

CHARACTERIZATION OF THE (IMMUNO)GENOME DIVERSITY IN DOMESTIC, WILD AND ANCIENT HYBRID OLD WORLD CAMELIDS



PhD thesis submitted to fulfil the requirement for the academic degree of

Doctor of Philosophy (PhD)

From the University of Veterinary Medicine Vienna

Submitted by

Sara Ribeiro Barbosa Almendra Lado, BSc MSc



From the Research Institute of Wildlife Ecology, Department of Interdisciplinary Life
Sciences, University of Veterinary Medicine, Vienna.

Vienna, June 2021

PhD Committee

1st supervisor

Dr. Pamela A. Burger
Research Institute of Wildlife Ecology
Department of Interdisciplinary Life Sciences
University of Veterinary Medicine, Vienna
Vienna, Austria
pamela.burger@vetmeduni.ac.at

2nd supervisor

Prof. Dr. Petr Hořín
Institute of Animal Genetics, CEITEC
University of Veterinary and Pharmaceutical Sciences
Brno, Czech Republic
horinp@vfu.cz

3rd supervisor

Dr. José Melo-Ferreira
CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos
InBIO Laboratório Associado
Universidade do Porto
Vairão, Portugal
jmeloferreira@cibio.up.pt

List of publications

Article 1. **Lado, S.**, Elbers, J. P., Duskocil, A., Scaglione, D., Trucchi, E., Banabazi, M. H., Almathen, F., Saitou, N., Ciani, E. & Burger, P. A. (2020). Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes. *Communications Biology*, 3(1), 1-8. <https://doi.org/10.1038/s42003-020-1098-7>

Impact factor: 4.165

Article 2. **Lado, S.**, Elbers, J. P., Rogers, M. F., Melo-Ferreira, J., Yadamsuren, A., Corander, J., Petr Horin & Burger, P. A. (2020). Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies. *BMC Genomics*, 21(606), 1-17. <https://doi.org/10.1186/s12864-020-06990-4>

Impact factor: 3.7

Article 3. **Lado, S.**, Elbers, J.P., Plasil, M., Loney, T., Weidinger, P., Camp, J. V., Kolodziejek, J., Futas, J., Kannan, D. O., Orozco-terWengel, P., Horin, P., Nowotny, N., Burger, P. A. (2021). Innate and adaptive immune genes associated with MERS-CoV infection in dromedaries. *Cells*, 10(6), 1291. <https://doi.org/10.3390/cells10061291>

Impact factor: 4.366

Article 4. **Lado, S.**, Elbers, J.P., Kilimci, F. S., Kara, M. E., Dabanoğlu, I., Hurk, Y., Brongers, T., Grigson, C., Lev-Tov, J., McClure, S., Davoudi, H., Mohaseb, A., Baker, P., Kühne, H., Kreppner, J., Haring, E., Berthon, R., Peters, J., Mashkour, M., Burger, P. A., Çakırlar, C. (*in preparation*). Hidden hybrids – detecting early hybridization between dromedary and Bactrian camels in a culture-historical context.

- Extra publication from an external collaboration during the PhD:

Mahtani-Williams, S., Fulton, W., Desvars-Larrive, A., **Lado, S.**, Elbers, J. P., Halpern, B., Herczeg, D., Babocsay G., Lauš B., Tamás Z., Jablonski D., Kukushkin O., Orozco-terWengel P., Vörös J. & Burger, P.A. (2020). Landscape Genomics of a Widely Distributed Snake, *Dolichophis caspius* (Gmelin, 1789) across Eastern Europe and Western Asia. *Genes*, 11(10), 1218. <https://doi:10.3390/genes11101218>

Impact factor: 3.3

Public outreach during the PhD

- 1) Invitation to write a “Behind the paper” text on *Nature Research Ecology & Evolution* blog after the *Article 1* was published in *Communications Biology*, entitled “Camel - the animal of the past, present and future” (in English).
<https://natureecoevocommunity.nature.com/posts/camel-the-animal-of-the-past-present-and-future>

- 2) Five articles were published in two German newspapers after *Article 1* was published in *Communications Biology* (Pamela Burger was interviewed, in German):
 - a. “Kamele statt Rinder” in *Berliner Zeitung* Nr. 21, Tuesday, January 26th 2021.
<https://www.berliner-zeitung.de/gesundheit-oekologie/kamele-statt-rinder-dromedare-sind-fuer-den-klimawandel-gut-geruestet-li.132152?pid=true>

 - b. “Die Zukunft der Wüstenschiffe” in *Schwäbische Zeitung* (Natur & Umwelt) Saturday, February 6th 2021.
https://www.schwaebische.de/ueberregional/panorama_artikel,-die-zukunft-der-wuestenschiffe-_arid,11324807.html

 - c. “Die Dromedare kommen” in *Stuttgarter Zeitung* Nr. 41, Friday, February 19th.
<https://www.stuttgarter-zeitung.de/inhalt.auswirkungen-des-klimawandels-welche-rolle-dromedare-in-zukunft-spielen-koennten.7cd506f5-79a7-4a3b-85d9-40ff4f570d92.html?reduced=true>

 - d. “Die Karawane zieht weiter” in *Kölner Stadt-Anzeiger*, Monday, February 15th (not available online).

 - e. “Ein alter Freund wird wieder gebraucht” in *Südkurier* Nr. 54, Saturday, March 6th (not available online).

- 3) Biology-related Podcast (“Trio.logia”) interview to talk about my academic background and research on camels (in Portuguese).
<https://open.spotify.com/episode/35eUZ7KnI3S8zpNvZs37vH>

TABLE OF CONTENTS

1. ACKNOWLEDGEMENTS	5
2. AUTHOR CONTRIBUTIONS.....	7
3. DECLARATION.....	9
4. SUMMARY	10
5. GENERAL INTRODUCTION.....	15
<i>References.....</i>	<i>23</i>
6. PUBLICATIONS.....	30
6.1 Article 1	30
6.2 Article 2	54
6.3 Article 3	78
6.4 Article 4	120
7. DISCUSSION & CONCLUSION.....	151
<i>References.....</i>	<i>163</i>
8. CURRICULUM VITAE	170

1. ACKNOWLEDGEMENTS

What a challenge to have the final year of the PhD during a global pandemic! Nevertheless, I have to thank my family, my boyfriend Sebastian and my friends (around the world) for all the support and for always believing in me. They have never failed me and I will be eternally grateful to them. Fábio e Filipa, seremos sempre o melhor trio.

I would like to thank Pamela for all the support, flexibility, (patience!) and opportunities she has offered me as my supervisor. Thanks to her, I was able to hold oral presentations at six conferences in different continents, to participate at IAEA-UNO meetings, attend workshops, as well as having collaborations with different external researchers. Not all supervisors invest in their students like Pamela did, and I am truly grateful for that. Also, as Pamela once said to me, “another camelologist is ‘born’”, and in big part, this passion I have now is because of the opportunities she provided me. I would also like to thank my two co-supervisors, Petr and José, for all the support and guidance throughout the whole PhD period. Their knowledge and help were invaluable. To Jean, I cannot thank him enough for all the guidance, support and friendship. Without a “co-supervisor” official title, he was a mentor and my growth as a researcher was also potentiated by him. I thank as well all my co-authors - it was a pleasure working with you all! To the Genetics Lab people, thank you so much for having that great environment and for all the genetic discussions we had. It was great sharing the lab with you all, as well as my “eureka!” (and “not so eureka”) PhD moments.

Last, but not least, I want to thank my (working) friends which I will keep for life, for all the support and good moments we have spent together throughout these three and a half years: Julia, Filipa, Ortal, Anouck, Ivan, Sara, Armando, Andrea and Otto, you have contributed to my PhD graduation success and to my wonderful life in Vienna. I would also like to thank all other working colleagues (both scientific and non-scientific staff), essentially everyone working at KLIVV and FIWI, for making me feel so welcome during this period. I am very grateful to have shared knowledge and experiences with such wonderful people that I will forever remember, and at the same time being so privileged to have worked in such a beautiful place, up the hill.

2. AUTHOR CONTRIBUTIONS

Article 1. **Lado, S.**, Elbers, J. P., Duskocil, A., Scaglione, D., Trucchi, E., Banabazi, M. H., Almathen, F., Saitou, N., Ciani, E. & Burger, P. A. (2020). Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes. *Communications Biology*, 3(1), 1-8. <https://doi.org/10.1038/s42003-020-1098-7>

S.L. performed analysis and wrote the first draft of the paper, J.P.E., A.D., D.S., E.T., E.C., and P.B. performed analyses, F.A., N.S., and M.H.B. contributed essential samples, E.C. and P.B. conceived and managed the project and wrote the paper. All authors provided valuable discussions, commented, and approved the final paper.

Article 2. **Lado, S.**, Elbers, J. P., Rogers, M. F., Melo-Ferreira, J., Yadamsuren, A., Corander, J., Petr Horin & Burger, P. A. (2020). Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies. *BMC Genomics*, 21(606), 1-17. <https://doi.org/10.1186/s12864-020-06990-4>

S.L. wrote the first draft of the manuscript, J.P.E. and M.F.R. performed analyses, P.H. and P.A.B. conceived and managed the project, J.M.F and J.C. revised the manuscript. All authors interpreted the results, provided valuable discussions, commented and approved the final manuscript.

Article 3. **Lado, S.**, Elbers, J.P., Plasil, M., Loney, T., Weidinger, P., Camp, J. V., Kolodziejek, J., Futas, J., Kannan, D. O., Orozco-terWengel, P., Horin, P., Nowotny, N., Burger, P. A. (2021). Innate and adaptive immune genes associated with MERS-CoV infection in dromedaries. *Cells*, 10(6), 1291. <https://doi.org/10.3390/cells10061291>

S.L. wrote the first draft of the manuscript. S.L., J.P.E. and M.P. analysed the data. S.L., P.B., D.O.K., N.N. and T.L. collected samples at the livestock market, UAE. T.L. assessed the camel chip demographic information. S.L., J.V.C., P.W. and J.K. performed wet lab work. P.O.W. and J.F. provided analytic methods and discussion on the data. P.B., P.H. and N.N. conceived and managed the project. SL and PB wrote the manuscript and all authors provided valuable discussions and approved the final manuscript.

*Article 4. **Lado, S.**, Elbers, J.P., Kilimci, F. S., Kara, M. E., Dabanoğlu, I., Hurk, Y., Brongers, T., Grigson, C., Lev-Tov, J., McClure, S., Davoudi, H., Mohaseb, A., Baker, P., Kühne, H., Kreppner, J., Haring, E., Berthon, R., Peters, J., Mashkour, M., Burger, P. A., Çakirlar, C. (*in preparation*). Hidden hybrids – detecting early hybridization between dromedary and Bactrian camels in a culture-historical context.*

S.L. performed wet lab work and wrote the first draft of the paper, I.D. extracted DNA from Aydin modern samples, S.L. and J.P.E. analyzed genetic data, C.C and P.B. conceived and managed the project. F.S.K., M. E. K., YvdH, T.B., C.G., J.L-T., S.C., H.D., A. M. P. B. H.K., J. K., R.B., contributed with essential samples. S.L., C.C. and P.B. wrote the manuscript. All authors provided valuable discussions.

3. DECLARATION

I, Sara Ribeiro Barbosa Almendra Lado, hereby declare that this PhD thesis, with the title “Characterization of the (immuno)genome diversity in domestic, wild and ancient hybrid Old World camelids” was carried out by myself as a candidate for graduation with the degree of Doctor of Philosophy in Science, under the supervision of Dr. Pamela Burger, Research Institute of Wildlife Ecology, Department of Interdisciplinary Life Sciences, University of Veterinary Medicine, Vienna; Prof. Petr Hořín, Institute of Animal Genetics, CEITEC, University of Veterinary and Pharmaceutical Sciences, Brno, Czech Republic; and Dr. José Melo-Ferreira, CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal. I confirm that this PhD thesis was done by the best of my abilities and in harmony with the rules, in all aspects, of Good Scientific Practice according to the faculty guideline and practice in the area.



Sara Lado

March 2021

4. SUMMARY

Background

Old World camels consist of three distinct species: one-humped dromedary (*Camelus dromedarius*), two-humped Bactrian camel (*Camelus bactrianus*), and the critically endangered two-humped wild camel (*Camelus ferus*). The two domestic species, the dromedary and the Bactrian camel, as well as dromedary-Bactrian camel hybrids (a human influenced crossing), are valuable animals since ancient times, not only for their production traits, but also for their power as prime vehicles of short- and long-distance caravan trade. Ancient trading routes, i.e., the Incense Route and Silk Road, acted as corridors of gene flow shaping genetic diversity and population structure in camels. Especially dromedary-Bactrian camel hybrids were highly valued in caravan trade, as crossbreds are stronger, more robust, handling harsher climates, and can have higher production traits compared to the any of the parental species. Yet, the starting point of this human-induced hybridisation practice is still unknown, although previous suggestions hint to pre-Roman times.

As camels have unique adaptations to diverse and extreme environments, they are in contact with different pathogenic pressures in diverse surroundings. Thus, there is not only great interest in understanding the general neutral and adaptive genetic diversity, but also the variation present in those parts of the genome, which encode the immune system. But camels are not just well adapted to harsh conditions – they can be resistant to devastating infections, which threaten other livestock species cohabiting the same regions (e.g., foot and mouth disease, in dromedaries). On the other hand, camels are potential reservoirs of (zoonotic) diseases, and due to an increased demand and consumption of camel meat and milk followed by a higher contact with the animals, they represent a significant source for disease transmission to humans, i.e., the Middle East Respiratory Syndrome (MERS). Moreover, the immune system of camels shows unique features, such as a special type of antibodies (nanobodies), somatic hypermutations in T-cell receptor genes or low major histocompatibility complex (MHC) polymorphisms, which makes them a special immunogenetic model.

Although camels have accompanied humans since ancient times, and their importance as livestock species is rising in view of increasing desertification and global climate change,

the scientific knowledge in genetic diversity on these organisms is still rather incomplete. At the start of this thesis, genetic inferences had either been based on few individuals or on a small number of (microsatellite) markers. Plus, fully annotated, chromosome-level assembled genomes (state-of-the-art in other livestock species) were missing, which also hampered the implementation of up-to-date methodology for large scale, genome-wide diversity or association studies. As such, a broader genetic approach using modern genetic tools was required to assess and characterize the global genetic diversity in camels, including immune response regions of the genome, as well as to understand their role in response to pathogens. This would have a great impact for camel research and camel husbandry, as genome-wide studies are necessary to study and understand specific genomic regions underlying economically important traits, as well as to understand the impact of camel-associated zoonotic diseases on public health and economy.

Aims of the thesis

On these grounds, the main objectives of this thesis were to fill existing knowledge gaps on camel (immuno)genetic diversity using a representative number of individuals and up-to-date methodology. For this, the patterns of (immuno)genetic diversity in Old World camelids were characterized using next generation sequencing, taking advantage of newly improved high-quality genome assemblies. Another aim of this thesis was to dissect the beginning of camel hybridisation within a zooarchaeological-paleogenetic framework. For this, camel bones from as early as Iron Age showing mixed-morphology were analysed with paleogenetic techniques and radiocarbon dating, together creating the foundation to understand early camel hybridisation in a culture-historical context.

Results

Previous studies on camel genetic diversity used only a limited number of microsatellite markers and/or mtDNA and detected only weak population structure in the global dromedary population. **The first article (“Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes”. *Communications Biology* 2020; 3:387)** of my thesis aimed at overcoming the challenges of low number of markers and investigating human-induced migration patterns, population structure and diversity in the global dromedary population applying a genome-wide approach. To that end, double-digest restriction-site associated DNA sequencing (ddRADseq) was used to

detect fine-scale population differentiation in dromedaries across Asia and Africa. Global patterns of effective migration rates revealed pathways of dispersal after domestication, following known ancient caravan routes.

Previously published genome assemblies from Old World camelids, which have been very useful for investigating genome-wide diversity, demography or population structure, show inconsistencies and gaps that limit analyses especially in repetitive and more complex genomic regions. Improved and more accurate genome assemblies and annotations were therefore needed, not only for large-scale genomic assessment, but also to study complex genomic regions such as adaptive and innate immune response (IR) genes. **My second article (“Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies”. *BMC Genomics* 2020; 21:606)** aimed at improving the existing genome assemblies of the three Old World camel species via different computational methods. These upgraded assemblies were then used as basis to assess nucleotide diversity of IR genes within and between species, and to compare the diversity found in immune genes with the rest of the genome. Differences in the nucleotide diversity were detected among the three Old World camelid species and between IR gene groups, being compatible with a combined role of population history and differential exposures to pathogens, and consequent different selective pressures.

In this context, emerging zoonotic diseases pose a serious threat not only to isolated animal populations, but also to humans. One recent example of an emerging *Coronaviridae* zoonotic pathogen, which lately has been receiving more attention due to the COVID-19 pandemic, is the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). This pathogen, which belongs to the beta-coronaviruses like the Severe Respiratory Syndrome (SARS) CoV-1/-2, has been identified in dromedaries from the Middle East that act as reservoir for transmission to humans. Just recently, also Bactrian camels and dromedary-Bactrian camel hybrids have been identified as potential carriers for this pathogen. Although some information is available on MERS-CoV prevalence, epidemiology, genetic diversity, and etiopathology in dromedaries and humans, little is known about the immune responses of camels to this zoonotic pathogen and its underlying genetic basis. The aim of **my third article (“Innate and adaptive immune genes associated with MERS-CoV infection in dromedaries”. *Cells*; 10:1291)** was to detect

patterns of variation that might be linked to MERS-CoV infection in dromedaries from the United Arab Emirates (UAE). Not only MERS-CoV shedding and antibody prevalence in three dromedary populations from UAE were characterized, but also phenotype-genotype association tests of MERS-CoV seropositive camels were performed by in-solution hybridisation capture and sequencing of 100 IR genes identified in the most up-to-date dromedary genome annotation (established in the second article). In this article, we report candidate IR genes with important functions in the adaptive and innate immune response, and in cilia coating the respiratory tract, which might be associated with MERS-CoV infection in dromedaries. This work, although it still needs support by functional studies on the identified candidate genes or large-scale genome-wide association studies (GWAS), opens doors for future immunogenetic research.

Finally, for the last chapter of this thesis, I went back to the ancient and long-lasting relationship between humans and camels. Domestic camels, their hybrids and backcrosses have facilitated short and long-distance trade routes for millennia across Eurasia. Since early empires achieved a high level of connectivity, it has been suggested that the practice of camel hybridization began sometime in early first millennium BC, shortly after the two species were domesticated and their geographic ranges started overlapping. By this time, important commercial networks were already established across Southwest Asia and North Africa, and other mammals were already being crossbred. **In my fourth article (“Hidden hybrids – detecting early hybridization between dromedary and Bactrian camels in a culture-historical context”; in preparation)**, the aim was to detect early hybrids and to investigate further the beginning of the hybridisation between the two domestic species, in a culture-historical context. To that end, several large camel bone assemblages dating to the Early Iron Age and more recent times were examined using morphological and ancient DNA techniques as well as low-coverage whole-genome shotgun sequencing. By radiocarbon dating of the genetically identified hybrids, I could detect the earliest evidence of dromedary-Bactrian hybridisation in an artifact dating to Early Iron I Age (1112 – 933 calBC) from Hasanlu in northwestern Iran, one of the very important trading regions in ancient times. With a specimen from Trier, I also show that hybrid camels were present in western Germany at the latest by the Medieval Period.

Conclusions

This thesis has combined different scientific fields (e.g., immunology, virology, population genetic and genomics, paleogenomics), using cutting edge molecular, population genomic and aDNA techniques, and enclosing a large framework spanning ancient hybridisation to modern genome-wide investigation of post-domestication migration routes and immune genetic diversity. Although there is still much to untangle concerning Old World camelids history and diversity, my thesis filled existing knowledge gaps by implementing modern methodology on a large number of samples representative of different populations. Altogether, this thesis not only provides improved genome assemblies which serve as reference for the scientific community, it also opens doors for future studies about (immune)genome diversity, disease association, and the long-lasting relationship between humans and camels. Going beyond, the achieved results can be included in comparative, evolutionary and complex functional molecular studies also in other species.

5. GENERAL INTRODUCTION

Human history is marked by the efforts to overcome obstacles, be they of geographical or cultural nature. Domestic animals have been linked to the process of human development and were essential for the successful implementation of human societies. In ancient times, by establishing trading routes and reusing them over millennia, corridors of gene flow were opened shaping genetic diversity and structure of (livestock) species, particularly camels (1-3). Old World camels (Artiodactyla, Tylopoda, Camelidae, Camelini) comprise three distinct species, the one-humped dromedary (*Camelus dromedarius*), the two-humped Bactrian camel (*Camelus bactrianus*), and the two-humped wild camel (*Camelus ferus*), the only remaining wild species within the Camelini tribe. It has been estimated that one- and two-humped camels diverged 4.4 (1.9 – 7.2) million years ago (Mya) (4), where the split between the ancestors of wild camel and domestic Bactrian camel is more recent, 1.1 (0.6 – 1.8) Mya (5, 6).

Although wild two-humped camels might have been distributed throughout Central Asia, nowadays their range has become severely reduced, being now restricted to four locations: three in China (Taklamakan desert, Gashun Gobi Desert and Arjin Mountains in the Lop Nur Lake region) and one in Mongolia (Great Gobi Strictly Protected Area ‘A’) (7). These are now the last refuges for the wild camel, which is listed as Critically Endangered IUCN Red List of Threatened Species (8). On the other hand, Bactrian camel domestication has been estimated to have begun in the late fourth and early third millennium before common era (BCE) (9, 10), even though the exact region of domestication is still a matter of debate. The two main hypotheses under debate are that the domestication process took place (I) in north-eastern Iran and the adjacent Kopet Dagh foothills in south-western Turkmenistan (region belonging to historical ‘Bactria’) (9) or (II) in the Asian steppe farther to the east where humans were familiar with wild camels over an extended period of time (e.g., in Kazakhstan or north-western Mongolia) (rev. in 7). Nowadays, Bactrian camels are distributed mainly in Central Asian countries, including Mongolia, China, Kazakhstan, north-eastern Afghanistan, Russia, Crimea and Uzbekistan, although with few populations in Northern Pakistan, Iran, Turkey and India (11, 12). Furthermore, dromedary’s domestication took place in the Arabian Peninsula most likely at the transition between the second and first millennia before the Common Era (13). After that, small numbers of dromedaries arrived in Mesopotamia and from there most likely were introduced into Africa either via the Sinai, possibly starting in the

first millennium BCE or transferred from the South of the Arabian Peninsula by boat via the Gulf of Aden to Eastern Africa or further north across the Red Sea to Egypt (10, 14). The southern sea route is supported by socio-ethological observations, as today's Eastern African dromedaries are used largely for milk production rather than for riding and transportation, and this could be rooted in practices associated with the early stages of dromedary husbandry on the southern Arabian Peninsula. Thus, cross-continental sharing of nuclear genotypes reflects an extensive gene flow between African and Asian dromedaries, notably with a panmictic population on a mitochondrial level and microsatellites (15). However, the most contemporary dromedary movement started in the 1860s, where several thousand camels were brought from India until the 1920s to develop the Australian outback (16, 17).

Old World camel species are very important animals for a number of reasons. Several countries depend on their use for production as live (milk, wool, manure) or slaughtered (meat, skin, fat) animals, or for their power (riding, packing, carting) (7). Nowadays, the two domestic species live in two distinct areas of the Old World, where their distribution overlap in western and central Asia, in a few countries such as Turkey, Iran, India, Afghanistan, and Kazakhstan. This coincides with the main region where the practice of anthropogenic-driven hybridisation between the two species is most common nowadays (18, 19). Although being classified as two closely related but distinct species, the dromedary and Bactrian camel are able to reproduce and have fertile offspring. Historically, the hybridisation between the two species was associated with the transportation of goods along multiple long-distance trade routes (i.e., Silk Road), aiming at producing animals with the robustness of the Bactrian camel, the endurance of the dromedary as well as the ability to tolerate sharply contrasting climatic conditions due to heterosis (improvement of biological characteristics due to interbreeding (19)). Previous studies revealed hybrids from a Roman archaeological site in Serbia, Viminacium, dated to approximately the late third to fourth centuries CE (18, 20), although new hypotheses point to even earlier hybridization events when the distribution of the two species might have started overlapping, soon after domestication (2000–3000 ya). Nowadays, first generation hybrids and their backcrosses are valued in many countries for increased milk or wool production, as well as in social activities like famous camel wrestling events (18, 19). Additionally, hybridisation between the wild and domestic two-humped camel

species also exists. As an example, occasionally the hybridisation of domestic females with wild bulls is initiated to enhance the fitness of domestic camels (21). Nevertheless, the movement of Bactrian camels into the habitat of the wild population leads to the transmission of potential pathogens (22) as well as introgression of domestic species' genes into wild camels (23-25).

Indeed, camels are unique animals; not only are they well adapted to harsh environments, but also their immune system shows very specific features (26). Whole genome studies can unravel the uniqueness of these adaptations (27) and identify the genetic basis of their particular immune response and resistance to devastating infections that threaten other livestock species in the same regions (e.g., foot and mouth disease in dromedaries (28)). In this context, important genomic regions accounting for part of the adaptive genetic variability like the MHC (29, 30), Natural Killer cells (NKR (31)) and T-cell receptors (32, 33) have been studied. Contrary to what has been observed in other mammalian models (34), MHC class II genes in camels have been shown to exhibit low genetic diversity (29). Also, observed somatic hypermutations in T-cell receptor genes make camels a special immunological model (35). Nevertheless, previous research on immune genetic diversity in Old World camels has mainly focused on these specific candidate gene regions, while other genomic regions involved in the immune response are still poorly studied. Especially, the to-date (at the start of my thesis) existing genome assemblies for the three Old World camel species (4, 36, 37) were scaffold based, lacking chromosomal information and the resolution for analysing repetitive and highly complex genomic regions like IR genes. Therefore, improved versions of these assemblies were needed to allow more detailed studies of the diversity in IR genes as well as to study the genetic basis for relevant camel-associated zoonotic diseases. Particularly in sub-Saharan Africa, the geographical expansion of dromedaries and their integration in semi-intensive crop-livestock or peri-urban farming systems has increased, accompanied with the risk of infections and (new emerging) zoonotic diseases (38, 39). Due to an increased consumption of their meat and milk, camels might represent a significant source for pathogen transmissions to humans as demonstrated by one relevant example, MERS-CoV (39, 40). Since the beginning of the SARS-CoV-2 pandemic, MERS-CoV has regained attention, as it is also a betacoronavirus with similar symptoms in humans (fever and

respiratory problems). However, the infection in dromedaries is characterized by minor clinical signs, from asymptomatic to moderate nasal discharge (41, 42).

Although there is this long-lasting close relationship between humans and camels, modern molecular tools have not been fully explored yet. Parts of published studies – although still relevant – are either lacking up-to-date methodology implementation or include few individuals and few markers for genetic inferences. Likewise, characterization of the genome-wide genetic diversity in immune regions of the genome has been still lacking, which might be due to the fact that high quality genome assemblies and annotations have not been available until recently (43-45) [**Article 2** in this thesis]. Although we have witnessed a livestock genomic revolution in the last two decades (46), only recently the number of studies taking advantage of Next Generation Sequencing (NGS) methods and genome-wide analyses has started increasing in camels. With the projects presented in this thesis, I contributed to characterize the (immuno)genome diversity in Old World camels, as well as to understand and create the foundation to place hybridisation in a culture-historical context after classifying early dromedary-Bactrian hybrids from pre-Roman times, using NGS methods. When compared to traditional DNA Sanger sequencing/genotyping methods, NGS presents numerous advantages such as higher sensitivity to detect low-frequency variants, with a faster turnaround time for higher number of sampling and comprehensive genomic coverage, at lower costs per sample (47). These new technologies enable us to examine camels' genetic variation in underlying economically relevant traits as well as diverse (immune) genes. The latter should be a priority, not only due to the fact that camels are well adapted to extreme environments and in contact with different pathogens, but also because both domestic species (and their hybrids) are economically important and have a high potential as productive livestock species (19).

With the projects accomplished during my PhD, integrated in a great research team, I aimed to contribute to overcoming some of the above identified knowledge gaps in camel research. Specifically, the aims of my thesis were to 1) assess patterns of diversity and population structure of the global dromedary population based on genome-wide markers; 2) estimate genetic diversity in IR gene groups in the three Old World species by analysing improved and more accurate genome assemblies; 3) characterize genetic diversity in IR genes and identify candidate genes associated with MERS-CoV infection

in dromedaries; and 4) classify early dromedary-Bactrian hybrids from pre-Roman times in a zooarchaeological – paleogenetic framework, creating the foundation to place hybridisation in a culture-historical context. The projects to reach aims 1 – 3 are described in detail in the subsection 5.1 “*Characterization of (immuno)genome diversity in Old World camels*”); while aim 4 is summarized in 5.2 “*Detection of early hybrids between dromedary and Bactrian camel in a culture-historical context*”).

5.1 “*Characterization of (immuno)genome diversity in Old World camels*”

Domestic and wild dromedary specimens cohabited together on the coastal Southeast of the Arabian Peninsula for nearly one millennium, until the beginning of the Common Era (CE) (13, 14, 48-50). After that, the early dispersal of modern dromedaries from the Arabian Peninsula to the Levant, North Africa and South Asia was followed by cross-continental back-and-forth movements along important commercial trading routes, which shaped genetic diversity. The camel-borne incense trade, from Arabia to the Levant, was an important element in the economy of the eastern Mediterranean region in the first millennium BC (51). Previous studies using a limited number of microsatellite markers and/or mtDNA detected only weak population structure in the global dromedary population (15, 52, 53). In this project, I hypothesised that fine-scale population structure might exist in the global dromedary population, which could not have been detected before due to lack of genome-wide information. Furthermore, we tested the hypothesis whether post-domestication patterns of gene flow in dromedaries would reflect the back-and-forth human-driven dispersal along ancient trading routes. Consequently, the aim of **Article 1** was to overcome the limitations of previous studies (low number of markers) and to understand if the application of a genome-wide approach covering many genomic regions would change the pattern of diversity and population structure seen in previous works. Also, we were interested in how human-induced migration patterns and historic demographic changes might have influenced population structure in the global dromedary population. For this, we used ddRADseq for SNP discovery and genotyping of a global dataset including dromedary samples spread over three continents. This technique is flexible and a highly cost-effective genotyping strategy for genome sampling (54).

As camels are such important animals, it is necessary to use and preserve the genomic diversity for future generations in a globally changing world. Maintaining genetic

diversity in populations is beneficial for reducing the spread of diseases by rapid adequate immune responses and for limiting, for example, parasite evolution (55). In fact, demographic changes might cause loss of genetic diversity, and particularly during domestication events (intensive selection and potential inbreeding) reduce diversity in certain genomic regions (56). In other regions such as IR genes, however, genetic diversity can be conserved due to diverse selective pressures of pathogens (57). Camels are able to cope with adverse climatic conditions in harsh environments, and they are in contact with different pathogenic pressures, which might affect the diversity in the coding regions of their immune genome. As previous studies mostly focused on candidate regions of the immune genome like MHC, NKR or T-cell receptor regions, we identified the need of a broader approach to capture and characterize the overall genetic diversity of the IR genes in all three camel species. As this knowledge gap was most likely a consequence of lack of high-quality chromosome-assembled genomes, higher-quality and more accurate genome assemblies were then needed. In my second project, I hypothesised that we would detect diversity patterns compatible with differential exposures to pathogens, and consequent different selective pressures in Old World camelids. Therefore, the aim of **Article 2** was first to apply computational efforts on existing genome assemblies to generate improved Old World camelid genome assemblies' versions – CamDro3, CamBac2 and CamFer2, for dromedaries, Bactrian camels and wild camels, respectively. Next, our goal was to take advantage of these upgraded genome assemblies to assess the genetic diversity in different groups of immune genes and compare them among species and to the rest of the intra-genic genomic diversity.

Even though the proximity of camels to humans has been always very tight, nowadays we see an increase in consumption of camel products such as camel milk or meat (38). This, together with the geographical expansion of camels, particularly dromedaries in sub-Saharan Africa, has increased the risk of infectious and zoonotic diseases, consequently with a big impact on public health and economy. Therefore, in-depth knowledge on camel-associated zoonoses, their clinical signs and modes of transmission is important to assess potential infection risks for humans as well as for other animals. Of particular importance is MERS-CoV, which was first isolated 2012 in Saudi Arabia. Retrospectively, specific antibodies were found in dromedary blood samples as early as 1992 (58). Since then, the virus has appeared in four continents (59), and transmission of

the virus from dromedary to Bactrian camel has been recently demonstrated (60). Although dromedaries remain the only documented source of human infection, MERS-CoV has been detected in bat species throughout the Middle East and Africa (rev. in 61). MERS-CoV prevalence, global and local epidemiology, genetic diversity (62), and the course of disease from experimental infections in dromedaries (see 63) have been well described. However, little is known about the underlying genetic basis of the infection in camels. In the need of improving knowledge about the underlying genetic variation in MERS-CoV infection in dromedaries, the aim of **Article 3** was to identify potential genetic diversity from significant IR genomic regions that might be in association with infected dromedaries. For this, a total of 121 dromedaries from United Arab Emirates (UAE) were characterized phenotypically, and a target enrichment approach with in-solution hybridisation capture of 100 annotated immune genes was applied. Targeted approaches using next-generation sequencing (NGS) allow a subset of genes or regions of interest of the genome are isolated and sequenced, generating a smaller, more manageable data set. Previously, targeted enrichment has successfully been applied in sequencing the IR genes of gopher tortoises (64), in assessing variants in immune genes associated with Hepatitis B virus infection (65) and in identifying somatic alterations in follicular dendritic cell sarcoma in known cancer-associated genes (66). I hypothesised that I would be able to detect evidence of genetic variation likely to be associated with recent MERS-CoV infection in seropositive dromedaries from UAE, stepping further on better understanding the disease dynamics in dromedaries.

5.2 “Detection of early hybrids between dromedary and Bactrian camel in a culture-historical context”

Finally, to close the circle of this thesis and going back to the long-lasting relationship between humans and camels, I was interested in investigating the beginning of hybridisation between the two domestic species in a culture-historical context. Textual and pictorial evidence show that the two species encountered each other in Mesopotamia as early as 1000 BCE, and both becoming common between the Caucasus and Arabia by the first century BCE (67-69). Thus, it is believed that human-mediated interbreeding between Bactrian and dromedary may have occurred soon after their domestication, in those regions where their geographical distribution overlapped, i.e., Central and West Asia. Iron Age in these region is estimated to be 1200 – 600 BCE, and long before this

period, mid- and late 3rd millennium BCE, humans would already cross equids, “kunga” understood as a hybrid between a hemione (*Equus hemionus*) and a domestic donkey (*Equus asinus*) (70). Later, the offspring of a jack (*Equus asinus*) and a mare (*Equus caballus*), “mule”, were often present in Mesopotamian art of the first millennium BCE (71). Mules thrive on cheaper food, have stronger working capacities and can carry more weight than horses, have longer life spans and are more resistant to disease (72). During Iron Age, humans were mostly breeders, and without a “species” concept, most likely they did experiments in a “try and fail” method by crossbreeding domestic species, including camels. The practice of crossbreeding between closely related species, and in particular between dromedary and Bactrian camel, can lead to heterosis showing an improvement of production traits, as well as increased growth/strength and fitness in harsh conditions (18). Unlike mules (generally sterile), different types of fertile hybrid and backcrosses are described today, e.g., F1 hybrids for resistance and strength, F1 X dromedary backcross for higher milk production, or F1 X Bactrian camel for wool and fat production (cold resistance) (19). Preliminary combined genetic and morphometric results revealed hybrids from a Roman archaeological site in Serbia (Viminacium), dated to approximately the late third to fourth centuries CE (18, 20). The general lack of osteomorphological characterization of hybrids based on recent skeletons with known parents makes it extremely difficult to identify archaeological hybrids, thus making them invisible. In order to detect hybrids with confidence in an archaeological context, ancient DNA analysis is the method of choice, as a complement to osteomorphological characterization, as bones are normally fragmented in several pieces and/or incomplete. Nevertheless, due to the nature of the archaeological samples, it can be very challenging to retrieve DNA from ancient samples. As endogenous DNA in ancient samples is usually present in low quantity, ancient DNA (aDNA) studies usually operate with low sequencing coverages. Due to fragmentation of DNA molecules a large proportion of the sequenced fragments are short, increasing the probability of multiple matching sites in the genome (73). Moreover, the proportion of DNA integrity is usually negatively correlated with the temperature to which the samples were exposed (74-76). Consequently, it is important to choose the right methods in order to obtain DNA from ancient sampling. The aim of **Article 4** was to detect early hybrids with confidence, possibly from pre-Roman times, to test the historic distribution, ubiquity, and cultural significance of hybrid camels in the culture-historical context. For this, large specimens

from key sites mainly in Iran, Syria and Israel, from Iron Age and more recent times were examined in order to identify possible hybrids in the archaeological record. For this we combined information from ancient DNA Sanger and next-generation low coverage genome-wide shotgun sequencing to accurately detect hybrids, as well as radio-carbon dating the samples considered as possible ancient hybrids. As camel hybridisation facilitated the high level of connectivity achieved by early empires across Eurasia, I hypothesized this practice might be already occurring as early as Iron Age, as important commercial networks were already established and other mammals were already being crossbred.

References

1. Edwards CJ, Baird JF, Machugh DE. Taurine and zebu admixture in Near Eastern cattle: a comparison of mitochondrial, autosomal and Y-chromosomal data. *Anim Genet.* 2007;38:520-524. doi:10.1111/j.1365-2052.2007.01638.x.
2. Fages A, Hanghøj K, Khan N, Gaunitz C, Seguin-Orlando A, Leonardi M, et al. Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell.* 2019;177:1419-1435. doi:10.1016/j.cell.2019.03.049.
3. Ming L, Yuan LY, Yi L, Ding GH, Hasi SR, Chen GL, et al. Whole-genome sequencing of 128 camels across Asia reveals origin and migration of domestic Bactrian camels. *Commun Biol.* 2020;3:1. doi:10.1038/s42003-019-0734-6.
4. Wu H, Guang X, Al-Fageeh MB, Cao J, Pan S, Zhou H, et al. Camelid genomes reveal evolution and adaptation to desert environments. *Nat Commun.* 2014;5(1):5188. doi:10.1038/ncomms6188.
5. Ji R, Cui P, Ding F, Geng J, Gao H, Zhang H, et al. Monophyletic origin of domestic bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild camel (*Camelus bactrianus ferus*). *Anim Genet.* 2009;40:377-382. doi:10.1111/j.1365-2052.2008.01848.x.
6. Mohandesan E, Fitak RR, Corander J, Yadamsuren A, Chuluunbat B, Abdelhadi O, et al. Mitogenome Sequencing in the Genus *Camelus* Reveals Evidence for Purifying Selection and Long-term Divergence between Wild and Domestic Bactrian Camels. *Sci Rep.* 2017;7(1):9970. doi:10.1038/s41598-017-08995-8.
7. Burger PA, Ciani E, Faye B. Old World camels in a modern world – a balancing act between conservation and genetic improvement. *Anim Genet.* 2019;50(6):598-612. doi:10.1111/age.12858.
8. Hare J. *Camelus ferus*, Bactrian Camel. The IUCN Red List of Threatened Species 2008. 2008:eT63543A12689285. Available from:

<https://www.iucnredlist.org/species/22712252/137668645> doi:
10.2305/IUCN.UK.2008.RLTS.T63543A12689285.en

9. Benecke N. Archäozoologische Studien zur Entwicklung der Haustierhaltung in Mitteleuropa und Südkandinavien von den Anfängen bis zum ausgehenden Mittelalter. Vol. 46. Akademie Verlag; 1994. 451 p. (Deutsches Archäologisches Institut, editor. Schriften zur Ur- und Frühgeschichte). <https://www.degruyter.com/document/doi/10.1515/9783050069456/html>.

10. Bulliet RW. The Camel and the Wheel. 19 ed. New York: Columbia University Press; 1975. 324 p.

11. Mirzaei F. Production and trade of camel products in some Middle East countries. J Sci Food Agric. 2012;16:153-160. doi:10.1533/9781845691721.2.287.

12. Vyas S, Sharma N, Sheikh FD, Singh S, Sena DS, Bissa UK. Reproductive status of *Camelus bactrianus* during early breeding season in India. Asian Pac J Reprod. 2015;4:61-64. doi:10.1016/S2305-0500(14)60060-9.

13. Uerpmann H-P, Uerpmann M. The appearance of the domestic camel in southeast Arabia. J Oman Stud. 2002;12:235–260.

14. Grigson C. Camels, Copper and Donkeys in the Early Iron Age of the Southern Levant: Timna Revisited. Levant. 2012;44:82-100. doi:10.1179/175638012X13285409187919.

15. Almathen F, Charruau P, Mohandesan E, Mwacharo JM, Orozco-terWengel P, Pitt D, et al. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. Proc Natl Acad Sci USA. 2016;113(24):6707-6712. doi:10.1073/pnas.1519508113.

16. Faye B, Grech S, Korchani T. Le dromadaire, entre féralisation et intensification. Anthropozoologica. 2004;39(2):7-14.

17. Rangan H, Kull C. The Indian Ocean and the making of outback Australia: an ecocultural odyssey. In: Moorthy S, Jamal Y, editors. Indian Ocean Studies: Cultural, Social and Political Perspectives. New York: Routledge; 2010. Chapter 3; p. 45-72.

18. Çakırlar C, Berthon R. Caravans, camel wrestling and cowrie shells: Towards a social zooarchaeology of camel hybridization in Anatolia and adjacent regions. Anthropozoologica. 2014;49(2):237-252. doi:10.5252/az2014n2a06.

19. Faye B, Konuspayeva G. The Encounter between Bactrian and Dromedary Camels in Central Asia. In: Knoll E-M, Burger PA, editors. Camels in Asia and North Africa. Interdisciplinary perspectives on their significance in past and present. Vienna: Austrian Academy of Sciences; 2012. p. 27-33. (Veröffentlichungen zur Sozialanthropologie).

20. Burger PA, Lado S, Mohandesan E, Vukovic-Bogdanovic S, Peters J, Çakırlar C. Ancient and modern hybridisation between one- and two-humped camels. Second

International Selçuk-Ephesus Symposium On Culture Of Camel-Dealing And Camel Wrestling; 2018 JAN 18-20, 2018; Selcuk, Turkey. 153-159 p. (vol. Volume II Natural And Applied Science Health And Medical Science). eng. Available from: http://www.selcuk.bel.tr/Files/dosyalar/sempozyum_bildiri/sempozyum_2018/2_Ulusla_rarasi_Selcuk_Efes_Devecilik_Kulturu_ve_Deve_Guresleri_Sempozyumu_Fen_ve%20Saglik_Bilimleri_2018.pdf.

21. Yadamsuren A, Dulamtseren E, Reading RP. The Conservation Status and Management of Wild Camels in Mongolia. In: Knoll E-M, Burger PA, editors. *Camels in Asia and North Africa. Interdisciplinary perspectives on their significance in past and present*. Vienna: Austrian Academy of Sciences; 2012. p. 45-54. (Veröffentlichungen zur Sozialanthropologie).

22. Walzer C, Kaczensky P, Dulamtseren E, Yadamsuren A. Working in a freezer: capturing and collaring wild Bactrian camels. In: Knoll E-M, Burger PA, editors. *Camels in Asia and North Africa. Interdisciplinary perspectives on their significance in past and present*. Vienna: Austrian Academy of Sciences Press; 2012. p. 61-68. (Veröffentlichungen zur Sozialanthropologie).

23. Felkel S, Wallner B, Chuluunbat B, Yadamsuren A, Faye B, Brem G, et al. A First Y-Chromosomal Haplotype Network to Investigate Male-Driven Population Dynamics in Domestic and Wild Bactrian Camels. *Front Genet.* 2019;10:423. doi:10.3389/fgene.2019.00423.

24. Silbermayr K, Orozco-terWengel P, Charruau P, Enkhbileg D, Walzer C, Vogl C, et al. High mitochondrial differentiation levels between wild and domestic Bactrian camels: a basis for rapid detection of maternal hybridization. *Anim Genet.* 2010;41(3):315-318. doi:10.1111/j.1365-2052.2009.01993.x.

25. Silbermayr K, Burger PA. Hybridization: A Threat to the Genetic Distinctiveness of the Last Wild Old World Camel Species. In: Knoll EM, Burger PA, editors. *Camels in Asia and North Africa. Interdisciplinary perspectives on their significance in past and present*. Vienna: Austrian Academy of Sciences; 2012. p. 69-76. (Veröffentlichungen zur Sozialanthropologie).

26. Hussen J, Schuberth H-J. Recent Advances in Camel Immunology. *Front Immunol.* 2021;11:3569. doi:10.3389/fimmu.2020.614150.

27. Ali A, Baby B, Vijayan R. From Desert to Medicine: A Review of Camel Genomics and Therapeutic Products. *Front Genet.* 2019;10:17. doi:10.3389/fgene.2019.00017.

28. Wernery U, Kinne J. Foot and mouth disease and similar virus infections in camelids: a review. *Rev Sci Tech Off Int Epiz.* 2012;31:907-918. doi:10.20506/rst.31.3.2160.

29. Plasil M, Mohandesan E, Fitak RR, Musilova P, Kubickova S, Burger PA, et al. The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes. *BMC Genom.* 2016;17(1):1-17. doi:10.1186/s12864-016-2500-1.

30. Plasil M, Wijkmark S, Elbers JP, Oppelt J, Burger P, Horin P. The major histocompatibility complex of Old World camelids: Class I and class I-related genes. *HLA Immune Response Genetics*. 2019;93(4):203-215. doi:10.1111/tan.13510.
31. Futas J, Oppelt J, Jelinek A, Elbers JP, Wijacki J, Knoll A, et al. Natural Killer Cell Receptor Genes in Camels: Another Mammalian Model. *Front Genet*. 2019;10:620. doi:10.3389/fgene.2019.00620.
32. Antonacci R, Bellini M, Castelli V, Ciccarese S, Massari S. Data characterizing the genomic structure of the T cell receptor (TRB) locus in *Camelus dromedarius*. *Data in Brief*. 2017;14:507-514. doi:10.1016/j.dib.2017.08.002.
33. Ciccarese S, Burger PA, Ciani E, Castelli V, Linguiti G, Plasil M, et al. The Camel Adaptive Immune Receptors Repertoire as a Singular Example of Structural and Functional Genomics. *Front Genet*. 2019;10:997. doi:10.3389/fgene.2019.00997.
34. Robinson J, Mistry K, McWilliam H, Lopez R, Marsh SGE. IPD—the Immuno Polymorphism Database. *Nucleic Acids Res*. 2010;38:D863-D869. doi:10.1093/nar/gkp879.
35. Vaccarelli G, Antonacci R, Tasco G, Yang F, Giordano L, El Ashmaoui HM, et al. Generation of diversity by somatic mutation in the *Camelus dromedarius* T-cell receptor gamma variable domains. *Eur J Immunol*. 2012;42:3416-3428. doi:10.1002/eji.201142176.
36. Fitak RR, Mohandesan E, Corander J, Burger PA. The *de novo* genome assembly and annotation of a female domestic dromedary of North African origin. *Mol Ecol Res*. 2015;16:314-324. doi:10.1111/1755-0998.12443.
37. Jirimutu, Wang Z, Ding G, Chen G, Sun Y, Sun Z, et al. Genome sequences of wild and domestic bactrian camels. *Nat Commun*. 2012;3(1):1202. doi:10.1038/ncomms2192.
38. Faye B. The camel, new challenges for a sustainable development. *Trop Anim Health Pro*. 2016;48(4):689-692. doi:10.1007/s11250-016-0995-8.
39. Zhu S, Zimmerman D, Deem SL. A Review of Zoonotic Pathogens of Dromedary Camels. *EcoHealth*. 2019;16(2):356-377. doi:10.1007/s10393-019-01413-7.
40. Hemida MG, Perera RA, Al Jassim RA, Kayali G, Siu LY, Wang P, et al. Seroepidemiology of Middle East respiratory syndrome (MERS) coronavirus in Saudi Arabia (1993) and Australia (2014) and characterisation of assay specificity. *Euro Surveill*. 2014;19:1-6. doi:10.2807/1560-7917.ES2014.19.23.20828.
41. Hemida M, Chu DKW, Poon LLM, Perera RAPM, Alhammadi M, Ng H-y, et al. MERS Coronavirus in Dromedary Camel Herd, Saudi Arabia. *Emerg Infect Dis*. 2014;20(7):1231. doi:10.3201/eid2007.140571.

42. Nowotny N, Kolodziejek J. Middle East respiratory syndrome coronavirus (MERS-CoV) in dromedary camels, Oman, 2013. *Euro Surveill.* 2014;19(16):20781. doi:10.2807/1560-7917.ES2014.19.16.20781.
43. Elbers JP, Rogers MF, Perelman PL, Proskuryakova AA, Serdyukova NA, Johnson WE, et al. Improving Illumina assemblies with Hi-C and long reads: An example with the North African dromedary. *Mol Ecol Res.* 2019;19(4):1015-1026. doi:10.1111/1755-0998.13020.
44. Lado S, Elbers JP, Rogers MF, Melo-Ferreira J, Yadamsuren A, Corander J, et al. Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies. *BMC Genom.* 2020;21(1):606. doi:10.1186/s12864-020-06990-4.
45. Ming L, Wang Z, Yi L, Batmunkh M, Liu T, Siren D, et al. Chromosome-level assembly of wild Bactrian camel genome reveals organization of immune gene loci. *Mol Ecol Res.* 2020;20(3):770-780. doi:10.1111/1755-0998.13141.
46. Rothschild MF, Plastow GS. Applications of genomics to improve livestock in the developing world. *Livest Sci.* 2014;166:76-83. doi:10.1016/j.livsci.2014.03.020.
47. Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, et al. Adaptation Genomics: the next generation. *Trends Ecol Evol.* 2010;25(12):705-712. doi:10.1016/j.tree.2010.09.002.
48. von den Driesch A, Obermaier H. The hunt for wild dromedaries during the 3rd and 2nd millennia BC on the United Arab Emirates coast. Camel bone finds from the excavations at Al Sufouh 2 Dubai, UAE. In: Grupe G, Peters J, editors. *Skeletal series and their socio-economic context. Rahden (Germany): Verlag Marie Leidorf GmbH; 2007. p. 133–167. (Documenta Archaeobiologiae).*
49. Grigson C. The history of the camel bone dating project. *Anthropozoologica.* 2014;49(2):225-235. doi:10.5252/az2014n2a05.
50. Uerpmann M, Uerpmann H-P. Archeozoology of camels in South-Eastern Arabia. In: Knoll EM, Burger PA, editors. *Camels in Asia and North Africa. Interdisciplinary perspectives on their significance in past and present. Vienna: Austrian Academy of Sciences; 2012. p. 109-122. (Veröffentlichungen zur Sozialanthropologie).*
51. Artzy M. Incense, Camels and Collared Rim Jars: Desert Trade Routes and Maritime Outlets in the Second Millennium. *Oxford Journal of Archaeology.* 1994;13:121-147. doi:10.1111/j.1468-0092.1994.tb00035.x.
52. Mburu DN, Ochieng JW, Kuria SG, Jianlin H, Kaufmann B, Rege JEO, et al. Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classification. *Anim Genet.* 2003;34(1):26-32. doi:10.1046/j.1365-2052.2003.00937.x.

53. Spencer PBS, Woolnough AP. Assessment and genetic characterisation of Australian camels using microsatellite polymorphisms. *Livest Sci.* 2010;129(1):241-245. doi:10.1016/j.livsci.2010.01.006.
54. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double Digest RADseq: An Inexpensive Method for *De Novo* SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE.* 2012;7(5):e37135. doi:10.1371/journal.pone.0037135.
55. Van Houte S, Ekroth AKE, Broniewski JM, Chabas H, Ashby B, Bondy-Denomy J, et al. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature.* 2016;532:385-388. doi:10.1038/nature17436.
56. Ramey HR, Decker JE, McKay SD, Rolf MM, Schnabel RD, Taylor JF. Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genom.* 2013;14:382. doi:10.1186/1471-2164-14-382.
57. Horrocks NPC, Matson KD, Tieleman BI. Pathogen Pressure Puts Immune Defense into Perspective. *Integr Comp Biol.* 2011;51:563-576. doi:10.1093/icb/icr011.
58. Corman VM, Jores J, Meyer B, Younan M, Liljander A, Said MY, et al. Antibodies against MERS Coronavirus in Dromedary Camels, Kenya, 1992–2013. *Emerg Infect Dis.* 2014;20:1319-1322. doi:10.3201/eid2008.140596.
59. Gossner C, Danielson N, Gervelmeyer A, Berthe F, Faye B, Kaasik Aaslav K, et al. Human–Dromedary Camel Interactions and the Risk of Acquiring Zoonotic Middle East Respiratory Syndrome Coronavirus Infection. *Zoon Publ Health.* 2016;63(1):1-9. doi:10.1111/zph.12171.
60. Lau SKP, Li KSM, Luk HKH, He ZR, Teng JLL, Yuen KY, et al. Middle East Respiratory Syndrome Coronavirus Antibodies in Bactrian and Hybrid Camels from Dubai. *Mosphere.* 2020;5(1):e00898-19. doi:10.1128/mSphere.00898-19.
61. Dawson P, Malik MR, Parvez F, Morse SS. What Have We Learned About Middle East Respiratory Syndrome Coronavirus Emergence in Humans? A Systematic Literature Review. *Vector-Borne Zoonot.* 2019;19(3):174-192. doi:10.1089/vbz.2017.2191.
62. Chu DKW, Hui KPY, Perera RAPM, Miguel E, Niemeyer D, Zhao J, et al. MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. *Proc Natl Acad Sci USA.* 2018;115:3144-3149. doi:10.1073/pnas.1718769115.
63. Adney DR, Clancy CS, Bowen RA, Munster VJ. Camelid Inoculation With Middle East Respiratory Syndrome Coronavirus: Experimental Models of Reservoir Host Infection. *Viruses.* 2020;12:1370. doi:10.3390/v12121370.
64. Elbers JP, Brown MB, Taylor SS. Identifying genome-wide immune gene variation underlying infectious disease in wildlife populations - a next generation sequencing approach in the gopher tortoise. *BMC Genom.* 2018;19:64. doi:10.1186/s12864-018-4452-0.

65. Yu F, Zhang X, Tian S, Geng L, Xu W, Ma N, et al. Comprehensive investigation of cytokine- and immune-related gene variants in HBV-associated hepatocellular carcinoma patients. *Biosci Rep*. 2017;37:1-8. doi:10.1042/BSR20171263.
66. Griffin GK, Sholl LM, Lindeman NI, Fletcher CDM, Hornick JL. Targeted genomic sequencing of follicular dendritic cell sarcoma reveals recurrent alterations in NF- κ B regulatory genes. *Modern Pathol*. 2016;29:67-74. doi:10.1038/modpathol.2015.130.
67. Al-Zaidi S-A. Betwixt and Between the Bactrian Camel and the Dromedary: The Semantic Evolution of the Lexeme *udru* during the 11th to 8th Centuries BCE. *Arabian Epigraphic Notes*. 2017;3:11-18.
68. Herles M. Kamele in assyrischen Quellen – Ein Exot wird zur Selbstverständlichkeit. In: Pietruschka U, Streck M, editors. *Symbolische Repräsentation und Wirklichkeit nomadischen Lebens*. Vol 12. Reichert; 2010. p. 127-167. (Nomaden und Sesshafte).
69. Potts DT. Camel hybridization and the role of *Camelus bactrianus* in the ancient Near East. *J Econ Soc Hist Orient*. 2004;47:143-165. doi:10.1163/1568520041262314.
70. Postgate JN. The equids of Sumer, again. In: Uerpmann H-P, Meadow RH, editors. *Equids in the Ancient World*. Wiesbaden: Ludwig Reichert Verlag; 1986. p. 194-206.
71. Clutton-Brock J. *A Natural History of Domesticated Mammals*. Cambridge University Press; 1999. 248 p. ISBN: 0521634954.
72. Tegetmeier WB, Sutherland CL. *Horses, asses, zebras, mules and mule breeding*. London: H. Cox; 1895.
73. Günther T, Nettelblad C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genet*. 2019;15(7):e1008302. doi:10.1371/journal.pgen.1008302.
74. Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc R Soc B*. 2012;279:4724-4733. doi:10.1098/rspb.2012.1745.
75. Hofreiter M, Paijmans JLA, Goodchild H, Speller CF, Barlow A, Fortes GG, et al. The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays*. 2015;37(3):284-293. doi:10.1002/bies.201400160.
76. Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, Collins MJ. Not just old but old and cold? *Nature*. 2001;410:771-772. doi:10.1038/35071177.

6. PUBLICATIONS

6.1 Article 1

Lado, S., Elbers, J. P., Duskocil, A., Scaglione, D., Trucchi, E., Banabazi, M. H., Almathen, F., Saitou, N., Ciani, E. & Burger, P. A. (2020). Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes. *Communications Biology*, 3(1), 1-8. <https://doi.org/10.1038/s42003-020-1098-7>

Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes

Sara Lado¹, Jean Pierre Elbers¹, Angela Duskocil¹, Davide Scaglione ², Emiliano Trucchi³,
Mohammad Hossein Banabazi⁴, Faisal Almathen^{5,6}, Naruya Saitou⁷, Elena Ciani⁸  & Pamela Anna Burger ¹ 

Dromedaries have been essential for the prosperity of civilizations in arid environments and the dispersal of humans, goods and cultures along ancient, cross-continental trading routes. With increasing desertification their importance as livestock species is rising rapidly, but little is known about their genome-wide diversity and demographic history. As previous studies using few nuclear markers found weak phylogeographic structure, here we detected fine-scale population differentiation in dromedaries across Asia and Africa by adopting a genome-wide approach. Global patterns of effective migration rates revealed pathways of dispersal after domestication, following historic caravan routes like the Silk and Incense Roads. Our results show that a Pleistocene bottleneck and Medieval expansions during the rise of the Ottoman empire have shaped genome-wide diversity in modern dromedaries. By understanding subtle population structure we recognize the value of small, locally adapted populations and appeal for securing genomic diversity for a sustainable utilization of this key desert species.

¹Research Institute of Wildlife Ecology, Department of Interdisciplinary Life Sciences, University of Veterinary Medicine Vienna, Savoyenstrasse 1, 1160 Vienna, Austria. ²IGA Technology Services, Via Jacopo Linussio, 51, 33100 Udine, Italy. ³Department of Life and Environmental Sciences, Marche Polytechnic University, Via Breccie Bianche, 60131 Ancona, Italy. ⁴Department of Biotechnology, Animal Science Research Institute of Iran (ASRI), Agricultural Research, Education & Extension Organization (AREEO), Karaj 3146618361, Iran. ⁵Department of Veterinary Public Health, College of Veterinary Medicine, King Faisal University, Al-Hasa, Saudi Arabia. ⁶The Camel Research Center, King Faisal University, Al-Hasa, Saudi Arabia. ⁷Population Genetics Laboratory, National Institute of Genetics, 1111 Yata, Mishima 411-8540, Japan. ⁸Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari Aldo Moro, Via Orabona, 4, 70125 Bari, Italy. email: Elena.Ciani@uniba.it; Pamela.Burger@vetmeduni.ac.at

As one of the most recently domesticated animals (~3000–4000 years before present; ybp), the dromedary (*Camelus dromedarius*) has a special position in human migration and trading¹. Its physiological adaptations to harsh and dry environments allowed humans to traverse hostile lands such as deserts like no other livestock, including the horse². Reversing a historian's observation “the sea unites rather than divides”³ to the desert, dromedaries facilitated the expansion of civilizations^{1,4,5}. Their superior and unique features predestined dromedaries for the use as military animals, and for the advance of international trading along ancient caravan routes, such as the Silk Road and Incense routes^{6,7}. Interspecific hybridization between dromedaries and the closely related two-humped Bactrian camels (*Camelus bactrianus*) produced even more robust and enduring animals with an aptitude for the extreme climatic conditions of the Silk Road^{8,9}. Nowadays, first generation hybrids and their backcrosses are valued in many countries for increased milk or wool production¹⁰, as well as in famous camel wrestling events¹¹. These remarkable commercial networks in human history facilitated domestic animal exchange across large geographical distances and acted as gene-flow corridors, not only for camels but also for other livestock¹². While camels were the chosen animals for transportation, horse movements along the complete network of the Silk Road mainly occurred in forms of tributes and gifts^{12,13}.

The early progenitors (*Protylopus*) of camelids emerged in the North American savannah during the Eocene (~45 Mya). After their split into New (Lamini) and Old World (Camelini) camels around 15 Mya, the ancestors of the Old World camels crossed into Eurasia via the Bering Land Bridge (~6.6 Mya) and further diverged into one- and two-humped camels (reviewed in Burger et al.¹⁴). Early-domestic dromedaries (second and first millennium before Common Era; BCE) cohabited the coastal Southeast of the Arabian Peninsula for nearly one millennium together with wild specimens, which did not survive the beginning of the Common Era (CE)^{15–19}. The early dispersal of modern dromedaries from the Arabian Peninsula to the Levant, North Africa, South Asia, and finally to Australia (introduced in the late 19th century)²⁰ was followed by cross-continental back-and-forth movements along historic trading routes. This led to a blurring of genetic stocks²¹ culminating in a panmictic dromedary population at the mitochondrial DNA level¹. Previous studies using a limited number of microsatellite markers and mtDNA detected only weak population structure in the global dromedary population^{1,20,22}. As few genomic studies have been completed with the dromedary, this major ungulate species has been left out of the livestock genomic revolution. However, two draft reference genomes at the scaffold level have been released^{23,24}, as well as a draft reference genome at the chromosome level²⁵, which will facilitate further genomic investigations.

We wanted to understand how human-induced migration patterns and historic demographic changes might have influenced population structure in the global dromedary population. We sequenced 22,721 SNP markers to overcome the limitations of previous studies using not more than 20 microsatellites. With a global dataset including samples spread over three continents, we describe effective migration rates of modern dromedaries that follow their human-driven dispersal along ancient trading routes⁷. Understanding subtle population structure, which has been shaped by past and recent demographic events, will help in recognizing the value of small populations and securing genomic diversity for a sustainable utilization of this important livestock species in a globally changing world.

Results and discussion

We performed double-digest restriction site associated DNA (ddRAD) sequencing on 122 dromedary DNA samples from 18

countries (Supplementary Data 1) representative of the species distribution range. We included one Bactrian camel to test for potential interspecific hybridization, as this continues to be a widespread practice in Central Asia that might have started as early as pre-Roman times¹¹. Higher numbers of reads mapping to the Bactrian camel were detected in three individuals from Iran and in six from Kazakhstan (see “Methods”), and we decided to remove these samples from downstream analysis due to potential introgression from Bactrian camel (Supplementary Data 2). After stringent filtering for genotype and individual missingness, minor allele frequency and relatedness, the final dataset consisted of 95 dromedaries and 22,721 SNPs present in at least 75% of the individuals.

Moderate genome-wide diversity and low population structure.

With 22,721 SNPs, we estimated expected ($H_E = 0.27 \pm 0.17$; mean \pm SD) and observed ($H_O = 0.25 \pm 0.17$) heterozygosities in the global dromedary population ($n_{\text{pop}} = 17$; $n_{\text{ind}} = 95$). Separating the samples according to their continental origins, both Asian ($n_{\text{ind}} = 49$, $H_E = 0.27 \pm 0.17/H_O = 0.25 \pm 0.17$) and African dromedaries ($n_{\text{ind}} = 46$, $H_E = 0.26 \pm 0.17/H_O = 0.25 \pm 0.18$) showed similar genomic diversity. The mean H_E ($t = -2.2641$, $df = 45,398$, $P = 0.02$) and inbreeding coefficients ($t = -2.5159$, $df = 43,024$, $P = 0.01$) were higher in Asian than African dromedaries, but mean H_O ($t = -1.2791$, $df = 45,385$, $P = 0.2$) was not different between continents, according to Welch's t test. Complete diversity and inbreeding values are given in Supplementary Table 1. In comparison with other domestic species, i.e., sheep ($H_E = 0.22–0.32$)²⁶ or cattle ($H_E = 0.24–0.30$)²⁷, we consider the genome-wide diversity in dromedaries as moderate at the best. Several bottlenecks during the last glacial period (see demographic analysis below, and Fitak et al.²⁴) and during domestication left modern dromedaries with a minimum of only six maternal lineages¹ and limited genome-wide diversity. This will have implications on future intensification of breeding and genomic selection in dromedaries from regions with increasing desertification.

In general, the genome-wide differentiation within the global dromedary population was very low. Analysis of Molecular Variance (AMOVA) revealed that most of the variation, ~94.3%, is explained within individuals (Supplementary Table 2). Allelic richness (AR) was similar between countries (AR = 0.25–0.27) with exception of Kenya which was lower (AR = 0.21). The pairwise fixation index between African and Asian individuals was very low ($F_{ST} = 0.006$; $P < 0.001$), and indices between dromedaries from different countries (if significant at all) were lowest in geographically close populations (e.g., Libya/Algeria: $F_{ST} = 0.0002$) and increased with geographic distance (Pakistan/Tunisia: $F_{ST} = 0.0328$) (Supplementary Table 3).

We screened for loci deviating from neutrality using BayeScan 2.1²⁸ and identified sixteen F_{ST} outliers to be putatively under selection (false discovery rate (FDR) < 0.05) between African and Asian dromedaries (Supplementary Fig. 1). We found it reasonable to investigate the biological functions of those genes harboring the SNPs as they might be relevant for the adaptation of dromedaries to their respective environments. We found SNPs in two genes, *CALN1* and *TREMI1*, which are responsible for calcium ion binding and amplifying inflammatory responses triggered by bacterial and fungal infections, respectively (scaffold:SNP-location; JWIN01030783.1:128274 and JWIN01033764.1:729703). In addition, we examined (potentially linked) regions 200 kbp upstream and downstream of the F_{ST} -outlier loci and detected fifty-three genes related to a number of biological functions (Supplementary Data 3). Interestingly, around one fifth of the detected genes had functions related to the immune system

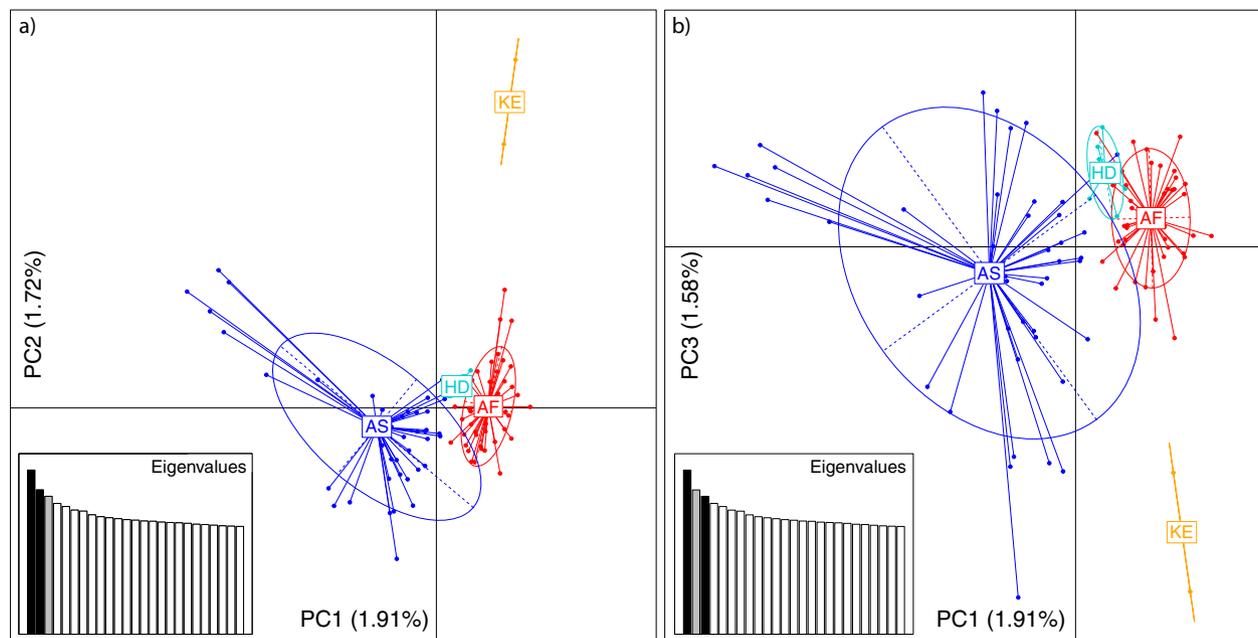


Fig. 1 Population substructure in the global dromedary population. Principal component (PC) analysis based on 22,721 SNPs to represent populations pre-classified according to phylogenetic clustering (Supplementary Fig. 2), Africa (AF), Asia (AS), *Hadhana* (HD), and Kenya (KE). Ellipses refer to the distribution of individuals within groups. The first three PCs explain 5.3% of the total variation and are shown in **a** PC1-PC2 and **b** PC1-PC3.

hinting to an adaptive process in response to different pathogens in the respective environments. Other protein coding gene functions were related to pathways such as circadian rhythm, (ga)lactose, metabolism, reproductive or various cellular and developmental processes. A full list of genes is presented in Supplementary Data 3. Signatures of selection related to photoperiod, metabolism, immunity and growth have also been observed in chicken²⁹, sheep³⁰, and cattle (*TGFB3*)³¹.

To understand the low genome-wide differentiation in dromedaries across their global range, we investigated population structure and admixture between populations. We projected the genetic variation of each dromedary on the first three axes inferred from a principal component analysis (PCA) and incorporated continental information (Africa/Asia) for each sample (Fig. 1). Principal component 1 (PC1) clearly separated African from Asian dromedaries, while PC2 and PC3 split Kenyan individuals from the rest of Africa and identified a single population from Saudi Arabia grouping closer to African than Asian individuals, although showing some cross-continental admixture (Fig. 1 and Supplementary Fig. 2). This separated population belongs to a specific breed, *Hadhana*, and is one of the twelve recognized dromedary ecotypes in Saudi Arabia, limited to mountain regions in the South of the Arabian Peninsula, Al-Baha³². In this case, the geographic accessibility might have an important role in the observed genetic distinctiveness. A possible explanation for the close relationship of *Hadhana* and African dromedaries might be the historic sea route from Jiddah in Saudi Arabia to Aydhab and Port Sudan. On the western coast of the Red Sea existed a trading route connecting the Horn of Africa to Petra and Damascus via Port Sudan, Aydhab and Myos Hormos, near today's Kosseir (Fig. 2)^{6,7}. In general, the Asian dromedary population showed higher genetic variability, although the genetic variation explained by the three first axes was rather low with only 5.3% (Fig. 1). While this could be a sign for ancestral variation (the Arabian Peninsula was a center of domestication¹), we cannot discard the hypothesis of post-domestication movements of camels or multiple origins of the founder populations as this would have left similar signals in the genomes.

We next inferred potential ancestry and admixture among Asian and African dromedaries using unsupervised genetic clustering in ADMIXTURE³³ (Fig. 3). Based on the lowest cross-validation error, the best clustering solution was 1 (Supplementary Fig. 3), which suggests a panmictic dromedary population and reflects the low genetic differentiation of 0.6% among individuals from different continents. Increasing the numbers of potential ancestral populations (K) from two to seven confirmed the already observed differentiation between African and Asian dromedaries ($K=2$), the clustering of the Saudi Arabian *Hadhana* breed with Africa ($K=4$), the separation of Kenyan and *Hadhana* individuals, and the higher number of distinct clusters on the Asian continent ($K=7$). We find a more homogenous gene pool in African animals with the exception of the East African group¹ (Figs. 1, 2 and Supplementary Fig. 2), represented in our dataset by the two Kenyan dromedaries. This can be a consequence of a random founder effect followed by lack of gene flow due to geographical, physiological (e.g., Trypanosome infestation) and/or cultural barrier, i.e., dromedaries in East Africa were dominantly used for milk production rather than transport or riding¹. There is a need to proceed with comprehensive analyses about the potential nature of natural and/or anthropogenic obstacles for gene flow between East African and other dromedaries.

Effective migration rates along ancient caravan routes. To formally test our qualitative observations of weak population structure among African and Asian dromedaries (Figs. 1, 3), we visualized the global spatial population structure using the Estimated Effective Migration Surfaces (EEMS) method³⁴. Based on a stepping-stone model, EEMS detected a corridor of significantly higher effective migration rates than the overall mean along the Mediterranean coast, connecting the Northern parts of Africa and the Arabian Peninsula until the border of the Arabian Desert (Fig. 2). This pattern shows a continuous gene flow throughout the coastal dromedary populations, and a lower than average migration in the inland desert populations. A known trading route which fits this observed effective migration pattern

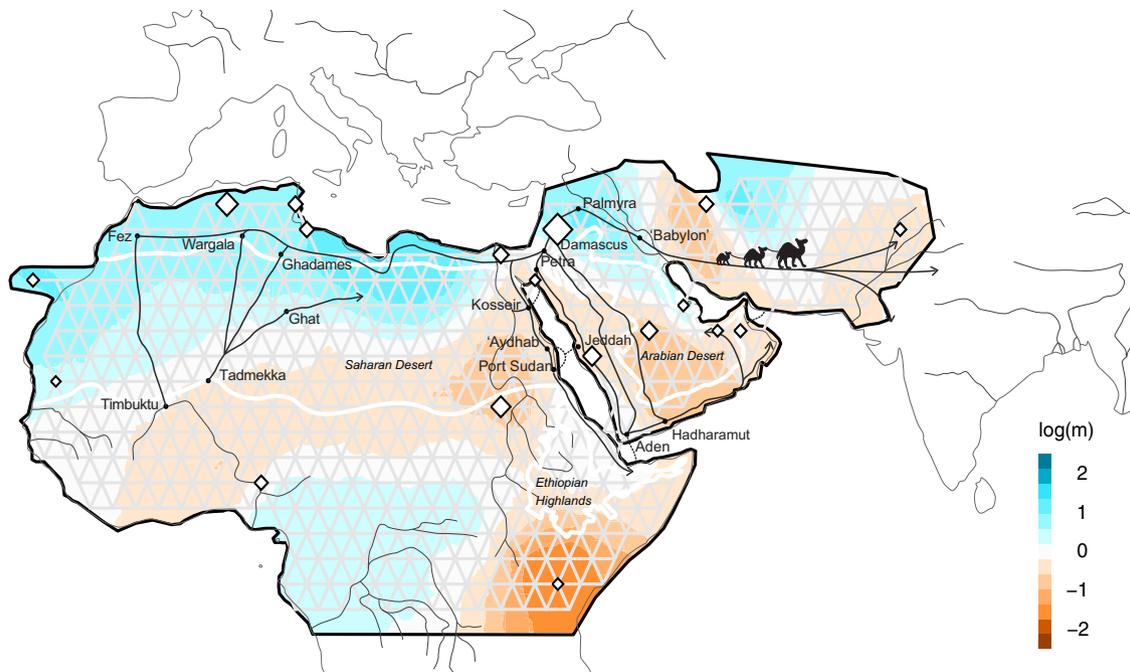


Fig. 2 Estimated Effective Migration Surfaces (EEMS) in the global dromedary population. EEMS plot representing the posterior mean of effective migration rates (m) (on a \log_{10} scale) across space. With this normalization, significantly higher than the overall average rates are represented in blue (“corridors”) and significantly lower than the overall average rate (“barriers”) are represented in brown. Zero corresponds to the overall mean migration rate. Samples are represented by diamonds and the size is proportional to the number of sampling. Approximate coordinates are used. Ethiopian Highlands and Arabian Desert are highlighted with white lines. Black lines represent historical network of caravan routes, i.e., Incense and Silk roads^{6,7} and main trans-Saharan gold trade networks³⁶ (adapted from²).

bordered the Mediterranean coast connecting Northwestern Africa to the North of the Arabian Peninsula from where caravans traveled toward Southern Asia along the Silk Road (Fig. 2)^{1,7}. The introduction of the dromedary into Northern Africa via the Sinai from Roman Egypt started in the early first millennium BCE and intensified in the Ptolemaic period^{6,15}. From there, dromedaries migrated along the Mediterranean coast, as archeological evidence dates their presence in Northwest Africa to the fourth to seventh century CE (Late Antiquity/Early Middle Ages)^{1,6}. Even earlier dispersal of taurine cattle along the Northern coast of Africa and through the Mediterranean sea to Europe was described during the Bronze age³⁵.

It is clear that camels, unlike other domesticated species, were able to penetrate deep into the Saharan desert and to connect trans-Saharan cultures. West Sahara belonged to an Islamic trading network classified as one of the major gold suppliers in the ninth to tenth centuries CE³⁶ (Fig. 2). Tadmekka, a territory located in the Southwestern Saharan desert and governed by the Tuareg, was operational by the eighth century CE and was one of the earliest towns established in the region where cross-Saharan camel caravans traded³⁷. These trades prolonged at least until the fourteenth century when Timbuktu, which similar to Tadmekka hosted large groups of Islamic traders, engaged in coin-based exchange economies across the Sahara³⁶.

While modern dromedaries along the western part of the Silk Road are still well connected today, the Iranian and Afghanistan deserts seem to present obstacles of effective migration. As EEMS assumes uniform migration rates the observed “barriers”, could however be assessed as areas of lower population density with fewer migrants exchanged per generation, producing an effective “barrier” to gene flow³⁴. The three main parallel itineraries of the Incense Routes through the Arabian Desert connecting the South of the Arabian Peninsula with the Levant⁷ also showed lower than average migration rates (Fig. 2), which could be interpreted

as lower population density or potential sampling gaps. These trading routes were essential during historical periods, not only for exchanging luxury products (e.g., incense or gems), but for trading everyday local products³⁸.

The strongest barrier detected in our dataset concerned dromedaries from the Horn of Africa, which had the lowest genetic effective migration rates (Fig. 2). Geographical isolation due to the Ethiopian highlands, which might disrupt gene flow with northern populations, and the Gulf of Aden would be the most likely explanation for the observed pattern. Genetic differentiation of livestock populations in East Africa has been described previously^{1,22,39}.

Late Pleistocene population decline and medieval expansions.

To complete our understanding of the moderate genome-diversity observed in the global dromedaries, we investigated the demographic history and inferred effective population size (N_E) over time with an Extended Bayesian Skyline approach (EBS)⁴⁰ (Fig. 4). As genetic differentiation between Asian and African dromedary populations was low (0.6%) but highly significant ($P < 0.001$), we tested whether the dromedary populations from the two continents experienced a similar demographic history. First, we investigated the global population and second, African and Asian individuals separately. With the latter approach, we accounted for a potential confounding effect of population structure for the inference of N_E ⁴⁰. Due to the observed substructure in Kenyan and *Hadhana* individuals, we excluded these two populations from the continental groups.

Irrespective of the continental origin, all inferences showed similar patterns of an initial population expansion from one million ybp until ~700,000 ybp (Fig. 4). Our genome-wide population approach confirmed previous N_E estimates based on a single dromedary whole genome sequence²³ using pairwise sequentially Markovian coalescent⁴¹. This population expansion

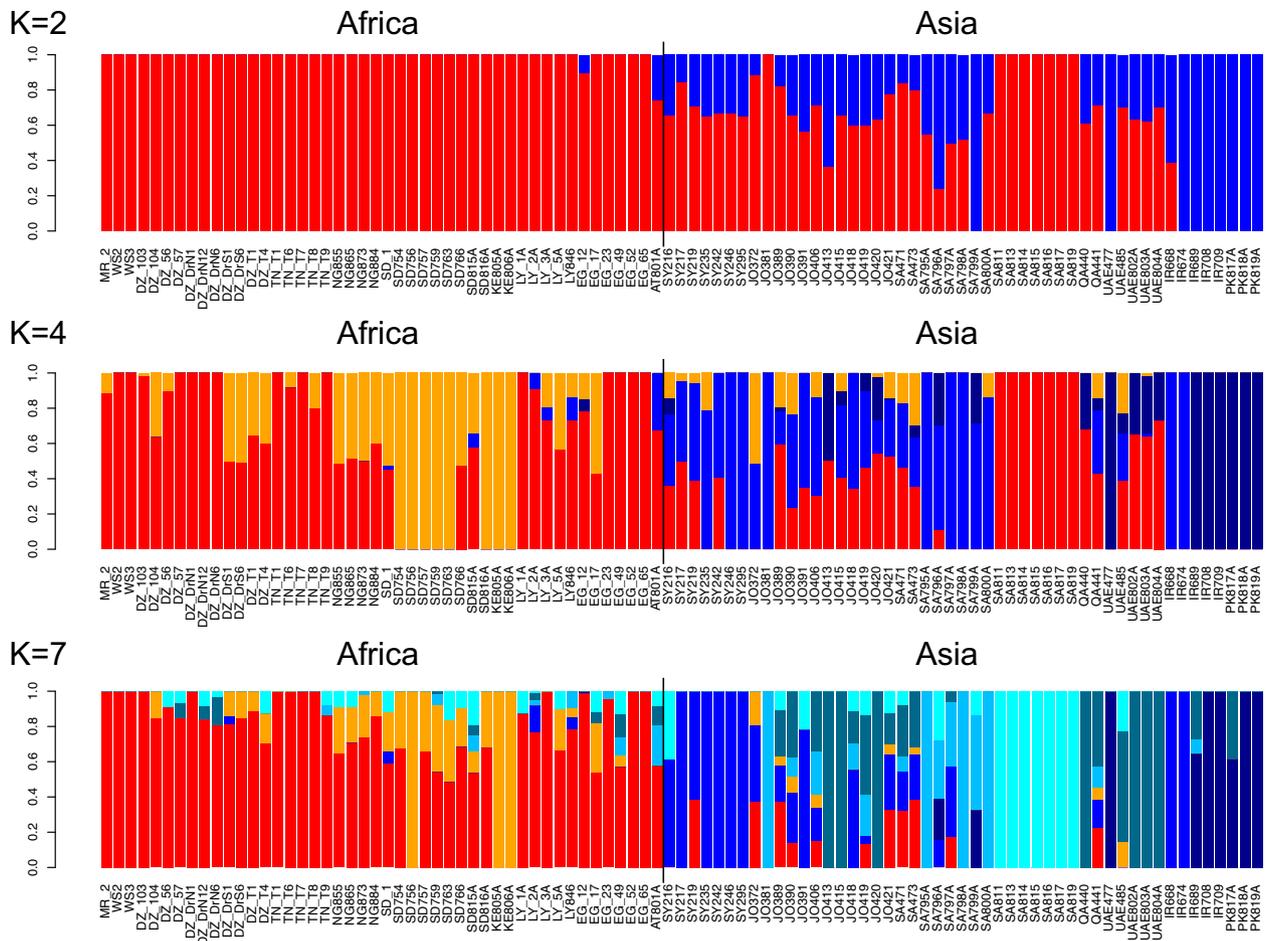


Fig. 3 Admixture analysis of the global dromedary population. Admixture analysis showing the proportion of potential ancestral populations ($K = 2; 4; 7$) for each individual (single bar). The geographical origins for each sample are represented below the bars in two-digit international country codes, where the middle line divides the first half (African countries) from the second half (Asian countries). Hadhana population is depicted in cyan blue ($K = 7$).

coincides with two remarkable periods: the middle Pleistocene transition (1.25–0.70 million ybp) characterized by climatic cycles, and the Galerian Mammal Age (1.2–0.60 million ybp), which influenced the distribution and evolution of biota and resulted in some species being adapted to arid, cold climates^{42,43}. Moreover, this timeframe also overlaps with the maximal diversity of the family Camelidae (early Galerian), supporting the adaptation of the dromedary ancestor to environmental changes with an expansion of its population during the middle Pleistocene transition^{23,44}. Population expansion was followed by a drastic decline in N_E beginning 700,000 ybp until the dromedary population collapsed during the last glacial period (LGP; 100,000–20,000 ybp)⁴⁵. This is a finding shared by previous Old World⁹ and New World camelid²³ studies and those focusing on Late Quaternary Megafauna⁴⁶.

Conversely, no bottleneck was picked up by any of the EBSPs during the time scale when dromedaries are predicted to have been first domesticated ~3000–4000 ybp^{15,16}. Previous BSP analysis using mtDNA likewise did not show a population decline during the time of domestication¹. It is possible that the detection of a bottleneck with the EBSP analysis related to domestication has been superimposed by the drastic decrease in N_E ending ~30,000 ybp (Fig. 4). Similar demographic changes were observed in alpacas²³, where three population bottlenecks were detected throughout the cold conditions of the LGM in South America, yet no bottleneck was visible during the domestication period.

After the Pleistocene bottleneck, the dromedary population slowly increased until reaching a stable N_E around 300 ybp, with a higher N_E present within Asia than in Africa. Demographic inferences based on mtDNA sequences described slightly earlier expansion of the maternal lineages around 600 ybp¹ associated with the rise of the Ottoman empire and the conquest of Constantinople (1453 CE), followed by the extension to Southern Asia and the Red Sea coasts⁴⁷.

Conclusions

Our study shows that assessing the evolutionary history of species using genome-wide approaches allows detailed inferences of population structure, migration, and potential signals of environmental adaptation. As the movements of dromedaries parallel those of humans, knowledge on dromedary spatial genetic signatures also sheds light into past human history. We detected genetic admixture across continental populations (Asia and Africa), which highlights the strong anthropogenic influence on these animals.

Human history is marked with the efforts for overcoming obstacles, be they of geographical (mountains, sea, and deserts) or cultural nature. Domestic animals, and in particular camels, have been linked to this process of human development and were essential for its success. By establishing trading routes and reusing them over millennia, corridors of gene flow were opened that shaped genetic diversity and structure not only in dromedaries,

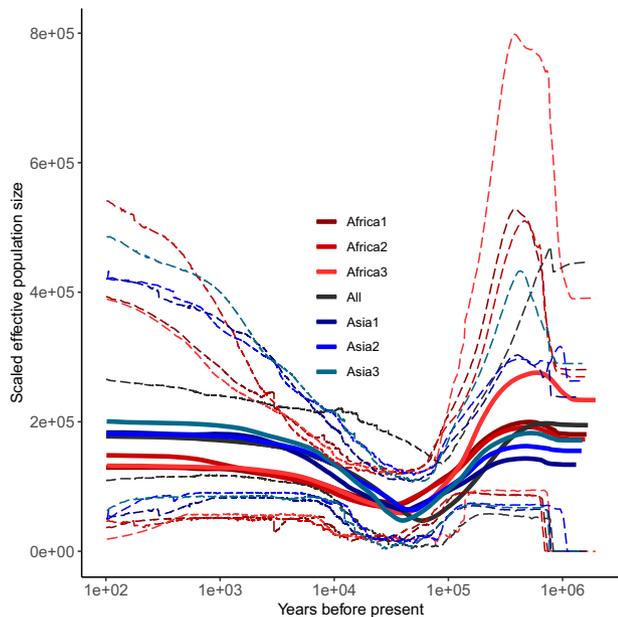


Fig. 4 Extended Bayesian Skyline Plot (EBSP) for the global, the African and the Asian dromedary population. EBSP for all dromedaries, with three independent runs per continent (Africa1, Africa2, and Africa3; Asia1, Asia2, and Asia3) and one for the global population (All). Solid lines represent median scaled effective population size and dashed lines represent 95% Highest Posterior Density (HPD) intervals. In each EBSP, 50 random ddRAD loci with at least four but not more than six SNPs across at least 75% individuals were used, and all EBSP runs were calibrated with a RAD locus-specific clock rate to calibrate the time scale. The lowest median effective population size for the different independent runs occurred 58,926 years before present (ybp) for the global population, 37,713, 26,869, and 38,802 ybp for African dromedaries and 41,923, 46,984, and 39,999 ybp for Asian dromedaries, respectively.

but also in Bactrian camels⁴⁸, and in many other (livestock) species^{49,50}. With a genome-wide dataset we detected patterns of high effective migration in the global dromedary population and revealed pathways of dispersal after domestication, mirroring ancient caravan routes (Fig. 3). While these served as corridors, deserts, and highlands (due to lower population density) represented possible barriers to gene flow. As we estimated an overall mean migration rate, the existing gene flow (below average) between populations along the caravan roads leading through desert regions might have been masked by the high connectivity (above average) of the coastal populations and/or lack of representative samples. Filling in population sampling gaps in future studies will provide a deeper understanding of the gene flow and genetic structure between populations. It will allow a powerful quantification of the magnitude of genetic isolation barriers that may persist.

Dromedaries and Bactrian camels are the most important livestock species in desert areas; their impact on land and water resources for food production is less than that of any other livestock. With increasing desertification and global climate change their importance will grow even more. For this reason, it is essential to understand the demographic history that has shaped modern dromedary populations. The genome-wide diversity present today, which we have characterized in this study, is a result of genetic shuffling due to historic and recent movements along trading routes. This constant mixing might have led to a unique genetic makeup that could make camels more resilient to global environmental changes, and that needs to be preserved.

Methods

DNA samples. We selected a total of 122 dromedary samples from a pool of previously extracted DNA collected during projects supported from the Austrian Science Foundation, FWF P24607-B25 (PI: P.B.) and the European Research Council, EU ENPI CBC MED PROCAMED I.B.1.1/493 (PI: E.C.) under all legal requirements. The samples originated from 18 countries, which were representative of the species' distributional range. In view of potential hybridization between one- and two-humped camels known to occur specifically in Central Asian regions, we included a Bactrian camel to control for introgression and as outgroup for phylogenetic analysis. Detailed information about all samples is provided in Supplementary Data 1.

Library preparation, sequencing, and initial data filtering. Library preparation and sequencing was performed at IGA Technology Services (Udine, Italy), to generate genome-wide data from ddRADseq. In silico analysis of the *C. dromedarius* genome assembly (NCBI accession: GCA_000803125.1)²⁴ highlighted *SphI-BstYI* as the best combination of restriction enzymes able to produce DNA fragments between 400 and 530 bp. ddRAD barcoded libraries were pooled and sequenced on an Illumina HiSeq 2500 in high output mode using 125 bp paired-end reads. Initial raw data analysis as well as SNP calling was performed by IGA Technology Services in-house bioinformatics pipeline. Briefly, all reads were trimmed to 110 bp, and quality controlled reads were aligned to the North African dromedary genome assembly (NCBI accession: GCA_000803125.1) using BWA-MEM⁵¹. ddRAD reads were then processed with Stacks v.1.35⁵². Out of nearly half a million RAD loci, 88,836 SNPs passed the imposed filtering criteria. The *pstack* module was run with a minimum coverage of 3 reads to call a haplotype, while SNPs were called using a bounded model to highest = 0.5 and alpha = 0.05. *Cstacks* and *sstacks* modules were run with default parameters. Population module was run requiring a minimum genotyping of 75% of individuals to score loci, along with a calling likelihood filtering threshold of -25.

The retrieved raw SNP data were stringently filtered for missing values with PLINK 1.07⁵³, first excluding individuals with more than 25% missing genotypes (--mind 0.25), next setting a threshold of 0.01 for minor allele frequency (--maf 0.01) and finally, removing SNPs with a missing genotype rate of more than 25% (--geno 0.25). Due to high individual missingness, 13 samples were removed from the dataset as well as 35,013 SNPs after filtering for minor allele frequency and missing genotype rate. To exclude any potentially related individuals, a symmetric identical by state matrix was created with PLINK with a cutoff value of 0.05; one sample from Jordan (JO434) and three from Nigeria (NG852, NG877 and NG887) were removed from further analyses (Supplementary Data 1).

Considering the practice of dromedary and Bactrian camel crossbreeding especially in Central Asian countries¹⁰, we screened for potential hybridization with Bactrian camel present in the dataset. Paired-end ddRAD reads from all 122 dromedaries were simultaneously mapped to either the dromedary (NCBI accession: GCA_000803125.2)²⁵ or Bactrian camel (NCBI accession: GCF_000767855.1)²³ genomes using BBSplit v. 38.79 (<https://sourceforge.net/projects/bbmap/>), with the following settings: minratio = 1.0, ambiguous = toss, ambiguous2 = toss. We preprocessed these two genome assemblies with dustmasker v. 1.0⁵⁴, and the percent Bactrian camel was estimated as the number of reads that unambiguously mapped to the Bactrian camel genome divided by the total number of unambiguously mapped reads to both dromedary and Bactrian camel multiplied by 100. We removed three individuals from Iran (IR715, IR717, and IR719) and six from Kazakhstan (KZ888, KZ889, KZ890, KZ891, KZ892, and KZ893) (Supplementary Data 2) that had "far-out" values, which are those greater than the third quartile plus the interquartile range multiplied by three⁵⁵.

Genome-wide summary statistics and population structure. SNPs were tested for Hardy-Weinberg Equilibrium and linkage disequilibrium using VCFTOOLS v0.1.15⁵⁶ and as no SNP exceeded a FDR of 0.05, all were retained. For subsequent file conversions, PGDSPIDER version 2.1.1.35⁷ was used. The expected (H_E) and observed (H_O) heterozygosities, AR, and inbreeding coefficients (F_{IS}) were calculated with the R package Hierfstat v0.04-22^{58,59}. We have used a parametric Welch *t* test implemented in R v.3.5.1 using the *t.test* function to compare mean H_E , H_O , and F_{IS} between African and Asian dromedaries. Pairwise F_{ST} and AMOVA⁵⁹ were analyzed with the program Arlequin 3.5.2.2⁶⁰.

We applied BayeScan 2.1²⁸ to identify F_{ST} outlier loci putatively under of selection using default settings with a FDR⁶¹ of 0.05. To understand if SNPs putatively detected under selection were linked to significant biological pathways, we screened the respective RAD sequences for genes using the annotation of the new CamDro2 assembly²⁵ and assessed their protein function with GeneCards (<http://www.genecards.org>). To consider also regions in potential linkage disequilibrium, we included genes 200 kbp upstream and downstream of the SNPs under selection in the analysis.

Individual-based PCA was performed with ADEGENET v2.1.1 using the *s.class* option to represent principal components of known groups. Furthermore we applied ADMIXTURE v1.3³³ to assess ancestry and possible structure among dromedary populations (i.e., countries of origin), using the lowest fivefold cross-validation error to choose the best number of clusters (K), from $K = 1$ to $K = 10$. To understand the phylogenetic relationship among individuals, we applied the Neighbor-Net method⁶² implemented in SplitsTree4⁶³, which is a neighbor-joining

algorithm for constructing phylogenetic networks from a genetic (allele sharing) distance matrix created in PLINK.

Estimating effective migration rates. We used the software estimated effective migration surfaces (EEMS)³⁴ to investigate effective migration patterns in the global dromedary population. Based on a stepping-stone model, and assuming that migration is symmetric, EEMS uses Markov Chain Monte Carlo to estimate the diversity and migration parameters and produces maps which represent the posterior mean of effective migration and effective diversity across space. We performed three distinct runs, each consisting of 10 million MCMC iterations, discarding the initial 5 million as burn-in and saving every 49,995 iterations for a grid with 500 demes. All runs reached convergence and results were similar across replicates. The habitat polygon was obtained using Google Maps API v3 Tool (<http://www.birdtheme.org/useful/v3tool.html>) and results were plotted using the R package rEEMSpots as suggested in Petkova et al.³⁴.

Demographic analysis of the global dromedary population. We assessed the demographic history of the species employing a coalescent-based multi-locus analysis with variable loci using BEAST2 v. 2.5.1⁶⁴, setting the Coalescent EBS⁴⁰ as a tree prior and following Huson and Bryant⁶³. EBS analysis was conducted on the global population where we randomly selected 50 RAD loci containing at least four but not more than six SNPs across at least 75% of the individuals using a custom R script (https://github.com/jelber2/RAD-Scripts/blob/master/RAD_Haplotypes.R)⁶⁵. We repeated the EBS analysis for African and Asian dromedaries separately, but due to slight population structuring, Kenya and *Hadhana* populations were excluded from this analysis. We ran EBS analyses three times per continent, with 50 random RAD loci used in each run. A RAD locus-specific clock rate (per generation) was estimated by calculating the average number of differences between dromedary and Bactrian camel sequences, dividing by the length of the RAD loci, taking the average among SNP classes (number of SNPs per RAD locus from 4 to 6), dividing by the split time between the Bactrian camel and dromedary of ~4.4 million ybp²³, and using a dromedary generation time of 5 years¹. Each EBS was run for 2,100,000,000 chains using the RAD locus-specific clock rate of 1.809442e-08 to calibrate the time scale.

Statistics and reproducibility. Sample size is outlined in Supplementary Data 1. We have used a parametric Welch *t* test implemented in R v.3.5.1 using the *t.test* function to compare mean H_E , H_O , and F_{IS} , and data are expressed as mean and SD. Pairwise F_{ST} values were performed including a minimum of three individuals per countries and their significance levels are represented with “+”. AMOVA were analyzed with the program Arlequin. *P* values below 0.05 are considered as statistically significant for all statistical tests in this work. All analyses are reproducible with access to genetic data (see “Data availability”).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All sequence files (.cram) are deposited at the European Nucleotide Archive with the accession number PRJEB38954 (<http://www.ebi.ac.uk/ena/data/view/PRJEB38954>). In addition, SNP data (.map and .ped) can be downloaded from Dryad⁶⁶ (<https://doi.org/10.5061/dryad.kh189322q>).

Code availability

Computer code and scripts for the various analyses are available at Dryad⁶⁶ (<https://doi.org/10.5061/dryad.kh189322q>).

Received: 17 December 2019; Accepted: 19 June 2020;

Published online: 16 July 2020

References

- Almathen, F. et al. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc. Natl Acad. Sci. USA* **113**, 6707–6712 (2016).
- de Barros Damgaard, P. et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, 1–9 (2018).
- Trevor-Roper, H. *The Invention of Tradition: the Highland Tradition of Scotland* (Cambridge University Press, Cambridge, 1983).
- Ramadan, S. & Inoue-Murayama, M. Advances in camel genomics and their applications: a review. *J. Anim. Genet.* **45**, 49–58 (2017).
- Sala, R. The domestication of camel in the literary, archaeological and petroglyph records. *J. Arid L. Stud.* **26**, 205–211 (2017).
- Bulliet, R. W. *The Camel and the Wheel* (Columbia University Press, 1975).
- Heiss, J. Caravans from South Arabia. in *Camels in Asia and North Africa. Interdisciplinary Perspectives on their Significance in Past and Present* (eds Knoll, E. M. & Burger, P.) 131–139 (Verlag der Österreichischen Akademie der Wissenschaften, 2012).
- Wilson, R. T. *The Camel* (Longman, London, 1984).
- Burger, P. A. The history of Old World camels in the light of molecular genetics. *Trop. Anim. Health Prod.* **48**, 905–913 (2016).
- Faye, B. & Konuspayeva, G. The encounter between Bactrian and dromedary camels in Central Asia. in *Camels in Asia and North Africa. Interdisciplinary Perspectives on their Significance in Past and Present* (eds Knoll, E. M. & Burger, P.) 29–36 (Verlag der Österreichischen Akademie der Wissenschaften, 2012).
- Çakırlar, C. & Berthon, R. Caravans, camel wrestling and cowrie shells: towards a social zooarchaeology of camel hybridization in Anatolia and adjacent regions. *Anthropozoologica* **49**, 237–252 (2014).
- Warmuth, V. M., Campaña, M. G., Eriksson, A. & Bower, M. I. M. Ancient trade routes shaped the genetic structure of horses in eastern Eurasia. *Mol. Ecol.* **22**, 5340–5351 (2013).
- Liu, X. *The Silk Road in World History* (Oxford University Press, 2010).
- Burger, P. A., Ciani, E. & Faye, B. Old World camels in a modern world—a balancing act between conservation and genetic improvement. *Anim. Genet.* **50**, 598–612 (2019).
- Uerpmann, H.-P. & Uerpmann, M. The appearance of the domestic camel in south-east Arabia. *J. Oman Stud.* **12**, 235–260 (2002).
- Uerpmann, H.-P. & Uerpmann, M. Archeozoology of camels in South-Eastern Arabia. in *Camels in Asia and North Africa. Interdisciplinary Perspectives on their Significance in Past and Present* (eds Knoll, E. M. & Burger, P.) 109–122 (Österreichischen Akademie der Wissenschaften, 2012).
- Von den Driesch, A. & Obermaier, H. The hunt for wild dromedaries during the 3rd and 2nd millennia BC on the United Arab Emirates coast. Camel bone finds from the excavations at Al Sufouh 2 Dubai, UAE. in *Documenta Archaeobiologiae Bd. 5—Skeletal series and their socio-economic context.* (eds Grupe, G. & Peters, J.) 133–167 (Verlag Marie Leidorf GmbH, Rahden (Germany), 2007).
- Grigson, C. Copper and Donkeys in the Early Iron Age of the Southern Levant: Timna Revisited. *Levant* **44**, 82–100 (2012).
- Grigson, C. The history of the camel bone dating project. *Anthropozoologica* **49**, 225–235 (2014).
- Spencer, P. B. S. & Woolnough, A. P. Assessment and genetic characterisation of Australian camels using microsatellite polymorphisms. *Livest. Sci.* **129**, 241–245 (2010).
- Jones, P. & Kenny, A. *Australia's Muslim Cameleers: Pioneers of the Inland, 1860s–1930s* (Wakefield Press, 2010).
- Mburu, D. N. et al. Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classification. *Anim. Genet.* **34**, 26–32 (2003).
- Wu, H. et al. Camelid genomes reveal evolution and adaptation to desert environments. *Nat. Commun.* **5**, 1–9 (2014).
- Fitak, R. R., Mohandesan, E., Corander, J. & Burger, P. A. The de novo genome assembly and annotation of a female domestic dromedary of North African origin. *Mol. Ecol. Resour.* **16**, 314–324 (2016).
- Elbers, J. P. et al. Improving Illumina assemblies with Hi-C and long reads: an example with the North African dromedary. *Mol. Ecol. Resour.* **19**, 1015–1026 (2019).
- Kijas, J. W. et al. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* **4**, 1–13 (2009).
- Makina, S. O. et al. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Front. Genet.* **5**, 1–7 (2014).
- Foll, M. & Gaggiotti, O. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993 (2008).
- Rubin, C. et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591 (2010).
- Larrañaga, O. R. et al. Genomic selection signatures in sheep from the Western Pyrenees. *Genet. Sel. Evol.* **50**, 9–21 (2018).
- Gurgul, A. et al. Identification of genome-wide selection signatures in the Limousin beef cattle breed. *J. Anim. Breed. Genet.* **133**, 264–276 (2016).
- Faye, B., Abdallah, H. R., Almathen, F. S., Harzallah, B. D. & Al-Mutairi, S. E. Camel biodiversity. *Camel phenotypes in the Kingdom of Saudi Arabia* (FAO, Rome, 2011).
- Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.* **48**, 94–100 (2016).
- Beja-Pereira, A. et al. The origin of European cattle: evidence from modern and ancient DNA. *PNAS* **103**, 8113–8118 (2006).

36. Nixon, S., Rehren, T. & Guerra, M. F. New light on the early Islamic West African gold trade: coin moulds from Tadmekka, Mali. *Antiquity* **85**, 1353–1368 (2011).
37. Nixon, S. Excavating Essouk-Tadmakka (Mali): new archaeological investigations of early Islamic trans-Saharan trade. *Azania Archaeol. Res. Afr.* **44**, 217–255 (2009).
38. Gingrich, A. Honig und tribale Gesellschaft: Historischer Hintergrund, sozialer Gebrauch und traditionelle Erzeugung im südlichen Hijaz. in *Tribale Gesellschaften der südwestlichen Regionen des Königreiches Saudi Arabien: sozialanthropologische Untersuchungen*, Vol. 732 (eds Gingrich, A., Heiss, J. & Zötl, J.) 173–215 (Verlag der Österreichischen Akademie der Wissenschaften, 2006).
39. Gifford-gonzalez, D. & Hanotte, O. Domesticating animals in Africa: implications of genetic and archaeological findings. *J. World Prehist.* **24**, 1–23 (2011).
40. Heled, J. & Drummond, A. J. Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* **8**, 289 (2008).
41. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
42. Head, M. J., Pillans, B. & Farquhar, S. A. The Early–Middle Pleistocene Transition: characterization and proposed guide for the defining boundary. *Episodes* **31**, 255–259 (2008).
43. Rook, L. & Martínez-Navarro, B. Villafranchian: the long story of a Plio-Pleistocene European large mammal biochronologic unit. *Quat. Int.* **219**, 134–144 (2010).
44. Vislobokova, I. A. Main stages in evolution of Artiodactyla Communities from the Pliocene—Early Middle Pleistocene of Northern Eurasia: Part 2. *Paleontol. J.* **42**, 414–424 (2008).
45. Yokoyama, Y., Lambeck, K., Deckker, P., De, Johnston, P. & Fifield, L. K. Timing of the last glacial maximum from observed sea-level minima. *Nature* **406**, 713–716 (2000).
46. Stuart, A. J. Late quaternary megafaunal extinctions on the continents: a short review. *Geol. J.* **50**, 338–363 (2015).
47. Kennedy, H. *The Great Arab Conquests: How the Spread of Islam Changed the World We Live in* (Da Capo Press, 2007).
48. Ming, L. et al. Whole-genome sequencing of 128 camels across Asia reveals origin and migration of domestic Bactrian camels. *Commun. Biol.* **3**, 1–9 (2020).
49. Fages, A. et al. Tracking five millennia of horse management with extensive ancient genome time series. *Cell* **177**, 1419–1435 (2019).
50. Edwards, C. J., Baird, J. F. & Machugh, D. E. Taurine and zebu admixture in Near Eastern cattle: a comparison of mitochondrial, autosomal and Y-chromosomal data. *Anim. Genet.* **38**, 520–524 (2007).
51. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
52. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
53. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
54. Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13**, 1028–1040 (2006).
55. Tukey, J. W. *Exploratory Data Analysis*, Vol. 2 (1977).
56. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
57. Lischer, H. E. L. & Excoffier, L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299 (2012).
58. Goudet, J. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
59. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application. *Genet. Soc. Am.* **131**, 479–491 (1992).
60. Excoffier, L. & Lischer, H. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
61. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
62. Bryant, D. & Moulton, V. Neighbor-Net, an agglomerative algorithm for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21**, 255–265 (2002).
63. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
64. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
65. Trucchi, E. et al. King penguin demography since the last glaciation inferred from genome-wide data. *Proc. R. Soc. B Biol. Sci.* **281**, 1–8 (2014).
66. Lado, S. et al. Dryad repository for computer codes and scripts. *Dryad* <https://doi.org/10.5061/dryad.kh189322q> (2020).

Acknowledgements

We are very grateful to all veterinarian colleagues and camel owners for their agreements to submit sample aliquots for scientific purposes. We thank Steve Smith for lab support and Jukka Corander for general discussion. Samples collection was performed during the Austrian Science Fund (FWF) project number P24607-B25 (to P.B.) and the EU project EU ENPI CBC MED Project PROCAMED. B.1.1/493 (to E.C.); ddRAD sequencing was performed in the frame of the EU project EU ENPI CBC MED Project PROCAMED. B.1.1/493 (to E.C.); S.L. and J.P.E. acknowledge funding from the FWF project number P29623-B25 (to P.B.).

Author contributions

S.L. performed analysis and wrote the first draft of the paper, J.P.E., A.D., D.S., E.T., E.C., and P.B. performed analyses, F.A., N.S., and M.H.B. contributed essential samples, E.C. and P.B. conceived and managed the project and wrote the paper. All authors provided valuable discussions, commented, and approved the final paper.

Competing interests

All authors declare no competing non-financial interests. Author D.S. is employed by the sequencing company IGA Technologies, Udine, Italy. All other authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-020-1098-7>.

Correspondence and requests for materials should be addressed to E.C. or P.A.B.

Reprints and permission information is available at <http://www.nature.com/reprints>

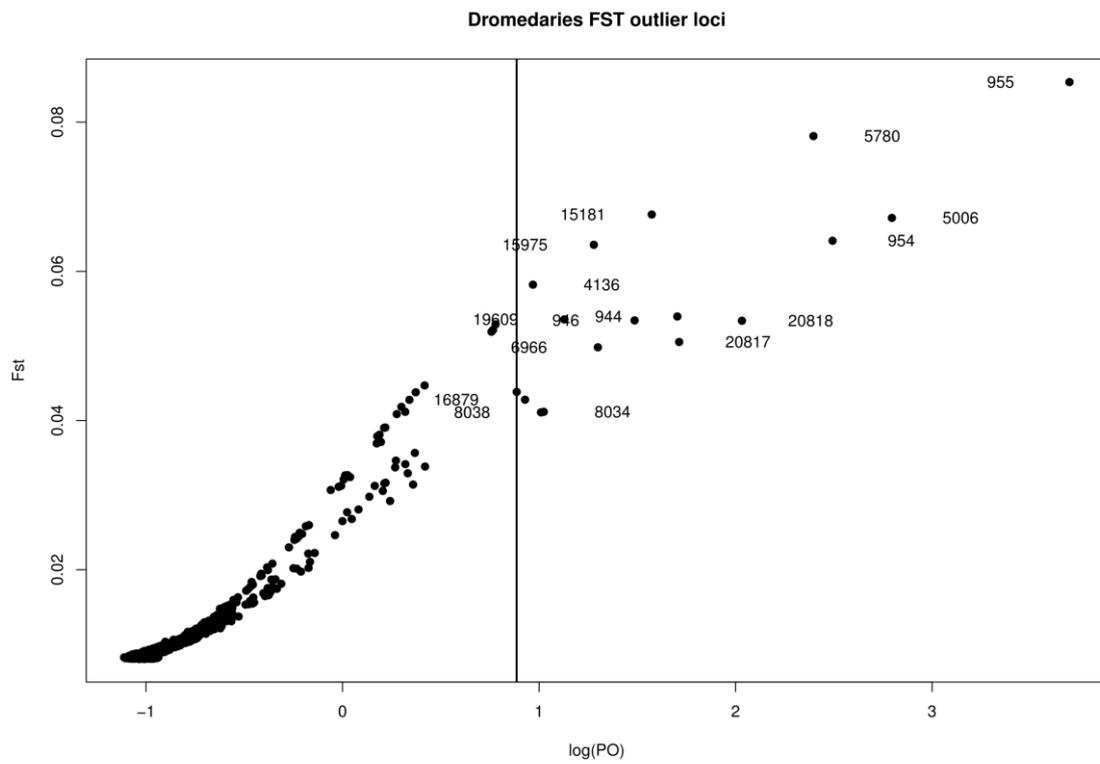
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



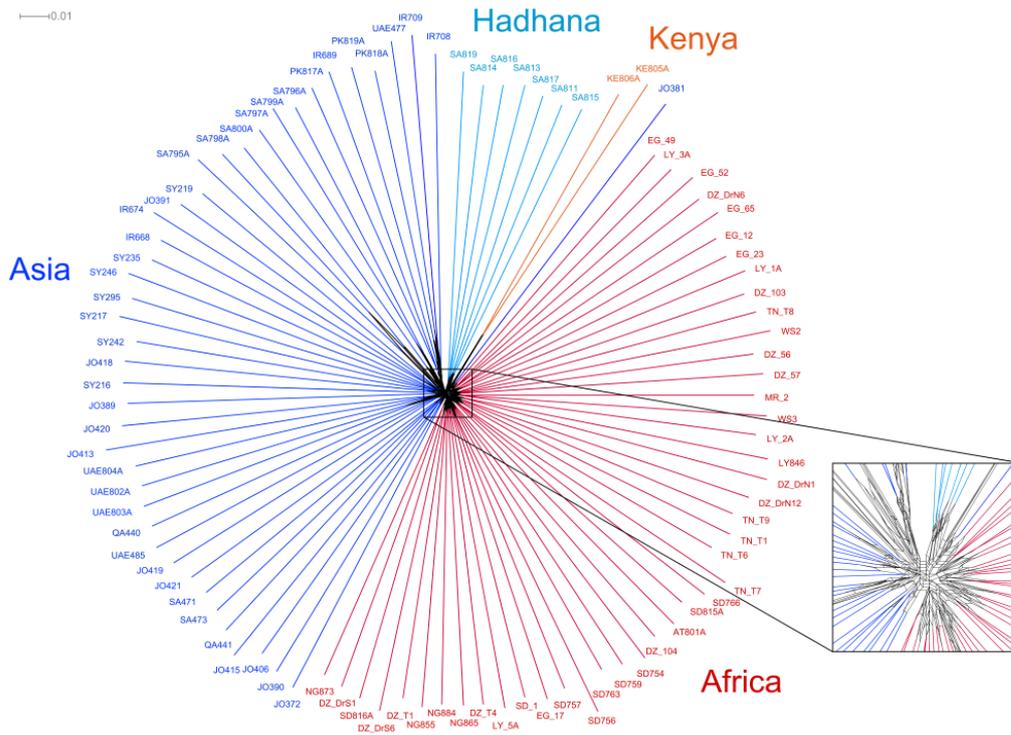
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

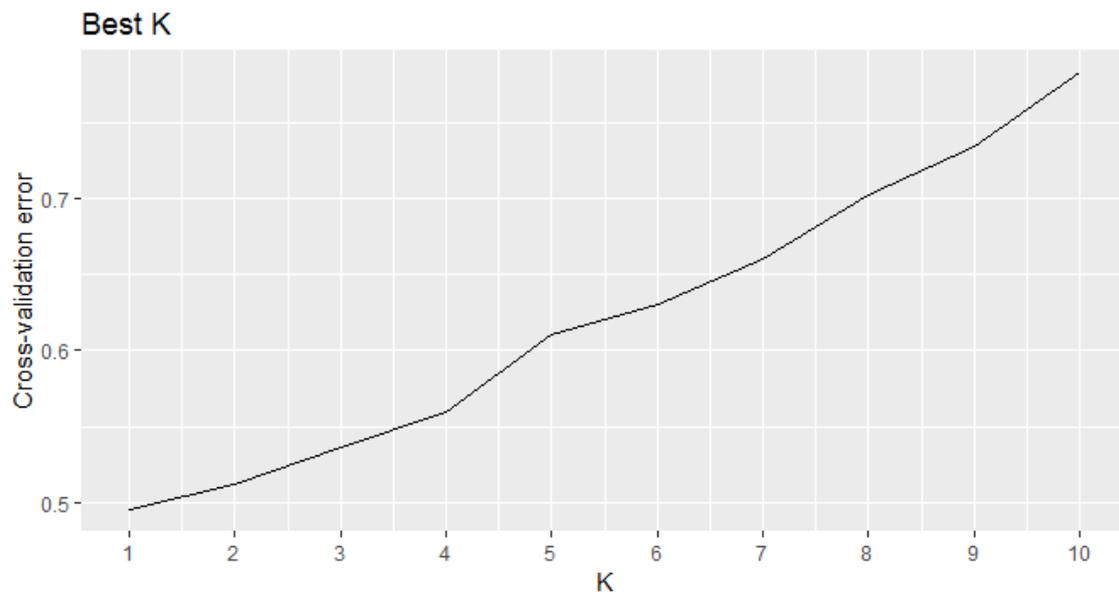
Supplementary Materials Supplementary Figures



Supplementary Figure 1. *F_{ST}* outliers. Sixteen loci putatively under selection detected using Bayescan (*q*-value lower than 5%). PO = posterior odds²⁸.



Supplementary Figure 2. SplitsTree network. Phylogenetic network calculated with SplitsTree using Neighbour-net, with a zoom to visualize the split between Africa, Asia, Hadhana and Kenyan dromedaries.



Supplementary Figure 3. Admixture's best K according to cross-validation error from K=1 to K=10.

Supplementary Table 1. Allelic richness (AR), observed and expected heterozygosities (H_o , H_E) and inbreeding coefficients (F_{IS}) for the total population, per continent and per country. AR calculations for Africa and Asia were corrected for 46 individuals minimum; AR calculations for countries were corrected for one individual minimum. SD corresponds to the standard deviation.

N	Countries	AR	SD	H_o	SD	H_E	SD	F_{IS}	SD
95	All	-	-	0.25	0.17	0.27	0.17	0.06	0.17
46	Africa	1.97	0.16	0.25	0.18	0.26	0.17	0.058	0.2
49	Asia	1.98	0.13	0.25	0.17	0.27	0.17	0.063	0.2
1	MR	1.26	0.44	0.26	0.44	NA	NA	NA	NA
2	WS	1.26	0.28	0.25	0.33	0.26	0.3	-0.06	0.56
11	DZ	1.26	0.19	0.25	0.21	0.26	0.19	0.03	0.3
5	TN	1.26	0.21	0.25	0.25	0.26	0.21	0.01	0.38
4	NG	1.26	0.22	0.25	0.27	0.26	0.23	-0.01	0.42
9	SD	1.25	0.2	0.23	0.21	0.26	0.2	0.07	0.34
2	KE	1.21	0.28	0.19	0.3	0.23	0.31	0.02	0.59
5	LY	1.26	0.21	0.26	0.25	0.27	0.22	0.01	0.4
6	EG	1.26	0.21	0.25	0.24	0.26	0.21	0.02	0.37
1	AT	1.26	0.44	0.26	0.44	NA	NA	NA	NA
7	SY	1.26	0.2	0.23	0.22	0.26	0.2	0.08	0.37
12	JO	1.26	0.19	0.24	0.2	0.27	0.19	0.07	0.3
15	SA	1.26	0.18	0.25	0.2	0.26	0.18	0.03	0.26
2	QA	1.27	0.29	0.27	0.34	0.27	0.3	-0.1	0.56
5	UAE	1.26	0.21	0.25	0.25	0.26	0.21	0.02	0.39
5	IR	1.27	0.21	0.25	0.24	0.27	0.21	0.05	0.4
3	PK	1.26	0.25	0.26	0.3	0.26	0.25	-0.04	0.48

NA due to low number of individuals

Supplementary Table 2. AMOVA. Groups correspond to Africa and Asia, populations correspond to countries.

Source of variation	d.f.	Sum of squares	Variance components	Percentage	<i>P</i> value
Among groups	1	3404.275	9.45527	0.45	< 0.005
Among populations within groups	15	35787.158	19.85303	0.94	< 0.005
Among individuals within populations	78	169914.609	91.32516	4.32	< 0.05
Within individuals	95	189595.5	1995.74211	94.30	< 0.005

Fixation Indices	
F_{IS}	0.04376
F_{SC}	0.00942
F_{CT}	0.00447
F_{IT}	0.05700

Supplementary Table 3. Pairwise F_{ST} values between dromedary populations from different countries including a minimum of three individuals (below the diagonal) and their significance levels (+) $P < 0.05$ (above the diagonal).

	DZ	TN	NG	SD	LY	EG	SY	JO	SA	UAE	IR	PK
DZ		-	-	-	-	+	+	+	+	+	+	+
TN	0.0045		-	+	-	-	+	+	+	-	-	-
NG	0.0008	0.0102		-	-	-	+	+	-	-	-	-
SD	0.0083	0.0130	0.0037		-	-	-	-	-	-	-	-
LY	0.0002	0.0021	0.0044	0.0082		-	+	-	-	-	-	-
EG	0.0051	0.0061	0.0085	0.0110	0.0015		+	-	+	+	+	-
SY	0.0121	0.0128	0.0122	0.0148	0.0060	0.0097		-	-	+	-	-
JO	0.0060	0.0070	0.0066	0.0111	0.0019	0.0052	0.0039		-	-	-	+
SA	0.0090	0.0101	0.0101	0.0141	0.0080	0.0106	0.0087	0.0050		-	-	-
UAE	0.0148	0.0158	0.0152	0.0184	0.0136	0.0160	0.0142	0.0047	0.0131		-	-
IR	0.0210	0.0223	0.0207	0.0240	0.0157	0.0201	0.0062	0.0074	0.0156	0.0070		-
PK	0.0290	0.0328	0.0311	0.0343	0.0230	0.0267	0.0217	0.0160	0.0251	0.0137	0.0030	

Description of additional supplementary files

Supplementary Data 1. Detailed sample information.

Supplementary Data 2. Number of paired-end ddRAD reads mapping unambiguously to either dromedary or Bactrian camel genome for all 123 dromedary and Bactrian camel samples.

Supplementary Data 3. Identification of genes located within 200kbp upstream and downstream of loci under putative selection, with respective functions of loci under putative selection. In bold are depicted the genes related to immunity.

Supplementary data

Supplementary Table 1. Samples' information list.

	Sample code	Original code	Species	Sampling area	Analysed
1	AT801A	Drom801A	Dromedary	Austria (North African origin)	X
2	DC156	DC156	Bactrian camel	Austria	
3	DZ_103	Algeria_103	Dromedary	Algeria	X
4	DZ_104	Algeria_104	Dromedary	Algeria	X
5	DZ_56	Algeria_56	Dromedary	Algeria	X
6	DZ_57	Algeria_57	Dromedary	Algeria	X
7	DZ_DrN1	Algeria_DrN1	Dromedary	Algeria	X
8	DZ_DrN12	Algeria_DrN12	Dromedary	Algeria	X
9	DZ_DrN6	Algeria_DrN6	Dromedary	Algeria	X
10	DZ_DrS1	Algeria_DrS1	Dromedary	Algeria	X
11	DZ_DrS6	Algeria_DrS6	Dromedary	Algeria	X
12	DZ_T1	Algeria_T1	Dromedary	Algeria	X
13	DZ_T4	Algeria_T4	Dromedary	Algeria	X
14	EG_106	Egypt_106	Dromedary	Egypt	
15	EG_12	Egypt_12	Dromedary	Egypt	X
16	EG_17	Egypt_17	Dromedary	Egypt	X
17	EG_23	Egypt_23	Dromedary	Egypt	X
18	EG_49	Egypt_49	Dromedary	Egypt	X
19	EG_52	Egypt_52	Dromedary	Egypt	X
20	EG_65	Egypt_65	Dromedary	Egypt	X
21	EG_66	Egypt_66	Dromedary	Egypt	
22	IR668	Drom668	Dromedary	Iran	X
23	IR674	Drom674	Dromedary	Iran	X
24	IR688	Drom688	Dromedary	Iran	
25	IR689	Drom689	Dromedary	Iran	X
26	IR708	Drom708	Dromedary	Iran	X
27	IR709	Drom709	Dromedary	Iran	X
28	IR715	Drom715	Dromedary	Iran	
29	IR717	Drom717	Dromedary	Iran	
30	IR719	Drom719	Dromedary	Iran	
31	JO372	Drom372	Dromedary	Jordan, Aqaba	X
32	JO381	Drom381	Dromedary	Jordan, Aqaba	X
33	JO389	Drom389	Dromedary	Jordan, Mafraq	X
34	JO390	Drom390	Dromedary	Jordan, Mafraq	X
35	JO391	Drom391	Dromedary	Jordan, Mafraq	X

36	JO405	Drom405	Dromedary	Jordan, Irbid	
37	JO406	Drom406	Dromedary	Jordan, Irbid	X
38	JO413	Drom413	Dromedary	Jordan, Irbid	X
39	JO415	Drom415	Dromedary	Jordan, Irbid	X
40	JO418	Drom418	Dromedary	Jordan, Irbid	X
41	JO419	Drom419	Dromedary	Jordan, Irbid	X
42	JO420	Drom420	Dromedary	Jordan, Irbid	X
43	JO421	Drom421	Dromedary	Jordan, Irbid	X
44	JO433	Drom433	Dromedary	Jordan, Mafraq	
45	JO434	Drom434	Dromedary	Jordan, Mafraq	
46	KE805A	Drom805A	Dromedary	Kenya	X
47	KE806A	Drom806A	Dromedary	Kenya	X
48	KZ888	Drom888	Dromedary	Kazakhstan	
49	KZ889	Drom889	Dromedary	Kazakhstan	
50	KZ890	Drom890	Dromedary	Kazakhstan	
51	KZ891	Drom891	Dromedary	Kazakhstan	
52	KZ892	Drom892	Dromedary	Kazakhstan	
53	KZ893	Drom893	Dromedary	Kazakhstan	
54	LY_1A	Libya_1A	Dromedary	Libya	X
55	LY_2A	Libya_2A	Dromedary	Libya	X
56	LY_3A	Libya_3A	Dromedary	Libya	X
57	LY_4A	Libya_4A	Dromedary	Libya	X
58	LY846	Drom846	Dromedary	Libya	X
59	MR_1	Mauretania_1	Dromedary	Mauritania	X
60	MR_2	Mauretania_2	Dromedary	Mauritania	
61	MR_4	Mauretania_4	Dromedary	Mauritania	
62	MR_5	Mauretania_5	Dromedary	Mauritania	
63	NG852	Drom852	Dromedary	Nigeria	
64	NG855	Drom855	Dromedary	Nigeria	X
65	NG865	Drom865	Dromedary	Nigeria	X
66	NG873	Drom873	Dromedary	Nigeria	X
67	NG877	Drom877	Dromedary	Nigeria	
68	NG884	Drom884	Dromedary	Nigeria	X
69	NG887	Drom887	Dromedary	Nigeria	
70	PK475	Drom475	Dromedary	Pakistan	
71	PK476	Drom476	Dromedary	Pakistan	
72	PK817A	Drom817A	Dromedary	Pakistan	X
73	PK818A	Drom818A	Dromedary	Pakistan	X

74	PK819A	Drom819A	Dromedary	Pakistan	X
75	QA440	Drom440	Dromedary	Qatar	X
76	QA441	Drom441	Dromedary	Qatar	X
77	SA471	Drom471	Dromedary	Saudi Arabia	X
78	SA472	Drom472	Dromedary	Saudi Arabia	
79	SA473	Drom473	Dromedary	Saudi Arabia	X
80	SA795A	Drom795A	Dromedary	Saudi Arabia, Magaheem	X
81	SA796A	Drom796A	Dromedary	Saudi Arabia, Homor	X
82	SA797A	Drom797A	Dromedary	Saudi Arabia, Wadda	X
83	SA798A	Drom798A	Dromedary	Saudi Arabia, Magaheem	X
84	SA799A	Drom799A	Dromedary	Saudi Arabia, Homor	X
85	SA800A	Drom800A	Dromedary	Saudi Arabia, Wadda	X
86	SA811	Drom811	Dromedary	Saudi Arabia, Al Baha	X
87	SA813	Drom813	Dromedary	Saudi Arabia, Al Baha	X
88	SA814	Drom814	Dromedary	Saudi Arabia, Al Baha	X
89	SA815	Drom815	Dromedary	Saudi Arabia, Al Baha	X
90	SA816	Drom816	Dromedary	Saudi Arabia, Al Baha	X
91	SA817	Drom817	Dromedary	Saudi Arabia, Al Baha	X
92	SA818	Drom818	Dromedary	Saudi Arabia, Guewla	
93	SA819	Drom819	Dromedary	Saudi Arabia	X
94	SD_1	Sudan_1	Dromedary	Sudan	X
95	SD_3	Sudan_3	Dromedary	Sudan	
96	SD754	Drom754	Dromedary	Sudan	X
97	SD756	Drom756	Dromedary	Sudan	X
98	SD757	Drom757	Dromedary	Sudan	X
99	SD759	Drom759	Dromedary	Sudan	X
100	SD763	Drom763	Dromedary	Sudan	X
101	SD766	Drom766	Dromedary	Sudan	X
102	SD814A	Drom814A	Dromedary	Sudan	
103	SD815A	Drom815A	Dromedary	Sudan	X
104	SD816A	Drom816A	Dromedary	Sudan	X
105	SY216	Drom216	Dromedary	Syria, Palmyra city	X
106	SY217	Drom217	Dromedary	Syria, Palmyra city	X
107	SY219	Drom219	Dromedary	Syria, Palmyra city	X
108	SY235	Drom235	Dromedary	Syria, Palmyra city	X
109	SY242	Drom242	Dromedary	Syria, Palmyra city	X
110	SY246	Drom246	Dromedary	Syria	X
111	SY295	Drom295	Dromedary	Syria	X

112	TN_T1	Tunisia_T1	Dromedary	Tunisia	X
113	TN_T6	Tunisia_T6	Dromedary	Tunisia	X
114	TN_T7	Tunisia_T7	Dromedary	Tunisia	X
115	TN_T8	Tunisia_T8	Dromedary	Tunisia	X
116	TN_T9	Tunisia_T9	Dromedary	Tunisia	X
117	UAE477	Drom477	Dromedary	UAE	X
118	UAE485	Drom485	Dromedary	UAE	X
119	UAE802A	Drom802A	Dromedary	UAE, Dubai	X
120	UAE803A	Drom803A	Dromedary	UAE, Dubai	X
121	UAE804A	Drom804A	Dromedary	UAE, Dubai	X
122	WS_2	Smara2	Dromedary	Smara	X
123	WS_3	Smara3	Dromedary	Smara	X

Supplementary Table 2. Number of paired-end ddRAD reads mapping unambiguously to either dromedary or Bactrian camel genome for all 123 dromedary and Bactrian camel samples.

Sample code	Unambiguously aligned reads		Percentage (%) Bactrian camel
	Dromedary	Bactrian camel	
DZ_103	2431408	156388	6.04329
DZ_104	3074644	192170	5.88249
DZ_56	2487196	153456	5.81129
DZ_57	2937092	182002	5.83509
DZ_DrN12	2650858	165892	5.88948
DZ_DrN1	1675008	106590	5.98283
DZ_DrN6	1746040	107172	5.78304
DZ_DrS1	927398	57052	5.79532
DZ_DrS6	2104830	132502	5.92232
DZ_T1	2239408	141188	5.93078
DZ_T4	2910924	179592	5.81107
SY216	3128626	199754	6.00154
SY217	2118784	133748	5.93767
SY219	3501274	231584	6.20393
SY235	2315158	148590	6.03106
SY242	4056102	261770	6.06248
SY246	3865120	248184	6.03369
SY295	1093358	70724	6.07552
JO372	2601388	165898	5.99497
JO381	2997392	198966	6.22477
JO389	2417914	152056	5.91664
JO390	1769980	113210	6.01161
JO391	1758452	111666	5.97107
JO405	492854	31756	6.05326
JO406	777444	48392	5.85976
JO413	1004740	63530	5.947
JO415	2186506	138312	5.94937
JO418	1707556	106930	5.89313
JO419	1115342	70308	5.92991
JO420	729906	45902	5.91667
JO421	3555536	229638	6.06678
JO433	217350	13784	5.96364
JO434	1046734	66884	6.00601
QA440	1997100	127856	6.01688
QA441	1532876	96046	5.89629
SA471	6414270	416238	6.09381
SA472	333646	21664	6.09721
SA473	4699842	303634	6.06846
PK475	22428	1424	5.97015
PK476	283840	19234	6.3463
UAE477	1182522	81588	6.45419

UAE485	2064346	133102	6.05712
IR668	1966310	125440	5.99689
IR674	1672368	106402	5.98177
IR688	1828422	127758	6.53099
IR689	1576960	110128	6.5277
IR708	1872026	131322	6.55513
IR709	1286870	87520	6.36792
SD754	983556	62058	5.93508
SD756	1939982	126850	6.13741
SD757	1138096	71950	5.94605
SD759	1307618	83766	6.02034
SD763	1406304	88658	5.93045
SD766	2962206	190984	6.05685
SA795A	1138096	69834	5.7813
SA796A	771590	50268	6.11639
SA797A	1359976	87116	6.02007
SA798A	1871728	122210	6.12908
SA799A	1466772	99168	6.33281
SA800A	2684034	171250	5.99765
AT801A	1605952	48940	2.95729
UAE802A	2835876	179888	5.96492
UAE803A	1992046	124706	5.89138
UAE804A	3367038	208080	5.82023
KE805A	3179714	205358	6.06658
KE806A	1787446	113864	5.98871
SA811	2943068	186332	5.95424
SA813	2541750	162902	6.02303
SD814A	400646	25376	5.9565
SA814	1676610	108058	6.0548
SD815A	2226780	140486	5.93453
SA815	3685552	229664	5.86593
SD816A	2212342	141248	6.00139
SA816	2194714	138556	5.93828
PK817A	1131588	76722	6.34953
SA817	3971444	257024	6.07842
PK818A	2148248	149354	6.50043
SA818	541346	33742	5.86728
PK819A	6308332	431714	6.40521
SA819	3037066	192094	5.94873
LY846	842626	52704	5.88654
NG852	2215904	135642	5.76821
NG855	2600874	158712	5.7513
NG865	2152412	135898	5.93879
NG873	2243294	137926	5.79224
NG877	2542884	156838	5.80941
NG884	2472972	153378	5.83997
NG887	3207654	199594	5.85792
EG_106	152972	9182	5.66252

EG_12	1629166	101334	5.85576
EG_17	1445716	88116	5.74483
EG_23	1263868	79126	5.89176
EG_49	1481266	92384	5.87068
EG_52	1885774	121414	6.04896
EG_65	1829238	114338	5.88287
EG_66	29060	1834	5.93643
LY_1A	2459846	156788	5.99197
LY_2A	1868654	116318	5.85993
LY_3A	1664384	104756	5.9213
LY_4A	2508432	163476	6.11832
MR_1	69314	4290	5.82849
MR_2	815368	50060	5.78442
MR_4	144278	9000	5.87168
MR_5	36142	2162	5.64432
WS_2	4058050	253044	5.8696
WS_3	2558202	161908	5.95226
SD_1	1103938	70606	6.01135
SD_3	348768	21866	5.89962
TN_T1	3137058	196098	5.88325
TN_T6	2028630	126294	5.86072
TN_T7	3374840	206528	5.76673
TN_T8	2841036	177652	5.88507
TN_T9	2406542	151748	5.93162
IR715*	1661398	171350	9.34935
IR717*	1310878	212914	13.9726
IR719*	1555956	168732	9.78333
KZ888*	779102	79550	9.26452
KZ889*	720310	78934	9.87608
KZ890*	1303594	1966262	60.133
KZ891*	958948	1160252	54.7495
KZ892*	1291388	1034492	44.4774
KZ893*	1421546	752598	34.6158
DC156 (Bactrian camel)	274292	1496774	84.5126

*The nine individuals marked with an asterisk were eliminated from downstream analysis due to potential introgression.

Supplementary Table 3. Identification of genes located within 200kbp upstream and downstream of loci under putative selection with its respective function from loci under putative selection. In bold are depicted the genes related to immunity.

Gene ID	Name	Protein function	Associated diseases
1	<i>AP4M1</i> Adaptor Related Protein Complex 4 Subunit Mu 1	Recognition and sorting of cargo proteins with tyrosine-based motifs from the trans-golgi network to the endosomal-lysosomal system	Nuclear Senile Cataract and Phototoxic Dermatitis
2	<i>APOBEC2</i> Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 2	C to U editing enzyme and zink ion binding. Role in embryonic patterning, cell lineage gene	Rift Valley Fever and Hermansky-Pudlak Syndrome 5
3	<i>ARID3C</i> AT-Rich Interaction Domain 3C	regulation, cell cycle control, transcriptional regulation	<i>No info</i>
4	<i>C11ORF53</i> Chromosome 11 Open Reading Frame 53	<i>No info</i>	<i>No info</i>
5	<i>C7ORF43</i> Chromosome 7 Open Reading Frame 43	<i>No info</i>	<i>No info</i>
6	<i>CARNMT1</i> Carnosine N-Methyltransferase 1	Methyltransferase that converts carnosine to anserine	<i>No info</i>
7	<i>CCL21C</i> C-C Motif chemokine 21	Involved in immunoregulatory and inflammatory processes (shows preferential activity towards naive T-cells)	<i>No info</i>
8	<i>CCL27</i> C-C Motif Chemokine Ligand 27	Involved in immunoregulatory and inflammatory processes (chemotactic for skin-associated memory T lymphocytes)	<i>No info</i>
9	<i>CNPY4</i> Canopy FGF Signaling Regulator 4	Plays a role in the regulation of the cell surface expression of TLR4.	<i>No info</i>
10	<i>CNTFR</i> Ciliary Neurotrophic Factor Receptor	Ligand-specific component of a tripartite receptor, which plays a critical role in neuronal cell survival, differentiation and gene expression	Cervix Endometriosis and Melanoacanthoma
11	<i>COLCA2</i> Colorectal Cancer Associated 2	<i>No info</i>	Spastic Paraplegia 50 and Autosomal Recessive
12	<i>COPS6</i> COP9 signalosome complex subunit 6	Involved in various cellular and developmental processes	<i>No info</i>
13	<i>DCTN3</i> Dynactin subunit 3	Together with dynein may be involved in spindle assembly and cytokinesis.	Gastric cancer
14	<i>DNAI1</i> Dynein Axonemal Intermediate Chain 1	Part of the dynein complex in respiratory cilia	<i>No info</i>
15	<i>FOXO3</i> Forkhead Box O3	Trigger apoptosis through gene expression	Chromosome 6Q deletion and Rhabdomyosarcoma
16	<i>GALT</i> Galactose-1-Phosphate Uridyltransferase	Plays an important role in galactose metabolism	<i>No info</i>
17	<i>GPC2</i> Glypican-2	Cell surface proteoglycan that bears heparan sulfate	<i>No info</i>

18	<i>GTF2H1</i>	General Transcription Factor IIH Subunit 1	Component of the general transcription and DNA repair factor IIH core complex	Sudden Infant Death Syndrome (SIDS)
19	<i>HPS5</i>	Hermansky-Pudlak syndrome 5 protein isoform 1	Component of the biogenesis of lysosome-related organelles complex-2	Gray platelet syndrome
20	<i>IFT43</i>	Intraflagellar transport protein 43	Involved in retrograde ciliary transport along microtubules from the ciliary tip to the base	Polycystic Lipomembranous Osteodysplasia With Sclerosing Leukoencephalopathy
21	<i>IL11RA</i>	Interleukin 11 Receptor Subunit Alpha	Essential for the normal development of craniofacial bones and teeth (Innate Immune System related pathway)	<i>No info</i>
22	<i>LAMTOR 4</i>	Late Endosomal/Lysosomal Adaptor, MAPK And MTOR Activator 4	Involved in amino acid sensing and activation of mTORC1	<i>No info</i>
23	<i>LDHA</i>	L-lactate dehydrogenase A	Catalyzes the conversion of L-lactate and NAD to pyruvate and NADH in the final step of anaerobic glycolysis	Colorectal Cancer
24	<i>LDHC</i>	L-Lactate Dehydrogenase C	Testis-specific gene with possible role in sperm mobility	<i>No info</i>
25	<i>MBLAC1</i>	Metallo-Beta-Lactamase Domain Containing 1	Hydrolase activity	<i>No info</i>
26	<i>MCM7</i>	Minichromosome Maintenance Complex Component 7	Mini-chromosome maintenance proteins essential for the initiation of eukaryotic genome replication	<i>No info</i>
27	<i>MEPCE</i>	7SK snRNA methylphosphate capping enzyme	S-adenosyl-L-methionine-dependent methyltransferase	Loeys-Dietz Syndrome 5 and Arrhythmogenic Right Ventricular Dysplasia
28	<i>NCR2</i>	Natural cytotoxicity triggering receptor 2	Cytotoxicity-activating receptor that may contribute to the increased efficiency of activated natural killer cells to mediate tumor cell lysis	<i>No info</i>
29	<i>NFYA</i>	Nuclear Transcription Factor Y Subunit Alpha	Component of the sequence-specific heterotrimeric transcription factor	<i>No info</i>
30	<i>NMRK1</i>	Nicotinamide Riboside Kinase 1	Catalyzes the phosphorylation of nicotinamide riboside and nicotinic acid riboside	Melanoma and dermatitis
31	<i>OARD1</i>	O-Acyl-ADP-Ribose Deacylase 1	Related to deacetylase activity and purine nucleoside binding	Craniosynostosis And Dental Anomalies
32	<i>OSTF1</i>	Osteoclast Stimulating Factor 1	Indirectly induces osteoclast formation and bone resorption (Innate Immune System related pathway)	Galactosemia and Galactokinase Deficiency
33	<i>POU2F1</i>	POU domain class 2-associating factor 1	Contains the POU domain, a 160-amino acid region necessary for DNA binding to a specific octameric sequence	Amyotrophic Lateral Sclerosis 16 and Spinal Muscular Atrophy
34	<i>PVRIG</i>	Transmembrane protein PVRIG	Negative regulation of T cell receptor signaling pathway	<i>No info</i>

35	<i>RASEF</i>	RAS And EF-Hand Domain Containing	Regulation of membrane traffic and with a potential role as tumor suppressor	<i>No info</i>
36	<i>RORB</i>	Nuclear receptor ROR-beta isoform 1	DNA binding, nuclear receptor, steroid hormone receptor (help regulate the expression of some genes involved in circadian rhythm)	Cold-Induced Sweating Syndrome and Attention Deficit-Hyperactivity Disorder
37	<i>RPP25L</i>	Ribonuclease P/MRP Subunit P25 Like	Small archaeobacterial proteins that may be a component of ribonuclease P or MRP	Ciliary Dyskinesia, Primary, 1 and Kartagener Syndrome
38	<i>SAALI</i>	Serum Amyloid A Like 1	Promotes the proliferation of synovial fibroblasts in response to proinflammatory stimuli	Fanconi-Bickel Syndrome and Myoglobinuria
39	<i>SERGEF</i>	Secretion Regulating Guanine Nucleotide Exchange Factor	Probable guanine nucleotide exchange factor, which may be involved in the secretion process	Premature Ovarian Failure 8 and Williams-Beuren Syndrome
40	<i>SESNI</i>	Sestrin 1	Induced by the p53 tumor suppressor protein, play a role in the cellular response to DNA damage and oxidative stress	<i>No info</i>
41	<i>SIGMAR1</i>	Sigma Non-Opioid Intracellular Receptor 1	Functions of various tissues associated with the endocrine, immune, and nervous systems	Posterolateral Myocardial Infarction and Skin Carcinoma In Situ
42	<i>STAG3</i>	Stromal Antigen 3	Subunit of the cohesin complex which regulates the cohesion of sister chromatids during cell division	Cranioectodermal Dysplasia 3 and Retinitis Pigmentosa 81
43	<i>TGFB3</i>	Transforming Growth Factor Beta 3	Involved in embryogenesis and cell differentiation	Herpes Simplex and Primary Amebic Meningoencephalitis
44	<i>TPHI</i>	Tryptophan Hydroxylase 1	Catalyzes the first and rate limiting step in the serotonin biosynthesis	Childhood-Onset Schizophrenia and Personality Disorder
45	<i>TREM2</i>	Triggering Receptor Expressed On Myeloid Cells 2	Functions in immune response and may be involved in chronic inflammation	<i>No info</i>
46	<i>TREML1</i>	Triggering Receptor Expressed On Myeloid Cells Like 1	Cell surface receptor that may play a role in the innate and adaptive immune response	<i>No info</i>
47	<i>TREML2</i>	Triggering Receptor Expressed On Myeloid Cells Like 2	Cell surface receptor that may play a role in the innate and adaptive immune response.	<i>No info</i>
48	<i>TRPM6</i>	Transient Receptor Potential Cation Channel Subfamily M Member 6	Epithelial magnesium transport, and magnesium absorption in the gut and kidney	Hypomagnesemia
49	<i>TSPO2</i>	Translocator Protein 2	Binds cholesterol and mediates its redistribution during erythropoiesis	<i>No info</i>
50	<i>UNC5CL</i>	Unc-5 Family C-Terminal Like	Inhibits NF-kappa-B-dependent transcription by impairing NF-kappa-B binding to its targets	<i>No info</i>
51	<i>ZCWPW1</i>	Zinc Finger, CW-Type With PWWP Domain 1		<i>No info</i>
52	<i>ZKSCAN1</i>	Zinc Finger With KRAB And SCAN Domains 1	Transcription factor that regulates the expression of GABA type-A receptors in the brain	<i>No info</i>
53	<i>ZNF3</i>	Zinc Finger Protein 3	Involved in cell differentiation and/or proliferation.	Xeroderma Pigmentosum, Complementation Group E

6.2 Article 2

Lado, S., Elbers, J. P., Rogers, M. F., Melo-Ferreira, J., Yadamsuren, A., Corander, J., Petr Horin & Burger, P. A. (2020). Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies. *BMC Genomics*, 21(606), 1-17. <https://doi.org/10.1186/s12864-020-06990-4>

RESEARCH ARTICLE

Open Access



Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies

Sara Lado¹, Jean P. Elbers^{1*}, Mark F. Rogers², José Melo-Ferreira^{3,4}, Adiya Yadamsuren⁵, Jukka Corander^{6,7,8}, Petr Horin^{9,10} and Pamela A. Burger^{1*} 

Abstract

Background: Immune-response (IR) genes have an important role in the defense against highly variable pathogens, and therefore, diversity in these genomic regions is essential for species' survival and adaptation. Although current genome assemblies from Old World camelids are very useful for investigating genome-wide diversity, demography and population structure, they have inconsistencies and gaps that limit analyses at local genomic scales. Improved and more accurate genome assemblies and annotations are needed to study complex genomic regions like adaptive and innate IR genes.

Results: In this work, we improved the genome assemblies of the three Old World camel species – domestic dromedary and Bactrian camel, and the two-humped wild camel – via different computational methods. The newly annotated dromedary genome assembly CamDro3 served as reference to scaffold the NCBI RefSeq genomes of domestic Bactrian and wild camels. These upgraded assemblies were then used to assess nucleotide diversity of IR genes within and between species, and to compare the diversity found in immune genes and the rest of the genes in the genome. We detected differences in the nucleotide diversity among the three Old World camelid species and between IR gene groups, i.e., innate versus adaptive. Among the three species, domestic Bactrian camels showed the highest mean nucleotide diversity. Among the functionally different IR gene groups, the highest mean nucleotide diversity was observed in the major histocompatibility complex.

(Continued on next page)

* Correspondence: jean.elbers@gmail.com; pamela.burger@vetmeduni.ac.at

¹Department of Interdisciplinary Life Sciences, Research Institute of Wildlife Ecology, Vetmeduni Vienna, Vienna, Austria

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



© The Author(s). 2020 **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.

(Continued from previous page)

Conclusions: The new camel genome assemblies were greatly improved in terms of contiguity and increased size with fewer scaffolds, which is of general value for the scientific community. This allowed us to perform in-depth studies on genetic diversity in immunity-related regions of the genome. Our results suggest that differences of diversity across classes of genes appear compatible with a combined role of population history and differential exposures to pathogens, and consequent different selective pressures.

Keywords: Chromosome mapping, Chromosome conformation capture, Dromedary, Genome assembly, Scaffolding, Genome annotation, Immune response genes, Genetic diversity

Background

Accurate genome assemblies provide an invaluable basis to assess genetic variation throughout the genome of species, to detect structural variants and to decipher complex genomic regions such as immune-response (IR) genes. Maintaining high genetic diversity in a population is important to reduce the spread of diseases, allowing rapid adequate immune responses and limiting, e.g., parasite evolution (see [1]). Even though demographic changes in general may cause important loss of genetic diversity, and particularly during domestication, due to intensive selection and potential inbreeding in many genomic regions [2], in other regions such as IR genes the genetic diversity can be conserved due to selective pressures of pathogens [3].

Old World camels (Artiodactyla, Tylopoda, Camelidae, Camelini) – the domesticated one-humped dromedaries (*Camelus dromedarius*) and two-humped Bactrian camels (*Camelus bactrianus*), as well as the critically endangered two-humped wild camels (*Camelus ferus*) – are valuable species not only for their production traits (e.g., meat, milk or wool), but for their power (e.g., riding or packing). Moreover, they are ungulate species with unique adaptations to diverse and extreme environments. Consequently, as they are in contact with different pathogenic pressures on different environments, there is great interest in understanding the general diversity in the part of the genome encoding their immune system. Previous research on immunogenome diversity in Old World camels focused mainly on the MHC genes (e.g., [4]), as due to its critical importance for individual survival, the MHC complex is the most intensively studied part of the vertebrate immunogenome [5]. MHC genes, however, account only for part of the genetic variability underlying resistance to infectious pathogens [6, 7]. A broader approach is required to capture the overall genetic diversity of the immune system and to understand its role in response to pathogens. On these grounds, high-quality genome assemblies are needed. Previous studies [8–12] developed high quality genome assemblies for the three Old World camel species. Although very useful for broad inferences of genome-wide diversity or demographic histories, an improved version

of these assemblies is needed to allow more detailed studies of the diversity in parts of the genome, such as IR genes. Access to different computational methods allows overcoming previous genome assemblies' limitations.

In this work, we describe our computational efforts to generate improved Old World camelid genome assemblies, and we present versions CamDro3, CamBac2 and CamFer2, for dromedaries, Bactrian camels and wild camels, respectively. Our goal was not only to provide novel assemblies for genomic analysis in camels, but also to take advantage of the upgraded genome assemblies to assess the genetic diversity in different groups of immune genes, and compare them among species and to the rest of the intra-genic genomic diversity.

Results

Improved *Camelus dromedarius* genome assembly

Despite the utility of the CamDro1 and CamDro2 assemblies, inconsistencies and gaps can limit analyses at various genomic scales. By using different bioinformatic methods, we were able to upgrade the available genome assemblies to CamDro3, which is more accurate, contiguous and show fewer scaffolds of increased size when compared to the previous ones. CamDro3 consistently had higher RNA-Seq read mapping rates than CamDro2, and these two assemblies had much higher mapping rates than the other assemblies (Supplemental Fig. 1). After CamDro3 and CamDro2, the assembly with the third highest mapping rates varied depending on the tissue and season analyzed, but *B. taurus* consistently had the lowest mapping rates. We were able to assign at least one super-scaffold to each of the 37 chromosomes except the Y chromosome as the dromedary used in CamDro1, CamDro2, and CamDro3 was female. Chromosomes are denoted by numbers 1–36 and X in the CamDro3 assembly. There were 113,944,958 bases in scaffolds not assigned to chromosomes (5.25% of the 2,169,346,739 base assembly).

In the CamDro3 annotation, we predicted 22,917 genes that produced 34,135 proteins, and 7.4% (1705) of genes had no assigned annotation. These numbers are slightly higher than for the CamDro2 assembly for which

we had predicted 22,534 genes that produced 34,024 proteins, and 7.7% (1730) of genes had no assigned annotation [11]. We assessed if predicted proteins were truncated due to uncorrected indels introduced by PacBio reads by comparing the predicted protein length hit distribution of the CamDro1 assembly (Illumina only data, Fig. 1, red line), which should lack such PacBio specific error, to that of the CamDro2 (Fig. 1, green line) and CamDro3 assemblies (Fig. 1, blue line). First, predicted proteins from the CamDro1 assembly had 21,257 protein hits against the UniProt/TrEMBL database, and 11,671 (55%) hits were between 0.85 and 1.15 (query sequence length/ subject sequence length; Fig. 1). Second, predicted proteins from the CamDro2 assembly had 32,297 protein hits, and 17,341 (54%) were between 0.85 and 1.15 (Fig. 1). Third, predicted proteins for CamDro3 assembly had 32,427 protein hits, and 17,006 (52%) were between 0.85 and 1.15 (Fig. 1). This suggests that CamDro3 is similar to CamDro2 with respect to proportion of uncorrected PacBio indels, but the proportions of uncorrected indels are very low when compared to CamDro1. AEDs were slightly higher in CamDro3 versus CamDro2 (Fig. 2). For example, CamDro2 had AED values ≤ 0.5 for 78.4% transcripts versus 79.1% transcripts

for CamDro3. Lower AED values indicate a better fit to the provided evidence when annotating a genome [15].

We predicted 22,223 genes that produced 33,153 proteins in CamDro3 using a more up to date set of proteins during annotation. These values were lower than when annotating CamDro3 using the same cDNA transcripts and proteins used for annotating CamDro2 possibly because there were fewer false genes predicted in the up-to-date annotation of CamDro3. Further, 8.46% (1879) genes produced proteins did not match UniProt/Swiss-Prot proteins. This value was higher than before, but we used UniProt/Swiss-Prot instead of the more comprehensive UniProt-TrEMBL protein database. The CamDro3 assembly and these annotations have been submitted to GenBank (GCA_000803125.3) and Dryad - see Data Accessibility Statement.

Improved *Camelus bactrianus* and *Camelus ferus* genomes via reference-guided assembly

CamBac2 increased in size by 46,927,041 bases and had 1862 fewer scaffolds than CamBac1, and CamBac2's N50 was nearly 8 times larger (Table 1). The longest contig in CamBac2 was more than 7 times larger than before. We have also predicted 19,491 genes that produced 25,

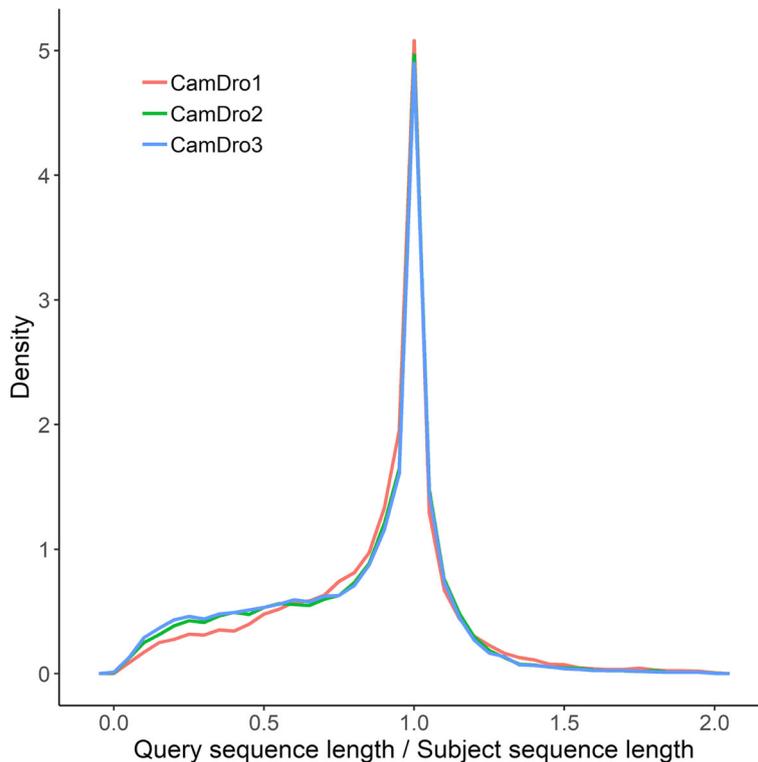


Fig. 1 Frequency polygons of query sequence length (predicted proteins) divided by subject (UniProt/TrEMBL) sequence length for DIAMOND [13] mapped MAKER [14] predicted proteins against UniProt/TrEMBL release 2018_07 database for: (red line) the original North African dromedary genome (CamDro1), ([8]; GenBank accession: GCA_000803125.1); (green line) the North African dromedary genome after adding $\sim 11\times$ PacBio sequencing reads (CamDro2); and (blue line) CamDro3

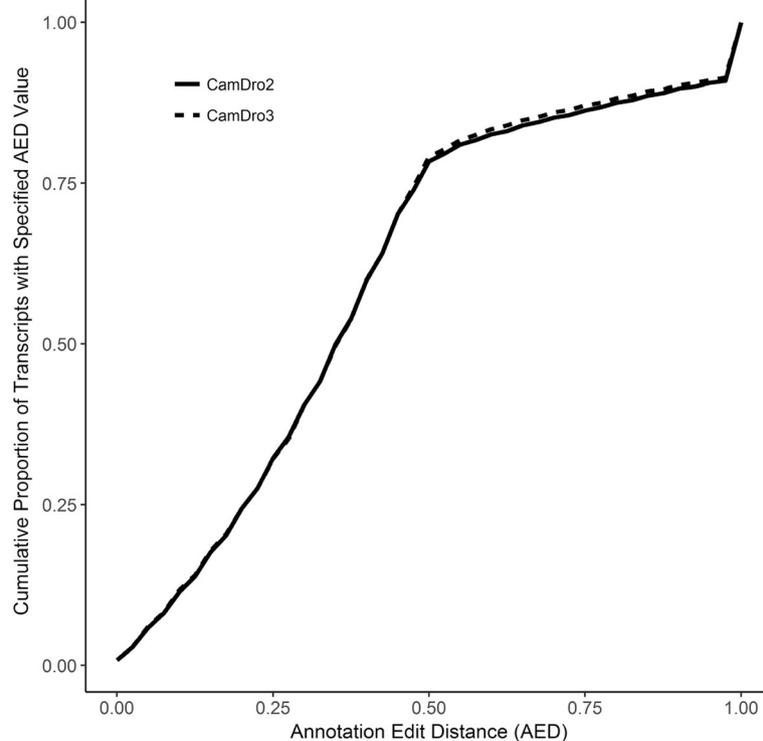


Fig. 2 Cumulative proportion of transcripts with specific or lower annotation edit distance (AED) for CamDro2 (solid line) and CamDro3 (dashed line). CamDro2 had AED ≤ 0.50 for 78.4% transcripts, whilst MAKER run 2 had 79.1% transcripts with AED ≤ 0.50 . Note that having a larger proportion of lower AED values indicates a genome annotation that is more congruent with the evidence used during the annotation process

95 proteins in CamBac2. Of these genes, 4.03% (786) did not match proteins from UniProt/Swiss-Prot. *Camelus bactrianus* had the second lowest mapping rates, after *B. taurus*. The CamBac2 assembly and these annotations have been submitted to Dryad - see Data Accessibility Statement.

CamFer2 was 77,064,279 bases larger and was organized into 4176 fewer scaffolds than CamFer1. CamFer2 had an N50 that was nearly 35 times larger than CamFer1's N50 (Table 1). CamFer2's longest contig was more than 2 times larger than CamFer1's largest contig.

We predicted 19,192 genes that produced 19,192 proteins in CamFer2. Of these genes, 3.69% (708) did not match proteins from UniProt/Swiss-Prot. There were many structural variations (inversions and repeats) when comparing the assembled chromosomes of CamFer2 and the *C. ferus* genome assembly from Ming et al., [12] (Supplemental Fig. 2). Ultimately, these latter genomes have similar scaffold N/L50 values, but CamFer2 has much smaller contig N/L50 values because of more abundant and larger gaps in assembled chromosomes (Supplemental Table 1). The CamFer2 assembly and these annotations have been submitted to Dryad - see Data Accessibility Statement.

Intra-specific genome-wide diversity

Mean coverage throughout the genomes of the three Old World camel species was not different among species ($F_{2,22} = 0.1871$, $P = 0.8307$; Table 2). The mean total number of SNPs was different among species ($F_{2,22} = 64.943$, $P < 0.0001$) as was the number of synonymous ($F_{2,22} = 66.99$, $P < 0.0001$) and non-synonymous SNPs ($F_{2,22} = 113.25$, $P < 0.0001$; Table 2). Mean total, synonymous, and non-synonymous SNPs were highest in Bactrian camels, followed by wild camels, then dromedaries. The mean number of insertions was different among species ($F_{2,22} = 31.269$, $P < 0.0001$) as was the mean number of deletions ($F_{2,22} = 16.407$, $P < 0.0001$; Table 2). Bactrian camels had a higher mean number of insertions than dromedaries and wild camels, which showed similar numbers of insertions. Bactrian camels had higher mean number of deletions, followed by wild camels, then dromedaries.

Heterozygosity rates in exons and introns

We assessed the heterozygosity rates in coding (exons) and noncoding (introns) regions, across multiple individuals. Heterozygosity means for all three species and coding/noncoding regions were all significantly different at the 0.05 level of significance. The results show that

Table 1 Assembly statistics for the CamBac1 (GCF_000767855.1) and CamFer1 (GCF_000311805.1) and after improvement (CamBac2 and CamFer2, respectively) with reference-guided assembly with Ragout [16] using Progressive Cactus [17] alignments to CamDro3 then filling in gaps with GapFiller [18]

Assembly	CamBac1	CamBac2	CamFer1	CamFer2
Total size	1,992,663,268	2,039,590,309	2,009,194,609	2,086,258,888
Gap length	13,666,687	57,965,943	23,778,176	99,159,843
Scaffolds				
Number	35,455	33,593	13,334	9158
Longest	46,538,883	122,729,119	15,735,958	123,639,755
N90 ^a	1,821,536	24,994,512	341,469	25,431,863
L90 ^b	255	29	1167	30
N50 ^a	8,812,066	68,446,253	2,005,940	69,671,486
L50 ^b	68	11	274	11
Contigs^c				
Number	67,435	56,044	68,872	66,352
Longest	1,143,031	2,938,098	853,441	1,096,594
N90	29,656	43,365	16,267	16,886
L90	15,603	10,214	25,475	23,951
N50	139,019	219,031	90,263	97,198
L50	3963	2415	5814	5272
Single-copy BUSCOs ^d	3827	3835	3796	3816
Duplicated BUSCOs	22	18	48	32
Fragmented BUSCOs	164	157	175	168
Missing BUSCOs	91	94	85	88

^aN90/N50 are the scaffold or contig lengths such that the sum of the lengths of all scaffolds or contigs of this size or larger is equal to 90/50% of the total assembly length

^bL90/L50 are the smallest number of scaffolds or contigs that make up at least 90/50% of the total assembly length

^cUsing minimum gap length of 10 bp

^dBUSCOs: Benchmarking Universal Single-Copy Orthologs [19] are mammalian BUSCOs from OrthoDB v. 9.1 genes [20]

exons have significantly lower mean heterozygosity compared to introns in all three species, and that the domestic camel had the highest heterozygosity, followed by the dromedary and lastly the wild camel (DC: exons = 0.00110; introns = 0.00316; Drom: exons = 0.000983; introns = 0.00217; WC: exons = 0.000941; introns = 0.00231). These results are in accordance with what was found in Fitak et al. (2020) [21], although in Jirimutu et al. (2012) [9] the domestic camel genome had lower heterozygosity in the exonic regions compared to wild camel genome (though in the latter study the authors based their estimates on single individuals).

Nucleotide diversity among Old World camels in immune response and intra-genic regions

After improving the three Old World camel genome assemblies, we assessed the nucleotide diversity in immune response and intra-genic (within gene) regions. When looking at non-synonymous and synonymous SNPs and indels altogether, mean nucleotide diversity was found not to differ significantly for adaptive, innate IR genes and the rest-of-genome genes, but to be higher in MHC class I and II genes in both dromedaries and domestic Bactrian camels (Fig. 3a; Table 3 for mean values and 95% bootstrap confidence limits). On the other hand, in wild camels, mean nucleotide diversity was not significantly different across gene types. When comparing nucleotide diversity per gene class in species pairs, mean MHC nucleotide diversity did not differ significantly for domestic Bactrian camels and dromedaries, as well as for wild camels and dromedaries, but differed between wild and domestic Bactrian camels, with the latter showing higher nucleotide mean diversity (Supplemental Fig. 3a; Table 3 for mean values and 95% bootstrap confidence limits). Innate and adaptive IR gene nucleotide diversity was statistically different between domestic Bactrian camels and the other two species, but the same between dromedaries and wild camels, while again Bactrian camels had a higher mean nucleotide diversity. Rest-of-genome gene nucleotide diversity was also higher for the Bactrian camel and different between this and the other two camel species.

On the other hand, when looking at only non-synonymous SNPs, dromedaries' mean nucleotide diversity patterns were more difficult to interpret. Mean innate gene nucleotide diversity was lower than mean rest-of-genome gene nucleotide diversity, but mean innate gene nucleotide diversity was statistically not different from mean adaptive or MHC nucleotide diversity nor was mean rest-of-genome nucleotide diversity different from mean adaptive or MHC nucleotide diversity (Fig. 3b; Table 3 for mean values and 95% bootstrap confidence limits). In domestic Bactrian camels, mean nucleotide diversity was the same for adaptive, innate and the rest-of-genome genes, but different in MHC genes where it was the highest. On the other hand, in wild camels, all gene groups had statistically the same mean nucleotide diversity. For both MHC and adaptive IR genes, mean nucleotide diversity was the same among the three camel species (Supplemental Fig. 3b). For innate IR genes, Bactrian and wild camels had the same mean nucleotide diversities, whereas dromedaries had a different mean nucleotide diversity from the other camel species, but the same compared to wild camels. Finally, for the rest-of-genome genes group, all species had statistically different mean nucleotide diversities, where

Table 2 Mean coverage and number of different types of variants per sample. DC for domestic Bactrian camel (*Camelus bactrianus*), Drom for dromedary (*Camelus dromedarius*), and WC for wild camel (*Camelus ferus*). SD for standard deviation

Sample	Mean Coverage	Total_SNPs	Synonymous SNPs	Non-synonymous SNPs	Insertions	Deletions
DC158	41.42	3,713,662	16,761	18,352	258,367	237,987
DC269	14.25	3,238,412	14,206	15,473	230,164	205,242
DC399	13.80	3,199,637	14,370	16,112	226,223	199,701
DC400	14.54	3,213,008	14,130	15,608	226,945	200,953
DC402	14.84	3,130,745	13,756	15,296	218,205	193,720
DC408	15.11	3,328,223	14,592	16,693	234,064	209,759
DC423	14.46	3,738,504	17,182	17,866	250,856	227,449
Drom439	14.30	1,929,784	8528	9135	163,100	147,765
Drom795	11.78	1,907,261	8600	9679	186,969	158,190
Drom796	14.23	1,991,649	8476	9193	170,719	156,795
Drom797	13.76	1,992,724	8945	9576	178,917	160,938
Drom800	40.73	1,500,998	6844	7255	140,148	122,312
Drom802	14.59	2,006,825	9311	10,122	188,392	166,360
Drom806	9.52	1,854,989	7944	8692	164,993	149,508
Drom816	10.33	1,929,982	8476	9263	173,380	154,757
Drom820	9.66	1,881,945	7694	8162	167,680	152,220
WC214	14.43	2,517,749	9919	10,071	157,630	162,297
WC216	12.86	2,654,274	11,040	10,871	170,009	176,405
WC218	14.22	1,825,617	7396	8026	109,795	107,655
WC219	14.04	2,707,996	11,187	11,038	173,685	179,297
WC220	14.92	2,707,716	11,067	10,982	170,579	179,365
WC247	14.06	2,956,856	11,567	11,235	189,010	196,986
WC303	41.54	2,937,692	11,625	11,313	189,408	204,838
WC304	14.67	2,748,380	11,047	10,844	180,435	186,048
WC305	14.05	2,704,263	10,599	10,520	176,820	181,412
Drom mean	15.43	1,888,462	8313	9009	170,478	152,094
Drom SD	9.7	154,355	729	867	14,512	12,552
DC mean	18.35	3,366,027	15,000	16,486	234,975	210,687
DC SD	10.2	252,904	1376	1210	14,409	16,125
WC mean	17.20	2,640,060	10,605	10,544	168,597	174,923
WC SD	9.1	334,004	1307	1017	24,154	28,002

domestic Bactrian camels showed to have the highest values.

There were 46 identified single-domain heavy-chain immunoglobulin genes in the *Camelus ferus* assembly of Ming et al. [12]. Of those 46, annotations for 43 could be lifted over to CamDro3, 36 to CamBac2, and 39 to CamFer2, which mapped on chromosome 6 and on other scaffolds. Mean nucleotide diversity was not significantly different among dromedaries, domestic camels, or wild camels when using either alignments made with all SNPs and indels or only non-synonymous SNPs (see Supplemental Table 2 and Supplemental Fig. 4).

Discussion

Despite its functional importance, the immunogenome of camels has received only limited attention, with work focusing on cytogenetic mapping in alpaca [22], the characteristics of single-domain heavy-chain antibodies [23] or specific mechanisms underlying the genetic diversity of T-cell receptors [24–26]. Dromedary and two-humped camels are important livestock species, well adapted to harsh conditions and resistant to devastating infections that threaten other livestock species in the same areas, like contagious pleuro-pneumonia [27] or foot-and-mouth disease in dromedaries [28]. Other infections have an important role in human health, such

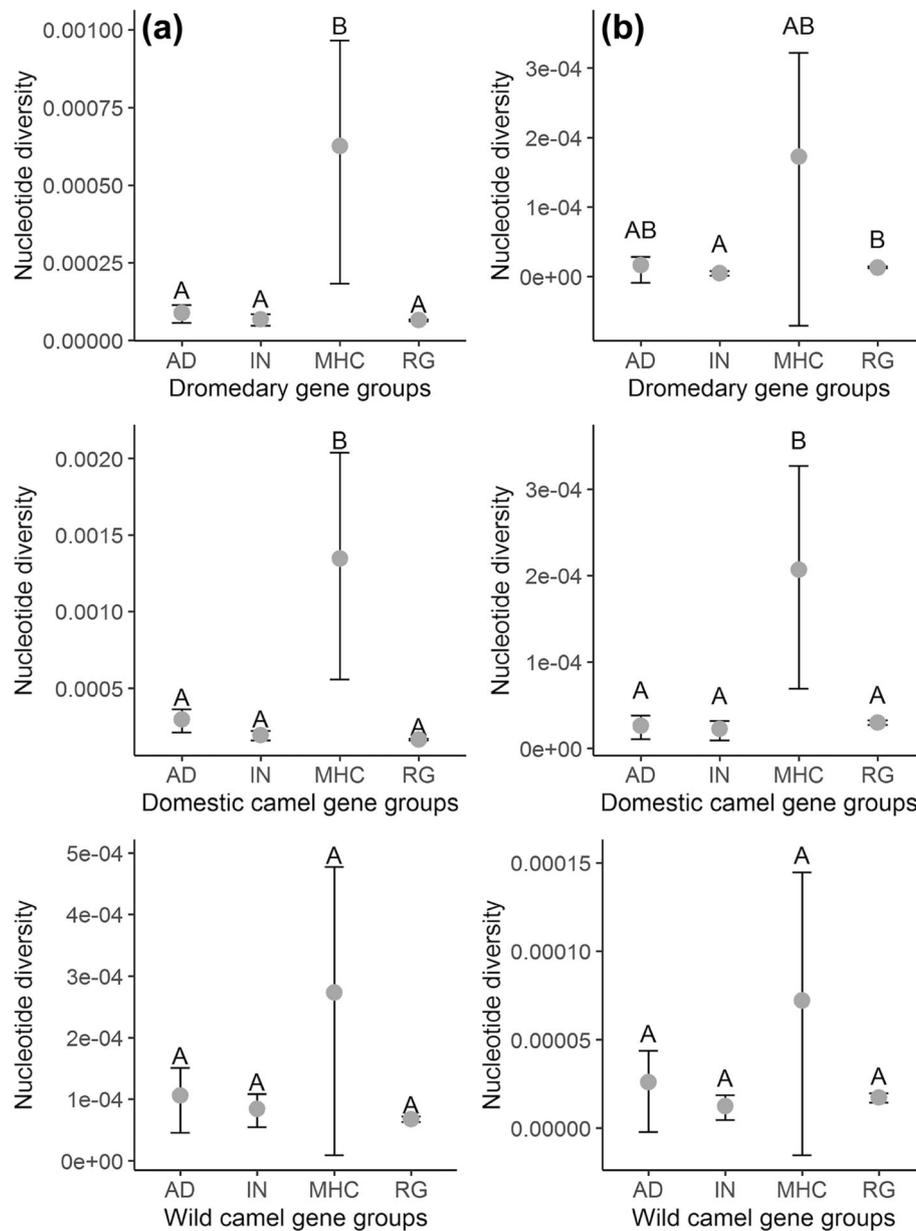


Fig. 3 Means with 95% bootstrap confidence intervals (see Methods) of nucleotide diversity for alignments made with non-synonymous and synonymous SNPs and indels **(a)** and only non-synonymous SNPs **(b)** for: dromedary (*C. dromedarius*; top panel), domestic Bactrian camel (*C. bactrianus*; middle panel), and wild camel (*C. ferus*; bottom panel) gene groups. AD for adaptive genes, IN for innate genes, MHC for MHC class I and II genes, and RG for rest-of-genome genes. Rest-of-genome genes are those not classified as adaptive or innate genes (see Methods). Uppercase letters above upper 95% confidence limits indicate groups have different (non-matching letters) or not different (matching letters) means based on non-overlapping confidence intervals

as the Middle East Respiratory Syndrome Coronavirus (MERS-CoV), for which dromedaries are potential reservoirs [29]. Variation in genetic diversity between innate and adaptive immunity genes is caused by differences in these gene groups' mechanisms. While innate immunity is less specific and more executive, its genes are subject to purifying rather than to positive/balancing selection,

whereas adaptive immunity is more focused on specific recognition of highly diverse antigens and its variability is exposed to different selective pressures [30, 31]. In this study, we compared the diversity in different groups of immune response genes with those found in intra-genic regions among the three Old World camel species, aiming to better understand to which selection pressures

Table 3 Means with 95% bootstrap confidence limits (CL, see Methods) of nucleotide diversity for alignments made with non-synonymous and synonymous SNPs and indels and only non-synonymous SNPs for: DROM (dromedary; *Camelus dromedarius*), DC (domestic Bactrian camel; *Camelus bactrianus*), and WC (wild camel; *Camelus ferus*) gene groups. AD for adaptive genes, IN for innate genes, MHC for MHC class I and II genes, and RG for rest of genome genes. Rest-of-genome-genes correspond to those genes which are not classified as adaptive or innate IR genes (see Methods)

Variant type	Species	Gene groups	Mean	95% lower CL	95% upper CL
SNPs and indels	DROM	MHC	6.26E-04	1.83E-04	9.65E-04
SNPs and indels	DROM	AD	8.81E-05	5.70E-05	1.14E-04
SNPs and indels	DROM	IN	6.81E-05	4.74E-05	8.49E-05
SNPs and indels	DROM	RG	6.55E-05	6.22E-05	6.87E-05
SNPs and indels	DC	MHC	1.35E-03	5.58E-04	2.04E-03
SNPs and indels	DC	AD	2.97E-04	2.11E-04	3.64E-04
SNPs and indels	DC	IN	1.94E-04	1.61E-04	2.23E-04
SNPs and indels	DC	RG	1.66E-04	1.60E-04	1.71E-04
SNPs and indels	WC	MHC	2.73E-04	9.06E-06	4.77E-04
SNPs and indels	WC	AD	1.06E-04	4.52E-05	1.51E-04
SNPs and indels	WC	IN	8.36E-05	5.45E-05	1.08E-04
SNPs and indels	WC	RG	6.71E-05	6.24E-05	7.13E-05
Non synonymous SNPs	DROM	MHC	1.72E-04	-7.09E-05	3.22E-04
Non synonymous SNPs	DROM	AD	1.58E-05	-8.83E-06	2.80E-05
Non synonymous SNPs	DROM	IN	4.79E-06	1.29E-06	7.42E-06
Non synonymous SNPs	DROM	RG	1.28E-05	1.13E-05	1.42E-05
Non synonymous SNPs	DC	MHC	2.07E-04	6.94E-05	3.27E-04
Non synonymous SNPs	DC	AD	2.63E-05	1.04E-05	3.80E-05
Non synonymous SNPs	DC	IN	2.26E-05	9.31E-06	3.17E-05
Non synonymous SNPs	DC	RG	2.97E-05	2.70E-05	3.25E-05
Non synonymous SNPs	WC	MHC	7.23E-05	-1.52E-05	1.45E-04
Non synonymous SNPs	WC	AD	2.61E-05	-2.17E-06	4.37E-05
Non synonymous SNPs	WC	IN	1.23E-05	4.52E-06	1.87E-05
Non synonymous SNPs	WC	RG	1.72E-05	1.45E-05	1.99E-05

they might have been exposed. For this purpose, we first improved the three available Old World camelid genome assemblies.

Old World Camelids genome assemblies' improvement

We applied several computational techniques to improve previous assemblies. To begin with, we were able to greatly improve CamDro3 genome assembly from CamDro2. Compared with the previous version, the number of predicted gene proteins in the CamDro3 were lower, possibly because there were fewer false genes predicted. After correcting mis-assemblies by re-scaffolding CamDro2 and by using a different indel-polishing method, CamDro3 is now more complete, with fewer gaps and likely more accurate. Additionally, the reference-guided assembly process significantly improved the quality and contiguity of CamBac2 and CamFer2, as they are now more contiguous, and have fewer and longer scaffolds. By using a closely related genome to improve a draft

assembly, it has a bigger impact on the final assembly, as well as the accuracy and completeness of a reference genome also contribute [32]. Although mean coverage throughout the genome was not different between species, mean total, synonymous, and non-synonymous SNPs, mean number of insertions and deletions were highest in domestic Bactrian camels compared to the other two species. These results might suggest that domestic Bactrian camels generally have higher genetic diversity than dromedaries and wild camels, as they might have experienced less severe demographic changes during domestication than dromedaries [33] and less recent population size reduction than the critically endangered wild camels [34].

Nucleotide diversity in important immune gene groups

Old World camels are known to be resistant to serious infectious diseases that threaten other livestock species inhabiting the same geographical regions, although they

may contract other poorly-studied diseases [35]. On the other hand, diseases of Camelidae are often difficult to deal with, having non-specific signs with a considerable economic impact [36]. Hence, as diversity in immune response gene regions may influence infectious disease susceptibility in populations, a better understanding of IR gene diversity will support camel breeding and sustainable management in countries of the Global South with large camel populations. As our data were not normally distributed and could not be transformed to approximate a normal distribution, we assessed differences in nucleotide diversity within species in different immune complexes of the genome by using a non-parametric bootstrapping method to estimate 95% confidence intervals of mean nucleotide diversity (Fig. 3 and Supplemental Fig. 3).

MHC class I and class II genes are amongst the most polymorphic genes studied in vertebrates [37]. Pathogen-mediated selection is widely held to be the major driving force in maintaining the high diversity at MHC loci [38]. In particular, the MHC diversity in populations is maintained by balancing selection [39]. According to the 95% confidence intervals derived from non-parametric bootstrap tests of mean nucleotide diversities, we observed that MHC (class I and II) genes had higher mean nucleotide diversity compared to all other gene groups, for two-humped camels, in both SNPs-indels and just non-synonymous SNPs analyses, and for dromedaries in SNP-indels analysis but not for only non-synonymous SNP analysis (Fig. 3). Previous research by Plasil et al., [4] showed that MHC nucleotide diversity within the three Old World species was generally low. In this case, the authors looked specifically into the antigen-binding sites and not to the complete genes where, according to our results, additional diversity appears to be present. The functional importance of this variation is currently unknown. However, it is important to acknowledge how particular pathogens affect immune genetic diversity and, vice versa, how genetic variation influences adaptation to emerging zoonosis, habitat fragmentation, and climate change [40]. MHC genes play an important role in the adaptive branch of the immune system and have been used extensively to estimate levels of adaptive genetic variation [41]. While innate immunity is an efficient first protection against many pathogens but rather less specific, adaptive (or acquired) immunity is a highly specific immune response, and its variability is subject to different selective pressures [30, 31]. Overall, mean nucleotide diversity was never different when comparing innate and adaptive IR gene groups in all three species, in both SNPs-indels and non-synonymous SNPs analyses.

When comparing nucleotide diversity among both two-humped camel species, wild camels had lower mean

nucleotide diversity for both SNP and indels and non-synonymous SNP analyses, except for the MHC class I and II genes and for adaptive genes with non-synonymous SNPs (Supplemental Fig. 3). Moreover, in general, the domestic Bactrian camel had higher mean nucleotide diversity compared to the wild camel, except for the mean nucleotide diversity in adaptive genes with non-synonymous SNPs. One possible explanation for these results is that the wild camel suffered strong population declines leading to the current status of “critically endangered” species (by the International Union for Conservation of Nature (IUCN)). Thus, with the number of individuals decreasing, loss of genetic diversity is unfortunately real [42, 43]. Another possible explanation is that domestic Bactrian camels are under higher pathogenic pressure compared to the wild species. For example, Bactrian camels can be raised and herded with other domestic species (e.g., sheep or goat and sometimes cattle) and due to this fact, the animals are in contact with different pathogens that would not be present in the wild camels’ natural habitat [44]. This pathogenic pressure might have selected for higher diversity in domestic Bactrian camels, explaining the higher diversity in the immunogenome as well as in the rest of the genome. Nevertheless, we cannot discard the possibility that the demographic dynamics influenced the mean nucleotide diversity levels compared among species. Patterns of demographic changes across all three species demonstrated widespread population declines during the Pleistocene [21]. Principally in dromedaries, according to Lado et al. [45] and Fitak et al. [21], long-term population bottlenecks were detected, which probably reduced the nucleotide diversity even more in this species. Furthermore, there is the assumption that dromedaries have been domesticated from a relatively small population of wild one-humped camels, which already have been declining in numbers in a limited geographical area at the Southeast coast of the Arabian Peninsula [33]. However, the domestication of Bactrian camels might have occurred over a much larger geographic region, involving (genetically) more distant and diverse wild two-humped camels [12]. Our results suggest that the IR genes follow the same pattern of rest-of-the-genome genes where domestic Bactrian camels are more diverse throughout all classes of genes when compared to the endangered wild camel.

We also assessed the nucleotide diversity of single-domain heavy-chain immunoglobulin genes in our data. For that, we lifted the 46 heavy-chain immunoglobulin gene annotations from the Ming et al. 2020 [12] *Camelus ferus* genome assembly over to CamFer2, CamDro3, and CamBac2. However, we could not detect all 46 gene annotations on chromosome 6 and on other scaffolds as compared to *Camelus ferus* [12]. We were only able to

recover 39 genes for CamFer2, 43 for CamDro3, and 36 for CamBac2. These lower numbers might be due to assembly quality as the contig (not scaffold) lengths are much longer in the Ming et al. [12] *Camelus ferus* assembly than in CamDro3, CamBac2, or CamFer2. Moreover, mean nucleotide diversity among dromedaries, domestic camels, and wild camels were not significantly different when using either alignments made with all SNPs and indels or only non-synonymous SNPs. In Ming et al. [12], the authors also compared the heavy-chain locus on chromosome 6 between the wild camel and alpaca (*Vicugna pacos*), and found that the gene content and order were very similar between the species. Interestingly, the alpaca, one of the four New World camel species, is evolutionarily the most closely related species to the Old World camels. Only recently, the most up-to-date chromosome-level reference genome assembly was released as VicPac3.1 [46]. Latest research shows that the genomic sequences of Natural Killer cell Receptor (NKR) genes were highly similar in both dromedary and domestic camel to alpaca sequences, as well as the organization of this genomic region [25]. Furthermore, high sequence similarity was observed for genes in the three different classes of MHC as well as MHC genes organization [46, 47].

Conclusions

In this study, using different computational methods, we were able to improve genomic resources for *Camelus dromedarius*, *C. bactrianus* and *C. ferus*. Our data provides high-quality genome assemblies, which are now more contiguous and have fewer and longer scaffolds than the previous version, and are promising resources for the scientific community. Moreover, our results give new insights into the differences in mean nucleotide diversity in immune response genes within and among the three Old World camel species. From the three species, domestic Bactrian camels had the highest mean nucleotide diversity, and from the different functional gene groups, MHC genes had the highest mean diversity. Examining genetic variation in diverse immune genes in camels should be a priority, not only because camels are well adapted to extreme environments even in contact with different pathogens, but also because both domestic species are economically very important, and the wild two-humped camel is critically endangered. The data also showed that studies focused on functionally important parts of the genes, combined with analyses of selection at the molecular and population level, will be helpful to improve the understanding of the biology and evolution of these species. Altogether, this work not only opens doors for future immunogenome studies, but also serves as a reference to further genome assembly improvements using computational methods.

Methods

Previous dromedary genome assemblies

CamDro1

The original North African dromedary genome assembly (CamDro1) was created from a female dromedary “Waris” ([8]; GenBank accession: GCA_000803125.1). Briefly, two types of Illumina libraries were generated and sequenced: 500 bp (short-insert, 100 bp paired-end reads) and 5 Kbp (long-insert/mate-pair, 50 bp paired-end reads) libraries. Short- and long-insert reads were trimmed and, after short-insert reads error-correction, de novo assembled with ABYSS [48] with a k-mer value of 64.

CamDro2

Dovetail Genomics (Santa Cruz, California, USA) created and sequenced Chicago and Dovetail Hi-C libraries derived from the same dromedary “Waris” used in CamDro1. First, the CamDro1 assembly was scaffolded using Dovetail Chicago data run through the HiRise pipeline [49]. Next, the Chicago assembly was scaffolded with Hi-C data. Using a PacBio Sequel sequencer, 11x long-read coverage were generated ([11]; Sequence Read Archive (SRA) accession: SRP050586) and PBJelly [50] was used to fill in gaps in the Hi-C assembly. PBJelly assembly was polished with Pilon [51] employing the same trimmed and error-corrected Illumina short-insert sequences used for the de novo assembly of CamDro1 by Fitak et al. ([8]; SRA accession: SRR2002493). Gaps present in the Pilon assembly were then filled with ABYSS Sealer [52]. Finally, the ABYSS assembly was polished with Pilon once again. This assembly is referred to as CamDro2 ([11]; GCA_000803125.2).

Improving the dromedary genome assembly: CamDro3

The CamDro2 assembly was re-scaffolded using the original Dovetail Chicago and Hi-C reads with the HiRise pipeline. We then filled in gaps using our PacBio long-reads ([11]; SRA accession: SRP050586), running PBJelly v. 15.8.24 twice. Instead of polishing the assembly with Pilon, we used a standard variant calling workflow, which increased RNA-Seq reads mapping rates relative to the Pilon-polished assembly (Table 4). Briefly, we first mapped trimmed and error-corrected Illumina short-insert sequences ([8]; Sequence Read Archive accession: SRR2002493) using BbMap v. 38.12 (<https://sourceforge.net/projects/bbmap/>) with the vslow and usejni settings to the PBJelly assembly. We then sorted and indexed the resulting BAM file with Sambamba v. 0.6.7 [55] and called variants with CallVariants v. 38.12 (<https://sourceforge.net/projects/bbmap/>). We finally used BCftools v. 1.2 (<http://samtools.github.io/bcftools/>) to generate a consensus sequence for which we filled in gaps using ABYSS Sealer v. 2.1.0 [52] using default settings except

Table 4 Assembly statistics for the CamDro2; CamDro3 (Pilon) using one round of Pilon [51] for polishing; and CamDro3 (BBMap) using one round of variant calling with BBMap (<https://sourceforge.net/projects/bbmap/>) for polishing. Note that CamDro3 (BBMap) was chosen over CamDro3 (Pilon) as the final version of CamDro3 because of better BUSCO and RNA-Seq mapping percentages

	Assembly		
	CamDro2	CamDro3 (Pilon)	CamDro3 (BBMap)
Total size	2,154,386,959	2,194,229,671	2,169,346,739
Gap length	20,603,579	17,930,821	17,043,352
Scaffolds			
Number	23,439	21,070	21,070
Longest	124,992,380	125,472,505	124,715,342
N90 ^a	4,922,612	25,062,887	24,767,672
L90 ^b	31	32	32
N50 ^a	75,021,453	70,557,636	70,369,702
L50 ^b	11	12	11
Contigs^c			
Number	45,969	41,934	53,085
Longest	9,490,880	14,412,615	2,012,572
N90	177,587	202,272	49,444
L90	1944	1436	10,023
N50	1,333,162	1,961,815	236,380
L50	423	303	2637
Single-copy BUSCOs ^d	3851	3853	3852
Duplicated BUSCOs	24	23	25
Fragmented BUSCOs	133	132	134
Missing BUSCOs	96	96	93
RNA-Seq Mapping Percentage ^e	88.30	90.36	92.04

^aN90/N50 are the scaffold or contig lengths such that the sum of the lengths of all scaffolds or contigs of this size or larger is equal to 90/50% of the total assembly length

^bL90/L50 are the smallest number of scaffolds or contigs that make up at least 90/50% of the total assembly length

^cUsing minimum gap length of 25 bp

^dBUSCOs: Benchmarking Universal Single-Copy Orthologs [19] are mammalian BUSCOs from OrthoDB v. 9.1 genes [20]

^eOverall mapping rates using HiSat v. 2.1.0 [53] of dromedary RNA-Seq reads from Sequence Read Archive accession: SRP017619 and Alim et al. [54]

for a bloom filter size of 40 GB and multiple K values from 90 to 20 in increments of 10. We refer to this as the CamDro3 assembly (GCA_000803125.3).

RNA-Seq analysis of dromedary

To assess the quality of the new assembly, we aligned 10 sets of paired-end RNA-Seq reads (Alim et al., 2019) to the original assembly (CamDro1), to CamDro2, the new assembly (CamDro3), and to several controls: *C. dromedarius* (RefSeq version - GCA_000767585.1), *C. bactrianus* (GCA_000767855.1), *C. ferus* (GCA_000311805.2) and *Bos taurus* (cattle) (GCA_000003055.3). The 10 RNA-Seq datasets were part of a 2 × 2 factorial experiment: summer vs. winter seasons and supraoptic nucleus (SON) vs. neurointermediate lobe (NIL) brain tissues, with $n = 3$ replicates in each class. Tissue was homogenized and extracted using Trizol/chloroform (ThermoFisher), and purified with the RNeasy MiniKit (Qiagen). The library template was prepared using

a ribosome depletion protocol (Ribo-Zero Gold; Illumina) and libraries prepared using TruSeq Stranded protocol (Illumina). Samples were multiplexed into lane pools with an 8pM concentration and sequenced (100 bp paired-end reads with an average 134 bp insert size) to a depth of > 35 million reads using an Illumina HiSeq 2500. Two of the 12 replicates were rejected for insufficient quality. We used Tophat v. 2.0.9 [56] with default settings to align reads to each genome and report overall alignment rate (default output of Tophat) within each class. For chromosome mapping we then used blastn v. 2.2.31+ [57] to map 4981 probe sequences assigned to *Vicuna (Lama) pacos* chromosomes [11, 22] to CamDro3 assembly scaffolds. We followed the same procedure as Elbers et al., [11].

Annotation to compare CamDro3 to CamDro2

To compare CamDro2 and CamDro3 assemblies, we annotated CamDro3 following the same steps used to

annotate CamDro2 [11]. Briefly, we annotated scaffolds greater than 10 Kbp with MAKER v. 2.31.9 [14, 58]. We masked repetitive regions with RepeatMasker v. open-4.0.7 against the entire Dfam_Consensus release 20,170,127 database. We included ab initio gene predictions from GeneMark-ES 4.33 [59], expressed sequence tag (EST) transcripts, and protein sequences. For ESTs, we assembled transcripts from two dromedary transcriptome experiments (SRA accession: SRP017619 and [54]). We performed adapter and quality trimming on raw demultiplexed paired-end reads using BBDuk v. 37.25, using the following settings: ktrim = r, k = 23, mink = 11, hdist = 1, tpe, tbo, qtrim = rl, trimq = 15. We then mapped quality and adapter trimmed reads to the CamDro3 assembly using HiSat v. 2.1.0 [53] using a maximum intron length of 100,000 and the “dta” option. Reads were assembled into transcripts using StringTie v. 1.3.3b [60] and extracted using Gffread v. 0.9.9 (<https://github.com/gperte/gffread>). For proteins, we combined predicted proteins from *B. taurus*, *C. bactrianus*, and *V. pacos* (GenBank accessions [NCBI annotation release]: GCF_000003055.6 [105], GCF_000311805.1 [100], and GCF_000164845.2 [101], respectively). We also included MAKER predicted proteins with an annotation edit distance (AED) < 0.75 from the CamDro1 assembly [8]. We trained Augustus v. 3.3 [61] using BUSCO v. 3.0.2 (Simão et al., 2015) searching for Eukaryota OrthoDB v. 9.1 genes [20]. We used a *C. dromedarius* specific repeat library created with RepeatModeler v. open-1.0.10 (<http://www.repeatmasker.org>) with the CamDro3 as input. We filtered the repeat library from RepeatModeler to remove known UniProt/SwissProt v. 2017_10 [62] proteins using ProtExcluder v. 1.1 [63]. We only retained genes, transcripts, and proteins with AED ≤ 0.50. Next, we predicted putative gene function with DIAMOND v. 0.9.19 [13] searches against the UniProt/TrEMBL release 2018_07 database using an e-value cutoff of 1e⁻⁶. For the CamDro1, CamDro2, and CamDro3 assemblies, we also mapped proteins predicted by MAKER against the same UniProt/TrEMBL database using DIAMOND and generated a frequency polygon of the query sequence length (predicted proteins) divided by the subject sequence length (UniProt/TrEMBL proteins) to assess if predicted proteins were truncated (query sequence length divided by the subject sequence length < 1.0) due to uncorrected insertions/deletions (indels) introduced by PacBio reads that might interrupt reading frames affecting protein translation [64].

Reference-guided assembly of the domestic Bactrian and wild camel genomes

We used CamDro3 in a reference-guided assembly strategy implemented with Ragout v. 2.0 [16] to upgrade the *C. bactrianus* (CamBac1, GCF_000767855.1, [10]) and *C.*

ferus (CamFer1, GCF_000311805.1, [9]) genome assemblies to chromosome-level scale. Briefly, we used default settings in Progressive Cactus v. Github commit c4bed56c0cd48d23411038acb9c19bcae054837e [17, 65] to generate HAL (hierarchical alignment format) alignments between CamDro3 and CamBac1 or CamDro3 and CamFer1, and then used Ragout with the “refine” and “small synteny block” settings to convert the alignments to FASTA, upgrading the CamBac1 and CamFer1 assemblies to CamBac2 and CamFer2, respectively. Before alignment with Progressive Cactus, we repeat-masked CamDro3 with RepeatMasker v. open-4.0.8 (<http://www.repeatmasker.org>) against the mammal repeats from RepBase RepeatMaskerEdition-20,181,026 [66]. We filled in gaps in CamBac2 and CamFer2 with GapFiller v. 1.10 [18] using default settings and BowTie [67] as the aligner. The paired-end reads used to fill in gaps were the original Illumina short-reads used in assembly with an insert size less than or equal to 800 bases (For CamBac2 SRA accessions: SRR1552325, SRR1552327, SRR1552330, SRR1552336, SRR1552341, SRR1552346, SRR1552347, and SRR1552348; for CamFer2 SRA accession: SRR671683), which we trimmed with BBDuk v. 37.76 (<https://sourceforge.net/projects/bbmap/>), using the following settings: ktrim = r, k = 23, mink = 11, hdist = 1, tpe, tbo, qtrim = rl, trimq = 15, ref. = bbmap-37.76/resources/adapters.fa. We used `assemblathon_stats.pl` (http://korflab.ucdavis.edu/Datasets/Assemblathon/Assemblathon2/Basic_metrics/assemblathon_stats.pl) to compare assembly statistics between CamFer2 and the *C. ferus* genome assembly from Ming et al. [12] using a genome size of 2.1 Gbp. To assess the level of disagreement between CamFer2 and *C. ferus* genome assembly from Ming et al. [12], we made a whole genome alignment with Minimap2 v. 2.17 [68] using the “asm5” preset. We then used D-GENIES [69] to generate a dot plot for the alignment by using the contig sorting function and filtering alignments for strong precision. Chromosomal synteny between the wild camel and dromedary was analyzed by Ming et al. [12] after whole-genome alignment between *C. ferus* genome assembly (new-CamFer) and CamDro3, where assignment of the chromosome nomenclature between these species was similar, with only few structural differences at the megabase (Mbp) scale. Synteny is likely highly conserved between wild camel and dromedary, and domestic Bactrian and dromedary.

Most up to date annotation for CamBac2, CamFer2, CamDro3

To get the most up to date annotation for CamBac2, CamFer2, and CamDro3, we annotated scaffolds greater than 10 Kbp in these assemblies with MAKER v. 2.31.10. We masked repetitive regions with RepeatMasker v.

open-4.0.7 against the entire Dfam_Consensus release 20,170,127 database. We included ab initio gene predictions from GeneMark-ES v. 4.38, EST transcripts, and protein sequences. For CamDro3 ESTs but CamBac2 and CamFer2 alternative ESTs, we assembled transcripts from two dromedary transcriptome experiments (SRA accession: SRP017619 and [54]). We performed adapter and quality trimming on raw demultiplexed paired-end reads using BBDuk v. 37.25, using the following settings: ktrim = r, k = 23, mink = 11, hdist = 1, tpe, tbo, qtrim = rl, trimq = 15. We then mapped quality and adapter trimmed reads to the CamDro3 assembly using HiSat v. 2.1.0 using a maximum intron length of 100,000 and the “dta” option. Reads were assembled into transcripts using StringTie v. 1.3.3b and extracted using Gffread v. 0.9.9. For CamBac2 ESTs but CamDro3 and CamFer2 alternative ESTs, we processed transcriptome reads from *C. bactrianus* (SRA accessions: SRP014573 and SRP148535) with HiSat, StringTie, and Gffread as before but mapped quality controlled reads to the CamBac2 assembly. For proteins, we combined predicted proteins from *B. taurus*, *C. bactrianus*, *C. dromedarius*, *C. ferus*, and *V. pacos* (GenBank accessions (NCBI annotation release): GCF_002263795.1 (106), GCF_000767585.1 (100), GCF_000767855.1 (100), GCF_000311805.1 (101), and GCF_000164845.2 (101), respectively). We trained Augustus v. 3.3.2 using BUSCO v. 3.0.2 searching for Eukaryota OrthoDB v. 9.1 genes in CamDro3, CamBac2, and CamFer2. We used a *C. dromedarius*, *C. bactrianus*, or *C. ferus* specific repeat library created with RepeatModeler open-1.0.10 with the CamDro3, CamBac2, or CamFer2 assemblies as input, respectively. We filtered each repeat library from RepeatModeler to remove known UniProt/Swiss-Prot release 2018_11 proteins using ProtExcluder v. 1.1. We only retained genes, transcripts, and proteins with AED \leq 0.50. Next, we predicted putative gene function with blastp v. 2.2.31+ [57] searches against the UniProt/Swiss-Prot release 2018_11 database using an e-value cutoff of $1e^{-6}$.

Variant calling

From whole-genome sequencing reads (100-bp Illumina paired end reads) of 25 Old World camels [21], we removed adapter sequences and reads with > 10% uncalled bases and/or > 50% of bases with a Phred-scaled quality score < 4. We also trimmed reads with PoPoolation v. 1.2.2 [70], where low-quality bases with a Phred score below 20 at the ends of the reads were removed. We converted base quality scores from Phred 64 to Phred 33 encoding and performed quality trimming with Repair v. 38.39 (<https://sourceforge.net/projects/bbmap/>) using the qtrim = rl and trimq = 15 settings. We mapped quality and adapter trimmed paired-end reads for *C. bactrianus*, *C. dromedarius*, and *C. ferus* individuals to the

CamBac2, CamDro3, and CamFer2 references, respectively with BWA-MEM v. 0.7.17 [71, 72]. We converted SAM files to BAM files with SAMtools v. 1.9 [73], then cleaned, sorted, added read groups, and marked duplicates with Picard v. 2.18.10 (<http://broadinstitute.github.io/picard>). We called variants for each species separately with CallVariants v. 38.39 (<https://sourceforge.net/projects/bbmap/>), keeping only SNPs and indels with quality scores greater than or equal to 27. We predicted what SNP alleles might be synonymous or non-synonymous using snpEff v 4.0e [74].

We calculated coverage metrics with mosdepth v. 0.2.6 [75] with the settings “-n --fast-mode and --by 500”. We used R v. 3.6.0 to test for differences in mean coverage, total number of SNPs, number of synonymous SNPs, number of non-synonymous SNPs, number of insertions, and number of deletions within species with the “lm” and “anova” base functions. For all models, we used a Benajimini-Hochberg post-hoc test [76] implemented in glht and summary functions in the R package multcomp v. 1.4–10 [77].

Heterozygosity rates in exons and introns

We predicted intron regions for gene annotations of CamDro3, CamBac2, and CamFer2 using Genome Tools v. 1.5.8 [78] with the gff3 function and -addinrons -retainids options. We then generated gene annotation files of only exons or introns for each camel species. We filtered the VCF files for each individual to retain only heterozygous SNPs. We then used BEDTools intersect v. 2.29.0 [79] to count the number of heterozygous SNPs for each individual ($n = 25$) in the exons or introns across the genome. We estimated heterozygosity as the number of heterozygous SNPs in the exons or introns of a given gene for a given individual divided by gene length.

We used the lm function in R 3.6.3 using heterozygosity as the dependent variable and the interaction of species and whether heterozygosity was estimated from exons or introns (hereafter exons or introns) as the independent variable. Residuals needed to be log10 transformed to be normally distributed. We used a generalized least squares variation of ANOVA (hereafter ANOVA [80]) as our transformed data did not have homogeneous variance. To control for heterogeneous variance, we used weights as “varIdent=(1|interaction of species, and exons or introns)” implemented with the gls function in the R package nlme v. 3.1–147 [81]. We used a Benajimini-Hochberg post-hoc test as before implemented with the glht and summary functions in the R package multcomp v. 1.4–13 and the cld function in multcomp with the options level = 0.05 and decreasing = T to determine if means for all species for exons and introns were significantly different at the 0.05 level.

Nucleotide diversity

Two comparisons of nucleotide diversity were made, (i) between functionally different gene groups within each species: innate immune response genes, adaptive immune response genes, MHC class I and II genes, and rest-of-genome genes, and (ii) between Old World camel species: domesticated dromedaries and Bactrian camels, and wild camels among gene groups.

To test for differences in genetic variation among functionally different gene groups, we performed blastp searches of CamBac2, CamFer2, and CamDro3 predicted proteins against UniProt/Swiss-Prot release_2018_11 proteins to assign gene ontology terms, and filtered these gene/GO term lists by the gene ontology terms “innate immune response” and “adaptive immune response” using the rGO2TR package [82]. For MHC class I and class II genes, we filtered the GFF3 (General Feature Format) files of gene annotations manually. For the rest-of-genome gene group, we examined genes that were not assigned to either the innate or adaptive immune response gene groups. We used BCFtools v. 1.9 to generate a consensus sequence with IUPAC codes for each individual against its respective reference genome for each gene being analyzed and made a multiple sequence alignment for each gene and species with FSA v. 1.15.9 [83] with MuMmer v. 4.0.0beta2 [84] for long alignments. Finally, we calculated nucleotide diversity for entire gene sequence multiple sequence alignments (each species separately) using the R package Pegas’s “nuc.div” function [85]. We used R v. 3.6.3 to test for differences in mean nucleotide diversity within species among gene groups. For this we compared the 95% confidence intervals of the mean estimated with the boot.ci function’s “basic” confidence interval method based on 1000 “ordinary” simulations (i.e., non-parametric bootstraps) implemented with the boot function from the R package boot v. 1.3–24 [86]. We chose to use non-parametric inference as the residuals could not be transformed to approximate a normal distribution, precluding the use of traditional ANOVA/linear model testing for differences in means.

For analyzing differences in mean nucleotide diversities within gene groups but among species, we used the same procedures as before but with the explanatory variable “species” (dromedary, domestic Bactrian camel, or wild camel) and response variable “nucleotide diversity” (adaptive, innate, MHC, or rest-of-genome genes). In addition to nucleotide diversity, estimated with gene consensus sequences made with non-synonymous and synonymous SNPs and indels, we also repeated all steps above using only non-synonymous SNPs (indels and synonymous SNPs were not included).

Interestingly, camels (New World and Old World camels) produce homodimeric heavy-chain

immunoglobulins (hcIGs [87]); without a light chain and with the antigen-binding fragment reduced to a single heavy-chain variable domain (VHH), in addition to the conventional antibodies [88]. To assess the nucleotide diversity of single-domain heavy-chain IG genes in our data, we first downloaded the scaffold.fasta.gz (Ming et al.’s [12] *Camelus ferus* genome assembly) and IGH.gff (heavy-chain immunoglobulin gene annotations) from https://figshare.com/articles/Data_from_Chromosome-level_assembly_of_wild_Bactrian_camel_genome_reveals_organization_of_immune_gene_loci/11297489. We then lifted over the *Camelus ferus* IGH.gff gene annotations assembly [12] to CamDro3, CamBac2, and CamFer2 using Liftoff Github commit #77b7c4c91b294737d18d7a76e3611d279bebea6e [89]. We repeated previous nucleotide diversity assessment steps as described above (see *Nucleotide diversity*) using the new lifted over annotations. As we could not transform data to have residuals with a normal distribution, we followed analysis steps as before, except that we used R v. 3.6.3 along with the R package boot v. 1.3–25 [86], and compared mean nucleotide diversity in heavy-chain immunoglobulin genes among dromedaries, domestic camels, and wild camels.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-06990-4>.

Additional file 1: Supplemental Table 1. Assembly statistics for the CamFer2 and the *Camelus ferus* genome (new-CamFer) assembly from Ming et al. (2020b) using a genome size of 2.1 Gbp.

Additional file 2: Supplemental Table 2. Means with 95% bootstrap confidence limits (CL, see Methods) of nucleotide diversity for alignments made with non-synonymous and synonymous SNPs and indels and only non-synonymous SNPs in HC (heavy-chain) immunoglobulin genes in DC (domestic camel), DROM (dromedary), and WC (wild camel).

Additional file 3: Supplemental Figure 1. RNA-Seq mapping rates.

Additional file 4: Supplemental Figure 2. D-GENIES (Cabanettes & Klopp, 2018) dot plot made with Minimap2 [68] whole-genome alignment between CamFer2 and the *Camelus ferus* genome (new-CamFer) assembly from Ming et al., [12]. Contigs are sorted and matches are filtered using the strong precision setting in D-GENIES.

Additional file 5: Supplemental Figure 3. Means with 95% bootstrap confidence intervals (see Methods) of nucleotide diversity for alignments made with non-synonymous and synonymous SNPs and indels (a) and only non-synonymous SNPs (b): MHC class I and II genes (top panel), innate (second panel), adaptive (third panel), and the rest of genome genes (bottom panel) for: DROM (dromedary, *C. dromedarius*), DC (domestic Bactrian camel, *C. bactrianus*), and WC (wild camel, *C. ferus*). Uppercase letters above upper 95% confidence limits indicate groups have different (non-matching letters) or not different (matching letters) means based on non-overlapping confidence intervals.

Additional file 6: Supplemental Figure 4. Means with 95% bootstrap confidence intervals (see Methods) of nucleotide diversity for alignments made with (left) non-synonymous SNPs, (right) all SNPs and indels in HC (heavy-chain) antibody (immunoglobulin) genes in DC (domestic camel), DROM (dromedary), and WC (wild camel). Uppercase letters above upper 95% confidence limits indicate groups have different (non-matching letters) or not different (matching letters) means based on non-overlapping confidence intervals.

Abbreviations

AED: Annotation edit distance; ANOVA: Analysis of Variance; DC: Domestic Bactrian camel (*Camelus bactrianus*); Drom: Dromedary (*Camelus dromedarius*); EST: Expressed sequence tag; hclGs: Homodimeric heavy-chain immunoglobulins; Indels: Insertions/deletions; IR: Immune-response; IUPAC: International Union of Pure and Applied Chemistry; Kbp: Kilo base pairs; Mbp: Mega base pairs; MHC: Major histocompatibility complex; NIL: Neurointermediate lobe; NKR: Natural Killer cell Receptor; RNA-Seq: Ribonucleic acid sequencing; SD: Standard deviation; SNP: Single nucleotide polymorphism; SON: Supraoptic nucleus; SRA: Sequence Read Archive; VHH: Single heavy-chain variable domain; WC: Wild camel (*Camelus ferus*)

Acknowledgements

We thank the CSC – IT Center for Science, Finland, for generous computational resources. We are also very grateful to all camel owners for their agreement to use the collected data for scientific purposes and to the Wild Camel Protection Foundation for continuous support.

Authors' contributions

S.L. wrote the first draft of the manuscript, J.P.E. and M.F.R. performed analyses, P.H. and P.A.B. conceived and managed the project, J.M.F. and J.C. revised the manuscript. All authors interpreted the results, provided valuable discussions, commented and approved the final manuscript.

Funding

S.L. and J.P.E. acknowledge funding from the Austrian Science Fund (FWF) project P29623-B25 to P.B. J.M.F. was supported by Fundação para a Ciência e a Tecnologia, FCT, Portugal (CEECIND/00372/2018 research contract).

Availability of data and materials

CamDro3 is available from NCBI GenBank (GCA_000803125.3) and NCBI RefSeq (GCF_000803125.2). Our CamDro3/CamBac2/CamFer2 gene annotations, predicted mRNA and proteins, and assemblies for gene annotations are available from Dryad (<https://doi.org/10.5061/dryad.qv9s4mwb3>). Raw VCF files (snp and indel variants) for each camel are also available in the Dryad repository. Example scripts and code for analyses are available from the Dryad repository.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Interdisciplinary Life Sciences, Research Institute of Wildlife Ecology, Vetmeduni Vienna, Vienna, Austria. ²Intelligent Systems Laboratory, University of Bristol, Bristol, UK. ³CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal. ⁴Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal. ⁵Wild Camel Protection Foundation Mongolia, Jukov avenue, Bayanzurh District, Ulaanbaatar 13343, Mongolia. ⁶Wellcome Sanger Institute, Hinxton, UK. ⁷Department of Mathematics and Statistics, Helsinki Institute for Information Technology, University of Helsinki, FIN-00014 Helsinki, Finland. ⁸Department of Biostatistics, University of Oslo, N-0317 Oslo, Norway. ⁹Department of Animal Genetics, Veterinary and Pharmaceutical University, Brno, Czech Republic. ¹⁰Ceitec VFU, RG Animal Immunogenomics, Brno, Czech Republic.

Received: 18 May 2020 Accepted: 13 August 2020

Published online: 03 September 2020

References

1. Van Houte S, Ekroth AKE, Broniewski JM, Chabas H, Ashby B, Bondy-denomy J, et al. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature*. 2016;532:385.

2. Ramey HR, Decker JE, McKay SD, et al. Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics*. 2013;14:382.
3. Horrocks NPC, Matson KD, Tieleman BI. Pathogen pressure puts immune defense into perspective. *Integr Comp Biol*. 2011;51:563–76.
4. Plasil M, Mohandesan E, Fitak RR, Musilova P, Kubickova S, Burger PA, et al. The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes. *BMC Genomics*. 2016;17:167. <https://doi.org/10.1186/s12864-016-2500-1>.
5. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet*. 2013;14:301–23.
6. Jepson A, Banya W, Sisay-Joof F, Hassan-King M, Nunes C, Bennett S, et al. Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens. *Infect Immun*. 1997;65:872–6.
7. Acevedo-Whitehouse K, Cunningham AA. Is MHC enough for understanding wildlife immunogenetics? *Trends Ecol Evol*. 2006;21:433–8.
8. Fitak RR, Mohandesan E, Corander J, Burger PA. The de novo genome assembly and annotation of a female domestic dromedary of north African origin. *Mol Ecol Resour*. 2016;16:314–24.
9. Jirimutu Wang Z, et al. Genome sequences of wild and domestic bactrian camels *Nat Commun*. 2012;3:1202.
10. Wu H, Guang X, Al-Fageeh MB, et al. Camelid genomes reveal evolution and adaptation to desert environments. *Nat Commun*. 2014;5:5188.
11. Elbers JP, Rogers MF, Perelman PL, Proskuryakova AA, Serdyukova NA, Johnson WE, et al. Improving Illumina assemblies with hi-C and long reads: an example with the north African dromedary. *Mol Ecol Resour*. 2019;19:1015–26.
12. Ming L, Wang Z, Yi L, Batmunkh M, Liu T, Siren D, et al. Chromosome-level assembly of wild Bactrian camel genome reveals organization of immune gene loci. *Mol Ecol Resour*. 2020;00:1–11.
13. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12:59–60.
14. Holt, C., Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 2011;12(491).
15. Yandell M, Ence D. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet*. 2012;13:329–42. <https://doi.org/10.1038/nrg3174>.
16. Kolmogorov M, Raney B, Paten B, Pham S. Ragout - a reference-assisted assembly tool for bacterial genomes. *Bioinformatics*. 2014;30:i302–9.
17. Paten B, Diekhans M, Earl D, John JS, Ma J, Suh B, et al. Cactus graphs for genome comparisons. *J Comput Biol*. 2011;18:469–81.
18. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. *Genome Biol*. 2012;13:R56.
19. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2.
20. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res*. 2017;45:D744–9.
21. Fitak RR, Mohandesan E, Corander J, Yadamsuren A, Chuluunbat B, Abdelhadi O, et al. Genomic signatures of domestication in Old World. *Commun Biol*. 2020;3:1–10. <https://doi.org/10.1038/s42003-020-1039-5>.
22. Avila F, Baily MP, Perelman P, Das PJ, Pontius J, Chowdhary R, et al. A comprehensive whole-genome integrated cytogenetic map for the alpaca (*Lama pacos*). *Cytogenet Genome Res*. 2014;144:196–207.
23. Muyllderms S, Baral TN, Retamozzo VC, De Baetselier P, De Genst E, Kinne J, et al. Camelid immunoglobulins and nanobody technology. *Vet Immunol Immunopathol*. 2009;128:178–83. <https://doi.org/10.1016/j.vetimm.2008.10.299>.
24. Antonacci R, Linguiti G, Burger PA, Castelli V, Pala A, Fitak R, et al. Comprehensive genomic analysis of the dromedary T cell receptor gamma (TRG) locus and identification of a functional TRGC5 cassette. *Dev Comp Immunol*. 2020;106:103614.
25. Futas J, Oppelt J, Jelinek A, Elbers JP, Wijacki J, Knoll A, et al. Natural killer cell receptor genes in camels: Another mammalian model. *Front Genet*. 2019;10 JUL:1–15.
26. Vaccarelli G, Antonacci R, Tasco G, Yang F, Giordano L, El Ashmaoui HM, et al. Generation of diversity by somatic mutation in the *Camelus dromedarius* T-cell receptor gamma variable domains. *Eur J Immunol*. 2012;42:3416–28.

27. Abbas B, Omer OH. Review of infectious diseases of the camel. *Vet Bull.* 2005;75:1–16.
28. Wernery U, Kinne J. Foot and mouth disease and similar virus infections in camelids: a review. *Rev Sci Tech - Off Int des épizooties.* 2012;31:907–18.
29. Hemida MG, Chu DKW, Poon LLM, Perera RAPM, Alhamadi MA, Ng HY, et al. MERS coronavirus in dromedary camel herd, Saudi Arabia. *Emerg Infect Dis.* 2014;20:1231–4.
30. Kurtz J, Kalbe M, Aeschlimann PB, Häberli MA, Wegner KM, Reusch TBH, et al. Major histocompatibility complex diversity influences parasite resistance and innate immunity in sticklebacks. *Proc R Soc B Biol Sci.* 2004; 271:197–204.
31. Uematsu S, Akira S. Toll-like receptors (TLRs) and their ligands. In: Bauer S, Hartmann G, editors. *Toll-like receptors (TLRs) and innate immunity.* Springer: Berlin Heidelberg; 2008. p. 1–20.
32. Gnerre S, Lander ES, Lindblad-toh K, Jaffe DB. Assisted assembly: how to improve a de novo genome assembly by using related species. *Genome Biol.* 2009;10:R88.
33. Almuthen F, Charruau P, Mohandesan E, Mwacharo JM, Orozco-terWengel P, Pitt D, et al. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc Natl Acad Sci.* 2016; 113:6707–12. <https://doi.org/10.1073/pnas.1519508113>.
34. Yadamsuren A, Dulamtseren E, Reading RP. The conservation status and Management of Wild Camels in Mongolia. In: Knoll E-M, Burger PA, editors. *Camels in Asia and North-Africa- interdisciplinary perspectives on their past and present significance.* Austrian Academy of Sciences Press: Wien; 2012. p. 45–54.
35. Dirie MF, Abdurahman O. Observations on little known diseases of camels (*Camelus dromedarius*) in the horn of Africa. *Rev Sci Tech - Off Int des épizooties.* 2003;22:1043–9.
36. Fassi-Fehri MM. Diseases of camels. *Rev Sci Tech Off Int des Epizoot.* 1987;6: 337–54.
37. Bontrop RE, Otting N, de Groot NG, Gaby G. M D. Major histocompatibility complex class II polymorphisms in primates. *Syst Lupus Erythematosus.* 1999;167:339–50.
38. Bernatchez L, Landry C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol.* 2003;16:363–77.
39. Janeway Jr CA, Travers P, Walport M, Shlomchik MJ. The complement system and innate immunity. In: *Immunobiology: The Immune System in Health and Disease.* 5th edition. New York: Garland Science; 2001.
40. Ujvari B, Belov K. Major histocompatibility complex (MHC) markers in conservation biology. *Int J Mol Sci.* 2011;12:5168–86.
41. Elbers JP, Clostio RW, Taylor SS. Neutral genetic processes influence MHC evolution in threatened gopher tortoises (*Gopherus polyphemus*). *J Hered.* 2017;108:515–23.
42. Ming L, Yi L, Sa R, Wang ZX, Wang Z, Ji R. Genetic diversity and phylogeographic structure of Bactrian camels shown by mitochondrial sequence variations. *Anim Genet.* 2017;48:217–20.
43. Ming L, Yuan L, Yi L, Ding G, Hasi S, Chen G, et al. Whole-genome sequencing of 128 camels across Asia reveals origin and migration of domestic Bactrian camels. *Commun Biol.* 2020;3:1–9.
44. Wells K, Gibson DI, Clark NJ, Ribas A, Morand S, McCallum HI. Global spread of helminth parasites at the human–domestic animal–wildlife interface. *Glob Chang Biol.* 2018;24:3254–65.
45. Lado S, Elbers JP, Doskocil A, Scaglione D, Trucchi E, Banabazi MH, et al. Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes. *Commun Biol.* 2020;3:1–8. <https://doi.org/10.1038/s42003-020-1098-7>.
46. Richardson MF, Munyard K, Croft LJ, Allnutt TR, Jackling F, Alshanbari F, et al. Chromosome-level alpaca reference genome VicPac3.1 improves genomic insight into the biology of new world camelids. *Front Genet.* 2019;10:1–15.
47. Plasil M, Wijkmark S, Elbers JP, Oppelt J, Burger PA, Horin P. The major histocompatibility complex of Old World camelids: class I and class I-related genes. *Hla.* 2019;93:203–15.
48. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 2009;19:1117–23.
49. Putnam NH, Connell BO, Stites JC, Rice BJ, Blanchette M, Calef R, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* 2016;26:342–50.
50. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, et al. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLoS ONE.* 2012;7(11):e47768.
51. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 2014;9:e112963.
52. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, et al. ABySS 2.0: resource-efficient assembly of large genomes using a bloom filter effect of bloom filter false positive rate. *Genome Res.* 2017;27:768–77.
53. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements Daehwan HHS public access. *Nat Methods.* 2015;12: 357–60.
54. Alim FZD, Romanova EV, Tay Y-L, Rahman AYBA, Chan KG, Hong KW, et al. Seasonal adaptations of the hypothalamo-neurohypophyseal system of the dromedary camel. *PLoS One.* 2019;14:1–33.
55. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics.* 2015;31:2032–4.
56. Kim D, Perteu G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14:R36.
57. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
58. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 2008;18:188–96.
59. Lomsadze A, Ter-Hovhannisyantsyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* 2005;33:6494–506.
60. Perteu M, Perteu GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33:290–5.
61. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 2006;34(Web Server Issue):W435–9.
62. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Methods Mol Biol.* 2016;1374: 23–54. <https://doi.org/10.1007/978-1-4939-3167-5>.
63. Campbell MS, Law MY, Holt C, Stein JC, Moghe GD, Hufnagel DE, et al. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* 2014;164:513–24.
64. Watson M, Warr A. Errors in long-read assemblies can critically affect protein prediction. *Nat Biotechnol.* 2019;37:124–6.
65. Paten B, Earl D, Nguyen N, Diekhans M, Zerbino D, Haussler D. Cactus: algorithms for genome multiple sequence alignment. *Genome Res.* 2011;21:1512–28.
66. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 2005;110:462–7.
67. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25.
68. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34:3094–100.
69. Cabanettes F, Klopp C. D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* 2018;2018(6):e4958.
70. Kofler R, Orozco-terWengel P, de Maio N, Pandey RV, Nolte V, Futschik A, et al. Popoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One.* 2011;6:e15925.
71. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio.GN]*; 2013.
72. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* 2009;25:1754–60.
73. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25: 2078–9.
74. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* 2012;6:80–92.
75. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics.* 2018;34:867–8.
76. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 1995;57:289–300.

77. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. *Biom J.* 2008;50:346–63.
78. Gremme G, Steinbiss S, Kurtz S. Genome tools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans Comput Biol Bioinforma.* 2013;10:645–56.
79. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
80. Venables WN, Ripley BD. *Modern applied statistics with S.* New York: Springer; 2002.
81. Pinheiro J, Bates D, DebRoy S, Sarkar D. nlme: linear and nonlinear mixed effects models. R package version 3.1–111; 2013.
82. Elbers JP, Taylor SS. GO2TR: a gene ontology-based workflow to generate target regions for target enrichment experiments. *Conserv Genet Resour.* 2015;7:851–7.
83. Bradley RK, Roberts A, Smoot M, et al. Fast statistical alignment. *PLoS Comput Biol.* 2009;5(5):e1000392.
84. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol.* 2018;14:1–14.
85. Paradis E. Pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics.* 2010;26:419–20.
86. Cauty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. R package version 1.3–24; 2019.
87. Ciccicarese S, Burger PA, Ciani E, Castelli V, Linguiti G, Plasil M, et al. The camel adaptive immune receptors repertoire as a singular example of structural and functional genomics. *Front Genet.* 2019;10(OCT):1–14.
88. Hamer-Casterman C, Atarouch T, Muyldermans S, Robinson G, Hamers C, Bajyana E, et al. Naturally occurring antibodies devoid of light chains. *Nature.* 1993;363:446–8 <https://www.nature.com/articles/363446a0.pdf>.
89. Shumate A, Salzberg SL. Liftoff: an accurate gene annotation mapping tool. *bioRxiv.* 2020;2020.06.24.169680. <https://doi.org/10.1101/2020.06.24.169680>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Supplementary Materials

Supplementary Tables

Supplemental Table 1. Assembly statistics for the CamFer2 and the *Camelus ferus* genome (new-CamFer) assembly from Ming et al. (2020b) using a genome size of 2.1 Gbp.

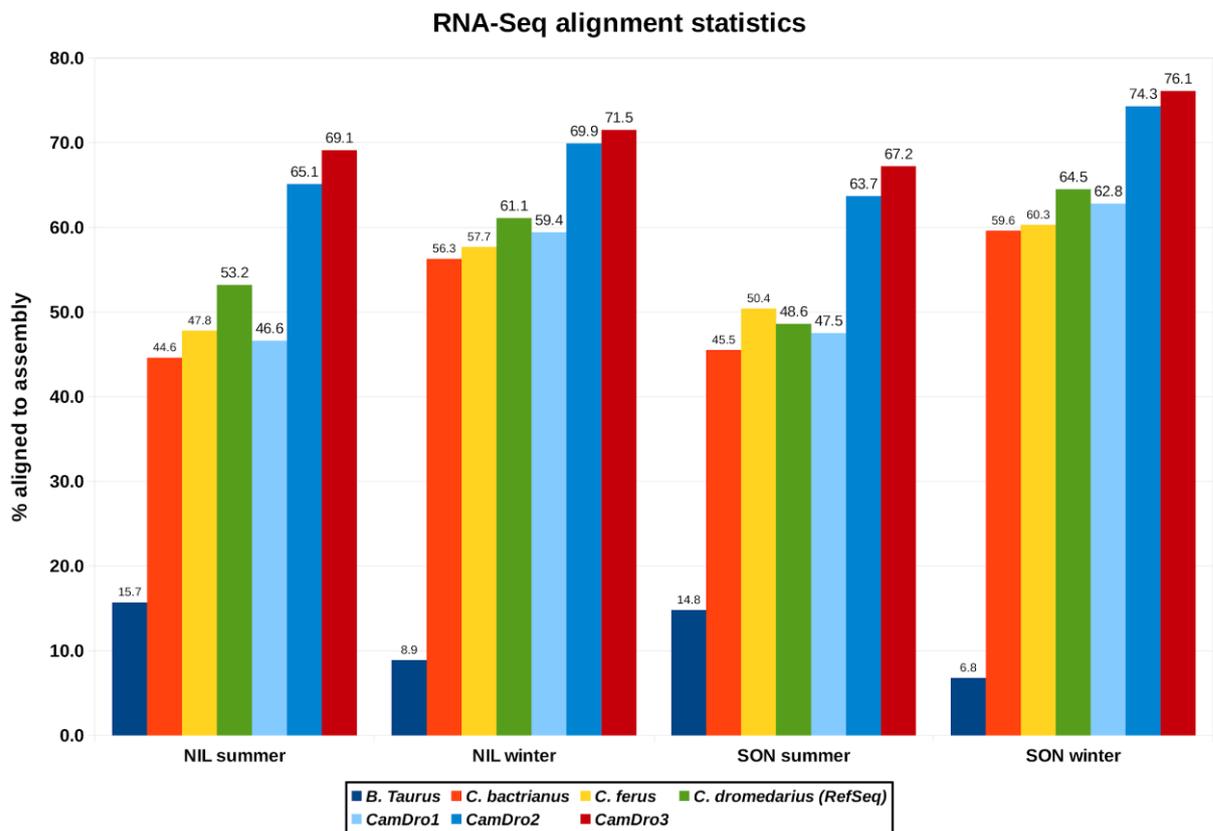
Assembly	new-CamFer	CamFer2
Number of scaffolds	2057	9158
Total size of scaffolds	2087077831	2086258888
Total scaffold length as percentage of assumed genome size	99.4	99.3
Longest scaffold	122453268	123639755
Shortest scaffold	5	200
Number of scaffolds > 1K nt	1997	8465
Percentage of scaffolds > 1K nt	97.1	92.4
Number of scaffolds > 10K nt	1175	619
Percentage of scaffolds > 10K nt	57.1	6.8
Number of scaffolds > 100K nt	99	61
Percentage of scaffolds > 100K nt	4.8	0.7
Number of scaffolds > 1M nt	43	42
Percentage of scaffolds > 1M nt	2.1	0.5
Number of scaffolds > 10M nt	36	36
Percentage of scaffolds > 10M nt	1.8	0.4
Mean scaffold size	1014622	227807
Median scaffold size	11758	1634
N50 scaffold length	76025729	69671486
L50 scaffold count	11	11
NG50 scaffold length	76025729	69671486
LG50 scaffold count	11	11
N50 scaffold - NG50 scaffold length difference	0	0
scaffold %A	29.17	27.96
scaffold %C	20.83	19.65
scaffold %G	20.82	19.66
scaffold %T	29.17	27.98
scaffold %N	0.01	4.75
scaffold %non-ACGTN	0	0
Number of scaffold non-ACGTN nt	0	0
Percentage of assembly in scaffolded contigs	97.6	99
Percentage of assembly in unscaffolded contigs	2.4	1
Average number of contigs per scaffold	2.1	6.7
Average length of break (>25 Ns) between contigs in scaffold	100	1884.878498
Number of contigs	4402	61715
Number of contigs in scaffolds	2382	53692
Number of contigs not in scaffolds	2020	8023
Total size of contigs	2086843331	1987191559

Longest contig	26533942	1096594
Shortest contig	5	4
Number of contigs > 1K nt	4342	54533
Percentage of contigs > 1K nt	98.6	88.4
Number of contigs > 10K nt	3469	29287
Percentage of contigs > 10K nt	78.8	47.5
Number of contigs > 100K nt	1123	5170
Percentage of contigs > 100K nt	25.5	8.4
Number of contigs > 1M nt	389	2
Percentage of contigs > 1M nt	8.8	0
Number of contigs > 10M nt	32	0
Percentage of contigs > 10M nt	0.7	0
Mean contig size	474067	32199
Median contig size	28692	8750
N50 contig length	5365398	104662
L50 contig count	112	4862
NG50 contig length	5355949	97024
LG50 contig count	113	5422
N50 contig - NG50 contig length difference	9449	7638
contig %A	29.17	29.36
contig %C	20.83	20.63
contig %G	20.82	20.64
contig %T	29.17	29.37
contig %N	0	0
contig %non-ACGTN	0	0
Number of contig non-ACGTN nt	0	0

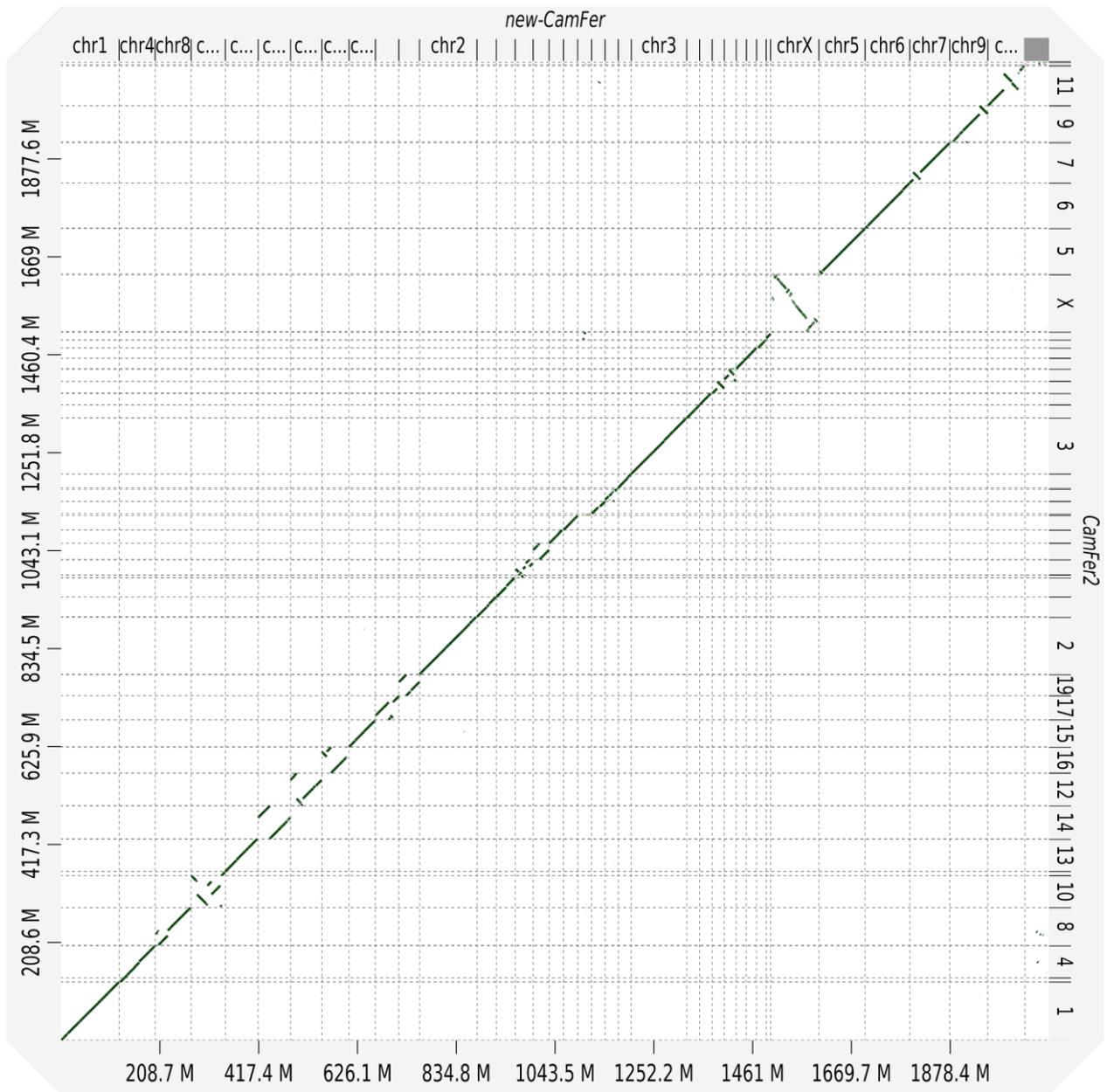
Supplemental Table 2. Means with 95 % bootstrap confidence limits (CL, see Methods) of nucleotide diversity for alignments made with non-synonymous and synonymous SNPs and indels and only non-synonymous SNPs in HC (heavy-chain) immunoglobulin genes in DC (domestic camel), DROM (dromedary), and WC (wild camel).

Non-synonymous SNPs	Mean	Upper 95% CL	Lower 95% CL	Species
	1.003664e-05	2.007329e-05	-1.003664e-05	DROM
	2.973608e-05	5.947216e-05	-2.632341e-05	DC
	1.080563e-04	2.036987e-04	-1.674009e-05	WC
All SNPs and indels	Mean	Upper 95% CL	Lower 95% CL	Species
	0.0004651567	0.0008144805	-1.595646e-05	DROM
	0.0005954219	0.0010261428	-6.131850e-05	DC
	0.0002285253	0.0003719989	5.000419e-05	WC

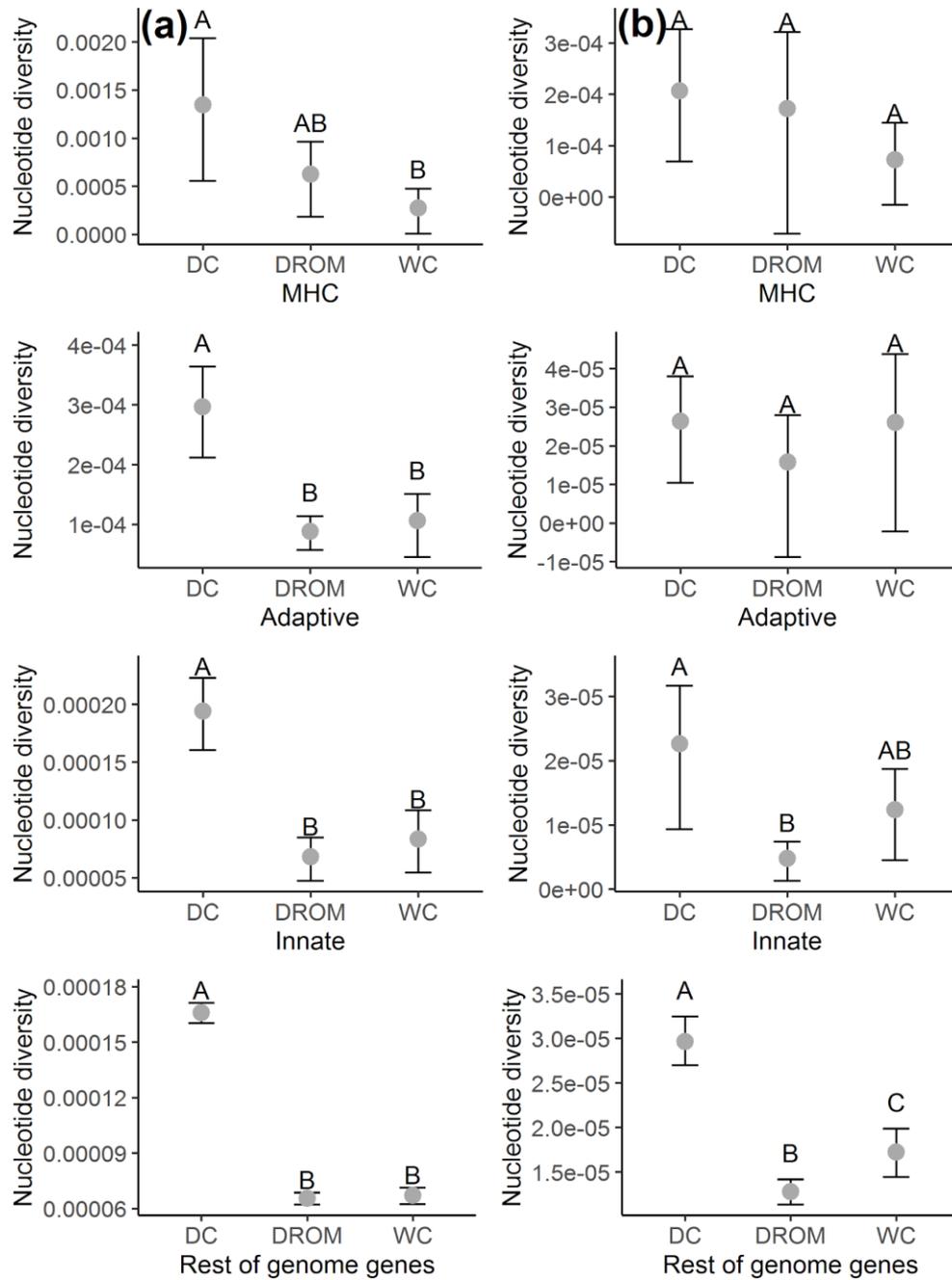
Supplementary Figures



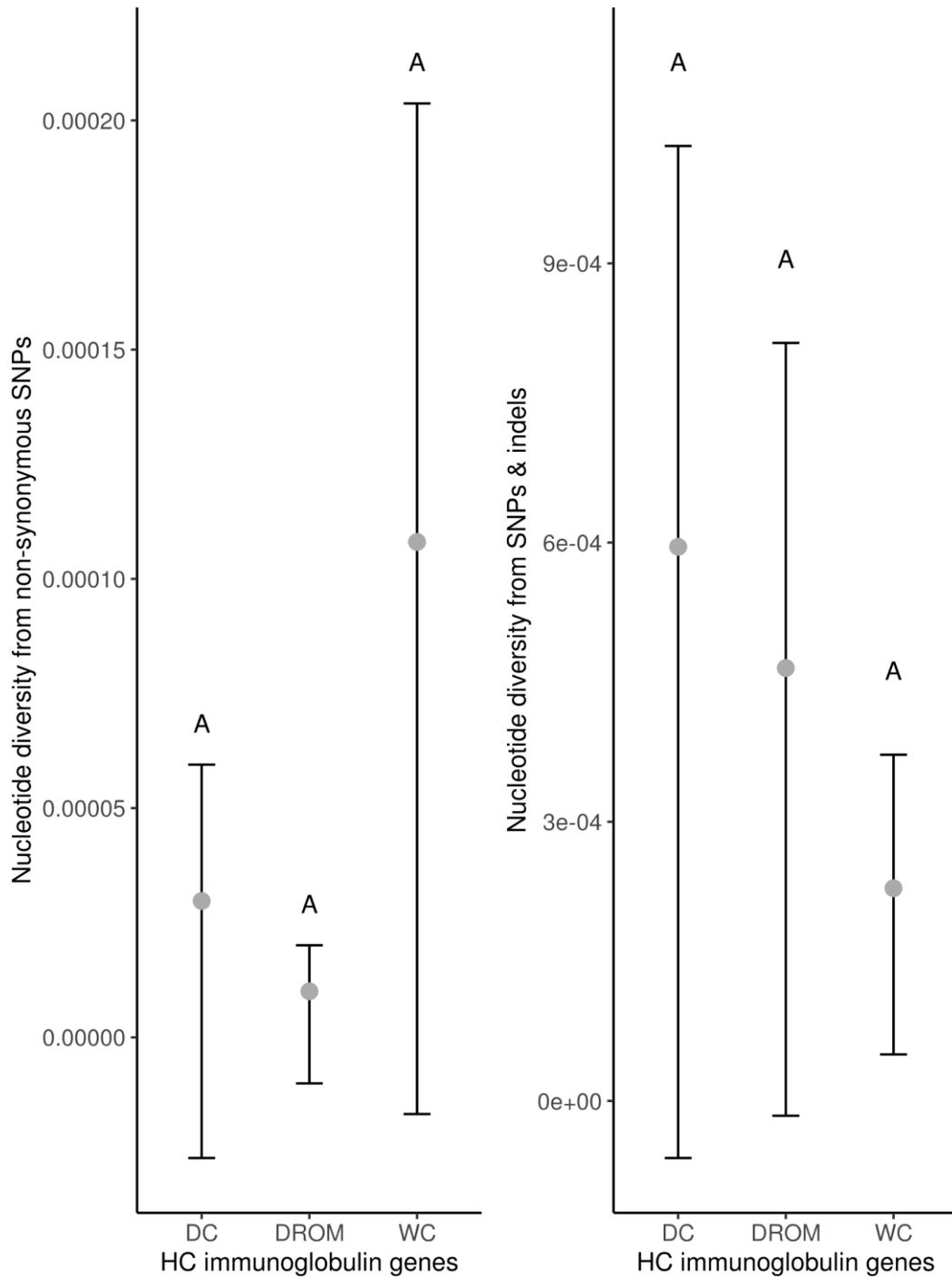
Supplemental Figure 1. RNA-Seq mapping rates.



Supplemental Figure 2. D-GENIES (Cabanettes & Klopp, 2018) dot plot made with Minimap2 [67] whole-genome alignment between CamFer2 and the *Camelus ferus* genome (new-CamFer) assembly from Ming et al., [12]. Contigs are sorted and matches are filtered using the strong precision setting in D-GENIES.



Supplemental Figure 3. Means with 95 % bootstrap confidence intervals (see Methods) of nucleotide diversity for alignments made with non-synonymous and synonymous SNPs and indels (a) and only non-synonymous SNPs (b): MHC class I and II genes (top panel), innate (second panel), adaptive (third panel), and the rest of genome genes (bottom panel) for: DROM (dromedary, *C. dromedarius*), DC (domestic Bactrian camel, *C. bactrianus*), and WC (wild camel, *C. ferus*). Uppercase letters above upper 95 % confidence limits indicate groups have different (non-matching letters) or not different (matching letters) means based on non-overlapping confidence intervals.



Supplemental Figure 4. Means with 95 % bootstrap confidence intervals (see Methods) of nucleotide diversity for alignments made with (left) non-synonymous SNPs, (right) all SNPs and indels in HC (heavy-chain) antibody (immunoglobulin) genes in DC (domestic camel), DROM (dromedary), and WC (wild camel). Uppercase letters above upper 95 % confidence limits indicate groups have different (non-matching letters) or not different (matching letters) means based on non-overlapping confidence intervals.

6.3 Article 3

Lado, S., Elbers, J.P., Plasil, M., Loney, T., Weidinger, P., Camp, J. V., Kolodziejek, J., Futas, J., Kannan, D. O., Orozco-terWengel, P., Horin, P., Nowotny, N., Burger, P. A. (2021). Innate and adaptive immune genes associated with MERS-CoV infection in dromedaries. *Cells*, 10(6), 1291. <https://doi.org/10.3390/cells10061291>

Article

Innate and Adaptive Immune Genes Associated with MERS-CoV Infection in Dromedaries

Sara Lado ¹, Jean P. Elbers ¹, Martin Plasil ^{2,3}, Tom Loney ⁴ , Pia Weidinger ⁵ , Jeremy V. Camp ^{5,6} , Jolanta Kolodziejek ⁵, Jan Futas ^{2,3} , Dafalla A. Kannan ⁷, Pablo Orozco-terWengel ⁸ , Petr Horin ^{2,3} , Norbert Nowotny ^{4,5}  and Pamela A. Burger ^{1,*} 

- ¹ Research Institute of Wildlife Ecology, Department of Interdisciplinary Life Sciences, University of Veterinary Medicine Vienna, 1210 Vienna, Austria; sara.lado@vetmeduni.ac.at (S.L.); JeanPierre.Elbers@vetmeduni.ac.at (J.P.E.)
- ² Department of Animal Genetics, University of Veterinary Sciences Brno, 61242 Brno, Czech Republic; m.plasil@gmail.com (M.P.); jfutas@vfu.cz (J.F.); horin@ics.muni.cz (P.H.)
- ³ RG Animal Immunogenomics, Ceitec Vetuni, 61242 Brno, Czech Republic
- ⁴ College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai 505055, United Arab Emirates; tom.loney@mbru.ac.ae (T.L.); Norbert.Nowotny@vetmeduni.ac.at (N.N.)
- ⁵ Viral Zoonoses, Emerging and Vector-Borne Infections Group, Institute of Virology, University of Veterinary Medicine Vienna, 1210 Vienna, Austria; Pia.Weidinger@vetmeduni.ac.at (P.W.); Jeremy.Camp@meduniwien.ac.at (J.V.C.); Jolanta.Kolodziejek@vetmeduni.ac.at (J.K.)
- ⁶ Center for Virology, Medical University of Vienna, 1090 Vienna, Austria
- ⁷ Al Ain City Municipality, Al Ain 15258, United Arab Emirates; Dafalla.Ahmed@aam.gov.ae
- ⁸ The Sir Martin Evans Building, Cardiff School of Biosciences, Cardiff University, Museum Ave, Cardiff CF10 3AX, UK; Orozco-terWengelPA@cardiff.ac.uk
- * Correspondence: pamelaburger@vetmeduni.ac.at; Tel.: +43-1-25077-7141



Citation: Lado, S.; Elbers, J.P.; Plasil, M.; Loney, T.; Weidinger, P.; Camp, J.V.; Kolodziejek, J.; Futas, J.; Kannan, D.A.; Orozco-terWengel, P.; et al. Innate and Adaptive Immune Genes Associated with MERS-CoV Infection in Dromedaries. *Cells* **2021**, *10*, 1291. <https://doi.org/10.3390/cells10061291>

Academic Editor: Marc Servant

Received: 23 March 2021

Accepted: 18 May 2021

Published: 23 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The recent SARS-CoV-2 pandemic has refocused attention to the *betacoronaviruses*, only eight years after the emergence of another zoonotic *betacoronavirus*, the Middle East respiratory syndrome coronavirus (MERS-CoV). While the wild source of SARS-CoV-2 may be disputed, for MERS-CoV, dromedaries are considered as source of zoonotic human infections. Testing 100 immune-response genes in 121 dromedaries from United Arab Emirates (UAE) for potential association with present MERS-CoV infection, we identified candidate genes with important functions in the adaptive, MHC-class I (*HLA-A-24*-like) and II (*HLA-DPB1*-like), and innate immune response (*PTPN4*, *MAGOHB*), and in cilia coating the respiratory tract (*DNAH7*). Some of these genes previously have been associated with viral replication in SARS-CoV-1/-2 in humans, others have an important role in the movement of bronchial cilia. These results suggest similar host genetic pathways associated with these *betacoronaviruses*, although further work is required to better understand the MERS-CoV disease dynamics in both dromedaries and humans.

Keywords: coronavirus; immune response genes; Old World camels; in-solution hybridization capture; zoonosis

1. Introduction

Emerging zoonotic diseases pose a serious threat not only to animal populations, but also to humans around the globe, as we experience with SARS-CoV-2 and the current COVID-19 pandemic [1]. A recent example of an emerging zoonotic pathogen in the family *Coronaviridae* is the Middle East respiratory syndrome coronavirus (MERS-CoV). It was first isolated in June 2012 from the sputum of a 60-year-old man from Saudi Arabia with acute pneumonia [2]. However, the April 2012 outbreak of acute respiratory illness in Jordan was retrospectively also diagnosed as MERS-CoV epidemic [3]. Similar to other emerging human coronaviruses, MERS-CoV is thought to have originated from bats; however, dromedaries (*Camelus dromedarius*) have been identified as reservoir hosts and the

primary source of human infections [4–7]. This *betacoronavirus*, similar to the severe acute respiratory syndrome coronaviruses SARS-CoV-1 and SARS-CoV-2, affects the respiratory tract in humans yielding generalized symptoms typical of acute respiratory viral infections. While the infection takes a mild course in dromedaries, ranging from asymptomatic to minor naso-ocular discharge [5,8,9], humans often suffer from a severe course of disease with a fatality rate of up to 35% [2,10].

MERS-CoV in dromedaries was retrospectively traced back to at least 1992, as specific antibodies were detected in dromedary blood samples collected from different African regions in that year [11]. The asymptomatic or mild course of the disease in camels suggests that MERS-CoV has never had a major impact on dromedaries. Meanwhile, MERS-CoV infection in dromedaries has been confirmed in more than 25 countries on all continents except Australia [7]. Recently, also Bactrian camels (*Camelus bactrianus*) and hybrids between dromedary and Bactrian camels have been identified as potential reservoirs of MERS-CoV [12]. New World camelids (alpacas (*Vicugna pacos*) and llamas (*Lama glama*)) are also susceptible to the virus [13].

Camelids (family Camelidae) are recognized not only as multipurpose livestock adapted to extreme environments, producing milk, meat and wool under harsh conditions, but also for their potential in combating infectious diseases. Camelids are unique among mammals in their ability to generate homodimeric immunoglobulins (Ig) in addition to conventional antibodies, which usually consist of two heavy and two light chains. In all New and Old World camel species, the antigen-binding fragment of these specific IgGs is reduced to a single variable domain lacking the light chain, which significantly reduces the size of the antibodies [14–16]. The so-called “nanobodies” can be used for human clinical applications by transporting therapeutic agents into different body parts, also crossing the blood–brain barrier [17]. Recent evidence has revealed that camelids can produce nanobodies that effectively neutralize *betacoronaviruses* [18,19] and block SARS-CoV-2 infection [20], which makes them promising candidates for antiviral (i.e., COVID-19) therapy.

Although much information is available on MERS-CoV prevalence, epidemiology, genetic diversity [21], and etiopathology from experimental infections of dromedaries [22], little is known about the immune responses of camels to this zoonotic pathogen and its underlying genetic basis. The genetic knowledge gap might be partially due to the fact that high quality chromosome-assembled genomes have only become available very recently [16,23,24]. Indeed, genomic approaches would be a helpful tool for immune studies of non-model animals [25]. However, only a few genome-wide analyses have been performed in Old World camels, and these have mainly addressed genomic diversity and selection during domestication [16,26]. Moreover, there is a dearth of research investigating genotype–phenotype associations, and recent studies were targeted to blood parameters [27] or production traits [28]. Immunogenetic studies in camelids have either characterized important candidate genes for the adaptive and innate immune response, such as the major histocompatibility complex [29,30], natural killer cells [31] and T-cell receptors [14], or identified genomic organization and diversity of immune response (IR) genes in few individuals [16,24]. Consequently, increased knowledge about the immune system in Old World camelids and its genetic basis will further improve our understanding of their role in spill-over of MERS-CoV to humans [22].

We decided to use a target enrichment approach with in-solution hybridization capture of 100 annotated immune genes [24] to genotype a larger number of dromedaries tested for a recent infection of MERS-CoV. Targeted enrichment has successfully been applied in sequencing the IR genes of gopher tortoises [32], in assessing variants in immune genes associated with Hepatitis B virus infection [33] and in identifying somatic alterations in follicular dendritic cell sarcoma in known cancer-associated genes [34].

In this study, we aimed to characterize the diversity in IR genes and to detect variants potentially associated with MERS-CoV infection in dromedaries. We sequenced 100 IR genes using targeted enrichment via in-solution hybridization capture in 121 dromedaries

collected in the field from three different sites in the United Arab Emirates (UAE). All dromedaries were assessed for MERS-CoV antibodies accounting for (past) infection by Enzyme-linked Immunosorbent Assay (ELISA) and Indirect Immunofluorescence Test (IIFT), as well as for the presence of the active virus (shedding) via molecular detection of viral nucleic acid using reverse transcription quantitative (RT-q) PCR. Most of the 121 dromedaries harbored MERS-CoV antibodies, which revealed that even young animals from the age of 2 months had already been exposed to the virus. Thus, we split all individuals positive for antibodies (seropositive) into two groups based on (i) virus presence (infection at the time of testing) and (ii) virus absence (no detectable infection at the time of testing). We also recorded the clinical status of the sampled dromedaries, however, due to the nonspecific signs of MERS-CoV, i.e., mild naso-ocular discharge, the clinical observations of the individuals were not convincing. As we are aware of the limitations related to a study conducted in the field (versus a controlled experiment), we accounted for (i) population structure, (ii) equal possibility of exposure to virus infection, and (iii) age and sex to reduce potential biases. Using phenotype–genotype association tests with univariate logistic regression, we identified candidate genes, which might be related to MERS-CoV infection in dromedaries. Although some of the gene variants associated to the presence of the disease have previously been related to SARS-CoV-1/-2 and other respiratory infectious diseases, further genomic and functional analyses will be necessary to broaden and corroborate our results. With our work, within all its limitations, we open doors for future novel research including large-scale screening for genes underlying defense mechanisms against an important zoonotic disease.

2. Materials and Methods

2.1. Ethics Statement

This study was approved by the Al Ain City Municipality and was part of ongoing public health surveillance in the UAE. The study was performed in accordance with the relevant laws and regulations that govern research in the UAE.

2.2. Sampling

The material for this study was collected during two field seasons (March/April 2019 and October 2019) from a total of 121 dromedaries, in three locations in the UAE: (1) the largest national livestock market (April 2019, $n = 37$; October 2019, $n = 39$); (2) a desert wildlife reserve (April 2019, $n = 30$); (3) a Bedouin family-owned farm (Al Mazrooei, Dubai, UAE) with camels primarily raised for racing and trading (March 2019, $n = 15$) (Table S1). Nasal swabs in RNA/DNA shield (ZymoResearch, Irvine, CA, USA) and serum samples of all dromedaries were collected and stored at $-80\text{ }^{\circ}\text{C}$ at the laboratory of the College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, UAE before shipment to the University of Veterinary Medicine Vienna, Austria. As backup, we also collected tail hair samples, which were stored in labeled paper envelopes. All dromedaries (aged ≥ 6 months) in the UAE have a subcutaneous identity microchip that is linked to a national database containing information on the camel's age, sex, and geographic origin within the UAE. All camels were scanned for these microchips and demographic data were extracted from the national database.

2.3. MERS-CoV Characterization

All nasal swab and serum samples were screened for MERS-CoV specific RNA. After thawing the samples were vortexed and centrifuged for 3 min at 6000 rpm. For automatic extraction with QIAcube (for 12 samples) or QIAcube HT (plate format device, both QIAGEN, Hilden, Germany) 140 or 200 μL , respectively, of each supernatant were taken. MERS CoV RT-qPCR in ORF1a gene region was performed using primers and probe as described previously [35]. However, this assay was adapted for qScript XLT One Step RT-qPCR ToughMix (QuantaBio, Beverly, MA, USA) on Applied Biosystems 7300 or 7500

Real-Time PCR systems (both Foster City, CA, USA). Samples with a Ct value equal or below 39.5 were considered positive.

To determine the presence of antibodies of immunoglobulin class IgG, IgA and IgM against MERS-CoV *in vitro*, serum samples were thawed, briefly centrifuged and subsequently screened using two camel specific serological assays: the Anti-MERS-CoV ELISA and the Anti-MERS-CoV IIFT (both EUROIMMUN, Lübeck, Germany) following the manufacturer's instructions. For the ELISA, samples with an extinction ratio higher than 1.1 in relation to the calibrator were considered positive, samples with an extinction ratio lower than 0.8 were considered negative, and samples within 0.8–1.1 were regarded as borderline. For the IIFT, antibody titers were determined according to the fluorescence of the different sample dilutions (1:10–1:1000). Here, the manufacturer's suggestions were adapted slightly to allow for a more conservative interpretation of the results (Table S2).

2.4. Camel DNA Extraction

DNA was extracted from a total of 82 nasal swabs, 11 hair and 28 blood samples with an improved salting-out method for high DNA yield [36] following a safety protocol (biosafety cabinet, FFP2 masks). DNA quantity was assessed by a spectrofluorometric assay using a fluorescence microplate reader (Twinkle; Berthold Technologies, Oak Ridge, TN, USA). Around 1 µg DNA from all 121 samples was sent to Daicel Arbor Biosciences (Ann Arbor, MI, USA) for library construction, hybridization capture and sequencing.

2.5. Probe Design and In-Solution Hybridization Capture Target Enrichment

We used a target enrichment approach based on in-solution hybridization with biotinylated RNA probes and selected 100 IR genes (Table S3) from the most up-to-date dromedary (CamDro3) annotations (<https://doi.org/10.5061/dryad.qv9s4mwb3> (accessed on 1 February 2021); [23]) for myBaits[®] design. The selected regions were provided to Daicel Arbor Biosciences (Ann Arbor, MI, USA) for bait design. A final total of 19,207 120-bp baits passed "Relaxed BLAST" analysis. For each 120-bp bait candidate, one BLAST hit against CamDro3 with the highest melting temperature was first discarded from the results (allowing for 1 hit in the genome), and only the top 500 hits (by bit score) were considered. Based on the distribution of remaining calculated melting temperatures, Daicel Arbor Biosciences filtered out nonspecific baits using the "Relaxed" (more nonspecific baits pass) criteria. Additional candidates were retained if they had at most 10 hits between 62.5 and 65 °C and 4 hits above 65 °C, and fewer than 2 passing baits on each flank.

Samples were sonicated and size selected following a protocol to produce an average insert length of approximately 300 bp. Up to 200 ng of sonicated and size selected DNA was taken into a library preparation method optimized for targeted capture. Unique dual-index combinations were added to each sample via 5–10 cycles of PCR amplification. The indexed libraries were quantified with both a spectrofluorimetric assay and a quantitative PCR assay. To prepare for capture, up to 80 ng of each library was pooled for capture (16- or 17-plex captures) and dried down to 7 µL by vacuum centrifugation. Captures were performed following the myBaits v4 protocol with an overnight hybridization. For each sample, half of the volume of beads in the elution buffer were amplified for 10 cycles. Final capture pools were quantified again with both a spectrofluorimetric and a quantitative PCR assay. Samples were sequenced on the Illumina NovaSeq 6000 platform on partial S4 flowcell lanes with 150-bp paired-end sequencing.

2.6. Variant Calling

We performed adapter and quality trimming using BBDuk v.38.75 (<https://sourceforge.net/projects/bbmap/> (accessed on 5 February 2020)), using "ref = resources/adapters.fa" that comes with BBMap/BBTools v. 38.75. For this, we selected the following settings: ktrim = r, k = 23, mink = 11, hdist = 1, tpe, tbo, qtrim = rl, trimq = 15. We then mapped quality and adapter trimmed reads to the CamDro3 ([23]; <https://doi.org/10.5061/dryad.qv9s4mwb3>) assembly using BBMap v. 38.75 (<https://sourceforge.net/projects/bbmap/>)

with the “usejni = t” setting. BAM files were cleaned, sorted, read groups added, and duplicates marked with Picard v. 2.21.7 (<http://broadinstitute.github.io/picard>). We called SNPs against CamDro3 [24] with CallVariants v. 38.39 (<https://sourceforge.net/projects/bbmap/>), keeping only SNPs with quality scores greater than or equal to 27 using the settings “ploidy =2 multisample minscore = 27.0 nopassdot = t duplicate = f minreadmapq = 30”. We then used BCFTools 1.9 (<http://samtools.github.io/bcftools/>) (accessed on 12 March 2019) to filter each individual’s raw VCF file to exclude sites with missing genotypes, kept only SNPs that passed “CallVariants”’s filters, and if a site was multiallelic, kept the genotype with the highest quality score. We also used BCFTools to merge VCF files for each individual into a single VCF file and finally employed BEDTools 2.29.0 [37] to keep only the SNPs that occurred in the target region where the 120-bp baits mapped using blastn v. 2.2.31+ [38].

2.7. Read-Based Imputation

We used STITCH v. 1.6.3 [39] to perform read-based imputation of SNPs. For this, first we selected scaffolds > 1 SNP and then extracted the positions of the SNPs. We performed adapter and quality trimming as before, and then we mapped the reads to CamDro3 with BBWrap v. 38.81 (<https://sourceforge.net/projects/bbmap/>) with “usejni = t” and “sam = 1.3” (output in the SAM 1.3 not 1.4 format for compatibility with STITCH 1.6.3). We finally ran STITCH 1.6.3 using the following number of ancestral haplotypes “k = {4, 6, 8, 10, 12, 14}” and the following number of generations ago “nGen = {100, 1000, 10,000, 100,000}” for each value of k and nGen, keeping only SNPs with INFO_SCORE > = 0.3. To determine the best combination of values of number of ancestral haplotypes and number of generations ago, we used Jvarkit’s “vcfcomparegt” version deaac59 [40] to compare non-imputed (CallVariants 38.75) and imputed (STITCH 1.6.3) genotypes from Drom1829 (which had the most non-imputed genotypes). We chose k=14 and nGen=100000 with best performance showing the highest number of genotypes that were the same between Drom 1829 imputed and Drom 1829 non-imputed sample (see Table S4 for full results). We filtered STITCH SNPs with k = 14 and nGen = 100,000 with VCFTools 0.1.15 with “max-missing 0.90”, “min-alleles 2”, “max-alleles 2” and to retain only SNPs with allele frequency < 1 (polymorphic sites). Our imputed dataset contained 3958 SNPs for 121 dromedary samples.

2.8. Data Filtering

Quality control of the data with $\leq 10\%$ missingness after read-based imputation was performed with PLINK 1.9 [41]. Relatedness was considered to detect samples with unexpectedly high value of identity by descent (IBS; i.e., >0.90) calculated in PLINK with the flags “-cluster” and “-matrix” to obtain the IBS matrix. We also filtered for a minor allele frequency (<1%) using the flag “-maf” and Hardy Weinberg equilibrium “-hwe” (p -value = 0.0000127, after FDR correction using “p.adjust” function on R 3.6.3 (R core team, 2019) based on the number of SNPs [42]).

2.9. Heterozygosity Associated in the Immune Response Genes

To assess the heterozygosity in the IR genes we used the GenomeTools v.1.5.8 [43] with the gff3 function and the ‘addintrons’ and ‘retainids’ options to predict intron regions from CamDro3 annotations [24]. We used BedTools 2.29.2 [37] to obtain only SNPs in exons, introns, or entire gene regions (it is possible for a SNP to be in both exon and intron regions of a gene, as transcript variants (isoforms) can differ by exon regions). We used Hierfstat 0.04-22 [44] with R 4.0.2 (R core team, 2020) to calculate observed (H_O) and expected heterozygosity (H_E) for each gene for exons, introns, and entire gene regions. We performed the same process to assess H_O/H_E in the total dataset, in positive and negative individuals, as well as in the different sampling sites. We assessed normality of residuals with the shapiro.test function and homogeneity of variance using the base R 4.0.2 (R core team, 2020) with the R package Car 3.0-10 [45]. We used a Welch t test (implemented in the base R “t.test” function) for testing H_O and H_E significance in genes, coding (exons)

and noncoding (introns) regions between cases and controls, without distinguishing gene groups. We used the base R `lm` and `ANOVA` functions to assess significance of main effects (H_O or $H_E \sim \text{Gene_Group}$) for gene, exon, and intron regions separately. If the `Gene_Group` was significant at the 0.05 level, we performed posthoc tests with Benjamini–Hochberg correction [42] with the R package `multcomp` version 1.4-15 [46].

2.10. Univariate Logistic Regression Analysis for Phenotype–Genotype Association

The association of SNPs (passing filtering criteria stated above) with the phenotype MERS-CoV-positive (case) and negative (control) was tested by univariate logistic regression analysis, accounting for sex, age and population structure. First, we included the first most informative PCs as covariates using PLINK 1.9 with the flags “`-pca`”. After, we used again PLINK 1.9 to perform the univariate logistic regression analysis by using “`-logistic`”, “`-covar`” and “`-adjust`”. Genomic inflation factor λ (lambda) was calculated in PLINK after applying logistic regressed p -values, and for values lower than 1, we calculated lambda on R (R core team). Graphical representations of Manhattan and Quantile-Quantile (QQ) plots were obtained with the R packages `qqman` v.0.1.4 [47] and `ggbio` v.1.36.0 [48]. We identified significant SNPs on a cut-off of $p < 0.05$ corrected for FDR [49]. Further SNPs located in genes with potential association with MERS-CoV infection were ranked by the lowest uncorrected significant p -values [50]. Gene names are based on functional annotations from Lado et al. ([25]; <https://doi.org/10.5061/dryad.qv9s4mwb3>), which we cross-referenced against GeneCards (<https://www.genecards.org/>). Finally, we used PLINK 1.9 to estimate allele frequencies and genotype counts, as well as to assess the significance differences, by using Fisher’s exact test with “`-fisher`” and “`-model`” “GENO” for allele frequencies and genotype counts between positives and negatives, respectively. We applied the “case-control for distinct traits” module in the Genetic Power Calculator ([51]; <https://zzz.bwh.harvard.edu/gpc/> (accessed on 12 May 2021)) and estimated the minimum required sample size to achieve adequate statistical power (80%; $\alpha = 0.05$; standard allelic test) for detecting evidence of an association in a candidate gene with significant minor allele frequency differences between cases and controls.

2.11. Linkage Disequilibrium (LD)-Based Gene-Set Test

We also performed a LD-based gene-set association analysis with PLINK v 1.9, using the SNPs in each of the 100 IR gene as a separate set. Set-based tests are particularly suited for large-scale candidate gene studies as opposed to whole genome association studies, as they can use permutation more efficiently. The empirical p -values were corrected for the multiple SNPs within a set (taking account of the LD between these SNPs). For this analysis we applied the default values of the standard r -squared (`-set-r2`) = 0.5, p -value (`-set-p`) = 0.05, max number SNPs (`-set-max`) = 5, and 10,000 permutations, representing a moderate setting of values.

3. Results

3.1. MERS-CoV Shedding and Antibody Prevalence in Dromedaries from the UAE

In this phenotype–genotype association study we investigated 121 dromedaries from three sites in the UAE: the largest national livestock market in the emirate of Abu Dhabi ($n = 76$), a desert wildlife reserve “Dubai Desert Conservation Reserve” ~60 km south-southeast of Dubai ($n = 30$), and a Bedouin owned camel farm ($n = 15$) ~70 km south of Dubai. Sex was equally represented within the samples (56 males, 57 females, 8 unknown sex) and the ages ranged from 2 months to almost 30 years (Table S1). We detected 107 and 117 seropositive dromedaries with MERS-CoV-specific Igs (IgG, IgA or IgM) by ELISA and IIFT, respectively, showing that most of these animals experienced a (past) MERS-CoV infection. Viral nucleic acids were detected by RT-qPCR in nasal swabs of 44 individuals (out of 76; 57.9%) from the livestock market exclusively (Table S1), including 18 females, 23 males and 3 unknown sex with an age between 2 months and 6 years. Almost half of the dromedaries ($n = 37$, 48.7%) were both RNA- and IgG-positive, which

shows animals were infected and were likely shedding virus due to new (re)infection or persistent infection, possibly with continuous or intermittent shedding. Three MERS-CoV RNA-positive dromedaries were negative by both serological assays indicating recent virus infection of these animals. Three further RNA-positive animals were ELISA negative (or borderline positive) but IIFT positive. Nasal swabs from dromedaries from the two other sites (wildlife reserve and camel farm), as well as all serum samples tested negative by the MERS-CoV specific RT-qPCR (Table S1).

3.2. In-Solution Hybridization Capture and Variant Calls in Dromedary IR Genes

For the 1,305,546 base target region composed of exons and introns of 100 IR genes, we generated 823,887,778 reads (121, 136, 372, 601 bases) passing filter of which 82.99% reads were not PCR or optical duplicates. Of the unique reads, 97.53% were successfully aligned to the most up-to-date dromedary (CamDro3) reference genome [23] (mean \pm standard deviation aligned reads per sample = $5,511,298 \pm 1,534,987$; minimum–maximum: 209,624–9,621,299). Mean coverage in the target region was generally high for each sample ($186 \times \pm 60 \times$; minimum–maximum: $6 \times$ – $320 \times$), resulting in the identification of 5768 raw SNPs in the target region. After variant filtering, filtering for genotype missingness (<25%) and removing non-polymorphic loci, we identified 760 SNPs. Due to the low number of SNPs in the target region after filtering we performed read-based imputation, which has successfully been used by animal and plant breeders [52]. Imputation enabled extending the set of SNPs identified to 5730 SNPs, however, upon filtering these markers with $\text{INFO_SCORE} \geq 0.3$ and removing non-polymorphic loci 3958 SNPs remained for further analyses.

We controlled for relatedness in the imputed dataset and no pair of individuals showed identity by descent higher than 0.88. Due to the high number of seropositive individuals, we continued the phenotype–genotype association analysis using univariate logistic regression including 101 dromedaries with antibodies confirmed by both ELISA and IIFT, after removing 14 borderline/negative samples for antibody prevalence, as well as three samples each with ambiguous virus infection results and missing age information. The seropositive dromedaries were split into two groups showing MERS-CoV presence (cases; $n = 36$) or absence (controls; $n = 65$). For the genotype data, we applied additional filtering steps to further reduce the possibility of capturing false positive variants and removed 13 SNPs out of Hardy–Weinberg equilibrium (HWE) exact test as well as 1003 SNPs with low minor allele frequencies of 1% or less ($\text{MAF} < 1\%$). The final dataset consisted of 2942 SNPs genotyped in 101 dromedaries including 54 females, 46 males and 1 unknown, grouped into 36 cases and 65 controls.

3.3. Diversity in the Targeted Immune Response (IR) Genes

To better understand the diversity of the 100 targeted IR genes, we organized them into functional groups, i.e., genes encoding MHC class I molecules, MHC class II molecules, toll-like receptors (TLR), granzymes, interleukins, genes expressed in natural killer (NK) cells (including natural killer cell complex (NKC) encoded killer cell lectin-like receptor (KLR) genes), and “other IR genes”. We estimated observed (H_O) and expected (H_E) heterozygosities in entire predicted genes, exons and introns separately. The average values calculated over all genes in the different IR gene groups ranged between 0.161–0.338 (H_O) and 0.193–0.343 (H_E), with the highest diversity (H_O) observed in entire killer cell genes, and the lowest in MHC class I genes (Table 1). The specific values estimated in the IR genes are provided in Table S3. ANOVA tests after posthoc correction with Benjamini–Hochberg (BH) only showed significant ($p < 0.05$) differences in the H_O between MHC class I and killer cell genes, while all other gene group comparisons were not significant (Figure 1, Table S5). Furthermore, no gene, intron or exon H_O or H_E differed significantly ($p < 0.05$) between MERS-CoV positive and negative individuals (Table S6).

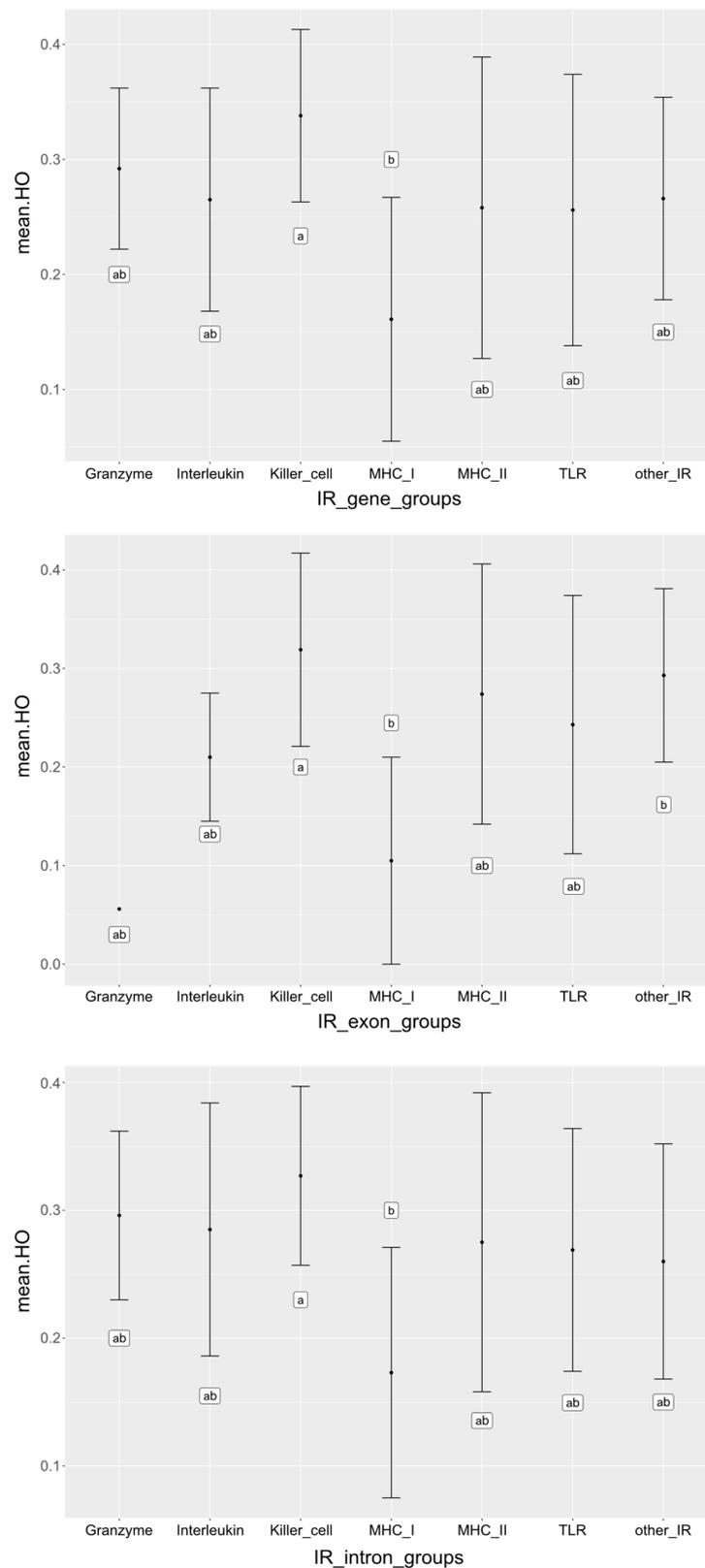


Figure 1. Observed heterozygosity (H_o) values of immune response gene groups. Means and standard deviations are shown for predicted genes, exon and introns separately. Results are only presented for gene, intron and exon H_o as only these showed significance for both ANOVA and posthoc correction with Benjamini–Hochberg (BH). Gene groups with different letters (‘a’ and ‘b’) indicate groups that had significantly different means whilst the same letters indicate nonsignificant different means.

Table 1. Average observed (H_O) and expected (H_E) heterozygosity in the captured immune response gene groups estimated over all genes, exons and introns.

IR Gene Group	Genes ^a			Exons ^a			Introns ^a		
	SNPs	H_O	H_E	SNPs	H_O	H_E	SNPs	H_O	H_E
Granzyme	54	0.292	0.311	5	0.056	0.066	50	0.296	0.315
Interleukin	121	0.265	0.284	21	0.210	0.218	107	0.285	0.305
Killer cells	282	0.338	0.342	38	0.319	0.313	253	0.327	0.336
MHC class I	435	0.161	0.193	96	0.155	0.185	353	0.173	0.205
MHC class II	596	0.258	0.277	140	0.274	0.290	456	0.275	0.294
TLR	124	0.256	0.298	55	0.243	0.272	74	0.269	0.311
Other IR genes	1253	0.266	0.280	235	0.293	0.307	1074	0.260	0.273

^a Genes, exons, or introns with fewer than two SNPs could not be included in calculations of averages.

3.4. Phenotype–Genotype Association in MERS-CoV Antibody-Positive Dromedaries

As our samples originated from three different locations, we corrected for population structure to avoid population stratification bias and possible false positive associations. The genetic variation in the population explained by the first six most informative principal components (PCs 1–6) summed up to 33%, and we included these as covariates in addition to sex and age (Figure 2). We performed a univariate logistic regression with the complete dataset of 2942 SNPs imputed over 100 IR genes from 101 dromedaries seropositive for MERS-CoV antibodies, including 36 virus shedders (cases) and 65 virus nonshedders (controls). The genomic inflation estimation lambda (based on median chi square) was lower than 1 ($\lambda = 0.82$). The quantile-quantile (QQ) plot (Figure 3a) with PCA correction showed that in general observed values followed the expected values, with an end tail characteristic of SNPs in potential association with the tested phenotypes (MERS-CoV presence or absence) (Figure 3b for Manhattan plot).

The selection of an appropriate statistical significance threshold in phenotype–genotype association studies is critical to differentiate true positives from false positives and false negatives. Therefore, we decided to present significant markers that were selected based on a cut-off of $p < 0.05$ corrected for a false discovery rate (FDR; [49]). In addition, we present the most significant SNPs ranked by the lowest uncorrected p -values. We detected 16 candidate SNPs (uncorrected $p < 0.01$), of which the top seven were significant using the FDR corrected $p < 0.0058$, as displayed in the Manhattan plot (Figure 3b) and Table 2. The seven top candidate SNPs were located within three genes on chromosomes (chr) 5, 20 and 34, respectively: Protein Tyrosine Phosphatase Non-Receptor Type 4 (*PTPN4*), which contained two intronic SNPs; an MHC class I human leukocyte-associated antigen (*HLA*) A-24-like sequence with one SNP potentially in an intron close to exon 4. However, due to an equivocal annotation of this locus in the CamDro3 reference genome, it is not clear whether it is a complete, potentially functional MHC class I sequence, and *Mago Homolog B* (*MAGOHB*), which harbored four intronic variants. The other SNPs were found in introns of the genes Dynein Axonemal Heavy Chain 7 (*DNAH7*; chr 17), Interleukin 10 Receptor Subunit Alpha (*IL10RA*; chr 33) and Coiled-Coil and C2 Domain Containing 2A (*CC2D2A*; chr 2). Among the potentially associated variants with slightly higher p -values (≤ 0.01548), we identified nine SNPs (seven in introns, one in exon 2 and another one in exon 4) located in the MHC class II gene *HLA-DPB1*-like (Table 2), which we thus consider as a strong candidate as well.

The genetic variation in the population explained by the first six most informative principal components (PCs 1–6) summed up to 33%, and we included these as covariates in addition to sex and age (Figure 2). We performed a univariate logistic regression with the complete dataset of 2942 SNPs imputed over 100 IR genes from 101 dromedaries seropositive for MERS-CoV antibodies, including 36 virus shedders (cases) and 65 virus nonshedders (controls). The genomic inflation estimation lambda (based on median chi square) was lower than 1 ($\lambda = 0.82$). The quantile-quantile (QQ) plot (Figure 3a) with PCA correction showed that in general observed values followed the expected values, with an end tail characteristic of SNPs in potential association with the tested phenotypes (MERS-CoV presence or absence) (Figure 3b for Manhattan plot).

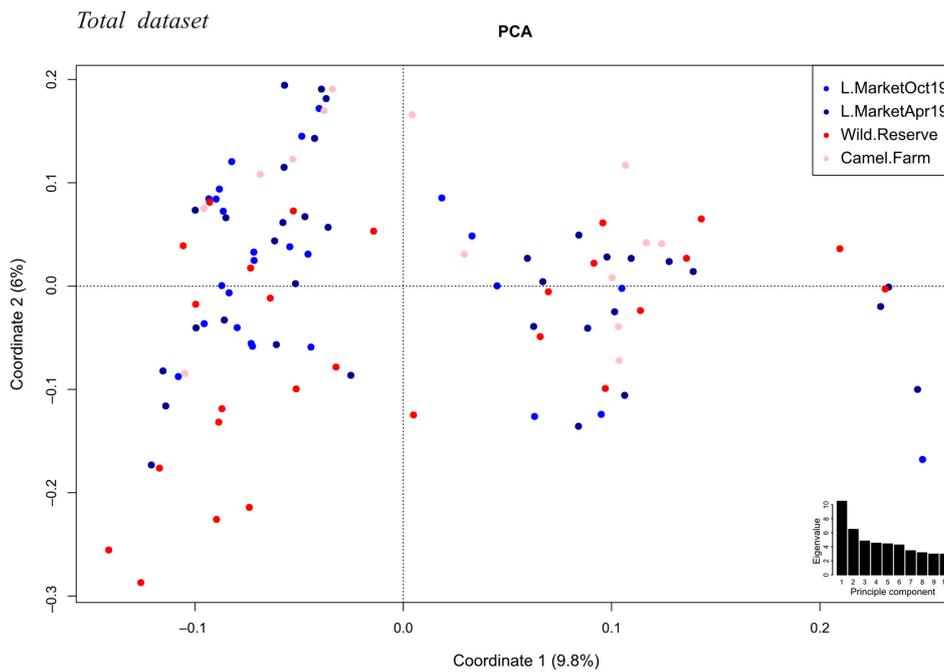


Figure 2. Principal component analysis of the population structure at three collection sites over two sampling periods. Variation explained by PC1 and PC2 are depicted in percentages. Individual animals are plotted on the first two principal components, colored by sampling site (livestock market (“L. Market”), dark and light blue, respectively); Dubai Desert Conservation Reserve (“Wild. Reserve”), dark red; a Bedouin camel farm (“Camel. Farm”), pink). The inset shows a bar plot of the eigenvalues for the first 10 principal components.

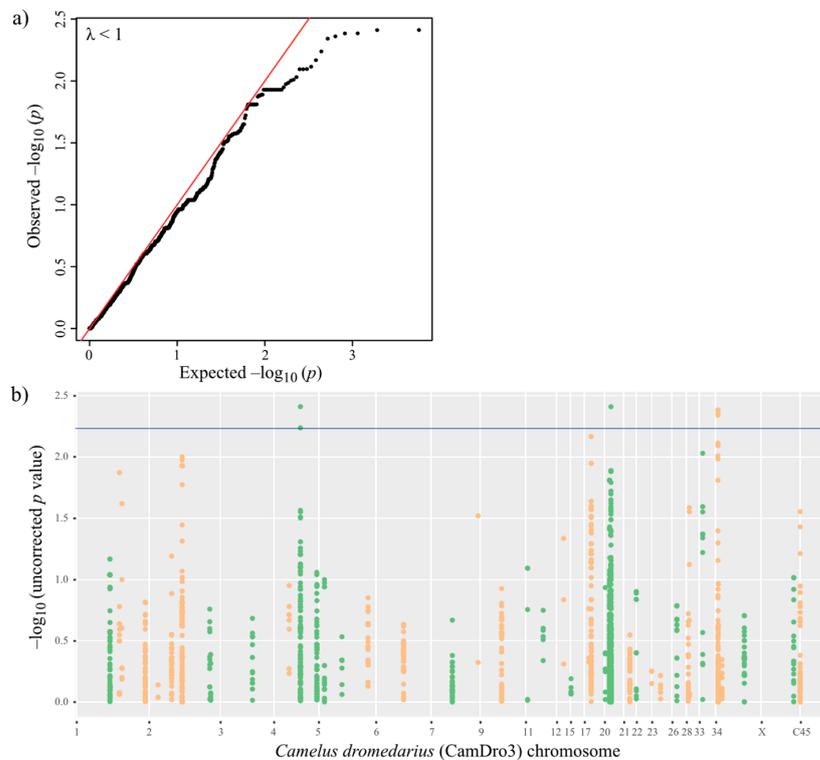


Figure 3. Univariate logistic regression results. (a) QQ plot and (b) Manhattan plot with FDR threshold depicted in blue. $-\log_{10}(p)$ values for SNPs alternate from green to orange to delineate chromosomes adjacent in the plots. C45 corresponds to Contig45, an unplaced scaffold in the CamDro3 reference.

The selection of an appropriate statistical significance threshold in phenotype–genotype association studies is critical to differentiate true positives from false positives and false negatives. Therefore, we decided to present significant markers that were selected based on a cut-off of $p < 0.05$ corrected for a false discovery rate (FDR; [49]). In addition, we present the most significant SNPs ranked by the lowest uncorrected p -values. We detected 16 candidate SNPs (uncorrected $p < 0.01$), of which the top seven were significant

Table 2. Significant SNPs located in candidate IR genes, allele frequencies of the minor alleles and genotype counts (homozygote minor/heterozygote/homozygote major allele) in MERS-CoV positive (cases) and negative (controls) camels.

Chr	Position (Minor/Major Allele)	Gene	Association Test <i>p</i> -Value	Allele Freq. Minor Allele				Genotype Counts		
				Cases	Controls	Exact <i>p</i> -Value	Odds Ratio	Cases	Controls	Exact <i>p</i> -Value
5	T8508361C intron 25	<i>PTPN4</i>	0.003873 **	0.08824	0.2	0.06096	0.3871	0/6/28	2/20/38	0.1241
20	A23100696G intron 4	<i>HLA-A-24-like</i>	0.003881 **	0.25	0.08475	0.00278 *	3.6	1/16/19	0/10/49	0.002554 *
34	G15362634A intron 1	<i>MAGOHB</i>	0.004123 **	0.3	0.4308	0.09343	0.5663	1/19/15	16/24/25	0.01201 *
34	A15363451G intron 2	<i>MAGOHB</i>	0.004129 **	0.2941	0.4308	0.06634	0.5506	1/18/15	16/24/25	0.01392 *
34	15367780	INTERGENIC	0.004364 **	-	-	-	-	-	-	-
34	G15361800A intron 1	<i>MAGOHB</i>	0.004553 **	0.3	0.4206	0.1239	0.5903	1/19/15	15/23/25	0.01467 *
5	C8506434T intron 26	<i>PTPN4</i>	0.005774 **	0.1	0.2097	0.07213	0.4188	0/7/28	2/22/38	0.1349
17	T23840747C intron 6	<i>DNAH7</i>	0.006791	0.1806	0.1154	0.2087	1.689	1/11/24	1/13/51	0.3252
34	C15363470T intron 2	<i>MAGOHB</i>	0.008005	0.3286	0.4385	0.173	0.6267	2/19/14	16/25/24	0.04715 *
20	A20676706C intron 4	<i>HLA-DPB1-like</i>	0.01548	0.04167	0.09231	0.2651	0.4275	0/3/33	2/8/55	0.6674
20	T20677126C exon 4	<i>HLA-DPB1-like</i>	0.01548	0.04167	0.09231	0.2651	0.4275	0/3/33	2/8/55	0.6674
20	G20678240T intron 2	<i>HLA-DPB1-like</i>	0.01548	0.04167	0.09231	0.2651	0.4275	0/3/33	2/8/55	0.6674
20	T20679052C exon 2	<i>HLA-DPB1-like</i>	0.01548	0.04167	0.09231	0.2651	0.4275	0/3/33	2/8/55	0.6674
20	A20679884C intron 1	<i>HLA-DPB1-like</i>	0.01548	0.04167	0.09231	0.2651	0.4275	0/3/33	2/8/55	0.6674
20	G20680467A intron 1	<i>HLA-DPB1-like</i>	0.01548	0.04167	0.09231	0.2651	0.4275	0/3/33	2/8/55	0.6674
20	T20680474C intron 1	<i>HLA-DPB1-like</i>	0.01548	0.04167	0.09231	0.2651	0.4275	0/3/33	2/8/55	0.6674
20	T20680741C intron 1	<i>HLA-DPB1-like</i>	0.01548	0.04167	0.08594	0.3868	0.4625	0/3/33	2/7/55	0.6561
20	G20681619C intron 1	<i>HLA-DPB1-like</i>	0.01548	0.04167	0.09231	0.2651	0.4275	0/3/33	2/8/55	0.6674

* significant $p < 0.05$; ** significant after FDR correction.

Calculating allele frequencies for the top seven candidate SNPs, we detected significantly ($p = 0.0028$) higher frequencies of the minor allele (chr20:23100696) in the *HLA-A-24-like* sequence in MERS-CoV positive dromedaries (Table 2). We estimated the required sample size to detect evidence for an association with a power of 80% in this candidate locus at a minimum of 15 samples. In addition, the homozygote genotype counts for the minor allele in *MAGOHB* were significantly higher ($p < 0.05$) in MERS-CoV negative camels than in positive ones (Table 2). To further reduce the potential selection of false positive markers our next step was to test the robustness of these results by accounting for spatial and temporal sampling.

3.5. Accounting for Spatial and Temporal Sampling

We attempted to account for the limitation that dromedaries sampled from the three different locations might not have had equal exposure possibilities to MERS-CoV. Although MERS-CoV antibodies were present in all camels included in the association analysis, indicating they had contact with the virus, we only detected active virus infection (shedding) in camels sampled at the livestock market. In this largest national livestock market, dromedaries from all over the UAE are sold, and this cohort was less structured in terms of their exposure to infection. The other two sampling locations included a wildlife reserve where the camels do not have regular contact with other dromedaries and a Bedouin owned camel farm, where the animals are relatively isolated. As such, we repeated the univariate logistic regression analysis only including 60 samples from the livestock market (35 females, 24 males, 1 unknown), split into 36 cases (MERS-CoV positive) and 24 controls (MERS-CoV negative), and genotyped for 2917 SNPs (after filtering for HWE and MAF as before). As the dromedaries had been transported to the livestock market from all over the UAE, we corrected for potential population structure and included the first five most informative

PCs (33% of the variation; Figure S1a) as a covariate in the univariate logistic regression analysis, along with sex and age. The genomic inflation estimation lambda (λ) was 0.78, and the QQ plot with PCA correction showed that observed values followed expected ones, with an end tail as observed in GWAS studies (Figure S2b). The most significant SNPs (Figure S2b) were located in the same four genes as detected in the initial test including complete dataset (101 individuals; 2942 SNPs).

Similarly, we accounted for the fact that we had two field seasons (spring and autumn 2019) and analyzed only those 75 samples collected in spring 2019 (36 females, 38 males, 1 unknown), divided into 22 cases and 53 controls. The univariate logistic regression with 3016 variants and PCs 1–6 (34.4%, Figure S1b) included as covariates resulted in a genomic inflation estimation lambda again lower than 1 ($\lambda = 0.77$) and a similar QQ plot (Figure S2c). Within the top most significant SNPs we identified three (*PTPN4*, *HLA-A-24-like*, *DNAH7*) out of the four previously detected genes (Figure S2c).

Finally, to better confirm the results from the imputed data set, we repeated the univariate logistic regression analysis with the initially called 760 SNPs after filtering for 25% of genotyping missingness (without imputation). Due to higher genotype missingness, we applied less stringent filtering for the relatedness threshold (0.95), but similar HWE (0.0000658) and MAF (<1%) thresholds as before, as well as taking into account population structure (first four PCs explaining 36% of the total variation). After filtering, 696 variants and 98 samples (35 cases and 63 controls) were included. With a lambda lower than 1 ($\lambda = 0.66$), *PTPN4* and *HLA-A-24-like* harbored the top SNPs (uncorrected $p < 0.02$) (Figure S2d).

3.6. Linkage Disequilibrium-Based Gene-Set Test

To further test the robustness of our results, we applied a complementary approach by means of a gene-set association test [41] using the complete dataset (101 individuals; 2942 SNPs). Overall, the gene-set results were similar to the univariate logistic regression SNP tests. From the 100 targeted IR genes, 20 had significant SNPs (uncorrected $p < 0.05$), including the genes *PTPN4*, *DNAH7*, *HLA-A-24-like*, *HLA-DPB1-like*, and *MAGOHB* (Table S7). *HLA-A-24-like* and *MAGOHB* were nominally significant ($p = 0.031$ and $p = 0.008$) and harbored 112 and 12 variants, of which 14 and 5 SNPs were significant, respectively. However, only two SNPs in *HLA-A-24-like* (chr20:23100696 | 23100503) and one in *MAGOHB* (chr34:15362634) passed the independent significance r-squared based threshold of 0.5. While these genes showed a stronger signal for potential genotype–phenotype association, none of the other genes were nominally significant (Table S7).

4. Discussion

It has long been established that a combination of population growth, biodiversity loss and land-use change drives the emergence and spread of zoonotic diseases [53]. The emergence of MERS-CoV over the past decade is no exception, being the likely outcome from such combined factors. As the consumption of camel milk and meat is increasing and camel products gain access to wider markets, the impact of camel-associated zoonotic diseases on public health and economics will also grow with advancing urbanization in African and Arabian countries. In this study we attend to this important zoonosis and target the immune response to MERS-CoV infection in a representative dromedary sample from the UAE.

4.1. High MERS-CoV Antibody Prevalence in Dromedaries from the UAE

In the course of this study, the assessment for the presence of MERS-CoV antibodies revealed a high prevalence (88%) of seropositive individuals ($n = 107$) within the 121 investigated dromedaries from the UAE. Additionally, another 94 dromedaries were screened as part of a public health surveillance and all showed seropositivity (N. Nowotny, personal communication). While sex was equally distributed over seropositive individuals in our study, age was as young as 2 months with an average of 6 years (Table S1). Previous studies

have shown that seropositivity is higher among adult dromedaries (2 years and older), indicating that the likelihood of exposure and subsequent infection increases with age [54]. Camels generally shed MERS-CoV for about 7 days and viral RNA is detectable with RT-qPCR up to 35 days post-infection. Virus-specific antibodies can be identified from 3 weeks after infection onwards, with anti-MERS-CoV IgMs present for at least 4 weeks (indicating a more recent infection), while anti-MERS-CoV IgGs can stay for many years [55]. High seroprevalence coupled with known instances of camel–human transmission provides a proxy for prospective epidemiological risks [56], as a human case study in Saudi Arabia with higher-than-expected prevalence of MERS-CoV seropositivity in dromedaries demonstrated [57]. Naturally infected dromedary camels shed virus from the upper respiratory tract, evidenced by the presence of RNA in the RT-qPCR from the nasal swabs we collected. No current MERS-CoV positives were detected among the Bedouin family-owned dromedaries or the wildlife reserve (Table S1), compared to the livestock market where animals from different regions in the UAE are mixed in medium-sized (~50–100 sqm) pens.

4.2. Different Diversity in IR Gene Groups

We observed that MHC class I mean diversity (H_O) was significantly lower compared to killer cell genes over all dromedaries. Low levels of genetic diversity in the MHC region have been also observed in wild and domestic two-humped camels [29]. Interestingly, a lower overall genomic heterozygosity was described in dromedaries compared to wild and domestic Bactrian camels [26], which could hint to a generally lower genetic diversity in dromedaries. However, recent genome-wide analyses of IR genes found that the mean nucleotide diversity in MHC class I and II genes in dromedaries and domestic Bactrian camels seemed to be higher compared to other adaptive or innate IR genes, as well as the rest-of-genome genes, at least for the MHC genes studied [24].

4.3. Candidate IR Genes Associated with MERS-CoV Infection in Dromedaries

Since 2002, three betacoronaviruses, i.e., SARS-CoV-1, MERS-CoV and the most recent SARS-CoV-2 emerged as human pathogens through possible zoonotic spill-over from animals, all associated with severe human respiratory infections. Both candidate gene and genome-wide sequencing approaches have offered relevant insights into the genetic basis of these zoonotic diseases [58]. Unlike for SARS-CoV-1 and -2, few studies have been conducted on host genetic variation underlying susceptibility to MERS-CoV, its pathogenesis, transmission, and mortality in humans [59,60]. In this field study, we identified candidate genes on four different chromosomes potentially associated to MERS-CoV infection in dromedaries from the UAE, *PTPN4*, *DNAH7*, MHC class-I (*HLA-A-24*-like sequence), MHC class II (*HLA-DPB1*-like) and *MAGOHB*. The SNPs significantly associated with the presence of MERS-CoV in seropositive camels were mainly distributed in intronic regions (Table 2) except for the MHC class II gene *HLA-DPB1*-like, where we found one SNP in exon 2 and another in exon 4 (Table 2). Exon 2 encodes the antigen-binding groove of the class II molecule and, therefore, its polymorphism is of functional importance. Exon 4 codes for the transmembrane domain that controls membrane domain partitioning and class II structure, both of which influence antigen presentation and T-cell activation [61]. In recent years, an important role of intronic polymorphisms has been established, either filling regulatory functions upstream of exons or being in linkage with other (exonic) variants (e.g., [62]). The *HLA-A-24*-like sequence harboring the significant chr20:23100696 and further variants (Table 2 and Figure S2) is incompletely annotated as fully functional classical MHC class I gene, with no exon 2 sequence, in the dromedary reference genome (CamDro3) as well as in the next closely related and chromosome-assembled genome of the wild camel (*Camelus ferus*; [16]). Therefore, we cannot exclude the possibility that we sequenced a pseudogene or a misassembled chimeric sequence for this locus. Although the (functional) impact of the identified candidate genes on MERS-CoV infection in dromedaries has yet to be determined, previous research associated those genes with viral replication in SARS-CoV-1/-2, MERS-CoV and other viruses causing

respiratory infectious diseases in humans, and with pathways involved in the movement of bronchial cilia.

PTPN4. *PTPN4* belongs to the protein tyrosine phosphatase (PTP) family and has an important role in the innate immune system. For instance, it inhibits the Toll-like receptor (TLR) 4 signaling pathway that triggers many immune proteins including proinflammatory cytokines and type I interferons [63]. Interestingly, while *PTPN4* acts as an inhibitor, the TLR4 signaling pathway is activated by the SARS-CoV-2 spike protein [64], and mice lacking TLR4 had more severe SARS-CoV infections than wild-type mice [65]. Meanwhile, the spike protein of MERS-CoV triggers the expression of negative regulators of the TLR signaling pathways [65]. Understanding the TLR signaling pathways in the context of MERS-CoV infection also in dromedaries would be an important contribution to mitigate the viral infection. *PTPN4* is related to predicted target functions of human micro(mi)RNAs that bind to the single-stranded (ss)-RNA such as SARS-CoV-2, and possibly to its spike protein gene. These predicted miRNA targets might destabilize the ss-RNA translation of SARS-CoV-2 in respiratory epithelial cells, which could explain successful antiviral defense [66]. As both polymorphisms identified in the UAE dromedaries are located in intronic regions (introns 25 and 26) of *PTPN4*, they might have regulatory functions that can influence the expression of the gene [62].

DNAH7. *DNAH7* encodes a force-generating protein that is an essential component of the inner dynein arm of axonemes in cilia coating the respiratory tract, which drive mucus along airway surfaces providing a critical defense mechanism of the pulmonary system [67]. It represents one of the most downregulated genes following SARS-CoV-2 infection of human bronchial epithelial cells in vitro [68]. *DNAH7* expression levels were also significantly downregulated in human bronchial epithelial cells infected with MERS-CoV and influenza A (H1N1), which induce apoptosis in these cells [69,70]. To recognize if similar mechanisms act in dromedaries with MERS-CoV infection, specific gene expression studies including camel bronchial epithelial cells would be necessary.

HLA-A-24-like (MHC class I A-24). The MHC (classes I, II, and III) has a dense clustering of immune relevant genes that can show extreme polymorphisms due to their main task of encoding cell surface proteins involved in antigen presentation [71]. Thus, HLA polymorphisms have been linked to susceptibility and pathogenesis of numerous infectious diseases including those caused by RNA viruses, especially SARS, influenza, AIDS, rabies, and West Nile fever [59]. For example, a protective effect of *HLA-A*02:01* against SARS-CoV-1 has been suggested in Asian patients [72,73], while *HLA-A*24:02* has been associated with COVID-19 susceptibility [74]. Large meta-analyses of allele frequency distributions in human traits showed that SNPs connected to disease susceptibility are generally skewed towards a higher minor allele frequency (>20%) [75]. In our study, a significantly higher (25%; $p = 0.0028$) frequency was also observed for the minor allele of the *HLA-A-24-like* chr20:23100696 in MERS-CoV positive dromedaries (Table 2). However, as this result concerns only a single SNP within a sequence of unclear status, several different interpretations are possible, including a false positive statistical artifact, an isolated finding due to an inaccurate annotation and assembly of the sequence and/or an effect of linkage with a causative MHC SNP variant. Therefore, our observation that UAE dromedaries with higher MAF within the *HLA-A-24-like* sequence might be more susceptible to MERS-CoV infection needs to be corroborated by additional (long-read) sequencing and haplotype analysis, including samples from other Arabian and African populations.

HLA-DPB1-like (MHC class II DPB1). Associations observed for two MHC class II SNPs located in exons 2 and 4 support the idea of an MHC effect on the phenotypes analyzed. Although antigen presentation of SARS-CoV-1 mainly depends on MHC class I molecules [76], class II genes can also contribute to *Betacoronaviridae* antigen presentation as suggested by the association of *HLA-DRB1*11:01* and *HLA-DQB1*02:02* alleles with susceptibility to MERS-CoV [60]. In COVID-19 patients from Italy, the allele frequency distributions for *HLA-DRB1*15:01* and *DQB1*06:02* showed significant correlations of the minor allele with higher susceptibility to the disease, while *DRB1*07:01* on the contrary

was negatively associated [77]. Interestingly, dromedaries with no current MERS-CoV infection were more often homozygote for the minor allele of *HLA-DPB1*-like, which is a paralog of *HLA-DRB1* in humans. The associations observed may, however, result from the effect of linkage with other MHC sequences.

MAGOHB. *MAGOHB* belongs to the mago nashi gene family and is required for pre-mRNA splicing. In macrophages—one of the effector cells of the innate immune system—the expression of *MAGOHB* increased rapidly after lipopolysaccharide (LPS) stimulation [78]. LPS is a natural adjuvant, which is synthesized by Gram-negative bacteria, and stimulates cells through TLR4 signaling pathway, causing the release of inflammatory cytokines and the upregulation of costimulatory molecules on antigen presenting cells [79]. Interestingly, *MAGOHB* is targeted by has-miR-20a-5p, one of six miRNAs that previously have been reported to be antiviral in respiratory diseases, and were found to be downregulated in lung tissues during viral infection [80,81]. From a network analysis, has-miR-20a-5p was identified among 38 miRNAs targeting host genes that interact with SARS-CoV-2 proteins [82]. The homozygote alternative (minor) genotype of *MAGOHB* was significantly ($p < 0.005$) more frequent in our control group, which might hint to a higher resistance to MERS-CoV infection in these dromedaries. It is also possible that *MAGOHB* represents positional markers in linkage with some genes of the natural killer complex (NKC). The dromedary NKC harbors, besides tested KLR genes, a number of other C-type lectin-like (CLEC) receptor genes downstream of *MAGOHB* [31] that were not included in the hybridization capture of IR genes. CLECs are expressed by myeloid cells and serve to monitor their environment and sense danger. In principle, they recognize a vast repertoire of (non-)glycan ligands from pathogens or modulate activity of cells. Dendritic cell CLECs, for example, by recognition and internalization of ligands start the process of antigen presentation to T cells and generation of an immune response [83].

In summary, we identified important candidate genes related to the innate and adaptive immune system in dromedaries from the UAE. The functional importance of these genes in response to MERS-CoV infection in dromedaries, similar to humans, needs to be investigated in more controlled in vitro and in vivo experiments.

4.4. Challenges and Impact of IR Gene Associations with Betacoronavirus Infections

Biological interpretations of statistical significance in association (field) studies have several limitations. While we did our best to select an equal distribution of age and sex and to account for potential population structure in the dromedaries sampled from three locations in the UAE, we cannot exclude that some of the associations presented here are false positives. Due to the nature of a field study, we cannot guarantee that all individuals had equally been exposed to MERS-CoV, although we included only dromedaries with antibodies indicating that they have had contact with the virus at some time in the past. We attempted to control for this fact by repeating the univariate logistic genotype–phenotype regression analysis only with dromedaries from the largest livestock market located in the emirate of Abu Dhabi, where animals from all over the UAE are traded. Our different tests accounting for spatial and temporal sampling showed in general good coherence of the results in the top selected candidate genes. The genetic power calculation for the required sample size to detect “true” association with a power of 80% in a candidate gene with significant allele frequency differences between case and controls resulted in a minimum number of 15 samples, which was reached throughout all subsampling strategies (e.g., 22 cases in the spring season samples, Figure S1b). While the analyzed virus assessment reflects presence/absence of the virus in seropositive camels, the nonrandom distribution of SNPs observed between actively infected/noninfected individuals indicates that they are genetically different, which consequently demands further investigations especially in terms of their immune mechanisms. Future case–case control studies need to include more dromedary populations from different African and Arabian countries and, though challenging, potential birth cohort studies or experimental MERS-CoV inoculation of dromedaries and other livestock herded together. This would help define predisposed

groups and support screening efforts for potential virus reservoirs. The next important step will be to investigate expression and functional pathways of the identified candidate IR genes to select for higher resistance to MERS-CoV. Finally, the innate and adaptive IR genes identified in dromedaries show high resemblances with human immune response to the zoonotic SARS-CoV-1, SARS-CoV-2 and MERS-CoV. Thus, understanding the underlying mechanisms to disease susceptibility/resistance in dromedaries and other animals will result in more effective strategies to combat *betacoronaviral* disease in human populations as well.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/cells10061291/s1>, Figure S1: Principal component analysis of the population structure at three collection sites over two sampling periods. Variation explained by PC1 and PC2 is depicted in percentages. Individual animals are plotted on the first two principal components, colored by sampling site (livestock market (“L.Market”) over two sampling periods (April and October 2019, dark and light blue, respectively); Dubai Desert Conservation Reserve (“Wild.Reserve”), dark red; a Bedouin camel farm (“Camel.Farm”), pink). The inset shows a barplot of the eigenvalues for the first 10 principal components. (a) Only livestock market samples; (b) only spring field season; (c) non-imputed dataset, Figure S2: Manhattan and QQ plot. Highlighted in bold are the four genes that are common in all analyses (*HLA-A-like*, *PTPN4*, *MAGOHB* and *DNAH7*). FDR corrected thresholds are represented in blue. (a) Total dataset; (b) only livestock market samples; (c) only spring field samples; (d) non-imputed dataset. C45 corresponds to Contig45, an unplaced scaffold, Table S1: Sample information and assessment of virus presence (swabs) and antibody prevalence (sera), Table S2: Scheme showing how IIFT antibody titers were determined according to the fluorescence of the different sample dilutions, Table S3. Observed (H_O) and expected (H_E) heterozygosity values depicted in immune response gene groups. Identified candidate genes *MAGOHB*, *HLA-A-24-like*, *HLA-DPB1-like*, *DNAH7* and *PTPN4* are highlighted in bold, Table S4. Read-based imputation performance. nAncestralHaplotypes (k) = number of ancestral haplotypes; nGen = number of generations ago, controls recombination rate; nDiff_from_non-imputed = number of genotypes that were not the same between imputed and non-imputed samples; nMatch_from_non-imputed = number of genotypes that were the same between imputed and non-imputed sample; nMissing_from_non-imputed = these are SNPs and hence genotypes that are missing because that SNP failed QC for imputation; nAdditional_SNPs_with_called_genotypes_from_non-imputed = these are genotypes newly added by imputation), Table S5: Statistical analysis of observed heterozygosity (H_O) for immune response gene groups in genes, exons and introns. Means and standard deviations are shown for genes, exon and introns separately. Results are only presented for gene, intron and exon H_O as only these showed significance for both ANOVA and posthoc correction with Benjamini–Hochberg (BH). Gene groups with different letters (‘a’ and ‘b’) indicate groups that had significantly different means whilst the same letters indicate nonsignificant different means, Table S6: Observed (H_O) and expected (H_E) heterozygosity in genes, exons and introns in MERS-CoV positive ($n = 36$) and negative ($n = 65$) individuals. p -values of mean differences were calculated with Welch t test, Table S7: Linkage disequilibrium-based haplotype (gene-set) test showing 20 genes with significant SNPs at $p < 0.05$. Identified candidate genes *MAGOHB*, *HLA-A-24-like*, *HLA-DPB1-like*, *DNAH7* and *PTPN4* are highlighted in bold. *HLA-A-24-like* and *MAGOHB* were nominally significant ($p < 0.05$) indicated with an asterisk. NSNP—number of SNPs in set; NSIG—total number of SNPs below p -value threshold; ISIG—number of significant SNPs also passing LD-criterion; STAT—average test statistic based on ISIG SNPs; EMP1—empirical set-based p -value; SNPs—positions of SNPs in the set.

Author Contributions: S.L. wrote the first draft of the manuscript. S.L., J.P.E. and M.P. analyzed the data. S.L., P.A.B., D.A.K., N.N. and T.L. collected samples at the livestock market, UAE. T.L. assessed the camel chip demographic information. S.L., J.V.C., P.W. and J.K. performed wet lab work. P.O.-t. and J.F. provided analytic methods and discussion on the data. P.A.B., P.H. and N.N. conceived and managed the project. S.L. and P.A.B. wrote the manuscript and all authors provided valuable discussions and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by research grants from the College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates (to N.N., grant No. MBRU-CM-RG2018-14; and T.L. grant No. MBRU-CM-RG2019-13). Next-Generation target

sequencing performed by Daicel Arbor Biosciences was funded by the Austrian Science Fund (FWF) project P29623-B25 to P.B., where S.L. and J.P.E. also acknowledge funding.

Institutional Review Board Statement: This study was approved by the Al Ain City Municipality and was part of ongoing public health surveillance in the UAE under the research projects reviewed by the College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates MBRU-CM-RG2018-14 (to N.N.) and MBRU-CM-RG2019-13 (to T.L.). The study was performed in accordance with the relevant laws and regulations that govern research in the UAE.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw FASTQ files were deposited on the European Nucleotide Archive (ENA) (ERS5621787 (SAMEA7874536)–ERS5621907 (SAMEA7874656)). VCF file, target region and bait sequences were deposited on dryad together with the scripts file (https://datadryad.org/stash/share/OfmXVPdiw1tj0LxeMZ_b7gky8ty3F0BOqfnGwz1qf1I (accessed on 27 January 2021)). Additional material requests can be addressed to Pamela.Burger@vetmeduni.ac.at.

Acknowledgments: The authors thank H.E. Matar Mohammed Saif Al Nuaimi (General Manager of Al Ain City Municipality) and his team including Babiker Mohammed Osman for supporting the study; Hashim Ahmed Saeed and Mohammed Helal Ahmed for assistance with sampling at the livestock market; Moayyed Sher Shah, Tamer Khafaga, Greg Simkins and the Dubai Desert Conservation Reserve staff for access and support during our sampling of the reserve; staff of the camel tour providers Al Maha, Arabian Adventures, Desert Star, Alpha Tours, and Travco Tours; Brigitte Howarth, Zayed University, Dubai, for her support; Hessa Mazrooei and her family for their support and generosity during the sampling on their farm; Nadine Wolf for veterinary assistance. We are grateful for the help of Noushad Karuvantevida, Athiq Ahmed Wahab and Abubakkar Babuhan in facilitating the study. Open Access Funding by the Austrian Science Fund (FWF).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Costagliola, A.; Liguori, G.; D'angelo, D.; Costa, C.; Ciani, F.; Giordano, A. Do animals play a role in the transmission of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)? a commentary. *Animals* **2021**, *11*, 16. [CrossRef] [PubMed]
2. Zaki, A.M.; Van Boheemen, S.; Bestebroer, T.M.; Osterhaus, A.D.M.E.; Fouchier, R.A.M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **2012**, *367*, 1814–1820. [CrossRef] [PubMed]
3. Hijawi, B.; Abdallat, M.; Sayaydeh, A.; Alqasrawi, S.; Haddadin, A.; Jaarour, N.; Alsheikh, S.; Alsanouri, T. Novel coronavirus infections in Jordan, April 2012: Epidemiological findings from a retrospective investigation. *East. Mediterr. Health J.* **2013**, *19* (Suppl. 1), 12–18. [CrossRef]
4. Reusken, C.B.E.M.; Haagmans, B.L.; Müller, M.A.; Gutierrez, C.; Godeke, G.J.; Meyer, B.; Muth, D.; Raj, V.S.; De Vries, L.S.; Corman, V.M.; et al. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: A comparative serological study. *Lancet Infect. Dis.* **2013**, *13*, 859–866. [CrossRef]
5. Nowotny, N.; Kolodziejek, J. Middle East Respiratory Syndrome coronavirus (MERS-CoV) in dromedary camels, Oman, 2013. *Eurosurveillance* **2014**, *19*, 1–5. [CrossRef]
6. Dawson, P.; Malik, M.R.; Parvez, F.; Morse, S.S. What Have We Learned about Middle East Respiratory Syndrome Coronavirus Emergence in Humans? A Systematic Literature Review. *Vector Borne Zoonotic Dis.* **2019**, *19*, 174–192. [CrossRef]
7. Gossner, C.; Danielson, N.; Gervelmeyer, A.; Berthe, F.; Faye, B.; Kaasik Aaslav, K.; Adlhoch, C.; Zeller, H.; Penttinen, P.; Coulombier, D. Human-Dromedary Camel Interactions and the Risk of Acquiring Zoonotic Middle East Respiratory Syndrome Coronavirus Infection. *Zoonoses Public Health* **2016**, *63*, 1–9. [CrossRef]
8. Omrani, A.S.; Shalhoub, S. Middle East respiratory syndrome coronavirus (MERS-CoV): What lessons can we learn? *J. Hosp. Infect.* **2015**, *91*, 188–196. [CrossRef] [PubMed]
9. Hemida, M.G.; Elmoslemany, A.; Al-Hizab, F.; Alnaeem, A.; Almathen, F.; Faye, B.; Chu, D.K.W.; Perera, R.A.P.M.; Peiris, M. Dromedary Camels and the Transmission of Middle East Respiratory Syndrome Coronavirus (MERS-CoV). *Transbound. Emerg. Dis.* **2017**, *64*, 344–353. [CrossRef]
10. Assiri, A.; McGeer, A.; Perl, T.M.; Price, C.S.; Al Rabeeah, A.A.; Cummings, D.A.T.; Alabdullatif, Z.N.; Assad, M.; Almulhim, A.; Makhdoom, H.; et al. Hospital outbreak of middle east respiratory syndrome coronavirus. *N. Engl. J. Med.* **2013**, *369*, 407–416. [CrossRef]
11. Corman, V.M.; Jores, J.; Meyer, B.; Younan, M.; Liljander, A.; Said, M.Y.; Gluecks, I.; Lattwein, E.; Bosch, B.J.; Drexler, J.F.; et al. Antibodies against MERS coronavirus in dromedary camels, Kenya, 1992–2013. *Emerg. Infect. Dis.* **2014**, *20*, 1319–1322. [CrossRef]
12. Lau, S.K.P.; Li, K.S.M.; Luk, H.K.H.; He, Z.; Teng, J.L.L.; Yuen, K.-Y.; Wernery, U.; Woo, P.C.Y. Middle East Respiratory Syndrome Coronavirus Antibodies in Bactrian and Hybrid Camels from Dubai. *mSphere* **2020**, *5*, e00898-19. [CrossRef]

13. Reusken, C.B.E.M.; Schilp, C.; Raj, V.S.; De Bruin, E.; Kohl, R.H.G.; Farag, E.A.B.A.; Haagmans, B.L.; Al-Romaihi, H.; Le Grange, F.; Bosch, B.J.; et al. MERS-CoV infection of alpaca in a region where MERS-CoV is endemic. *Emerg. Infect. Dis.* **2016**, *22*, 1129–1131. [[CrossRef](#)]
14. Ciccarese, S.; Burger, P.A.; Ciani, E.; Castelli, V.; Linguiti, G.; Plasil, M.; Massari, S.; Horin, P.; Antonacci, R. The camel adaptive immune receptors repertoire as a singular example of structural and functional genomics. *Front. Genet.* **2019**, *10*, 997. [[CrossRef](#)]
15. Muyltermans, S.; Baral, T.N.; Retamozzo, V.C.; De Baetselier, P.; De Genst, E.; Kinne, J.; Leonhardt, H.; Magez, S.; Nguyen, V.K.; Revets, H.; et al. Camelid immunoglobulins and nanobody technology. *Vet. Immunol. Immunopathol.* **2009**, *128*, 178–183. [[CrossRef](#)] [[PubMed](#)]
16. Ming, L.; Wang, Z.; Yi, L.; Batmunkh, M.; Liu, T.; Siren, D.; He, J.; Jurant, N.; Jambl, T.; Li, Y.; et al. Chromosome-level assembly of wild Bactrian camel genome reveals organization of immune gene loci. *Mol. Ecol. Resour.* **2020**, *20*, 770–780. [[CrossRef](#)] [[PubMed](#)]
17. Jovčevska, I.; Muyltermans, S. The Therapeutic Potential of Nanobodies. *BioDrugs* **2020**, *34*, 11–26. [[CrossRef](#)]
18. Hanke, L.; Perez, L.V.D.J.S.; Das, H.; Schulte, T.; Morro, A.M.; Corcoran, M.; Achour, A.; Hedestam, G.K.; Hällberg, B.M.; Murrell, B.; et al. An alpaca nanobody neutralizes SARS-CoV-2 by blocking receptor interaction. *Nat. Commun.* **2020**, *11*, 4420. [[CrossRef](#)] [[PubMed](#)]
19. Wrapp, D.; De Vlieger, D.; Corbett, K.S.; Torres, G.M.; Wang, N.; Van Breedam, W.; Roose, K.; van Schie, L.; Hoffmann, M.; Pöhlmann, S.; et al. Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies. *Cell* **2020**, *181*, 1004–1015.e15. [[CrossRef](#)]
20. Koenig, P.-A.; Das, H.; Liu, H.; Kümmerer, B.M.; Gohr, F.N.; Jenster, L.-M.; Schiffelers, L.D.J.; Tesfamariam, Y.M.; Uchima, M.; Wuerth, J.D.; et al. Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape. *Science* **2021**, *371*, eabe6230. [[CrossRef](#)]
21. Chu, D.K.W.; Hui, K.P.Y.; Perera, R.A.P.M.; Miguel, E.; Niemeyer, D.; Zhao, J.; Channappanavar, R.; Dudas, G.; Oladipo, J.O.; Traoré, A.; et al. MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3144–3149. [[CrossRef](#)] [[PubMed](#)]
22. Adney, D.R.; Clancy, C.S.; Bowen, R.A.; Munster, V.J. Camelid Inoculation with Middle East Respiratory Syndrome Coronavirus: Experimental Models of Reservoir Host Infection. *Viruses* **2020**, *12*, 1370. [[CrossRef](#)] [[PubMed](#)]
23. Elbers, J.P.; Rogers, M.F.; Perelman, P.L.; Proskuryakova, A.A.; Serdyukova, N.A.; Johnson, W.E.; Horin, P.; Corander, J.; Murphy, D.; Burger, P.A. Improving Illumina assemblies with Hi-C and long reads: An example with the North African dromedary. *Mol. Ecol. Resour.* **2019**, *19*, 1015–1026. [[CrossRef](#)] [[PubMed](#)]
24. Lado, S.; Elbers, J.P.; Rogers, M.F.; Melo-Ferreira, J.; Yadamsuren, A.; Corander, J.; Horin, P.; Burger, P.A. Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies. *BMC Genom.* **2020**, *21*, 606. [[CrossRef](#)]
25. Pedersen, A.B.; Babayan, S.A. Wild immunology. *Mol. Ecol.* **2011**, *20*, 872–880. [[CrossRef](#)]
26. Fitak, R.R.; Mohandesan, E.; Corander, J.; Yadamsuren, A.; Chuluunbat, B.; Abdelhadi, O.; Raziq, A.; Nagy, P.; Walzer, C.; Faye, B.; et al. Genomic signatures of domestication in Old World. *Commun. Biol.* **2020**, *3*, 316. [[CrossRef](#)]
27. Guo, F.; Ming, L.; Si, R.; Yi, L.; He, J.; Ji, R. A genome-wide association study identifies quantitative trait loci affecting hematological traits in camelus bactrianus. *Animals* **2020**, *10*, 96. [[CrossRef](#)] [[PubMed](#)]
28. Bahbahani, H.; Musa, H.H.; Wragg, D.; Shuiep, E.S.; Almathen, F.; Hanotte, O. Genome Diversity and Signatures of Selection for Production and Performance Traits in Dromedary Camels. *Front. Genet.* **2019**, *10*, 893. [[CrossRef](#)]
29. Plasil, M.; Mohandesan, E.; Fitak, R.R.; Musilova, P.; Kubickova, S.; Burger, P.A.; Horin, P. The major histocompatibility complex in Old World camels and low polymorphism of its class II genes. *BMC Genom.* **2016**, *17*, 117. [[CrossRef](#)]
30. Plasil, M.; Wijkmark, S.; Elbers, J.P.; Oppelt, J.; Burger, P.A.; Horin, P. The major histocompatibility complex of Old World camels: Class I and class I-related genes. *Hla* **2019**, *93*, 203–215. [[CrossRef](#)]
31. Futas, J.; Oppelt, J.; Jelinek, A.; Elbers, J.P.; Wijacki, J.; Knoll, A.; Burger, P.A.; Horin, P. Natural killer cell receptor genes in camels: Another mammalian model. *Front. Genet.* **2019**, *10*, 620. [[CrossRef](#)] [[PubMed](#)]
32. Elbers, J.P.; Brown, M.B.; Taylor, S.S. Identifying genome-wide immune gene variation underlying infectious disease in wildlife populations—A next generation sequencing approach in the gopher tortoise. *BMC Genom.* **2018**, *19*, 64. [[CrossRef](#)]
33. Yu, F.; Zhang, X.; Tian, S.; Geng, L.; Xu, W.; Ma, N.; Wang, M.; Jia, Y.; Liu, X.; Ma, J.; et al. Comprehensive investigation of cytokine- and immune-related gene variants in HBV-associated hepatocellular carcinoma patients. *Biosci. Rep.* **2017**, *37*. [[CrossRef](#)] [[PubMed](#)]
34. Griffin, G.K.; Sholl, L.M.; Lindeman, N.I.; Fletcher, C.D.M.; Hornick, J.L. Targeted genomic sequencing of follicular dendritic cell sarcoma reveals recurrent alterations in NF- κ B regulatory genes. *Mod. Pathol.* **2016**, *29*, 67–74. [[CrossRef](#)] [[PubMed](#)]
35. Corman, V.M.; Müller, M.A.; Costabel, U.; Timm, J.; Binger, T.; Meyer, B.; Kreher, P.; Lattwein, E.; Eschbach-Bludau, M.; Nitsche, A.; et al. Assays for laboratory confirmation of novel human coronavirus (HCoV-EMC) infections. *Eurosurveillance* **2012**, *17*, 1–9. [[CrossRef](#)] [[PubMed](#)]
36. De Volo, S.B.; Reynolds, R.T.; Douglas, M.R.; Antolin, M.F. An improved extraction method to increase DNA yield from molted feathers. *Condor* **2008**, *110*, 762–766. [[CrossRef](#)]
37. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)] [[PubMed](#)]

38. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
39. Davies, R.W.; Flint, J.; Myers, S.; Mott, R. Rapid genotype imputation from sequence without reference panels. *Nat. Genet.* **2016**, *48*, 965–969. [[CrossRef](#)] [[PubMed](#)]
40. Lindenbaum, P. Jvarkit: Java-based utilities for Bioinformatics. *FigShare* **2015**, *10*, m9.
41. Purcell, S.; Neale, B.; Todd-brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)]
42. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [[CrossRef](#)]
43. Gremme, G.; Steinbiss, S.; Kurtz, S. Genome tools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2013**, *10*, 645–656. [[CrossRef](#)]
44. Goudet, J. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **2005**, *5*, 184–186. [[CrossRef](#)]
45. Fox, J.; Weisberg, S. *An R Companion to Applied Regression*, 3rd ed.; Sage Publications Inc.: New York, NY, USA, 2019.
46. Hothorn, T.; Bretz, F.; Westfall, P. Simultaneous Inference in General Parametric Models. *Biom. J.* **2008**, *50*, 346–363. [[CrossRef](#)]
47. Turner, S.D. qqman: An R package for visualizing GWAS results using QQ and manhattan plots. *J. Open Source Softw.* **2018**, *3*, 731. [[CrossRef](#)]
48. Yin, T.; Cook, D.; Lawrence, M. ggbio: An R package for extending the grammar of graphics for genomic data. *Genome Biol.* **2012**, *13*, R77. [[CrossRef](#)] [[PubMed](#)]
49. Benjamini, Y.; Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **2001**, *29*, 1165–1188. [[CrossRef](#)]
50. Narum, S.R. Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conserv. Genet.* **2006**, *7*, 783–787. [[CrossRef](#)]
51. Purcell, S.; Cherny, S.S.; Sham, P.C. Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **2003**, *19*, 149–150. [[CrossRef](#)]
52. VanRaden, P.M.; Sun, C.; O’Connell, J.R. Fast imputation using medium or low-coverage sequence data. *BMC Genet.* **2015**, *16*, 82. [[CrossRef](#)]
53. Engering, A.; Hogerwerf, L.; Slingenbergh, J. Pathogen-host-environment interplay and disease emergence. *Emerg. Microbes Infect.* **2013**, *2*, 1–7. [[CrossRef](#)]
54. Alagaili, A.N.; Briese, T.; Mishra, N.; Kapoor, V.; Sameroff, S.C.; de Wit, E.; Munster, V.J.; Hensley, L.E.; Zalmout, I.S.; Kapoor, A.; et al. Middle east respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. *MBio* **2014**, *5*. [[CrossRef](#)]
55. Adney, D.R.; van Doremalen, N.; Brown, V.R.; Bushmaker, T.; Scott, D.; de Wit, E.; Bowen, R.A.; Munster, V.J. Replication and shedding of MERS-CoV in upper respiratory tract of inoculated dromedary camels. *Emerg. Infect. Dis.* **2014**, *20*, 1999–2005. [[CrossRef](#)] [[PubMed](#)]
56. Zhu, S.; Zimmerman, D.; Deem, S.L. A Review of Zoonotic Pathogens of Dromedary Camels. *Ecohealth* **2019**, *16*, 356–377. [[CrossRef](#)] [[PubMed](#)]
57. Mohd, H.A.; Al-Tawfiq, J.A.; Memish, Z.A. Middle East Respiratory Syndrome Coronavirus (MERS-CoV) origin and animal reservoir. *Virol. J.* **2016**, *13*, 1–7. [[CrossRef](#)] [[PubMed](#)]
58. Shelton, J.F.; Shastri, A.J.; Ye, C.; Weldon, C.H.; Filshtein, T.; Coker, D.; Symons, A.; Esparza-gordillo, J.; The 23andMe COVID-19 Team; Aslibekyan, S.; et al. Trans-ethnic analysis reveals genetic and non-genetic associations with COVID-19 susceptibility and severity. *medRxiv* **2020**. [[CrossRef](#)]
59. Ovsyannikova, I.G.; Haralambieva, I.H.; Crooke, S.N.; Poland, G.A.; Kennedy, R.B. The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. *Immunol. Rev.* **2020**, *296*, 205–219. [[CrossRef](#)]
60. Hajeer, A.H.; Balkhy, H.; Johani, S.; Yousef, M.Z.; Arabi, Y. Association of human leukocyte antigen class II alleles with severe Middle East respiratory syndrome-coronavirus infection. *Ann. Thorac. Med.* **2016**, *11*, 211–213. [[CrossRef](#)]
61. Harton, J.; Jin, L.; Hahn, A.; Drake, J. Immunological Functions of the Membrane Proximal Region of MHC Class II Molecules. *F1000Research* **2016**, *5*. [[CrossRef](#)] [[PubMed](#)]
62. Cooper, D.N. Guest Editorial Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. *Hum. Genom.* **2010**, *4*, 284–288. [[CrossRef](#)]
63. Huai, W.; Song, H.; Wang, L.; Li, B.; Zhao, J.; Han, L.; Gao, C.; Jiang, G.; Zhang, L.; Zhao, W. Phosphatase PTPN4 Preferentially Inhibits TRIF-Dependent TLR4 Pathway by Dephosphorylating TRAM. *J. Immunol.* **2015**, *194*, 4458–4465. [[CrossRef](#)] [[PubMed](#)]
64. Brandao, S.C.S.; Ramos, J.d.O.X.; Dompieri, L.T.; Godoi, E.T.A.M.; Figueiredo, J.L.; Sarinho, E.S.C.; Chelvanambi, S.; Aikawa, M. Is Toll-like receptor 4 involved in the severity of COVID-19 pathology in patients with cardiometabolic comorbidities? *Cytokine Growth Factor Rev.* **2020**, *5*, 135–143. [[CrossRef](#)]
65. Mubarak, A.; Alturaiki, W.; Hemida, M.G. Middle east respiratory syndrome coronavirus (MERS-CoV): Infection, immunological response, and vaccine development. *J. Immunol. Res.* **2019**, *2019*, 6491738. [[CrossRef](#)] [[PubMed](#)]
66. Haddad, H.; Al-Zyoud, W. miRNA target prediction might explain the reduced transmission of SARS-CoV-2 in Jordan, Middle East. *Non Coding RNA Res.* **2020**, *5*, 135–143. [[CrossRef](#)] [[PubMed](#)]

67. Zhang, Y.J.; O'Neal, W.K.; Randell, S.H.; Blackburn, K.; Moyer, M.B.; Boucher, R.C.; Ostrowski, L.E. Identification of dynein heavy chain 7 as an inner arm component of human cilia that is synthesized but not assembled in a case of primary ciliary dyskinesia. *J. Biol. Chem.* **2002**, *277*, 17906–17915. [[CrossRef](#)]
68. Nunnari, G.; Sanfilippo, C.; Castrogiovanni, P.; Imbesi, R.; Li Volti, G.; Barbagallo, I.; Musumeci, G.; Di Rosa, M. Network perturbation analysis in human bronchial epithelial cells following SARS-CoV2 infection. *Exp. Cell Res.* **2020**, *395*, 112204. [[CrossRef](#)]
69. Tao, X.; Hill, T.E.; Morimoto, C.; Peters, C.J.; Ksiazek, T.G.; Tseng, C.-T.K. Bilateral Entry and Release of Middle East Respiratory Syndrome Coronavirus Induces Profound Apoptosis of Human Bronchial Epithelial Cells. *J. Virol.* **2013**, *87*, 9953–9958. [[CrossRef](#)]
70. Yuen, K.M.; Chan, R.W.; Mok, C.K.; Wong, A.C.; Kang, S.S.; Nicholls, J.M.; Chan, M.C. Differential onset of apoptosis in avian influenza H5N1 and seasonal H1N1 virus infected human bronchial and alveolar epithelial cells: An in vitro and ex vivo study. *Influenza Other Respir. Viruses* **2011**, *5*, 437–438.
71. Dendrou, C.A.; Petersen, J.; Rossjohn, J.; Fugger, L. HLA variation and disease. *Nat. Rev. Immunol.* **2018**, *18*, 325–339. [[CrossRef](#)]
72. Wang, S.F.; Chen, K.H.; Chen, M.; Li, W.Y.; Chen, Y.J.; Tsao, C.H.; Yen, M.Y.; Huang, J.C.; Chen, Y.M.A. Human-Leukocyte Antigen Class I Cw 1502 and Class II DR 0301 Genotypes Are Associated with Resistance to Severe Acute Respiratory Syndrome (SARS) Infection Sheng-Fan. *Viral Immunol.* **2011**, *24*, 421–426. [[CrossRef](#)] [[PubMed](#)]
73. Li, X.; Geng, M.; Peng, Y.; Meng, L.; Lu, S. Molecular immune pathogenesis and diagnosis of COVID-19. *J. Pharm. Anal.* **2020**, *10*, 102–108. [[CrossRef](#)] [[PubMed](#)]
74. Warren, R.L.; Birol, I. HLA Predictions from the Bronchoalveolar Lavage Fluid Samples of Five Patients at the Early Stage of the Wuhan Seafood Market COVID-19 Outbreak. *Bioinformatics.* **2020**, *36*, 5271–5273. [[CrossRef](#)]
75. Park, J.H.; Gail, M.H.; Weinberg, C.R.; Carroll, R.J.; Chung, C.C.; Wang, Z.; Chanock, S.J.; Fraumeni, J.F.; Chatterjee, N. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18026–18031. [[CrossRef](#)]
76. Liu, J.; Wu, P.; Gao, F.; Qi, J.; Kawana-Tachikawa, A.; Xie, J.; Vavricka, C.J.; Iwamoto, A.; Li, T.; Gao, G.F. Novel Immunodominant Peptide Presentation Strategy: A Featured HLA-A*2402-Restricted Cytotoxic T-Lymphocyte Epitope Stabilized by Intrachain Hydrogen Bonds from Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein. *J. Virol.* **2010**, *84*, 11849–11857. [[CrossRef](#)] [[PubMed](#)]
77. Novelli, A.; Andreani, M.; Biancolella, M.; Liberatoscioli, L.; Passarelli, C.; Colona, V.L.; Rogliani, P.; Leonardis, F.; Campana, A.; Carsetti, R.; et al. HLA allele frequencies and susceptibility to COVID-19 in a group of 99 Italian patients. *Hla* **2020**, *96*, 610–614. [[CrossRef](#)] [[PubMed](#)]
78. Singh, K.K.; Wachsmuth, L.; Kulozik, A.E.; Gehring, N.H. Two mammalian MAGOH genes contribute to exon junction complex composition and nonsense-mediated decay. *RNA Biol.* **2013**, *10*, 1291–1298. [[CrossRef](#)] [[PubMed](#)]
79. McAleer, J.P.; Vella, A.T. Understanding how lipopolysaccharide impacts CD4 T-cell immunity. *Crit. Rev. Immunol.* **2008**, *28*, 281–299.
80. Zhu, Z.; Qi, Y.; Ge, A.; Zhu, Y.; Xu, K.; Ji, H.; Shi, Z.; Cui, L.; Zhou, M. Comprehensive characterization of serum microRNA profile in response to the emerging avian influenza A (H7N9) virus infection in humans. *Viruses* **2014**, *6*, 1525–1539. [[CrossRef](#)]
81. Leon-Icaza, S.A.; Zeng, M.; Rosas-Taraco, A.G. microRNAs in viral acute respiratory infections: Immune regulation, biomarkers, therapy, and vaccines. *ExRNA* **2019**, *1*, 1. [[CrossRef](#)]
82. Sardar, R.; Satish, D.; Gupta, D. Identification of Novel SARS-CoV-2 Drug Targets by Host MicroRNAs and Transcription Factors Co-regulatory Interaction Network Analysis. *Front. Genet.* **2020**, *11*, 1–9. [[CrossRef](#)] [[PubMed](#)]
83. Hooper, J.K.; Eggink, L.L.; Cote, R. Stories from the Dendritic Cell Guardhouse. *Front. Immunol.* **2019**, *10*, 2880. [[CrossRef](#)] [[PubMed](#)]

Supplementary Materials for
Innate and adaptive immune genes associated with MERS-CoV infection in dromedaries

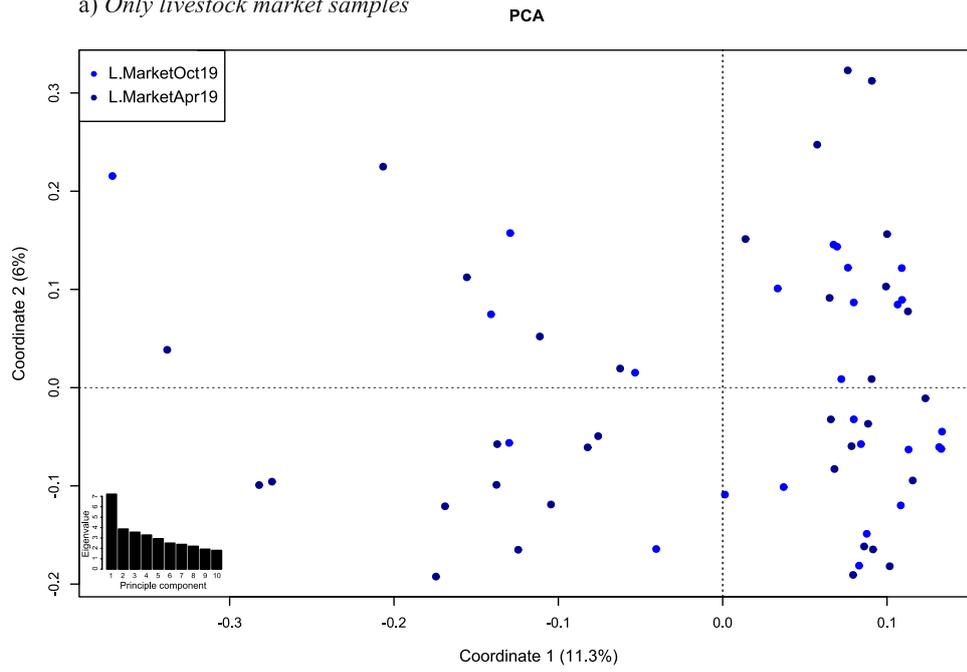
Sara Lado, Jean P. Elbers, Martin Plasil, Tom Loney, Pia Weidinger, Jeremy V. Camp,
Jolanta Kolodziejek, Jan Futas, Dafalla O. Kannan, Pablo Orozco-terWengel, Petr Horin,
Norbert Nowotny, Pamela A. Burger*

*Corresponding author. pamela.burger@vetmeduni.ac.at

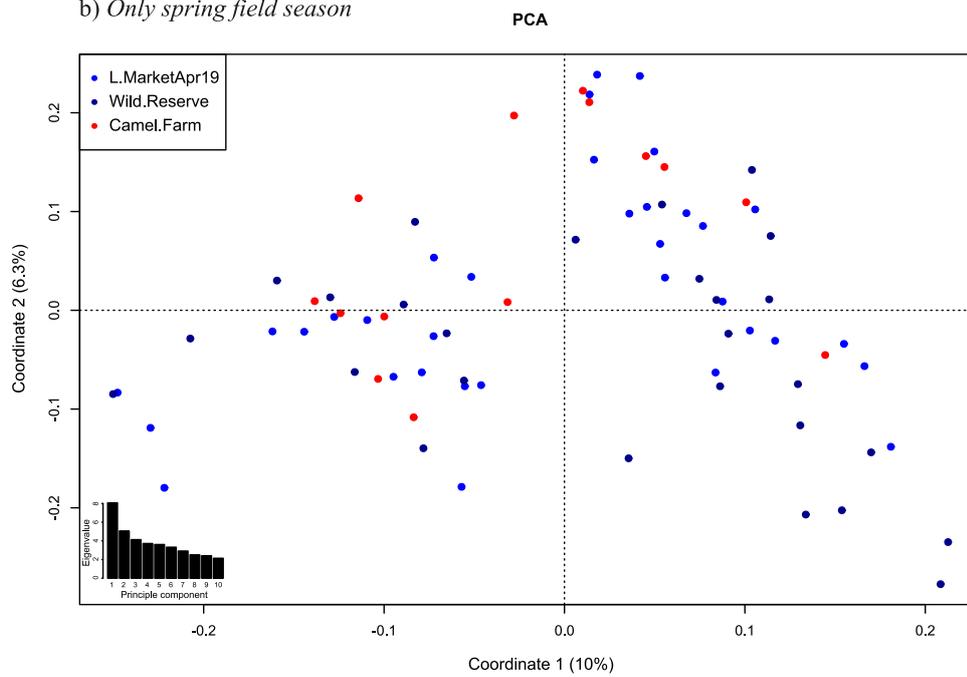
This PDF file includes:

Figures S1 and S2
Tables S1 to S7

a) *Only livestock market samples*



b) *Only spring field season*



c) *Non-imputed dataset*

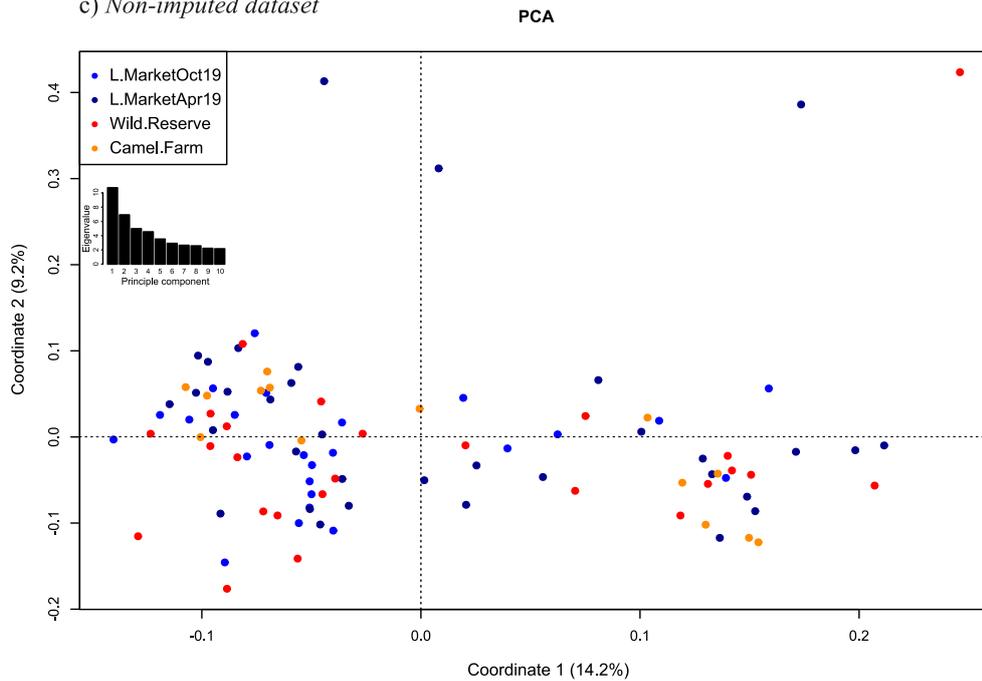
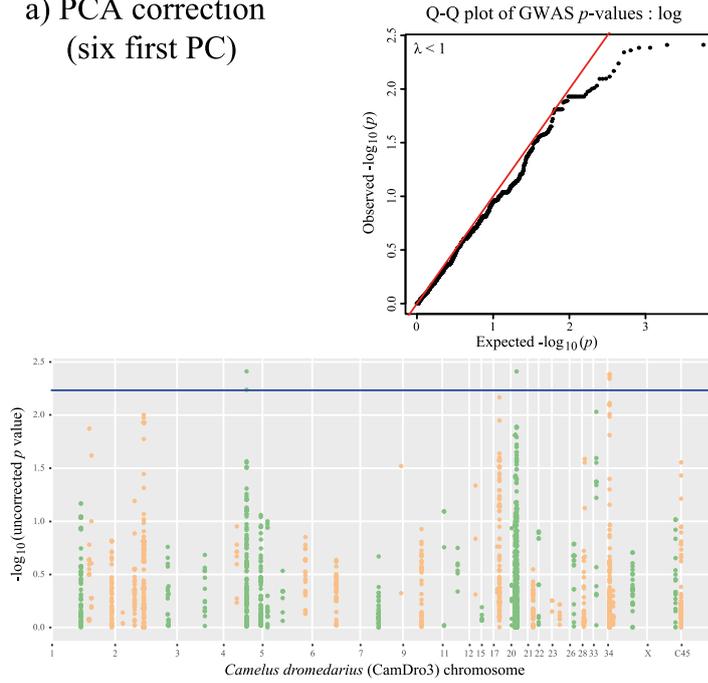


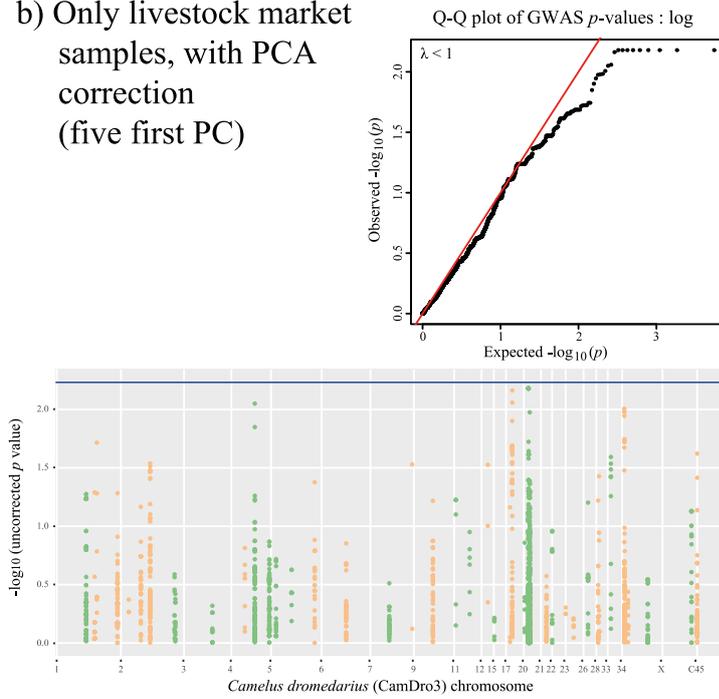
Figure S1. Principal Component Analysis of the population structure at three collection sites over two sampling periods. Variation explained by PC1 and PC2 are depicted in percentages. Individual animals are plotted on the first two principal components, colored by sampling site (livestock market [“L. Market”], over two sampling periods (April and October 2019, dark and light blue, respectively); Dubai Desert Conservation Reserve [“Wild. Reserve”], dark red; and a Bedouin camel farm [“Camel. Farm”], pink). The inset shows a barplot of the eigenvalues for the first 10 principal components. a) Only livestock market samples; b) Only spring field season: c) Non-imputed dataset.

a) PCA correction
(six first PC)



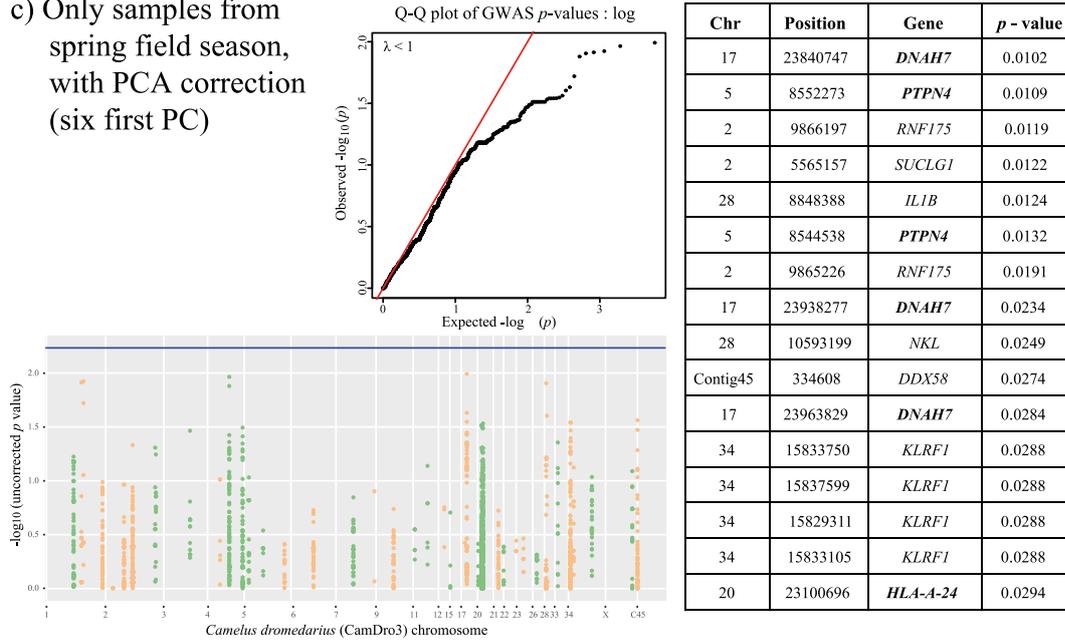
Chr	Position	Gene	p - value
5	8508361	<i>PTPN4</i>	0.0039
20	23100696	<i>HLA-A-24</i>	0.0039
34	15362634	<i>MAGOHB</i>	0.0041
34	15363451	<i>MAGOHB</i>	0.0041
34	15367780	INTERGENIC	0.0044
34	15361800	<i>MAGOHB</i>	0.0046
5	8506434	<i>PTPN4</i>	0.0058
17	23840747	<i>DNAH7</i>	0.0068
34	15371299	INTERGENIC	0.0077
34	15363470	<i>MAGOHB</i>	0.0080
34	15369030	INTERGENIC	0.0080
34	15371264	INTERGENIC	0.0080
33	12210072	<i>IL10RA</i>	0.0093
34	15370956	INTERGENIC	0.0098
2	113136710	<i>CC2D2A</i>	0.0099
34	15371123	INTERGENIC	0.0104

b) Only livestock market
samples, with PCA
correction
(five first PC)



Chr	Position	Gene	p - value
20	20676706	<i>HLA-DPB1</i>	0.0066
20	20677126	<i>HLA-DPB1</i>	0.0066
20	20678240	<i>HLA-DPB1</i>	0.0066
20	20679052	<i>HLA-DPB1</i>	0.0066
20	20679884	<i>HLA-DPB1</i>	0.0066
20	20680467	<i>HLA-DPB1</i>	0.0066
20	20680474	<i>HLA-DPB1</i>	0.0066
20	20680741	<i>HLA-DPB1</i>	0.0066
20	20681619	<i>HLA-DPB1</i>	0.0066
17	23963829	<i>DNAH7</i>	0.0069
17	23948208	<i>DNAH7</i>	0.0087
5	8508361	<i>PTPN4</i>	0.0089
34	15363451	<i>MAGOHB</i>	0.0098
34	15362634	<i>MAGOHB</i>	0.0104
34	15367780	INTERGENIC	0.0105
20	23100696	<i>HLA-A-24</i>	0.0106

c) Only samples from spring field season, with PCA correction (six first PC)



d) Non-imputed dataset (first four PC)

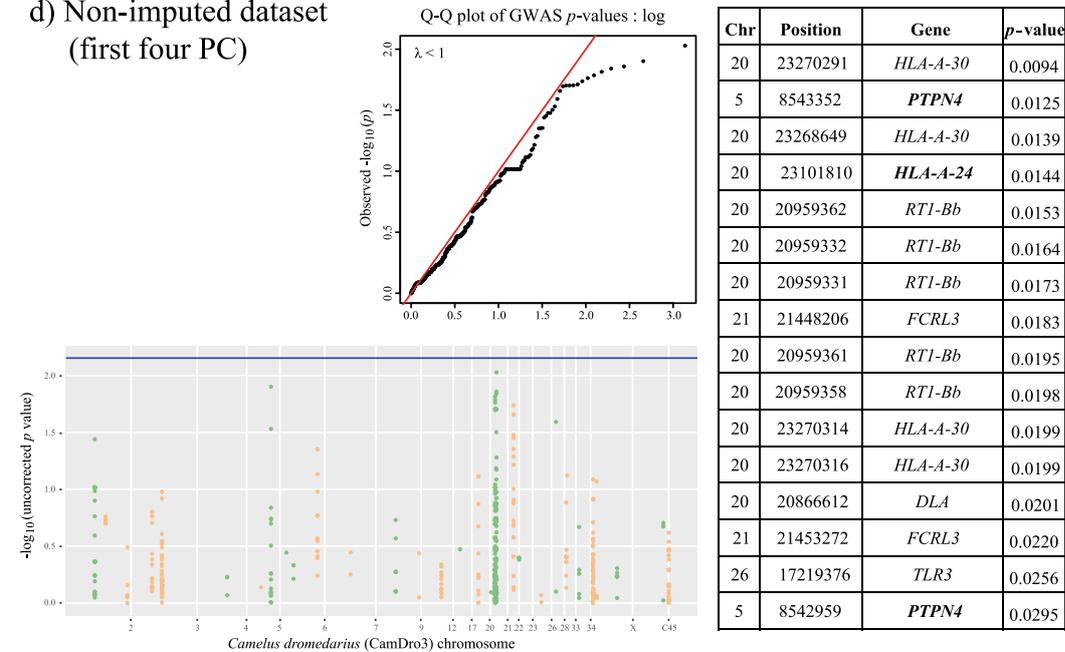


Figure S2. Manhattan and QQ plot. Highlighted in bold are the four genes that are common in all analyses (*HLA-A*-like, *PTPN4*, *MAGOHB* and *DNAH7*). FDR corrected thresholds are represented in blue. a) Total dataset; b) only livestock market samples; c) only spring field samples; d) non-imputed dataset. C45 corresponds to Contig45, an unplaced scaffold.

Supplementary tables

Table S1. Sample information and phenotype characterization for virus shedding (swabs) and antibody prevalence (sera).

No	Phenotype	Original code	Lab code	Sampling site & date	Sex (M/F)	Age (yr)	Chip ID	Farm location (UAE)	Nasal Swabs	Camel serum samples			Samples used in the association tests
									MERS CoV ORF1a RT-qPCR (ct value)	MERS CoV ORF1a RT-qPCR	Anti-MERS CoV ELISA (extinction ratio)	Anti-MERS-CoV-IIFT (antibody titer)	
1	V- AB-	51	Drom1508	Livestock market, Oct 19	F	2	985007841357609	Marmoom	neg.	neg.	borderline 1.1	1:100	
2	V- AB-	88	Drom1545	Livestock market, Oct 19		6 m	NO CHIP INFO		neg.	neg.	borderline 0.9	1:100	
3	V- AB-	NSB14	Drom1831	Livestock market, April 19	M	2	900111881038113	Al Saad	neg.	neg..	borderline 1.02	≥ 1:1,000	
4	V- AB-	M12	Drom1940	Bedouin farm, Al Mazrooei, March 19	M	2 m	NO CHIP INFO		neg.	neg.	borderline 1.00	1:1,000	
5	V- AB-	NSw53	Drom1903	Wildlife Reserve, ArabAdv, April 19	F	2	NO CHIP INFO		neg.	neg.	neg. 0.07	neg.	
6	V- AB-	NSw54	Drom1904	Wildlife Reserve, ArabAdv, April 19	M	6	NO CHIP INFO		neg.	neg.	borderline 1.09	1:100	
7	V- AB-	NSw57	Drom1907	Wildlife Reserve, DesertStar, April 19	M	8-9	NO CHIP INFO		neg.	neg.	neg. 0.64	1:100	
8	V- AB+	13	Drom1470	Livestock market, Oct 19	F	1	784010050046584	Abu Samra	doubtful 39.6	neg.	pos. 4.0	≥ 1:1,000	
9	V- AB+	14	Drom1471	Livestock market, Oct 19	F	6	992001000331305		neg.	neg.	pos. 3.4	≥ 1:1,000	X
10	V- AB+	25	Drom1482	Livestock market, Oct 19	F	3	900057600121775	Dubai	doubtful 40.4	neg.	pos. 2.5	≥ 1:1,000	
11	V- AB+	31	Drom1488	Livestock market, Oct 19	M	2	908182001493742		neg.	neg.	pos. 4.7	≥ 1:1,000	X
12	V- AB+	35	Drom1492	Livestock market, Oct 19	F	6	900057600126791	Malaghat	neg.	neg.	pos. 4.8	≥ 1:1,000	X
13	V- AB+	37	Drom1494	Livestock market, Oct 19	F	5	968000002916753		neg.	neg.	pos. 4.5	≥ 1:1,000	X
14	V- AB+	38	Drom1495	Livestock market, Oct 19	M	6	985007841229562	Al Wagan	neg.	neg.	pos. 5.9	≥ 1:1,000	X
15	V- AB+	39	Drom1496	Livestock market, Oct 19	F	6	784010050348103		neg.	neg.	pos. 4.1	≥ 1:1,000	X
16	V- AB+	42	Drom1499	Livestock market, Oct 19	F	3	784019000006947	Sweihan	neg.	neg.	pos. 3.5	≥ 1:1,000	X
17	V- AB+	43	Drom1500	Livestock market, Oct 19	F	6	985007841219484	Marmoom	neg.	neg.	pos. 4.8	≥ 1:1,000	X
18	V- AB+	47	Drom1504	Livestock market, Oct 19	F	6	784010050073661	Marakh	neg.	neg.	pos. 5.1	≥ 1:1,000	X

19	V- AB+	48	Drom1505	Livestock market, Oct 19	F	5	784019000006158	Mezyad	neg.	neg.	pos. 3.2	1:1,000	X
20	V- AB+	69	Drom1526	Livestock market, Oct 19	F	6	784010050079111		neg.	neg.	pos. 4.6	≥ 1:1,000	X
21	V- AB+	72	Drom1529	Livestock market, Oct 19			991001001739927		neg.	neg.	pos. 4.2	≥ 1:1,000	
22	V- AB+	79	Drom1536	Livestock market, Oct 19	F	6	784010050079139		neg.	neg.	pos. 4.4	≥ 1:1,000	X
23	V- AB+	85	Drom1542	Livestock market, Oct 19	F	4	784010050140963	Dubai	doubtful 39.8	neg.	pos. 3.5	≥ 1:1,000	
24	V- AB+	NSB5	Drom1823	Livestock market, April 19	M	2	900111881038114	Al Saad	neg.	neg..	pos. 2.39	1:1,000	X
25	V- AB+	NSB8	Drom1826	Livestock market, April 19	M	2	900111881038106	Al Saad	neg.	neg..	pos. 1.74	1:1,000	X
26	V- AB+	NSB12	Drom1829	Livestock market, April 19			NO CHIP INFO		neg.	neg..	pos. 2.02	1:1,000	
27	V- AB+	NSB13	Drom1830	Livestock market, April 19	F	4	784010050423283	Bida Bint Saud / Bad' Bint Sa'oud	neg.	neg..	pos. 3.49	≥ 1:1,000	X
28	V- AB+	NSB16	Drom1833	Livestock market, April 19	F	4-5	634078000075327	Sweihan	neg.	neg..	pos. 3.14	≥ 1:1,000	X
29	V- AB+	NSB17	Drom1834	Livestock market, April 19	F	8	784010050077291	Bida Bint Saud / Bad' Bint Sa'oud	neg.	neg..	pos. 4.09	≥ 1:1,000	X
30	V- AB+	NSB18	Drom1835	Livestock market, April 19	F	2	900182001414799	Al Jabeeb	neg.	neg..	pos. 3.01	≥ 1:1,000	X
31	V- AB+	NSB19	Drom1836	Livestock market, April 19	M	1	991001002575031	Marakh	neg.	neg..	pos. 5.10	≥ 1:1,000	X
32	V- AB+	NSB20	Drom1837	Livestock market, April 19	M	2	900111880935178	Marakh	neg.	neg..	pos. 3.67	≥ 1:1,000	X
33	V- AB+	NSB21	Drom1838	Livestock market, April 19	M	3	985007841400429	Zaid	neg.	neg..	pos. 1.61	≥ 1:1,000	X
34	V- AB+	NSB41	Drom1858	Livestock market, April 19	F	4	985007841359958		neg.	neg..	pos. 3.61	1:1,000	X
35	V- AB+	NSB42	Drom1859	Livestock market, April 19	F	4	784010050067053	Al Kowah	neg.	neg..	pos. 4.14	≥ 1:1,000	X
36	V- AB+	NSB46	Drom1863	Livestock market, April 19	F	6	991001002575819	Mulakat	neg.	neg..	pos. 4.46	≥ 1:1,000	X
37	V- AB+	M1	Drom1929	Bedouin farm, Al Mazrooei, March 19	F	4	NO CHIP INFO		neg.	neg.	pos. 3.84	≥ 1:1,000	X
38	V- AB+	M2	Drom1930	Bedouin farm, Al Mazrooei, March 19	F	4	NO CHIP INFO		neg.	neg.	pos. 3.96	≥ 1:1,000	X
39	V- AB+	M3	Drom1931	Bedouin farm, Al Mazrooei, March 19	F	4	NO CHIP INFO		neg.	neg.	pos. 2.01	1:1,000	X
40	V- AB+	M4	Drom1932	Bedouin farm, Al Mazrooei, March 19	F	15	NO CHIP INFO		neg.	neg.	pos. 1.35	1:1,000	X
41	V- AB+	M5	Drom1933	Bedouin farm, Al Mazrooei, March 19	F	8	NO CHIP INFO		neg.	neg.	pos. 4.06	≥ 1:1,000	X

42	V- AB+	M6	Drom1934	Bedouin farm, Al Mazrooei, March 19	F	25-30	NO CHIP INFO		neg.	neg.	pos. 3.89	≥ 1:1,000	X
43	V- AB+	M7	Drom1935	Bedouin farm, Al Mazrooei, March 19	F	5	NO CHIP INFO		neg.	neg.	pos. 4.17	≥ 1:1,000	X
44	V- AB+	M8	Drom1936	Bedouin farm, Al Mazrooei, March 19	F	12	NO CHIP INFO		neg.	neg.	pos. 4.03	≥ 1:1,000	X
45	V- AB+	M9	Drom1937	Bedouin farm, Al Mazrooei, March 19	F	12	NO CHIP INFO		neg.	neg.	pos. 2.89	1:1,000	X
46	V- AB+	M10	Drom1938	Bedouin farm, Al Mazrooei, March 19	F	12	NO CHIP INFO		neg.	neg.	pos. 2.96	≥ 1:1,000	X
47	V- AB+	M11	Drom1939	Bedouin farm, Al Mazrooei, March 19	F	14	NO CHIP INFO		neg.	neg.	pos. 2.91	1:1,000	X
48	V- AB+	M13	Drom1941	Bedouin farm, Al Mazrooei, March 19	M	10	NO CHIP INFO		neg.	neg.	pos. 4.04	≥ 1:1,000	X
49	V- AB+	M14	Drom1942	Bedouin farm, Al Mazrooei, March 19	M	14	NO CHIP INFO		neg.	neg.	pos. 4.11	≥ 1:1,000	X
50	V- AB+	M15	Drom1943	Bedouin farm, Al Mazrooei, March 19	F	10	NO CHIP INFO		neg.	neg.	pos. 2.90	≥ 1:1,000	X
51	V- AB+	NSw20	Drom1872	Wildlife Reserve, Al Maha, April 19	M	15	NO CHIP INFO		neg.	neg.	pos. 2.67	1:1,000	X
52	V- AB+	NSw21	Drom1873	Wildlife Reserve, Al Maha, April 19	M	13	NO CHIP INFO		neg.	neg.	pos. 3.54	≥ 1:1,000	X
53	V- AB+	NSw22	Drom1874	Wildlife Reserve, Al Maha, April 19	M	17	NO CHIP INFO		neg.	neg.	pos. 3.09	1:1,000	X
54	V- AB+	NSw23	Drom1875	Wildlife Reserve, Al Maha, April 19	F	14	NO CHIP INFO		neg.	neg.	pos. 3.40	1:1,000	X
55	V- AB+	NSw25	Drom1877	Wildlife Reserve, Al Maha, April 19	M	9	NO CHIP INFO		neg.	neg.	pos. 3.25	1:1,000	X
56	V- AB+	NSw26	Drom1878	Wildlife Reserve, Al Maha, April 19	M	7	NO CHIP INFO		neg.	neg.	pos. 2.19	≥ 1:1,000	X
57	V- AB+	NSw27	Drom1879	Wildlife Reserve, Al Maha, April 19	F	5	NO CHIP INFO		neg.	neg.	pos. 2.12	1:1,000	X
58	V- AB+	NSw30	Drom1882	Wildlife Reserve, Al Maha, April 19	F	14	NO CHIP INFO		neg.	neg.	pos. 4.26	≥ 1:1,000	X
59	V- AB+	NSw32	Drom1884	Wildlife Reserve, Al Maha, April 19	M	12	NO CHIP INFO		neg.	neg.	pos. 3.95	≥ 1:1,000	X
60	V- AB+	NSw34	Drom1886	Wildlife Reserve, Al Maha, April 19	M	7	NO CHIP INFO		neg.	neg.	pos. 3.01	1:100	X
61	V- AB+	NSw35	Drom1887	Wildlife Reserve, Al Maha, April 19	M	2 y 1 m	NO CHIP INFO		neg.	neg.	pos. 1.87	1:1,000	X
62	V- AB+	NSw38	Drom1890	Wildlife Reserve, Alpha, April 19	M	15	NO CHIP INFO		neg.	neg.	pos. 3.89	1:1,000	X
63	V- AB+	NSw40	Drom1892	Wildlife Reserve, Travco, April 19	M	12	NO CHIP INFO		neg.	neg.	pos. 4.58	≥ 1:1,000	X
64	V- AB+	NSw42	Drom1893	Wildlife Reserve, Travco, April 19	M	9	NO CHIP INFO		neg.	neg.	pos. 2.86	1:100	X
65	V- AB+	NSw43	Drom1894	Wildlife Reserve, ArabAdv, April 19	F	16	NO CHIP INFO		neg.	neg.	pos. 4.42	≥ 1:1,000	X
66	V- AB+	NSw46	Drom1897	Wildlife Reserve, ArabAdv, April 19	M	16	NO CHIP INFO		neg.	neg.	pos. 3.26	≥ 1:1,000	X

67	V- AB+	NSw48	Drom1899	Wildlife Reserve, ArabAdv, April 19	M	12	NO CHIP INFO		neg.	neg.	pos. 4.55	≥ 1:1,000	X
68	V- AB+	NSw51	Drom1901	Wildlife Reserve, ArabAdv, April 19	F	16	NO CHIP INFO		neg.	neg.	pos. 2.76	≥ 1:1,000	X
69	V- AB+	NSw52	Drom1902	Wildlife Reserve, ArabAdv, April 19	M	11	NO CHIP INFO		neg.	neg.	pos. 3.24	1:1,000	X
70	V- AB+	NSw55	Drom1905	Wildlife Reserve, ArabAdv, April 19	F	18	NO CHIP INFO		neg.	neg.	pos. 3.37	≥ 1:1,000	X
71	V- AB+	NSw56	Drom1906	Wildlife Reserve, ArabAdv, April 19	F	6 m	NO CHIP INFO		neg.	neg.	pos. 1.28	1:100	X
72	V- AB+	NSw59	Drom1909	Wildlife Reserve, DesertStar, April 19	M	5	NO CHIP INFO		neg.	neg.	pos. 2.13	1:1,000	X
73	V- AB+	NSw60	Drom1910	Wildlife Reserve, DesertStar, April 19	M	5	NO CHIP INFO		neg.	neg.	pos. 3.10	1:1,000	X
74	V- AB+	NSw63	Drom1913	Wildlife Reserve, DesertStar, April 19	M	15	NO CHIP INFO		neg.	neg.	pos. 2.35	1:1,000	X
75	V- AB+	NSw67	Drom1917	Wildlife Reserve, DesertStar, April 19	M	18	NO CHIP INFO		neg.	neg.	pos. 3.02	1:1,000	X
76	V- AB+	NSw69	Drom1919	Wildlife Reserve, DesertStar, April 19	M	18	NO CHIP INFO		neg.	neg.	pos. 3.34	≥ 1:1,000	X
77	V- AB+	NSw79	Drom1928	Wildlife Reserve, DesertStar, April 19	M	7	NO CHIP INFO		neg.	neg.	pos. 3.42	1:1,000	X
78	V+ AB-	2	Drom1459	Livestock market, Oct 19	M	4 m	NO CHIP INFO		pos. 29.3	neg.	neg. 0.1	neg.	
79	V+ AB-	5	Drom1462	Livestock market, Oct 19	M	1	900057600126219	Marakh	pos. 19.9	neg.	neg. 0.1	neg.	
80	V+ AB-	6	Drom1463	Livestock market, Oct 19	M	3 m	NO CHIP INFO		pos. 25.1	neg.	neg. 0.1	neg.	
81	V+ AB-	7	Drom1464	Livestock market, Oct 19	M	2 m	NO CHIP INFO		pos. 34.2	neg.	neg. 0.5	1:100	
82	V+ AB-	54	Drom1511	Livestock market, Oct 19	M	2	900057600121585		pos. 33.4	neg.	borderline 0.9	1:1,000	
83	V+ AB-	87	Drom1544	Livestock market, Oct 19	M	1	900057600121595		Pos. 39.2	neg.	neg. 0.3	1:100	
84	V+ AB-	NSB40	Drom1857	Livestock market, April 19		2 m	NO CHIP INFO		pos. 26.1	neg..	neg. 0.33	1:100	
85	V+ AB+	1	Drom1458	Livestock market, Oct 19	M	2	900182001697895		pos. 34.8	neg.	pos. 3.6	≥ 1:1,000	X
86	V+ AB+	3	Drom1460	Livestock market, Oct 19	M	2	900215000003532		pos. 30.3	neg.	pos. 4.1	≥ 1:1,000	X
87	V+ AB+	4	Drom1461	Livestock market, Oct 19			900111881027297		pos. 36.7	neg.	pos. 1.9	1:1,000	
88	V+ AB+	8	Drom1465	Livestock market, Oct 19	F	3	900057600122200	Abu Dhabi	pos. 36.2	neg.	pos. 5.0	≥ 1:1,000	X
89	V+ AB+	9	Drom1466	Livestock market, Oct 19	F	3	900215000005008	Abu Dhabi	pos. 36.8	neg.	pos. 5.3	≥ 1:1,000	X
90	V+ AB+	10	Drom1467	Livestock market, Oct 19	F	5	985007841277728	Marakh	pos. 37.1	neg.	pos. 2.6	≥ 1:1,000	X
91	V+ AB+	11	Drom1468	Livestock market, Oct 19	F	6	784010050031772	Mulakat	pos. 36.3	neg.	pos. 4.8	≥ 1:1,000	X

92	V+ AB+	23	Drom1480	Livestock market, Oct 19	F	6	784010050267117		pos. 36.9	neg.	pos. 4.5	≥ 1:1,000	X
93	V+ AB+	27	Drom1484	Livestock market, Oct 19	F	6m	784010050465212	Al Yahar	pos. 35.1	neg.	pos. 5.0	≥ 1:1,000	X
94	V+ AB+	28	Drom1485	Livestock market, Oct 19	M	5 m	NO CHIP INFO		pos. 27.2	neg.	pos. 4.5	≥ 1:1,000	X
95	V+ AB+	29	Drom1486	Livestock market, Oct 19	M	6 m	NO CHIP INFO		pos. 35.4	neg.	pos. 1.8	≥ 1:1,000	X
96	V+ AB+	32	Drom1489	Livestock market, Oct 19	M	2	991001002574519		pos. 38.6	neg.	pos. 4.0	≥ 1:1,000	X
97	V+ AB+	45	Drom1502	Livestock market, Oct 19	M	5	985007841209387	Khushaba	pos. 38.0	neg.	pos. 4.6	≥ 1:1,000	X
98	V+ AB+	56	Drom1513	Livestock market, Oct 19	F	2	784010050551133	Jabib	pos. 36.1	neg.	pos. 4.2	≥ 1:1,000	X
99	V+ AB+	90	Drom1547	Livestock market, Oct 19	F	6	784010050294118	Marakh	pos. 35.8	neg.	pos. 4.3	≥ 1:1,000	X
100	V+ AB+	NSB10	Drom1827	Livestock market, April 19	F	2	784010050550508	Seih Sabra/Sih Sabra/Seeh Sabra	pos. 38.2	neg.	pos. 1.87	1:1,000	X
101	V+ AB+	NSB11	Drom1828	Livestock market, April 19	M	2	784010050550691	Mezyad	pos. 34.9	neg.	pos. 2.73	1:1,000	X
102	V+ AB+	NSB23	Drom1840	Livestock market, April 19	M	1	991001002570111		pos. 32.6	neg.	pos. 4.34	≥ 1:1,000	X
103	V+ AB+	NSB30	Drom1847	Livestock market, April 19	F	4	784010050229794	Badr Zaid	pos. 39.5	neg.	pos. 1.96	1:1,000	X
104	V+ AB+	NSB31	Drom1848	Livestock market, April 19	F	6	784010050243508		pos. 39.0	neg.	pos. 4.69	≥ 1:1,000	X
105	V+ AB+	NSB33	Drom1850	Livestock market, April 19	M	1-2	992001000330620	Malaqāt	pos. 32.3	neg.	pos. 4.67	≥ 1:1,000	X
106	V+ AB+	NSB34	Drom1851	Livestock market, April 19	F	2	991001002575035		pos. 35.1	neg.	pos. 4.44	1:1,000	X
107	V+ AB+	NSB35	Drom1852	Livestock market, April 19	F	3	991001002574462	Sweihan	pos. 33.4	neg.	pos. 1.37	1:100	X
108	V+ AB+	NSB36	Drom1853	Livestock market, April 19	M	3	991001002575745	Jebayeb	pos. 35.4	neg.	pos. 1.33	1:100	X
109	V+ AB+	NSB37	Drom1854	Livestock market, April 19	M	2-3	992001000330472	Al Saad	pos. 34.1	neg.	pos. 1.84	1:100	X
110	V+ AB+	NSB38	Drom1855	Livestock market, April 19	M	3	991001002575920		pos. 37.5	neg.	pos. 4.02	≥ 1:1,000	X
111	V+ AB+	NSB39	Drom1856	Livestock market, April 19	F	6	784010050365484	RAK	pos. 33.5	neg.	pos. 4.67	≥ 1:1,000	X
112	V+ AB+	NSB43	Drom1860	Livestock market, April 19	M	1-2	992001000330719	Marakh	pos. 37.4	neg.	pos. 2.40	1:100	X
113	V+ AB+	NSB44	Drom1861	Livestock market, April 19	F	6	784010050516433		pos. 37.5	neg.	pos. 4.82	≥ 1:1,000	X
114	V+ AB+	NSB45	Drom1862	Livestock market, April 19	F	3	784010050028578	Jabib – Al Faqa	pos. 38.9	neg.	pos. 3.53	≥ 1:1,000	X
115	V+ AB+	NSB47	Drom1864	Livestock market, April 19	M	3	991001002575899	Marmoon	pos. 35.2	neg.	pos. 1.76	1:1,000	X
116	V+ AB+	NSB48	Drom1865	Livestock market, April 19	M	1-2	990001000053606	Sweihan	pos. 33.7	neg.	pos. 4.82	≥ 1:1,000	X

117	V+ AB+	NSB49	Drom1866	Livestock market, April 19	F	3	900111881038306		pos. 36.1	neg.	pos. 5.47	≥ 1:1,000	X
118	V+ AB+	NSB50	Drom1867	Livestock market, April 19		2m	NO CHIP INFO		pos. 38.5	neg.	pos. 3.17	1:1,000	X
119	V+ AB+	NSB51	Drom1868	Livestock market, April 19	M	2	991001002574752	RAK	pos. 30.0	neg.	pos. 2.04	1:1,000	X
120	V+ AB+	NSB52	Drom1869	Livestock market, April 19	M	1-2	900074001585559	Shabiah	pos. 39.3	neg.	pos. 4.91	≥ 1:1,000	X
121	V+ AB+	NSB54	Drom1871	Livestock market, April 19	F	5	784010050359111		pos. 36.3	neg.	pos. 3.89	1:1,000	X

Table S2. Observed (H_O) and expected (H_E) heterozygosity values depicted in immune response gene groups. Identified candidate genes *MAGOHB*, *HLA-A-24*-like, *HLA-DPB1*-like, *DNAH7* and *PTPN4* are highlighted in bold.

Gene ID	Genes- No.SNPs	Genes- H_O	Genes- H_E	Exons- No.SNPs	Exons- H_O	Exons- H_E	Introns- No.SNPs	Introns- H_O	Introns- H_E	Name	Description
Granzyme											
Cadr_00004168	4	0.272	0.329	0	NA	NA	4	0.272	0.329	<i>GZMA</i>	Granzyme A (Bos taurus OX=9913)
Cadr_00004169	1	NA	NA	0	NA	NA	1	NA	NA	<i>GZMA</i>	Granzyme A (Homo sapiens OX=9606)
Cadr_00005822	0	NA	NA	0	NA	NA	0	NA	NA	<i>GZMB</i>	Granzyme B (Homo sapiens OX=9606)
Cadr_00005823	22	0.24	0.246	1	NA	NA	22	0.24	0.246	<i>GZMB</i>	Granzyme B (Homo sapiens OX=9606)
Cadr_00005821	5	0.411	0.38	0	NA	NA	5	0.411	0.38	<i>GZMH</i>	Granzyme H (Homo sapiens OX=9606)
Cadr_00004167	6	0.29	0.343	1	NA	NA	5	0.27	0.32	<i>GZMK</i>	Granzyme K (Homo sapiens OX=9606)
Cadr_00025032	16	0.246	0.255	3	0.056	0.066	13	0.289	0.299	<i>GZMM</i>	Granzyme M (Homo sapiens OX=9606)
Mean		0.29	0.31		0.06	0.07		0.30	0.31		
Interleukin											
Cadr_00001885	3	0.423	0.484	0	NA	NA	3	0.423	0.484	<i>CXCL8</i>	Interleukin-8 (Canis lupus familiaris OX=9615)
Cadr_00023412	7	0.197	0.197	2	0.158	0.146	5	0.213	0.217	<i>IL10</i>	Interleukin-10 (Lama glama OX=9844)
Cadr_00028914	14	0.201	0.228	11	0.187	0.22	9	0.246	0.278	<i>IL10RA</i>	Interleukin-10 receptor subunit alpha (Homo sapiens OX=9606)
Cadr_00001098	61	0.165	0.168	1	NA	NA	61	0.165	0.168	<i>IL10RB</i>	Interleukin-10 receptor subunit alpha (Homo sapiens OX=9606)
Cadr_00029940	16	0.32	0.318	5	0.19	0.188	11	0.379	0.377	<i>IL1A</i>	Interleukin-1 alpha (Lama glama OX=9844)
Cadr_00029941	20	0.285	0.308	2	0.306	0.317	18	0.283	0.307	<i>IL1B</i>	Interleukin-1 beta (Lama glama OX=9844)
Mean		0.27	0.28		0.21	0.22		0.28	0.31		
Killer cell											
Cadr_00029273	19	0.277	0.295	3	0.169	0.179	16	0.297	0.317	<i>Klra2</i>	Killer cell lectin-like receptor 2 (Mus musculus OX=10090)
Cadr_00029303	4	0.229	0.22	0	NA	NA	4	0.229	0.22	<i>KLRB1</i>	Killer cell lectin-like receptor subfamily B member 1 (Homo sapiens OX=9606)
Cadr_00029300	2	0.448	0.399	2	0.448	0.399	0	NA	NA	<i>Klrb1b</i>	Killer cell lectin-like receptor subfamily B member 1B allele A (Camelus bactrianus XP_010944886.1)
Cadr_00029489	18	0.348	0.389	1	NA	NA	17	0.344	0.382	<i>KLRC2</i>	NKG2-C type II integral membrane protein (Homo sapiens OX= 9606)
Cadr_00029281	10	0.315	0.349	8	0.289	0.323	7	0.312	0.34	<i>KLRD1</i>	Natural killer cells antigen CD94 (Bos taurus OX=9913)
Cadr_00029283	13	0.345	0.35	1	NA	NA	12	0.341	0.347	<i>Klre1</i>	Killer cell lectin-like receptor subfamily E member 1 (Mus musculus OX=10090)
Cadr_00029297	73	0.458	0.435	3	0.374	0.367	70	0.462	0.438	<i>KLRF1</i>	Killer cell lectin-like receptor subfamily F member 1 (Macaca fascicularis OX=9541)
Cadr_00029295	48	0.266	0.244	3	0.294	0.261	47	0.266	0.244	<i>KLRF2</i>	Killer cell lectin-like receptor subfamily F member 2 (Homo sapiens OX=9606)
Cadr_00008447	42	0.352	0.354	11	0.414	0.411	31	0.33	0.334	<i>Klrg2</i>	Killer cell lectin-like receptor subfamily G member 2 (Mus musculus OX=10090)
Cadr_00029277	28	0.369	0.403	1	NA	NA	28	0.369	0.403	<i>Klri1</i>	Killer cell lectin-like receptor subfamily I member 1 (Mus musculus OX=10090)
Cadr_00029276	11	0.406	0.419	1	NA	NA	11	0.406	0.419	<i>KLRK1</i>	NKG2-D type II integral membrane protein (Pongo Pygmaeus OX=9600)
Cadr_00029279	14	0.244	0.251	4	0.248	0.254	10	0.242	0.249	<i>KLRK1</i>	NKG2-D type II integral membrane protein (Sus scrofa OX=9823)
Mean		0.34	0.34		0.32	0.31		0.33	0.34		
MHC Class I											
Cadr_00022140	112	0.09	0.116	15	0.061	0.067	97	0.094	0.124	<i>HLA-A-24-like</i>	HLA class I histocompatibility antigen, A-24 alpha chain (Homo sapiens OX=9606)

Cadr_00022145	30	0.276	0.273	8	0.318	0.322	22	0.261	0.256	<i>HLA-A-11</i>	HLA class I histocompatibility antigen, A-11 alpha chain (Homo sapiens OX=9606)
Cadr_00022149	14	0.216	0.279	12	0.22	0.285	5	0.181	0.215	<i>HLA-A-69</i>	HLA class I histocompatibility antigen, A-69 alpha chain (Homo sapiens OX=9606)
Cadr_00022150	134	0.191	0.263	29	0.196	0.288	116	0.191	0.261	<i>HLA-A-30</i>	HLA class I histocompatibility antigen, A-30 alpha chain (Homo sapiens OX=9606)
Cadr_00022148	20	0.225	0.329	10	0.21	0.301	10	0.241	0.357	<i>HLA-C</i>	HLA class I histocompatibility antigen, Cw-6 alpha chain (Homo sapiens OX=9606)
Cadr_00022156	4	0.028	0.038	2	0.026	0.045	2	0.031	0.03	<i>Patr</i>	class I histocompatibility B-1 alpha chain (Fragment) (Pan troglodytes OX=9598)
Cadr_00022105	18	0.197	0.187	10	0.203	0.188	8	0.191	0.184	<i>Patr-A</i>	Patr class I histocompatibility antigen, A-126 alpha chain (Pan troglodytes OX=9598)
Cadr_00022139	67	0.315	0.318	3	0.239	0.241	64	0.318	0.322	<i>Patr-A</i>	Patr class I histocompatibility antigen, A-126 alpha chain (Pan troglodytes OX=9598)
Cadr_00022147	5	0.02	0.032	4	0.025	0.035	1	NA	NA	<i>Patr-A</i>	Patr class I histocompatibility antigen, A-126 alpha chain (Pan troglodytes OX=9598)
Cadr_00022160	31	0.047	0.09	3	0.048	0.075	28	0.047	0.092	<i>Patr-A</i>	Patr class I histocompatibility antigen, A-126 alpha chain (Pan troglodytes OX=9598)
Cadr_00022155	0	NA	NA	0	NA	NA	0	NA	NA	<i>Popy</i>	class I histocompatibility antigen A-1 alpha chain (Pongo pygmaeus OX=9600)
Mean		0.16	0.19		0.15	0.18		0.17	0.20		
MHC Class II											
Cadr_00022027	28	0.35	0.343	2	0.48	0.45	26	0.339	0.334	<i>BoLA-DQB</i>	BoLa class II histocompatibility antigen, DQB*0101 beta chain (Bos taurus OX=9913)
Cadr_00004894	11	0.324	0.299	2	0.397	0.378	9	0.307	0.281	<i>CD74</i>	HLA class II histocompatibility antigen gamma chain (Homo sapiens OX=9606)
Cadr_00022030	95	0.412	0.402	37	0.437	0.416	58	0.396	0.393	<i>DLA</i>	class II histocompatibility antigen, DR-1 beta chain (Canis lupus familiaris OX=9615)
Cadr_00022020	5	0.27	0.307	2	0.21	0.245	3	0.31	0.348	<i>HLA-DMA</i>	HLA class II histocompatibility antigen, DM alpha chain (Homo sapiens OX=9606)
Cadr_00022021	33	0.287	0.319	8	0.292	0.335	25	0.285	0.314	<i>HLA-DMB</i>	HLA class II histocompatibility antigen, DM alpha chain (Homo sapiens OX=9606)
Cadr_00022018	25	0.27	0.295	17	0.268	0.293	8	0.274	0.3	<i>HLA-DOA</i>	HLA class II histocompatibility antigen, DO alpha chain (Homo sapiens OX=9606)
Cadr_00022026	72	0.328	0.328	10	0.245	0.267	62	0.342	0.338	<i>HLA-DOB</i>	HLA class II histocompatibility antigen, DO alpha chain (Homo sapiens OX=9606)
Cadr_00022017	32	0.274	0.296	5	0.253	0.274	27	0.278	0.3	<i>HLA-DPA1</i>	HLA class II histocompatibility antigen, DP alpha chain (Homo sapiens OX=9606)
Cadr_00022016	41	0.273	0.331	9	0.253	0.307	32	0.279	0.338	<i>HLA-DPBI-like</i>	HLA class II histocompatibility antigen, DP alpha chain (Homo sapiens OX=9606)
Cadr_00022036	89	0.304	0.372	13	0.245	0.386	76	0.314	0.369	<i>HLA-DRB1</i>	HLA class II histocompatibility antigen, DRB1-4 alpha chain (Homo sapiens OX=9606)
Cadr_00022037	63	0.414	0.43	6	0.32	0.318	57	0.424	0.442	<i>HLA-DRB1</i>	HLA class II histocompatibility antigen, DRB1-1 alpha chain (Homo sapiens OX=9606)
Cadr_00022038	34	0.219	0.236	13	0.241	0.26	21	0.206	0.221	<i>Mamu-DRA</i>	Mamu class II histocompatibility antigen, DR alpha chain (Macac mulata OX=9544)
Cadr_00022032	7	0.048	0.101	2	0	0.022	5	0.067	0.133	<i>RT1-Bb</i>	Rano class II histocompatibility antigen, B-1 beta chain (Rattus norvegicus OX=10116)
Cadr_00022034	36	0.172	0.202	2	0.304	0.266	34	0.164	0.198	<i>RT1-Bb</i>	Rano class II histocompatibility antigen, B-1 beta chain (Rattus norvegicus OX=10116)
Cadr_00022028	15	0.423	0.378	9	0.426	0.383	6	0.417	0.372	<i>SLA</i>	class II histocompatibility antigen, DQ haplotype D alpha chain (Sus scrofa =X=9823)
Cadr_00022033	6	0.003	0.023	0	NA	NA	6	0.003	0.023	<i>SLA</i>	class II histocompatibility antigen, DQ haplotype D alpha chain (Sus scrofa =X=9823)
Cadr_00022035	4	0.012	0.047	3	0.013	0.046	1	NA	NA	<i>SLA</i>	class II histocompatibility antigen, DQ haplotype D alpha chain (Sus scrofa =X=9823)
Mean		0.26	0.28		0.27	0.29		0.28	0.29		
TLR											
Cadr_00002152	5	0.192	0.19	5	0.192	0.19	0	NA	NA	<i>TLR1</i>	Toll-like receptor 1 (Homo sapiens OX=9606)
Cadr_00002153	23	0.274	0.304	16	0.291	0.317	9	0.222	0.256	<i>TLR10</i>	Toll-like receptor 10 (Bos taurus OX=9913)
Cadr_00001385	10	0.398	0.437	10	0.398	0.437	3	0.355	0.363	<i>TLR2</i>	Toll-like receptor 2 (Equus caballus OX=9796)
Cadr_00026583	31	0.384	0.398	6	0.24	0.253	25	0.419	0.433	<i>TLR3</i>	Toll-like receptor 3 (Boselaphus tragocamelus OX=9917)
Cadr_00016120	8	0.222	0.21	1	NA	NA	7	0.224	0.213	<i>TLR4</i>	Toll-like receptor 4 (Sus scrofa OX=9823)
Cadr_00023195	4	0.453	0.46	4	0.453	0.46	0	NA	NA	<i>TLR5</i>	Toll-like receptor 5 (Homo sapiens OX=9606)
Cadr_00002151	15	0.195	0.217	2	0.112	0.124	13	0.208	0.231	<i>TLR6</i>	Toll-like receptor 6 (Homo sapiens OX=9606)
Cadr_00003728	18	0.186	0.376	1	NA	NA	17	0.186	0.37	<i>TLR7</i>	Toll-like receptor 7 (Homo sapiens OX=9606)
Cadr_00003726	7	0.166	0.302	7	0.166	0.302	0	NA	NA	<i>TLR8</i>	Toll-like receptor 8 (Homo sapiens OX=9606)
Cadr_00020415	3	0.088	0.083	3	0.088	0.083	0	NA	NA	<i>TLR9</i>	Toll-like receptor 9 (Sus scrofa OX=9823)

Other IR												
Cadr_00030052	11	0.278	0.295	11	0.278	0.295	7	0.293	0.309	<i>ACO1</i>	Cytoplasmic aconitate hydratase (Bos taurus OX=9913)	
Cadr_00020478	5	0.127	0.135	5	0.127	0.135	5	0.127	0.135	<i>APPL1</i>	DCC-interacting protein 13-alpha (Homo sapiens OX=9606)	
Cadr_00002239	297	0.17	0.171	66	0.248	0.251	239	0.149	0.15	<i>CC2D2A</i>	Coiled-coil and C2 domain-containing protein 2A (Homo sapiens OX=9606)	
Cadr_00029296	27	0.335	0.316	2	0.319	0.284	27	0.335	0.316	<i>CLEC2B</i>	C-type lectin domain family 2 member B (Homo sapiens OX=9606)	
Cadr_00007342	11	0.226	0.215	4	0.282	0.267	7	0.194	0.184	<i>CXCR2</i>	C-X-C chemokine receptor type 2 (Bos taurus OX=9913)	
Cadr_00030053	90	0.284	0.312	11	0.278	0.295	86	0.286	0.314	<i>DDX58</i>	Probable ATP-dependent RNA helicase DDX58 (Sus scrofa OX=9823)	
Cadr_00020479	129	0.269	0.28	15	0.251	0.261	120	0.264	0.276	<i>DNAH7</i>	Dynein heavy chain 7 axonemal (Homo sapiens OX=9607)	
Cadr_00006877	64	0.169	0.176	1	NA	NA	63	0.169	0.176	<i>DPP4</i>	Dipeptidyl peptidase 4 (Bos taurus OX=9913)	
Cadr_00012213	12	0.352	0.352	4	0.425	0.443	8	0.316	0.307	<i>FCAR</i>	Immunoglobulin alpha Fc receptor (Homo sapiens OX=9606)	
Cadr_00024638	47	0.292	0.306	16	0.319	0.343	36	0.296	0.31	<i>FCRL3</i>	Fc receptor-like protein 3 (Homo sapiens OX=9606)	
Cadr_00011189	2	0.291	0.275	1	NA	NA	1	NA	NA	<i>HP</i>	Haptoglobin (Sus scrofa OX=9823)	
Cadr_00006880	29	0.17	0.201	1	NA	NA	28	0.162	0.193	<i>IFIH1</i>	Interferon-induced helicase C domain-containing protein 1 (Homo sapiens OX=9606)	
Cadr_00015578	0	NA	NA	0	NA	NA	0	NA	NA	<i>IFNB2</i>	Interferon beta-2 (Bos taurus OX=9913)	
Cadr_00017035	3	0.177	0.205	0	NA	NA	3	0.177	0.205	<i>IFNG</i>	Interferon gamma (Camelus bactrianus OX=9837)	
Cadr_00001103	50	0.279	0.258	0	NA	NA	50	0.279	0.258	<i>IFNGR2</i>	Interferon gamma receptor 2 (Homo sapiens OX=9606)	
Cadr_00029272	12	0.26	0.279	4	0.126	0.129	8	0.327	0.355	MAGOHB	Protein mago nashi homolog 2 (Bos taurus OX=9913)	
Cadr_00004186	13	0.241	0.256	3	0.293	0.316	10	0.226	0.238	<i>MAP3K1</i>	Mitogen-activated protein kinase kinase kinase 1 (Homo sapiens OX=9606)	
Cadr_00005819	1	NA	NA	0	NA	NA	1	NA	NA	<i>Mast</i>	cell protease 3 (Ovis aries OX=9940)	
Cadr_00012215	15	0.306	0.354	3	0.334	0.391	12	0.299	0.345	<i>NCR1</i>	Natural cytotoxicity triggering receptor 1 (Bos taurus OX=9913)	
Cadr_00021869	9	0.174	0.186	3	0.099	0.095	6	0.212	0.232	<i>NCR2</i>	Natural cytotoxicity triggering receptor 2 (Homo sapiens OX=9606)	
Cadr_00001692	77	0.266	0.254	2	0.243	0.248	75	0.267	0.254	<i>NFKB1</i>	Nuclear factor NF-kappa-B p105 subunit (Canis lupus familiaris OX=9615)	
Cadr_00009474	6	0.273	0.27	2	0.352	0.34	4	0.234	0.235	<i>NFKB2</i>	Nuclear factor NF-kappa-B p100 subunit (Homo sapiens OX=9606)	
Cadr_00029278	15	0.426	0.47	2	0.45	0.503	13	0.422	0.465	<i>NKG2A</i>	NKG2-A/NKG2-B type II integral membrane protein (Macaca mulatta OX=9544)	
Cadr_00029993	12	0.087	0.098	1	NA	NA	11	0.093	0.102	<i>NKL</i>	Antimicrobial peptide NK-lysin (Fragment) (Sus scrofa OX=9823)	
Cadr_00009139	6	0.371	0.378	5	0.348	0.357	1	NA	NA	<i>PRF1</i>	Perforin-1 (Homo sapiens OX=9606)	
Cadr_00007027	22	0.243	0.239	3	0.272	0.273	20	0.244	0.238	<i>PRKRA</i>	Interferon-inducible double-stranded RNA dependent protein kinase activator A (Homo sapiens OX=9606)	
Cadr_00009475	0	NA	NA	0	NA	NA	0	NA	NA	<i>Psd</i>	PH and SEC7 domain-containing protein 1 (Mus musculus OX=10090)	
Cadr_00024639	2	0.399	0.432	2	0.399	0.432	2	0.399	0.432	<i>PTMA</i>	Prothymosin alpha (Pongo abelii OX=9601)	
Cadr_00006681	163	0.213	0.222	17	0.299	0.318	153	0.207	0.215	PTPN4	Tyrosine-protein phosphatase non-receptor type 4 (Homo sapiens OX=9606)	
Cadr_00001384	10	0.398	0.437	10	0.398	0.437	3	0.355	0.363	<i>RNF175</i>	RING finger protein 175 (Homo sapiens OX=9606)	
Cadr_00004895	4	0.333	0.3	0	NA	NA	4	0.333	0.3	<i>Rps14</i>	40S ribosomal protein S14 (Mus musculus OX=10090)	
Cadr_00017710	6	0.41	0.45	2	0.335	0.386	4	0.448	0.481	<i>Rps7</i>	40S ribosomal protein S7 (Rattus norvegicus OX=10116)	
Cadr_00001327	32	0.275	0.273	14	0.26	0.259	18	0.287	0.284	<i>Suclg1</i>	Succinate-CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial (Mus musculus OX=10090)	
Cadr_00002785	18	0.132	0.265	1	NA	NA	17	0.131	0.261	<i>Tmem255a</i>	Transmembrane protein 255A (Mus musculus OX=10090)	
Cadr_00022101	7	0.188	0.178	4	0.226	0.211	3	0.138	0.134	<i>TNF</i>	Tumor necrosis factor (Camelus bactrianus OX=9837)	
Cadr_00006503	45	0.372	0.4	19	0.365	0.393	29	0.377	0.407	<i>Traf3</i>	TNF receptor-associated factor 3 (Mus musculus OX=10090)	
Cadr_00011190	1	NA	NA	1	NA	NA	0	NA	NA	<i>TXNL4B</i>	Thioredoxin-like protein 4B (Homo sapiens OX=9606)	
Mean		0.27	0.28		0.29	0.31		0.26	0.27			

Table S3. Statistical analysis of observed heterozygosity (H_o) for immune response gene groups in genes, exons and introns. Means and standard deviations are shown for genes, exon and introns separately. Results are only presented for gene, intron and exon H_o as only these showed significance for both ANOVA and posthoc correction with Benjamini-Hochberg (BH). Gene groups with different letters ('a' and 'b') indicate groups that had significantly different means whilst the same letters indicate non-significant different means.

Gene H_o

Granzyme	Interleukin	Killer Cell	MHC I	MHC II	TLR	Other IR
ab	ab	a	b	ab	ab	ab

	Group	H_o mean	St. Deviation
1	Granzyme	0.292	0.070
2	Interleukin	0.265	0.097
3	Killer Cell	0.338	0.075
4	MHC I	0.161	0.106
5	MHC II	0.258	0.131
6	TLR	0.256	0.118
7	Other IR	0.266	0.088

Exon H_o

Granzyme	Interleukin	Killer Cell	MHC I	MHC II	TLR	Other IR
ab	ab	a	b	ab	ab	b

	Group	H_o mean	St. Deviation
1	Granzyme	0.056	NA
2	Interleukin	0.210	0.065
3	Killer Cell	0.319	0.098
4	MHC I	0.155	0.105
5	MHC II	0.274	0.132
6	TLR	0.243	0.131
7	Other IR	0.293	0.088

Intron H_o

Granzyme	Interleukin	Killer Cell	MHC I	MHC II	TLR	Other IR
ab	ab	a	b	ab	ab	ab

	Group	H_o mean	St. Deviation
1	Granzyme	0.296	0.066
2	Interleukin	0.285	0.099
3	Killer Cell	0.327	0.070
4	MHC I	0.173	0.098
5	MHC II	0.275	0.117
6	TLR	0.269	0.095
7	Other IR	0.260	0.092

Table S4. Observed (H_O) and expected (H_E) heterozygosity in genes, exons and introns in MERS-CoV positive (n = 36) and negative (n = 65) individuals. P -values of mean differences were calculated with Welch t test.

	Genes_ H_O	Genes_ H_E	Exons_ H_O	Exons_ H_E	Introns_ H_O	Introns_ H_E
Negative	0.26	0.28	0.25	0.27	0.26	0.29
Positive	0.27	0.28	0.27	0.27	0.28	0.28
<i>p</i>-value	0.50	0.79	0.58	0.96	0.42	0.78
Welch t	-0.68	0.27	-0.55	0.05	-0.80	0.28
df	180.29	183.97	137.67	139.77	162.80	165.79

Table S5. Linkage Disequilibrium-based haplotype (gene-set) test showing 20 genes with significant SNPs at $p < 0.05$. Identified candidate genes *MAGOHB*, *HLA-A-24-like*, *HLA-DPBI-like*, *DNAH7* and *PTPN4* are highlighted in bold. *HLA-A-24-like* and *MAGOHB* were nominally significant ($p < 0.05$) indicated with an asterisk. NSNP - Number of SNPs in set; NSIG - Total number of SNPs below p -value threshold; ISIG - Number of significant SNPs also passing LD-criterion; STAT - Average test statistic based on ISIG SNPs; EMP1 - Empirical set-based p -value; SNPs - positions of SNPs in the set.

SET	NSNP	NSIG	ISIG	EMP1	SNPs	Name	Description
Cadr_00029272	12	5	1	0.008*	chr34:15362634	MAGOHB	Protein mago nashi homolog 2 (Bos taurus OX=9913)
Cadr_00022140	112	14	2	0.031*	chr20:23100696 23100503	HLA-A-24-like	HLA class I histocompatibility antigen, A-24 alpha chain (Homo sapiens OX=9606)
Cadr_00001327	32	1	1	0.032*	chr2:5565157	<i>Suc1g1</i>	Succinate--CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial (Mus musculus OX=10090)
Cadr_00022016	41	9	1	0.058	chr20:20681619	HLA-DPBI-like	HLA class II histocompatibility antigen, DP alpha chain (Homo sapiens OX=9606)
Cadr_00028914	14	7	2	0.060	chr33:12210072 12210460	<i>IL10RA</i>	Interleukin-10 receptor subunit alpha (Homo sapiens OX=9606)
Cadr_00011189	2	1	1	0.063	chr9:33578918	<i>HP</i>	Haptoglobin (Sus scrofa OX=9823)
Cadr_00029993	12	2	2	0.085	chr28:10593199 10591217	<i>NKL</i>	Antimicrobial peptide NK-lysin (Fragment) (Sus scrofa OX=9823)
Cadr_00001384	10	1	1	0.093	chr2:9866197	<i>RNF175</i>	RING finger protein 175 (Homo sapiens OX=9606)
Cadr_00001385	10	1	1	0.093	chr2:9866197	<i>TLR2</i>	Toll-like receptor 2 (Equus caballus OX=9796)
Cadr_00017035	3	1	1	0.095	chr12:24456808	<i>IFNG</i>	Interferon gamma (Camelus bactrianus OX=9837)
Cadr_00020479	129	21	4	0.117	chr:1723840747 23963829 23948208 23854332	DNAH7	Dynein heavy chain 7 axonemal (Homo sapiens OX=9607)
Cadr_00006681	163	8	3	0.163	chr5:8508361 8569590 8531515	PTPN4	Tyrosine-protein phosphatase non-receptor type 4 (Homo sapiens OX=9606)
Cadr_00022139	67	6	2	0.188	chr20:23039666 23044464	<i>Patr-A</i>	Patr class I histocompatibility antigen, A-126 alpha chain (Pan troglodytes OX=9598)
Cadr_00022038	34	1	1	0.220	chr20:21059892	<i>Mamu-DRA</i>	Mamu class II histocompatibility antigen, DR alpha chain (Macac mulata OX=9544)
Cadr_00022027	28	1	1	0.260	chr20:20837533	<i>BoLA-DQB</i>	BoLa class II histocompatibility antigen, DQB*0101 beta chain (Bos taurus OX=9913)
Cadr_00029273	19	1	1	0.278	chr34:15371765	<i>Klra2</i>	Killer cell lectin-like receptor 2 (Mus musculus OX=10090)

Cadr_00022145	30	2	1	0.314	chr20:23134732	<i>HLA-A-11-like</i>	HLA class I histocompatibility antigen, A-11 alpha chain (Homo sapiens OX=9606)
Cadr_00022026	72	3	1	0.345	chr20:20830433	<i>HLA-DOB</i>	HLA class II histocompatibility antigen, DO alpha chain (Homo sapiens OX=9606)
Cadr_00030053	90	2	2	0.427	Contig45:329958 334608	<i>DDX58</i>	Probable ATP-dependent RNA helicase DDX58 (Sus scrofa OX=9823)
Cadr_00002239	297	17	3	0.427	chr2:113136710 113141381 113168889	<i>CC2D2A</i>	Coiled-coil and C2 domain-containing protein 2A (Homo sapiens OX=9606)

Table S6. Scheme showing how IIFT antibody titers were determined according to the fluorescence of the different sample dilutions.

Sample dilutions / Fluorescent signal			Antibody Titer
1:10	1:100	1:1,000	
weak	negative	negative	1:10
moderate	negative	negative	1:10
strong	weak	negative	1:100
strong	moderate	negative	1:100
strong	strong	weak	1:1,000
strong	strong	moderate	≥ 1:1,000
strong	strong	strong	≥ 1:1,000

Table S7. Read-based imputation performance. nAncestralHaplotypes (k) = number of ancestral haplotypes; nGen = number of generations ago, controls recombination rate; nDiff_from_non-imputed = number of genotypes that were not the same between imputed and non-imputed samples; nMatch_from_non-imputed = number of genotypes that were the same between imputed and non-imputed sample; nMissing_from_non-imputed = these are SNPs and hence genotypes that are missing because that SNP failed QC for imputation; nAdditional_SNPs_with_called_genotypes_from_non-imputed = these are genotypes newly added by imputation).

nAncestralHaplotypes	nGen	nDiff_from_non-imputed	nMatch_from_non-imputed	nMissing_from_non-imputed	nAdditional_SNPs_with_called_genotypes_from_non-imputed
14	100000	36	1767	356	3057
8	1000	52	1765	342	3015
8	10000	37	1765	357	3010
8	100000	33	1757	369	2977
10	100000	35	1753	371	3013
10	10000	35	1752	372	3035
14	1000	41	1748	370	3077
10	1000	43	1741	375	3060
12	10000	39	1741	379	3052
14	10000	46	1739	374	3044
12	100000	34	1722	403	3041
6	100000	48	1721	390	2903
12	1000	48	1715	396	3092
14	100	46	1707	406	3197
6	1000	43	1698	418	2977
10	100	48	1693	418	3102
6	10000	42	1685	432	2952

6	100	65	1683	411	3031
8	100	53	1673	433	3079
12	100	53	1671	435	3122
4	1000	58	1659	442	2742
4	10000	69	1630	460	2772
4	100	103	1615	441	2776
4	100000	71	1581	507	2715

6.4 Article 4

*Article 4. **Lado, S.**, Elbers, J.P., Kilimci, F. S., Kara, M. E., Dabanoğlu, I., Hurk, Y., Brongers, T., Grigson, C., Lev-Tov, J., McClure, S., Davoudi, H., Mohaseb, A., Baker, P., Kühne, H., Kreppner, J., Haring, E., Berthon, R., Peters, J., Mashkour, M., Burger, P. A., Çakırlar, C. (*in preparation*). Hidden hybrids – detecting early hybridization between dromedary and Bactrian camels in a culture-historical context.*

Hidden hybrids – detecting early hybridisation between dromedary and Bactrian camels in a culture-historical context

Sara Lado¹, Jean P. Elbers¹, Figen Sevil Kilimci², M. Erkut Kara², İlknur Dabanoğlu²
Youri van den Hurk³, Thom Brongers³, Caroline Grigson⁴, Justin Lev-Tov⁵, Sarah McClure⁶, Hossein Davoudi⁷, Azadeh Mohaseb⁸, Polydora Baker⁹, Hartmut Kühne¹⁰,
Janoscha Kreppner¹¹, Elisabeth Haring¹² Rémi Berthon⁸, Joris Peters¹³, Marjan Mashkour^{8,14}, Pamela A. Burger^{1,*}, Canan Çakirlar^{4*}

¹Research Institute of Wildlife Ecology, Department of Interdisciplinary Life Sciences, Vetmeduni Vienna, Vienna, Austria

²Faculty of Veterinary Medicine, Aydın Adnan Menderes University, Turkey

³Institute of Archaeology, University of Groningen, Groningen, Netherlands

⁴Institute of Archaeology, University College London, UK

⁵R. Christopher Goodwin and Associates, Inc. in Frederick, Maryland

⁶Department of Anthropology, University of California, Santa Barbara, USA

⁷Archaeozoological section, Bioarchaeological laboratory, University of Tehran, Iran

⁸Archéozoologie, Archéobotanique: Sociétés, Pratiques et Environnements. CNRS / Muséum National d'Histoire Naturelle Département Ecologie et Gestion de la Biodiversité, Paris - France

⁹Historic England, UK

¹⁰Institute for Near Eastern Archaeology, Free University Berlin, Germany

¹¹Westfälische Wilhelms Universität Münster, Institut für Altorientalistik und Vorderasiatische Archäologie Department, Germany

¹²Central Research Laboratories, Natural History Museum Vienna, Austria

¹³Institut für Paläoanatomie, Domestikationsforschung und Geschichte der Tiermedizin, Tierärztliche Fakultät der LMU München, Munich, Germany

¹⁴Bioarchaeology Laboratory, Central Laboratory, University of Tehran, Tehran, Iran

* Corresponding authors: c.cakirlar@rug.nl and pamela.burger@vetmeduni.ac.at

Abstract

Hybridisation between dromedary and the closely related Bactrian camel produces more robust and enduring animals, adapted to a wider range of climatic conditions and able to traverse large geographical distances. Domestic camels, their hybrids and backcrosses have facilitated short and long-distance trade routes for millennia across Eurasia. As early empires reached a high level of connectivity, it has been suggested that the practice of camel interbreeding began in the early first millennium before Common Era (BCE), shortly after the two species were domesticated and their geographic ranges started overlapping. By this time, important commercial and political networks were well established across southwest Asia and North Africa, where other mammals have already been crossbred. In this study, several camel bone assemblages dating to the early Iron Age and more recent times were examined using morphological and ancient DNA techniques as well as low-coverage whole-genome shotgun sequencing. To detect possible hybrids from ancient samples, we created our own reference database, from modern whole-genome data. By radiocarbon dating of the genetically identified hybrids, we could detect the earliest evidence of dromedary-Bactrian hybridisation in a specimen dating to Early Iron I Age (1112 – 933 calibrated years before Common Era, calBC) from Hasanlu in northwestern Iran, one of the very important trading regions in ancient times. With a specimen from Trier, we also show that latest by the Medieval Period hybrid camels were present in Europe as today's western Germany. A direct-radiocarbon date on a purebred dromedary from Kinet (Turkey) confirms the occurrence of this species in the northern Levant (Bay of Alexandretta) as early as the 7th century BCE, during the Neo-Assyrian Period. Finally, large mixed-morphology camel specimens not clear-cut assigned by zooarchaeologists showed no genetic signs of hybridisation but turned out to be purebred dromedaries. This fundamental work lays the foundation to understanding the origins of these large mixed morphology phenotypes and the beginning of camel hybridization. In the future, it will be necessary to examine the chronological distribution and frequencies of early dromedaries, Bactrian camels and their hybrids across Eurasia.

Introduction

Natural hybridisation between two taxa is an important driver of evolution (*1*). In livestock, anthropogenic hybridisation between different populations, breeds, or species aims at creating a heterosis or hybrid vigour effect, *i.e.*, an increase in body mass, fertility

and production traits in the first (F1) generation (2). Consecutive interbreeding of F1/ F2 hybrids, however, often results in a reduction or even loss of fertility as well as other desirable traits, due to genetic incompatibilities and outbreeding depression (3, 4). Similar to other species, camel anthropogenic hybridisation is a human mediated crossing between two closely related Old World camel species, the one-humped dromedary (*Camelus dromedarius*) and the two-humped Bactrian camel (*Camelus bactrianus*). The crossing between these largest animals humans have ever domesticated follows very elaborate schemes of crossing and backcrossing with either parental species for high milk and wool yield (5, 6). The desired vigor effect, however, is only partly retained in further (back-) crossed individuals, and entirely lost in third-generation hybrids, F2 x F2 crosses (5). Hybrid camels have been instrumental in caravan trade and military campaigns in Medieval and Early Modern Era, and in some regions such as western Turkey, until the introduction of trucks in the 1960s (7). But hybrids are also “beloved pets”, considered as family members (8), like in western Turkey where F1 hybrids (*Tulus*) are trained to compete against each other in heavily regulated “camel wrestling” fights (8, 9).

While there is high interest in camels in the archaeological and historical records as well as in understanding the development and operation of cultural networks across Eurasia, one of the most remarkable actors of these networks – the hybrid camel – remains hidden (8, 10–14). In Old World camels, the practice of crossbreeding between dromedaries and Bactrian camel might have started as early as or maybe even pre-dating Roman times (8, 13, 15, 16) and this practice was associated with the transport of goods along the Silk Road. They were preferred over normal breeds, not only because they could carry significantly larger cargos over longer distances, but also because they are adapted to a wider variety of environments well beyond arid regions.

Bactrian camel domestication has been estimated to begin in the late fourth and early third millennium before common era either in Northeastern Iran and the adjacent Kopet Dagh foothills in southwestern Turkmenistan (17, 18) or in the Asian steppe farther to the East where humans were familiar with wild camels over an extended period of time (e.g., in Kazakhstan or northwestern Mongolia; rev. in (19)). Furthermore, dromedary’s domestication took place in the Arabian Peninsula most likely at the transition between the second and first millennia before the Common Era (BCE) (20). By the late 1th millennium BCE, dromedaries were present in the Negev Desert (20). Pictures, figurines,

bones and teeth of camels start appearing in the archaeological record from around 2500 BCE, with evidence for Bactrian camels occurring earliest in Eastern Iran and then early-domesticated dromedaries in Southern Arabia (20, 21). In principle, human-mediated inter-breeding between the two species may have occurred soon after their geographical distribution overlapped. Textual and pictorial evidence show that the two species encountered each other in Mesopotamia as early as 1000 BCE, with the linguistic distinction between them getting blurred by 800 BCE, and both becoming common between the Caucasus and Arabia by the first century BCE (6, 13, 22). Camels are commonly depicted in ancient art from this vast region, and pictorial evidence may very likely include depictions of hybrid camels (23). Yet, the earliest genetically deduced hybrid so far originated from a Roman archaeological site in Serbia, Viminacium, dated to approximately the late third to fourth centuries CE (24). The camel-borne incense trade, from Arabia to the Levant, was an important element in the economy of the eastern Mediterranean region in the 1st millennium BCE (25). Moreover, the idea of hybridizing domesticated animals has a longer history in southwest Asia. Long before Iron Age, mid- and late 3rd millennium BCE humans would already cross equids, like the “kunga”, a hybrid between a hemione (*Equus hemionus*) and a domestic donkey (*Equus asinus*) (26). Later, the offspring of a jack (*Equus asinus*) and a mare (*Equus caballus*), “mule”, were often present in Mesopotamian art of the 1st millennium BCE (27). Mules thrive on cheap food, have stronger working capacities and can carry more weight than horses, have longer life spans and are more resistant to disease (28). During the Iron Age, we can predict that humans were mostly breeders and likely experimented in a “try and fail” method by crossbreeding domestic species, including camels.

Hybrids are difficult to detect in the archaeological record. Zooarchaeology, which uses osteomorphology and osteometry (linear or geometric, GMM) of animal bone remains, investigates often-fragmented animal bones using mixed morphology as a trait (16). Detecting hybrids by using ancient DNA (aDNA) techniques is another possibility. Interestingly, a methodological framework, Zonkey pipeline, was developed to exploit high-throughput sequencing data retrieved from archaeological material, originally designed to identify F1-equine hybrids, although it can be applied to the identification of F1-hybrids in other groups (29). However, for archaeological specimens, the retrieval of aDNA sequences is far from routine, as DNA normally becomes degraded by nucleases and there might in fact very little and often no endogenous (target species specific) DNA

surviving the ancient sampling process (30). Some environmental conditions might preserve DNA for longer time, such as low temperature and dry environments, contrary to warm or wet environments (30). Recently, the oldest ancient DNA ever retrieved was reported, where genome-wide data from two mammoth specimens dating to more than one million years old was the recovered (31).

Fascinatingly, DNA that has been recovered from archaeological remains makes it possible to go back in time. As such, successfully accessing and extracting ancient samples from key sites, i.e., important ancient commercial networks located in present-day Iran, Turkey, Israel, or Syria, will provide a better perspective of the hybridisation history between the dromedary and Bactrian camel in a culture-historical context. This study investigates the early origins and geographic spread of hybrid camels in the 1st millennium BCE and afterwards. The aim of this study was to understand the historic distribution, ubiquity, and cultural significance of hybrid camels using archaeozoological and palaeogenetic methodologies. For this, we re-examined a number of relatively large and mixed-morphology camel bone assemblages from key archaeological sites from as early as the Iron I Age. Finally, and most importantly, zooarchaeological specimens were dated using relative and radiocarbon dating, and integrated in the culture-historical and archaeological context. This study is a step further into elucidating the evolution of an important technology in animal breeding, that of hybridisation.

Materials and Methods

Sample preparation and DNA extraction from ancient and historical and modern camel bones

A total of 41 camel bones from archaeological deposits dating from the Iron Age to the Medieval period in Turkey, Iran, Syria, Israel and Germany, were analyzed in the Paleogenetic Core Facility, ArchaeoBioCenter, LMU Munich, Germany (Supplementary Table S1). Sample selection was based on the combination of availability, relatively and date, preservation state, relatively large size, and mixed morphology (32). Using a drill, approximately 250 mg of bone powder was obtained from each sample for extraction. Standard precautions for ancient DNA work were taken, including the use of negative controls in all extraction steps and sterilization of equipment plastic wares and reagents using bleach and UV light. Bone samples were extracted following a combined protocol

from Rohland & Hofreiter (33) and Rohland et al. (34), as implemented in Almathen et al. (35). Extractions were conducted in batches of 10 samples including the presence of blank controls. For those 26 samples that were successfully amplified for mitochondrial DNA (mtDNA), a second extraction was performed at the ancient DNA lab facility at the Natural History Museum (NHM) Vienna and sent for whole-genome low-coverage sequencing to Daicel ArborBioscience (Ann Arbor, Michigan, USA). Authentication criteria for aDNA studies, such as multiple DNA extractions, independent PCR amplification and parallel extraction/ PCR controls were performed.

Additionally, we included four historical samples – two bones (GIA-2401, GIA-Hybrid-01), one tooth (GIA-5245) and one museum tissue (AC1908101). The three historical bone and tooth samples were drilled at the Paleogenetic Core Facility, ArchaeoBioCenter, LMU Munich, Germany. The following DNA extraction was performed inside the “clean room” at the Genetics Laboratory, Konrad Lorenz Institute of Comparative Behavioral Ecology (KLIVV), Vetmeduni, Vienna, using the same protocol as for the ancient samples. A modified salting-out extraction method (36) was applied on the museum tissue. Finally, 13 modern muscle tissue samples from Aydin, Turkey (from slaughterhouse) were processed in the Recombinant DNA and Recombinant Protein Center (REDPROM), Aydin Adnan Menderes University, Turkey, and the extracted genomic DNA was sent to the KLIVV genetics lab, Vetmeduni, Vienna for further genetic analysis. Total DNA was extracted from muscle tissue using a commercial kit (Qiagen, QIAamp DNA Micro Kit, Hilden Germany) according to the manufacturer’s manual (see Supplementary Table S1 for full samples information).

PCR amplification of mitochondrial and nuclear markers

To genetically confirm the ancestry of the camel samples, both the maternally inherited mtDNA and the nuclear DNA were recovered. For all the 58 ancient, historical and modern samples, a 148-base pair fragment of mitochondrial control region (nt 15345-115493; numbering according to GenBank accession number NC_009849.1) was amplified using two short overlapping PCRs (Ancient_mtDNA_1 and 2; Supplementary Table S2; (35)). For the identification of the species status of the *Camelus* specimen, the software FinchTV 1.4.0 (Geospiza, Inc.; Seattle, WA, USA; <http://www.geospiza.com>) was used to view DNA sequence chromatogram data. The sequences were then imported into BioEdit (37), aligned and compared the obtained mtDNA sequences with sequences

from dromedary (JX946206.1, KF719283.1 - KF719290.1, JX946273.1, KT334323.1, KT334322.1, KT334321.1, NC_009849.1) and Bactrian camel mitochondrion (MH109975.1, MH109982.1, MH109997.1, NC_009628.2). In addition, 11 nuclear species-specific SNP markers (38) were used to differentiate dromedaries, Bactrian camels and hybrids (F1 and F1 backcrosses) in historical and modern samples. Mitochondrial and nuclear primers information for species assignment based on mtDNA, and genotyped SNPs are indicated in Supplementary Table S2. PCR mixes for ancient and historical samples were performed in a PCR cabinet inside the “clean room” at the Genetics Laboratory, Konrad Lorenz Institute of Comparative Behavioral Ecology (KLIVV), Vetmeduni Vienna, and amplified using a Thermal cycler outside the clean room. Four replicate PCRs per ancient sample were sequenced to ensure consistency of sequence determination and one negative control was included in each PCR. Amplicons were purified using Exonuclease I and Shrimp Alkaline Phosphatase (USB® ExoSAP-IT® PCR Product Cleanup, Affymetrix). PCR products were Sanger sequenced in both directions on an automatic sequencer ABI3130xl Genetic Analyzer (Applied Biosystems) using BigDye v.1.1 (Thermo Fisher Scientific) chemistry, which is more adequate for short PCR products. Each sequence position was determined from two independent PCR amplifications (forward and reverse) to avoid sequence errors caused by template damage. For the SNP genotyping, the homozygote (alternative alleles fixed between dromedary and Bactrian camel) or heterozygote (alleles from both species) state of each locus was determined to identify pure dromedaries or Bactrian camels and hybrid individuals.

Low coverage whole genome shotgun sequencing

Genomic DNA (gDNA) from the 26 ancient DNA extractions performed at the NHM Vienna from samples, which were successfully sequenced for mtDNA (except for TNM-6108 which was not successfully sequenced, but showed mtDNA bands on the agarose gel) were sent to Daicel Arbor Biosciences’s for whole-genome shotgun sequencing. In addition, three historic bone samples, one museum tissue sample and two modern reference samples (one dromedary, three F1 hybrids, and one dromedary and one Bactrian camel backcross) were also sent to Daicel Arbor Biosciences’s. The total DNA was quantified via a spectrofluorimetric assay. Historical and modern samples were sonicated and size selected to produce an average insert length of approximately 175 bp. Long

fragments were removed from the degraded samples via bead purification. For the degraded samples, a double-stranded library preparation chemistry appropriate for short and degraded fragments was applied to the samples in a cleanroom setting. Unique dual-index combinations were added to each sample via 5-10 cycles of PCR amplification. The indexed libraries were quantified with both a spectrofluorimetric assay and a quantitative PCR assay. Samples were prepared into an equimolar pool for shotgun sequencing. For each sample, half of the volume of beads in the elution buffer were amplified for 10 cycles (modern/historical) or 12 cycles (ancient). Final pools were quantified again with both a spectrofluorimetric assay and a quantitative PCR assay. Samples were sequenced on the Illumina NovaSeq 6000 platform with 150-bp paired-end reads on partial S4 lanes. Demultiplexed Fastq data was delivered.

Hybrid detection from shotgun sequencing

Quality and adapter trimming was performed with BBDuk 38.86 (part of BBDuk/BBTools; <https://sourceforge.net/projects/bbmap/>) using the following settings “ktrim=r k=23 mink=11 hdist=1 tpe tbo qtrim=r trimq=15”, setting the reference as the adapters.fa file that comes with BBDuk. Optical duplicates were removed with Clumpify 38.86 (part of BBTools) using the following settings “dupedist=12000 dedupe=t optical=t”. PCR duplicates were removed with Dedupe 38.86 (part of BBTools) using default settings. In order to remove potential contamination, reads were mapped to a concatenated CamDro3 (dromedary)/ CamBac2 (Bactrian camel) reference using BBWrap 38.86 (part of BBTools) with the vslow setting, outputting reads mapping to the concatenated reference with the outm option as unaligned reads. BBWrap 38.86 runs BBDuk 38.86 but only loads the reference once for each sample. We repeated mapping reads from ancient samples to a concatenated CamDro3 (dromedary)/CamBac2(Bactrian camel) reference using BBWrap 38.86 with vslow setting, outputting reads mapping to the concatenated reference with the out (not outm) option as aligned reads. After, we ran mapDamage 2.2.1 (39) with the option –rescale to rescale base quality scores to identify DNA damage patterns typical for ancient DNA. Nucleotide misincorporations are commonly observed in ancient DNA. These are characterized by elevated C → T substitution near sequencing starts, and complementary increased G → A rates near the end. As shown in a previous study, for samples up to 117 years old, mapDamage analysis detected no signature of deamination in the mapped reads (40). Due to this, we have only

run this software on ancient samples. To quantify the percentage of dromedary for each sample we used modified versions for CamDro3 and CamBac2 (hereafter modified CamDro3 and modified CamBac2 references). For this, we computed high frequency 15-mers for CamDro3 with computeHighFreqKmers from Winnomap (Github commit d547331c2e0e843a7837b9d2f804dc920936ae6c

<https://github.com/marbl/Winnomap>). Next, we made pairwise alignments between CamDro3 and CamBac2 with Winnomap using -k 15 and -x asm5, and keeping only primary alignments with divergence values <0.02. We used dustmasker (1.0.0 Package: blast 2.7.1; (41)) to dustmask CamDro3 and CamBac2, extracted dustmasked alignments (made by Winnomap) with bedtools ‘getfasta’ 2.25.0 and converted dustmasked bases to Ns. We then mapped the dromedary reference mitochondrial sequence (GenBank accession: NC_009849.1) against the modified dromedary reference, and Bactrian camel mitochondrial reference (GenBank accession: NC_009628.2) against the modified Bactrian reference using blastn 2.2.31+. We converted bases where the mitochondria mapped in both modified dromedary and Bactrian camel references to N’s with bedtools ‘maskfasta’ 2.29.2. After, we mapped each sample’s reads simultaneously with BBSplit 38.86 (part of BBTools) to either modified CamDro3 and CamBac2 references using the following settings “minratio=1.0 qtrim=lr trimq=15 untrim=false ambiguous2=toss”. We calculated the percentage of dromedary as number of unambiguously mapped reads to the modified CamDro3 genome / (number of unambiguously mapped reads to modified CamBac2 genome + number of unambiguously mapped reads to modified CamDro3 genome).

We used an approach inspired in the Zonkey pipeline (29) on our shotgun sequencing and created our own reference database to detect hybrids, from modern whole-genome data (42). We started by downloading the first 1250000 reads from nine dromedary and seven Bactrian camel whole-genome sequencing samples (<https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP056721>). Then we performed quality and adapter trimming with BBDuk 38.86 using the following settings “ktrim=r k=23 mink=11 hdist=1 tpe tbo qtrim=rl trimq=15” and by setting the reference as the adapters.fa file that comes with BBDuk. We removed optical duplicates with Clumpify 38.86 using the following settings “dedupe=t optical=t”. We also removed PCR duplicates with Dedupe 38.86 using default settings. For each combination of nine dromedaries and seven Bactrian camels (9x7=63) we randomly subsampled 1000 reads:

1) 0% dromedary 100% Bactrian camel (Bactrian camel purebred); 2) 25% dromedary 75% Bactrian camel (Bactrian backcross); 3) 50% dromedary 50% Bactrian camel (F1 hybrid); 4) 75% dromedary 25% Bactrian camel (Dromedary backcross); 5) 100% dromedary 0% Bactrian camel (Dromedary purebred). We chose 1000 random input simulation reads, as the total number of unambiguously mapped reads to either the modified CamDro3 or CamBac2 genomes was between 268-402 reads for simulation, which was less than the lowest value seen in the real samples. By using BBSplit 38.86 (part of BBTtools) we mapped each simulation's and actual samples' reads simultaneously to either modified CamDro3 and CamBac2 references using the following settings "minratio=1.0 qtrim=lr trimq=15 ambiguous2=toss". We then calculated the percentage dromedary as before. We then compared actual samples to simulated samples by calculating Tukey's fences for "extreme values" for each group of 0, 25, 50, 75, and 100% dromedary simulations values for percent dromedary as upper outlier value = 3rd quartile + (3rd quartile – 1st quartile) x 1.5 and lower outlier value = 1st quartile – (3rd quartile – 1st quartile) x 1.5. We compared the percentage of dromedary from actual samples to upper and lower outlier values and actual samples were considered as 1) 0% dromedary 100% Bactrian camel (Bactrian camel purebred); 2) 25% dromedary 75% Bactrian camel (Bactrian backcross); 3) 50% dromedary 50% Bactrian camel (F1 hybrid); 4) 75% dromedary 25% Bactrian camel (Dromedary backcross); 5) 100% dromedary 0% Bactrian camel (Dromedary purebred), and see if they fit within the upper and lower ranges for 0,25,50,75, or 100% dromedary. We used R 3.6.3 (R core team 2019) to determine if the actual values fit within the ranges of outliers based on the simulations.

Radiocarbon dating

We selected the specimens from well-stratified archaeological contexts, but they all come from multi-period sites where mixing of contexts is the norm. As the camel dating project showed (21), claims for early domestic forms should be confirmed by direct radiocarbon dating. In this study, five specimens (KT001, MM505, MM302, HAS822, HAS2023) of high interest in terms of aDNA, osteomorphological characteristics, location and possible date were probed for radiocarbon dating at the Centre for Isotope Research of the University of Groningen. For the samples MM302, HAS822, HAS2023, ¹⁴C ages (in yrBP) are calibrated to calendar years with software program OxCal, version 4.3 (43) and used the calibration curve IntCal13 (44). For the sample KT001, calibrated dating results

^{14}C ages (in yrBP) are calibrated to calendar years with software program OxCal, version 4.4 (43) and used the calibration curve IntCal20 (45). Moreover, two specimens from Tell Nebi Mend already were radiocarbon dated by the camel dating project (21).

Results

Sanger sequencing success of ancient and historical camel bone samples

Working with ancient DNA can be very challenging; especially in samples collected from warm climatic regions the DNA is very poor in quality and quantity due to molecular damage and contamination (30). From the 41 ancient samples, 25 were successfully sequenced for mtDNA, and one additional showed a specific (weak) PCR product band on agarose gel. Those 26 samples included mainly originated from Syria, Turkey, Israel, Iran, and Germany, and were dated earliest to the Early Iron Age. Alignment results confirmed that all ancient samples showed dromedary maternal ancestry, except two (HAS2845 and HAS3467 from Iran dated in Iron Age III; Table 1, Supplementary Table S3), which had the maternal lineage of Bactrian camel. Similarly, we found dromedary maternal ancestry in the historical and modern samples, except for one zoo specimen (GIA2401) from the Netherlands and a museum sample (AC1908101) from Kazakhstan (Supplementary Table S3).

For the historical and modern samples, the eleven diagnostic SNPs presented the homozygous pattern of pure dromedary for the modern controls (GIA 5245 and Aydin 16) (Supplementary Table S3). Furthermore, we detected eleven F1 hybrids with a dromedary maternal lineage in samples from Turkey (GIA-Hybrid-01 and Aydin2 – Aydin 10 and Aydin 14, dromedary female x Bactrian male) and one with a Bactrian camel female Kazakhstan (AC1908101, dromedary male x Bactrian female), expressing the expected heterozygous profile for all loci (except for one locus in five Aydin samples). The dromedary backcrosses, Aydin 1 and Aydin 15, showed a homozygous profile for only five and six loci respectively, remaining heterozygous in the other loci, while the Bactrian backcross (GIA-2401) showed heterozygosity for two loci (18%) remaining homozygous for the Bactrian allele in the other loci (Supplementary Table S3)

Detecting ancient dromedary-Bactrian camel hybridisation using shotgun sequencing

We shotgun sequenced 26 ancient samples at low coverage which showed to have endogenous DNA based on mtDNA, as well as six historical/ modern samples. The initial

total number of reads for all the 26 ancient samples, and the six modern and historic samples was 187,147,290, where after quality adapter trimming, optical and PCR duplicate removal, we ended up with a total of 170,696,362. In the end, the total number of reads mapped was 11,845,334, where the average percentage of endogenous nuclear DNA for the ancient samples was 0.54% (SD = 1.38) (see Supplementary Table S4 for proportion of reads mapping unambiguously to modified dromedary or Bactrian camel genomes). Shotgun sequencing results of nuclear DNA for historical/modern samples confirmed the results from Sanger sequencing of eleven nuclear species-specific SNPs, confirming the presence of three F1 hybrids, one purebred dromedary, and one dromedary and one Bactrian backcrosses, respectively (Supplementary Table S4). Within the 26 ancient samples that were shotgun sequenced, we detected four samples showing different levels of hybridisation. Three were classified as dromedary backcrosses: two samples from Iran in Iron I age (HAS822) and Parthian time (MM302) and one from Germany in Roman/Medieval Age (GIA5356), and one as Bactrian camel backcross (HAS2023) (Table 1, Supplementary Table S4). All the others were identified as dromedary except for HAS2845 and HAS3467 from Iron Age III, which were assigned as Bactrian camels. To authenticate the sequences obtained as nuclear endogenous DNA, we ran the 26 ancient samples with and without mapDamage 2.2.1 (39) to identify DNA damage patterns typical for ancient or degraded DNA. Estimated species classification – with and without running mapDamage software – were the same for all samples except for HAS3467 which did not fit within the upper and lower ranges for any of the categories without mapDamage (Supplementary Table S4).

Table 1. Mitochondrial and nuclear DNA species assignment, approximate age and radiocarbon dating for the ancient samples which were successfully amplified.

Ancient sample	Locality	mtDNA	nuDNA (shotgun seq)	Approximate age period	Radiocarbon dating
85/553/0006	Tell Sheikh Hamad, Syria	Dromedary	Dromedary	Middle/ Late Iron Age	-
85/6151/0039	Tell Sheikh Hamad, Syria	Dromedary	Dromedary	Middle/ Late Iron Age	-
88/9377/0056	Tell Sheikh Hamad, Syria	Dromedary	Dromedary	Middle/ Late Iron Age	-
89/8981/0337	Tell Sheikh Hamad, Syria	Dromedary	Dromedary	Middle/ Late Iron Age	-
AS-001	Ashkelon, Israel	Dromedary	Dromedary	Medieval Period	-
GIA5356	Trier, Germany	Dromedary	Dromedary backcross	Roman/ Medieval Period	-
HAS822	Tepe Hasanlu, NW Iran	Dromedary	Dromedary backcross	Iron I	1051 – 908 calBC
HAS2023	Tepe Hasanlu, NW Iran	Dromedary	Bactrian backcross	Iron I	1112 – 933 calBC
HAS2845	Tepe Hasanlu, NW Iran	Bactrian	Bactrian	Iron III	-
HAS3467	Tepe Hasanlu, NW Iran	Bactrian	Bactrian	Iron III	-
KT-001	Kinet, Turkey (close to Syria)	Dromedary	Dromedary	Middle / Late Iron Age	790 – 544 calBC
KT-002	Kinet, Turkey (close to Syria)	Dromedary	Dromedary	Medieval Period	-
MM302	Shahr-i-Qumis, Iran	Dromedary	Dromedary backcross	Parthian Period	347 – 535 calAD
MM315	Shahr-i-Qumis, Iran	Dromedary	Dromedary	Parthian Period	-
MM322	Shahr-i-Qumis, Iran	Dromedary	Dromedary	Parthian Period	-
MM505	Shahr-i-Qumis, Iran	Dromedary	Dromedary	Parthian Period	No collagen
Tepe3465-164	Kilise Tepe, Turkey	Dromedary	Dromedary	Late Iron Age	-
TJ-007	Tell Jemmeh, Israel	Dromedary	Dromedary	Late Iron Age	-
TJ-129	Tell Jemmeh, Israel	Dromedary	Dromedary	Late Iron Age	-
TJ-159	Tell Jemmeh, Israel	Dromedary	Dromedary	Late Iron Age	-
TNM-4170	Tell Nebi Mend, Syria	Dromedary	Dromedary	Hellenistic/ Roman	359 calBC - 71 calAD (Grigson 2014)
TNM-4171	Tell Nebi Mend, Syria	Dromedary	Dromedary	Late Iron Age/ Hellenistic	506 - 5 calBC (Grigson 2014)
TNM-6108	Tell Nebi Mend, Syria	-	Dromedary	Iron Age / Hellenistic	-
TNM-6717	Tell Nebi Mend, Syria	Dromedary	Dromedary	Hellenistic	-
TUP-001	Tupras, Turkey (close to Syria)	Dromedary	Dromedary	Islamic Period in Hatay	-
ZTf00441	Ziyarettepe, Turkey (close to Syria)	Dromedary	Dromedary	Medieval Period	-

Radiocarbon dating

We radiocarbon dated three early-hybrids (HAS822, HAS2023, MM302) and two dromedaries (MM505, KT001). Results showed that one of the five samples (MM505) did not have enough and high quality collagen for proper analysis. One as Bactrian camel backcross Iron I Age from Iran (1112 – 933 calBC, HAS2023), and two samples which were classified as dromedary backcrosses from Iran were confirmed to be in Iron I Age (1051 – 908 calBC, HAS822) and the Parthian Period (347 – 535 calAD, MM302). One dromedary sample from the Mediterranean coast of Turkey close to Syria confirmed to be in Middle/ Late Iron Age (790 – 544 cal BC, KT-001) (Figure 1, Supplementary Table

S5). Two dromedary samples from western Syria which were previously radiocarbon dated previously in Grigson (2014) (359 calBC - 71 calAD, TNM4170; 506 - 5 calBC, TNM4171), date to the Persian/ Hellenistic periods (Supplementary Table S5).

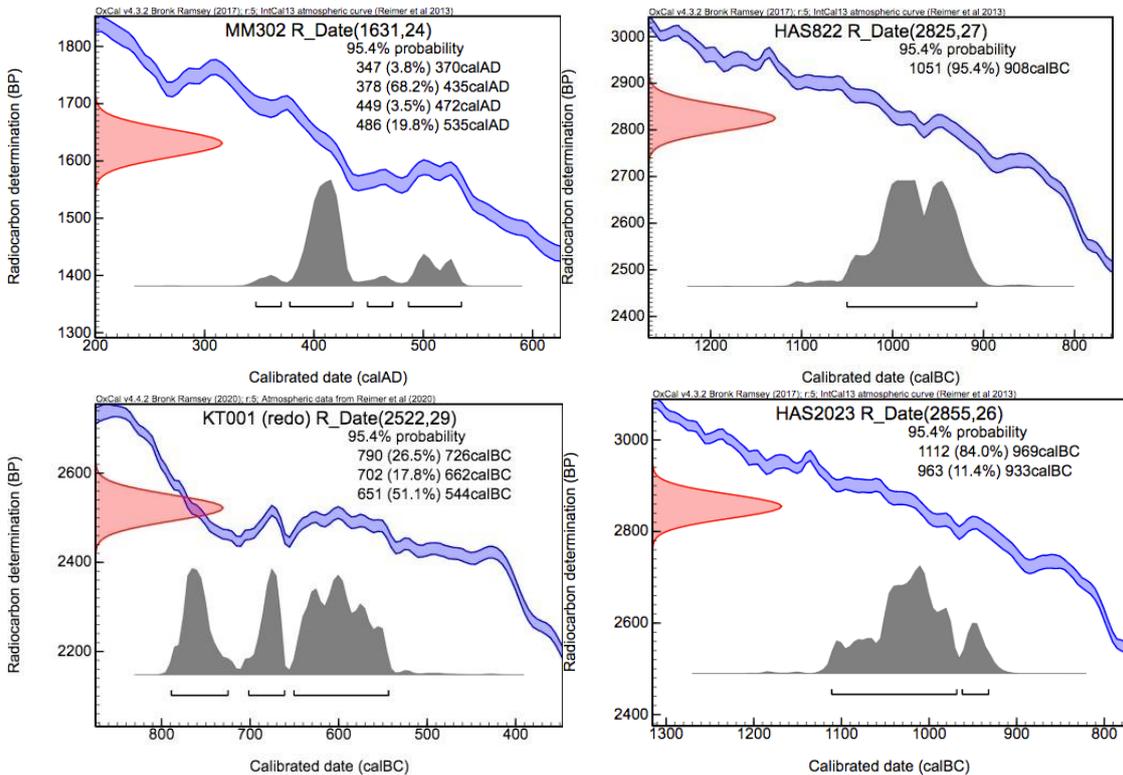


Figure 1 - Calibrated dating results, for the samples MM302, HAS822, KT001 and HAS2023. Ages are calibrated to calendar years. The 95.4% (2σ) probability range demonstrated in the figures is based on the ^{14}C measurement results. The date range indicates the time periods matching the measured ^{14}C value at this level of probability.

Discussion

The aim of camel hybridisation was producing animals combining the robustness of the Bactrian camel and the endurance of the dromedary, capable to adapt to a wide variety of environments. Nowadays, hybrid camels are important among pastoralists in Central Asia and Iran, as well as pets for camel wrestling in western Turkey (8, 46). The two domestic species' distribution areas currently overlap in Central and western Asian countries such as Iran, Afghanistan, Kazakhstan, and India. Today, hybridisation between dromedary and Bactrian camel is a widespread practice in Central Asia, not only as a way to breed

more robust individuals, but also to increase milk, wool and meat production compared to parental species (5, 13, 47). There are a wide number of types of crossing (Bactrian camel male or female x dromedary female or male) with important consequences on the behavior and physiology of the animals. Also, alternate crossing is possible in order to maximize the heterosis and increase milk production (dromedary effect), fatty matter (Bactrian effect), wool productivity (Bactrian effect) and cold (Bactrian effect) or high temperature (dromedary effect) resistance (5).

History of hybrid camel

The camel-borne incense trade, from Arabia to the Levant, was an important element in the economy of the eastern Mediterranean region in the 1st millennium BCE (25). Historical studies show that camel hybrids have been prized beasts of burden in short- and long-distance trade, and were taken to other continents in military campaigns, leading to the formation of camel cultures across Eurasia (5, 7, 8, 46). Several zooarchaeologists have considered the presence of hybrids among archaeological camel remains from southwest Asia without aDNA tools (12, 16, 48–50). The earliest of these claims is on a camel burial from Eastern Arabia, dating to the Hellenistic Period (ca. 300 BCE) (16). Potts (13) suggested that hybridisation could have taken place earlier in the Neo-Assyrian period (800 – 700 BCE) in Iran, facilitated by the far-reaching imperial routes. Potts (2004) took his inspiration from Uerpmann (16), and based his argument mainly on textual and artistic records. In this study, we tested this hypothesis using palaeogenetic methods, and investigated the presence and spread of hybrid camels across southwest Asia during the Iron Age and to a lesser extent later.

We genotyped modern (from Aydin, Turkey) and historical hybrids based on eleven nuclear SNPs fixed in each species (38). Although these eleven sites are expected to be species-specific, they represent a very small part of the genome. Nevertheless, we retrieved coherent results between the species assignment from these eleven sites and from the low-coverage whole-genome shotgun sequencing approach (which covers more areas of the nuclear genome) in those samples, for which both methods were applied. Moreover, we used archaeozoological methods to identify hybrids in bone remains, i.e., four samples identified with different levels of hybridisation (backcrosses) were directly radiocarbon dated to reveal the chronological sequence of the emergence and early spread of hybrid camels.

Detection of early hybrids from Iran in the Iron Age

Notably, our earliest evidence for hybrid camels was identified as Bactrian camel backcross from Iran dated to the Early Iron I Age (1112 – 933 calBC, HAS2023), in Iran, Hassanlu, where important commercial networks were located. This is a few centuries earlier than previous assumptions that pointed at camels hybridizing in the Iranian plateau around the Neo-Assyrian Period (ca. 800 BCE), based on interpretation of various textual and picture sources (13). Trading routes from Asia via northern Iran and Anatolia towards Europe were cold routes, even when it was not winter season, due to high altitude (e.g., Silk road crossing in Elburz Mountains in Iran – North and South route). Thus, caravan animals operating this route would have been adapted to cold environments. In that sense, the here-detected frequency of cold-adapted Bactrian camels (HAS2845, HAS3467 in Iron III Age) and Bactrian camel backcrosses in northern Iran can be expected. On the other hand, we also identified three samples as dromedary backcrosses, of which two originated from Iran: one dated to Iron I Age (1051 – 908 calBC, HAS822) and the other to the Parthian Period (347 – 535 calAD, MM302). Dromedary backcrosses could provide a higher milk yield, which would be a direct benefit for nomadic people. Also, in Iran, zooarchaeologists have suggested that dromedary was already present in the late 2nd millennium BCE levels at Tepe Sagzabad (northwest Iran) (51). Moreover, the presence of Bactrian camel in Assyria from the end of the 2nd through the middle of the 1st millennium BCE evidence suggests that Bactrian camels were imported into Assyria from various parts of Iran at this time (Potts 2004). Dromedary domestication seems to have taken place around 1200 BCE on the southeast coast of the Arabian Peninsula (35), only one century earlier than what we have detected the hybrids in northern Iran. One possible explanation is that one century might be enough so the camels can travel all the way up to the northern Iran. But our radiocarbon date also indicates that one possibility that needs to be investigated further is that Bactrian domestication influenced dromedary domestication. According to Potts (13), the information passing could have been through the Persian Gulf.

Presence of large dromedaries with mixed-morphologies

From the ancient samples included in this study, all other camels were identified as dromedaries. One of them (KT-001) originated from Kinet (South Turkey close to Syria) and was direct-radiocarbon dated confirming the occurrence of this species in the northern

Levant (Bay of Alexandretta) as early as 7th century BCE, during the Neo-Assyrian Period. Moreover, four dromedary samples from the Middle/ Late Iron Age were located in Tell Sheikh Hamad (Syria), not far from Hasanlu (Iran). Although Tell Sheikh Hamad and Hasanlu were connected via imperial trade routes, they were located in very different environmental zones. Based on osteomorphology, these samples were previously identified as Bactrian camels (52) due to their very large mixed-morphology, different from early-domesticated dromedaries in Arabia (e.g., (35)). Additionally, camel specimens that we sampled from Tell Jemmeh (Israel), located farther in Southern Levant, also show mixed-morphologies, although they were not so large (not shown; (53)). Our genetic analysis revealed them as purebred dromedaries dating to the Persian Period, around the 5th century BCE.

In general, the extinct ancestor of modern dromedaries was shown to be larger than the modern individuals (54). One possible explanation for the presence of these large and mixed-morphology phenotypes in our samples might be that they resemble a former wild type, or an intermediate phenotype between wild and domestic dromedaries, possibly resulting from introgression during an early domestication phase when restocking from the wild was likely (35). Previous studies show this pattern in other animals such as cattle, sheep or pig (e.g., (55, 56)), in which through time (before breeds were formed), animals fluctuated in size, showing regional differences. Thus, further research is necessary to understand the origins of these very large dromedary phenotypes that no longer exist.

Presence of hybrid camels in medieval times in Europe

Archaeological camel finds in Central Europe are not as unusual as one might expect. They cover a chronological span from the Roman period (50 BCE – 700 CE) until the early Modern Age (ca. 17th CE) (12). The latest hybrid detected in our study (GIA5356, dromedary backcross) stems from an old excavation in Trier (western Germany) and might date to the Roman Period or the Medieval Period. Exact dates need to be confirmed by radiocarbon dating because contextual evidence for this bone is missing. The hybridisation process allows hybrid camels to be more adaptable to different environments, which accelerates the spread of these camels across climate zones, including Europe. Based on osteological and zooarchaeological information, previous studies suggested that parts of the camels found in the northern provinces of the Roman Empire, would possibly be hybrid camels (12). As dromedaries are desert animals, the

most plausible reason of why camels could travel so far northern, in wetter and colder environments outside their normal expected habitat, would be due to their hybrid acquired aptitudes. As such, the Trier specimen is the northernmost hybrid camel confirmed with ancient DNA techniques. Previously, a 17th century camel hybrid was detected in Tulln (Austria), indicating that camel hybrids could indeed exist in northern regions (57). As Trier was a Roman town, not a military city, future studies should focus on retracing whether hybrids were either bred in northern areas or only brought to those regions.

Conclusions

Once the camel hybridisation through human mediation happened, the spread of camels via the Levant into other (harsh) environments has been rapid. Part of the bio-cultural success of camels might stem from this hybridisation process. This study was important to reveal early hybridisation combining interdisciplinary methods and enlightening a common journey between men and camels. Here, we detected the earliest camel hybrid by analyzing archaeological camel bones from the Iron Age in southwest Asia by using a combination of methods such as zooarchaeology, palaeogenetics, radiocarbon dating. Corroborated by historical and archaeological evidence, our findings suggest that hybridisation occurred as early as the 11th century BCE, prior to the Neo-Assyrian Period, and shortly after the domestication of the dromedary. We have also detected a specimen from Roman/ Medieval Trier in Germany, showing that hybrid camels were probably common and widespread in long-distance exchange in the Roman Period. Nevertheless, osteomorphological characterization of hybrids based on modern skeletons with known parentage can help identifying archaeological hybrids when endogenous DNA is not well-preserved or too fragmented. In the near future, osteomorphology and osteometry of hybrid camels will be assessed using recent specimens from western Turkey, including the individuals from Aydin (Turkey) we genotyped in this study. Future work should also be dedicated to untangling the domestication history of the Bactrian camel, which is still largely unknown. Research is also necessary to examine in detail the chronological distribution and frequencies of dromedaries, Bactrian camels and their hybrids across Eurasia. Finally, this work opens doors to future large-scale studies that include a larger suite of methods like isotopic analyses and a much larger sampling, and a more comprehensive network.

Bibliography

1. E. J. Baack, L. H. Rieseberg, A genomic view of introgression and hybrid speciation. *Curr. Opin. Genet. Dev.* **17**, 513–518 (2007).
2. J. A. Birchler, H. Yao, S. Chudalayandi, Unraveling the genetic basis of hybrid vigor. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12957–12958 (2006).
3. S. Edmands, Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution (N. Y.)*. **53**, 1757–1768 (1999).
4. F. W. Allendorf, G. Luikart, *Conservation and the genetics of populations* (John Wiley & Sons, 2009).
5. B. Faye, G. Konuspayeva, in *Camels in Asia and North Africa*, E.-M. Knoll, P. A. Burger, Eds. (Academy of Science Press, Vienna, Austria, 2012), pp. 27–33.
6. S.-A. Al-Zaidi, Betwixt and Between the Bactrian Camel and the Dromedary: The Semantic Evolution of the Lexeme *udru* during the 11th to 8th Centuries bce. *Arab. Epigr. Notes.* **3**, 11–18 (2017).
7. O. İnal, One-humped history: The camel as historical actor in the late ottoman empire. *Int. J. Middle East Stud.*, 1–16 (2020).
8. C. Çakırlar, R. Berthon, Caravans, camel wrestling and cowrie shells: towards a social zooarchaeology of camel hybridization in Anatolia and adjacent regions. *Anthropozoologica.* **49**, 237–252 (2014).
9. S. Manav, A. Koc, A. Cagli, M. Yilmaz, in *5th Conference of ISOCARD 'Recent Advances in Camelids: Biology, Health and Production'* (Ed. by A. Sghiri & F. Kichou), pp. 483–5. *Institut Agronomique et Veterinaire Hassan II, Rabat.* (2018).
10. R. Berthon, M. Mashkour, P. Burger, C. Çakırlar, in *Les vaisseaux du désert et des steppes: Les camélidés dans l'Antiquité (Camelus dromedarius et Camelus bactrianus)*, D. Agut-Labordère, B. Redon, Eds. (MOM Editions, 2020; <https://doi.org/10.4000/books.momeditions.8457>), pp. 21–26.
11. P. A. Burger, The history of Old World camelids in the light of molecular genetics. *Trop. Anim. Health Prod.* **48**, 905–913 (2016).
12. F. Pigière, D. Henrotay, Camels in the northern provinces of the Roman Empire. *J. Archaeol. Sci.* **39**, 1531–1539 (2012).
13. D. T. Potts, Camel hybridization and the role of *Camelus bactrianus* in the ancient Near East. *J. Econ. Soc. Hist. Orient.* **47**, 143–165 (2004).
14. M. Mashkour, The funeral rites at Mleiha (Sharja - U.A.E.) ; the camelid graves. *Anthropozoologica.* **25**, 725–736 (1997).
15. P. Steinkeller, in *Exploring the Longue Durée : Essays in Honor of Lawrence E. Stager*, J. D. Schloen, Ed. (Winona Lake, Indiana, Eisenbrauns, 2009), vol. 31, pp. 5–17.
16. H. P. Uerpmann, Camel and horse skeletons from protohistoric graves at Mleiha in the Emirate of Sharjah (U.A.E.). *Arab. Archaeol. Epigr.* **10**, 102–118 (1999).

17. R. W. Bulliet, *The Camel and the Wheel* (Columbia University Press, ed. 19, 1975).
18. N. Benecke, *Archäozoologische Studien zur Entwicklung der Haustierhaltung in Mitteleuropa und Südkandinavien von den Anfängen bis zum ausgehenden Mittelalter (Vol. 46)*. (Deutsches Archäologisches Institut., 1994).
19. P. A. Burger, E. Ciani, B. Faye, Old World camels in a modern world – a balancing act between conservation and genetic improvement. *Anim. Genet.* **50**, 598–612 (2019).
20. L. Sapir-Hen, E. Ben-Yosef, The introduction of domestic camels to the southern Levant: Evidence from the Arava Valley. *Tel Aviv.* **40**, 277–285 (2013).
21. C. Grigson, The history of the camel bone dating project. *Anthropozoologica.* **49**, 225–235 (2014).
22. M. Herles, in *Symbolische Repräsentation und Wirklichkeit nomadischen Lebens.*, U. Pietruschka, M. Streck, Eds. (Nomaden un., 2010), pp. 127–167.
23. N. O. Brusgaard, *Carving interactions: Rock art in the nomadic landscape of the Black Desert, north-eastern Jordan* (Archaeopress Publishing Ltd., 2019).
24. P. A. Burger, S. Lado, E. Mohandesan, S. Vuković-Bogdanović, J. Peters, C. Çakırlar, in *Second International Selçuk-Ephesus Symposium On Culture Of Camel-Dealing And Camel Wrestling; JAN 18-20, 2018; Selcuk, Turkey. Volume II Natural And Applied Science Health And Medical Science; (ISBN: 978-605-88682-9-8)* (2018), pp. 153–159.
25. M. Artzy, Incense, Camels and Collared Rim Jars: Desert Trade Routes and Maritime Outlets in the Second Millennium. *Oxford J. Archaeol.* **13**, 121–147 (1994).
26. J. . Postgate, in *Equids in the Ancient World.*, H.-P. Uerpmann, R. H. Meadow, Eds. (Ludwig Reichert Verlag, Wiesbaden, 1986), p. 194e206.
27. J. Clutton-Brock, *A natural history of domesticated mammals* (Cambridge University Press, 1999).
28. W. B. Tegetmeier, C. L. Sutherland, *Horses, asses, zebras, mules and mule breeding* (H. Cox, London, 1895).
29. M. Schubert, M. Mashkour, C. Gaunitz, A. Fages, A. Seguin-Orlando, S. Sheikhi, A. H. Alfarhan, S. A. Alquraishi, K. A. S. Al-Rasheid, R. Chuang, L. Ermini, C. Gamba, J. Weinstock, O. Vedat, L. Orlando, Zonkey: A simple, accurate and sensitive pipeline to genetically identify equine F1-hybrids in archaeological assemblages. *J. Archaeol. Sci.* **78**, 147–157 (2017).
30. M. Hofreiter, D. Serre, H. N. Poinar, M. Kuch, S. Pääbo, Ancient DNA. *Nat. Rev. Genet.* **2**, 353 (2001).
31. T. van der Valk, P. Pečnerová, D. Díez-del-Molino, A. Bergström, J. Oppenheimer, S. Hartmann, G. Xenikoudakis, J. A. Thomas, M. Dehasque, E. Sağlıcan, F. R. Fidan, I. Barnes, S. Liu, M. Somel, P. D. Heintzman, P. Nikolskiy, B. Shapiro, P. Skoglund, M. Hofreiter, A. M. Lister, A. Götherström, L. Dalén, Million-year-old DNA sheds light on the genomic history of

- mammoths. *Nature*. **591**, 265–269 (2021).
32. C. Çakırlar, S. Kilimci, in *Second International Selçuk-Ephesus Symposium on Culture of Camel-Dealing and Camel Wrestling: Social Sciences. vol. 1, Selcuk, pp. 46-50. D Ertürk & Ö Gökdemir (eds)*. (2018).
 33. N. Rohland, M. Hofreiter, Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2**, 1756–1762 (2007).
 34. N. Rohland, D. Reich, S. Mallick, M. Meyer, R. E. Green, N. J. Georgiadis, A. L. Roca, M. Hofreiter, Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biol.* **8**, 16–19 (2010).
 35. F. Almathen, P. Charruau, E. Mohandesan, J. M. Mwacharo, P. Orozco-Wengel, D. Pitt, A. M. Abdussamad, M. Uerpmann, H.-P. Uerpmann, B. De Cupere, P. Magee, M. a. Alnaqeeb, B. Salim, A. Raziq, T. Dessie, O. M. Abdelhadi, M. H. Banabazi, M. Al-Eknah, C. Walzer, B. Faye, M. Hofreiter, J. Peters, O. Hanotte, P. A. Burger, Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc. Natl. Acad. Sci.* **113**, 6707–6712 (2016).
 36. S. B. De Volo, R. T. Reynolds, M. R. Douglas, M. F. Antolin, An improved extraction method to increase DNA yield from molted feathers. *Condor.* **110**, 762–766 (2008).
 37. T. A. Hall, BioEdit : a user-friendly biological sequence alignment editor and analysis program for Windows 95 / 98 / NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).
 38. E. Ruiz, E. Mohandesan, R. R. Fitak, P. Burger, Diagnostic single nucleotide polymorphism markers to identify hybridization between dromedary and Bactrian camels Diagnostic single nucleotide polymorphism markers to identify hybridization between dromedary and Bactrian camels (2015), doi:10.1007/s12686-015-0420-z.
 39. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* **29**, 1682–1684 (2013).
 40. A. McGaughan, Effects of sample age on data quality from targeted sequencing of museum specimens: What are we capturing in time? *BMC Genomics.* **21**, 1–10 (2020).
 41. A. Morgulis, E. M. Gertz, A. A. Schäffer, R. Agarwala, A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13**, 1028–1040 (2006).
 42. R. R. Fitak, E. Mohandesan, J. Corander, A. Yadamsuren, B. Chuluunbat, O. Abdelhadi, A. Raziq, P. Nagy, C. Walzer, B. Faye, P. A. Burger, Genomic signatures of domestication in Old World. *Commun. Biol.* **3**, 1–10 (2020).
 43. C. B. Ramsey, Bayesian analysis of radiocarbon dates. *Radiocarbon.* **51**, 337–360 (2009).

44. P. Reimer, E. Bard, A. Bayliss, J. W. Beck, P. Blackwell, C. B. Ramsey, C. Buck, H. Cheng, R. L. Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, H. Hafliðason, I. Hajdas, C. Hatté, T. J. Heaton, D. L. Hoffmann, A. G. Hogg, K. A. Hughen, K. F. Kaiser, B. Kromer, S. W. Manning, M. Niu, R. W. Reimer, D. A. Richards, E. M. Scott, J. R. Southon, R. A. Staff, C. S. M. Turney, J. van der Plicht, *Radiocarbon*. **55**(4) 1869-1887 (2013).
45. P. J. Reimer, W. E. N. Austin, E. Bard, A. Bayliss, P. G. Blackwell, C. Bronk Ramsey, M. Butzin, H. Cheng, R. L. Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, I. Hajdas, T. J. Heaton, A. G. Hogg, K. A. Hughen, B. Kromer, S. W. Manning, R. Muscheler, J. G. Palmer, C. Pearson, J. Van Der Plicht, R. W. Reimer, D. A. Richards, E. M. Scott, J. R. Southon, C. S. M. Turney, L. Wacker, F. Adolphi, U. Büntgen, M. Capano, S. M. Fahrni, A. Fogtmann-Schulz, R. Friedrich, P. Köhler, S. Kudsk, F. Miyake, J. Olsen, F. Reinig, M. Sakamoto, A. Sookdeo, S. Talamo, The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0-55 cal kBP). *Radiocarbon*. **62**, 725–757 (2020).
46. M. Dioli, Dromedary (*Camelus dromedarius*) and Bactrian camel (*Camelus bactrianus*) crossbreeding husbandry practices in Turkey and Kazakhstan: An in-depth review. *Pastoralism*. **10** (2020), doi:10.1186/s13570-020-0159-3.
47. V. Blagovescenskii, Reserves in the production of milk and meat. *Konevod. Konnyi Sport*. **33**, 8–9 (1963).
48. J. Studer, A. Schneider, in *Archaeozoology of the Near East VIII Tome II*, L. G. E. Vila, A. M. Choyke, H. Buitenhuis, Eds. (Paris, 2008), pp. 581–596.
49. I. Köhler-Rollefson, in *Pella of the Decapolis*, R. H. Smith, I. P. Day, Eds. (College of Wooster Art Museum, Wooster, 1989), vol. 2, pp. 142-163.
50. F. M., Archaeozoological Remains from the Lower Sanctuary. A Preliminary Report on the 1994 Excavations. *Stud. Troica*, 217–227 (1996).
51. M. Mashkour, Chasse et élevage au nord du Plateau central iranien entre le Néolithique et l'Âge du Fer. *Paléorient*. **28**, 27–42 (2002).
52. C. Becker, in *Umwelt und Subsistenz der assyrischen Stadt Dur-Katlimmu am unteren Habur, Berichte der Ausgrabung von Tall Seh Hamad / Dur-Katlimmu (BATSH) 8*, K. H, Ed. (Harrassowitz Verlag, Wiesbaden:, 2008), pp. 61–131.
53. P. Wapnish, Camel caravans and camel pastoralists at Tell Jemmeh. *J. Anc. Near East. Soc. Columbia Univ. New York, NY*. **13**, 101–121 (1981).
54. E. Mohandesan, C. F. Speller, J. Peters, H. P. Uerpmann, M. Uerpmann, B. De Cupere, M. Hofreiter, P. a. Burger, Combined hybridization capture and shotgun sequencing for ancient DNA analysis of extinct wild and domestic dromedary camel. *Mol. Ecol. Resour*. **17**, 300–313 (2017).
55. M. MacKinnon, Cattle ‘breed’ variation and improvement in Roman Italy: connecting the zooarchaeological and ancient textual evidence. *World Archaeol*. **42**, 55–73 (2010).
56. U. Abarella, in *Bones and the Man: Studies in honour of Don Brothwell*, K. Dobney, T. O’Connor, Eds. (Oxbow Books , Oxford, 2002), pp. 51–62.

57. A. Galik, E. Mohandesan, G. Forstenpointner, U. M. Scholz, E. Ruiz, M. Krenn, P. Burger, A sunken ship of the desert at the river danube in Tulln, Austria. *PLoS One*. **10**, 1–16 (2015).

Acknowledgements

C.C. acknowledges funding from Wenner-Gren Foundation for Anthropological Research Project #9519. We would like to thank technician Christian Küchelmann (then of GIA), Jens Rohde (then of Free University Berlin). S.L. and J.P.E. acknowledge funding from the Austrian Science Funds (FWF) project P29623-B25 to P.B.

Author contributions

S.L. performed wet lab work and wrote the first draft of the paper, I.D. extracted DNA from Aydin modern samples, S.L. and J.P.E. analyzed genetic data, C.C and P.B. conceived and managed the project. F.S.K., M. E. K., YvdH, T.B., C.G., J.L-T., S.C., H.D., A. M. P. B. H.K., J. K., R.B., contributed with essential samples. S.L., C.C. and P.B. wrote the manuscript. All authors provided valuable discussions.

Data availability

Raw reads from whole-genome shotgun sequencing will be deposited at ENA.

Additional material requests can be addressed to c.cakirlar@rug.nl and pamela.burger@vetmeduni.ac.at

Supplemental Material**Supplementary Table S1.** Complete samples' information.

	Sample ID	Assigned Species	Dating/period	Location	Donor	Bone type	Details	mtDNA	Shotgun Seq	RC Dating
	Ancient									
1	85/553/0006	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Calcaneous	Broken pieces of shaft	X	X	
2	85/6151/0039	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Phalanx 2	Complete	X	X	
3	85/6151/0115	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Tibia	broken pieces			
4	85/9177/0026	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Radius	Broken piece of the shaft			
5	86/9377/0024	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Mandibula				
6	88/9377/0056	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Phalanx 1	Complete	X	X	
7	89/8779/0200	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Calcaneous	a piece of the infused proximal shaft			
8	89/8981/0337	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Phalanx 1	Complete	X	X	
9	97/6949/0078	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Scapula	a piece of broken blade			
10	97/6951/0099	Camelidae	Iron Age	Tell Sheikh Hamad, Syria	Çakırlar/Kreppner/Kühne	Humerus	2 fragments of epiphysis			
11	AS-001	Camelidae	Persian Period	Ashkelon, Israel	Çakırlar	Metatarsus		X	X	
12	GIA5356	Camelidae	Roman/Medieval Period	Trier, Germany	Çakırlar	Metatarsus	Complete	X	X	
13	KT-001	Camelidae	Middle/ Late Iron Age	Kinet Höyük, Turkey	Çakırlar	Cranial + petrous bone	Half	X	X	X
14	KT-002	Camelidae	Medieval Period	Kinet Höyük, Turkey	Çakırlar	Phalanx1	Almost complete	X	X	
15	Qu-001	Camelidae	Late Roman/Medieval	Jebel Qurma, Jordan	Çakırlar	Calcaneous	Distal bit			
16	TJ-008	Camelidae	Persian Period	Tell Jemmeh, Israel	Çakırlar	Astragalus	Complete			
17	TJ-027	Camelidae	Persian Period	Tell Jemmeh, Israel	Çakırlar	Phalanx 2	Complete			
18	TJ-078	Camelidae	Persian Period	Tell Jemmeh, Israel	Çakırlar	Radius	Distal			
19	TJ-103	Camelidae	Persian Period	Tell Jemmeh, Israel	Çakırlar	Phalanx 1	complete			
20	TJ-115	Camelidae	Persian Period	Tell Jemmeh, Israel	Çakırlar	Calcaneous	Complete			
21	TJ-127	Camelidae	Persian Period	Tell Jemmeh, Israel	Çakırlar	Metatarsus	Proximal			
22	TUP-001	Camelidae	Islamic Period in Hatay	Tupras	Çakırlar	Phalanx 1	Complete	X	X	
23	ZTR00441	Camelidae	Medieval		Labo Paris	32 Rd	Epi	X	X	
24	TNM-4170	Camelidae	Hellenistic/ Roman	Tell Nebi Mend, Syria	Grigson/ Çakırlar	Lunate	Complete	X	X	Grigs on 2014
25	TNM-4171	Camelidae	Late Iron Age/ Hellenistic	Tell Nebi Mend, Syria	Grigson/ Çakırlar	Humerus	Distal	X	X	Grigso n 2014
26	TNM-6108	Camelidae	Iron Age / Hellenistic	Tell Nebi Mend, Syria	Grigson/ Çakırlar	Phalanx 1	Prox		X	
27	TNM-6717	Camelidae	Hellenistic	Tell Nebi Mend, Syria	Grigson/ Çakırlar	Phalanx 2	Distal	X	X	
28	TB-6034-777	Camelidae	Medieval/ Ottoman	Tell Burak, Lebanon	Çakırlar	Metatarsus	Distal			

29	AK-01	Camelidae	Persian Period	Akko, Israel	Çakırlar					
30	HAS822	Camelidae	Iron I	Tepe Hasanlu, NW Iran	Marjan/ Azadeh/ Davoudi	Ph1		X	X	X
31	HAS2023	Camelidae	Iron I	Tepe Hasanlu, NW Iran	Marjan/ Azadeh/ Davoudi	Ph1		X	X	X
32	HAS2845	Camelidae	Iron III	Tepe Hasanlu, NW Iran	Marjan/ Azadeh/ Davoudi	Hm		X	X	
33	HAS3467	Camelidae	Iron III	Tepe Hasanlu, NW Iran	Marjan/ Azadeh/ Davoudi	Ph1		X	X	
34	MM302	Camelidae	Parthian period	Iran, Shahr-i-Qumis, Iran	Marjan/ Azadeh/ Davoudi	Humerus	R	X	X	X
35	MM315	Camelidae	Parthian period	Iran, Shahr-i-Qumis, Iran	Marjan/ Azadeh/ Davoudi	Radius	L	X	X	
36	MM322	Camelidae	Parthian period	Iran, Shahr-i-Qumis, Iran	Marjan/ Azadeh/ Davoudi	Radius	L	X	X	
37	MM505	Camelidae	Parthian period	Iran, Shahr-i-Qumis, Iran	Marjan/ Azadeh/ Davoudi	Phalanx		X	X	X
38	TJ-007	Camelidae	Persian Period	Tell Jemmeh, Israel	Çakırlar	Os carpale secundum	Complete, left	X	X	
39	TJ-129	Camelidae	Persian Period	Tell Jemmeh, Israel	Çakırlar	Os carpi accessorium	Complete, Right	X	X	
40	TJ-159	Camelidae	Persian Period	Tell Jemmeh, Israel	Çakırlar	Tibia	Distal, right	X	X	
41	Tepe3465-164	Camelidae	Late Iron Age	Kilise Tepe, Turkey	Polydora Baker	Carpal Hamatum	Left	X	X	
	Historic/Modern									
42	GIA-2401	Bactrian	1984	Zoo specimen	Çakırlar	Sesamoid		X	X	
43	GIA-5245	Dromedary	1974	Jordan, roadkill	Çakırlar	Tooth		X	X	
44	GIA-Hybrid-01	Hybrid	2008	Aydin, Turkey	Çakırlar	Astragalus		X	X	
45	AC1908101	F1 hybrid	1908	Kazakhstan	Berthon	Tissue	Female	X	X	
46	Aydin1	Drom backcross	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
47	Aydin2	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
48	Aydin3	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
49	Aydin4	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
50	Aydin5	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
51	Aydin6	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
52	Aydin7	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
53	Aydin8	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
54	Aydin9	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
55	Aydin10	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		
56	Aydin14	F1 hybrid	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X	X	
57	Aydin15	Drom backcross	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X	X	
58	Aydin16	Dromedary	Modern	Aydin, Turkey	Dabanoglu/ Kilimci/ Kara	Tissue		X		

Supplementary Table S2. Mitochondrial and nuclear primers information.

Primer Name	Sequence (5' to 3')	Tm ° C	Product size (bp)
Ancient_mtDNA_F1	RCCACACCCTCCCTAAGACT	60.51	92
Ancient_mtDNA_R1	CGGAGGTCAGGGGGTAGT	59.91	
Ancient_mtDNA_F2	CACCCAAAGCTGGAATTCTT	59.17	100
Ancient_mtDNA_R2	GGCATGAYATGTGGTTTTTAG	58.01	

Primer	Scaffold	Position	Reference	Alternate	Left Start	Left Length	Left Tm ° C	Left GC%	Left Sequence	Right Start	Right Length	Right Tm ° C	Right GC%	Right Sequence	Length
HP405	NW_006211106 .1	5030405	G	C	126	20	59.22	55	CCAGGAGCTTTTCGAGTAGC	250	20	57.22	55	CAGCACAGAGAACTCACTGC	125
HP900	NW_006211075 .1	294900	G	C	121	20	59.11	55.00	CCACATGCTCAGGTATCTGG	245	20	59.70	55.00	GGGATTCCTTGTGCTACAGC	125
HP288	NW_006211252 .1	214288	T	A	121	20	59.84	55.00	GTCTATGAGGGCGTTTCTGC	245	20	58.65	50.00	CAGCCTTCTTGTCTGTTCG	125
HP597	NW_006211126 .1	6065597	A	C	116	20	59.69	45.00	ATGAACAGTTTGGGTTGGG	240	20	58.27	50.00	CGCGATGTCACCTTATAGG	125
HP264	NW_006211022 .1	519264	T	C	124	22	59.23	45.45	TGGACAGAACTTGTGTCTCC	248	20	59.76	45.00	TTTGGTAAGGCATGAATCC	125
HP206	NW_006210212 .1	501206	C	G	124	21	58.94	47.62	TGTCAGACTGTTAGGCATTC	248	22	59.83	50.00	CATCCAAGTCTCCATCTAATCC	125
HP429	NW_006210666 .1	2288429	A	C	155	20	57.24	50.00	GCAGGCATACAACTAATCC	279	20	59.19	50.00	GCTTTTCTTCTGGCTCAGG	125
HP633	NW_006210489 .1	4279633	G	C	126	22	58.79	45.45	GCATGTAGAAGGTTGCATAGG	250	19	59.53	52.63	CAGCCTTCTTGCATCTGG	125
HP930	NW_006210745 .1	3459930	C	A	123	20	58.62	50.00	CTCCAGGAAACAAAAGTCC	247	20	58.85	50.00	TTTGGGAGTGTCTGTCTGC	125
HP379	NW_006210464 .1	1151379	G	A	111	19	59.44	52.63	AGGATGCCATCATGTCAGG	260	21	57.96	52.38	GAGGGAGCTCATGAATAGG	150
HP501	NW_006211169 .1	218501	A	T	142	20	58.74	50	GAATAGATTGGGGAGCAAGC	266	20	58.69	55	CTCTTCTCCATCCCTATGGC	125

Supplementary Table S3. Sanger sequencing results. The parentheses indicate the reference alleles (dromedary allele | Bactrian camel allele) identified from the whole-genome sequencing.

Sample type	Sample ID	mtDNA	Nuclear	HP206	HP288	HP379	HP405	HP429	HP501	HP597	HP633	HP930	HP264	HP900
				(G C)	(A T)	(A G)	(C G)	(C A)	(T A)	(C A)	(C G)	(A C)	(C T)	(C G)
Historical/ modern														
	GIA-2401	Bact	Bactrian Backcross	CC	TT	AG	GG	AA	AA	AA	GG	AC	TT	GG
	GIA-5245	Drom	Dromedary	GG	AA	AA	CC	CC	TT	CC	CC	AA	-	CC
	GIA-Hybrid-01	Drom	F1 Hybrid	GC	AT	AG	CG	CA	TA	CA	CG	AC	CT	CG
	AC1908101	Bact	F1 Hybrid	GC	AT	AA	CG	CA	TA	CA	CG	AC	CT	CG
	Aydin1	Drom	Dromedary Backcross	GG	AT	AG	CG	CC	TA	CA	CC	AA	CT	CC
	Aydin2	Drom	F1 Hybrid	GC	AT	AG	CG	AC	TA	CA	CG	AC	CT	CG
	Aydin3	Drom	F1 Hybrid	GC	AT	AG	CG	AC	TA	CA	CG	AC	CT	CG
	Aydin4	Drom	F1 Hybrid	GC	AT	AG	CG	AC	TA	CA	CG	AC	TT	CG
	Aydin5	Drom	F1 Hybrid	GC	AT	AG	CC	AC	TA	CA	CG	AC	CT	CG
	Aydin6	Drom	F1 Hybrid	GC	AT	AG	CG	AA	TA	CA	CG	AC	CT	CG
	Aydin7	Drom	F1 Hybrid	GC	AT	AA	CG	AC	TA	CA	CG	AC	CT	CG
	Aydin8	Drom	F1 Hybrid	GC	AT	AG	CG	AC	TA	CA	CG	AC	CT	CG
	Aydin9	Drom	F1 Hybrid	GC	AT	AG	CG	AC	TA	CA	CG	AC	CT	CG
	Aydin10	Drom	F1 Hybrid	GC	AT	AG	CG	AC	TA	CA	CG	AC	CT	GG
	Aydin14	Drom	F1 Hybrid	GC	AT	AG	CG	AC	TA	CA	CG	AC	CT	CG
	Aydin15	Drom	Dromedary Backcross	GG	AT	AA	CG	CC	TT	CA	CC	AA	CT	CG
	Aydin16	Drom	Dromedary	GG	AA	AA	CC	CC	TT	CC	CC	AA	CC	CC
Ancient														
	85/6151/0039	Drom												
	89/8981/0337	Drom												
	AS-001	Drom												
	GIA5356	Drom												

TUP-001	Drom												
ZTf00441	Drom												
KT-001	Drom												
KT-002	Drom												
85/553/0006	Drom												
88/9377/0056	Drom												
TNM-4170	Drom												
TNM-4171	Drom												
TNM-6108	-												
TNM-6717	Drom												
HAS822	Drom												
HAS2023	Drom												
HAS2845	Bact												
HAS3467	Bact												
MM302	Drom												
MM315	Drom												
MM322	Drom												
MM505	Drom												
TJ-007	Drom												
TJ-129	Drom												
TJ-159	Drom												
Tepe3465-164	Drom												

Supplementary Table S4. Hybrid detection from low-coverage whole-genome shotgun sequencing for modern, historical and ancient samples (with and without mapdamage analysis).

Sample	Sample reads	Sample bases	Unambiguously mapped reads Dromedary	Unambiguously mapped reads Bactrian camel	Unambig Mapped reads Total	Percent Dromedary	Classification
Historical/ modern							
AC1908101	70956	7991020	13690	13486	27176	50.3753	F1 hybrid
Aydin14	3266208	485042190	912252	782986	1695238	53.8126	F1 hybrid
Aydin15	6490084	960251909	2274156	1056538	3330694	68.2787	Dromedary backcross
GIA-2401	1032646	150067044	134996	398622	533618	25.2982	Bactrian backcross
GIA-5245	4246	319802	2512	104	2616	96.0245	Dromedary
GIA-Hybrid-01	694010	99958279	186980	153490	340470	54.9182	F1 hybrid
Ancient							
85_553_0006	2480	133094	1430	42	1472	97.1467	Dromedary
85_553_0006-mapdamage	2520	113617	1470	42	1512	97.2222	Dromedary
85_6151_0039	5216	343142	3060	138	3198	95.6848	Dromedary
85_6151_0039-mapdamage	5230	300143	3054	142	3196	95.5569	Dromedary
88_9377_056	3238	207397	1848	72	1920	96.25	Dromedary
88_9377_056-mapdamage	3238	179641	1852	72	1924	96.2578	Dromedary
89_8981_0337	2278	137211	1362	44	1406	96.8706	Dromedary
89_8981_0337-mapdamage	2328	123720	1406	48	1454	96.6988	Dromedary
AS-001	4052	303767	2110	128	2238	94.2806	Dromedary
AS-001-mapdamage	4320	288075	2226	132	2358	94.402	Dromedary
GIA5356	3534	331119	1456	432	1888	77.1186	Dromedary backcross
GIA5356-mapdamage	4270	396547	1704	546	2250	75.7333	Dromedary backcross
HAS822	3714	308557	1384	536	1920	72.0833	Dromedary backcross
HAS822-mapdamage	4358	324413	1470	714	2184	67.3077	Dromedary backcross
HAS2023	48174	6136074	6268	16088	22356	28.0372	Bactrian backcross
HAS2023-mapdamage	60326	7356098	6806	19502	26308	25.8705	Bactrian backcross
HAS2845	116682	14094999	8472	40010	48482	17.4745	Bactrian
HAS2845-mapdamage	154772	18625983	9730	51814	61544	15.8098	Bactrian
HAS3467	32232	3994813	3304	10792	14096	23.4393	Unknown
HAS3467-mapdamage	42310	5113778	3724	13972	17696	21.0443	Bactrian
KT-001	7446	596698	3852	186	4038	95.3938	Dromedary
KT-001-mapdamage	8104	607430	4082	206	4288	95.1959	Dromedary
KT-002	4140	274126	2362	86	2448	96.4869	Dromedary
KT-002-mapdamage	4146	247649	2374	84	2458	96.5826	Dromedary
MM302	8308	690841	2896	910	3806	76.0904	Dromedary backcross
MM302-mapdamage	10564	862526	3368	1268	4636	72.6488	Dromedary backcross
MM315	2428	140504	1438	48	1486	96.7699	Dromedary
MM315-mapdamage	2494	124681	1508	48	1556	96.9152	Dromedary
MM322	5482	384414	3088	100	3188	96.8632	Dromedary
MM322-mapdamage	5504	336533	3086	98	3184	96.9221	Dromedary
MM505	4638	281658	2656	110	2766	96.0231	Dromedary

MM505-mapdamage	4778	263127	2784	120	2904	95.8678	Dromedary
T3465164	2994	225372	1640	72	1712	95.7944	Dromedary
T3465164-mapdamage	3202	217644	1736	78	1814	95.7001	Dromedary
TJ-007	790	47353	452	12	464	97.4138	Dromedary
TJ-007-mapdamage	812	43536	466	10	476	97.8992	Dromedary
TJ-129	2364	151796	1250	56	1306	95.7121	Dromedary
TJ-129-mapdamage	2392	131370	1276	58	1334	95.6522	Dromedary
TJ-159	5196	342818	2868	134	3002	95.5363	Dromedary
TJ-159-mapdamage	5222	300435	2876	140	3016	95.3581	Dromedary
TNM-4170	2838	172698	1586	56	1642	96.5895	Dromedary
TNM-4170-mapdamage	3354	230330	1902	64	1966	96.7447	Dromedary
TNM-4171	2676	230648	1396	64	1460	95.6164	Dromedary
TNM-4171-mapdamage	3116	252262	1604	76	1680	95.4762	Dromedary
TNM-6108	1046	16933	708	28	736	96.1957	Dromedary
TNM-6108-mapdamage	1012	19958	726	22	748	97.0588	Dromedary
TNM-6717	5532	345487	3104	128	3232	96.0396	Dromedary
TNM-6717-mapdamage	5738	330893	3220	150	3370	95.549	Dromedary
TUP-001	5942	510773	3086	176	3262	94.6045	Dromedary
TUP-001-mapdamage	6570	523670	3346	206	3552	94.2005	Dromedary
ZTf00441	3764	257006	2148	116	2264	94.8763	Dromedary
ZTf00441-mapdamage	3834	221392	2178	116	2294	94.9433	Dromedary

Lower Outlier Percent	Upper Outlier Percent	Simulated Percent Dromedary - Classification
8.246	22.807	0 – Bactrian camel
24.945	44.756	25 – Bactrian backcross
45.696	63.285	50 – F1 hybrid
66.164	82.262	75 – Dromedary backcross
86.580	99.556	100 - Dromedary

Supplementary Table S5. Radiocarbon dating analysis results including information on collagen quality and stable isotope ratios. MM505 did not yield sufficient amount of collagen. GrM nr = Groningen Measurement number; BP = before present; C:N = carbon and nitrogen ratio; IRMS = isotope-ratio mass spectrometer; F14C = fraction of radiocarbon (¹⁴C); Isotope-ratio mass spectrometer (IRMS); ±1σ = probability range.

Sample name	Dated material	GrM nr	F14C	±1σ	14C Age (yrBP)	±1σ	Collagen Yield (%)	% C	% N	C:N	δ13C (‰;IRMS)	±1σ	δ15N (‰;IRMS)	±1σ
KT001	Collagen	24061	0.731	0.003	2522	29	5	36.9	13.6	3.2	-19.32	0.15	6.84	0.3
MM302	Collagen	22153	0.816	0.002	1631	24	1.8	38	14	3.2	-19.2	0.15	8.32	0.3
HAS822	Collagen	22154	0.704	0.002	2825	27	2.2	39.6	14.7	3.2	-19.01	0.15	7.87	0.3
HAS2023	Collagen	22155	0.701	0.002	2855	26	11.1	41.6	15.4	3.2	-18.86	0.15	8.3	0.3

7. DISCUSSION & CONCLUSION

The history of camels has been influenced by humans from their (early) domestication until today. Yet, the value of camels as productive livestock animal in times of global warming and growing deserts has been neglected so far by scientists and policy makers alike. Camels are multipurpose animals, and no other domestic animal is able to provide such a variety of services (meat, milk and wool production, leisure, transportation) to the society, especially in harsh environmental conditions (1). The camel population has also experienced steady growth in Asian countries, such as Syria, Saudi Arabia, Oman, as well as Sahel countries and Horn of Africa, as the demand and production of camel meat and milk intensifies (2). In this context, the combination of human-related climate change, population growth, decline in biodiversity, and land-use change are major drivers for the evolution and spread of zoonotic disease (3). In my thesis, with four articles, I describe in detail how I followed up on previously identified knowledge gaps and characterized patterns of (immuno)genome diversity in camels influenced by ancient human trading, as well as the response to pathogens, and how I was able to identify early-hybrids in an archaeological and culture-historical context. Now, in this thesis discussion, I will highlight not only the importance of (hybrid) camels in the (*present*) modern society, but also show their relevance from a (*past*) culture-historical context to a (*future*) global warming era. As it is imperative to create awareness on the (hybrid) camel value, I will also discuss the importance in maintaining the present (immuno)genetic diversity for their capability to respond to future challenges like increasing temperatures, growing deserts or emerging diseases and how this can influence human life in marginal economic zones.

Importance of camels in immunology and public health

The camel genome anchors several unique variations, being the main reason behind camel's ability to survive under extreme environmental conditions (4). The role of camels as a mean of transport and as food resource is appreciated daily by people living in the desert in Asia and Africa. Moreover, in these regions, camel products such as milk or meat (or urine) are not only consumed as food, but are perceived as remedies in several human diseases (5). Camel milk and urine are used for the treatment of skin problems, chronic hepatitis, stomach infections, infectious diseases, certain cardiovascular conditions, strengthen the human immune system, to reduce the growth of cancer cells or

to cure autism (5, 6). Furthermore, the dromedary, as well as the dromedary-Bactrian camel hybrid (hereafter hybrid camel), are adapted to arid lands and low nutritive natural resources, making it one of the less sensitive animals during drought. Their impact on land and water resources for food production is less than that of any other livestock species, and with increasing desertification and global climate change their importance will grow even more (7) [**Article 1**]. The promotion of the (hybrid) camel as an animal with a deep history also has relevance for efforts to conserve its genetic and phenotypic diversity for a sustainable utilization by livestock pastoralists who are challenged by climate change and competition from global markets. Protecting and maintaining livestock pastoralism is crucial to human well-being, food security, and the sustainability of cultural diversity.

Camelids are not only characterized by their remarkable adaptation to harsh environments and production trait potential or alternative medicine benefits, but also by their extraordinary immunology and important role in fighting infectious diseases (8). Using a genome-wide approach in Lado et al. (7) [**Article 1**], we screened for loci deviating from neutrality and identified sixteen F_{ST} outliers to be putatively under selection between African and Asian dromedaries. We examined (potentially linked) regions 200 kb upstream and downstream of the F_{ST} -outlier loci and detected fifty-three genes related to a number of biological functions where around one fifth of the detected genes had functions related to the immune system, hinting to an adaptive process in response to different pathogens in the respective environments.

Notably, camel immunology is unique and supports the fight against *Coronaviridae* (including the Covid-19 pandemic) with one fascinating feature: the unusual configuration of their single-domain antibodies, so called *nanobodies*. All Old World and New World (*Lama glama*, *Lama guanicoe*, *Vicugna pacos* and *Vicugna vicugna*) camelid species produce IgG homodimeric heavy-chain immunoglobulins without a light chain and with the antigen-binding fragment reduced to a single heavy-chain variable domain, in addition to the conventional antibodies (9). The unique features of these nanobodies, which are more stable types of antibodies, make them especially useful in biotechnology and for clinical applications that could be more effective for fighting diseases (10). Coronaviruses make use of a large envelope protein called “spike” to engage host cell receptors and catalyse membrane fusion. A previous study showed that generated llama

nanobodies were able to bind to the spike protein of the SARS-CoV-1 and MERS-CoV (11). What was particularly encouraging was that researchers also demonstrated that engineered nanobodies could neutralize these viruses, and also SARS-CoV-2, *in vitro*. Because of the vital role that these spike proteins play, they represent a vulnerable target for the development of therapeutics. Also, the isolation and characterization of a nanobody derived from alpaca has been reported that specifically targets the receptor-binding domain of the SARS-CoV-2 spike glycoprotein and effectively neutralizes the virus (12). As such, camelid nanobodies are promising candidates for antiviral therapy to neutralise betacoronaviruses.

The visible impacts of climate change on camels include the expansion of the geographical distribution of the species, a higher integration of camels in mixed crop-livestock systems and the increased risk of emerging (zoonotic) diseases (13-15). Different studies demonstrated a high seroprevalence of antibodies to a variety of zoonotic pathogens in camel populations along with current and past examples of camel-human transmission (rev. in 16). Back in 1998, a famous American epidemiologist mentioned the *Coronaviridae* as the most dangerous viral family due to their ability to mutate and reassert (17). This statement turned out to be very true, as new coronavirus diseases have since emerged, like SARS-CoV-1 in 2003, MERS-CoV in 2012, and more recently in 2020, the COVID-19 pandemic took the world by surprise. As the consumption of camel milk and meat is rising and camel products enter wider markets, the impact of camel-associated zoonotic diseases on public health and economy also grows. In this context, the emergence of MERS-CoV might have been caused by a combination of such factors. Since camels have been pointed out as important disease reservoirs, they need to receive attention. The MERS-CoV transmission from camels to humans takes multiple ways, such as airborne (droplet) infection (18), via camel urine, and food-borne, through the consumption of unpasteurised milk and raw meat (14). Regarding MERS-CoV epidemiology, a variety of host factors associated with disease susceptibility and virus transmission have been identified, including the virus entry receptor (dipeptidyl peptidase-4 [DPP4]), presumed attachment factors, sialic acids, host proteases, interferons, interferon-stimulated genes, and adaptive immune response (rev. in 19, 20-22). While a number of studies have been conducted in humans, the actual reservoir species, the dromedary is still poorly studied regarding susceptibility, resistance or immune genetic response towards MERS-CoV infection. Recent releases of

chromosome assembled camel genomes (23) include also detailed information on immune response genes, as I show in my second Article (24) [**Article 2**]. Furthermore, in the third Article (25) [**Article 3**] presented in this thesis, we show high seroprevalence of virus-specific antibodies which I aimed at understanding patterns of immunogenetic diversity in dromedaries in response to MERS-CoV infection. For this, 100 IR genes identified in the most up-to-date dromedary genome annotation CamDro3 (24) [**Article 2**]) were assessed for genetic variation potentially associated with MERS-CoV recent infection in seropositive dromedaries from the UAE. Specifically, I have detected variation in candidate genes with important functions in the adaptive – MHC-class I (*HLA-A-24*-like) and II (*HLA-DPB1*-like) – and innate immune response (*PTPN4*, *MAGOHB*), and in cilia coating the respiratory tract (*DNAH7*). Some of these genes have previously been associated with viral replication in SARS-CoV-1/-2 in humans, others have an important role in the movement of bronchial cilia.

Due to the selective pressure exerted by host immune systems, many viruses have evolved proteins that interfere with the antigen presentation by MHC class I molecules, by using a whole variety of strategies to inhibit the MHC class I pathway (26). The MHC is divided into three distinct classes: the class I, II and III. Especially the human leukocyte-associated antigen (HLA) genes, belonging to class I and II, have been recognized for their importance in both disease risk and resistance in humans (20, 27). HLA polymorphisms have been linked to susceptibility and pathogenesis of numerous infectious diseases including those caused by RNA viruses, such as influenza or HIV, as well as diseases that also affect camels such as rabies or West Nile fever (20). The influence of HLA gene polymorphisms for SARS-CoV susceptibility, pathogenesis, and outcome has been investigated (predominantly in Asian human populations) and associations between HLA genes and the development and/or severity of SARS-CoV have been found in certain populations (20). More specifically, a protective effect of *HLA-A*02:01* against SARS-CoV-1 has been suggested in Asian patients (28, 29), while *HLA-A*24:02* has been associated with COVID-19 susceptibility (30). Similarly, as described in Lado et al. (25) [**Article 3**], the *HLA-A-24*-like sequence harbours significant variants in potential association with MERS-CoV infection in dromedaries. However, the identified sequence might display a not fully functional classical MHC class I gene as the exon 2 sequence is missing in both currently available chromosome assembled reference genomes, the dromedary (CamDro3; (24) [**Article 2**]) and the wild camel (*Camelus ferus*

(23)). Therefore, I could not exclude the possibility that we sequenced a pseudogene or a misassembled chimeric sequence. The SNPs significantly associated with the presence of the MERS-CoV in seropositive camels were mainly distributed in intronic regions except for the MHC class II gene *HLA-DPBI*-like, where we found one SNP in exon 2 and another in exon 4, respectively (25) [Article 3]. Exon 2 encodes the antigen-binding groove of the class II molecule and therefore, its polymorphism is of functional importance. Exon 4 codes for the transmembrane domain that controls membrane domain partitioning and class II structure, both of which influence antigen presentation and T-cell activation (31). Although antigen presentation of SARS-CoV-1 mainly depends on MHC class I molecules (32), class II genes can also contribute to betacoronaviridae antigen presentation as suggested by the association of *HLA-DRB1**11:01 and *HLA-DQB1**02:02 alleles with susceptibility to MERS-CoV (33).

It is important to acknowledge how particular pathogens affect immune genetic diversity as well as how genetic variation influences adaptation to emerging zoonosis, habitat fragmentation, and climate change (34). MHC genes play an important role in the adaptive branch of the immune system and have been used extensively to estimate levels of adaptive genetic variation (35). In Lado et al. (25) [Article 3], I estimated that MHC class I mean diversity (H_0) was significantly lower compared to killer cell genes over all dromedaries. Low levels of genetic diversity in the MHC region have also been observed in wild and domestic two-humped camels (36). Although in this study, the authors looked specifically into the antigen-binding sites and not into complete genes where, according to Lado et al. (24) [Article 2], additional diversity appears to be present. Interestingly, a lower overall genomic heterozygosity was described in dromedaries compared to wild and domestic Bactrian camels (37), which could hint to a generally lower genetic diversity in dromedaries. However, in Lado et al. (24) [Article 2] genome-wide analyses of IR genes found a higher mean nucleotide diversity in MHC class I and II genes of dromedaries and domestic Bactrian camels compared to other adaptive or innate IR genes, as well as to the rest-of-genome genes. While adaptive (or acquired) immunity is a highly specific immune response and its variability is subject to different selective pressures, innate immunity is an efficient first protection against many pathogens but rather less specific (38, 39). In Lado et al. (25) [Article 3], I have also detected two innate candidate IR genes (*PTPN4*, *MAGOHB*) potentially associated to MERS-CoV infection in dromedaries, both with important roles in virus defense mechanisms. *PTPN4* is related to

predicted target functions of human micro(mi)RNAs that bind to the single-stranded (ss)-RNA such as SARS-CoV-2; and possibly to its spike protein gene. These predicted miRNA targets might destabilize the ss-RNA translation of SARS-CoV-2 in respiratory epithelial cells, which could explain successful antiviral defense (40). Interestingly, *MAGOHB* is targeted by has-miR-20a-5p, one of six miRNAs that previously have been reported to be anti-viral in respiratory diseases, and were found to be down regulated in lung tissues during viral infection (41, 42), as well as has-miR-20a-5p was identified among 38 miRNAs targeting host genes that interact with SARS-CoV-2 proteins (43). Lastly, I detected a candidate gene involved in cilia coating the respiratory tract (*DNAH7*). It represents one of the most down-regulated genes following SARS-CoV-2 infection of human bronchial epithelial cells in vitro (44). *DNAH7* expression levels were also significantly down regulated in human bronchial epithelial cells infected with MERS-CoV and influenza A (H1N1), which induce apoptosis in these cells (45, 46).

In line with other studies, my results show that greater attention should be given to better understanding MERS-CoV dynamics in dromedaries, as it is another relevant zoonotic disease belonging to the *Coronaviridae* like the still on-going SARS-CoV-2 pandemic. Since the transfection from dromedary to Bactrian camel has been proven recently (47), it is necessary to better delineate the geographical distribution of camel involvement in MERS-CoV (as well as other zoonotic diseases). Also, large-scale serological screening of human populations in areas where MERS-CoV is endemic in dromedary camels should be considered. Diversity characterization and genome-wide association studies need to be performed by scanning markers across the genome to find genetic variations associated with relevant zoonotic diseases. Likewise, measures should be delineated to prevent putative food-borne transmission of MERS-CoV. Further studies might provide important insight for understanding factors potentially contributing to effective management strategies to combat MERS-CoV and other zoonotic diseases in camel populations and consequently in human populations as well.

Human influence on the history of camels since domestication

Camels have been linked to the process of human development and were essential for its success. By establishing trading routes and reusing them over millennia, corridors of gene flow were opened shaping genetic diversity and structure. Over the past 11,000 years humans have brought a wide variety of animals under domestication, for food, secondary

products, labour, and companionship, with impact in human economy, society, and religion (48). Compared with other species that can travel for long distances, such as the horse (~5,500 ya; (49)) or the donkey (~5,000 ya; (50)), the domestication of the dromedary and Bactrian camel started rather late, most likely between the second and first millennia, and in the late fourth and early third millennium respectively (51, 52). After the domestication of dromedary in the coastal areas of the Arabian Peninsula, the introduction of the dromedary into Northern Africa via the Sinai from Roman Egypt started in the early first millennium BCE and intensified in the Ptolemaic period (52, 53). Although there have been strong bottlenecks during the late Pleistocene in all three extant camel species, dromedaries show a general lower genome-wide diversity than Bactrian camels (37). Moreover, I could not detect a secondary bottleneck during the domestication period – it is possible that the detection of a bottleneck related to domestication has been superimposed by the drastic decrease in the effective population size ending around 30,000 years BCE (Lado et al (7) [**Article 1**]). Similar demographic changes were observed in alpacas (New World camel) (54), where three population bottlenecks were detected throughout the Late Glacial Maximum in South America, nevertheless no bottleneck was visible during the domestication period.

After domestication, movements of dromedaries were influenced by humans and vice versa, therefore, knowledge on dromedary spatial genetic signatures also sheds light into past human history (53). In Lado et al. (7) [**Article 1**], I have assessed the genome-wide differentiation within the global dromedary population, which was shown to be very low. I could perceive that the traditional usage of dromedaries as pack animals, their exchange and movements along transcontinental caravan routes accounted for the observed lack of global population structure. Indeed, I could detect genetic admixture across continental populations (Asia and Africa), which highlights the strong anthropogenic influence on these animals from the cross-continental back-and-forth movements. In general, the Asian dromedary population showed higher genetic variability, which could be a sign for ancestral variation (with the Arabian Peninsula being a centre of domestication), although I cannot discard the hypothesis of post-domestication movements of camels or multiple origins of the founder populations. Also, I could detect within Asia a separated population belonging to a specific breed, Hadhana – one of the twelve recognized dromedary ecotypes in Saudi Arabia, limited to mountain regions in the South of the Arabian Peninsula, Al-Baha (55). In this case, the geographic accessibility might have an

important role in the observed genetic distinctiveness. Also, a possible explanation for the close relationship detected between Hadhana and African dromedaries might be the historic red sea route from Jiddah in Saudi Arabia to Aydhab and Port Sudan, an ancient known trading route (Lado et al. (7) [Article 1]). On the other hand, I found a more homogenous gene pool in African animals with the exception of the East African group, represented in our dataset by the two Kenyan dromedaries. This can be a consequence of a random founder effect followed by lack of gene flow due to geographical, physiological (e.g., Trypanosome infestation) and/ or cultural barrier, i.e., dromedaries in East Africa were dominantly used for milk production rather than transport or riding (56). Nevertheless, by understanding subtle population structure, we recognize the value of small, locally adapted populations and appeal for securing genomic diversity for a sustainable utilization of this key desert species. Finally, the global patterns of effective migration rates revealed pathways of dispersal after domestication, following historic caravan routes like the Silk and Incense Roads. In my first Article, I detected a corridor of significantly higher effective migration rates than the overall mean along the Mediterranean coast, connecting north-western Africa to the North of the Arabian until the border of the Arabian Desert (56, 57).

Camels promote trade and human interaction at extraordinary levels. Not only dromedaries were known as “Ship of the desert”, but hybrid camels were also the chosen animals as beasts of burden, as caravans, capable of travelling short and long ancient trading routes and are adapted to a wider range of climatic conditions and resist droughts. Although taxonomically distinct species, dromedaries and Bactrian camels are capable of interbreeding with each other. This produces a fertile F1 (first filial generation) hybrid that, due to heterosis arising from the interaction between parental genomes, has a better growth rate and productivity than that of their purebred parents (58, 59). Hybrid camels can carry greater weight, adapt to a wider variety of environments well beyond arid regions and better capable of withstanding rough terrains than both parents, being preferred over normal breeds as a better pack camel. Hybrid camels have been instrumental in caravan trade and military campaigns in Medieval and Early Modern Era, and in some regions such as western Turkey, until the introduction of trucks in the 1960s (60). In general, it is believed that the best first-generation hybrids are the products of male Bactrian camels crossed with female dromedaries, although female Bactrian camels crossed with male dromedaries are also attested (61). In Lado et al. (7) [Article 1],

although we were mainly interested in the modern-day global dromedary population, we included one Bactrian camel to test for potential interspecific hybridization, as this continues to be a widespread practice in Central Asia. Nevertheless, hybrid backcrosses might not be that different from pure-bred camels to the unfamiliar eye. In Lado et al. (7) [**Article 1**], although samples were identified as dromedaries, we detected potential hybrids from Iran and Kazakhstan, where camel hybridization practice is very common nowadays. Sustainability researchers working on camels recommend hybrids and backcrosses as they produce more milk, meat and wool than standard camels as well as the ability to survive the harsh climate (62). In several places in Kazakhstan, dromedaries, Bactrian camels, and their hybrids are free ranging and naturally intermingling (59). In such conditions, unsupervised mating may occur and go unrecorded producing animals with an unknown dromedary-Bactrian camel genetic make-up. The identification of hybrid camels based only on phenotypic characters is unreliable, particularly for later stage hybrids. Rigorous pedigree recording combined with a camel identification system (e.g., ear tag or ID chip) might be a good tool to identify individual hybrids. However, if the exact genetic make-up is required, then high density genome-wide studies are necessary.

Yet, almost nothing is known about the origins of hybrid camels and the common journey they share with humans. Although archaeological camel hybrids are rare, several zooarchaeologists have considered the presence of hybrids among archaeological camel remains from southwest Asia (without aDNA tools) (63-67). The earliest of these claims is on a camel burial from Eastern Arabia, dating to the Hellenistic Period (ca. 300 BCE) (66). Potts (61) suggested that hybridisation could have taken place earlier in the Neo-Assyrian period (800 – 700 BCE) in Iran, facilitated by the far-reaching imperial routes. The same author took his inspiration from Uerpmann (66), and based his argument mainly on textual and artistic records. However, genetic evidence about the early-rise of hybridization is still missing.

In a cultural-palaeogenetic context, the identification of isolated archaeological faunal remains often entails difficulties and uncertain results. Even if the material is well preserved, the hybrid species identification can be very problematic due to its intermediate position, in our case between Bactrian camel and dromedary. In Lado et al. (68) [in prep., **Article 4**], I bring together different research areas in (zoo)archaeology,

genetics, and history to investigate the presence and spread of hybrid camels across southwest Asia during the Iron Age and to a lesser extent later. I was interested in the various pathways in a geographical and chronological context, which humans followed to establish successful hybridization systems and a close relationship to their animal partners. Previous studies have revealed hybrids from a Roman archaeological site in Serbia (Viminacium), dated approximately to the late third to fourth centuries CE, using genetic tools (58, 69). However, following Potts (61) suspicion of starting during Neo-Assyrian period (800 – 700 BCE), I predicted that hybridization might have started even earlier, soon after the domestication of dromedaries (1000 – 2000 BCE) when the distribution of the two species started overlapping, during the Iron Age in Anatolia (1200 – 600 BCE). Long before this period, in mid- and late 3rd millennium BCE, humans would already cross other species (see *introduction* section).

To test my hypothesis, I shotgun sequenced aDNA samples, extracted from archaeological camel bones from the Middle East, and I could identify different levels of hybrids and backcrosses next to pure dromedaries and Bactrian camels. Thus, the palaeogenetically confirmed hybrids were directly radiocarbon dated. The earliest hybrid I detected was identified as Bactrian camel backcross from Hasanlu, Iran, dated to the Early Iron I Age (1112 – 933 calibrated years before Common Era (calBC); HAS2023) in one of the very important trading areas in northwestern Iran. This dates to a few centuries earlier than previous assumptions that pointed at camels hybridizing in the Iranian plateau at the time of the Assyrian Empire (61). Trading routes from Asia via northern Iran and Anatolia towards Europe were cold routes, even when it was not winter season, due to high altitude (e.g., Silk road crossing in Elburz Mountains in Iran – North and South route). Thus, caravan animals operating this route would have been adapted to cold environments. In that sense, the here-detected frequency of cold-adapted Bactrian camels (HAS2845, HAS3467 in Iron III Age) and Bactrian camel backcrosses in northern Iran can be expected. On the other hand, we also identified three samples as dromedary backcrosses, of which two originated from Iran: one dated to Iron I Age (1051 – 908 calBC, HAS822) and the other to the Parthian Period (347 – 535 calAD, MM302). Dromedary backcrosses could provide a higher milk yield, which would be a direct benefit for nomadic people. The demand for the vigorous hybrid camels promoted specialized trade across the known world until the end of the Middle Ages. As such, producing and maintaining first generation hybrid camels and their backcrosses always

has relied on encounters of peoples from diverse cultural backgrounds and lead to unique forms of cultural hybridization.

As I show in the fourth Article of this thesis (68) [in prep., **Article 4**], we finally have established archaeological, genetic and radiocarbon-dated evidence that this practice has indeed occurred as early as Early Iron I Age, sited in Hasanlu (Iran) where important commercial networks were located. This is also supported by the fact that dromedaries were already present at Tepe Sagzabad (northwest Iran) in the late second millennium BCE levels (70). From the ancient samples included in this study, all other camels were identified as dromedaries. One of them (KT-001) originated from Kinet (South Turkey close to Syria) and was direct-radiocarbon dated confirming the occurrence of this species in the northern Levant (Bay of Alexandretta) as early as seventh century BCE, during the Neo-Assyrian Period. Moreover, samples revealed as purebred dromedaries (after genetic analysis) from the Middle/ Late Iron Age located in Tell Sheikh Hamad (Syria) and from Tell Jemmeh (Israel) from the Persian Period (around the fifth century BCE), were very large or showing mixed-morphology, different from early-domesticated dromedaries in Arabia (e.g., 56). In previous studies, the extinct ancestor of modern dromedaries was shown to be larger than the modern individuals (71). One possible explanation for the presence of these large and mixed-morphology phenotypes in our samples might be that they resemble a former wild type, or an intermediate phenotype between wild and domestic dromedaries, possibly resulting from introgression during an early domestication phase when restocking from the wild was likely (56). In other animals such as cattle, sheep or pig, it has been shown that this pattern was common in which through time (before breeds were formed), animals fluctuated in size, showing regional differences (e.g., 72, 73).

Finally, the latest hybrid detected in our study (GIA5356, dromedary backcross) stems from an old excavation in Trier (western Germany) and might date to the Roman Period or the Medieval Period. Nevertheless, exact dates need to be confirmed by radiocarbon dating, because contextual evidence for this bone is missing. Based on osteological and zooarchaeological information, previous studies suggested that parts of the camels found in the northern provinces of the Roman Empire, would possibly be hybrid camels (64). In general, archaeological camel finds in Central Europe cover a chronological span from the Roman period (50 BCE – 700 CE) until the early Modern Age (ca. 17th century CE)

(64). As the hybridisation process allows hybrid camels to adapt to wider environments, this accelerates the spread of hybrid camels across climate zones, including Europe. As such, this Trier specimen is the northernmost hybrid camel confirmed with ancient DNA techniques. Earlier, a 17th century camel hybrid was detected in Tulln (Austria), indicating that camel hybrids could indeed exist in northern regions (74).

Although hybridization between Bactrian camel and dromedary was associated with the transportation of goods along multiple routes in earlier times, today hybridization schemes are very well established, mostly in order to facilitate improved milk, meat and wool yield from Middle Eastern and Central Asian countries. Besides being praised for their production traits, camel hybrids are also valued pets and a source of entertainment in social events in western parts of Turkey, such as camel beauty contests or camel wrestling during rutting season. These cultural events, in which I had the privilege to be part of in January 2019, in Izmir (Turkey), have existed for at least some centuries. Although textual evidence on the historical background of camel wrestling is scarce, it is difficult to imagine that this modern-day event does not have a long past tradition. Since one of the main aims of domesticating camels and breeding hybrids was to create powerful beasts of burden, it is highly likely that earlier forms of spectacle involving rutting camels were part of a mechanism that enhanced selective breeding. Thus, such events might have evolved synchronously with the history of camel hybridization (58). Besides being a source of entertainment, cross-cultural socialization or perhaps an exchange of exotic and local items for caravan owners and investors, wrestling would be an efficient way to select the strongest camels to purchase. Indeed, the strong human influence and interference in the history of camels from domestication until today cannot be passed unnoticed.

Concluding remarks

Although there is still much to untangle concerning Old World camels' history and diversity, my thesis has filled existing knowledge gaps by implementing modern methodologies on a large number of samples representative of different populations. In particular, this thesis comprises different scientific fields, i.e., immunology, virology, population genetics, and palaeogenomics using cutting edge molecular, population genomic and ancient DNA techniques. As such, my work encloses a large framework spanning ancient hybridisation to modern genome-wide investigation of post-

domestication migration routes and (immune) genetic diversity. This thesis also provides a starting point for understanding the genomic basis underlying the immune response to MERS-CoV infection in dromedaries, which is still poorly understood, and perhaps to other zoonotic diseases. Importantly, the diversity was characterized in relevant immune response genomic regions and compared between the three Old World species. Furthermore, with this thesis we moved a step further in explaining human influence in camel genetic diversity and history by revealing dromedaries and hybrid camels as a fundamental element of long-distance exchange networks such as the Silk Road. I also describe how we were able to detect the earliest evidence of dromedary-Bactrian hybridisation in an artifact dating to Early Iron I Age via a combination of scientific fields (archaeology, palaeogenomics, radiocarbon dating and history). With a specimen from Trier, we also show that hybrid camels were distributed as west as western Germany at the latest by the Medieval Period. This has provided an entirely new window into the archaeological and historic study of cultural encounters, frontiers, and hybridity in a large region of Eurasia, which has been populated by nomadic pastoralists charged with symbolism, but little textual remains. Altogether, this thesis not only opens doors for future (immune) genome diversity and disease-related studies, it also provides improved genome assemblies at chromosome-level, which serve as reference for future genome-wide association studies. Although this thesis created the foundation to place camels in a past culture-historical and present context, I truly believe that camels are *the* animals of the future. Despite the discovery of camel nanobodies as a great hope for future biotechnology and medical treatment, the value of (hybrid) camels as productive livestock in times of global warming is still being neglected by both scientists and policy makers. My mission as being part of the scientific community is also to create awareness of the importance of camels and of maintaining the present (immuno)genetic diversity for their capability to respond to future challenges like increasing temperatures, growing deserts or emerging diseases.

References

1. Nagy P, Wernery U, Burger P, Juhasz J, Faye B. The impact of COVID-19 on Old World Camelids and their potential role to combat a human pandemic. *Anim Front.* 2021;11(1):60-66. doi:10.1093/af/vfaa048.

2. Faye B. How many large camelids in the world? A synthetic analysis of the world camel demographic changes. *Pastoralism*. 2020;10(1):25. doi:10.1186/s13570-020-00176-z.
3. Engering A, Hogerwerf L, Slingenbergh J. Pathogen-host-environment interplay and disease emergence. *Emerg Microbes Infect*. 2013;2:e5. doi:10.1038/emi.2013.5.
4. Ali A, Baby B, Vijayan R. From Desert to Medicine: A Review of Camel Genomics and Therapeutic Products. *Front Genet*. 2019;10:17. doi:10.3389/fgene.2019.00017.
5. Kaskous S. Importance of camel milk for human health. *Emir J Food Agric*. 2016;28:158-163. doi:10.9755/ejfa.2015-05-296.
6. Alkhamees OA, Alsanad SM. A review of the therapeutic characteristics of camel urine. *Afr J Tradit Complement Altern Med*. 2017;14(6):120-126. doi:10.21010/ajtcam.v14i6.12.
7. Lado S, Elbers JP, Doskocil A, Scaglione D, Trucchi E, Banabazi MH, et al. Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes. *Commun Biol*. 2020;3:387. doi:10.1038/s42003-020-1098-7.
8. Hussen J, Schuberth H-J. Recent Advances in Camel Immunology. *Front Immunol*. 2021;11:3569. doi:10.3389/fimmu.2020.614150.
9. Hamer-Casterman C, Atarchouch T, Muyltermans S, Robinson G, Hamers C, Bajyana E, et al. Naturally occurring antibodies devoid of light chains. *Nature*. 1993;363:446-448. doi:10.1038/363446a0.
10. Muyltermans S, Baral TN, Retarnozzo VC, De Baetselier P, De Genst E, Kinne J, et al. Camelid immunoglobulins and nanobody technology. *Vet Immunol Immunopathol*. 2009;128(1-3):178-183. doi:10.1016/j.vetimm.2008.10.299.
11. Wrapp D, De Vlieger D, Corbett KS, Torres GM, Wang N, Van Breedam W, et al. Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies. *Cell*. 2020;181(5):1004-1015.e15. doi:10.1016/j.cell.2020.04.031.
12. Hanke L, Perez LV, Sheward DJ, Das H, Schulte T, Moliner-Morro A, et al. An alpaca nanobody neutralizes SARS-CoV-2 by blocking receptor interaction. *Nat Commun*. 2020;11(1):4420. doi:10.1038/s41467-020-18174-5.
13. Faye B, Chaibou M, Vias G. Integrated Impact of Climate Change and Socioeconomic Development on the Evolution of Camel Farming Systems. *Br J Environ Clim Change*. 2012;2:227-244. doi:10.9734/BJECC/2012/1548.
14. Gossner C, Danielson N, Gervelmeyer A, Berthe F, Faye B, Kaasik Aaslav K, et al. Human–Dromedary Camel Interactions and the Risk of Acquiring Zoonotic Middle East Respiratory Syndrome Coronavirus Infection. *Zoon Publ Health*. 2016;63(1):1-9. doi:10.1111/zph.12171.

15. Megersa B, Biffa D, Abunna F, Regassa A, Bohlin J, Skjerve E. Epidemic characterization and modeling within herd transmission dynamics of an "emerging trans-boundary" camel disease epidemic in Ethiopia. *Trop Anim Health Pro.* 2012;44:1643-1651. doi:10.1007/s11250-012-0119-z.
16. Zhu S, Zimmerman D, Deem SL. A Review of Zoonotic Pathogens of Dromedary Camels. *EcoHealth.* 2019;16(2):356-377. doi:10.1007/s10393-019-01413-7.
17. Burke DS. Evolvability of emerging viruses. In: Nelson AM, Horsburgh Jr CR, editors. *Pathology of emerging infections.* 2nd ed. ed. Washington (DC): American Society for Microbiology; 1998. p. 1-12.
18. Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel-human interface. *eLife.* 2018;7:e31257. doi:10.7554/eLife.31257.
19. van Doremalen N, Miazgowicz KL, Milne-Price S, Bushmaker T, Robertson S, Scott D, et al. Host Species Restriction of Middle East Respiratory Syndrome Coronavirus through Its Receptor, Dipeptidyl Peptidase 4. *J Virol.* 2014;88:9220-9232. doi:10.1128/jvi.00676-14.
20. Ovsyannikova IG, Haralambieva IH, Crooke SN, Poland GA, Kennedy RB. The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. *Immunol Rev.* 2020;296:205-219. doi:10.1111/imr.12897.
21. Widagdo W, Na Ayudhya SS, Hundie GB, Haagmans BL. Host Determinants of MERS-CoV Transmission and Pathogenesis. *Viruses.* 2019;11:280. doi:10.3390/v11030280.
22. Widagdo W, Okba NMA, Li W, de Jong A, de Swart RL, Begeman L, et al. Species-Specific Colocalization of Middle East Respiratory Syndrome Coronavirus Attachment and Entry Receptors. *J Virol.* 2019;93(16):e00107-19. doi:10.1128/jvi.00107-19.
23. Ming L, Wang Z, Yi L, Batmunkh M, Liu T, Siren D, et al. Chromosome-level assembly of wild Bactrian camel genome reveals organization of immune gene loci. *Mol Ecol Res.* 2020;20(3):770-780. doi:10.1111/1755-0998.13141.
24. Lado S, Elbers JP, Rogers MF, Melo-Ferreira J, Yadamsuren A, Corander J, et al. Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies. *BMC Genom.* 2020;21(1):606. doi:10.1186/s12864-020-06990-4.
25. Lado S, Elbers JP, Plasil M, Loney T, Weidinger P, Camp JV, et al. Innate and Adaptive Immune Genes Associated with MERS-CoV Infection in Dromedaries. *Cells.* 2021;10(6):1291. doi:10.3390/cells10061291.
26. Hewitt EW. The MHC class I antigen presentation pathway: strategies for viral immune evasion. *Immunology.* 2003;110(2):163-169. doi:10.1046/j.1365-2567.2003.01738.x.

27. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nat Rev Immunol.* 2018;18:325-339. doi:10.1038/nri.2017.143.
28. Li X, Geng M, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. *Journal of Pharmaceutical Analysis.* 2020;10:102-108. doi:10.1016/j.jpha.2020.03.001.
29. Wang SF, Chen KH, Chen M, Li WY, Chen YJ, Tsao CH, et al. Human-Leukocyte Antigen Class I Cw 1502 and Class II DR 0301 Genotypes Are Associated with Resistance to Severe Acute Respiratory Syndrome (SARS) Infection. *Viral Immunology.* 2011;24:421-426. doi:10.1089/vim.2011.0024.
30. Warren RL, Birol I. HLA predictions from the bronchoalveolar lavage fluid samples of five patients at the early stage of the Wuhan seafood market COVID-19 outbreak. *ArXiv.* 2020:2004.07108v3. doi:10.1093/bioinformatics/btaa756.
31. Harton J, Jin L, Hahn A, Drake J. Immunological Functions of the Membrane Proximal Region of MHC Class II Molecules. *F1000Research.* 2016;5:368. doi:10.12688/f1000research.7610.1.
32. Liu X. *The Silk Road in World History.* Oxford University Press; 2010. 176 p. (New Oxford World History). ISBN: 0199713723.
33. Hajeer AH, Balkhy H, Johani S, Yousef MZ, Arabi Y. Association of human leukocyte antigen class II alleles with severe Middle East respiratory syndrome-coronavirus infection. *Annals of Thoracic Medicine.* 2016;11:211-213. doi:10.4103/1817-1737.185756.
34. Ujvari B, Belov K. Major Histocompatibility Complex (MHC) Markers in Conservation Biology. *Int J Mol Sci.* 2011;12(8):5168-5186. doi:10.3390/Ijms12085168.
35. Elbers JP, Clostio RW, Taylor SS. Neutral Genetic Processes Influence MHC Evolution in Threatened Gopher Tortoises (*Gopherus polyphemus*). *J Hered.* 2017;108:515-523. doi:10.1093/jhered/esx034.
36. Plasil M, Mohandesan E, Fitak RR, Musilova P, Kubickova S, Burger PA, et al. The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes. *BMC Genom.* 2016;17(1):1-17. doi:10.1186/s12864-016-2500-1.
37. Fitak RR, Mohandesan E, Corander J, Yadamsuren A, Chuluunbat B, Abdelhadi O, et al. Genomic signatures of domestication in Old World camels. *Commun Biol.* 2020;3(1):316. doi:10.1038/s42003-020-1039-5.
38. Kurtz J, Kalbe M, Aeschlimann PB, Haberli MA, Wegner KM, Reusch TBH, et al. Major histocompatibility complex diversity influences parasite resistance and innate immunity in sticklebacks. *Proc R Soc Lond B.* 2004;271(1535):197-204. doi:10.1098/rspb.2003.2567.

39. Uematsu S, Akira S. Toll-Like Receptors (TLRs) and Their Ligands. In: Bauer S, Hartmann G, editors. Toll-Like Receptors (TLRs) and Innate Immunity. Vol 183. Berlin Heidelberg: Springer; 2008. p. 1-20. (Handbook of Experimental Pharmacology).
40. Haddad H, Al-Zyoud W. miRNA target prediction might explain the reduced transmission of SARS-CoV-2 in Jordan, Middle East. *Non-coding RNA Research*. 2020;5:135-143. doi:10.1016/j.ncrna.2020.08.002.
41. Leon-Icaza SA, Zeng M, Rosas-Taraco AG. microRNAs in viral acute respiratory infections: immune regulation, biomarkers, therapy, and vaccines. *ExRNA*. 2019;1:1-7. doi:10.1186/s41544-018-0004-7.
42. Zhu Z, Qi Y, Ge A, Zhu Y, Xu K, Ji H, et al. Comprehensive Characterization of Serum MicroRNA Profile in Response to the Emerging Avian Influenza A (H7N9) Virus Infection in Humans. *Viruses*. 2014;6:1525-1539. doi:10.3390/v6041525.
43. Sardar R, Satish D, Gupta D. Identification of Novel SARS-CoV-2 Drug Targets by Host MicroRNAs and Transcription Factors Co-regulatory Interaction Network Analysis. *Front Genet*. 2020;11:571274. doi:10.3389/fgene.2020.571274.
44. Nunnari G, Sanfilippo C, Castrogiovanni P, Imbesi R, Li Volti G, Barbagallo I, et al. Network perturbation analysis in human bronchial epithelial cells following SARS-CoV2 infection. *Exp Cell Res*. 2020;395:112204. doi:10.1016/j.yexcr.2020.112204.
45. Tao X, Hill TE, Morimoto C, Peters CJ, Ksiazek TG, Tseng C-TK. Bilateral Entry and Release of Middle East Respiratory Syndrome Coronavirus Induces Profound Apoptosis of Human Bronchial Epithelial Cells. *J Virol*. 2013;87:9953-9958. doi:10.1128/jvi.01562-13.
46. Yuen KM, Chan RWY, Mok CK, Wong ACN, Kang SSR, Nicholls JM, et al. Differential onset of apoptosis in avian influenza H5N1 and seasonal H1N1 virus infected human bronchial and alveolar epithelial cells: an *in vitro* and *ex vivo* study. *Influenza and Other Respiratory Viruses*. 2011;5;Suppl. 1:437-438. doi:10.1038/jid.2014.371.
47. Lau SKP, Li KSM, Luk HKH, He ZR, Teng JLL, Yuen KY, et al. Middle East Respiratory Syndrome Coronavirus Antibodies in Bactrian and Hybrid Camels from Dubai. *Msphere*. 2020;5(1):e00898-19. doi:10.1128/mSphere.00898-19.
48. Zeder MA. The domestication of animals. *Journal of Anthropological Research*. 2012;68(2):161-190. doi:10.3998/jar.0521004.0068.201.
49. Outram AK, Stear NA, Bendrey R, Olsen S, Kasparov A, Zaibert V, et al. The Earliest Horse Harnessing and Milking. *Science*. 2009;323(5919):1332-1335. doi:10.1126/science.1168594.
50. Rossel S, Marshall F, Peters J, Pilgram T, Adams MD, O'Connor D. Domestication of the donkey: Timing, processes, and indicators. *Proc Natl Acad Sci USA*. 2008;105:3715-3720. doi:10.1073/pnas.0709692105.

51. Sala R. The Domestication of Camel in the Literary, Archaeological and Petroglyph Records. *Journal of Arid Land Studies*. 2017;26. doi:10.14976/jals.26.4_205.
52. Uerpmann H-P, Uerpmann M. The appearance of the domestic camel in southeast Arabia. *J Oman Stud*. 2002;12:235–260.
53. Bulliet RW. *The Camel and the Wheel*. 19 ed. New York: Columbia University Press; 1975. 324 p.
54. Wu H, Guang X, Al-Fageeh MB, Cao J, Pan S, Zhou H, et al. Camelid genomes reveal evolution and adaptation to desert environments. *Nat Commun*. 2014;5(1):5188. doi:10.1038/ncomms6188.
55. Faye B, Abdallah HR, Almathen FS, Harzallah BD, Al-Mutairi SE. Camel biodiversity. Camel phenotypes in the Kingdom of Saudi Arabia. Rome; 2011. 69 p. FAO.
56. Almathen F, Charruau P, Mohandesan E, Mwacharo JM, Orozco-terWengel P, Pitt D, et al. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc Natl Acad Sci USA*. 2016;113(24):6707-6712. doi:10.1073/pnas.1519508113.
57. Heiss J. Caravans from South Arabia. In: Knoll E-M, Burger P, editors. *Camels in Asia and North Africa. Interdisciplinary Perspectives on their Significance in Past and Present*. Vienna: Austrian Academy of Sciences; 2012. p. 131-139. (Veröffentlichungen zur Sozialanthropologie).
58. Çakırlar C, Berthon R. Caravans, camel wrestling and cowrie shells: Towards a social zooarchaeology of camel hybridization in Anatolia and adjacent regions. *Anthropozoologica*. 2014;49(2):237-252. doi:10.5252/az2014n2a06.
59. Dioli M. Dromedary (*Camelus dromedarius*) and Bactrian camel (*Camelus bactrianus*) crossbreeding husbandry practices in Turkey and Kazakhstan: An in-depth review. *Pastoralism*. 2020;10:6. doi:10.1186/s13570-020-0159-3.
60. İnal O. One-Humped History: The Camel as Historical Actor in the Late Ottoman Empire. *International Journal of Middle East Studies*. 2020;53;1:57-52. Epub 2020/12/22. doi:10.1017/S0020743820000987.
61. Potts DT. Camel hybridization and the role of *Camelus bactrianus* in the ancient Near East. *J Econ Soc Hist Orient*. 2004;47:143-165. doi:10.1163/1568520041262314.
62. Faye B, Konuspayeva G. The Encounter between Bactrian and Dromedary Camels in Central Asia. In: Knoll E-M, Burger PA, editors. *Camels in Asia and North Africa. Interdisciplinary perspectives on their significance in past and present*. Vienna: Austrian Academy of Sciences; 2012. p. 27-33. (Veröffentlichungen zur Sozialanthropologie).
63. Köhler-Rollefson I. Zoological analysis of camel skeletons. In: Smith RH, Day LP, editors. *Pella of the Decapolis. Vol 2*. Wooster: College of Wooster Art Museum; 1989. p. 142-164.

64. Pigière F, Henrotay D. Camels in the northern provinces of the Roman Empire. *J Archaeol Sci*. 2012;39:1531-1539. doi:10.1016/j.jas.2011.11.014.
65. Studer J, Schneider A. Camel use in the Petra region, Jordan: 1st century BC to 4th century AD. In: Vila E, Gourichon L, Choyke AM, Buitenhuis H, editors. *Archaeozoology of the Near East VIII.Tome II*, Vol 49. Paris: 2008. p. 581–596. (Travaux de la maison de l'orient et de la méditerranée).
66. Uerpmann H-P. Camel and horse skeletons from protohistoric graves at Mleiha in the Emirate of Sharjah (U.A.E.). *Arabian Archaeology and Epigraphy*. 1999;10(1):102-118. doi:10.1111/j.1600-0471.1999.tb00131.x.
67. Fabiš M. Archaeofaunal Remains from the Lower Sanctuary. A Preliminary Report on the 1994 Excavations. *Studia Troica*. 1996;6:217-227.
68. Lado S, Elbers JP, Kilimci FS, Kara ME, Dabanoğlu I, Hurk Y, et al. Hidden hybrids – detecting early hybridization between dromedary and Bactrian camels in a culture-historical context. In preparation.
69. Burger PA, Lado S, Mohandesan E, Vukovic-Bogdanovic S, Peters J, Çakirlar C. Ancient and modern hybridisation between one- and two-humped camels. Second International Selçuk-Ephesus Symposium On Culture Of Camel-Dealing And Camel Wrestling; 2018 JAN 18-20, 2018; Selcuk, Turkey. 153-159 p. (vol. Volume II Natural And Applied Science Health And Medical Science). eng. Available from: http://www.selcuk.bel.tr/Files/dosyalar/sempozyum_bildiri/sempozyum_2018/2_Ulusla_rarasi_Selcuk_Efes_Devecilik_Kulturu_ve_Deve_Guresleri_Sempozyumu_Fen_ve%20Saglik_Bilimleri_2018.pdf.
70. Mashkour M. Chasse et élevage au nord du Plateau central iranien entre le Néolithique et l'Âge du Fer. *Paléorient*. 2002;28(1):27-42. doi:10.3406/paleo.2002.4737.
71. Mohandesan E, Speller CF, Peters J, Uerpmann H-P, Uerpmann M, De Cupere B, et al. Combined hybridization capture and shotgun sequencing for ancient DNA analysis of extinct wild and domestic dromedary camel. *Mol Ecol Res*. 2017;17(2):300-313. doi:10.1111/1755-0998.12551.
72. Albarella U. 'Size matters': how and why biometry is still important in zooarchaeology. In: Dobney K, O'Connor T, editors. *Bones and the Man: Studies in Honour of Don Brothwell*. Oxford: Oxbow Books; 2002. Chapter 7; p. 51-62.
73. MacKinnon M. Cattle 'breed' variation and improvement in Roman Italy: connecting the zooarchaeological and ancient textual evidence. *World Archaeology*. 2010;42(1):55-73. doi:10.1080/00438240903429730.
74. Galik A, Mohandesan E, Forstenpointner G, Scholz UM, Ruiz E, Krenn M, et al. A Sunken Ship of the Desert at the River Danube in Tulln, Austria. *PLoS ONE*. 2015;10(4):e0121235. doi:10.1371/journal.pone.0121235.

8. CURRICULUM VITAE

Sara Ribeiro Barbosa Almendra Lado, BSc MSc

 Vienna, Austria

 sara.lado@vetmeduni.ac.at

Sex Female | Date of birth 18/04/1992 | Nationality Portuguese

Education

From November 2017- March 2021: PhD candidate at the Veterinary Medicine University Vienna, Austria. Field: (Immuno)genetics and genomics in camels - “Characterization of the (immuno)genome diversity in domestic, wild and ancient hybrid old world camelids”. PhD defense on June 9th 2021.

September 2013 – December 2015: Master degree in Biodiversity, Genetics and Evolution at CIBIO/InBio – UP, Portugal.

September 2010 - July 2013: Bachelor degree in Biology at the Faculty of Sciences of the University of Porto, Portugal.

Publications in SCI journals

Lado, S., Elbers, J.P., Kilimci, F. S., Kara, M. E., Dabanoğlu, I., Hurk, Y., Brongers, T., Grigson, C., Lev-Tov, J., McClure, S., Davoudi, H., Mohaseb, A., Baker, P., Kühne, H., Kreppner, J., Haring, E., Berthon, R., Peters, J., Mashkour, M., Burger, P. A., Çakırlar, C. (*in preparation*). Hidden hybrids – detecting early hybridization between dromedary and Bactrian camels in a culture-historical context.

Lado, S., Elbers, J.P., Plasil, M., Loney, T., Weidinger, P., Camp, J. V., Kolodziejek, J., Futas, J., Kannan, D. O., Orozco-terWengel, P., Horin, P., Nowotny, N., Burger, P. A.

(2021). Innate and adaptive immune genes associated with MERS-CoV infection in dromedaries. *Cells*, 10(6), 1291. <https://doi.org/10.3390/cells10061291>

Impact factor: 4.366

Lado, S., Elbers, J. P., Duskocil, A., Scaglione, D., Trucchi, E., Banabazi, M. H., ... & Burger, P. A. (2020). Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes. *Communications Biology*, 3(1), 1-8. <https://doi.org/10.1038/s42003-020-1098-7>

Impact factor: 4.165

- Invitation from Nature Ecology and Evolution blog to write a “behind the paper” post: <https://natureecoevocommunity.nature.com/posts/camel-the-animal-of-the-past-present-and-future>

Lado, S., Elbers, J. P., Rogers, M. F., Melo-Ferreira, J., Yadamsuren, A., Corander, J., ... & Burger, P. A. (2020). Nucleotide diversity of functionally different groups of immune response genes in Old World camels based on newly annotated and reference-guided assemblies. *BMC Genomics*, 21(606), 1-17. <https://doi.org/10.1186/s12864-020-06990-4>.

Impact factor: 3.7

Mahtani-Williams, S., Fulton, W., Desvars-Larrive, A., **Lado, S.**, Elbers, J. P., Halpern, B., ... & Burger, P.A. (2020). Landscape Genomics of a Widely Distributed Snake, *Dolichophis caspius* (Gmelin, 1789) across Eastern Europe and Western Asia. *Genes*, 11(10), 1218. <https://doi.org/10.3390/genes11101218>.

Impact factor: 3.3

Lado, S., Alves, P. C., Islam, M. Z., Brito, J. C., & Melo-Ferreira, J. (2019). The evolutionary history of the Cape hare (*Lepus capensis sensu lato*): insights for systematics and biogeography. *Heredity*, 123(5), 634-646. <https://doi.org/10.1038/s41437-019-0229-8>.

Impact factor: 3.38

Lado, S., Farelo, L., Forest, V., Acevedo, P., Dalén, L., & Melo-Ferreira, J. (2018). Post-glacial range revolutions in South European hares (*Lepus spp.*): Insights from ancient DNA and ecological niche modelling. *Journal of biogeography*, 45(12), 2609-2618. <https://doi.org/10.1111/jbi.13454>.

Impact factor: 3.88

Training and professional experience

November 2017 – Spring 2021: PhD candidate at Vetmeduni, Vienna, Austria.

March 2021: “Resilience & Well-being in Academia”, Vetmeduni, Vienna, Austria (WEBINAR).

February 2021: One-week advanced course: “Art & design for Scientists”, CIBIO-InBIO, Vairão, Portugal (ONLINE).

June 2020: Practical peer review training with one-to-one guidance from Publons Academy – graduation (ONLINE).

Winter semester 2019: “Didactics & Methods of Teaching” course for employees at Vetmeduni.

October 2019: Consultancy Meeting “Advances in Nuclear and Genomic Tools to Improve Livestock Productivity – Technology Gaps and New Approaches for Application in Developing Countries” one-week workshop as camel expert, Vienna International Centre (VIC by IAEA, UNO).

July 2019: “Bioinformatics analysis of genomic data to assess population structure, genotype-phenotype association and genomic prediction” two-week hands-on workshop, Seibersdorf (IAEA, UNO).

April 2019: Attendance at gene drive workshop “Evaluation of spatial and temporal control of Gene Drives”, BOKU, Vienna.

August 2018: “Hands-on Workshop on Exome-capture Library preparation” 3 days at Vienna Biocenter (INTERREG project).

6 – 10 November 2017 – “Landscape genomics workshop” by PR STATISTICS (Wales, UK).

September – October 2017: Internship at Max Planck Institute (Plön), in Bioinformatics with Dr. Julien Dutheil.

December 2016 – September 2017: Active member of the World Lagomorph Society, under supervision of Prof. Klaus Hacklander (BOKU).

December 2016 – November 2017: WS thermo X tele project (Sparkling Science) at BOKU, Vienna and National Park Donau-Auen (Schloss Eckartsau), Austria with Robin Sandfort.

April – September 2016: ERASMUS+ Traineeship working for the World Lagomorph Society in different projects (e.g., recording European Museums’ information on coat color change species; Lagomorph book project) at BOKU, Vienna with Dr. Klaus Hackländer.

February - March 2016: Participation in two distinct projects: 1) species ID of hares from South of France; and 2) infer population structure/ demography from Greenland hares (extraction, amplification/ genotyping and sequencing of ancient DNA samples) at the department of Bioinformatics and Genetics of the Swedish Natural History Museum, Stockholm with Dr. Love Dalén.

September 2014 – December 2015: Master thesis in “Population history and taxonomy of African hares (genus *Lepus*) inferred from genetic variation” at CIBIO-InBIO/UP, Portugal with Paulo C. Alves and J. Melo-Ferreira.

February 2014 – August 2014: ERASMUS exchange on the second semester of the first year of Master degree - Master on Wildlife Ecology and Wildlife Management at BOKU, Vienna, Austria.

September 2012 – July 2013: Scientific investigation as internship on the last year of Bachelors in “Population genetics and systematics of North African hares” at CIBIO-InBIO/UP, Portugal.

February 2012 – August 2012: Volunteer research assistance at the wet-lab in a scientific investigation with *Jaculus sp.*, using non-invasive genetics at CIBIO-InBio-UP, Portugal.

October - December 2011: “XI Curso Introdução à medicina Legal e Outras Ciências Forenses” (Legal Medicine and Forensic Sciences Research Introduction) at the Medicine University of Porto - Abel Salazar.

Conference presentations

26 – 28 February **2020:** **SPEAKER. Lado, S**; Burger, PA; Elbers, J; Peters, J; Çakirlar. “*Detecting ancient dromedary-Bactrian and recent domestic-wild camel hybridization using shotgun sequencing.*” 4th Annual Meeting in Conservation Genetics from Genomes to Application; FEB 26-28, 2020; Frankfurt, Germany.

26 – 28 February **2020:** Burger, PA; Mathani-Williams, S; Desvars-Larrive, A; Fulton, W; **Lado, S**; Elbers, JP; Halpern, B; Barbocsay, G; Nagy, ZT; Orozco-terWengel, P; Herczeg, D; Vörös, J. “*Landscape genomics of a widely distributed racer (*Dolichophis caspius*, Gmelin 1789) across eastern Europe and western Asia.*” 4th Annual Meeting in Conservation Genetics from Genomes to Application; Frankfurt, Germany.

11 – 15 January **2020:** **SPEAKER. Lado, S**, Burger, PA, Elbers, J, Peters, J, Çakirlar, C. Plant and Animal Genomics XXVIII international conference “*Applying shotgun sequencing to detect early Dromedary-Bactrian camel hybrids*”, in San Diego, California, USA.

2 – 6 September **2019:** Mahtani-Williams, S; Fulton, W; Desvars-Larrive, A; **Lado, S**; Elbers, J; Halpern, B; Babocsay, G; Laus, B; Nagy, ZT; Orozco-terWengel, P; Herczeg, D; Vörös, J; Burger, PA. “*Landscape genomics of the Caspian whipsnake (*Dolichophis caspius*) across Eastern Europe and Western Asia.*” XX European Congress of Herpetology; Milano, Italy. (ISBN: 979-12-200-5284-9).

3 – 7 June **2019:** Çakirlar, C; Berthon, R; Burger, P; Kara, ME; Kilimci, FS; Kreppner, J; **Lado, S**; Mashkour, M; McClure, S; Peters, J. “*Hidden hybrids: Camels and cultural*

blending in the Ancient Near East.” The Archaeozoology of Southwest Asia and Adjacent Areas (ASWA) conference; Barcelona, Spain. 2019.

17 – 20 January **2019**: **SPEAKER. Lado, S.** Burger, PA, Elbers, J, Peters, J, Çakırlar, C. 3rd International Selçuk-Ephesus Symposium on Culture of Camel Dealing and Camel Wrestling “*Detection of early Dromedary-Bactrian camel hybrids through ancient sampling*”, in Izmir, Turkey.

17 November **2018**: **SPEAKER. Lado, S.** Elbers, J; Burger, P. 3rd International Conference. “*On-going projects on Dromedary (Camelus dromedarius) genomics*”, in Casablanca, Morocco.

17 November **2018**: Burger, PA; **Lado, S.** Doskocil, A; Mohandesan, E; Elbers, J; Fitak, RR. “*Genomic signals of selection related to domestication and adaptation in Old World camels.*” 3rd Conference Impact of Climatic and Environment Changes on Animal Productions: Advantages, Constraints and Perspectives in the Camel; Casablanca, Morocco. 2018.

12 – 15 November **2018**: **SPEAKER. Lado, S.** Elbers, JP; Doskocil, A; Ciani, E; Burger, PISOCARD conference – “*Genome-wide diversity and demographic history in the global dromedary population*”, in Laayoune, Morocco.

26 – 28 February **2018**: **SPEAKER. Lado, S.** Elbers, JP; Doskocil, A; Ciani, E; Burger, P3rd Annual Meeting in Conservation Genetics 2018 “*The need to recognize and conserve genetic diversity in the global dromedary population*”, in Vienna, Austria.

January **2018**: Burger, P.A., **Lado, S.** Mohandesan, E., Vukovic-Bogdanovic S., Peters, J. & Çakırlar, C. “*Ancient and Modern Hybridisation between one- and two-humped camels.*” In: A. KOÇ & O.U.H. Erdogan (Eds.) Second International Selcuk-Eephesus Symposium on Culture of Camel-Dealing and Camel Wrestling. Vol. II, pp. 153-159.

4 – 8 September **2016**: **SPEAKER.** Oral presentation given at the 90th Annual Meeting of the German Society for Mammalian Biology “*Population history and taxonomy of African hares (genus Lepus) inferred from genetic variation*” in Berlin.

11 – 15 July **2016**: **Poster presentation** given at the 5th World Lagomorph Conference “*Population history and taxonomy of African hares (genus Lepus) inferred from genetic variation*” at California State University, California.

3 – 6 May **2016**: **Poster presentation** given at ConGenOmics 2016 “*Few loci are not enough for accessing phylogenetic relationships of species complex: the case of the cape hare, Lepus capensis.*” CIBIO-InBIO/UP.

17 – 21 August **2015**: **SPEAKER**, Oral Presentation given at the 7th European Congress of Mammalogy “*Population history and taxonomy of North African hares (genus Lepus) inferred from genetic variation*” at Stockholm University, Sweden.

29 – 30 April **2013**: Attendance at Advances in Ecological Speciation (AES) at CIBIO-InBIO/UP, Portugal.

6 – 7 December **2012**: Attendance at Trends in Biodiversity and Evolution (TIBE) “*Integrative Approaches in Evolutionary Biology*” at CIBIO-InBIO/UP, Portugal.

5 – 6 December **2011**: Attendance at Trends in Biodiversity and Evolution (TIBE) “*New Challenges in Conservation Genetics*” at CIBIO-InBIO/UP, Portugal.

Teaching experience

- Lectures given in the “Population and Immunogenetics” course for Master students from BOKU and Vetmeduni universities (two subjects taught by me: 1. “Maintenance of genetic diversity”; 2. “Genetic restoration and captive management”). This teaching was given for three years, two Winter Semesters and one Summer Semester, during my PhD.
- Supervision of several undergraduate students in the wet lab during MSc and PhD degrees.

Scholarships and grants

November 2017: Partfunded scholarship for the “Landscape genetic data analysis using R” (LNDG02) course, by PR STATISTICS (Wales, UK).

September – October 2017: Max Planck Institute, Plön, Germany, scholarship for Bioinformatics internship.

September 2016: Travel grant for the 90th Annual Meeting of the German Society for Mammalian Biology.

April – October 2016: *ERASMUS* + Internship scholarship at Institute of Wildlife Biology and Game Management, BOKU, Vienna, Austria.

Summer semester 2014: *ERASMUS* scholarship for the Master Program at Institute of Wildlife Biology and Game Management, BOKU, Vienna, Austria.
