254. Successful trio-based reverse genetic screen in an endangered local cattle breed

I.M. Häfliger[1*], F.R. Seefried[2] and C. Drögemüller[1]

[1]*Institute of Genetics, University of Bern, Bremgartenstrasse 109a, 3000 Bern, Switzerland;*
[2]*Qualitas AG, Chamerstrasse 56, 6300 Zug, Switzerland;* irene.haefliger@vetsuisse.unibe.ch

Abstract

We scanned the genomic data of almost one thousand Evolèner cattle, representing the by far smallest local cattle population of Switzerland, whereof only pedigree records are available. A reverse genetic screen using 94k SNP markers in 585 trios mapped recessive deleterious loci using homozygous haplotype deficiency. We found two haplotypes, EH1 and EH2, with moderate allele frequencies of >0.10. Mining for candidate variants was carried out by linkage analysis of the predicted haplotype status and whole-genome sequencing variant catalogue of seven Evolèner bulls. This led to the detection of two perfectly linked missense variants affecting conserved residues: *GBE1*:p.Arg437Gln for EH1, and *LRRC8A*:p.Ala73Val for EH2. No homozygous animals were observed in >5,100 cattle of various breeds including Evolèner. The presented study showed that very limited data can lead to the identification of candidate variants and thereby help to improve reproduction success in an endangered indigenous breed of cattle.

Introduction

The name of the Evolèner (EV) breed originates from the village called Evolèner in the municipality Val d'Hérens in the Valais; however, genetically the breed evolved from the local Swiss breed Eringer (in French Hérens). EV cattle are characterized by their small body size and by white markings on their foreheads, while their basic coat colour ranges from red to black. The breeds current population is spread across Switzerland and includes 1,175 animals (Dec 31, 2021), of which 823 animals are registered in the herdbook (A. Barenco, personal communication), whereof a proportion of 44% is genotyped. Together with genotypes from previous generations, almost 1000 genotype records were available based on a public SNP array encompassing 150k markers.

Recessive lethal variants had been identified by mapping haplotypes showing significant deviation from the Hardy-Weinberg equilibrium (HWE) in population-wide single nucleotide polymorphism (SNP) array data and subsequent whole-genome sequencing (WGS) data analysis of selected individuals (e.g. Schwarzenbacher *et al.*, 2016; VanRaden *et al.*, 2011). Usually, this approach is applied on large, economically important populations, where the SNP data is generated under the umbrella of genomic selection. Therefore, it is more challenging for smaller populations to identify harmful recessive alleles that impair fertility. Recently, population-wide SNP genotypes have been generated within the framework of a national genetic diversity project aiming at conserving the EV breed. Even though the power for haplotype detection was limited due to the relatively small dataset, the trio-based approach applied allowed the identification of two haplotypes showing significant depletion in homozygosity and the associated candidate causal variants affecting genes important for reproductive success.

Materials & methods

Genotype and pedigree data was provided by the breeding organisation (swissherdbook) and included all genotyped EV cattle born later than 2009 as well as their ancestors. Altogether 964 genotype records were available based on a publicly available SNP array encompassing 150K markers. The genomic positions relate to the latest cattle reference sequence ARS-UCD1.2.

https://www.wageningenacademic.com/doi/pdf/10.3920/978-90-8686-940-4_254 - Thursday, February 16, 2023 6:59:12 AM - Universität Bern IP Address:130.92.90.193
source: https://doi.org/10.48350/178889 | downloaded: 3.5.2024

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_254, © I.M. Häfliger *et al.* 2022

https://www.wageningenacademic.com/doi/pdf/10.3920/978-90-8686-940-4_254 - Thursday, February 16, 2023 6:59:12 AM - Universität Bern IP Address:130.92.90.193

SNP data analysis. As the initial dataset included different batches, genotype data was imputed using Fimpute v3 software (Sargolzaei *et al.*, 2014) with default parameters. SNP data was pruned before imputation on animal and on marker level. Samples were used if the sample call rate was higher than 0.95. Marker were kept in the dataset if each of the following criteria were achieved: minor allele frequency >0.01, call rate per SNP>0.99, and <1% mendelian errors. The final dataset contained 94,345 markers and 964 animals. Imputed genotypes were screened for haplotypes showing a deviation from HWE by depletion of homozygosity in a dataset that includes complete trios (sire, dam and offspring; n=585). Thereby a 50 SNP sliding window approach with overlap within the software snp1101 was applied (Sargolzaei, 2014). Within each region the most significant haplotype that also passed the Bonferroni correction ($P<1.68E^{-4}$) was used for further analysis in WGS data.

WGS data. For each of the selected haplotypes, diplotypes were predicted for the genotyped population to represent an individuals' haplotype status. Aiming at three carrier animals selected per haplotype, five bulls were chosen for whole-genome sequencing. WGS data was prepared for these five samples and additional 196 available samples from other projects as previously described (Häfliger *et al.*, 2021; data published in the European Nucleotide Archive: PRJEB18113). Further genomic information was provided by the international consortium of the 1000 Bull Genomes project run 9, which included 5,116 animals from various breeds (Hayes and Daetwyler, 2019). Finally, the combined WGS dataset encompassed 5,317 genomes, including eight Evolèner cattle.

Variant interpretation. A linkage disequilibrium (LD) approach was performed to identify haplotype-associated variants in WGS data by using plink v1.9 (Purcell *et al.*, 2007). LD (r2) was calculated using the diplotype for each haplotype and WGS-derived protein-changing genotypes for seven EV animals that had both types of genomic information available. For improving the understanding of associated variants, the effects of the amino acid changes were estimated using PROVEAN (Choi *et al.*, 2012) and conservation scores PhyloP and PhastCons (Pollard *et al.*, 2010; Siepel *et al.*, 2005).

Results

With the above described approach two different haplotypes were identified that showed significant deviation from HWE: Evolèner Haplotype 1 (EH1), on chromosome 1 between the positions 29,769,321 and 30,160,645 and Evolèner Haplotype 2 (EH2) on chromosome 11 from 97,965,320 to 99,366,886. EH1 and EH2 were expected to be observed in homozygous state in the analysed dataset in 13 and 16 animals, respectively. The haplotype frequency for EH1 and EH2 in EV was estimated to be >0.1, while a positive or negative trend was not observed over the last 14 year of births. Table 1 provides a summary of the two identified missense variants that show perfect LD ($r^2=1$) with the haplotypes.

Firstly, the EH1-associated variant in *GBE1* shows high conservation scores at DNA level (PhyloP=7.4; PhastCons=1) and the predicted amino acid exchange is most likely deleterious (PROVEAN score=-3.7), as it is situated in the functionally important catalytic domain of the GBE1 enzyme (Figure 1A). Secondly, the EH2-associated genomic variant in *LRRC8A* is not highly conserved across species (PhyloP=0.19;

Table 1. Candidate causal variants for deficiency of homozygotes in Evolèner.

Haplotype	Gene	OMIM	Transcript[1]	Coding DNA change	Predicted protein change
EH1	*GBE1*	607839	NM_001122729.1	c.1310G>A	p.Arg437Gln
EH2	*LRRC8A*	608360	NM_001076807.1	c.218C>T	p.Ala73Val

[1] According to the NCBI Annotation Release 106.

https://www.wageningenacademic.com/doi/pdf/10.3920/978-90-8686-940-4_254 - Thursday, February 16, 2023 6:59:12 AM - Universität Bern IP Address:130.92.90.193

PhastCons=0.002) and is predicted to be neutral (PROVEAN score=-0.49) on protein level, although the residue is located in the first extracellular loop of the LRRC8A protein (Figure 1B).

Both candidate variants were validated in 5,317 individuals with WGS data (Table 2). The EH1-associated *GEB1* variant could only be observed in EV animals, while the EH2-associated LRRC8A variant was observed in EV, as well as in Eringer and Gelbvieh cattle. These results lead to the conclusion that only the EH1-associated variant is breed specific, but both variants could not be observed in homozygous state in more than 5K animals.

Discussion

By applying a reverse genetic approach on a limited number of 595 completely genotyped trios, the SNP data of less than 1000 animals led to the identification of two homozygous deficient haplotypes, and the WGS data of seven animals pinpoint to two candidate causative variants. As far as we know, this is the first successful analysis based on such a limited dataset, concerning a small population for which not any phenotypic information is available. The EH1-related candidate variant in the essential *GBE1* gene, represents a most likely embryonic lethal mutation (Froese *et al.*, 2015). The gene is associated with the recessive, sometimes fatal, glycogen storage disorder type IV known in human, horse, and cat (OMIM: 607839; OMIA: 000420-9796, 000420-9685). Variants in the catalytic domain had been shown to destabilize the enzyme (Froese *et al.*, 2015). The exactly corresponding missense variant in the human *GBE1* gene is reported as rs770819412 and predicted to be likely disease causing; however, no homozygous variant carriers were observed and thereby no phenotype was assigned. These indications support our hypothesis, that we cannot observe

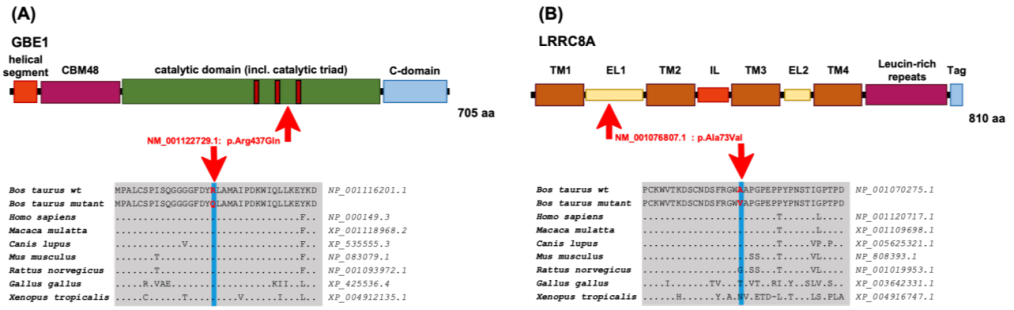

Figure 1. Protein domains and multiple species alignments of the amino acid sequences around the candidate variants. (A) The EH1-related variant in the *GBE1* gene (Froese *et al.*, 2015) and (B) the EH2-related variant in *LRRC8A* (Choi *et al.*, 2021). The abbreviated domains in LRRC8A mean transmembrane domain (TM), extracellular loop (EL), and intracellular loop (IL).

Table 2. Genotype distribution of the candidate variants in all whole-genome sequencing data, including Evolèner cattle.

	WGS data			
Variant	**ref/ref**	**ref/var**[1]	**var/var**	**./.**
EH1 (*GEB1*:c.1310G>A)	5,312	5	0	0
EH2 (*LRRC8A*: c.218C>T)	5,187	10	0	120

[1] Carrier animals of the EH1-associated variant are all of the breed EV; carrier animals of EH2-associated variant are of the breeds EV (4), Eringer (1), Gelbvieh (4), or unknown breed (1).

homozygous individuals or EH1-related cases, due to its embryonic lethal effect. The EH2-related candidate variant in *LRRC8A* gene, represents a likely candidate for a fatal postnatal phenotype. The protein is an essential compound of the volume-regulated anion channel, that maintains cellular homeostasis and is associated with the dominant disorder agammaglobulinemia 5 in human (OMIM: 608360; Choi *et al.*, 2021). In Lrrc8a knockout mice, increased pre- and postnatal mortality, growth retardation and tissue anomalies were observed (Kumar *et al.*, 2014). The extracellular loops were shown to be important oxidation sites during immune responses (Choi *et al.*, 2021). In conclusion, we identified two novel haplotypes and candidate variants potentially affecting fertility in the Evolèner population. Both protein-changing variants affecting candidate genes for developmental defects were never observed homozygous in more than 5,000 animals, and show perfect LD with the deficient homozygous haplotypes. Further steps include validation using farm questionnaires and *in vivo* data. Nevertheless, avoidance of risk matings can be immediately performed due to the ongoing comprehensive genotyping strategy in that population.

References

Choi H., Rohrbough J.C., Nguyen H.N., Dikalova A., and Lamb F.S. (2021) Oxidant-resistant LRRC8A/C anion channels support superoxide production by NADPH oxidase 1. J. Physiol. 599, 3013–3036. https://doi.org/10.1113/JP281577

Choi Y., Sims G.E., Murphy S., Miller J.R., and Chan A.P. (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS One 7, e46688. https://doi.org/10.1371/journal.pone.0046688

Froese D.S., Michaeli A., McCorvie T.J., Krojer T., Sasi M., *et al.* (2015) Structural basis of glycogen branching enzyme deficiency and pharmacologic rescue by rational peptide design. Hum. Mol. Genet. 24, 5667–5676. https://doi.org/10.1093/hmg/ddv280

Häfliger I.M., Seefried F.R., Spengeler M., and Drögemüller C. (2021) Mining massive genomic data of two Swiss Braunvieh cattle populations reveals six novel candidate variants that impair reproductive success. Genet. Sel. Evol. 53, 1–16. https://doi.org/10.1186/s12711-021-00686-3

Hayes B.J., and Daetwyler H.D. (2019) 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: Applications and outcomes. Annu. Rev. Anim. Biosci. 7, 89–102. https://doi.org/10.1146/annurev-animal-020518-115024

Kumar L., Chou J., Yee C.S.K., Borzutzky A., Vollmann E.H., *et al.* (2014) Leucine-rich repeat containing 8A (*LRRC8A*) is essential for T lymphocyte development and function. J. Exp. Med. 211, 929–942. https://doi.org/10.1084/jem.20131379

Pollard K.S., Hubisz M.J., Rosenbloom K.R., and Siepel A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 20, 110–121. https://doi.org/10.1101/gr.097857.109

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795

Sargolzaei M. (2014) SNP1101 User's Guide. Version 1.0. Guelph HiggsGene Solut. Inc.

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014). A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15, 478. https://doi.org/10.1186/1471-2164-15-478

Schwarzenbacher H., Burgstaller J., Seefried F.R., Wurmser C., Hilbe M., *et al.* (2016) A missense mutation in TUBD1 is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. BMC Genomics 17, 400. https://doi.org/10.1186/s12864-016-2742-y

Siepel A., Bejerano G., Pedersen J.S., Hinrichs A.S., Hou M., *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15, 1034–1050. https://doi.org/10.1101/gr.3715005

VanRaden P.M., Olson K.M., Null D.J., and Hutchison J.L. (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy Sci. 94, 6153–6161. https://doi.org/10.3168/jds.2011-4624

https://www.wageningenacademic.com/doi/pdf/10.3920/978-90-8686-940-4_254 - Thursday, February 16, 2023 6:59:12 AM - Universität Bern IP Address:130.92.90.193