among the cattle breeds (Gautier *et al.*, 2009; Gautier *et al.*, 2010b;

Matukumalli *et al.*, 2009), but it only explains 45.7% of the variance in relatedness between populations (Figure S5 A). An admixture graph obtained after adding 15 migration events optimally improved the model fit since it explained more than 99% of the genetic variance across populations (Figure 3, Figure S5 C). The inferred migration edges essentially confirmed several of the admixture events already suggested by the exploratory analyses (see above). These include, for instance, migration edges between i) African taurine and BAL or CYP; ii) European taurine and North African breeds (i.e. TID, GUE, CHE branch, CHF or BAT); and iii) Zebu and breeds from North Africa (i.e. CHE, GUE, TID), from Italy (RMG and MAM), from Cyprus (CYP) or from Egypt (BAL). It should also be noticed that one migration event was added between the Cyprus (CYP) and the Greek (GRE) breed. In addition, *f3* based tests showed that (i) BAL, CYP, CHE and GUE derived from presumably three-way admixture between breeds representative of EUT, AFT and ZEB ancestries, (ii) BAT, OUL, CHF, TID are EUTxAFT admixed breeds and (iii) GRE and Cinesara (CIN) are EUTxZEB admixed (Table S3).

Identification and functional annotation of genes associated with climatic covariables

In order to identify genes involved in adaptation to the Mediterranean climate, we focused on the MED-Set and performed GEA with BayPass (Gautier, 2015) for several climatic covariables representative of the Mediterranean climate corresponding to i) the four climatic covariables (PC 1 to 4) summarizing the 35 climatic parameters of the Climond database (Tables S5 and S6); ii) five original climatic covariables (Bio01, Bio07, Bio12, Bio20 and Bio28), representative of the main climatic covariable categories and related to temperature, precipitation, radiation and humidity (Figure 1); and iii) the temperature humidity index, THI. Bayes Factor estimates of the support for association of each SNP with

the different covariables are plotted in Figure S9. We thus identified 5, 19, 2, 3, 10, 6, 7, 6, 8 and 15 genes associated with PC1, PC2, PC3, PC4, Bio01, Bio07, Bio12, Bio20, Bio28 and THI, respectively (Table 2 and Table S6). Full gene names are detailed in Table S9.

Several of these genes were actually associated with at least two climatic variables like i) *VDAC1* (associated with PC1, Bio12 and Bio20), *TCF7* (associated with PC1, Bio01, Bio12, Bio20, Bio28 and THI) and *SKP1* (associated with PC1, Bio01, Bio12, Bio20, Bio28 and THI), located on the same region on BTA7, ii) *ANTXR2* on BTA6 and *GLDC* on BTA8 associated with both Bio01 and THI, iii) *RETSAT* and *ELMOD3* located within the same region on BTA11 and *LAMC1* on BTA16 both associated with PC2 and Bio07, iv) *KCNH1* on BTA16 associated with PC2, Bio07, Bio12 and Bio28, v) *HSPB3* located on BTA20 associated with Bio12 and Bio28, and vi) *TRAM2* on BTA23 associated with PC2, Bio20 and Bio28.

To obtain a more global view of the main gene functions targeted by the Mediterranean climate, the 55 genes associated with at least one climatic covariable were then functionally annotated using IPA (Ingenuity®Systems). The main functional categories, in which the 49 genes that mapped to IPKB were involved, are listed in Table 3 (see Table S7 for an exhaustive list of their functional annotation). The top five significant functions belonging to the three IPA categories “Physiological System Development and Function”, “Molecular and Cellular Functions” and “Diseases and Disorders” were related to i) the development and functioning of the nervous system (“Nervous system development and function”), ii) the tissue and organ development and morphology (“Tissue morphology”, “Organ morphology”, “Organismal development”, “Organismal injury and abnormalities”), iii) to cancer (“Cancer”, “Gastrointestinal disease”, “Tumor morphology”) and to cell development and proliferation (i.e. “Cell morphology”, “Cellular assembly and organization”, “Cellular development”,

"Cellular function and maintenance", "Cellular growth and proliferation" and Cellular development and proliferation").

Furthermore, these 49 genes participated to two significant networks (Tables S8 and S9) interconnected by two molecules, DUSP4 and GLDC, the latter being also associated with Bio01 and THI. The first network was found involved in i) "Cellular movement"; ii) "Cellular development" and iii) "Cellular growth and proliferation", while the second one was found involved in i) "Cancer"; ii) "Cell morphology"; iii) "Organismal injury and abnormalities". These two networks were then merged in a global network comprising 278 molecules in Figure 4. Among the central nodes of this global network, five molecules belonged to the set of genes associated with at least one climatic variable: i) the two heat shock proteins HSPB3 and HSP27 (note that HSPB3 is included in the HSP27 complex), ii) alpha-catenin, iii) SMYD3, and iv) NRG1.

Overall, the functional annotation of genes associated with climatic variable pointed towards the following main physiological functions involved with i) cancer and cellular proliferation; ii) morphology; and iii) nervous system development and functioning. A detailed inspection of the gene functional annotation (Table S7) also revealed additional interesting functions related to i) skeletal and muscular development and function (involving *FBN1*, *GPR87*, *DISP1* and *VDAC1*); and ii) immune and anti-infectious responses, (involving in particular *LEF1*, *TCF7*, *MAP3K8*, *MLST8*, *NRG1*, *SMYD3* and *ANTXR2*).

Identification of selection footprints in the genome of the 21 Mediterranean cattle breeds

If climate was assumed to represent a critical adaptive constraint, other selective pressures may have significantly impacted Mediterranean cattle. We thus scanned the genome for footprints of adaptive differentiation focusing on the MED-Set based on the *XtX* statistic (Gautier, 2015). A total of 17 candidate genes under selection could be identified according to their function or to their proximity with the *XtX* peak (Table 4, Figure S9A). These included i) *RACGAP1*, *AQP6*, *TIGAR* and *CCND2* on BTA05; ii) RPL34, *LEF1*, *LAP3*, *NCAPG*, *LCORL* and *ANTXR2* on BTA06; iii) *VDAC1*, *TCF7* and *SKP1* on BTA07; iv) *B3GLCT* and *RXFP2* on BTA12; and *ZFPM* and *MC1R* on BTA18 (full gene names are detailed in Table 4).

Interestingly, three of the nine identified regions also contained candidate genes associated with climatic covariables, i) *LEF1* (Region#3 on BTA6) associated with THI; ii) *ANTXR2* (Region #5 on BTA 6) associated with both Bio01 and THI; and iii) *VDAC1*, *TCF7* and *SKP1* (Region #6 on BTA 7) all three associated with PC1, Bio12 and Bio20 (note that *TCF7* was also found associated with Bio1, Bio28 and THI et and *SKP1* with Bio28 et THI). In addition, two other genes (*C7H5ORF15* and *SAR1B*), located within Region #6 but not proposed as candidate genes under selection, were also associated with Bio20 and THI, respectively.

Six regions harboring footprints of selection (#1, #2, #4, #7, #8 and #9 in Table 4) did not contain any gene associated with any of the considered climatic covariables and might thus result from other selective pressures.

Discussion

Our newly developed medium density SNP genotyping data set demonstrates a clear structuring of the genetic diversity of local breeds across the Mediterranean basin resulting from a complex recent demographic history with admixture events involving the three major cattle groups. Yet, most of the local breeds we considered can be viewed as stabilized in the sense that they displayed a high degree of within population genetic homogeneity. This led us to consider the population or breed as the appropriate level of organization to draw a bovine genomic map of short term adaptive response to local Mediterranean climate variation that was mainly captured by climatic covariables related to temperature, precipitation, radiation and humidity.

Strikingly, the in-depth functional analysis we carried out based on the 55 candidate genes found associated with climatic covariables highlighted several functions or physiological pathways (e.g., cancer, pigmentation, metabolism or infection and immunity) already reported in other similar GEA performed in human and sheep to be important in climate adaptation (Hancock *et al.*, 2011; Hancock *et al.*, 2008; Lv *et al.*, 2014).

More specifically, our results suggested a central role of genes involved in cancer, cellular growth and proliferation while highlighting the impact of heat stress and UV exposure on physiology. Some plausible candidate genes were identified, such as *METTL3*, which regulates the UV-induced DNA damage response (Xiang *et al.*, 2017) and *LEF1*, which could be related to thermotolerance and UV protection, as suggested by its central role in hair pigmentation (Guenther *et al.*, 2014)and its location in a QTL for UV-protective eye area

pigmentation in cattle (Pausch *et al.*, 2012). In addition, LEF1 was also found under adaptive differentiation across Mediterranean breeds in our study and across Chinese local cattle breeds (Gao *et al.*, 2017). The functional annotation of the candidate genes also identifies genes involved in the development and function of the nervous system such as *CTNNA2*, *NRG1* and *RFX4* (Abe *et al.*, 2004; Dominguez *et al.*, 2018), the latter being also reported as associated with climate covariable in Chinese cattle (Gao *et al.*, 2017). These genes might be important to adapt to varying Mediterranean climatic conditions via their action on thermal regulation, behavioral and sensory system development. Other candidate genes play also a role in development such as NDUFB3 and *FBN1* that are involved in morphology and stature (Alston *et al.*, 2016; Hirano *et al.*, 2012). This might be viewed as support for the hypothesis that climate strongly influences body size with positive correlations between heat and aridity, and smaller size (Franks & Hoffmann, 2012). Finally, several candidate genes were found to be involved in amino acid metabolism (e.g. GADL1, *GLDC*, *NRG1*, *SLC46A1*) and in cardiovascular system or in development of urinary tract (e.g. *ALDH1A2*, *AMER1*, *CDH4*, *EYA1*, *FBN1*, *LAMC1*, *MST1*, *MYC*, NOG, *PTGS2*, *PTPRF*, *SHH*, *SLC19A1*, *SMO*, *SST*, *TP53*, *VEGFA*, *VHL*) which could be speculatively interpreted as resulting from physiological adaptations required to cope with drought in some Mediterranean areas.

Interestingly, some of the genes found associated with climate were actually involved in infectious disease resistance (e.g. *ANTXR2*, *MAP3K8*, *MLST8* and *SMYD3*), perhaps implicating pathogens with distributions that are influenced by climate. The most striking example is represented by *ANTXR2* that was found both under selection and associated with climate covariables, suggesting that anthrax could have exerted a strong selective pressure on Mediterranean cattle breeds. Indeed, this gene encodes the major receptor mediating *in*

vivo lethality of the toxin produced by *Bacillus anthracis* (Arevalo *et al.*, 2014; Liu *et al.*, 2012) responsible for anthrax, the oldest known zoonosis with a world-wide distribution, severely affecting human and ruminants (Velimirovic, 1984). This disease, is thought to have originated in Egypt and Mesopotamia and to be depicted in ancient writings since 1491 B.C (Schwartz, 2009). In Europe, it was covered in a 10th century collection of veterinary writings, was responsible for enormous domestic livestock losses in Europe from the 17th to the 19th century (Schwartz, 2009) and also occurred more recently, at the end of the last century, in the Mediterranean countries of Southern Europe, especially in Spain, Italy, Turkey and Greece. A link between climatic factors (i.e. temperature, precipitation patterns and solar radiation) and the onset of anthrax outbreaks have been established, as the spores of *Bacillus anthracis* are especially resistant in contaminated soils, where they can survive for years, (Dorofeev & Blagoveshchenskaya, 1978; Hugh-Jones & Blackburn, 2009). In particular, areas in Europe with a pronounced dry season, such as Mediterranean countries have a higher prevalence of animal anthrax (Velimirovic, 1984). This disease could also have exerted strong selective pressure in cattle from other areas, *ANTXR2* being found under selection in West-African cattle (Gautier *et al.*, 2009), and in other susceptible species since this gene was also found associated with climate variables in humans and sheep (Hancock *et al.*, 2011; Lv *et al.*) and correlated with pathogen diversity in humans (Fumagalli *et al.*, 2011).

In addition to *ANTXR2* and *LEF1*, we identified other genes, *VDAC1*, *TCF7* and *SKP1* that were all associated with climatic covariables and located in a genomic region displaying strong evidence of adaptive differentiation on BTA7. Interestingly, the strongest differentiation signal in this region is obtained close to *TCF7*, which, as *LEF1*, is a main downstream effector

of a signaling pathway (i.e. Wnt/β catenin) and both genes have recently been found to be involved in a regulatory feedback loop controlling taste cell renewal in the circumvallate papilla epithelium and loss of gustatory nerve fibers in mice (Gaillard *et al.*, 2017). Similarly, a SNP proximal to *LEF1* is also associated with feeding behavior and eating efficiency in Duroc pigs (Ding *et al.*, 2017).

Although difficult to characterize, it is nevertheless important to note that other human-mediated or natural selective pressures may also have driven the adaptive genetic diversity of the Mediterranean local cattle breeds. Hence, six genomic regions among the nine harboring evidence of adaptive differentiation did not contain any markers associated with climate. In particular, Region #4 that contains the genes *NCAPG* and *LCORL* associated with body weight and height (Takasuga, 2016) and Region #9 surrounding the coloration locus *MC1R* were previously identified in other cattle breeds as under selection (Flori *et al.*, 2009; Gautier, 2015; Gutierrez-Gil *et al.*, 2015; Qanbari *et al.*, 2014; Xu *et al.*, 2015) or associated with morphology and pigmentation scores respectively (Gautier, 2015).

Overall, our genome scan for association with climate covariable on breeds representative of various local Mediterranean conditions provided a global picture of the main targeted candidate physiological adaptations, which might be refined by analyzing a high-density SNP dataset in future studies. This illustrates in turn the originality and genetic potential of Mediterranean cattle breeds in particular in the context of global warming.(Hewitson *et al.*, 2014; Harris *et al.*, 2013; Segnalini *et al.*, 2013).

In addition to the identification of the genetic variants underlying adaptive response of Mediterranean cattle breeds to local climatic variation, characterizing their origin may thus be critical to promote the conservation of genetic resources and the associated traditional

herding systems that are threatened by the increasing use of a small number of commercial breeds (Bruford *et al.*, 2015; Savolainen *et al.*, 2013). The structuring of genetic diversity of Mediterranean local cattle breeds and their recent post-domestication history makes it unlikely that locally favorable genetic variants arose by new mutations. Instead, adaptation to local climate conditions may have rather relied on standing genetic variation that existed in both domesticated and wild European cattle during the period when domestication took place and as a consequence of early farmer migration, or may be the result of more recent adaptive introgression of indicine origin (both hypotheses being of course mutually non-exclusive). Zebu that diverged from taurine between 84,000 and 275,000 YBP (Bradley *et al.*, 1996; Ho *et al.*, 2008), have been subjected for a longer time to tropical and arid conditions and are now well adapted to these specific conditions while African taurine, which diverged from European taurine several thousand years ago (Bradley *et al.*, 1996; Stock & Gifford-Gonzalez, 2013), are also well adapted to tropical humid climate. It is also tempting to speculate that wild adapted aurochs sub-populations that lived in the corresponding area may have interbred with more recently arrived domesticated individuals (Achilli *et al.*, 2008; Park *et al.*, 2015; Upadhyay *et al.*, 2017). Although the current lack of genetic data on local aurochs individuals prevents us from assessing their contribution, our characterization at the genome-wide scale demonstrates an admixture with i) zebu ancestry in Southern Europe with an East to West decreasing gradient for Cyprus, Greek, Maremmana and Romagnola Italian and Corsican breeds, which confirms and refines previous studies (Cymbron *et al.*, 2005; Gautier *et al.*, 2010a); ii) African taurine ancestry in Maghreb (Tidili, Oulmes Zaër, Brune de l'Atlas and Chelifienne breeds); and iii) both zebu and African taurine ancestries in Maghreb (Cheurfa and Guelmoise breeds), Egypt and Cyprus. Taken together, this observed pattern of cattle genetic ancestry for the Mediterranean breeds remains

concordant with the known migration history of Neolithic farmers from the taurine domestication center in the Fertile Crescent toward the West through the Mediterranean coasts and major islands (e.g., cattle arriving in Corsica around 5,000 YBP), following the *Mediterranean route* (Vignes, 1999); the migration of taurine from North-Africa to Spain after their introduction in Africa through Egypt 6500 BC; and the crossing in Egypt of the various migration routes taken by settled communities towards Europe and Africa presumably leading to interbreeding of the cattle populations (Payne & Hodges, 1997). Cattle with indicine ancestry were probably brought in Southern Europe by the Silk Road route (~200 BC to 1720 CE) connecting Asia to the Mediterranean sea (and stopping in Italy), in agreement with the decreasing gradient of indicine ancestry observed from Sicily to Italy mainland and Corsica. The weak indicine ancestry detected in some Algerian breeds (i.e. Cheurfa and Guelmoise) probably resulted from admixture of African taurine with African zebu while EUT ancestry detected more generally in some North African breeds provides evidence of the more recent admixture with European taurine cattle breeds during the two last centuries (Porter, 1991).

Previous studies on different cattle breeds showed that both selection on standing genetic variation of taurine origin or indicine adaptive introgression are encountered. For instance, in the Caribbean Creole cattle breed from Guadeloupe island which originates from a three-way admixture between zebu, European and African taurine, the three ancestries contributed in selected variants (Gautier & Naves, 2011). More strikingly, in the tropical Senepol breed the slick locus variant involved in thermotolerance was found to be of European taurine ancestry (Flori *et al.*, 2012). Similarly, in our study, while no indicine ancestry was detected in breeds originating from Spain or mainland France, as already

observed for other European continental breeds (e.g., Gautier et al., 2010) genetic variants favorable to arid or hot climatic conditions are thus likely to be of taurine origin in these regions. Overall, due to their various origins, Mediterranean cattle breeds represent attractive models to further assess the relative contribution of the major bovine ancestries in climate adaptation.

Acknowledgements

This work was supported by the INRA Metaprogramme ACCAF grant 2012 (GALIMED project) and the Animal Genetics Division of INRA. We would like to thank all the breeders of the Algerian, Cyprus, Egyptian, French, Greek, Italian, Moroccan and Spanish cattle breeds, included in this study, for their help during animal sampling. We also would like to acknowledge Laurent Avon (retired from the Institut de l'Elevage), for his help and advices about the genetic characterization of the Raço di Biòu and Corsican cattle breeds.

References

Abe K, Chisaka O, Van Roy F, Takeichi M (2004) Stability of dendritic spines and synaptic contacts is controlled by alpha N-catenin. *Nat Neurosci*, **7**, 357-363.

Abebe TD, Naz AA, Leon J (2015) Landscape genomics reveal signatures of local adaptation in barley (Hordeum vulgare L.). *Front Plant Sci*, **6**, 813.

Achilli A, Olivieri A, Pellecchia M*, et al.* (2008) Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr Biol*, **18**, R157-158.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, **19**, 1655-1664.

Alston CL, Howard C, Olahova M*, et al.* (2016) A recurrent mitochondrial p.Trp22Arg NDUFB3 variant causes a distinctive facial appearance, short stature and a mild biochemical and clinical phenotype. *J Med Genet*, **53**, 634-641.

Arevalo MT, Navarro A, Arico CD*, et al.* (2014) Targeted silencing of anthrax toxin receptors protects against anthrax toxins. *J Biol Chem*, **289**, 15730-15738.

Bohmanova J, Misztal I, Cole JB (2007) Temperature-humidity indices as indicators of milk production losses due to heat stress. *J Dairy Sci*, **90**, 1947-1956.

Boushaba N, Boujenane I, Moazami-Goudarzi K*, et al.* (2018) Genetic diversity and relationships among six local cattle populations in semi-arid areas assessed by a bovine medium-density single nucleotide polymorphism data. *Animal*, 1-7.

Bradley DG, Loftus RT, Cunningham P, MacHugh DE (1998) Genetics and domestic cattle origins. *Evolutionary Anthropology*, **6**, 79-86.

Bradley DG, MacHugh DE, Cunningham P, Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci U S A*, **93**, 5131-5135.

Bruford MW, Ginja C, Hoffmann I*, et al.* (2015) Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025. *Front Genet*, **6**, 314.

Cherry JF (1981) Pattern and process in the earliest colonization of the Mediterranean islands. *Proc. Prehist. Soc.*, **47**, 41-68.

Chessel D (2004) The ade4 package. I. One-table methods. *R News*, **4**, 5-10.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411-1423.

Cymbron T, Freeman AR, Isabel Malheiro M, Vigne JD, Bradley DG (2005) Microsatellite diversity suggests different histories for Mediterranean and Northern European cattle populations. *Proc Biol Sci*, **272**, 1837-1843.

De Villemereuil P, Gaggiotti OE (2015) A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, **6**, 1248-1258.

Decker JE, McKay SD, Rolf MM*, et al.* (2014) Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet*, **10**, e1004254.

Ding R, Quan J, Yang M*, et al.* (2017) Genome-wide association analysis reveals genetic loci and candidate genes for feeding behavior and eating efficiency in Duroc boars. *PLoS One*, **12**, e0183244.

Dominguez SL, Hegde GV, Hanson JE*, et al.* (2018) Antibody-mediated stabilization of NRG1 induces behavioral and electrophysiological alterations in adult mice. *Sci Rep*, **8**, 8239.

Dorofeev AA, Blagoveshchenskaya GS (1978) Solar activity and anthrax epizootics. *Veterinariya*, **10**, 51-55.

Epstein HE, Mason IL (1984) Cattle. In: *Evolution of domesticated animals* (ed. Mason IL), pp. 6-27. Longman, London.

Felius M, Beerling ML, Buchanan DS*, et al.* (2014) On the History of Cattle Genetic Resources. *Diversity*, **6**, 705-750.

Felius M, Koolmees PA, B. T, Consortium ECGD, Lenstra JA (2011) On the Breeds of Cattle-Historic and Current Classifications. *Diversity*, **3**, 660-692.

Ficetola GF, Mazel F, Thuiller W (2017) Global determinants of zoogeographical boundaries. *Nat Ecol Evol*, **1**, 89.

Flori L, Fritz S, Jaffrezic F*, et al.* (2009) The genome response to artificial selection: a case study in dairy cattle. *PLoS One*, **4**, e6595.

Flori L, Gonzatti MI, Thevenon S*, et al.* (2012) A quasi-exclusive European ancestry in the Senepol tropical cattle breed highlights the importance of the slick locus in tropical adaptation. *PLoS One*, **7**, e36133.

Flori L, Thevenon S, Dayo GK*, et al.* (2014) Adaptive admixture in the West African bovine hybrid zone: insight from the Borgou population. *Mol Ecol*, **23**, 3241-3257.

Förstner W, Moonen B (2003) A metric for covariance matrices. In: *Geodesy-The Challenge of the 3rd Millennium*, pp. 299–309. Springer, Berlin Heidelberg.

Fournier-Level A, Korte A, Cooper MD*, et al.* (2011) A map of local adaptation in Arabidopsis thaliana. *Science*, **334**, 86-89.

Frachon L, Bartoli C, Carrere S*, et al.* (2018) A genomic map of climate adaptation in Arabidopsis thaliana at a micro-geographic scale. *Frontiers in Plant Science*.

Franks SJ, Hoffmann AA (2012) Genetics of climate change adaptation. *Annu Rev Genet*, **46**, 185-208.

Frichot E, Schoville SD, Bouchard G, Francois O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol*, **30**, 1687-1699.

Fumagalli M, Sironi M, Pozzoli U*, et al.* (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet*, **7**, e1002355.

Gaillard D, Bowles SG, Salcedo E*, et al.* (2017) beta-catenin is required for taste bud cell renewal and behavioral taste perception in adult mice. *PLoS Genet*, **13**, e1006990.

Gao Y, Gautier M, Ding X*, et al.* (2017) Species composition and environmental adaptation of indigenous Chinese cattle. *Sci Rep*, **7**, 16196.

Gautier M (2015) Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, **201**, 1555-1579.

Gautier M, Faraut T, Moazami-Goudarzi K*, et al.* (2007) Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics*, **177**, 1059-1070.

Gautier M, Flori L, Riebler A*, et al.* (2009) A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics*, **10**, 550.

Gautier M, Hocking TD, Foulley JL (2010a) A Bayesian outlier criterion to detect SNPs under selection in large data sets. *PLoS One*, **5**, e11913.

Gautier M, Laloe D, Moazami-Goudarzi K (2010b) Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One*, **5**.

Gautier M, Naves M (2011) Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol Ecol*, **20**, 3128-3143.

Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM (2014) A molecular basis for classic blond hair color in Europeans. *Nat Genet*, **46**, 748-752.

Gunther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205-220.

Gutierrez-Gil B, Arranz JJ, Wiener P (2015) An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. *Front Genet*, **6**, 167.

Hancock AM, Witonsky DB, Alkorta-Aranburu G*, et al.* (2011) Adaptations to climate-mediated selective pressures in humans. *PLoS Genet*, **7**, e1001375.

Hancock AM, Witonsky DB, Gordon AS*, et al.* (2008) Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet*, **4**, e32.

Harris GR, Sexton DMH, Booth BBB, Collins M, J.M. M (2013) Probabilistic projections of transient climate change. *Climate Dynamics*, **40**, 2937-2972.

Hewitson B, Janetos AC, Carter TR*, et al.* (2014) Regional context. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects* (ed. Press CU), pp. 1133-1197, Cambridge, United Kingdom and New York, NY, USA.

Hijmans RJ, Van Etten J (2012) raster: Geographic analysis and modeling with raster data.

Hirano T, Matsuhashi T, Kobayashi N, Watanabe T, Sugimoto Y (2012) Identification of an FBN1 mutation in bovine Marfan syndrome-like disease. *Anim Genet*, **43**, 11-17.

Ho SY, Larson G, Edwards CJ*, et al.* (2008) Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biol Lett*, **4**, 370-374.

Hugh-Jones M, Blackburn J (2009) The ecology of Bacillus anthracis. *Mol Aspects Med*, **30**, 356-367.

Jeffreys H (1961) *Theory of probability*, third edition edn. Oxford University Press.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405.

Joost S, Bonin A, Bruford MW*, et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol Ecol*, **16**, 3955-3969.

Kriticos DJ, Webber BL, Leriche A*, et al.* (2012) CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, **3**, 53-64.

Liu S, Zhang Y, Hoover B, Leppla SH (2012) The receptors that mediate the direct lethality of anthrax toxin. *Toxins (Basel)*, **5**, 1-8.

Liu Y, Qin X, Song XZ*, et al.* (2009) Bos taurus genome assembly. *BMC Genomics*, **10**, 180.

Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P (1994) Evidence for two independent domestications of cattle. *Proc Natl Acad Sci U S A*, **91**, 2757-2761.

Lv FH, Agha S, Kantanen J*, et al.* (2014) Adaptations to climate-mediated selective pressures in sheep. *Mol Biol Evol*, **31**, 3324-3343.

MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos taurus and Bos indicus). *Genetics*, **146**, 1071-1086.

Matukumalli LK, Lawley CT, Schnabel RD*, et al.* (2009) Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*, **4**, e5350.

Müllner D (2013) fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. *Journal of Statistical Software*, **53**, 1-18.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289-290.

Park SD, Magee DA, McGettigan PA*, et al.* (2015) Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biol*, **16**, 234.

Patterson N, Moorjani P, Luo Y*, et al.* (2012) Ancient admixture in human history. *Genetics*, **192**, 1065-1093.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet*, **2**, e190.

Pausch H, Wang X, Jung S*, et al.* (2012) Identification of QTL for UV-protective eye area pigmentation in cattle by progeny phenotyping and genome-wide association analysis. *PLoS One*, **7**, e36346.

Payne WJA, Hodges J (1997) *Tropical Cattle: Origins, Breeds and Breeding Policies* Wiley-Blackwell.

Pebesma E, Bivand R (2005) Classes and methods for spatial data in R. *R News*, **5**.

Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, **11**, 1633-1644.

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*, **8**, e1002967.

Porter V (1991) *Cattle, A Handbook to the Breeds of the World* The Crowood Press Ltd, Ramsbury, Wiltshire.

Qanbari S, Pausch H, Jansen S*, et al.* (2014) Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet*, **10**, e1004148.

Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nat Rev Genet*, **14**, 807-820.

Schwartz M (2009) Dr. Jekyll and Mr. Hyde: a short history of anthrax. *Mol Aspects Med*, **30**, 347-355.

Segnalini M, Bernabucci U, Vitali A, Nardone A, Lacetera N (2013) Temperature humidity index scenarios in the Mediterranean basin. *Int J Biometeorol*, **57**, 451-458.

Sempéré G, Moazami-Goudarzi K, Eggen A*, et al.* (2015) WIDDE: a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC Genomics*, **16**, 940.

South A (2011) rworldmap: A New R package for Mapping Global Data. *The R Journal*, **3**, 35-43.

Stock F, Gifford-Gonzalez D (2013) Genetics and African cattle domestication. *Afr. Archaeol. Rev.*, **30**, 51-72.

Takasuga A (2016) PLAG1 and NCAPG-LCORL in livestock. *Anim Sci J*, **87**, 159-167.

Upadhyay MR, Chen W, Lenstra JA*, et al.* (2017) Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. *Heredity (Edinb)*, **119**, 469.

Velimirovic B (1984) Anthrax in Europe. *Rev. Sci. Tech. Off. int. Epiz.*, **3**, 527-559.

Vignes JD (1999) The large "true" Mediterranean islands as a model for the Holocen human impact on the European vertebrate fauna? Recent data and new reflections, 295-321.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **19**, 395-420.

Xiang Y, Laurent B, Hsu CH*, et al.* (2017) RNA m(6)A methylation regulates the ultraviolet-induced DNA damage response. *Nature*, **543**, 573-576.

Xu L, Bickhart DM, Cole JB*, et al.* (2015) Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol Biol Evol*, **32**, 711-725.

Data accessibility

SNP genotyping data are available in the WIDDE database (http://widde.toulouse.inra.fr/widde/) and in the Data Inra Dataverse portal (https://data.inra.fr/).

Author contributions

LF, KMG, MG and DL conceived the study. LF, KMG, MG, ST, VA, AA, IB, NB, FC, SC, RC, ACA, JVD, AE, GH, EJ, VL, AL, PL, CL, AM, SM, DM, CHM, MAO, OP, BP, CR, NSM, TS, GS and ST participated to sampling and/or laboratory work. LF and GS developed the database were genotyping data were stored. LF performed breed genetic characterization, identification of footprints of selection and association tests. DL performed analysis of climatic covariables. MG provided methodological support. LF drafted the manuscript with the input of KMG, DL and MG. All authors read and approved the final manuscript.

Figures

Figure 1: Cattle breed location and the geographic pattern of the annual mean values of the five main climatic variables

A. Location of the 62 breeds of the WORLD-Set, including the 21 breeds of the MED-Set. EUT, AFT, ZEB are indicated in red, blue, green, respectively, and the breeds located in the Mediterranean area (from the MED-Set), in orange. ANA and TUR hybrids breeds, not considered in the MED-Set are in black.

B. to F. Geographic pattern of the annual mean temperature (B.), the temperature annual range (C.), the annual precipitation (D.), the annual mean radiation (E.) and the annual mean moisture index (F.), with the location of Mediterranean breeds (MED-Set).

Figure 2: Results of the principal component analysis (PCA) and of the unsupervised clustering including 1704 individuals from 62 breeds genotyped for 33301 SNPs. (A) PCA results. Individuals are plotted on the first two principal components according to their coordinates. Ellipses characterize the dispersion of each breed around its center of gravity. EUT, AFT and ZEB breeds are plotted in red, blue and green, respectively, the breeds located in the Mediterranean area (Med-Set), in orange and the hybrids breeds ANA and TUR, not considered in the Med-Set, in black. (B) Unsupervised hierarchical clustering results with and inferred number of clusters k=3 with ADMIXTURE 1.04. For each individual the proportions of each cluster (*y* axis) which were interpreted as representative of EUT, AFT and ZEB ancestries are plotted in red, blue and green, respectively.

Figure 3: Inferred Mediterranean cattle tree. The structure of the graph inferred by *TreeMix* for the 21 Mediterranean cattle breeds of the MED-Set and for ANG, NDA and NEL breeds (representative of EUT, AFT and ZEB population groups, respectively), using NEL as outgroup and allowing 15 migration events, is plotted. Migration arrows are colored according to their weight. Population IDs are colored according to their geographic location (in orange for North-African breeds, black for South-French breeds, dark grey for Spanish breeds, turquoise for Italian breeds and purple for breeds from the east of the Mediterranean basin) or known ancestry (in blue, green and red for ANG, NDA and NEL, respectively).

Figure 4: Global gene network including genes associated with at least one climatic covariable.

Tables

Table 1: Description of cattle breeds of the WORLD and MED datasets and measure of inbreeding.

The breed ID, breed name, type (i.e. EUT, AFT, ZEB or HYB), land of origin, number of individuals, origin of the data, data set name and inbreeding measures (***Fis***) are indicated for each breed.

Breed code	**Breed Name**	**Type**	**Land of Origin**	**Nb of indivi-duals**	**Origin of the Data**	**Data set**	F_{is}
ABO	Abondance	EUT	France	22	*Gautier et al, 2010*	WORLD-Set	-0.0383
ANA	Anatolian[†]	HYB	Turkey	31	*Decker et al, 2014*	WORLD-Set	0.0349
ANG	Angus	EUT	Scotland	61	*Matukumalli et al, 2009*	WORLD-Set	-0.0028
AST	Asturiana	EUT	Spain	13	*this study*	WORLD-Set	0.0038
AUB	Aubrac	EUT	France (S)	22	*Gautier et al, 2010*	WORLD-Set	0.0113
AYR	Finnish Ayrshire	EUT	Scotland /Sweden	18	*Decker et al, 2014*	WORLD-Set	-0.0024
BAL	Baladi	HYB	Egypt	30	*this study*	WORLD-Set & MED-Set	0.03
BAO	Baoulé	AFT	Burkina Faso	29	*Gautier et al, 2009*	WORLD-Set	-0.012
BAT	Brune de l'Atlas	HYB	Morocco	20	*this study*	WORLD-Set & MED-Set	0.008
BIS	Biskra	EUT	Algeria	30	*Boushaba et al, 2018*	WORLD-Set	0.0237
BPN	Bretonne Black Pied	EUT	France	18	*Gautier et al, 2010*	WORLD-Set	-0.0115
BRM	Brahman	ZEB	USA (India)	25	*Matukumalli et al, 2009*	WORLD-Set	0.0265

BSW	Brown Swiss	EUT	Switzerland	24	*Matukumalli et al, 2009*	WORLD-Set	-0.0397
CHA	Charolais	EUT	France	20	*Gautier et al, 2010*	WORLD-Set	0.001
CHE	Cheurfa	HYB	Algeria	31	*Boushaba et al, 2018*	WORLD-Set & MED-Set	0.0176
CHF	Chelifienne	HYB	Algeria	30	*Boushaba et al, 2018*	WORLD-Set & MED-Set	0.0285
CHI	Chianina	EUT	Italy	9	*Decker et al, 2014*	WORLD-Set	0.0042
CIN	Cinisara	HYB	Italy	71	*Mastrangelo et al, 2014*	WORLD-Set & MED-Set	0.0308
COR	Corsican	EUT	France	33	*this study*	WORLD-Set & MED-Set	0.0128
CYP	Cyprus	HYB	Cyprus	9	*this study*	WORLD-Set & MED-Set	0.0259
GAS	Gascon	EUT	France	22	*Gautier et al, 2010*	WORLD-Set	0.0019
GBV	Gelbvieh	EUT	Germany	23	*Matukumalli et al, 2009 ; Decker et al, 2014*	WORLD-Set	-0.014
GIR	Gir	ZEB	India	24	*Matukumalli et al, 2009*	WORLD-Set	0.0075
GNS	Guernsey	EUT	Channel Islands	21	*Matukumalli et al, 2009*	WORLD-Set	0.0187
GRE	Brachykeratiki	HYB	Greece	19	*this study*	WORLD-Set & MED-Set	0.1155
GUE	Guelmoise	HYB	Algeria	30	*Boushaba et al, 2018*	WORLD-Set & MED-Set	0.0166
HFD	Hereford	EUT	England	31	*Matukumalli et al, 2009*	WORLD-Set	0.0617
JER	Jersey	EUT	Channel Islands	21	*Gautier et al, 2010*	WORLD-Set	0.0006
LAG	Lagune	AFT	Benin	44	*Gautier et al, 2009*	WORLD-Set	0.0393
LMS	Limousin	EUT	France	44	*Matukumalli et al, 2009*	WORLD-Set	0.0039
MAA	Marismena	EUT	Spain	22	*this study*	WORLD-Set & MED-Set	-0.029

MAL	Mallorquina	EUT	Spain	30	*this study*	WORLD-Set & MED-Set	-0.0282
MAM	Maremmana	EUT	Italy	24	*this study*	WORLD-Set & MED-Set	-0.0245
MAN	Maine-Anjou (Rouge des Près)	EUT	France	46	*Gautier et al, 2010*	WORLD-Set	-0.0135
MAR	Maraichine (Parthenaise)	EUT	France	19	*Gautier et al, 2010*	WORLD-Set	-0.0132
MEN	Menorquina	EUT	Spain	30	*this study*	WORLD-Set & MED-Set	-0.0357
MOD	Modicana	EUT	Italy	71	*Mastrangelo et al, 2014*	WORLD-Set & MED-Set	-0.0027
MON	Montbéliard	EUT	France	30	*Flori et al, 2009*	WORLD-Set	-0.0379
NAN	Negra Andaluza	EUT	Spain	32	*this study*	WORLD-Set & MED-Set	0.1112
NAV	Casta Navarra	EUT	Spain	30	*this study*	WORLD-Set	-0.0302
NDA	N'Dama	AFT	Guinea	25	*Matukumalli et al, 2009*	WORLD-Set	0.0078
NEL	Nelore	ZEB	India	21	*Matukumalli et al, 2009*	WORLD-Set	-0.0197
NOR	Normande	EUT	France	30	*Flori et al, 2009*	WORLD-Set	-0.0321
NRC	Norvegian Red Cattle	EUT	Norway	21	*Matukumalli et al, 2009*	WORLD-Set	-0.0067
OUL	Oulmes Zaër	HYB	Morocco	52	*Gautier et al, 2009; Boushaba et al, 2018*	WORLD-Set & MED-Set	-0.0015
PMT	Piemontese	EUT	Italy	24	*Matukumalli et al, 2009*	WORLD-Set	-0.0097
PRP	French Red Pied Lowland	EUT	France	22	*Gautier et al, 2010*	WORLD-Set	-0.0037
RDB	Raço di Biou	EUT	France	29	*this study*	WORLD-Set & MED-Set	-0.0284
RMG	Romagnola	EUT	Italy	24	*Matukumalli et al, 2009*	WORLD-Set & MED-Set	-0.0025
SAL	Salers	EUT	France	22	*Gautier et al, 2010*	WORLD-Set	0.0169

SAM	Sardo-Modicana	EUT	Italy	11	*this study*	WORLD-Set & MED-Set	0.00008
SAR	Sarda	EUT	Italy	12	*this study*	WORLD-Set & MED-Set	0.0334
SCH	Scottich Highland	EUT	Scottland	7	*Decker et al, 2014*	WORLD-Set	0.0096
SHO	Shorthorn‡	EUT	England	34	*Decker et al, 2014*	WORLD-Set	0.0455
SIM	Simmental	EUT	Switzerland	13	*Matukumalli et al, 2009 ; Decker et al, 2014*	WORLD-Set	0.022
SOM	Somba	AFT	Togo	44	*Gautier et al, 2009*	WORLD-Set	0.0406
TAR	Tarentaise	EUT	France	18	*Gautier et al, 2010*	WORLD-Set	-0.0089
TID	Tidili	HYB	Morocco	30	*Boushaba et al, 2018*	WORLD-Set & MED-Set	0.0338
TUR	Turkish Grey	HYB	Turkey	8	*Decker et al, 2014*	WORLD-Set	0.0011
VOS	Vosgienne	EUT	France	20	*Gautier et al, 2010*	WORLD-Set	-0.0136
ZFU	Zebu Fulani	ZEB	West-Africa	43	*Gautier et al, 2009*	WORLD-Set	0.0028
ZMA	Zebu from Madagascar	ZEB	Madagascar	35	*Gautier et al, 2009*	WORLD-Set	0.0127

† pool of four breeds: Anatolian Black, Anatolian Southern Yellow, East Anatolian red and South Anatolian red

‡ Pool of three breeds: Lincoln Red, Milking Shorthon, Beef Shorthorn

Table 2: Positions and candidate genes associated with climatic covariables.

The climatic covariables Bio01, 07, 12, 20 and 28, extracted from the Climond database, correspond to the annual mean temperature, temperature annual range, annual precipitation, annual mean radiation and annual mean moisture index, respectively. PC1 to PC4 correspond to the first four components and summarize 57.22%, 16.56%; 10.23% and 6.13% of the information respectively. Regions harboring footprints of selection based on the *XtX* measure of differentiation are indicated.

Covariable	BTA	Position (kb)	*BFmc*	RefGene	Gene symbol	Region (*XtX*)
PC1	7	47252135	21.30470640	NM_174485	*VDAC1*	#6
PC1	7	47252135	21.30470640	NM_001035299	*C7H5orf15*	#6
PC1	7	47349962	46.93735740	NM_001099186	*VDAC1*	#6
PC1	7	47384327	29.17102876	NM_001099186	*TCF7*	#6
PC1	7	47384327	29.17102876	NM_001034781	*SKP1*	#6
PC1	26	23461479	25.27788831	NM_001098083	*SUFU*	
PC2	1	117722957	25.08228767	NM_001205451	*GPR87*	
PC2	1	117722957	25.08228767	NM_001206505	*MED12L*	
PC2	3	67884386	21.80793474	NM_001192804	*ST6GALNAC5*	
PC2	3	103021431	22.04560307	NM_001101079	*PTPRF*	
PC2	7	21610660	20.93327016	NM_001077082	*C7H19orf71*	
PC2	7	21610660	20.93327016	NM_001037609	*FZR1*	

PC2	7	49104721	25.65825551	NM_001205402	*TGFBI*	
PC2	11	47916623	20.40842597	NM_001105344	*CD8B*	
PC2	11	49473033	52.96447989	NM_001102279	*RETSAT*	
PC2	11	49473033	52.96447989	NM_001015661	*ELMOD3*	
PC2	16	27164390	20.14747236	NM_001038161	*DISP1*	
PC2	16	65669824	52.96447989	NM_001206817	*LAMC1*	
PC2	16	74158269	52.96447989	NM_174372	*KCNH1*	
PC2	17	57251180	22.87576484	NM_001192597	*CUX2*	
PC2	19	18494850	22.49287196	NM_001192650	*CRLF3*	
PC2	19	18494850	22.49287196	NM_001205587	*SUZ12*	
PC2	19	20439349	20.30397301	NM_001079585	*SLC46A1*	
PC2	19	20439349	20.30397301	NM_001102518	*SARM1*	
PC2	23	24667121	35.36143428	NM_001193150	*TRAM2*	
PC3	3	63026914	21.82611597	NM_181013	*ADGRL2*	
PC3	5	31739928	21.23361942	NM_001076992	*C5H12orf54*	
PC4	13	35614532	24.47639126	NM_001099071	*MAP3K8*	
PC4	15	34342385	21.18041595	NM_001098043	*CLMP*	
PC4	21	30671618	21.05662152	NM_001206968	*OTUD7A*	

Bio1	2	90085321	30.74346170	NM_176667	*NDUFB3*	
Bio1	3	51399357	20.23439535	NM_001105200	*BTBD8*	
Bio1	6	96493695	24.16640507	NM_001076826	*ANTXR2*	#5
Bio1	7	47349962	46.93735740	NM_001099186	*TCF7*	#6
Bio1	7	47717578	28.86164065	NM_001035315	*SAR1B*	#6
Bio1	8	38476077	22.77906444	NM_001192951	*GLDC*	
Bio1	10	25732248	49.95200683	NM_001075707	*TOX4*	
Bio1	10	25732248	49.95200683	NM_001102238	*METTL3*	
Bio1	10	62038632	22.83700213	NM_174053	*FBN1*	
Bio1	19	36395869	23.92967373	NM_001076925	*SPAG9*	
Bio7	5	70175036	26.08635172	NM_001192789	*POLR3B*	
Bio7	11	49473033	52.96447989	NM_001102279	*RETSAT*	
Bio7	11	49473033	52.96447989	NM_001015661	*ELMOD3*	
Bio7	16	65669824	21.07427809	NM_001206817	*LAMC1*	
Bio7	16	74158269	52.96447989	NM_174372	*KCNH1*	
Bio7	23	24667121	37.32474177	NM_001193150	*TRAM2*	
Bio12	5	70338965	22.47398776	NM_001191182	*RFX4*	
Bio12	7	47252135	27.06173048	NM_174485	*VDAC1*	#6

Bio12	7	47252135	27.06173048	NM_001035299	*C7H5orf15*	#6
Bio12	7	47349962	42.14870329	NM_001099186	*TCF7*	#6
Bio12	7	47384327	29.64551211	NM_001099186	*TCF7*	#6
Bio12	7	47384327	29.64551211	NM_001034781	*SKP1*	#6
Bio12	16	74158269	21.46532469	NM_174372	*KCNH1*	
Bio12	20	24670287	26.00414634	NM_001046570	*HSPB3*	
Bio20	7	47252135	36.23801922	NM_174485	*VDAC1*	#6
Bio20	7	47252135	36.23801922	NM_001035299	*C7H5orf15*	#6
Bio20	7	47349962	42.14870329	NM_001099186	*TCF7*	#6
Bio20	7	47384327	29.17102876	NM_001099186	*TCF7*	#6
Bio20	7	47384327	29.17102876	NM_001034781	*SKP1*	#6
Bio20	22	5310074	29.49877704	NM_001102281	*GADL1*	
Bio20	23	24667121	23.17021341	NM_001193150	*TRAM2*	
Bio28	7	47349962	23.49230054	NM_001099186	*TCF7*	#6
Bio28	7	47384327	28.09802651	NM_001099186	*TCF7*	#6
Bio28	7	47384327	28.09802651	NM_001034781	*SKP1*	#6
Bio28	14	34236943	22.28638188	NM_001192526	*PREX2*	
Bio28	16	32341971	23.20999936	NM_001076406	*SMYD3*	

Bio28	16	74158269	28.17383675	NM_174372	*KCNH1*	
Bio28	20	24670287	26.75094601	NM_001046570	*HSPB3*	
Bio28	23	24667121	25.20398084	NM_001193150	*TRAM2*	
Bio28	27	27804403	20.60028489	NM_174128	*NRG1*	
THI	6	18332881	20.14747236	NM_001192856	*LEF1*	#3
THI	6	64487002	20.09533804	NM_001205725	*KCTD8*	
THI	6	96493695	23.74002296	NM_001076826	*ANTXR2*	#5
THI	7	47349962	42.94488271	NM_001099186	*TCF7*	#6
THI	7	47384327	29.12576293	NM_001099186	*TCF7*	#6
THI	7	47384327	29.12576293	NM_001034781	*SKP1*	#6
THI	7	47717578	20.25178607	NM_001035315	*SAR1B*	#6
THI	8	38476077	49.95200683	NM_001192951	*GLDC*	
THI	11	55229674	26.25366613	NM_001102110	*CTNNA2*	
THI	13	46841024	21.32250582	NM_001205573	*LARP4B*	
THI	22	16407075	25.76297041	NM_001206812	*ZKSCAN7*	
THI	22	16407075	25.76297041	NM_001205819	*ZNF445*	
THI	25	1737669	20.72269252	NM_001038172	*PGP*	
THI	25	1737669	20.72269252	NM_001035411	*MLST8*	

THI	25	1737669	20.72269252	NR_129529	*MIR1842*	
THI	25	1737669	20.72269252	NR_030807	*MIR2382*	

Table 3: Top five significant functions of the putative candidate genes associated with at least one climatic covariable, presented by functional category.

Functional Category	Name	p-value	Number of focus genes
Physiological System Development and Function	Nervous system development and function	4.93E-02 – 1.96E-05	15
	Tissue development	4.93E-02 – 1.96E-05	18
	Tissue morphology	4.65E-02 – 5.10E-04	7
	Organ morphology	4.65E-02 – 4.15E-03	13
	Organismal development	4.93E-02 – 4.15E-03	13
Molecular and Cellular Functions	Cell morphology	4.65E-02 – 1.96E-05	13
	Cellular assembly and organization	4.23E-02 – 1.96E-05	12
	Cellular development	4.91E-02 – 1.96E-05	19
	Cellular function and maintenance	4.65E-02 – 1.96E-05	12

	Cellular growth and proliferation	4.91E-02 – 1.96E-05	15
Diseases and Disorders	Cancer	4.96E-02 - 4.22E-04	43
	Gastrointestinal disease	4.96E-02 - 4.22E-04	38
	Organismal injury and abnormalities	4.96E-02 - 4.22E-04	43
	Tumor morphology	2.93E-02 – 2.19E-03	4
	Hematological disease	4.93E-02 – 2.88E-03	21

Table 4: Regions harboring footprints of selection based on the *XtX* measure of differentiation. For each region, the peak *XtX* value, its position in Mb, the putative candidate genes and the candidate functions determined by association test using BayPass are indicated.

Region	BTA	Start-End in Mb	Peak position (Mb)	*XtX* value at the peak position	Candidate genes	Associated phenotypes/ covariables
#1	5	29.5-30.5	30.06	43.24	Rac GTPase activating protein 1 (*RACGAP1),* aquaporin 6 (*AQP6)*	
#2	5	105.5-107	106.26	48.31	TP53 induced glycolysis regulatory phosphatase *(TIGAR),* cyclin D2 (*CCND2)*	
#3	6	17-19	17.9	54.72	ribosomal protein L34 (*RPL34),* lymphoid enhancer binding factor *1 (LEF1)*	Climate†
#4	6	38-40.5	39.0	51.78	leucine aminopeptidase 3 (*LAP3),* non-SMC condensin I complex subunit G *(NCAPG),* ligand dependent nuclear receptor corepressor like (*LCORL)*	Morphology‡
#5	6	95.5-97.	96.4	46.91	anthrax toxin receptor 2 *(ANTXR2)*	Climate†
#6	7	46.5-48	47.2	68.34	voltage dependent anion channel 1 *(VDAC1),* transcription factor 7 *(TCF7),* S-phase kinase associated protein 1 *(SKP1)*	Climate†
#7	10	32.5-34	33.0	45.97	*NA*	
#8	12	29-30.	29.7	43.62	beta 3-glucosyltransferase *(B3GLCT),* relaxin family peptide receptor 2 *(RXFP2)*	Horn development

#9	18	13.5-15	13.8	56.31	zinc finger protein, FOG family member 1 *(ZFPM1),* melanocortin 1 receptor *(MC1R)*	Pigmentation[‡]

[†]this study

[‡](Gautier, 2015)

Supplementary Information

Figure S1: Unsupervised hierarchical clustering results with an inferred number of clusters k=2 to 10 (A. to I.), 15 (J.), 20 (K), 30 (L.) and 40 (M.) using ADMIXTURE 1.04.

Figure S2: Cross-validation plot obtained using the ADMIXTURE procedure (5-fold CV).

Figure S3: Neighbor-Joining tree relating the 1,704 individuals

The tree was constructed using allele sharing distances averaged over 39,921 SNPs. Edges are colored according to the individual breed of origin.

Figure S4: Hierarchical clustering (UPGMA) of the pairwise *Fst* estimated between breeds of the WORLD-SET.

Figure S5: Cattle tree inferred with *TreeMix* using the MED-Set.

A. Maximum likelihood tree without migration events. Population IDs are colored according to their geographic location (in orange for North-African breeds, black for South-French breeds, dark grey for Spanish breeds, turquoise for Italian breeds and purple for breeds from the east of the Mediterranean basin) or known ancestry (in blue, green and red for ANG, NDA and NEL, respectively).

B. Plot of the residual fit from the maximum likelihood tree.

C. Plot of the percentage of variance in relatedness between breeds explained by the inferred graphs with a number of migration events which ranges from 0 to 20.

Figure S6: Results of the principal component analysis (PCA) of the climatic covariables

A. Projection on a map of the colorplot synthesizing the breed coordinates on the three first PCA principal components. This plot can show up to three coordinates at the same time by recoding each coordinate into intensities of a given color channel of the RGB system (i.e. Red, Green and Blue for the first, second and third PC, respectively). Each color-recoded population score is then plotted onto a map using population geographical coordinates.

B. Correlation circle describing the importance of each bioclimatic variable along the first two PCA axes. Variables are colored according to their group (red for temperature, green for precipitation, turquoise for moisture and blue for radiation). Longitude and Latitude are projected as supplementary variables on the circle.

Figure S7: Representation of the scaled covariance matrix $\hat{\Omega}$ among the 21 cattle breed of the MED-Set as estimated from BayPass under the core model with $\rho = 1$.

$\hat{\Omega}$ is represented inthe form of a correlation plot (A) and of a hierarchical clustering tree (B).

Figure S8: Extent of Linkage Disequilibrium, measured by r^2 computed with *plink* 1.9 (Purcell et al., 2007), across and within the MED-Set populations.

Across-populations r^2 were computed following Gautier et al. (2007) on artificially constructed composite populations of 21 individuals (1 individual randomly drawn per population). To limit sampling biases, results from 100 samples were averaged.

Figure S9: Plots over the genome of the *XtX* (A) and *BFmc* estimates analyzing component 1 (B), 2 (C), 3 (D), 4 (E), THI (F), Bio01 (G), Bio07 (H), Bio12 (I), Bio20 (J), Bio28 (K).

The threshold of 38.67 determined using PODS (corresponding to 0,1% of the PODS) are indicated in dotted line on the *XtX* plot (A). Threshold corresponding to BF_{mc} value of 20 (decisive evidence) is indicated in dotted lines on each BF_{mc} plot (B to K).

Table S1: Proportion of cluster 1, 2 and 3, interpreted as representative of EUT, ZEB and AFT ancestries, respectively, for each individual belonging to the WORLD-SET cattle populations.

Table S2: Paiwise Fst between WORLD-Set cattle populations.

Table S3: Results of formal tests of admixture in Mediterranean cattle breeds using the f3 statistics (Patterson *et al.*, 2012).

Table S4: List of GPS coordinates, values of 35 climatic covariables extracted from Climond database (https://www.climond**.org, (Kriticos *et al.*, 2012)), coordinates of each breed from the first four axes of a PCA including the 35 climatic covariates and values of THI for the 21 breeds of the MED-Set.**

Table S5: List of the 35 climatic covariables extracted from the Climond database (https://www.climond**.org) and correlations of these variables with the first four axes of the PCA.**

Table S6: *BFmc* values calculated for each SNP position and for each of the 10 climatic covariables considered under the AUX model implemented in BayPass (Gautier, 2015). For each SNP, are indicated the chromosome number, the position (in bp), the Refseq and the symbol of the gene, the boundaries (+/-15kb) of which include the SNP position.

Table S7: Significant functions of the candidate genes associated with at least one climatic covariable.

Table S8: Significant gene networks including genes associated with at least one climatic covariable. The complete list of gene participating to each network, the network score, the number of genes associated with climatic covariables (focus genes) and the top three diseases and functions in which gene networks are involved, are indicated. Results were obtained using the *Ingenuity Pathway Analysis* software (Ingenuity®Systems, www.ingenuity.com).

Table S9: List of the molecules included in the global network. For each molecule, the symbol, the Entrez gene name, the association with climatic variables, the network ID, the cellular location and the molecule type are indicated**.** Results were obtained using the *Ingenuity Pathway Analysis* software (Ingenuity®Systems, www.ingenuity.com).