Received: 28 November 2023 | Accepted: 29 November 2023

DOI: 10.1111/age.13387

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

Genetic basis of ear length in sheep breeds sampled across the region from the Middle East to the Alps

Jürgen Klawatsch[1,2] | Dimitris Papachristou[3] | Panagiota Koutsouli[3] | Maulik Upadhyay[1] | Doris Seichter[2] | Ingolf Russ[2] | Boro Mioč[4] | Mojca Simčič[5] | Iosif Bizelis[3] | Ivica Medugorac[1]

[1]Population Genomics Group, Department of Veterinary Sciences, Ludwig Maximillians University Munich, Munich, Germany

[2]Tierzuchtforschung e.V. Munich, Grub, Germany

[3]Department of Animal Science, Agricultural University of Athens, Athens, Greece

[4]Department of Animal Science and Technology, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia

[5]Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

Correspondence
Ivica Medugorac, Population Genomics Group, Department of Veterinary Sciences, Ludwig Maximillians University Munich, Munich, Germany.
Email: i.medugorac@lmu.de

Funding information
Deutscher Akademischer Austauschdienst, Grant/Award Number: BI-DE/20-21-003, DAAD ID-57418683 and DAAD ID-57514909; Public Service for Animal Genetic Slovenia; Tierzuchtforschung e.V. Munich

Abstract

Ear length in sheep (*Ovis aries*) shows a wide range of natural variation, from the absence of an outer ear structure (*anotia*), to small outer ears (*microtia*), to regular ear length. Up until now, the underlying genetics of this phenotype has been studied in four sheep breeds from China, Jordan and Italy. These studies revealed a broad range of genes significantly associated with ear length, potentially indicating genetic heterogeneity across breeds or geographic regions. In the current study, we performed genome-wide SNP genotyping and haplotype-based mapping, in a population of 340 individuals, to identify loci influencing ear length variation in additional sheep breeds from Slovenia, Croatia, Cyprus and Greece. Additionally, two previously described candidate variants were also genotyped in our mapping population. The mapping model without candidate variant genotypes revealed only one genome-wide significant signal, which was located next to *HMX1* on OAR6. This region was previously described as being associated with ear length variation in the Altay and Awassi sheep breeds. The mapping model including the candidate duplication genotype near *HMX1* as a fixed effect explained the phenotypic variance on OAR6 and revealed an additional genome-wide significant locus on OAR13 associated with ear length. Our results, combined with published evidence, suggest that a duplication in the evolutionarily conserved region near *HMX1* is the major regulator of ear length in sheep breeds descended from a larger region from Central Asia, to the Middle East, Cyprus, Greece and to the Alps. This distribution suggests an ancient origin of the derived allele.

KEYWORDS
ear length, genetic mapping, ruminants, sheep

INTRODUCTION

Natural variation of outer ear length in sheep (*Ovis aries*) was first characterised as a Mendelian trait (Ritzman, 1916) and clusters into three distinct categories (Lush, 1930). These categories include: (i) animals without an outer ear structure, referred to as *anotia*; (ii) animals with small ears, referred to as *microtia*; and (iii) animals with regular sized ears. The first description of a field observation on ear length variation described the 'earless' state as 'nearly straight lines running from the base and forming an abrupt, sharp point' (Ritzman, 1916). Later, the designation of an 'earless' state was further widened by describing it as ears visible

when the sheep has been shorn, but not easily seen if the sheep's fleece has grown for more than 2 months. For some breeds, efforts have been made to more clearly distinguish between the described three categories. An early paper on Karakul sheep (Lush, 1930) reported the maximum ear length of sheep with *anotia* as 5.0 cm and that of sheep with *microtia* as 12.1 cm. For comparison, that paper reported the minimum length of 'long-eared' sheep to be 15.4 cm, which is also comparable with the animals studied here.

In other species, such as humans, it has been discussed whether *anotia* and *microtia* constitute symptoms of syndromes and are thus associated with disease (Harris et al., 1996). However, in sheep, these phenomena have been described as isolated traits without pathological links. Since the proposition that ear length is a Mendelian trait, further genetic analyses of this and related traits have added a deeper complexity to the picture. For example, in cattle it has been shown that the gene regulatory network (GRN) of one keratinous outgrowth structure on the head (e.g. horn) may overlap with the GRN that controls a different keratinous outgrowth structure on the head (e.g. the eyelash-and-eyelid phenotype) (Allais-Bonnet et al., 2013). These results raise the possibility that a similar phenomenon could also apply to presence and length variation in sheep horns and ears. This possibility, taken with the fact that the genetic basis of *anotia* and *microtia* is still not fully understood, led us to further analyse the underlying genetics of this phenotype in this study.

In 2016, the first findings of the molecular basis on ear length variation were reported in Awassi sheep sampled in Jordan, southwest Asia (Jawasreh et al., 2016). In Awassi sheep, a single nucleotide polymorphism (SNP) at OAR23 in exon 7 of *GATA6* (GATA binding protein 6) was found to be significantly associated with ear length variation. In a subsequent study, another genome-wide significant SNP, on OAR23, at the *DCC* gene (deleted in colorectal carcinoma), was reported in Duolang sheep sampled in Xinjiang province, China (Gao et al., 2018). In another study, in Valle del Belice sheep of southern Italy, a mutation with suggestive significance was described at OAR1, located within intron 3 of the Clarin 1 (*CLRN1*) gene (Mastrangelo et al., 2018). More recently, He et al. (2020) identified a duplication in the evolutionary conserved region (ECR) close to *HMX1* (H6 family homeo-box 1) on OAR6, as a potential candidate associated with ear length variation in Altay sheep, which were also sampled in the Xinjiang province, China. He et al. (2020) defined a chromosomal region, beginning from position 114173 075 bp with a length of 317 bp, which is located upstream of *HMX1*, as an ECR because it was found to be highly conserved in an alignment of 12 species from mammals to *Xenopus tropicalis* to zebrafish. Another study found that the same duplication in the ECR of *HMX1* was present in the Awassi sheep breed from Jordan and was associated with ear length variation (Jawasreh & Al-Omari, 2020). Moreover, this study suggested a significant association and interaction between the duplication in the ECR near to *HMX1* (OAR6) and the missense mutation in exon 7 of *GATA6* (OAR23). Interestingly, a study (Si et al., 2020) reported that a duplication in the enhancer region of the *HMX1* locus is associated with a type of *microtia* in humans. Furthermore, a study of this trait in Highland Cattle, utilising SNP-chip genotyping and sequencing-based fine mapping, also yielded *HMX1* as a candidate locus (Koch et al., 2013).

Overall, the diversity of the studied breeds, their geographic locations and the mapping results of the previous studies, suggested a possible locus heterogeneity in the regulation of ear length across sheep breeds. To gain further insights to this question, here, we studied the genome-wide association on ear length in two Slovenian sheep breeds (the Bovec sheep and Improved Bovec sheep), Rab sheep from Croatia, Cyprus fat-tailed sheep and 22 Greek sheep strains, by utilising a combined linkage disequilibrium and linkage analysis (cLDLA).

METHODS

Animals, breeds, populations and phenotype recording

In this study, 340 animals from two Slovenian, one Croatian, one Cypriot and 22 Greek sheep populations were studied. These animals were distributed across several localities (n=45, Figure 1). The number of animals per flock ranged from 1 to 67 with a mean of 13.1 (Table 1). The Greek samples constitute a highly fragmented population because sampling was carried out across distant islands. These islands have a low exchange of animals between them and therefore the strains are isolated.

Phenotypes were recorded from all animals of the sampled populations. Since the same person did not examine all individuals, we employed a pre-designed and internally standardised phenotyping form and used a self-built, custom measurement instrument (see Figure S1). At the time of data collection, the 340 investigated animals (265 females and 75 males) were between 2 and 22 years old and their ear lengths spanned from about 2 to 17 cm (see Figures S2 and S3). The measured ear lengths allowed for a division of sheep phenotypes into three categories: (i) ear lengths of 2–5 cm; (ii) ear lengths of 6–10 cm; and (iii) ear lengths of 10–17 cm. In the early stage of this project, ear length was assessed visually in categories for some (~35.9%) animals. Later, the ear lengths of all remaining animals were quantitatively measured in centimetres (Figure S1). In order to maximise the mapping population, we converted the ear length categories from the early project phase into quantitative measures. For a precise conversion, we referred to

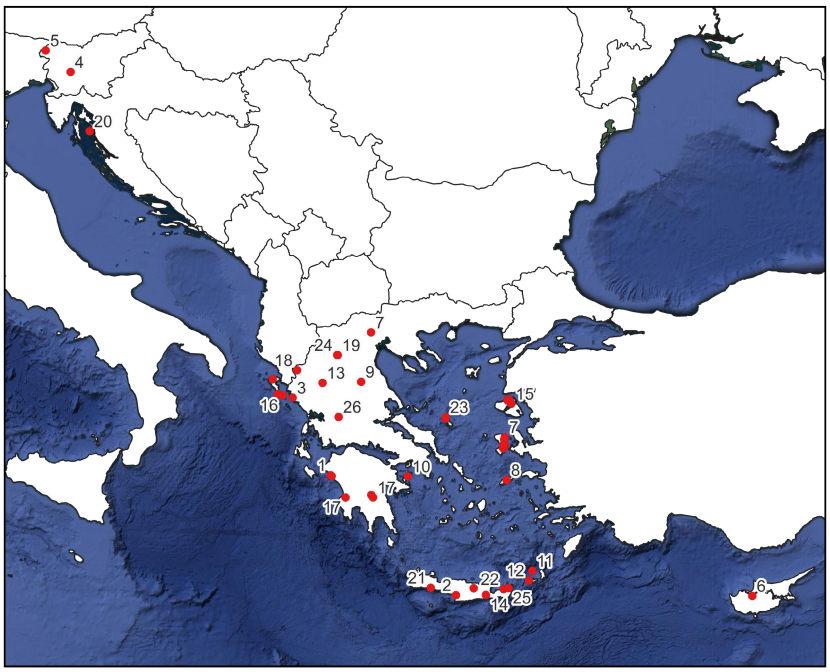

FIGURE 1 Geographic positions of the localisations from the 26 sheep breeds sampled in this study. The red dots designate one population each and belong to different breeds, indicated by position IDs given in Table 1. The map was created with QGIS version 3.22.7 (QGIS.org, 2021).

TABLE 1 Characterisation of the 26 studied sheep breeds with total number of animals per breed (c), number of animals with normal (n), *microtic* (m) and *anotic* (a) outer ears.

Position ID[a]	Breed name[b]	Breed abbreviation	c	a	m	n
1	Argos fat tailed	ARG	20	9	9	2
2	Asterousia Crete	AST	13	1	8	4
3	Boutsiko Epirus	BTS	6	0	0	6
4	Improved Bovec sheep	BOVI	34	0	0	34
5	Bovec sheep	BOVO	34	1	32	1
6	Cyprus fat tailed	CFT	14	0	0	14
7	Chios semi-fat tailed	HIOS	33	1	3	29
8	Ikaria semi-fat tailed	IKAR	4	0	0	4
9	Kalarrytiko	KLR	10	0	1	9
10	Kokovitiko	KOK	1	1	0	0
11	Karpathos	KRPT	15	1	3	11
12	Kasos	KSO	9	0	1	8
13	Katsika Ioannina	KTS	5	0	0	5
14	Lasithi Crete	LST	2	1	1	0
15	Semi-fat tailed	LESB	67	3	13	51
16	Lefkimmi Kerkyra	LEK	10	0	0	10
17	Local Peloponnese's (DRS, SKAR, SVRE)[b]	LPLS	13	0	5	8
18	Local Pogoni	LPG	1	0	1	0
19	Pelagonia	PLG	3	0	2	1
20	Rab	RAB	8	0	2	6
21	Sfakia Crete	SFK	12	2	9	1
22	Sitia Crete	SHT	2	0	0	2
23	Skyros	SKR	10	1	3	6
24	Sarakatsaniko	SRK	10	5	2	3
25	Toplu Crete	TOPS	2	0	0	2
26	Tsipoureika	TSP	2	0	1	1

[a] ID refers to the positions of the flocks for the respective breeds in the map (Figure 1).
[b] Population abbreviations are given in brackets.

images of categorised and quantitatively measured animals. The measured and derived quantitative ear lengths (Figure S1) were used in the genetic mapping models.

It is noteworthy that the proportion of females in our sample material is significantly higher than that of males. However, this higher proportion applies to all ear length categories in the same way (see Figure S3c) and simply results from the herd structure. This is clearly confirmed by a Mann–Whitney *U*-test for quantitative continuous ear length measurements between females and males ($p = 0.6962$). Therefore, the high proportion of females in our mapping design should not be taken as evidence for sex as an influencing factor on ear length.

Sampling of biological material, DNA isolation and SNP genotyping

Blood samples were drawn from all studied animals during routine examinations under good veterinary practice. Sampling took place in the period from 2018 to 2021 and samples were stored at −20°C until lab analysis. DNA was isolated using the ReliaPrep™ Blood gDNA Miniprep System (Promega, Fitchburg, MA, USA) according to the manufacturer's instructions. DNA concentration and the OD260/OD280 ratio was measured using a NanoDrop-2000 (Thermo Fisher Scientific, Wilmington, DE, USA). SNP genotyping was conducted on the OvineSNP50 BeadChip (Illumina, San Diego, CA, USA), covering 54 241 SNPs, using standard procedures (http://www.illumina.com). The positions of all markers, QTL, sequences and candidate genes in this work correspond to the Oar_4.0 reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2).

SNP quality control, haplotyping and unified additive relationships

Only individuals with a genotype call rate higher than 0.95 were retained in the data set. SNPs were excluded from the downstream analysis, if: (i) they mapped to unknown positions; (ii) they had a genotyping rate of less than 90%; and (iii) they had a minor allele frequency lower than 2.5%. Thus, after quality control, 340 animals and 46 733 SNPs remained in the dataset.

In order to impute missing genotypes and estimate haplotype phase, the BEAGLE software package (v 5.0) was used (Browning et al., 2018; Browning & Browning, 2007). For this step, the genotype and pedigree information of all available ovine data from an in-house database was used. Note that this database also contained animals that were not phenotyped for ear length and thus are not relevant for the further mapping analyses performed here but improve haplotype inference. These non-phenotyped animals are diverse, comprise different breeds and were previously utilised for the mapping of other traits and for phylogenetic analyses (Deniskova et al., 2018, 2019; Lagler et al., 2022; Luhken et al., 2016). The genetic structure of our mapping population was inspected with a principal component analysis. To account for the apparent population stratification and family structure, a genome-wide additive relationship was estimated between all pairs of animals, which was then merged into a unified additive relationship matrix. This was done with the method developed by Yang et al. (2010) using the implementation in the R package SNPREADY (Granato et al., 2018).

Genotyping of candidate variants at two specifically selected loci

As mentioned above, several broader genomic regions were found to be possibly associated with ear length in sheep. However, only two loci with known candidate causal variants have been described. These two candidate variants, located proximate to *HMX1* and *GATA6*, were genotyped in our entire mapping population. *HMX1* was described by He et al. (2020) and a significant interaction with *GATA6* was found in Jawasreh and Al-Omari (2020). In general, for the candidate SNP in *GATA6* Exon 7, we performed genotyping by PCR-RFLP and electrophoresis on a 1% agarose gel on all 340 sampled animals. The PCR primers are described in Jawasreh and Al-Omari (2020). The reaction mixtures used 2 μL of a prepared dilution of 25 ng/μL genomic DNA, 10 μM of each forward and reverse primer and 3 μL of 5× Colourless GoTaq® Reaction Buffer, and nuclease-free water was used to reach a final reaction volume of 15 μL. Thermal cycling conditions of 2 min denaturation at 95°C followed by 35 cycles of 30 s at 95°C, then 1 min annealing at 58°C and a 1 min extension at 72°C with a final elongation at 72°C for 10 min were applied. A 2 μL aliquot of PCR products was digested with the restriction enzyme StyI, in the following reaction mix: 2 μL of 1× rCutSmart™ Buffer, 0.2 μL of StyIHF (20 U/μl) (New England Biolabs, MA, USA) and 15.8 μL of water. The mixture was incubated for 3 h at 37°C. Finally, we separated the DNA fragments by size and visualised them by agarose gel electrophoresis (1% agarose gel in 1× TBE buffer) with GelRed as a dye. For the candidate variant close to *HMX1*, we also used the primers described in Jawasreh and Al-Omari (2020) and applied a similar PCR protocol as above, with the only difference being an annealing temperature of 63°C. Since the *HMX1* alleles differ in size (duplication of 76 bp), an agarose gel electrophoresis of the PCR products and genotyping by fragment size were performed directly.

Genetic mapping of loci affecting ear length

For genetic mapping, cLDLA (Meuwissen & Goddard, 2001) was performed as described previously

(Gehrke et al., 2020; Lagler et al., 2022). Briefly, ASREML (Gilmour et al., 2009) was used to simultaneously estimate the maximum likelihood, variance components, and fixed and random effects, based on a genome-wide additive relationship matrix and a locus-specific identity by descent (IBD) matrix. For the estimation of the locus-specific IBD matrix, a sliding window approach was used. Different window sizes had previously been tested, and it was concluded that with increased window size, likelihood ratio test (LRT) curves became smoother and less peaked. However, the computation time to estimate the IBD matrices increased considerably with increased window size (Gehrke et al., 2020). Thus, for this study, a window width of 40 SNPs was chosen as the optimal compromise. In general, the following mixed linear model was applied:

$$\mathbf{y} = X\boldsymbol{\beta} + Z_1\mathbf{u} + Z_2\mathbf{q} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of the ear length in cm; $\boldsymbol{\beta}$ is the vector of fixed effects including the overall mean (μ) age at phenotyping and genotypes of candidate genes (Models II, III and IV); $\mathbf{u}$ is the vector of n random polygenic effects for each animal; and $\mathbf{q}$ is the vector of random additive genetic effects of the putatively associated quantitative trait locus (QTL). Random residual effects were included in the vector $\mathbf{e}$. The matrices $\mathbf{X}$, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the incidence matrices for the fixed and random effects.

In total, four cLDLA models were fitted: Model I without genotypes of any candidate genes as fixed effects; Model II with *HMX1* genotypes as fixed effect; Model III with *GATA6* genotypes as fixed effect; and Model IV with *GATA6* and *HMX1* genotypes simultaneously as fixed effects. Finally, an LRT for the goodness-of-fit between the null hypothesis (H_0, model without random QTL effects) and the alternative hypothesis (H_1, model including QTL effects) at each SNP interval midpoint (flanked by 20 SNPs on the left and 20 SNPs on the right) was estimated. As a measure of significance, threshold LRT values, which were corrected for multiple testing at the chromosome-wide and genome-wide levels (p-values of 0.05), were introduced. The LRT values follow a χ^2 distribution with one degree of freedom (Heuven et al., 2005). The ratio of the number of non-overlapping (40 SNP) windows in a given chromosome or the genome to the window width, multiplied by the right-tailed probability from the χ^2 distribution of a given LRT value gave the corresponding p-value. Considering the SNP chip, our quality control parameters, our chromosomes of interest and the length of the entire genome, we set the chromosome-wide threhsold to an LRT value of 12.59 and the genome-wide significance threshold to 16.70. All maxima of the LRT curve (LRT_{max}) that passed the genome-wide significance threshold were considered signal peaks. Closely located LRT peaks were viewed as part of the same QTL according to the rules described by Muller et al. (2017), which also consider haplotype patterns in neighbouring peaks. For LRT_{max} values of each of the detected QTL, diplotype effects of the studied individuals were estimated and animals were ranked according to their diplotype effect. Here, the diplotype effect refers to (a) which haplotype pairs on homologous chromosomes contribute and (b) the extent to which they contribute (effect size; explained variability) to a given phenotype. The pipeline described above used GNU-PARALLEL (v. 2021) (Tange, 2011) to parallelise running of some intermediate steps.

Analysis of mapped QTL for gene presence

To select genomic regions to be analysed for the presence of putatively causative genes for natural ear length variation, confidence intervals (CIs) were calculated. A CI of 97% was obtained at the chromosomal positions between LRT values that crossed a threshold of one logarithm of the odds (LOD) score point lower than the LRT_{max} of a given LRT signal, with 1 LOD being equivalent to a likelihood-ratio-test statistic of 4.605 (Visscher & Goddard, 2004). Genes in these regions were identified from the UCSC Genome Browser (Oar_v4.0/oviAri4) (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2).

Diplotype effect analyses at significant QTL to examine primarily contributing animals to the diplotype effects

To estimate diplotype effects for each individual at significantly detected loci, we analysed ASREML outputs at marker intervals with the highest LRT value. To observe haplotype structure, potential haplotype clustering and the relationships of given haplotypes with the studied phenotype, the animals with their geographic information were sorted according to their diplotype effect. The presence of potential geographic clustering of haplotypes was examined by considering the geographic information of the animals with highest and lowest diplotype effects estimated by ASREML. Animals with the 5% most extreme diplotype effects per detected QTL were considered.

RESULTS

Genetic mapping analysis of outer ear length variation

The cLDLA, conducted without modelling any candidate loci as fixed effects (Model I), yielded one genome-wide significant mapping result (Figure 2a). The strongest signal with an LRT_{max}-value of 59.51 ($p = 1.41 \times 10^{-11}$) is on ovine chromosome 6 (OAR6) at position OAR6:114002547. The previously described candidate gene *HMX1* (He et al., 2020) is located 0.32 Mb

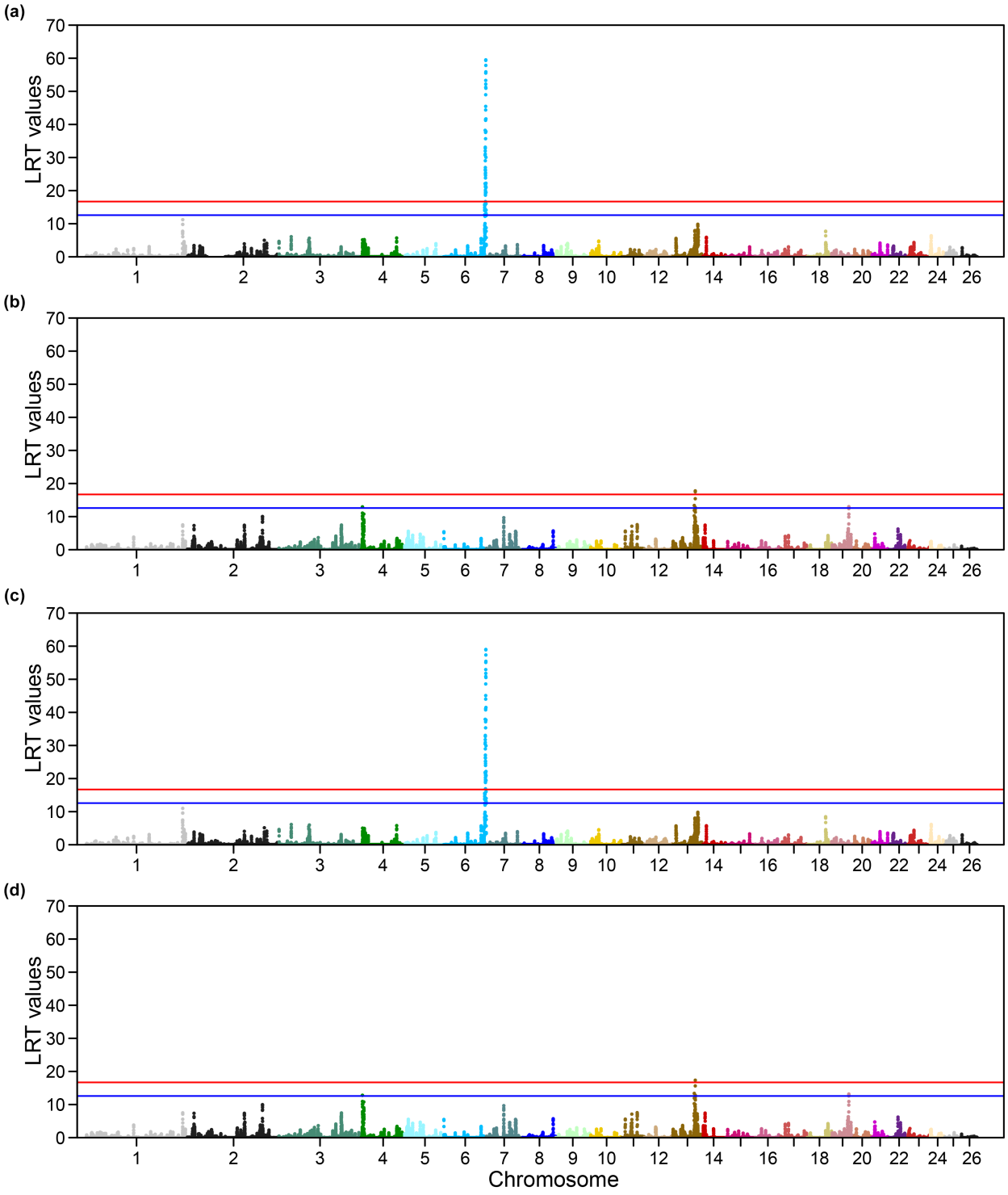

FIGURE 2 Genome-wide genetic mapping of outer ear length variation through combined linkage disequilibrium and linkage analysis (cLDLA). Genome-wide significance (red) and suggestive (chromosome-wide) significance (blue) thresholds are each indicated as straight lines. Likelihood ratio test (LRT) values of the different cLDLA models: (a) the result of model I, i.e. without any candidate gene genotypes as fixed effects; (b) the result of model II, with *HMX1* genotypes as fixed effect; (c) the result of model III with *GATA6* genotypes as fixed effect; and (d) the result of model IV with *GATA6* and *HMX1* genotypes simultaneously as fixed effects, are shown.

downstream of the position with the maximum LRT value. Unlike the previous studies, we did not identify any significant signals at the putative candidate gene *GATA6*, which is located on OAR23 (34.4 Mb) (Jawasreh et al., 2016), at the *DCC* gene (OAR23, 52.5 Mb) (Gao et al., 2018) or at the *CLRN1* gene (OAR1, 235.1 Mb) (Mastrangelo et al., 2018; Figure 2a). In Model II, the genotypes of the putative candidate gene *HMX1* (OAR6)

were included as fixed effects. Consequently, the comparison of the mapping results with Models I (Figure 2a) and II (Figure 2b) reflects the effects of the *HMX1* genotype. This comparison clearly shows that the peak on OAR6 disappears completely, while one peak on OAR13 (OAR13:60778138) increases above genome-wide significance with an LRT value of 17.77 (p=0.0285). In Model III, which considered the *GATA6* genotypes as a fixed effect, the signal on OAR6 seen in Model I remained, with slight variation in LRT values (Figure 2c). This indicates minimal or no influence of *GATA6*, as the difference between the Models I and III was the inclusion of the *GATA6* genotype as a fixed effect in the latter model.

Model IV fit the genotype of both putative candidate genes *GATA6* and *HMX1* (Figure 2d), which Jawasreh and Al-Omari (2020) found to be in significant epistatic interaction. The mapping results for Models II and IV were similar, with genome-wide significances at nearby positions on OAR13 (Model II, OAR13:60778138; Model IV, OAR13:60825335) and neither model showed a signal in the region of *GATA6*. There are nine genes located within the CI97 of the mapping signals on OAR13 (see Table S1).

The relationship between ear length, the genotyped duplication at the target locus of *HMX1* and breed was also analysed (Figure 3). This analysis of ear length, relative to the *HMX1* genotype allowed the observation of clustering of breeds from different geographic origins. Also, it clearly allows for the interpretation that the *HMX1* allele with the duplication is associated with *anotia* and the allele without the duplication is associated with regular ear length. Heterozygous individuals show *microtia*. The within-breed effect of the genotypes can be illustrated with breeds like LESB (Table 1), where all three genotypes of the *HMX1* duplication segregate (see Figure S5). The means of ear lengths with standard deviations do not show ear length overlaps across the genotype groups and thus indicate *HMX1* as the main effect on ear length more clearly. Benjamini–Hochberg corrected pairwise Mann–Whitney U tests show that ear lengths between the *HMX1* genotypes, both across breeds as well as within the example breed LESB, differ significantly from each other (Table S2).

Diplotype effect analysis of both significant QTLs and examination of the geographic origin of animals, primarily contributing to the diplotype effects

Diplotype effects were estimated for each individual at the significant QTL close to *HMX1*, from mapping Models I and III, and at the detected signal on chromosome 13, from mapping Models II and IV. Animals with haplotypes of strong negative diplotype effect (i.e. reducing ear length) in one locus do not show geographic clustering with individuals analysed at the second locus (Figure S6). Figure S6 shows the geographic origins of animals with the highest-ranking diplotype effects. The sheep with haplotypes showing the strongest negative diplotype effect at the OAR6 signal are shown in cyan and those at the OAR13 locus in yellow. It can be seen that no clear geographic pattern emerges, i.e. no specific spatial grouping of animals or breeds with strong effect on the QTL at OAR6 or OAR13 is visible.

DISCUSSION

Literature background, breeds, phenotypes and aim

The ear length phenotypes *anotia*, which is characterised by the absence of an outer ear structure and *microtia*, which is characterised by small outer ears,

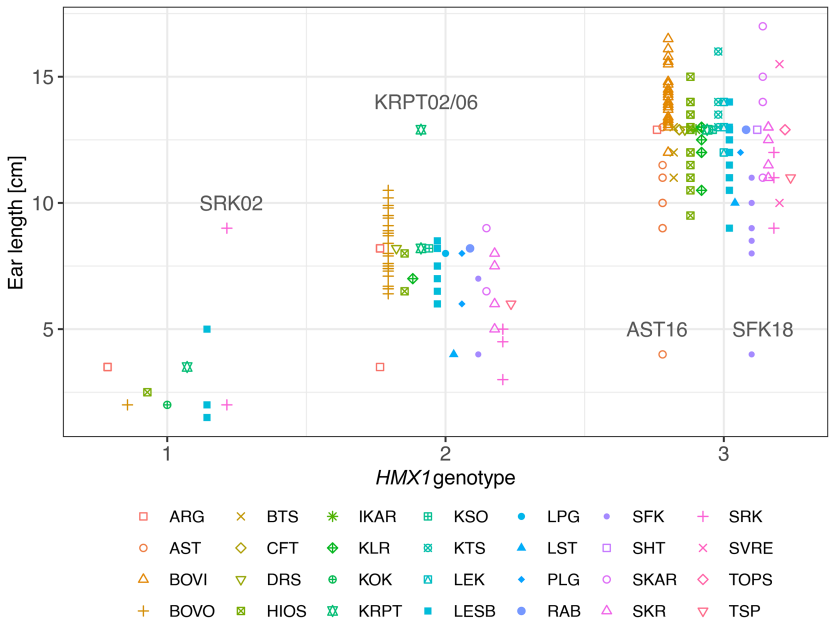

FIGURE 3 Ear lengths of studied animals depicted according to their *HMX1* genotype and colour coded based on their breeds. The numerically coded *HMX1* genotypes are: 1, homozygous duplication; 2, heterozygous; and 3, homozygous reference sequence. Extreme ear length values per genotype are labelled with in-house individual IDs.

have been seen in numerous Asian and European sheep breeds and previous genetic analyses found independently four candidate genes associated with this phenotype in sheep: *GATA6* (Jawasreh et al., 2016), *DCC* (Gao et al., 2018), *CLRN1* (Mastrangelo et al., 2018) and *HMX1* (He et al., 2020). Possible epistatic interactions between two of these genes were also postulated (Jawasreh & Al-Omari, 2020). In order to confirm the previously proposed variants and to look for new ones, we studied ear length variation in a wide range of local breeds, considering the whole genome as a potential candidate. This was done by genome-wide SNP genotyping in animals with diverse geographic, and presumably genetic, backgrounds. In addition to the strong stratification of the mapping population (Figure S4), sampling of related animals within herds also contributes to family structure. Thus, the four mixed QTL mapping models used here incorporate the covariance structure between individuals and thus account for both population stratification and family structure at the same time (see Price et al., 2010 for a review).

Genetic mapping models

The cLDLAs with four different models presented here detected one major locus and mapped one additional locus, masked by the major locus. Similar to the previous studies in Asian Altay and Awassi sheep (He et al., 2020; Jawasreh & Al-Omari, 2020), in the current study, the duplication close to *HMX1* was confirmed to be significantly associated with ear length variation in European breeds. Therefore, the same duplication is causal for ear length variation in sheep breeds from Central Asia (Altay), the Middle East (Awassi), South-East Europe (Greek and Cyprus breeds) and the Alps (Croatian and Slovenia breeds). This distribution suggests a very old mutation that had enough time for diffusion across vast distances eastwards to Central Asia and westwards to the Alps.

Genotyping of previously described candidate genes

To test the association of *GATA6* with *microtia* (Jawasreh et al., 2016) and a possible interaction between the candidate variants of *GATA6* and the duplication near *HMX1* (Jawasreh & Al-Omari, 2020), we genotyped these variants and included these as fixed effects in one of our mapping models. Based on the resulting LRT values, we can rule out a direct effect of *GATA6* or an epistatic interaction between variants at the two loci. However, Models I and III support a role for the candidate variant at *HMX1*, where the allele carrying the duplication close to *HMX1* is associated with shorter ears and the reference allele is associated with regular ear size. For each *HMX1* genotype group, our data show one or two individuals with noticeably deviating ear lengths compared with the other observations (Figure 3). The most pronounced outlier from the duplication homozygous group is SRK02, which has *microtia* typical ear length. This could be due to the presence of phenocopies, sampling errors or additional genetic variants somewhere in the genome, although we tried to avoid sampling errors in our data through rigorous quality control measures, e.g. photo series of each animal together with corresponding sample tubes and of completed forms.

Biological function and potential molecular mechanisms of *HMX1*

Data from our own and previous network regulatory studies (e.g. Boulling et al., 2013) indicate that HMX1 might act on higher hierarchical regulatory control structures in a given GRN, which assigns initial ear length fate in distinct groups, relative to which further downstream regulatory circuits can then act upon. This reasoning may further explain the variance within each *HMX1* genotype group. This is also in line with current characterisations of *HMX1* in the literature. *HMX1* is one of four members of the homeobox (HMX) family of transcription factors. The family is characterised by the presence of a homeobox domain consisting of 60 amino acids. The other three family members are *HMX2*, *HMX3* and sensory organ homeobox 1 (*SOHo1*) (Stadler et al., 1995). *Hmx1* is known to be expressed during mouse development in the trigeminal ganglion and in the second branchial arches as early as on embryonic day 9.5 (E9.5) and in the dorsal root ganglia at E10.5. Later in development, *Hmx1* is expressed in the lens, in the neural epithelium of the eye, in the sympathic and vagal nerve ganglia and in the mesenchyme near the developing ear (Quina et al., 2012; Yoshiura et al., 1998). Mice carrying mutations at *Hmx1* show microphthalmia, in addition to ear and cranial malformations (Munroe et al., 2009). We checked in our cLDLA analyses the signal strengths of variants in the previously described target genes of *HMX1*, which comprise *SLC6A9*, *EPHA6*, *SGCG*, *TSHZ2*, *PTPRO* and *SEMA3F* (Boulling et al., 2013), but none of these loci contain variants close to the suggestive significant threshold for association with ear length.

Characterisation of the detected OAR13 locus

The model comparisons performed here suggest that the major variant near *HMX1* masks at least one additional locus (OAR13) that influences variation in ear length in our mapping population. Two transcription

factors and two cell signalling-related genes were identified from the nine genes within the CI97 around the OAR13 signal. The remaining genes do not have functions clearly associated with development (Table S1). The transcription factors and cell-signalling-related genes can be viewed as potential candidate genes for ear length regulation. *FOXS1* (*Forkhead box protein S1*), for example, is known to be important for neural innervation (Heglind et al., 2005), which in turn is important for tissue growth and growth guidance. Furthermore, two other members of the large and complex '*forkhead box*' transcription factor family, *FOXL2* and *FOXC2*, are associated with horn development in ruminants (see Aldersey et al., 2020 and Aldersey, 2022) and are thus part of a gene regulatory network that affects a specific outgrowth structure on the head. Given the evolutionary conservation of the forkhead box transcription factor family (e.g. Laissue, 2019), *FOXS1* may also affect the development of head appendages. Regardless of this potential functional relevance of genes within the CI97 around the OAR13 signal discovered here, follow-up analyses are currently hampered by the diffuse haplotype structure in the candidate region. Indeed, large diversity of haplotypes at CI97 of OAR13 does not lead us to a clear identification of breeds, or group of animals, or haplotypes, that could be promising candidates for whole genome sequencing or other deeper genetic analyses.

Possible weakness of the study

One of the weaknesses of this study is the limited number of individuals with very short ears ($n=26$), i.e. the category from 2 to 5 cm, owing to the fact that *anotia* is a trait with a very low frequency across sheep breeds. Only certain herds within a breed show a moderate frequency of *microtia*, and within these herds there are sporadic animals with *anotia*. This makes it challenging to sample larger numbers of animals with this phenotype. Such comparable circumstances could be the reason why the animal set used in He et al. (2020) focuses on the difference between *microtia* and regular ear length. From the 129 Altay sheep studied by He et al. (2020), only three were homozygous for the *HMX1* duplication. Jawasreh and Al-Omari (2020) compared Awassi sheep with *anotia*, *microtia* and regular ear length and discovered mainly individuals that were either *HMX1* heterozygotes or homozygotes for the reference allele. In our set, 17 animals were homozygous for the *HMX1* duplication. In previous studies (Kunz et al., 2016; Lagler et al., 2022), we have shown that cLDLA is able to correctly map the causal mutation in the presence of phenocopies and in designs with a relatively small number of phenotyped animals.

The use of low-density SNP markers is also a weakness of the current study. Utilising high density SNP chips or whole genome sequencing would allow for a more precise genetic mapping. However, in previous studies (Lagler et al., 2022), we compared GWAS and cLDLA with comparable designs and could show that haplotype-based cLDLA is able to detect true linkage signals even in situations where GWAS approaches fail. Therefore, we are confident that the signals mapped here are not false positives and it is unlikely that we missed other causal loci with noticeable frequencies in our mapping population.

CONCLUSIONS

A duplication at the *HMX1* locus (OAR6) shows the strongest association with ear length in our study cohort of sheep. It was shown that *anotia* and *microtia* in sheep from Central Asia to the Alps are primarily associated with the same, presumably very old, genetic variant and that the two character states reflect different genotypes at the same locus. In addition, a region on OAR13 (including, amongst others, the *FOXS1* gene) showed a weaker association with the phenotype. Together, these findings help to explain the genetic basis of outer ear growth.

AUTHOR CONTRIBUTIONS

Jürgen Klawatsch: Formal analysis; investigation; visualization; writing – original draft. **Dimitris Papachristou:** Data curation; resources. **Panagiota Koutsouli:** Data curation; resources. **Maulik Upadhyay:** Formal analysis; writing – review and editing. **Doris Seichter:** Resources. **Ingolf Russ:** Resources. **Boro Mioč:** Data curation; resources. **Mojca Simčič:** Data curation; resources. **Iosif Bizelis:** Data curation; resources. **Ivica Medugorac:** Conceptualization; funding acquisition; project administration; resources; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

We acknowledge all breeders and breeding organisations that provided samples. We also want to acknowledge AMALTHIA – the Network for the Protection of Greek Indigenous Farm Animals – for its valuable help in collection of samples. We thank R. Damian of the laboratory of the Population Genomics Group for preparation of DNA samples used in the analyses. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

Authors are thankful to DAAD funding programmes for supporting our work through the project numbers BI-DE/20-21-003, DAAD ID-57514909 and DAAD ID-57418683. Further, we thank the Public Service for Animal

Genetic resources in Slovenia. J.K. was funded by a PhD Research Fellowship from the Tierzuchtforschung e.V. München.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request and through https://doi.org/10.6084/m9.figshare.24434533.

ETHICS STATEMENT

All samples were collected by trained personnel according to best veterinary practice and local legislation.

ORCID

Jürgen Klawatsch https://orcid.org/0009-0007-5244-7276
Ivica Medugorac https://orcid.org/0000-0002-6891-3039

REFERENCES

Aldersey, J.E. (2022) *Investigation into the genetic basis of bovine horn development*. Adelaide, SA: University of Adelaide.

Aldersey, J.E., Sonstegard, T.S., Williams, J.L. & Bottema, C.D.K. (2020) Understanding the effects of the bovine POLLED variants. *Animal Genetics*, 51, 166–176.

Allais-Bonnet, A., Grohs, C., Medugorac, I., Krebs, S., Djari, A., Graf, A. et al. (2013) Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. *PLoS One*, 8, e63512.

Boulling, A., Wicht, L. & Schorderet, D.F. (2013) Identification of *HMX1* target genes: a predictive promoter model approach. *Molecular Vision*, 19, 1779–1794.

Browning, B.L., Zhou, Y. & Browning, S.R. (2018) A one-penny imputed genome from next-generation reference panels. *American Journal of Human Genetics*, 103, 338–348.

Browning, S.R. & Browning, B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, 81, 1084–1097.

Deniskova, T., Dotsev, A., Lushihina, E., Shakhin, A., Kunz, E., Medugorac, I. et al. (2019) Population structure and genetic diversity of sheep breeds in the Kyrgyzstan. *Frontiers in Genetics*, 10, 1311.

Deniskova, T.E., Dotsev, A.V., Selionova, M.I., Kunz, E., Medugorac, I., Reyer, H. et al. (2018) Population structure and genetic diversity of 25 Russian sheep breeds based on whole-genome genotyping. *Genetics, Selection, Evolution*, 50, 29.

Gao, L., Xu, S.S., Yang, J.Q., Shen, M. & Li, M.H. (2018) Genome-wide association study reveals novel genes for the ear size in sheep (*Ovis aries*). *Animal Genetics*, 49, 345–348.

Gehrke, L.J., Capitan, A., Scheper, C., Konig, S., Upadhyay, M., Heidrich, K. et al. (2020) Are scurs in heterozygous polled (Pp) cattle a complex quantitative trait? *Genetics, Selection, Evolution*, 52, 6.

Gilmour, A., Gogel, B., Cullis, B. & Thompson, R. (2009) *ASReml user guide*. Hemel Hempstead: VSN International Ltd.

Granato, I.S.C., Galli, G., de Oliveira Couto, E.G., eSouza, M.B., Mendonça, L.F. & Fritsche-Neto, R. (2018) snpReady: a tool to assist breeders in genomic analysis. *Molecular Breeding*, 38, 102.

Harris, J., Kallen, B. & Robert, E. (1996) The epidemiology of anotia and microtia. *Journal of Medical Genetics*, 33, 809–813.

He, S., Zhang, Z., Sun, Y., Ren, T., Li, W., Zhou, X. et al. (2020) Genome-wide association study shows that microtia in Altay sheep is caused by a 76 bp duplication of *HMX1*. *Animal Genetics*, 51, 132–136.

Heglind, M., Cederberg, A., Aquino, J., Lucas, G., Ernfors, P. & Enerback, S. (2005) Lack of the central nervous system- and neural crest-expressed forkhead gene Foxs1 affects motor function and body weight. *Molecular and Cellular Biology*, 25, 5616–5625.

Heuven, H.C., Bovenhuis, H., Janss, L.L. & van Arendonk, J.A. (2005) Efficiency of population structures for mapping of Mendelian and imprinted quantitative trait loci in outbred pigs using variance component methods. *Genetics, Selection, Evolution*, 37, 635–655.

Jawasreh, K., Boettcher, P.J. & Stella, A. (2016) Genome-wide association scan suggests basis for microtia in Awassi sheep. *Animal Genetics*, 47, 504–506.

Jawasreh, K.I.Z. & Al-Omari, H.D. (2020) Duplication in ECR near *HMX1* and a SNP in *GATA6* genes regulate microtia in Awassi sheep. *Genes (Basel)*, 11, 597.

Koch, C.T., Bruggmann, R., Tetens, J. & Drogemuller, C. (2013) A non-coding genomic duplication at the HMX1 locus is associated with crop ears in highland cattle. *PLoS One*, 8, e77841.

Kunz, E., Rothammer, S., Pausch, H., Schwarzenbacher, H., Seefried, F.R., Matiasek, K. et al. (2016) Confirmation of a non-synonymous SNP in PNPLA8 as a candidate causal mutation for Weaver syndrome in Brown Swiss cattle. *Genetics, Selection, Evolution*, 48, 21.

Lagler, D.K., Hannemann, E., Eck, K., Klawatsch, J., Seichter, D., Russ, I. et al. (2022) Fine-mapping and identification of candidate causal genes for tail length in the Merinolandschaf breed. *Communications Biology*, 5, 918.

Laissue, P. (2019) The forkhead-box family of transcription factors: key molecular players in colorectal cancer pathogenesis. *Molecular Cancer*, 18, 5.

Luhken, G., Krebs, S., Rothammer, S., Kupper, J., Mioc, B., Russ, I. et al. (2016) The 1.78-kb insertion in the 3′-untranslated region of RXFP2 does not segregate with horn status in sheep breeds with variable horn status. *Genetics, Selection, Evolution*, 48, 78.

Lush, J.L. (1930) Earlessness in Karakul sheep. *Journal of Heredity*, 21, 107–112.

Mastrangelo, S., Sottile, G., Sutera, A.M., Di Gerlando, R., Tolone, M., Moscarelli, A. et al. (2018) Genome-wide association study reveals the locus responsible for microtia in Valle del Belice sheep breed. *Animal Genetics*, 49, 636–640.

Meuwissen, T.H. & Goddard, M.E. (2001) Prediction of identity by descent probabilities from marker-haplotypes. *Genetics, Selection, Evolution*, 33, 605–634.

Muller, M.P., Rothammer, S., Seichter, D., Russ, I., Hinrichs, D., Tetens, J. et al. (2017) Genome-wide mapping of 10 calving and fertility traits in Holstein dairy cattle with special regard to chromosome 18. *Journal of Dairy Science*, 100, 1987–2006.

Munroe, R.J., Prabhu, V., Acland, G.M., Johnson, K.R., Harris, B.S., O'brien, T.P. et al. (2009) Mouse H6 Homeobox 1 (Hmx1) mutations cause cranial abnormalities and reduced body mass. *BMC Developmental Biology*, 9, 27.

Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. (2010) New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11, 459–463.

QGIS.org. (2021) *QGIS geographic information system*. Chicago, IL: Open Source Geospatial Foundation Project.

Quina, L.A., Kuramoto, T., Luquetti, D.V., Cox, T.C., Serikawa, T. & Turner, E.E. (2012) Deletion of a conserved regulatory element required for Hmx1 expression in craniofacial mesenchyme in the dumbo rat: a newly identified cause of congenital ear malformation. *Disease Models & Mechanisms*, 5, 812–822.

Ritzman, E.G. (1916) Mendelism of short ears in sheep. *Journal of Agricultural Research*, 6, 797–798.
Si, N., Meng, X., Lu, X., Liu, Z., Qi, Z., Wang, L. et al. (2020) Duplications involving the long range HMX1 enhancer are associated with human isolated bilateral concha-type microtia. *Journal of Translational Medicine*, 18, 244.
Stadler, H.S., Murray, J.C., Leysens, N.J., Goodfellow, P.J. & Solursh, M. (1995) Phylogenetic conservation and physical mapping of members of the H6 homeobox gene family. *Mammalian Genome*, 6, 383–388.
Tange, O. (2011) GNU Parallel - The Command-Line Power Tool. *;login: The USENIX Magazine*, 36, 42–47.
Visscher, P.M. & Goddard, M.E. (2004) Prediction of the confidence interval of quantitative trait loci location. *Behavior Genetics*, 34, 477–482.
Yang, J., Benyamin, B., Mcevoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R. et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42, 565–569.
Yoshiura, K., Leysens, N.J., Reiter, R.S. & Murray, J.C. (1998) Cloning, characterization, and mapping of the mouse homeobox gene Hmx1. *Genomics*, 50, 61–68.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Klawatsch, J., Papachristou, D., Koutsouli, P., Upadhyay, M., Seichter, D., Russ, I. et al. (2024) Genetic basis of ear length in sheep breeds sampled across the region from the Middle East to the Alps. *Animal Genetics*, 55, 123–133. Available from: https://doi.org/10.1111/age.13387