Check for updates

Received: 22 December 2023 | Accepted: 11 March 2024

DOI: 10.1111/age.13421

SHORT COMMUNICATION

ANIMAL GENETICS WILEY

Population structure and genomic diversity of the Einsiedler horse

Abstract
The breeding history of the Einsiedler horse is closely connected with the Benedictine cloister Einsiedeln. In the mid-nineteenth century, it was decided to use European Warmblood stallions for cross-breeding and to abandon the selection of stallions. Since that time, it has only been possible to trace back the origin of Einsiedler horses using maternal ancestry information. Here, we collected high-density genotype data for European Warmblood horses (Selle Français, Swiss Warmblood and Einsiedler) and Franches-Montagnes horses, the last native Swiss horse breed, to unravel the current population structure of the Einsiedler horse. Using commonly applied methods to ascertain fine-scale population structures, it was not possible to clearly differentiate the Einsiedler from other European Warmblood horses. However, by means of runs of homozygosity (ROH) we were able to detect breed-specific ROH islands for the Einsiedler horse, including genes involved in domestication and adaptation to high altitude. Therefore, future breeding activities should involve the screening of these breed-specific ROH segments, the revival of cryopreserved sperm and the selection of Einsiedler stallions.

The Benedictine cloister Einsiedeln, located in central Switzerland, is considered the oldest stud farm in Europe with a proven written record, first mentioned AD 1064 (Frei, 2015). Through its geographical location at the beginning of the Gotthard Pass it became an important transport hub for goods from the northern Germanic regions to southern Italy and created a high demand for resilient horses (Frei, 2015). Initially, horses from all over Europe passed through the region, thus the origin of the Einsiedler (ES) horse is not clear. Previously, it has been demonstrated that ES mares shared some mitochondrial haplotypes with baroque breeds (Andalusians and Lusitanos), highly influential in European breeding during the Renaissance (Riggenbach et al., 2005). To date, the selection of ES horses is solely based on pedigree information, and the mares are classified into three maternal lines, namely Klima, Quarta and Sella (Weiss & Waldmeier, 2015). Hence, the ES is currently not recognised as a domestic horse breed by the Swiss authorities, while the Franches-Montagnes (FM) is described as the last native horse breed in Switzerland.

To assess the current population structure and genome diversity of the ES horse, we collected Axiom™ Equine genotype data (Schaefer et al., 2017; Schaefer & McCue, 2020) for 604 horses including 52 ES and 152 Swiss Warmblood (WB) horses, as well as 200 unrelated Selle Français (SF) and FM, respectively. For the subsequent analyses we excluded single nucleotide polymorphisms (SNPs) positioned on the sex chromosome or with unknown position resulting in 602 131 genome-wide SNPs according to the EquCab3.0 reference genome (Kalbfleisch et al., 2018) and the most recent annotation file (https://www.thermofisher.com/order/catalog/product/550583). Furthermore, missing genotype data were phased with BEAGLE 5.2 (Browning et al., 2018) separately for each breed. For the population structure analysis, we further excluded SNPs with a minor allele frequency <5% ensuing in 421 988 genome-wide SNPs.

To assess the population structure of the horses, we calculated pairwise F_{ST} estimates between breeds and allele sharing distances between individuals using the *fst* and *distance* function as implemented in PLINK v1.9 (Chang et al., 2015), while SPLITSTREE (www.splitstree.org) was used to create the respective neighbour joining trees (Kloepper & Huson, 2008). Furthermore, we applied a previously described three-step approach, which combines admixture and genomic inbreeding (F_{ROH}) of individuals into a high-resolution network visualisation. The different components involved in this approach are described in detail by Neuditschko et al. (2017) and Druml et al. (2018). Briefly, we computed the admixture levels of each horse using the program ADMIXTURE 1.23 (Alexander et al., 2009) increasing the number of clusters (K) from 1 to 10. To determine to optimal number of clusters, cross-validation error estimation was performed for each K. Furthermore, we derived supervised admixture proportions of ES horses, using WB, SF and FM horses as reference individuals. Supervised (K=3) and unsupervised admixture results increasing K from 2 to 6 were visualised with the program DISTRUCT 1.1 (Rosenberg, 2004) and integrated in the high-resolution network visualisation. The F_{ROH} were calculated by dividing the total length

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2024 The Authors. *Animal Genetics* published by John Wiley & Sons Ltd on behalf of Stichting International Foundation for Animal Genetics.

of runs of homozygosity (ROH) segments (S_{ROH}) by the length of the autosomal genome (L_{AUTO}), which was set to 2280.92 Mb; see the description below. Finally, we performed high-resolution network visualisations of the horses using NetView (Neuditschko et al., 2012; Steinig et al., 2016). Therefore, we used the aforementioned allele sharing distances between horses and applied the algorithm in its default setting (number of k nearest neighbours = 10). To illustrate the genetic relatedness between neighbouring horses, we associated the thickness of edges (connecting lines) with the magnitude of the genetic distance, while thicker edges corresponded to lower genetic distances. To identify highly inbred and outcrossed horses within the respective population networks we scaled the node size of each horse based on the individual F_{ROH}, while the node colour represents the admixture level at the given number of K clusters, respectively.

Runs of homozygosity segments were determined with an overlapping window approach implemented in PLINK v1.9 (Chang et al., 2015). The following settings were applied: a minimum SNP density of one SNP per 50 kb; a maximum gap length of 100 kb; and a minimum length of homozygous segments of 500 kb (including more than 80 homozygous SNPs), while one heterozygote SNP was permitted per segment. The total number of ROH (N_{ROH}), the total length of ROH segments (S_{ROH}) and the average length of ROH (L_{ROH}) were summarised for each breed. Differences between breeds were investigated using ANOVA, and *post hoc* Tukey's honestly significant difference tests at a significance level of $\alpha < 0.05$ as implemented in the R package MULTCOMPVIEW (Graves et al., 2015). Putative ROH islands were determined based on overlapping homozygous regions, shared by more than 50% of studied

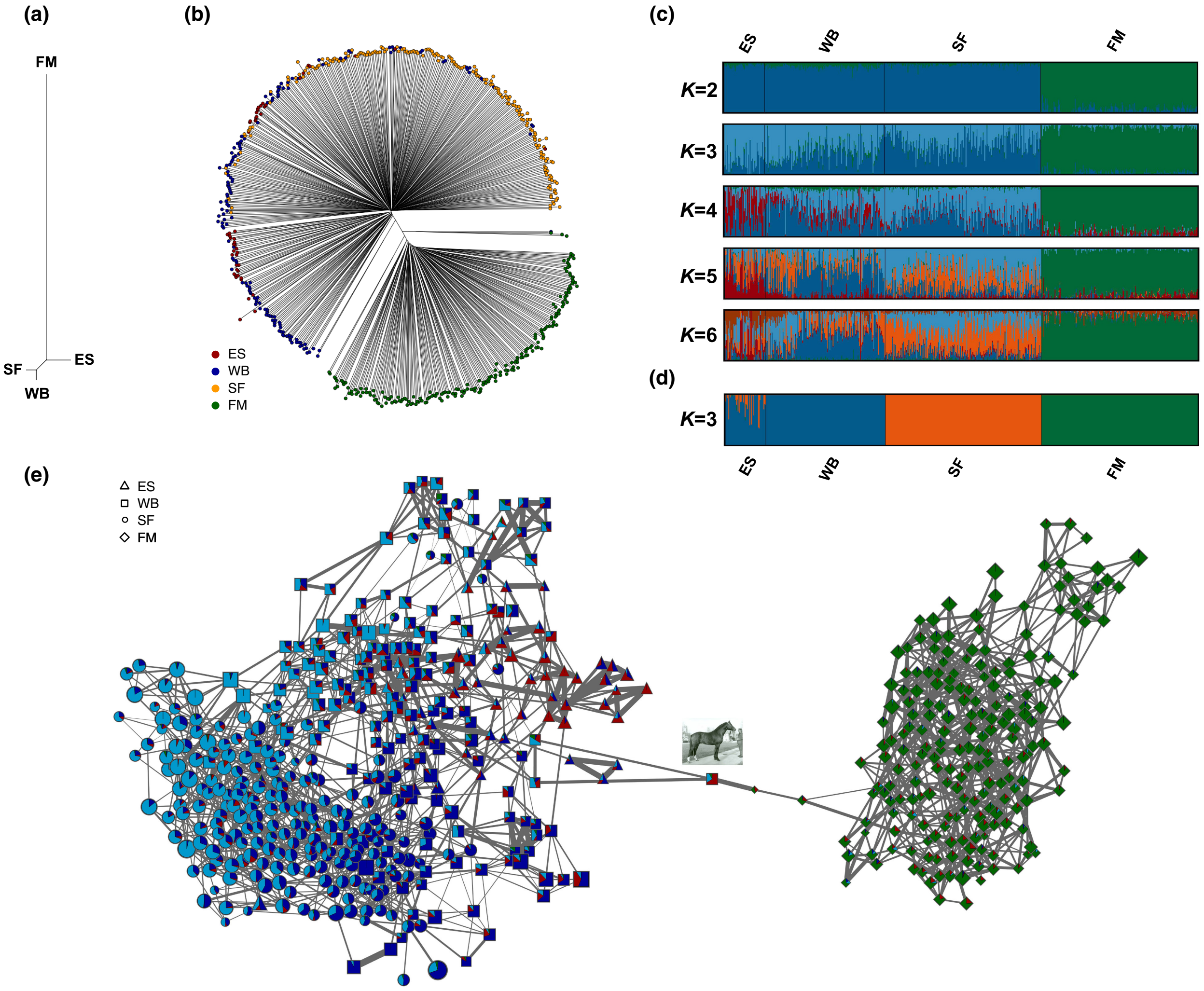

FIGURE 1 Population structure results of the sampled horses. (a) UPGMA (unweighted pair group method with arithmetic mean) tree of pairwise F_{ST} distances between breeds. (b) Neighbour joining tree of allele sharing distances between horses. (c) Unsupervised clustering results of horses increasing the number of clusters (K) from 2 to 6. (d) Supervised clustering result at $K=3$, using Swiss Warmblood (WB), Selle Français (SF) and Franches-Montagnes (FM) horses as reference individuals. (e) High-resolution population network, where each horse is illustrated by a node, with individual node size proportional to F_{ROH}, while the node shape refers to the breed origin and the node colour represents the individual levels of admixture at $K=4$.

13652052, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13421 by Schweizerische Akademie Der, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

horses within each breed and were identified with the R package DETECTRUNS (Biscarini et al., 2019). Runs of homozygosity islands which occurred only in one single breed were defined as 'breed-specific' or 'private'. Resulting ROH islands were checked for overlaps between ES and the other breeds, whereas a threshold of two SNPs for the minimal overlapping length size was applied. We used the NCBI genome data viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/) and the reference genome assembly EquCab3.0 (Kalbfleisch et al., 2018) to identify genes embedded in ROH islands.

The lowest F_{ST} distance between breeds was observed for WB and SF (0.005) followed by WB and ES (0.013), while FM horses showed the highest pairwise F_{ST} distances to all other breeds (Figure 1a). The topology of the neighbour joining tree coincides with the F_{ST} result and clearly affirms that the Warmblood horses cannot be differentiated from each other according to their genetic origin (Figure 1b). Increasing the K from 1 to 10, the cross-validation error estimation did not provide an optimal cluster solution (Figure S1). The first level of model-based clustering (K=2) FM horses were differentiated from the other three breeds (Figure 1c). At K=3 the three Warmblood breeds (ES, WB and SF) were further sub-structured without a discernible pattern associated with the origin of the horses. At the next level (K=4) the majority of ES horses were allocated in a distinct population cluster, while at additional levels of clustering (K=5 and K=6), further substructures were detected within the Warmblood horses. Supervised clustering at K=3 revealed a distinct admixture pattern of ES horses, simultaneously highlighting ES horses clustering with WB horses (Figure 1d). Integrating the unsupervised admixture proportions at K=4 into the high-resolution network visualisation, it can be noted that the presumed ES cluster was associated with family structures within the ES data (Figure 1e). In this context, it was also interesting to see that the WB stallion Noé, building the hub between FM and the Warmblood breeds, also falls into this cluster, as this stallion was used for cross-breeding within the FM breed and also had ES ancestry (Poncet & Wermeille, 2009; Weiss & Waldmeier, 2015). Concerning the F_{ROH} we did not observe a distinctive pattern of ES horses, while highly inbred horses were identified within FM and SF (Figure 1e). On average, SF had slightly more ROH segments (~15) compared with the other three breeds (Table 1). Selle-Français and FM had longer total ROH segments and higher F_{ROH} than ES and WB, while the mean segment length (L_{ROH}) of FM significantly differed from those in the other breeds. In total we identified three ROH islands within the ES, whereas one was partially shared with SF (Table S1). The most notable gene in a private ROH island for ES was *ZFPM1*. This gene is associated with mood regulation, aggressive behaviour, anxiety behaviour and contextual fear memory

TABLE 1 Total number of runs of homozygosity (N_{ROH}), total length of runs of homozygosity (ROH) segments (S_{ROH}), average length of ROH (L_{ROH}) and genomic inbreeding coefficients (F_{ROH}) for the Einsiedler, Swiss Warmblood, Selle Français and Franches-Montagnes.

Breed	*n*	Mean	SD	Median	Minimum	Maximum
Einsiedler	52					
N_{ROH}		185.80^a	22.72	182.50	108.00	257.00
S_{ROH} (Mb)		242.90^a	63.38	244.60	83.30	416.70
L_{ROH} (Mb)		1.29^a	0.24	1.28	0.77	1.83
F_{ROH} (%)		12.44^a	3.08	12.23	2.44	23.69
Swiss Warmblood	152					
N_{ROH}		179.70^a	24.00	178.00	128.00	280.00
S_{ROH} (Mb)		244.80^a	55.57	243.30	115.30	431.40
L_{ROH} (Mb)		1.36^a	0.22	1.35	0.90	2.33
F_{ROH} (%)		10.73^a	2.44	10.67	5.06	18.91
Selle Français	200					
N_{ROH}		200.70^b	26.43	202.50	105.00	301.00
S_{ROH} (Mb)		298.94^b	78.26	289.01	85.47	599.49
L_{ROH} (Mb)		1.44^b	0.29	1.44	0.81	2.56
F_{ROH} (%)		13.11^b	3.43	12.67	3.75	26.28
Franches-Montagnes	200					
N_{ROH}		177.50^a	29.05	177.00	69.00	318.00
S_{ROH} (Mb)		283.71^b	70.34	279.00	55.59	540.38
L_{ROH} (Mb)		1.60^c	0.33	1.60	0.81	2.84
F_{ROH} (%)		12.44^b	3.08	12.23	2.44	23.69

Note: $^{a–c}$ not sharing any letter indicates significant difference between breeds based on the Tukey honestly significant difference test at the 5% level of significance.

13652052, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13421 by Schweizerische Akademie Der, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and therefore considered important for domestication (Librado et al., 2021) and selection of ES horses. The second gene, *WNK1*, embedded in a shared ROH island for ES and SF, was previously associated with familial hyperkalemic hypertension (Vidal-Petiot et al., 2013) and adaptation to high altitudes in Tibetan people (Shi et al., 2018). Furthermore, *WNK1* was associated with cold tolerance in Tibetan cattle (Liu et al., 2023). This gene was also found in ROH islands for the Bosnian Mountain and Posavina horse (Grilz-Seger et al., 2019), supporting the hypothesis that ES horses are adapted to mountainous regions of higher altitudes. However, this gene is also linked with skeletal muscle hypertrophy (Mandai et al., 2018) in horses, which is of high importance for the performance of sports horses, like SF.

Based on the high-resolution population network visualisation we were able to demonstrate that putative sub-structures, as defined by model-based clustering, can be associated with family structures. Therefore, the current ES horse cannot be considered as a distinct breed. However, we also identified ROH islands for ES horses involving genes associated with domestication and adaptation to higher altitudes. To re-establish the ES breed we suggest screening putative candidates for breed-specific ROH segments and initiating a selection programme for stallions, including the revival of cryopreserved sperm.

KEYWORDS

admixture, horse, inbreeding, population network, runs of homozygosity

AUTHOR CONTRIBUTIONS

Annik Gmel: Conceptualization; formal analysis; funding acquisition; investigation; methodology; writing – original draft. **Anne Ricard:** Data curation; writing – review and editing. **Vinzenz Gerber:** Data curation; writing – review and editing. **Markus Neuditschko:** Conceptualization; data curation; funding acquisition; methodology; resources; software; supervision; visualization; writing – original draft.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of Esther Weiss from the Stiftung zur Förderung der Einsiedler Marstallzucht – Für das Einsiedler Pferd for organising hair sampling of the Einsiedler horses, and the Fondation Sur-La-Croix for funding part of the genotyping for the Franches-Montagnes horses. Open access funding provided by Agroscope.

FUNDING INFORMATION

This study was funded by the Swiss Federal Office for Agriculture under contract number 625000469 and Fondation Sur-La-Croix (internal contract number 6510263).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

Anonymised data are available for academic purposes on signing a material transfer agreement.

ETHICS STATEMENT

This study was conducted on data sampled under permits VD3527b, VD2976.1 and VD2227.2, each approved by the cantonal veterinary office of Vaud, Switzerland.

Annik Gmel[1,2]
Anne Ricard[3]
Vinzenz Gerber[4]
Markus Neuditschko[1]

[1] *Animal GenoPhenomics, Agroscope, Posieux, Fribourg, Switzerland*
[2] *Equine Department, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland*
[3] *Institut National de la Recherche Agronomique, Domaine de Vilvert, Jouy-en-Josas, France*
[4] *Vetsuisse Faculty, University of Bern, Bern, Switzerland*

Correspondence
Markus Neuditschko, Animal GenoPhenomics, Agroscope, Posieux, Fribourg, Switzerland.
Email: markus.neuditschko@agroscope.admin.ch

ORCID

Annik Gmel https://orcid.org/0000-0002-6259-2377
Markus Neuditschko https://orcid.org/0000-0001-7824-701X

REFERENCES

Alexander, D.H., Novembre, J. & Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664.

Biscarini, F., Cozzi, P., Gaspa, G. & Marras, G. (2019) *detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes*, in *the R project*, CRAN (the comprehensive R archive network).

Browning, B.L., Zhou, Y. & Browning, S.R. (2018) A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, 103(3), 338–348.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 7.

Druml, T., Neuditschko, M., Grilz-Seger, G., Horna, M., Ricard, A., Mesarič, M. et al. (2018) Population networks associated with runs of homozygosity reveal new insights into the breeding history of the Haflinger horse. *Journal of Heredity*, 109(4), 384–392.

Frei, T. (2015) Einsiedeln als Pferdezentrum der Innerschweiz. In: Frei, T. et al. (Eds.) *Das Kloster Einsiedeln und seine Pferde*. Einsiedeln: Schwyzer Hefte, pp. 15–26.

13652052, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13421 by Schweizerische Akademie Der, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Graves, S., Piepho, H.-P. & Selzer, M.L. (2015) *Package 'multcompView'*, in *visualizations of paired comparisons*, CRAN (the comprehensive R archive network).

Grilz-Seger, G., Neuditschko, M., Ricard, A., Velie, B., Lindgren, G., Mesarič, M. et al. (2019) Genome-wide homozygosity patterns and evidence for selection in a set of European and near eastern horse breeds. *Genes (Basel)*, 10(7), 491.

Kalbfleisch, T.S., Rice, E.S., DePriest, M.S., Jr., Walenz, B.P., Hestand, M.S., Vermeesch, J.R. et al. (2018) Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications Biology*, 1(1), 197.

Kloepper, T.H. & Huson, D.H. (2008) Drawing explicit phylogenetic networks and their integration into SplitsTree. *BMC Evolutionary Biology*, 8, 1–7.

Librado, P., Khan, N., Fages, A., Kusliy, M.A., Suchan, T., Tonasso-Calvière, L. et al. (2021) The origins and spread of domestic horses from the Western Eurasian steppes. *Nature*, 598(7882), 634–640.

Liu, X., Chen, J., Xu, X., Liu, J., Zhang, J., Cheng, H. et al. (2023) A missense mutation of the WNK1 gene affects cold tolerance in Chinese domestic cattle. *Animal Biotechnology*, 34, 1–6.

Mandai, S., Mori, T., Nomura, N., Furusho, T., Arai, Y., Kikuchi, H. et al. (2018) WNK1 regulates skeletal muscle cell hypertrophy by modulating the nuclear localization and transcriptional activity of FOXO4. *Scientific Reports*, 8(1), 9101.

Neuditschko, M., Khatkar, M.S. & Raadsma, H.W. (2012) NetView: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PLoS One*, 7(10), e48375.

Neuditschko, M., Raadsma, H.W., Khatkar, M.S., Jonas, E., Steinig, E.J., Flury, C. et al. (2017) Identification of key contributors in complex population structures. *PLoS One*, 12(5), e0177638.

Poncet, P.-A. & Wermeille, V. (2009) *Le cheval des Franches-Montagnes à travers l'histoire*. Porrentruy: Société jurassienne d'émulation.

Riggenbach, C., Stranzinger, G., Poncet, P.A., Glowatzki, M.L., Muntwyler, J., Gaillard, C. et al. (2005) Molekulargenetische Untersuchungen beim Einsiedler-Pferd. *Schweizer Archiv für Tierheilkunde*, 147(5), 195–204.

Rosenberg, N.A. (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137–138.

Schaefer, R.J. & McCue, M.E. (2020) Equine genotyping arrays. *Veterinary Clinics: Equine Practice*, 36(2), 183–193.

Schaefer, R.J., Schubert, M., Bailey, E., Bannasch, D.L., Barrey, E., Bar-Gal, G.K. et al. (2017) Developing a 670k genotyping array to tag~2M SNPs across 24 horse breeds. *BMC Genomics*, 18(1), 565.

Shi, R., Li, J., He, J., Meng, Q., Qian, Z., Shi, D. et al. (2018) Association of with-no-lysine kinase 1 and serine/threonine kinase 39 gene polymorphisms and haplotypes with essential hypertension in Tibetans. *Environmental and Molecular Mutagenesis*, 59(2), 151–160.

Steinig, E.J., Neuditschko, M., Khatkar, M.S., Raadsma, H.W. & Zenger, K.R. (2016) Netview p: a network visualization tool to unravel complex population structure using genome-wide SNP s. *Molecular Ecology Resources*, 16(1), 216–227.

Vidal-Petiot, E., Elvira-Matelot, E., Mutig, K., Soukaseum, C., Baudrie, V., Wu, S. et al. (2013) WNK1-related familial hyperkalemic hypertension results from an increased expression of L-WNK1 specifically in the distal nephron. *Proceedings of the National Academy of Sciences*, 110(35), 14366–14371.

Weiss, E. & Waldmeier, E. (2015) Einsiedler Pferde gestern – heute – morgen. In: Frei, T. et al. (Eds.) *Das Kloster Einsiedeln und seine Pferde*. Einsiedeln: Schwyzer Hefte, pp. 15–26.

SUPPORTING INFORMATION

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

13652052, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13421 by Schweizerische Akademie Der, Wiley Online Library on [13/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License