

RESEARCH

Open Access



# Landscape genomics reveals regions associated with adaptive phenotypic and genetic variation in Ethiopian indigenous chickens

Fasil Getachew Kebede<sup>1,2\*</sup> , Martijn F.L. Derks<sup>1</sup>, Taddelle Dessie<sup>2</sup>, Olivier Hanotte<sup>2,3</sup>, Carolina Pita Barros<sup>1</sup>, Richard P.M.A. Crooijmans<sup>1</sup>, Hans Komen<sup>1</sup> and John W.M. Bastiaansen<sup>1</sup>

## Abstract

Climate change is a threat to sustainable livestock production and livelihoods in the tropics. It has adverse impacts on feed and water availability, disease prevalence, production, environmental temperature, and biodiversity. Unravelling the drivers of local adaptation and understanding the underlying genetic variation in random mating indigenous livestock populations informs the design of genetic improvement programmes that aim to increase productivity and resilience. In the present study, we combined environmental, genomic, and phenotypic information of Ethiopian indigenous chickens to investigate their environmental adaptability. Through a hybrid sampling strategy, we captured wide biological and ecological variabilities across the country. Our environmental dataset comprised mean values of 34 climatic, vegetation and soil variables collected over a thirty-year period for 260 geolocations. Our biological dataset included whole genome sequences and quantitative measurements (on eight traits) from 513 individuals, representing 26 chicken populations spread along 4 elevational gradients (6–7 populations per gradient). We performed signatures of selection analyses ( $F_{ST}$  and XP-EHH) to detect footprints of natural selection, and redundancy analyses (RDA) to determine genotype-environment and genotype-phenotype-associations. RDA identified 1909 outlier SNPs linked with six environmental predictors, which have the highest contributions as ecological drivers of adaptive phenotypic variation. The same method detected 2430 outlier SNPs that are associated with five traits. A large overlap has been observed between signatures of selection identified by  $F_{ST}$  and XP-EHH showing that both methods target similar selective sweep regions. Average genetic differences measured by  $F_{ST}$  are low between gradients, but XP-EHH signals are the strongest between agroecologies. Genes in the calcium signalling pathway, those associated with the hypoxia-inducible factor (HIF) transcription factors, and sports performance (*GALNTL6*) are under selection in high-altitude populations. Our study underscores the relevance of landscape genomics as a powerful interdisciplinary approach to dissect adaptive phenotypic and genetic variation in random mating indigenous livestock populations.

\*Correspondence:

Fasil Getachew Kebede  
fasilgetachew7@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Keywords** Local adaptation, Poultry production, Signatures of selection, Redundancy analysis, Environmental predictors, Quantitative traits, Genetic improvement

## Background

The genetics of local adaptation and climate resilience in livestock has become more relevant in the face of climate change [1–4]. Local adaptation refers to the response of individuals to differential selective pressure leading to higher genetic fitness in their environment than individuals from elsewhere [5, 6]. Resilient animals have the capacity to be minimally affected by environmental disturbances (e.g., temperature stress, disease pressure, introduction to a new habitat) if they occur, or can return rapidly to the state pertained before exposure to the disturbance [7, 8]. Animals that combine high production potential with resilience to external stressors in a wide variety of environmental conditions are regarded as ‘robust’ [9].

Phenotypic differentiation represents the fraction of phenotypic variance between populations over the total phenotypic variance and helps understand evolutionary processes shaping populations [10–12]. Environmental differences acting as a natural selective force can result in exceptionally strong genetic differentiation in genomic regions containing loci subjected to selection [13]. Phenotypic and genetic differentiation along environmental gradients, or across contrasting habitat types, can be indicative of local adaptation [6, 14, 15]. For instance, alleles providing adaptation to high elevation are found in high frequency in populations at high elevation but in low frequency in populations at low elevation in humans [16, 17], in chickens [16, 18, 19], in pigs [20, 21], and small ruminants [22].

Understanding the genetic basis of phenotypic variation and local adaptation in livestock in response to environmental variation helps to enhance productivity and mitigate climate change [2, 23–25]. Randomly mating indigenous livestock populations are raised in stressful environmental conditions for many generations and harbour genomic regions conferring local adaptation that need to be exploited [26, 27]. Once identified, beneficial alleles/variants in indigenous chickens can be introduced into commercial chickens through breeding programmes [28, 29] or genome editing [30–32] to develop animals with desirable phenotypic attributes.

The effects of environmental selective pressures as drivers of local adaptation and specially their influences on phenotypic and genetic differentiation in Ethiopian chicken populations have not been investigated enough to shape our understanding of environmental adaptation. Ethiopian indigenous chickens also called local, village, scavenging, backyard, or family chickens are widely adapted, random mating, nondescript domesticated

chicken populations. They are managed in extensive (low-input) systems in their natural environment, without selective breeding programmes in place [33]. Ethiopia has one of the earliest evidences for chicken domestication and dispersal in Africa [34]. Ethiopian chickens are distributed in all agroecologies [33] and show substantial phenotypic and genetic diversity [35–39]. Large genetic diversity of present-day Ethiopian chicken populations might be attributed to their multiple waves of introduction into the country [40, 41] and the presence of highly diverse environment (e.g., climate, vegetation, elevation) [42]. As such, the country can be considered an ideal place for studying adaptive phenotypic and genetic variation in chickens.

Certain phenotypes in Ethiopian indigenous chickens (e.g., comb shape, parasitic resistance) are related with local adaptation [26]. Genomic regions conferring adaptation to environmental challenges (e.g., elevation, temperature, water scarcity, and feed availability) have been identified among African indigenous chickens [43–45]. Important insights on local adaptation of Ethiopian indigenous chickens were obtained in previous studies: the association between environmental predictors and phenotypic differentiation in quantitative traits was reported as an evidence for adaptive variation [46]. Another important study claimed that environmental conditions may have driven genomic variation in indigenous chicken populations [47].

At an interface between ecology and population genetics, landscape genomics provides an analytical framework useful to investigate the underlying evolutionary processes behind phenotypic and genetic differentiation of random mating populations raised in heterogeneous environments. Landscape genomics seeks to understand the influences of geographic and environmental features on selectively neutral and adaptive loci, and underlying micro-evolutionary processes such as gene flow, selection, and genetic drift [48, 49]. Landscape genomic approaches were followed in studies of adaptive genetic variation in different farm animal species animals including chickens, sheep, goats, swine, and cattle [45, 50–57].

Popular tools being used in landscape genomic studies include species distribution modelling, signatures of selection analyses, and genotype-environment analyses. Multivariate methods that simultaneously account for multiple drivers of phenotypic and environmental divergence, are recently being applied in landscape genomic studies to identify quantitative trait loci (QTL) associated with environment predictors [58–61] and with phenotypic variables [60, 62–65].

Species distribution models (SDMs), also known as environmental (ecological) niche models (ENMs) or habitat distribution models [66], use computer algorithms to analyse environmental data and to predict the distribution of a species across geographic space and time. SDMs are a popular tool in quantitative ecology because of their low data requirement, availability of many software packages and guidelines, and their higher predictive abilities [67, 68]. The central concept in SDMs is the niche theory [69, 70], which delineates the environment into fundamental and realized niches. In recent years, the conceptual framework for SDMs has been extended by livestock scientists and used to identify environmental predictors associated with habitat suitability and local adaptation [45, 46, 56, 71, 72].

Signatures of selection analysis are useful to identify regions of the genome that have differentiated between populations, possibly in response to selective pressure [73, 74]. Positive selection leaves conspicuous footprints or selective sweeps on the genome that can be detected using several approaches ranging from summary statistics such as Tajima's D, to maximum likelihood and machine learning [75]. Cross-population Extended Haplotype Homozygosity (XP-EHH) detects differential selection between two populations [76]. Pairwise comparison of fixation index ( $F_{ST}$ ) reveals differentiation of populations in different environments due to differences in evolutionary history [77].  $F_{ST}$  and XP-EHH approaches are complementary to each other and lead to a more comprehensive understanding of signatures of selection.  $F_{ST}$  is more suited for detection of positive selection in the distant past [78] while XP-EHH is more useful for detection of entirely or approximately fixed loci [76].

Another statistical method that is being used in landscape genomics is Redundancy Analysis (RDA). RDA is useful to investigate association between genomic and environmental variability. RDA combines regression and principal component analysis (PCA) and it is an extremely powerful tool for ecologists to model multivariate response data [79, 80]. RDA determines how groups of loci covary in response to the multivariate environment, and can better detect processes that result in weak, multilocus molecular signatures relative to univariate tests [81]. It accounts for population structures, demographic histories, and polygenic interactions [59, 82]. Multivariate methods like RDA, that simultaneously account for multiple drivers of phenotypic and environmental divergence are being used to identify quantitative trait loci (QTL) associated with environment predictors [58–61]. Multivariate ordination methods such as RDA have outperformed mixed-model-based methods and machine learning-based methods (e.g., Random Forest) in detecting loci associated with environmental variation

[59, 82]. Despite its ability to investigate genotype-phenotype associations RDA is mostly neglected in GWAS studies in livestock, while it became a standard in genotype-environment association studies in wildlife [62, 83]. In the present study we follow a landscape genomic approach to dissect adaptive genetic and phenotypic variation in Ethiopian indigenous chickens. We use SDMs to identify the most relevant environmental predictors driving local adaptation and produce habitat suitability maps. We perform signatures of selection analyses ( $F_{ST}$  and XP-EHH) to detect genetic differentiation between populations and selective sweeps. RDA is applied to identify outlier SNPs associated with environmental and phenotypic variation. Together, we explain variations in the genome by using key environmental drivers and identify candidate genes and genomic regions linked with environmental adaptation in Ethiopian indigenous chickens.

## Materials and methods

### Sampling strategy

Sampling design is a fundamental aspect of landscape genomic studies. As such, we implemented a robust sampling strategy, considering environmental gradation (e.g., elevational clines) and geographic (latitudinal and longitudinal) variation in the country [46] to avoid biases in discovery of genomic regions under selection. A hybrid strategy combines maximization of geographic distance (based on coordinates) and climatic distance between chosen sites. The landscape is divided into distinct environmental regions before choosing sites within each region that maximizes spatial distance [84]. A hybrid sampling strategy ensures environmental and geographic representativeness of sampling sites and increases statistical power by reducing false discovery rates of statistically significant loci in signatures of selections analysis and genome wide association studies [84–86]. The strategy also prevents the sampling of neighbouring sites with similar conditions and avoids the superposition between adaptive and neutral genetic variation [87].

The spatial distribution of our samples considered environmental (e.g., geography, climate) and biotic processes (e.g., domestication, routes of introduction) influencing the chicken populations. A total of 513 chickens were sampled from four environmental gradients (*gradient-I, -II, -III, and -IV*) with a minimum distance between gradients of 500 km. Environmental or ecological gradients refer to gradual changes in abiotic environmental factors (such as elevation, temperature, soil, vegetation, and precipitation) with consequences on the species' distribution and local adaptation [88]. *Gradient-I* stretches from the Rift valley lowlands of northeastern Ethiopia along the territories of Afar region to the highlands of Wollo province within Amhara region. *Gradient-II*, starts from the Rift valley lowlands in central Ethiopia, crosses the

highlands of Hararghe, including Mount Gara Muleta, and stretches to eastern Ethiopia within Oromia region. *Gradient-III* stretches from the highlands of northwestern Ethiopia and goes down to the lowlands along the Ethiopian-Sudanese border within Benishangul-Gumuz region. *Gradient-IV* extends from the highlands of western Ethiopia in Oromia region to the lowlands along the Ethiopian-Kenyan border in Southern region. Areas around the national borders of Ethiopia have low elevation, which gradually culminates to highland plateau in the center of the country creating a striking contrast in agroecology.

Each gradient comprised three environmental clusters or agroecologies, primarily delineated based on elevation in meters above sea level (m.a.s.l.). These are *lowland* (400–1800 m.a.s.l.); *midaltitude/midland* (1800–2400 m.a.s.l.); and *highland* (2400–3500 m.a.s.l.) according to the conventional agroecological classification in Ethiopia [89, 90]. Clusters within a gradient were distant by at least 100 km and farmers keeping target chicken populations within a cluster visited separate livestock markets. Each cluster along the spatial gradient constituted of 2–3 populations. In the context of the present study, a population of Ethiopian indigenous chickens refers to individuals that are kept within a specific geographic area (at *village* level) and which are assumed to be similarly influenced by environmental (ecological) and socio-economic factors. A village (*kebele*) is the smallest administrative unit in Ethiopia. The metadata of 513 individual samples representing 26 chicken populations is presented in Supplementary Table 1. The topographic map of Ethiopia showing the Ethiopian indigenous sample populations and their environmental gradients is presented in Fig. 1.

The chicken populations from different geographies of Ethiopia may be the result of different evolutionary histories. We controlled for the potential confounding effects between demographic processes (e.g., domestication history, migration) and adaptive variation in our analysis by performing signatures of selection analyses at three different analytical layers (*layer-I*, *layer-II*, and *layer-III*). Figure 2 shows the sampling and analytical framework used in the present study.

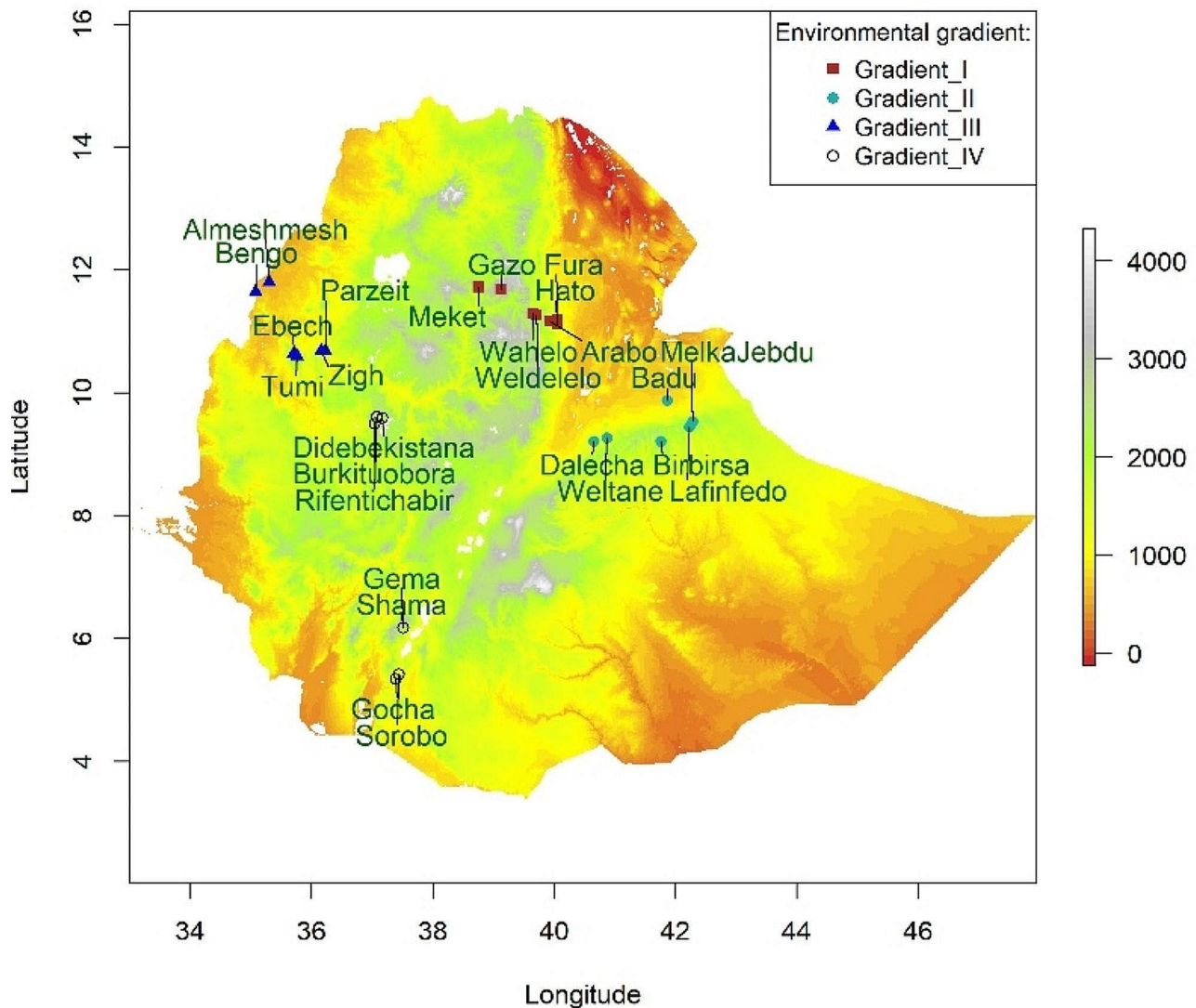
### Environmental data

For every population, a single geographic coordinate was taken at the center of the village during sampling of chickens. Coordinates from nine additional grids (1.44 km<sup>2</sup>), covering a total of 12.96 km<sup>2</sup>, were then drawn around a recorded location and extracted using Google Earth Pro v 7.3.2 to ensure high representation of environmental variability affecting the population. The total number of ‘presence’ or ‘occurrence’ points used in SDMs for the 26 sample populations comprised 260 coordinates. Out

of 34 environmental predictors, 9 predictors identified through SDMs for their association with habitat suitability and adaptive evolution of chickens in Ethiopia [46] were included in the present study for genotype-environment association analysis with RDA. The 9 predictors are isothermality, temperature seasonality, mean temperature of the coldest quarter, precipitation of the warmest quarter, precipitation of the coldest quarter, solar radiation of the month of May, water vapour pressure of the month of May, water vapour pressure of the month of August, and soil clay content (Supplementary Table 2). Values for bioclimatic variables (temperature, precipitation, soil radiation, and water vapour pressure) in different seasons were obtained from WorldClim database (<http://www.worldclim.org/>; version 2) at a spatial resolution of 30 s (~1 km<sup>2</sup>) [91] based on mean values of 30 years (1970–2000). Additionally, considering the importance of elevation in the conventional definition of agroecologies in Ethiopia [33], its link with certain adaptive traits in chickens [92], and our sampling design that takes into account elevational clines, we incorporated elevation as a tenth environmental predictor. All the ten predictors were used to produce habitat suitability maps for the 26 sample chicken populations with MaxEnt computer algorithm (version 3.4.1) [93]. Configuration of model parameters for MaxEnt was set based on a previous study [46].

### Quantitative trait data

Collection of phenotypic data was performed on adult chickens (about 20 chickens sampled from each of the 26 villages). These chickens were selected randomly by walking along a defined path (transect) across an administrative village and sampling one chicken from each farming household until a total of 15 hens and 5 cocks (roosters) were measured. The age of the chickens was estimated by interviewing owners to confirm that females were in their second clutch (7 to 8 months-of-age) and males were above 12 months-of-age. The researchers also visually appraised roosters for the presence of well-developed spurs. To minimize the risk of inbreeding, one chicken was sampled per household. Under rare circumstances ( $n=9$ ), two chickens were sampled per household when farmers expressed that their animals have no family relationship, for instance when they were obtained from different sources (e.g., one is hatched at home while the other was bought from the market). 19 quantitative traits were initially measured on each of the 513 adult chickens. Out of these 19 quantitative traits, we used the five traits identified by [46] for their putative roles in local adaptation and usefulness in phenotypic classification of Ethiopian chicken populations (Supplementary Table 3). These are live body weight, beak length, comb width, wattle width, and earlobe width. Live body weight (total mass of an individual in grams before slaughter) was



**Fig. 1** Topographic map of Ethiopia depicting the 26 Ethiopian indigenous chicken sample populations and their environmental gradients. Range of numbers with different colours in the legend indicate elevation (m.a.s.l.)

taken using a digital scale in the morning when the animal was fasting i.e., before it was released to scavenge in the backyard. The phenotypic measurements for the other traits were read from pictures of individual chickens and analysed using ImageJ software (version 1.52a) [94]. To reduce systematic error, the same operator measured all chickens, which were held in the same position by a technician. A steel ruler was placed in every picture as a distance reference.

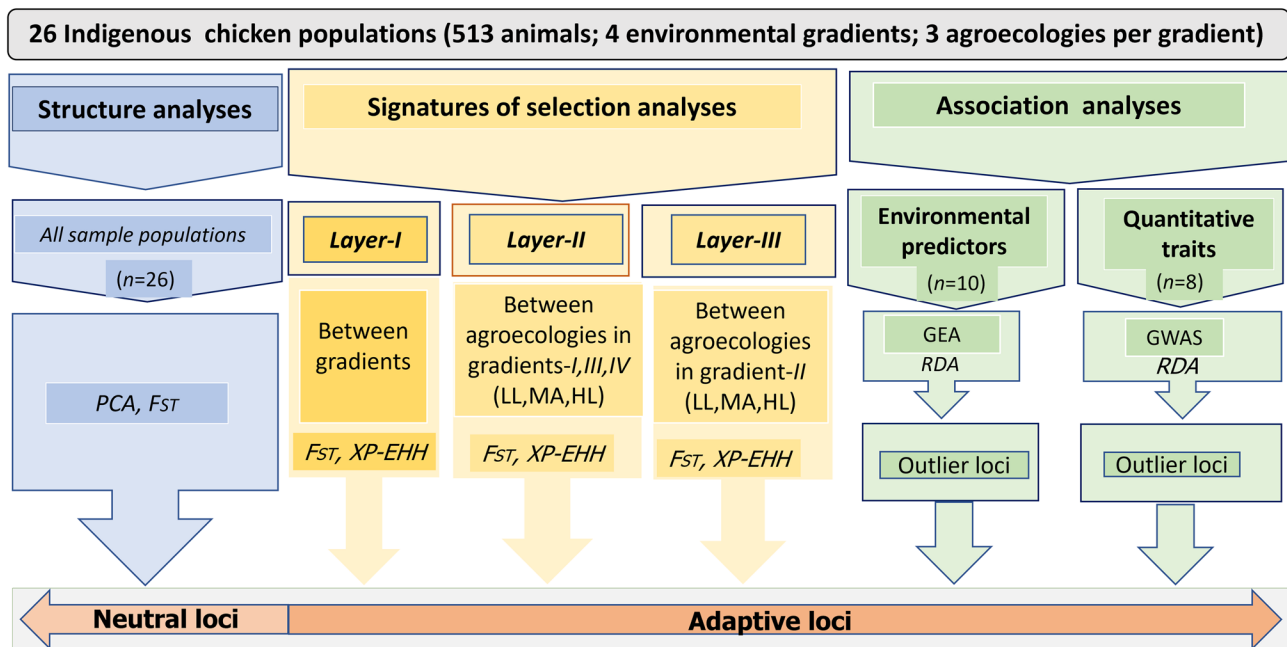
#### Blood sampling

Whole blood samples were taken from the wing vein of individual chickens in line with standard procedures [95].

A volume of 50–250  $\mu$ l of whole blood with anticoagulant (K2EDTA) per sample was put into a cryo-tube filled with 1.5 ml absolute ethanol (100%). Samples were preserved at  $-20^{\circ}\text{C}$  until DNA extraction and processing.

#### Whole genome sequence and data processing

WGS data was generated on Illumina HiSeq2000 platform in paired-end mode with a read length of 150 bp. Reads were quality trimmed using Trimmomatic (Version 0.39) (*ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36*) [96]. The average depth of coverage was 8.63 (range: 5.47–14.12) with an average mapping rate of



**Fig. 2** Sampling and analytical framework in landscape genomics study to detect adaptive phenotypic and genetic variation in Ethiopian indigenous chicken populations LL=lowland; MA=midaltitude; and HL=highland. Adaptive loci are the result of natural selection and contribute to fitness while neutral loci are due to other evolutionary process (e.g., gene flow, genetic drift, demographic history)

99.2% (97.05–99.6) and a mapping quality of 33.6 (28.77–34.45) to the GRCg6a reference assembly (Ensemble *Gallus\_gallus.GRCg6a.dna.toplevel.fa*). Genomic analysis was performed on autosomes and non-autosomes.

Freebayes variant calling was run on processed bam files (alignment data) to generate VCF file with the following setting: min-base-quality 10 --min-alternate-fraction 0.2 --haplotype-length 0 --ploidy 2 --min-alternate-count 2 [97]. The ‘min-base-quality 10’ specifies the minimum base quality required for a base to be considered during variant calling. The ‘min-alternate-fraction 0.2’ sets the minimum fraction of reads to 20% supporting the alternate allele for a variant to be called. The ‘haplotype-length 0’ disables the haplotype extension feature, ensuring that all reads are considered independently during variant calling. The ‘ploidy 2’ defines the ploidy of the organism being analysed, with a value of 2 indicating diploid. The ‘min-alternate-count 2’ establishes a minimum number of 2 observations supporting an alternate allele required for a variant to be called.

Post processing was performed in BCFtools [98] using vcf-filter module of with the setting ‘-f ‘QUAL>20’. SNPs with low phred quality score (<20), low call rate (<0.7), and those within 3 bp of an insertion-deletion (indel) were discarded. Filtering of genotypes was performed [99] prior to downstream analyses to improve data quality. Genotypes were filtered for SNPs not in Hardy-Weinberg equilibrium ( $p < 5 \times 10^{-6}$ ), with minor allele frequency (MAF) <5%. After applying stringent quality filtration on genomes of 513 individuals, we used a clean

dataset of 25 M (autosomal and non-autosomal SNPs) from 466 individuals for all downstream analyses (see **Materials and Methods**). Information on genome coverage, mapping rate and quality of samples is presented in Supplementary Table 4.

**Population structure analysis**

PCA was performed using the Eigenstrat method, with the smartpca function from Eigensoft v 6.1.4 software [100, 101] to understand the structure of the 26 populations. The VCF files containing the found variants were converted to the eigenstrat format with a python script ([https://github.com/CarolinaPB/Bioinfo\\_scripts/blob/main/vcf2eigenstrat.py](https://github.com/CarolinaPB/Bioinfo_scripts/blob/main/vcf2eigenstrat.py)).

**Signatures of selection analysis (SSA)**

The search for signals of positive selection ( $F_{ST}$  or XP-EHH) was carried out on SNP data ( $n=25$  M) that were obtained from WGS after stringent quality filtering. Haplotypes were phased using FastPhase software [102] prior to signatures of selection analyses. To identify candidate loci and genomic regions linked with local adaptation, we performed signatures of selection analyses ( $F_{ST}$  and XP-EHH) at three different analytical layers (Fig. 2).

In *layer-I*, we classified the indigenous chicken populations into four gradients (without regard to their agroecologies) and analysed them to detect genetic differentiation between them. The populations across *gradients (-I, -III, and -IV)* were then categorized by agroecology and analysed (lowland, midaltitude, and

highland) in *layer-II*. Chickens have a complex history of introduction and dispersal in Africa at large and in Ethiopia in particular through multiple maritime and/or terrestrial routes [40]. Considering the geographic closeness of *gradient-II* to the Arabian Peninsula, we analysed populations from this gradient (*layer-III*) separately by agroecology (lowland, midaltitude, and highland) to account for possible differences in evolutionary processes from the other three gradients.

Fixation test ( $F_{ST}$ ) analysis was conducted using VCFtools v0.1.16, [103], to identify regions of increased genomic differentiation between the classifications defined in the analytical layers.  $F_{ST}$  value of 0 indicates no differentiation between populations while a value of 1 indicates complete differentiation. Previous works in diverse species including chicken suggest that sliding-window analyses between 20 kb (with 10 kb overlap) and 400 kb (with 200 kb overlap) have considerable power to detect changes in allele frequencies and genomic regions with significant divergence between populations in signatures of selection analysis [104, 105].

We calculated the average  $F_{ST}$  values with overlapping windows of 50 kb (25 kb overlapping). We calculated the average XP-EHH values for the classifications defined in each of the analytical layers. Analysis of genomic regions with signs of recent positive selection with XP-EHH was based on the concept of *extended haplotype homozygosity* (EHH) [76, 106] and was performed on phased haplotypes using R package *rehh* (Version 3.2.2) [107]. First, the *data2haplohh* function was used to convert the VCF files to a suitable format to be used to compute XP-EHH. Then, XP-EHH was calculated with the *ies2xpehh* function from the same package.

The same size of overlapping bins (50 kb) was used for XP-EHH analysis to allow comparison with  $F_{ST}$ . First, the average ( $F_{ST}$  or XP-EHH) values for all bins in each pairwise comparison in an analytical layer were sorted on their significance. Empirical  $p$ -values were calculated for both  $F_{ST}$  and XP-EHH by ranking the windows based on each metric and dividing the rank by the total number of windows. The same approach was used by a previous study on Ethiopian chickens [45]. Only the 1% most significant windows ( $p < 0.01$   $F_{ST}$  or XP-EHH) were retained as significant. Significant windows which were commonly identified by the two methods were counted as overlapping.

#### Pathway enrichment analysis

We used ShinyGO with chicken as background to perform pathway enrichment analysis and identify genes that are under selection in specific agroecologies [108].

#### Association analyses

Association analyses in RDA were performed with the R package ‘vegan’ using RDA function [109] to identify environmental predictors and quantitative traits associated with genomic variation. RDA is a multivariate multiple regression followed by a PCA of the table of fitted values and presents relationships between variables in two-dimensional space using ordination plots [80, 110, 111]. Environmental predictors and quantitative traits were analysed separately according to [112].

RDA was performed using a set of genome-wide LD-pruned SNPs to keep a subset of SNPs that are nearly uncorrelated with each other and keep a subset of markers that are in approximate linkage equilibrium. Pruning of genotypes for high LD reduces redundant loci and improves efficiency of models in association analysis [113]. The cleaned dataset with 25 M SNPs filtered for SNPs not in Hardy-Weinberg equilibrium ( $p < 5 \times 10^{-6}$ ) and MAF < 5%, was LD pruned using PLINK [114]. We used the following setting for pruning: `plink2 --vcf --set-all-var-ids @:# --chr-set 38 --allow-extra-chr --indep-pairwise 100 10 0.5 --maf 0.05 --recode vcf-iid --out --indep-pairwise`. The ‘100 10 0.5’ instructs PLINK to perform LD pruning by evaluating LD in sliding windows of 100 variants, removing variants within each window if more than 10 are correlated, and considering variants to be correlated if their LD correlation ( $r^2$ ) exceeds 0.5. The ‘--recode vcf-iid’ modifier produces sample IDs in the last header row of VCF file. The LD pruning resulted in a subset of markers comprising 1,070,305 SNPs from 466 individuals, which was large enough for RDA. The dataset was structured as a matrix of 466 chickens by ~1 million SNP markers.

#### Genotype-environment association (GEA) analysis with RDA

Correlated predictors cause problems for regression-based models like RDA and variable reduction was done when correlation coefficients between ecological predictors exceeded an acceptable threshold ( $r > |0.7|$ ) [115]. We fitted partial RDA with the 10 selected environmental predictors conditioned on (i.e., controlling for the effects of) geography as explanatory variables and the genetic dataset as response variable [81]. SNPs exhibiting RDA loadings greater than 3.5 standard deviations (two-tailed  $p$ -value = 0.0005) from the mean were identified as selection signals. This threshold is very conservative and helps to identify loci under strong selection (i.e., minimizes false positive rates) [59]. After a visual inspection of the scree plots, we extracted SNP loadings from the first three canonical axes.

**Genotype-phenotype association analysis with RDA**

We fitted partial RDA with the five least correlated and most explanatory quantitative traits selected by correlation analysis. The RDA were fitted with the quantitative traits as explanatory variables, conditioned on geography, and the genetic dataset as response variable. SNPs exhibiting RDA loadings greater than three and half standard deviations from the mean were identified as association signals [59]. After a visual inspection of the scree plots, we extracted SNP loadings from the first three canonical axes.

**Results**

**Habitat suitability**

The suitability of an environmental niche for a population depends on which environmental predictors are influencing the species. The habitat suitability maps produced by SDMs suggests that the 26 populations have different niches (Fig. 3).

**Genomic diversity of Ethiopian indigenous chickens**

PCA based on the filtered variants provides information on the structure and relatedness of the 26 Ethiopian indigenous chicken sample populations (Figs. 4, 5 and 6). The PCA shows no clear separation among populations ( $n=20$ ) sampled from *gradient-I*, *III*, and *gradient-IV*, while populations ( $n=6$ ) sampled from *gradient-II* have distinctly separated from the other three gradients (Fig. 4).

The PCA based on gradients (Fig. 5) illustrates clear separation between chickens sampled from *gradient-II* and the other three gradients. Admixture is seen between *gradients-I* and *-IV*, and between *gradients-III* and *-IV*.

Chicken populations sampled from the three agroecologies (lowland, midlatitude, and highland) did not clearly differentiate except in *gradient-II* where populations sampled from the lowlands were distinct from those

sampled from the midlands and highlands of the same gradient (Fig. 6).

After carefully looking at the three PCA plots and understanding the genetic structure (Figs. 4, 5 and 6) of Ethiopian chickens, we decided that the sampled populations from *gradient-II* should not be analysed together with populations from the other three gradients.

**Signatures of selection for environmental adaptation**

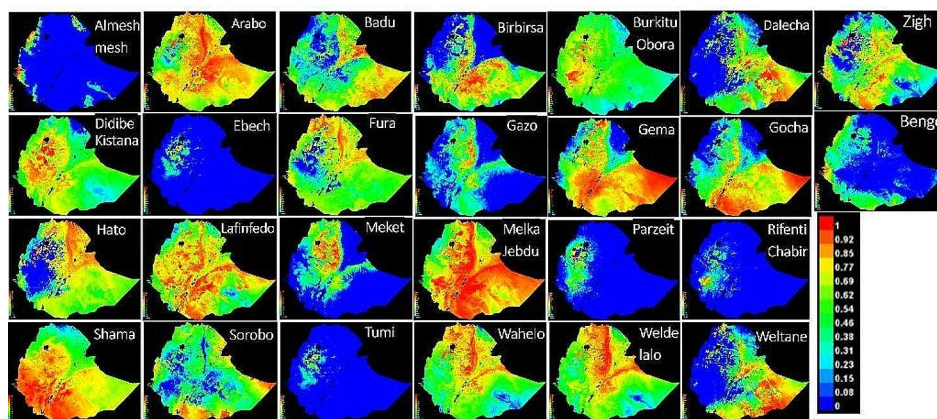
**Genetic differentiation between gradients**

The mean  $F_{ST}$  values between any two gradients was low (Table 1). This suggests that genetic differentiation between geographies among Ethiopian indigenous chicken populations is very little. A complete list of significant genes ( $p<0.01$ ) from overlapping windows jointly identified by  $F_{ST}$  and XP-EHH in each gradient-wise comparisons is presented in Supplementary Table 5.

The Manhattan plots of  $F_{ST}$  analyses show pairwise comparison between populations sampled from environmental *gradients-I*, *-III*, and *-IV* (Supplementary Fig. 1). Some regions of the genome show genetic differentiation across gradients.

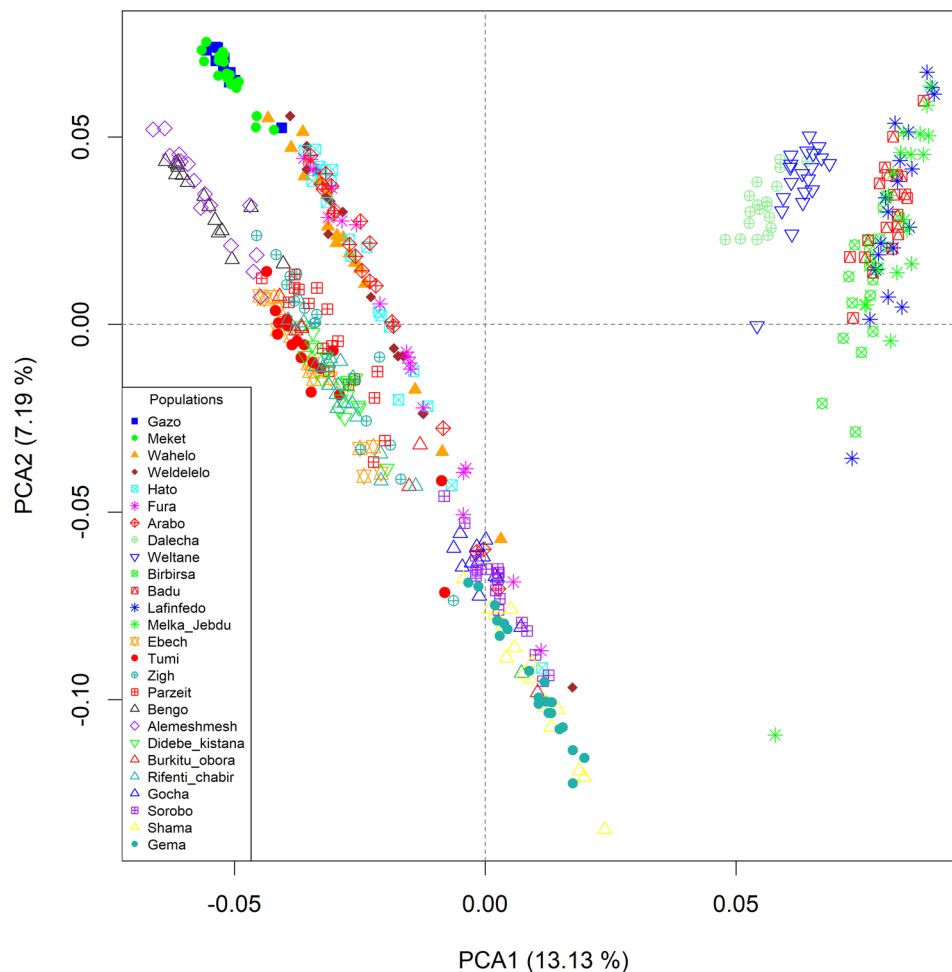
**Genetic differentiation between agroecologies across gradients**

The mean  $F_{ST}$  values between any two agroecologies across gradients was lower than values obtained for comparisons between any two gradients (Table 2). The Manhattan plots on  $F_{ST}$  and their scores for comparisons across agroecologies (Supplementary Fig. 2; Supplementary Table 6) show very low values across the genome. The  $F_{ST}$  scores for comparisons between agroecologies within *gradient-II* (Table 2, Supplementary Fig. 3, Supplementary Table 6) are relatively higher than between agroecologies across gradients, owing to differences in genetic background evidenced by the population structure analysis, Figs. 4, 5 and 6.



**Fig. 3** Habitat suitability maps of the 26 Ethiopian chicken populations. Colours towards red spectrum indicate more suitable conditions



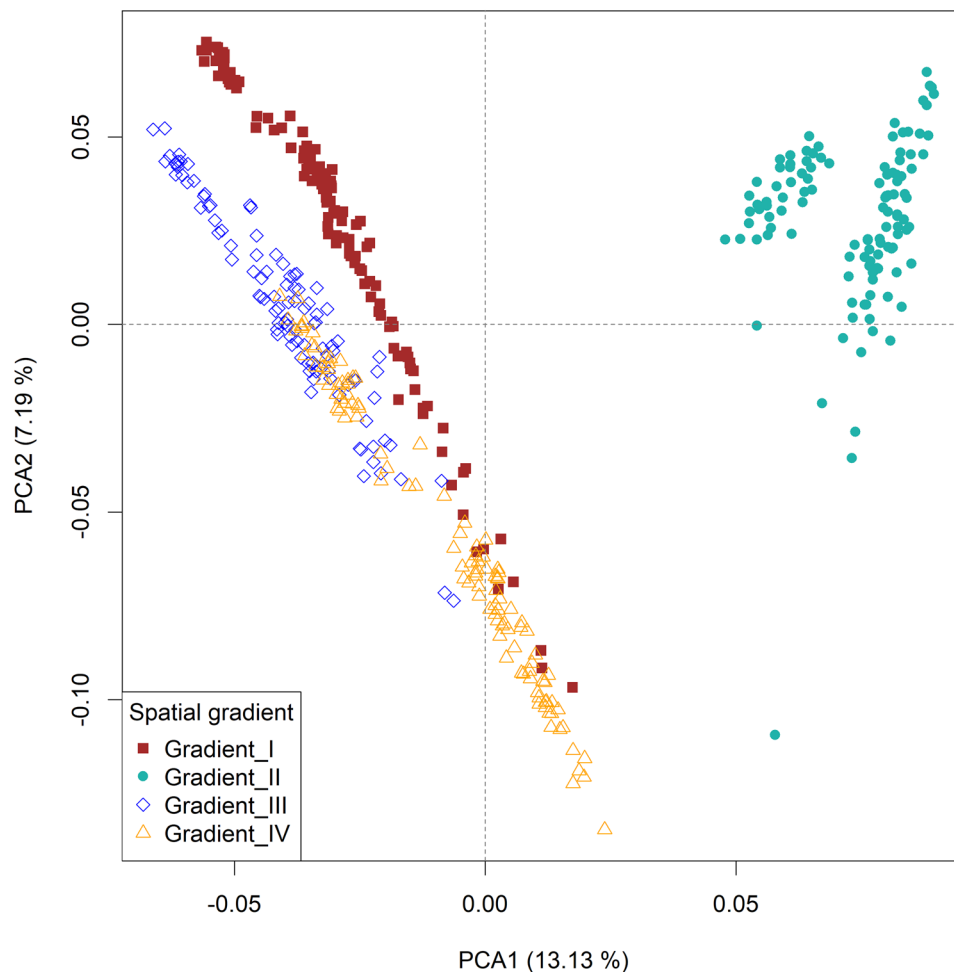


**Fig. 4** PCA plots of 26 Ethiopian indigenous chickens by population based on 25 million autosomal SNPs

#### Selection signatures between agroecologies across gradients

We performed XP-EHH analysis to identify genomic loci associated with high-altitude adaptation. Figure 7 shows the most significant selective sweeps for highland vs. lowland populations. All SNPs with a  $-\log(p\text{-value})$  above 2 or below  $-2$  from the green line are significantly selected ( $p < 0.01$ ) in one agroecology but not in the other. The most significant window under selection in the highland populations was found on chromosome four, overlapping the *GALNTL6* gene (XP-EHH=4.16). Variants in this gene have been associated with power performance in humans [116] possibly by showing a positive effect on anaerobic metabolism. In addition, several genes under selection (XP-EHH>2.7) are part of the calcium signalling pathway which has been associated with high altitude adaptation and hypoxia in previous studies including Tibetan chickens [19]. The genes identified in this pathway include *ERBB4* [117], *PLCB2*, *STIM2* [118], and *GNAS* [119]. The *MOAA* and *MOAB* genes are also under strong selection in the high altitude populations

(XP-EHH>3.5), these genes correlate with the expression of *HIF-1 $\alpha$*  and with transcription factors *Sp1* and *Sp3* which are master regulators of the cellular and developmental response to hypoxia [120]. Other genes under selection include the *RIPPLY2*, associated with body length [121], the *SGCZ* gene which response to the HIF-1 transcription activity during hypoxia [122], the *SPNS2* gene that regulated hypoxia-inducible factor 2alpha [123], and the *BRINP3* gene under selection in high-altitude Andeans [124]. A complete list of genes from overlapping windows jointly identified by *F<sub>ST</sub>* and XP-EHH in agroecological comparisons in lowland vs. highland, lowland vs. midland, and midland vs. highland respectively across the three gradients (*layer-II*) are presented respectively in Supplementary Table 7. XP-EHH scores for comparisons between different agroecologies across the three gradients (*layer-II*) and between gradients are presented in Supplementary Tables 8 and Supplementary Table 9 respectively.



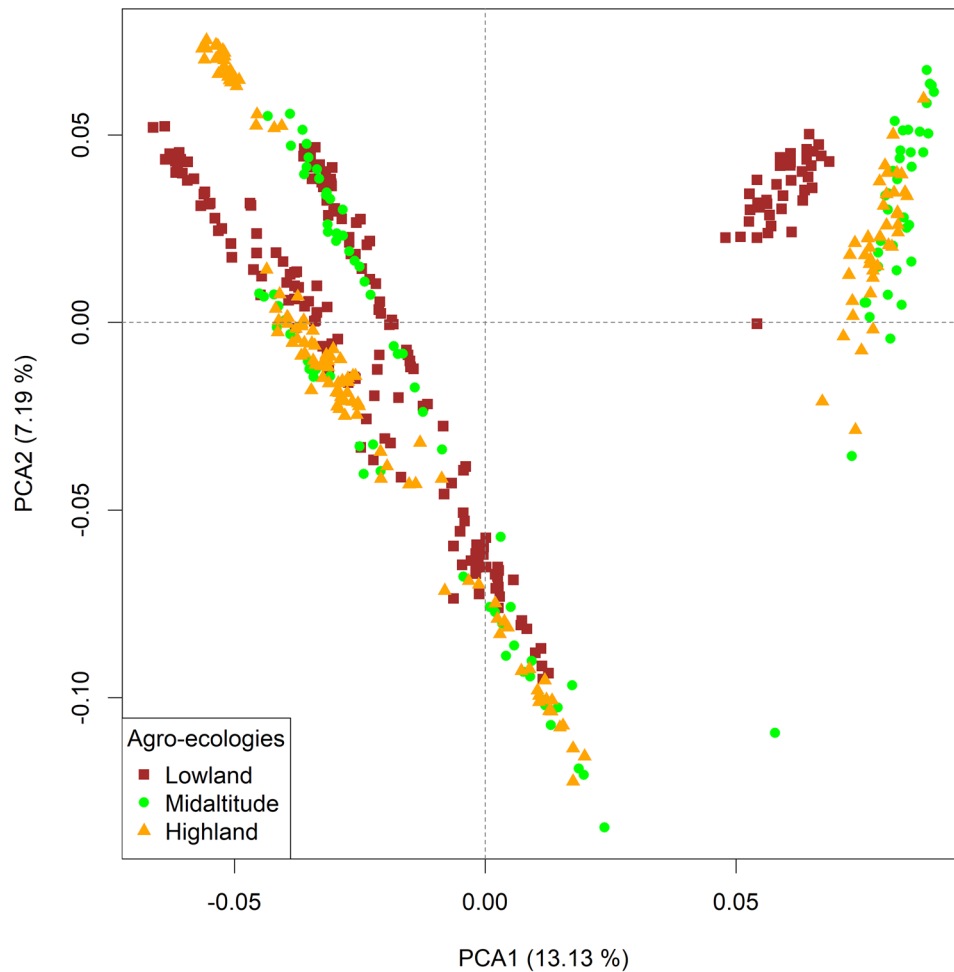
**Fig. 5** PCA plots of 26 Ethiopian indigenous chickens by gradient based on 25 million autosomal SNPs

### Selection signatures between agroecologies within *gradient-II* (analytical layer-III)

Figure 8 shows the most significant selective sweeps for highland vs. lowland populations within *gradient-II*. The most significant window was found on chromosome 3 overlapping the follicle stimulating hormone receptor (*FSHR*) gene, which is an activator of the hypoxia-inducible factor-1 protein [125], a key regulator of oxygen homeostasis. The second strongest signal is found on chromosome 6 overlapping the *CHAT* gene which has a direct interaction with the hypoxia-inducible factor (HIF)-1 $\alpha$  protein [126]. Interestingly, the third strongest peak overlaps with the *RYR2* gene. This gene is well known to be associated with high altitude adaptation in Tibetan chickens [19]. Other notable genes under selection include *HIGDIA* (hypoxia inducible domain family, member 1 A), *IGFBP1* (insulin like growth factor), *CAP2*, and *HERC4*, of which the latter two have been found to be differentially expressed under hypoxic environments [18].

There is no significant enrichment for the highland genes found. However, the genes under selection in the lowland populations are enriched for the ECM-receptor interaction and focal adhesion which serve as an important role in tissue and organ morphogenesis and in the maintenance of cell and tissue structure and function [127].

A significant overlap (13.4%) was observed between significant windows ( $p < 0.01$ ) identified by  $F_{ST}$  and XP-EHH analyses in the pairwise agroecological comparisons across gradients and within *gradient-II* (20.9%), indicating that the two methods target the same regions and hence are good predictors of selection signatures. Additional details are given under Supplementary Fig. 5 for overlaps between the two methods across gradients and under Supplementary Fig. 6 for overlaps within *gradient-II*. Complete list of genes from overlapping windows jointly identified by  $F_{ST}$  and XP-EHH in agroecological comparisons across gradients and within *gradient-II* are also presented in Supplementary Table 10 A–C and Supplementary Table 10D–E, respectively.



**Fig. 6** PCA plots of 26 Ethiopian indigenous chickens by agroecology based on 25 million autosomal SNPs

**Table 1** Mean  $F_{ST}$  scores between chicken populations sampled from different gradients

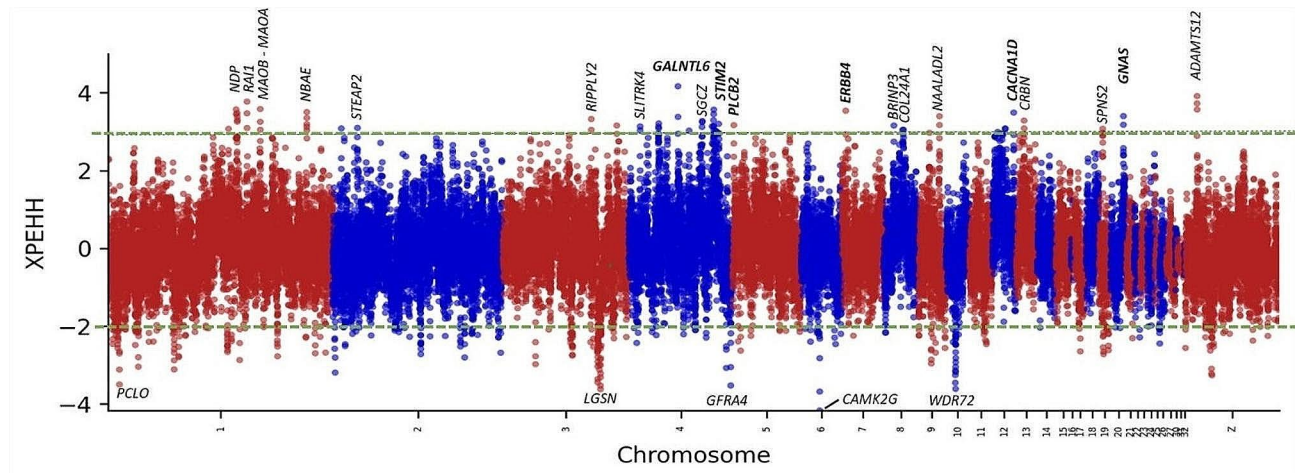
Between gradients ( <i>layer-I</i> )	
	Mean $F_{ST}$
<i>I</i> vs. <i>III</i>	0.021
<i>I</i> vs. <i>IV</i>	0.028
<i>III</i> vs. <i>IV</i>	0.021

**Table 2** Mean  $F_{ST}$  scores between chicken populations sampled from different agroecologies

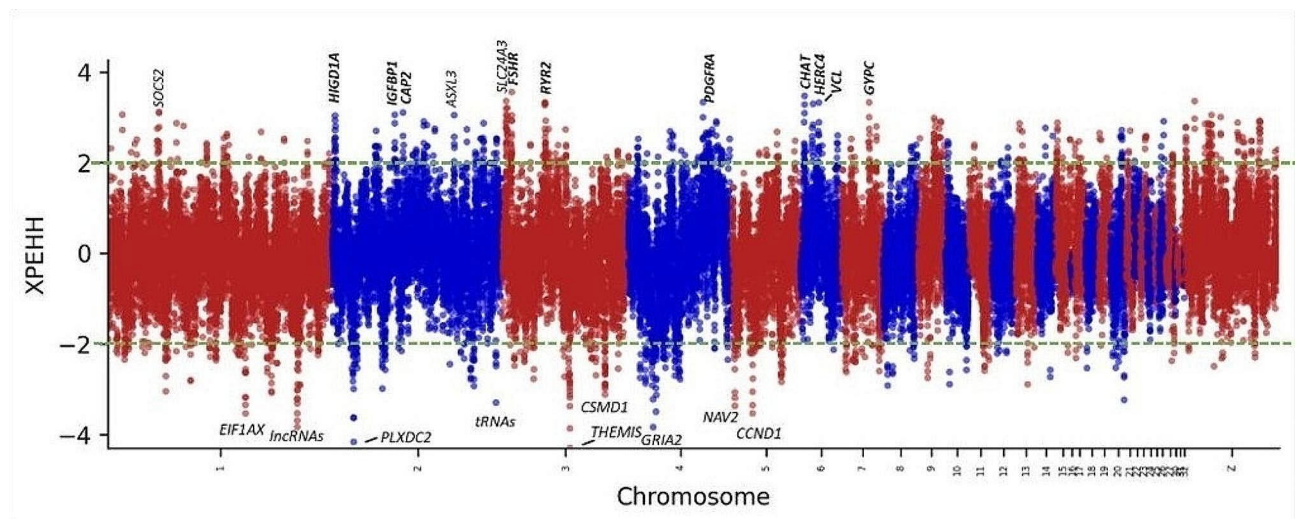
Between agroecologies		
	Mean $F_{ST}$ between agroecologies across gradients ( <i>layer-II</i> )	Mean $F_{ST}$ between agroecologies within gradient-II ( <i>layer-III</i> )
Lowland-Midland	0.009	0.021
Lowland-Highland	0.004	0.018
Midland-Highland	0.011	0.009

**Pathway enrichment analysis**

We assessed whether the genes under selection in the highland population are enriched for specific KEGG pathways. In total 150 bins have a XP-EHH value greater than 2.7 ( $p < 0.01$ ). These bins include a total of 95 genes. Only the “Calcium signalling pathway” was significantly enriched (q-value < 0.1) and in which five genes were found (*ERBB4*, *GRIN2A*, *STIM2*, *GNAS*, *PLCB2*) to be under selection in the highland populations. Several candidate genes in the calcium-signalling pathway were found to be under directional selection in adaptation to the hypoxia experienced by two Tibetan chicken populations [19], suggesting a potential genetic mechanism underlying high altitude adaptation to be similar in Ethiopian highland chicken compared to Tibetan chickens [19].  $Ca^{2+}$  are signalling molecules that regulate the response to hypoxia, which modulates cell contraction, cell proliferation and growth [128, 129]. Moreover, calcium signalling stimulates the translation of HIF-alpha, a transcription factors that mediates adaptation to hypoxia [130]. The candidate selected genes identified in this



**Fig. 7** XP-EHH plots for overlapping bins of 50 kb indicating positive selection in the highland populations while negative values indicating selection in lowland populations of Ethiopian indigenous chicken populations sampled across three gradients (-I, -III, and -IV). All SNPs with a  $-\log(p\text{-value})$  above 2 or below  $-2$  from the green line are significantly selected ( $p < 0.01$ ) in one agroecology but not in the other. Genes indicated in bold have been associated with the calcium signalling pathway or hypoxia



**Fig. 8** XP-EHH plot for highland vs. lowland in *gradient-II*. All SNPs with a  $-\log(p\text{-value})$  above 2 or below  $-2$  from the green line are significantly selected ( $p < 0.01$ ) in one agroecology but not in the other. Positive XP-EHH values indicate positive selection in the highland populations while negative values indicate selection in lowland populations. Genes indicated in bold have been associated with the hypoxia related pathways

study, and their variants, may be useful targets for clarifying our understanding of high-altitude adaptation in chicken. In addition, the “ribosome pathways” is enriched for windows under selection in the lowland populations ( $XP\text{-EHH} < -2$ ) from a total of 212 windows and 112 genes.

XP-EHH detected signatures of selection between populations sampled from any two gradients were also strong (Supplementary Fig. 4.). However, they were not as strong as signatures detected across agroecologies.

**Pathway enrichment analysis**

We assessed whether the genes under selection in the highland population are enriched for specific KEGG

pathways. In total 150 bins have a XP-EHH value greater than 2.7 ( $p < 0.01$ ). These bins include a total of 95 genes. Only the “Calcium signalling pathway” was significantly enriched ( $q\text{-value} < 0.1$ ) in which five genes were found (*ERBB4*, *GRIN2A*, *STIM2*, *GNAS*, *PLCB2*) to be under selection in the highland populations. Several candidate genes in the calcium-signalling pathway were found to be under directional selection in adaptation to the hypoxia experienced by two Tibetan chicken populations [19], suggesting a potential genetic mechanism underlying high altitude adaptation to be similar in Ethiopian highland chicken compared to Tibetan chickens [19].  $Ca^{2+}$  are signalling molecules that regulate the response to hypoxia, which modulates cell contraction,

cell proliferation and growth [128, 129]. Moreover, calcium signalling stimulates the translation of HIF-alpha, a transcription factors that mediates adaptation to hypoxia [130]. The candidate selected genes identified in this study, and their variants, may be useful targets for clarifying our understanding of high-altitude adaptation in chicken. In addition, the “ribosome pathways” is enriched for windows under selection in the lowland populations ( $XP-EHH < -2$ ) from a total of 212 windows and 112 genes.

XP-EHH detected signatures of selection between populations sampled from any two gradients were also strong (Supplementary Fig. 4). However, they were not as strong as signatures detected across agroecologies.

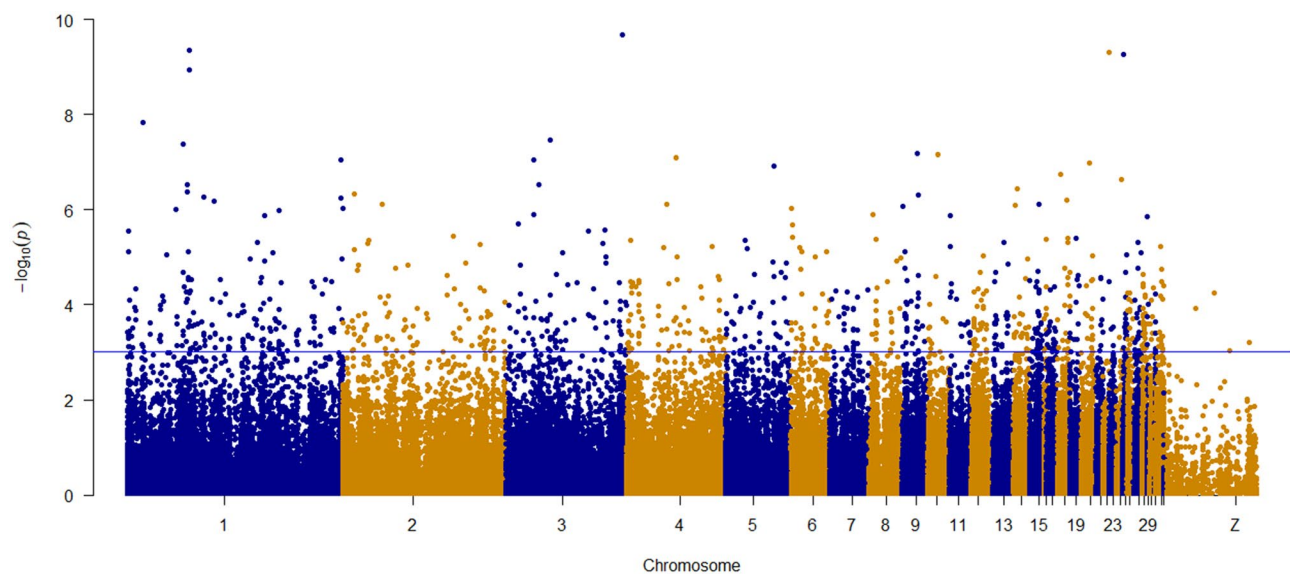
#### Genotype-environment associations (GEA)

Of the ten predictors identified through MaxEnt-based SDMs for their association with habitat suitability of chickens [46] and elevation (added as a tenth predictor), 6 less correlated ( $r \leq |0.7|$ ) predictors were retained for RDA (Supplementary Fig. 5). These predictors were: precipitation of the warmest quarter, precipitation of the coldest quarter, solar radiation of May, elevation, soil clay content and temperature seasonality. We had as many RDA axes as we had predictors ( $n = 6$ ) in our model. The first three RDA axes explained more than half (68.1%) of the variance in the environmental predictors (Supplementary Table 11). The significance of models in RDA is based on  $F$ -statistics [131]. The adjusted  $R^2$  considering the number of environmental predictors was 0.02, meaning that our constrained ordination explains about 2% of the variation or that 2% of the SNP variation is associated with the environmental predictors. Based on the magnitude of the

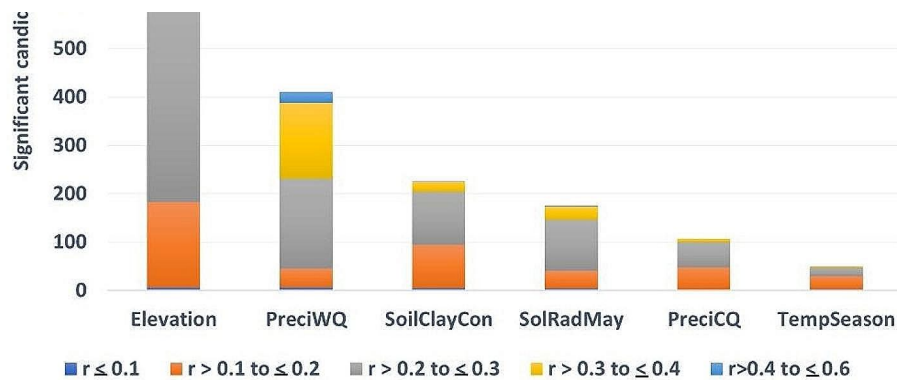
arrows in PCA plots based on RDA axes 1 and 2 (Supplementary Fig. 8) elevation, precipitation of the warmest quarter, and soil clay content had the highest contributions to genotypic variation, while temperature seasonality and solar radiation had the lowest contributions.

The SNP loadings for environmental predictors on each of the three RDA axes show a relatively normal distribution (Supplementary Fig. 9). The 1,909 SNPs from the two extreme ends of the loading distribution with standard deviation  $> 3.5$  (two-tailed  $p$ -value = 0.0005) for each significant axis were taken as outlier SNPs that are associated with environmental variation. The list of candidate SNPs which have significant association ( $p < 0.001$ ) with the six environmental predictors and considered to be under selection are presented in Supplementary Table 12. SNPs associated with the combined set of environmental predictors in *gradients -I, III, and -IV* do not show a clear clustering but are more or less evenly spread across the genome (Fig. 9).

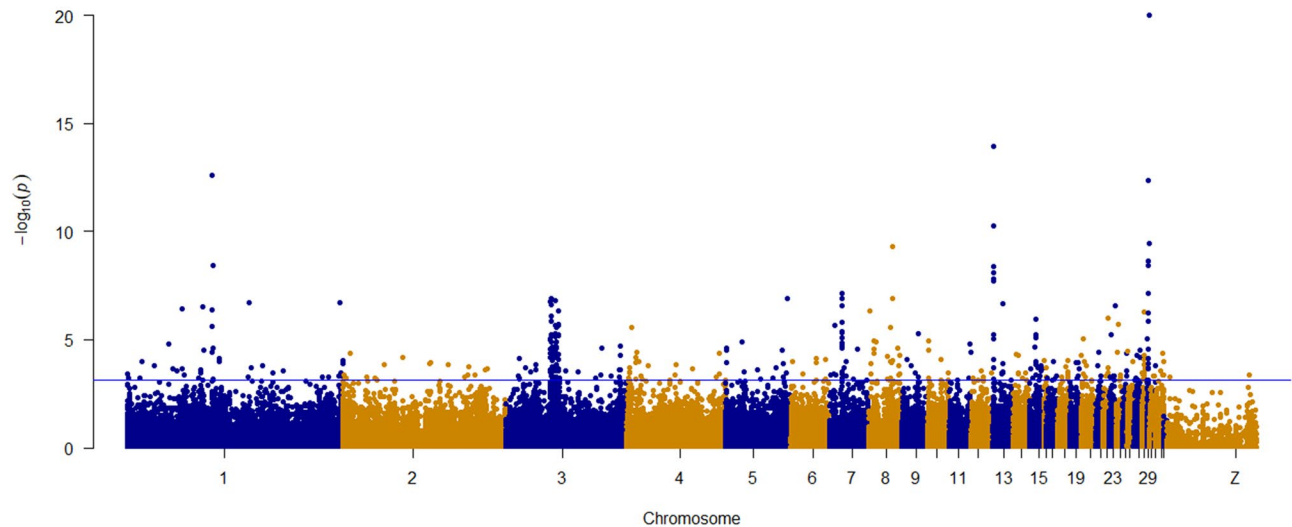
Some of the highest  $-\log_{10}(p)$  values are found on chromosomes 1 and 3 (Fig. 9). Only the peak on chromosome 1 shows additional significant SNPs near the top SNP. The significant candidate SNPs ( $n = 1,909$ ) that are associated with the combined set of environmental predictors are assigned to individual predictors based on the correlation values estimated by partial RDA analysis (Fig. 10). Most candidate SNPs (942 or 49.3%) have their highest correlation with elevation. Elevation has also the highest number ( $n = 321$  or 57.4%) of the moderately to highly associated SNPs ( $n = 559$ ) ( $0.3 < r < 0.6$ ). The second environmental predictor most associated with candidate SNPs is precipitation of the warmest quarter. It has correlation with 410 candidate SNPs (21.47%). The other 4



**Fig. 9** Manhattan plot of RDA showing the association of SNPs with the combined set of six environmental predictors in the three *gradients (-I, -III, and -IV)* as explanatory variables. The y-axis indicates  $-\log_{10}(p)$ . Horizontal blue line indicates the significance threshold ( $p < 0.001$ )



**Fig. 10** Number of significant candidate SNPs ( $p < 0.001$ ) that are most correlated with each of the six selected environmental predictors grouped by absolute magnitude of their correlation



**Fig. 11** Manhattan plot of RDA showing the association of SNPs with phenotypic variation in the five quantitative traits in *gradients -I, -III, and -IV*. The y-axis indicates  $-\log_{10}(p)$ . Horizontal blue line indicates the significance threshold ( $p < 0.001$ )

environmental predictors have the highest correlation for a smaller number of SNPs ( $n=557$  or 29.17%), but for all predictors, a considerable number of SNPs ( $n=59$ ) are found with correlations above  $|0.3|$  and only two SNPs have correlations above  $|0.4|$ .

**Genotype-phenotype association**

Out of a total of 8 phenotypic variables identified through MaxEnt-based SDMs for their utility in phenotypically discriminating study populations [46], five least correlated ( $|r| \leq 0.72$ ) quantitative traits were selected to be used for RDA (Supplementary Fig. 10). These five traits were live body weight, beak length, comb width, wattle width and earlobe width. The correlation between comb width and wattle width was 0.72 which is slightly higher than the common threshold ( $|r| > 0.7$ ) used to reduce variables. However, we decided to keep both traits because of their adaptive roles documented in literature

related with thermoregulation in tropical chickens. The first three RDA axes explained most of the variance (62.1%) in the phenotypic predictors (Supplementary Table 13). The adjusted  $R^2$  for the partial RDA was 0.002. This shows that only 0.2% of the SNPs variation is associated with quantitative traits.

The SNP loadings for quantitative traits on each of the three RDA axes show a relatively normal distribution (Supplementary Fig. 11). Based on the magnitude of the arrows in the PCA plots based on RDA axes 1 and 2 (Supplementary Fig. 12), comb width, wattle width and body weight were most useful in explaining SNP variation. SNPs associated with the combined set of quantitative traits in *gradients -I, -III, and -IV* show strong supportive peaks on chromosomes 1, 3, 4, 7, 8, 13, 15, and 29 indicating probable regions of quantitative trait loci (QTL) associated with phenotypic variation (Fig. 11). The peaks were more diffused across the genome for the Manhattan

plot showing association between SNPs and quantitative traits for populations sampled from *gradient-II* (Supplementary Fig. 13).

A stacked bar chart showing the number of outlier SNPs ( $p < 0.001$ ) that are most correlated with each of the five quantitative traits is presented in Fig. 12. The significant candidate SNPs (were assigned to individual traits based on correlation values estimated by partial RDA analysis. Partial RDA identified 1340 candidate SNPs that have significant association with the five quantitative traits (Supplementary Table 14). A total of 19 SNPs show moderate to high correlation with body weight ( $0.3 < r \leq 0.6$ ). Most candidate SNPs, 39%, were associated with comb width ( $n=519$ ) and 27% with body weight ( $n=360$ ) (Fig. 12 and Supplementary Table 13). Higher association was also seen between comb width and candidate SNPs for populations sampled from *gradient-II* (Supplementary Fig. 14).

The list of significant ( $p < 0.001$ ) candidate SNPs identified by RDA and their respective  $F_{ST}$  and XP-EHH values across gradients are presented in Supplementary Tables 15 and in Supplementary Table 16.

### Discussion

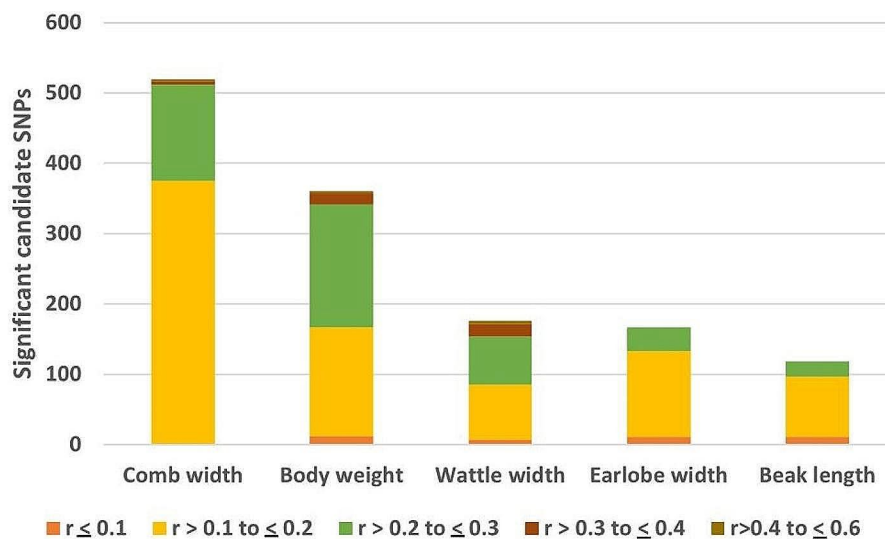
Adverse effects of climate change and increasing demand for animal source proteins, particularly in the tropics (particularly in Africa and Southeast Asia), necessitate that we properly understand the genetic architecture of environmental adaptation and develop productive and environmentally resilient breeds [132, 133]. Investigation of molecular pathways indicate that indigenous chickens are more adapted to the environment in which they live compared to specialized chickens [132]. Important

insights were obtained from earlier studies on local adaptation of African chickens [45, 134, 135] by applying SDMs and signatures of selection analyses. However, previous studies did not adequately relate genomic variation with environmental and phenotypic variation. Analysing genomic data without relating it environmental and phenotypic variation does not provide a complete picture of adaptive variation.

In the present study, we followed a landscape genomic approach to study adaptive and phenotypic variation among Ethiopian chickens. We applied an environmental-gradation approach to survey chicken populations across all possible agroclimatic clines in the country. Our sample size of 513 animals from four environmental gradients was large enough to capture adaptive variation across populations. For species with limited dispersal, sample sizes above 200 units are generally sufficient to detect most adaptive signals in landscape genomics, while in random mating populations this threshold should be increased to 400 units [84]. After applying stringent quality filtration, we had 25 M SNPs (autosomal and non-autosomal) and 466 individuals for downstream genomic analyses. The dataset used in the present study is substantially larger than previous genomic studies on Ethiopian chickens (which sampled a maximum of 225 birds per study) [45, 136, 137].

We combined different techniques including SDMs, genetic differentiation test ( $F_{ST}$ ), cross-population Extended Haplotype Homozygosity (XP-EHH), and RDA.

SDMs were used in our study to identify the most relevant environmental predictors influencing habitat suitability for chickens. The environmental predictors



**Fig. 12** Number of significant candidate SNPs ( $p < 0.001$ ) that are most correlated with each of the five quantitative traits, grouped by absolute magnitude of their correlation

identified in the present study (related with elevation, precipitation, soil clay content, solar radiation, and temperature) were reported in earlier studies for their influences on availability of feed, productivity, prevalence of diseases and parasites [26, 46, 138, 139]. The habitat suitability maps produced by SDMs suggest that the 26 Ethiopian indigenous chicken sample populations may have gone through different environmental selective pressures which give rise to phenotypic and genetic differentiation.

The gradients and agroecologies show low differentiation, as evidenced by the low  $F_{ST}$  values. Lower level of genetic differentiation was also detected among Ethiopian chickens by [45]. In contrast to the  $F_{ST}$  results, strong signals of selection ( $p < 0.01$ ) were detected by XP-EHH in pairwise agroecological comparisons. XP-EHH results show that selective pressure in Ethiopian chickens is stronger between agroecologies.

A large overlap was observed between significant windows identified by  $F_{ST}$  and XP-EHH analyses, suggesting that both methods identified similar regions in the genome are under selection. The overlap between  $F_{ST}$  and XP-EHH analyses ranged from 13.4 to 20.9% between agroecologies which is considerably higher than the 4.9% overlap reported by [45] between  $F_{ST}$  and XP-EHH for Ethiopian chickens. The large overlap between  $F_{ST}$  and XP-EHH in the present study might be due to our sampling strategy. Firstly, the sampling design captured a wide range of geographic and environmental variation and helped to survey most of the ecotypes and agroecologies in the country. Secondly, the design may have minimized confounding between neutral and adaptive processes which result from mixing of populations that have different demographic histories. By classifying the populations by gradients, we controlled for the effects of population genetic structure associated with specific geographies. For instance, a very high overlap between the  $F_{ST}$  and XP-EHH results was found in agroecological comparisons within *gradient-II*. The decision to analyse this gradient on its own was informed by PCA, which clearly separated populations of *gradient-II* from the other three *gradients* (-I, -II, and -III). *Gradient-II* represents chicken populations from eastern parts of Ethiopia which have a distinct evolutionary history and route of introduction into the country [38, 40, 136], in contrast to populations representing the other three gradients. Combining *gradient-II* with the other three would have reduced the overlap of  $F_{ST}$  and XP-EHH results.

Our results based on XP-EHH show that genes in the calcium signalling pathway are under selection in high-altitude adapted Ethiopian chicken populations as well as genes associated with the hypoxia-inducible factor (HIF) transcription factors. The gene under strongest selection is the *GALNTL6* gene associated with sports performance in multiple human studies. It is hypothesized that

this gene is expressed in the gut microbiome regarding regulation of short-chain fatty acids and their anti-inflammatory and resynthesis functions causing a positive effect on anaerobic metabolism. The *ERBB4* gene, found to be under selection in high altitude Ethiopian chicken populations, is also under selection in human Tibetan populations [117]. *ERBB4* is strongly associated with vascular wall stability, and possibly with the production of erythrocytes and belongs to the epidermal growth factor receptor subfamily. We also identified the *MAOB* and *MAOA* genes to be under selection, where *MAOB* has been shown to be correlated with HIF-1 $\alpha$  (tumour grade and hypoxia-inducible transcription factor) [120]. Inhibition of *MAOA* in cells may exert antitumour activity in the treatment of prostate cancer [140]. The roles of *MAOA* and *MAOB* genes in local adaptation of chicken need to be further investigated.

Results from signatures of selection analyses with the two methods ( $F_{ST}$  and XP-EHH) can be used complementarily with RDA to shed light on the relationship between genomic, phenotypic, and environmental variation in local adaptation studies in indigenous chickens. With RDA, we identified 83 candidate SNPs in regions on chromosomes 1,3, 4, 7,8, 13, 15, and 19 that have a moderate to high correlation ( $0.3 < r < 0.6$ ) with live body weight. Conventional GWAS studies in the past identified body weight associated SNPs and QTLs on chromosomes 1,4, 8, 11, 19 in Chinese, Rwandan, and Ethiopian chicken breeds [27, 141–144]. Our results demonstrate that RDA can be used as an alternative approach to GWAS in random mating, indigenous livestock populations which have sufficiently interacted with the environment.

Candidate SNPs associated with the six SDM-identified environmental predictors contributing to habitat suitability were identified by RDA. The RDA found only 2% of the SNP variation to be associated with the six environmental predictors. This is a small value but not unexpected because most of the SNPs are under neutral and therefore not show a relationship with the environmental predictors. SNPs that do show association with the environmental predictors are likely to be under selection. This selection can be in response to these selected predictors that were used in the model or some other environmental variable that is correlated with these predictors.

Candidate SNPs associated with environmental predictors (Fig. 9) were evenly spread across the genome without obvious overlap with the peaks from genotype-phenotype association (Fig. 11). Genotype-phenotype associations had very distinct suggesting that phenotypic variation is present among populations for selection to act on it. The environmental drivers could increase haplotypes related to adaptive phenotypic plasticity and morphological variation in indigenous chickens. The pea-comb, a dominant mutation in chickens, drastically



reduces the size of the comb and wattle, decreasing heat loss and making the chicken less susceptible to frost lesions [145]. Histological section analysis of dermal papillary layer has revealed that red earlobes have many more blood vessels and were associated with thinner skin than that of white earlobes [146] indicating the role of earlobes in thermoregulation. The total amount of SNP variation associated with phenotypic variation was only 0.2%, in contrast with 2% of the SNP variation associated with environmental variation. The underlying mechanisms of genotype-phenotype associations are well studied and understood in livestock, but this is not the case for genotype-environment associations. Finding 2% of SNP variation related to environment variation is promising for further investigation of the mechanisms leading to these associations.

In summary, in this manuscript we reported the first study integrating phenotypic, genomic, and environmental information on Ethiopian indigenous chickens. Our findings on genomic and phenotypic variability associated with environmental adaptation (e.g., genes selected in highland populations, genes associated with body weight and ecological variables) are useful in the design of breeding programmes aiming at developing more productive and resilient chicken strains (lines) suitable for smallholder systems in the face of climate crisis. The landscape genomic approach followed in the present study can also be used to study adaptive variation in other random mating indigenous livestock populations that are managed extensively to better understand organismal response to environmental variables and develop better breeding strategies.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10193-6>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

## Acknowledgements

The authors thank Nego Huruma and Million Abebe (veterinary technicians), and Michael Temesgen and Anteneh Hailemariam (drivers) who participated in the field work under challenging circumstances. The authors also thank all farmers and agricultural experts who cooperated in the study by providing essential information.

## Author contributions

FK, JB, HK, TD, RC, and OH conceived the ideas and designed the study. FK developed the sampling frame, collected biological samples and environmental data, trained field technicians, and drafted the manuscript. FK, JB, HK, and RC facilitated DNA extraction and whole genome sequencing. FK, MD, and CPB performed bioinformatic analyses. FK performed environmental data analyses. JB, MD, HK, TD, CPB and OH critically reviewed the manuscript and provided additional comments. TD, JB, HK, RC and OH secured funding. All authors have read and approved the final manuscript for publication.

## Funding

This work was supported by Animal Breeding and Genomics (ABG) of Wageningen University and Research (WUR) and the Koepon Foundation. The African Chicken Genetic Gains (ACGG) project and the Tropical Poultry Genetic Solutions (TPGS) Program of the International Livestock Research Institute (ILRI) co-financed the work with grant received from the Bill and Melinda Gates Foundation (Investment ID: OPP1112198 and INV-042538). Additional funding was obtained from the CGIAR Initiative on Sustainable Animal Productivity or SAPLING. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation, WUR, ILRI and the Koepon Foundation. WUR, the Koepon foundation, and ILRI jointly sponsored Fasil Getachew Kebede as a PhD student at ABG, WUR.

## Data availability

The datasets analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

The experimental protocols were approved by Institutional Animal Use and Care Committee of the International Livestock Research Institute (ILRI), Ref no. IACUC2019-12. All methods were carried out in accordance with ARRIVE guidelines.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Animal Breeding and Genomics, Wageningen University & Research, Droevendaalsesteeg 1, Wageningen PB-6708, The Netherlands  
<sup>2</sup>International Livestock Research Institute, P.O. Box 5689, Addis Ababa, Ethiopia  
<sup>3</sup>School of Life Sciences, The University of Nottingham, Nottingham NG7 2RD, UK

Received: 3 October 2023 / Accepted: 5 March 2024

Published online: 18 March 2024

## References

- Doekes HP, Bovenhuis H, Berghof TV, Peeters K, Visscher J, Mulder HA. Research note: genome-wide association study for natural antibodies and resilience in a purebred layer chicken line. *Poult Sci.* 2023;102(1):102312.
- Rovelli G, Ceccobelli S, Perini F, Demir E, Mastrangelo S, Conte G, Abeni F, Marletta D, Ciampolini R, Cassandro M. The genetics of phenotypic plasticity in livestock in the era of climate change: a review. *Italian J Anim Sci.* 2020;19(1):997–1014.
- Silpa MV, König S, Sejian V, Malik PK, Nair MRR, Fonseca VF, Maia ASC, Bhatta R. Climate-resilient dairy cattle production: applications of genomic tools and statistical models. *Front Veterinary Sci.* 2021;8:625189.
- Sánchez-Molano E, Kapsona VV, Ilska JJ, Desire S, Conington J, Mucha S, Banos G. Genetic analysis of novel phenotypes for farm animal resilience to weather variability. *BMC Genet.* 2019;20(1):1–10.
- Kawecki TJ, Ebert D. Conceptual issues in local adaptation. *Ecol Lett.* 2004;7(12):1225–41.
- Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. *Nat Rev Genet.* 2013;14(11):807.
- Colditz IG, Hine BC. Resilience in farm animals: biology, management, breeding and implications for animal welfare. *Anim Prod Sci.* 2016;56(12):1961–83.
- Berghof TV, Poppe M, Mulder HA. Opportunities to improve resilience in animal breeding programs. *Front Genet.* 2019;9:692.
- Knap P, Su G. Genotype by environment interaction for litter size in pigs as quantified by reaction norms analysis. *Animal: Int J Anim Bioscience.* 2008;2(12):1742.

10. Storz JF. Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. *Mol Ecol*. 2002;11(12):2537–51.
11. Leinonen T, Cano JM, Mäkinen H, Merilä J. Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *J Evol Biol*. 2006;19(6):1803–12.
12. Schmid M, Guillaume F. The role of phenotypic plasticity on population differentiation. *Heredity* (Edinb). 2017;119(4):214–25.
13. Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*. 1973;74(1):175–95.
14. Turesson G. The species and the variety as ecological units. *Hereditas*. 1922;3(1):100–13.
15. Conover DO, Duffy TA, Hice LA. The covariance between genetic and environmental influences across ecological gradients. *Ann NY Acad Sci*. 2009;1168(1):100–29.
16. Lorenzo FR, Huff C, Myllymäki M, Olenchock B, Swierczek S, Tashi T, Gordeuk V, Wuren T, Ri-Li G, McClain DA. A genetic mechanism for tibetan high-altitude adaptation. *Nat Genet*. 2014;46(9):951–6.
17. Hackinger S, Kraaijenbrink T, Xue Y, Mezzavilla M, van Driem G, Jobling MA, de Knijff P, Tyler-Smith C, Ayub Q. Wide distribution and altitude correlation of an archaic high-altitude-adaptive EPAS1 haplotype in the Himalayas. *Hum Genet*. 2016;135(4):393–402.
18. Zhang Y, Zheng X, Zhang Y, Zhang H, Zhang X, Zhang H. Comparative transcriptomic and proteomic analyses provide insights into functional genes for hypoxic adaptation in embryos of tibetan chickens. *Sci Rep*. 2020;10(1):1–13.
19. Wang M-S, Li Y, Peng M-S, Zhong L, Wang Z-J, Li Q-Y, Tu X-L, Dong Y, Zhu C-L, Wang L. Genomic analyses reveal potential independent adaptation to high altitude in tibetan chickens. *Mol Biol Evol*. 2015;32(7):1880–9.
20. Ai H, Yang B, Li J, Xie X, Chen H, Ren J. Population history and genomic signatures for high-altitude adaptation in tibetan pigs. *BMC Genomics*. 2014;15(1):1–14.
21. Ma Y-F, Han X-M, Huang C-P, Zhong L, Adeola AC, Irwin DM, Xie H-B, Zhang Y-P. Population genomics analysis revealed origin and high-altitude adaptation of tibetan pigs. *Sci Rep*. 2019;9(1):11463.
22. Friedrich J, Wiener P. Selection signatures for high-altitude adaptation in ruminants. *Anim Genet*. 2020;51(2):157–65.
23. Merilä J, Hendry AP. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evol Appl*. 2014;7(1):1–14.
24. Sgro CM, Terblanche JS, Hoffmann AA. What can plasticity contribute to insect responses to climate change? *Annu Rev Entomol*. 2016;61:433–51.
25. Kelly M. Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. *Philosophical Trans Royal Soc B*. 2019;374(1768):20180176.
26. Bettridge JM, Psifidi A, Terfa ZG, Desta TT, Lozano-Jaramillo M, Dessie T, Kaiser P, Wigley P, Hanotte O, Christley RM. The role of local adaptation in sustainable production of village chickens. *Nat Sustain*. 2018;1(10):574–82.
27. Psifidi A, Banos G, Matika O, Desta TT, Bettridge J, Hume DA, Dessie T, Christley R, Wigley P, Hanotte O. Genome-wide association studies of immune, disease and production traits in indigenous chicken ecotypes. *Genet Selection Evol*. 2016;48(1):74.
28. Horst P. Native fowl as reservoir for genomes and major genes with direct and indirect effects on the adaptability and their potential for tropically orientated breeding plans. *Archiv fuer Gefluegelkunde (Germany, FR)* 1989.
29. Sheng Z, Pettersson ME, Hu X, Luo C, Qu H, Shu D, Shen X, Carlborg Ö, Li N. Genetic dissection of growth traits in a Chinese indigenous commercial broiler chicken cross. *BMC Genomics*. 2013;14(1):1–12.
30. Oishi I, Yoshii K, Miyahara D, Kagami H, Tagami T. Targeted mutagenesis in chicken using CRISPR/Cas9 system. *Sci Rep*. 2016;6(1):1–10.
31. Khwatenge CN, Nahashon SN. Recent advances in the application of CRISPR/Cas9 gene editing system in Poultry species. *Front Genet*. 2021;12:127.
32. Ballantyne M, Woodcock M, Doddamani D, Hu T, Taylor L, Hawken RJ, McGrew MJ. Direct allele introgression into pure chicken breeds using Sire Dam Surrogate (SDS) mating. *Nat Commun*. 2021;12(1):1–10.
33. Kebede FG, Komen H, Alemayehu TD, Hanotte O, Kemp S, Alemu SW, Bastiaansen JW. Agroecologies defined by species distribution models improve model fit of genotype by environment interactions to identify the best performing chicken breeds for smallholder systems. *Front Sustainable Food Syst*. 2023;7:1305799.
34. Woldekios H, D'Andrea A. Early evidence for domestic chickens (*Gallus gallus Domesticus*) in the Horn of Africa. *Int J Osteoarchaeology*. 2017;27(3):329–41.
35. Dessie T. Phenotypic and genetic characterization of local chicken ecotypes in Ethiopia. Humboldt-Universität zu Berlin; 2003.
36. Hassen H, Nesor F, De Kock A, van Marle-Köster E. Study on the genetic diversity of native chickens in northwest Ethiopia using microsatellite markers. *Afr J Biotechnol* 2009, 8(7).
37. Dana N. Breeding programs for indigenous chicken in Ethiopia analysis of diversity in production systems and chicken populations. Wageningen University; 2011.
38. Mwacharo JM, Björnstad G, Mobegi V, Nomura K, Hanada H, Amano T, Jianlin H, Hanotte O. Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Mol Phylogenet Evol*. 2011;58(2):374–82.
39. Adebabay K. Whole genome based characterization of indigenous chicken populations in Ethiopia. *Dissertation Addis Ababa: Addis Ababa University*; 2018.
40. Mwacharo JM, Björnstad G, Han J, Hanotte O. The history of African village chickens: an archaeological and molecular perspective. *Afr Archaeol Rev*. 2013;30(1):97–114.
41. Lyimo C, Weigend A, Msoffe P, Eding H, Simianer H, Weigend S. Global diversity and genetic contributions of chicken populations from african, a sian and E uropean regions. *Anim Genet*. 2014;45(6):836–48.
42. Billi P. Landscapes and landforms of Ethiopia. Springer; 2015.
43. Fleming DS, Weigend S, Simianer H, Weigend A, Rothschild M, Schmidt C, Ashwell C, Persia M, Reecy J, Lamont SJ. Genomic comparison of indigenous African and northern European chickens reveals putative mechanisms of stress tolerance related to environmental selection pressure. *G3: Genes Genomes Genet*. 2017;7(5):1525–37.
44. Elbeltagy AR, Bertolini F, Fleming DS, Van Goor A, Ashwell CM, Schmidt CJ, Kugonza DR, Lamont SJ, Rothschild MF. Natural selection footprints among African Chicken breeds and Village Ecotypes. *Front Genet*. 2019;10:376.
45. Gheyas AA, Vallejo-Trujillo A, Kebede A, Lozano-Jaramillo M, Dessie T, Smith J, Hanotte O. Integrated environmental and genomic analysis reveals the drivers of local adaptation in African indigenous chickens. *Mol Biol Evol* 2021.
46. Kebede FG, Komen H, Dessie T, Alemu SW, O H JaB. Species and phenotypic distribution models reveal Population differentiation in Ethiopian indigenous chickens. *Front Genet*. 2021;12:723360.
47. Gheyas AA, Trujillo AV, Kebede A, Lozano-Jaramillo M, Dessie T, Smith J, Hanotte O. Integrated environmental and genomic analysis reveals the drivers of local adaptation in African indigenous chickens. *bioRxiv* 2020.
48. Balkenhol N, Cushman SA, Waits LP, Storfer A. Current status, future opportunities, and remaining challenges in landscape genomics [Chapter 14]. In: *Balkenhol, Niko; Cushman, Samuel A; Storfer, Andrew T; Waits, Lisette P, eds Landscape Genetics: Concepts, Methods, Applications, First Edition John Wiley and Sons Ltd p 247–255* 2016:247–255.
49. Storfer A, Patton A, Fraik AK. Navigating the Interface between Landscape Genetics and Landscape Genomics. *Front Genet*. 2018;9:68.
50. Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, Taberlet P. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol Ecol*. 2007;16(18):3955–69.
51. Pariset L, Joost S, Gargani M, Valentini A. Landscape genomics in livestock. In: *Analysis of Genetic Variation in Animals* IntechOpen; 2012.
52. Mdladla K. Landscape genomic approach to investigate genetic adaptation in South African indigenous goat populations. 2016.
53. Roffler GH, Schwartz MK, Pilgrim KL, Talbot SL, Sage GK, Adams LG, Luikart G. Identification of landscape features influencing gene flow: how useful are habitat selection models? *Evol Appl*. 2016;9(6):805–17.
54. Li Y, Zhang X-X, Mao R-L, Yang J, Miao C-Y, Li Z, Qiu Y-X. Ten years of landscape genomics: challenges and opportunities. *Front Plant Sci*. 2017;8:2136.
55. Cesconeto RJ, Joost Sp, McManus CM, Paiva SR, Cobuci JA, Braccini J. Landscape genomic approach to detect selection signatures in locally adapted Brazilian swine genetic groups. *Ecol Evol*. 2017;7(22):9544–56.
56. Vallejo-Trujillo A, Kebede A, Lozano M, Dessie T, Sparks N, Smith J, Hanotte O, Gheyas A. Ecological niche modelling applies for the characterization of indigenous livestock species. the example of Ethiopian village chickens; 2018.
57. Goitom S, Gicheha MG, Njonge FK, Kiplangat N. Landscape genomics and selection signatures of local adaptation of Eritrean indigenous cattle along environmental gradients. *Trop Anim Health Prod*. 2021;53:1–8.
58. Harrison KA, Amish SJ, Pavlova A, Narum SR, Telonis-Scott M, Rourke ML, Lyon J, Tonkin Z, Gilligan DM, Ingram BA. Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. *Mol Ecol*. 2017;26(22):6253–69.
59. Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol Ecol*. 2018;27(9):2215–33.

60. Kess T, Boulding EG. Genome-wide association analyses reveal polygenic genomic architecture underlying divergent shell morphology in Spanish *Littorina saxatilis* ecotypes. *Ecol Evol.* 2019;9(17):9427–41.
61. Torrado H, Carreras C, Raventos N, Macpherson E, Pascual M. Individual-based population genomics reveal different drivers of adaptation in sympatric fish. *Sci Rep.* 2020;10(1):1–14.
62. Valette T, Leitwein M, Lascaux J-M, Desmarais E, Berrebi P, Guinand B. Spotting genome-wide pigmentation variation in a brown trout admixture context. *bioRxiv* 2020.
63. Talbot B, Chen T-W, Zimmerman S, Joost S, Eckert AJ, Crow TM, Semizer-Cuming D, Seshadri C, Manel S. Combining genotype, phenotype, and environment to infer potential candidate genes. *J Hered.* 2017;108(2):207–16.
64. Vangestel C, Eckert AJ, Wegryzn JL, St. Clair JB, Neale DB. Linking phenotype, genotype and environment to unravel genetic components underlying cold hardiness in coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*). *Tree Genet Genomes* 2018, 14(1).
65. Carvalho CS, Forester BR, Mitre SK, Alves R, Imperatriz-Fonseca VL, Ramos SJ, Resende-Moreira LC, Siqueira JO, Trevelin LC, Caldeira CF. Combining genotype, phenotype, and environmental data to delineate site-adjusted provenance strategies for ecological restoration. *Mol Ecol Resour.* 2021;21(1):44–58.
66. Elith J, Leathwick JR. Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Syst.* 2009;40:677–97.
67. Elith J, Franklin J. Species distribution modeling. *Encyclopedia of Biodiversity: Second Edition.* Elsevier Inc.; 2013. pp. 692–705.
68. Guisan A, Thuiller W, Zimmermann NE. Habitat suitability and distribution models: with applications in R. Cambridge University Press; 2017.
69. Hutchinson GE. Concluding remarks. *Cold Spring Harb Symp Quant Biol.* 1957. Cold Spring Harbor Laboratory Press; 1957. pp. 415–27.
70. Soberón J. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol Lett.* 2007;10(12):1115–23.
71. Vajana E, Barbato M, Colli L, Milanese M, Rochat E, Fabrizi E, Mukasa C, Del Corvo M, Masembe C, Muwanika VB, et al. Combining Landscape Genomics and Ecological Modelling to investigate local adaptation of indigenous Ugandan cattle to East Coast Fever. *Front Genet.* 2018;9:385.
72. Lozano-Jaramillo M. Predicting breed by environment interaction using ecological modelling. Wageningen: Wageningen University; 2019.
73. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4(3):e72.
74. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varily P, Shamovsky O, Palma A, Mikkelsen T, Altshuler D, Lander E. Positive natural selection in the human lineage. *Science.* 2006;312(5780):1614–20.
75. Pavlidis P, Alachiotis N. A survey of methods and tools to detect recent and strong positive selection. *J Biol Research-Thessaloniki.* 2017;24(1):1–17.
76. Sabeti PC, Varily P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R. Genome-wide detection and characterization of positive selection in human populations. *Nature.* 2007;449(7164):913–8.
77. Nei M. Definition and estimation of fixation indices. *Evolution.* 1986;40(3):643–5.
78. Cadzow M, Boocock J, Nguyen HT, Wilcox P, Merriman TR, Black MA. A bioinformatics workflow for detecting signatures of selection in genomic data. *Front Genet.* 2014;5:293.
79. Legendre P, Gallagher ED. Ecologically meaningful transformations for ordination of species data. *Oecologia.* 2001;129(2):271–80.
80. Borcard D, Gillet F, Legendre P. Numerical ecology with R. Volume 2. Springer; 2011.
81. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. *Mol Ecol.* 2015;24(17):4348–70.
82. Capblanq T, Luu K, Blum MG, Bazin E. Evaluation of redundancy analysis to identify signatures of local adaptation. *Mol Ecol Resour.* 2018;18(6):1223–33.
83. Jombart T, Pontier D, Dufour A-B. Genetic markers in the playground of multivariate analysis. *Heredity (Edinb).* 2009;102(4):330–41.
84. Selmoni O, Vajana E, Guillaume A, Rochat E, Joost S. Sampling strategy optimization to increase statistical power in landscape genomics: a simulation-based approach. *Mol Ecol Resour.* 2020;20(1):154–69.
85. De Mita S, Thuillet AC, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol Ecol.* 2013;22(5):1383–99.
86. Lotterhos KE, Whitlock MC. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol.* 2015;24(5):1031–46.
87. Manel S, Albert CH, Yoccoz NG. Sampling in landscape genomics. *Methods Mol Biology (Clifton NJ).* 2012;888:3–12.
88. Garnier E, Navas M-L, Grigulis K. Plant functional diversity: organism traits, community structure, and ecosystem properties. Oxford University Press; 2016.
89. Dove K. Kulturzonen Von Nord-Abessinien. *J. Perthes;* 1890.
90. MoA. Agro-ecological zonation of Ethiopia. In: Edited by Agriculture Mo; 2000.
91. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol.* 2017;37(12):4302–15.
92. Huang S, Zhang L, Rehman MU, Iqbal MK, Lan Y, Mehmood K, Zhang H, Qiu G, Nabi F, Yao W. High altitude hypoxia as a factor that promotes tibial growth plate development in broiler chickens. *PLoS ONE.* 2017;12(3):e0173698.
93. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Modell.* 2006;190(3–4):231–59.
94. Rasband WS. ImageJ 1.52a. in: National Institute of Health, USA; 1997.
95. Grimes SE. A basic Laboratory Manual for the small-scale production and testing of 1–2 Newcastle disease vaccine. 2002.
96. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
97. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv Preprint arXiv:12073907* 2012.
98. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011;27(21):2987–93.
99. Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. A spectrum of free software tools for processing the VCF variant call format: vcfliib, bio-vcf, cyvcf2, hts-nim and slivar. *PLoS Comput Biol.* 2022;18(5):e1009123.
100. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2(12):e190.
101. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904–9.
102. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet.* 2006;78(4):629–44.
103. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. The variant call format and VCFtools. *Bioinformatics.* 2011;27(15):2156–8.
104. Klimentidis YC, Aissani B, Shriver MD, Allison DB, Shrestha S. Natural selection among Eurasians at genomic regions associated with HIV-1 control. *BMC Evol Biol.* 2011;11(1):1–11.
105. Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature.* 2010;467(7315):587–90.
106. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ. Detecting recent positive selection in the human genome from haplotype structure. *Nature.* 2002;419(6909):832–7.
107. Gautier M, Klassmann A, Vitalis R. Reh2.0: a reimplementation of the R package reh2 to detect positive selection from haplotype structure. *Mol Ecol Resour.* 2017;17(1):78–90.
108. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics.* 2020;36(8):2628–9.
109. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H. Package 'vegan'. *Community Ecol Package Version.* 2013;2(9):1–295.
110. Rao CR. The use and interpretation of principal component analysis in applied research. *Sankhyā: Indian J Stat Ser A* 1964:329–58.
111. Van Den Wollenberg AL. Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika.* 1977;42(2):207–19.
112. Forester B. Detecting multilocus adaptation using Redundancy Analysis (RDA). *Population Genetics in R* Retrieved from [https://popgen.nesce.nt.org/2018-03-27\\_RDA\\_GEA.html](https://popgen.nesce.nt.org/2018-03-27_RDA_GEA.html). 2019.
113. Ye S, Gao N, Zheng R, Chen Z, Teng J, Yuan X, Zhang H, Chen Z, Zhang X, Li J. Strategies for obtaining and pruning imputed whole-genome sequence data for genomic prediction. *Front Genet.* 2019;10:673.
114. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75.
115. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Cárre G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ. Collinearity: a review of methods to deal

- with it and a simulation study evaluating their performance. *Ecography*. 2013;36(1):27–46.
116. Ramírez JD, Alvarez-Herms J, Castaneda-Babarro A, Larruskain J, de la Piscina XR, Borisov OV, Semenova EA, Kostyrykova ES, Kulemin NA, Andryushchenko ON. The GALNTL6 gene rs558129 polymorphism is associated with power performance. *J Strength Cond Res*. 2020;34(11):3031.
117. Zhao Y, Zhang Z, Liu L, Zhang Y, Fan X, Ma L, Li J, Zhang Y, He H, Kang L. Associations of high altitude polycythemia with polymorphisms in EPAS1, ITGA6 and ERBB4 in Chinese Han and Tibetan populations. *Oncotarget*. 2017;8(49):86736.
118. Berna-Erro A, Braun A, Kraft R, Kleinschnitz C, Schuhmann MK, Stegner D, Wulsch T, Eilers J, Meuth SG, Stoll G. STIM2 regulates capacitive Ca<sup>2+</sup> entry in neurons and plays a key role in hypoxic neuronal cell death. *Sci Signal*. 2009;2(93):ra67–67.
119. Burke B, Giannoudis A, Corke KP, Gill D, Wells M, Ziegler-Heitbrock L, Lewis CE. Hypoxia-induced gene expression in human macrophages: implications for ischemic tissues and hypoxia-regulated gene therapy. *Am J Pathol*. 2003;163(4):1233–43.
120. Sharpe MA, Baskin DS. Monoamine oxidase B levels are highly expressed in human gliomas and are correlated with the expression of HIF-1 $\alpha$  and with transcription factors Sp1 and Sp3. *Oncotarget*. 2016;7(3):3379.
121. McInerney-Leo AM, Sparrow DB, Harris JE, Gardiner BB, Marshall MS, O'Reilly VC, Shi H, Brown MA, Leo PJ, Zankl A. Compound heterozygous mutations in RIPPLY2 associated with vertebral segmentation defects. *Hum Mol Genet*. 2015;24(5):1234–42.
122. Cimmino F, Avitabile M, Lasorsa VA, Montella A, Pezone L, Cantalupo S, Visconte F, Corrias MV, Iolascon A, Capasso M. HIF-1 transcription activity: HIF1A driven response in normoxia and in hypoxia. *BMC Med Genet*. 2019;20:1–15.
123. Bouquerel P, Gstalder C, Müller D, Laurent J, Brizuela L, Sabbadini R, Malavaud B, Pyronnet S, Martineau Y, Ader I. Essential role for SphK1/S1P signaling to regulate hypoxia-inducible factor 2 $\alpha$  expression and activity in cancer. *Oncogenesis*. 2016;5(3):e209–209.
124. Crawford JE, Amaru R, Song J, Julian CG, Racimo F, Cheng JY, Guo X, Yao J, Ambale-Venkatesh B, Lima JA. Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. *Am J Hum Genet*. 2017;101(5):752–67.
125. Alam H, Maizels ET, Park Y, Ghaey S, Feiger ZJ, Chandel NS, Hunzicker-Dunn M. Follicle-stimulating hormone activation of hypoxia-inducible factor-1 by the phosphatidylinositol 3-kinase/AKT/Ras homolog enriched in brain (Rheb)/mammalian target of rapamycin (mTOR) pathway is necessary for induction of select protein markers of follicular differentiation. *J Biol Chem*. 2004;279(19):19431–40.
126. Kakinuma Y, Tsuda M, Okazaki K, Akiyama T, Arikawa M, Noguchi T, Sato T. Heart-specific overexpression of Choline Acetyltransferase Gene protects Murine Heart against Ischemia through Hypoxia-Inducible Factor-1 $\alpha$ -Related defense mechanisms. *J Am Heart Association*. 2013;2(1):e004887.
127. Rozario T, DeSimone DW. The extracellular matrix in development and morphogenesis: a dynamic view. *Dev Biol*. 2010;341(1):126–40.
128. Shimoda LA, Udem C. Interactions between calcium and reactive oxygen species in pulmonary arterial smooth muscle responses to hypoxia. *Respir Physiol Neurobiol*. 2010;174(3):221–9.
129. Wang Y-X, Zheng Y-M. ROS-dependent signaling mechanisms for hypoxic Ca<sup>2+</sup> responses in pulmonary artery myocytes. *Antioxid Redox Signal*. 2010;12(5):611–23.
130. Hui AS, Bauer AL, Striet JB, Schnell PO, Czyzyk-Krzeska MF. Calcium signaling stimulates translation of HIF- $\alpha$  during hypoxia. *FASEB J*. 2006;20(3):466–75.
131. Legendre P, Oksanen J, ter Braak CJ. Testing the significance of canonical axes in redundancy analysis. *Methods Ecol Evol*. 2011;2(3):269–77.
132. Perini F, Cendron F, Rovelli G, Castellini C, Cassandro M, Lasagna E. Emerging genetic tools to investigate molecular pathways related to heat stress in chickens: a review. *Animals*. 2020;11(1):46.
133. Porto-Neto LR, Reverter A, Prayaga KC, Chan EK, Johnston DJ, Hawken RJ, Fordyce G, Garcia JF, Sonstegard TS, Bolormaa S. The genetic architecture of climatic adaptation of tropical cattle. *PLoS ONE* 2014, 9(11).
134. Trujillo AV, Kabede A, Lozano-Jaramillo M, Dessie T, Smith J, Hanotte O, Gheyas A. Ecological niche modelling for delineating livestock ecotypes and exploring environmental genomic adaptation: the example of Ethiopian village chicken. *Front Ecol Evol* 2022:1–21.
135. Gebru G, Belay G, Vallejo-Trujillo A, Dessie T, Gheyas A, Hanotte O. Ecological niche modelling as a tool to identify candidate indigenous chicken ecotypes of Tigray (Ethiopia). *Front Genet* 2022, 13.
136. Lawal RA, Martin SH, Vanmechelen K, Vereijken A, Silva P, Al-Atiyat RM, Aljumaah RS, Mwacharo JM, Wu D-D, Zhang Y-P. The wild species genome ancestry of domestic chickens. *BMC Biol*. 2020;18(1):1–18.
137. Bettridge JM, Psifidi A, Terfa ZG, Desta TT, Lozano-Jaramillo M, Dessie T, Kaiser P, Wigley P, Hanotte O, Christley RM. The role of local adaptation in sustainable production of village chickens. In: *Nature Sustainability 1* (2018) 10 2018; 2018.
138. Lozano-Jaramillo M, Bastiaansen J, Dessie T, Komen H. Use of geographic information system tools to predict animal breed suitability for different agro-ecological zones. *Animal*. 2019;13(7):1536–43.
139. Alemu SW, Hanotte O, Kebede FG, Esatu W, Abegaz S, Bruno JE, Abrar B, Alemayehu T, Mrode R, Dessie T. Evaluation of live-body weight and the number of eggs produced for introduced and local chickens in Ethiopia. *Acta Agriculturae Scand Sect A—Animal Sci*. 2021;70(2):71–7.
140. Chen C-H, Wu BJ. Monoamine oxidase A: an emerging therapeutic target in prostate cancer. *Front Oncol*. 2023;13:1137050.
141. Xie L, Luo C, Zhang C, Zhang R, Tang J, Nie Q, Ma L, Hu X, Li N, Da Y. Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. *PLoS ONE*. 2012;7(2):e30910.
142. Liu R, Sun Y, Zhao G, Wang F, Wu D, Zheng M, Chen J, Zhang L, Hu Y, Wen J. Genome-wide association study identifies loci and candidate genes for body composition and meat quality traits in Beijing-You chickens. *PLoS ONE*. 2013;8(4):e61172.
143. Habimana R, Ngeno K, Okeno TO, Hirwa C, Keambou Tiambo C, Yao NK. Genome-Wide Association Study of Growth Performance and Immune Response to Newcastle Disease Virus of Indigenous Chicken in Rwanda. *Front Genet* 2021:1438.
144. Cha J, Choo H, Srikanth K, Lee S-H, Son J-W, Park M-R, Kim N, Jang GW, Park J-E. Genome-Wide Association Study Identifies 12 Loci Associated with Body Weight at Age 8 weeks in Korean native chickens. *Genes*. 2021;12(8):1170.
145. Wright D, Boije H, Meadows JR, Bed'Hom B, Gourichon D, Vieaud A, Tixier-Boichard M, Rubin C-J, Imsland F, Hallböök F. Copy number variation in intron 1 of SOX5 causes the pea-comb phenotype in chickens. *PLoS Genet* 2009, 5(6).
146. Luo W, Xu J, Li Z, Xu H, Lin S, Wang J, Ouyang H, Nie Q, Zhang X. Genome-Wide Association Study and Transcriptome Analysis Provide New Insights into the White/Red Earlobe Color formation in Chicken. *Cell Physiol Biochem*. 2018;46(5):1768–78.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.