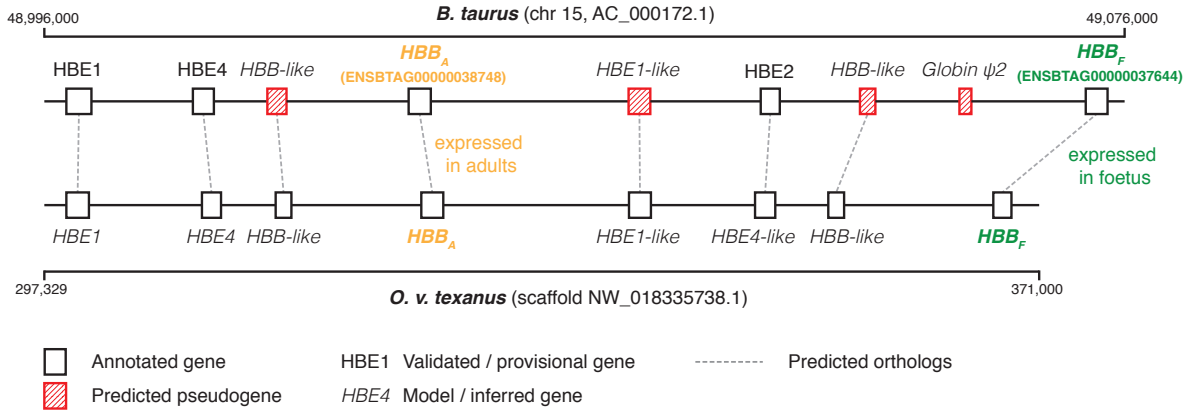
Supplementary Material

for

The genetic basis and evolution of red blood cell sickling in deer

A Esin, LT Bergendahl, V Savolainen, JA Marsh & T Warnecke



Supplementary Figure 1

The β-globin locus. Comparison of the *Bos taurus* (UMD 3.1.1) and *Odocoileus virginianus texanus* (Ovir.te 1.0) β-globin locus. Genetic elements for both loci are shown as predicted in the NCBI genome browser. Pseudogene labels are assigned as in REF. 26. Orthology relationships are supported by highest BLAST matches of mRNAs (or gene models where no mRNA is available) between the two genomes. *B. taurus* genes in this locus have no additional matches in the *O. v. texanus* assembly outside of the locus presented here, suggesting that the locus has not duplicated further in deer. As adult and foetal β-globin genes are not classically denoted HBB_A and HBB_F in cattle, we provide Ensembl gene IDs for unambiguous identification.

a

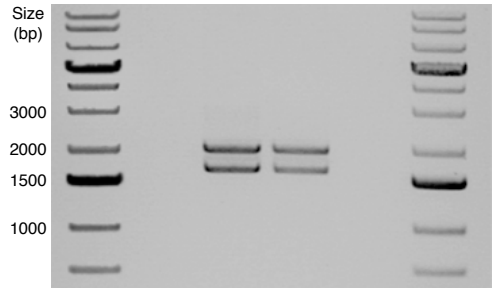
Name	Sequence (5' to 3')	Tm (°C)
Ovirg_F1	ATTTGTATTGCTGGAATGACTGAGA	64.5
Ovirg_R1	CCATCTGAGTAAGAGACAGTGAAAT	61.5
Ovirg_F2	CTTACACTTGCTTCTGACACAACCG	67.7
Ovirg_R2	GGAAAGCAGGAAAGGGAGCT	65.7
Ccap_R1	CCATCTGAGTAAGAGATCATGAAAT	61.3
Ovirg_Fmid2	GTGATAAGCTGCACGTGGAT	62.7
MTCB_F	CCHCCATAAATAGGNGAAGG	57.6
MTCB_R	WAGAAYTTCAGCTTTGGG	55.3

b

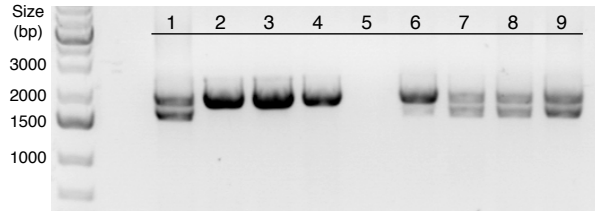
Sample	Adult β-globin	Foetal β-globin
WTD	Ovirg_F1 + Ovirg_R1	GE
SD	Ovirg_F1 + Ovirg_R1	GE
RED	Ovirg_F1 + Ovirg_R1	GE
BACT	Ovirg_F1 + Ovirg_R1	GE
WLD	Ovirg_F1 + Ovirg_R1	GE
BFD	Ovirg_F1 + Ovirg_R1	Not amplified
PDD	Ovirg_F1 + Ovirg_R1	Not amplified
SIKA	Ovirg_F1 + Ovirg_R1	Not amplified
ELK	Ovirg_F1 + Ovirg_R1	Not amplified
MUNT	Ovirg_F1 + Ovirg_R1	Not amplified
PUDU	Ovirg_F1 + Ovirg_R1	Not amplified
REIN	Ovirg_F1 + Ccap_R1	Ovirg_F2 + Ovirg_R1
ROE	Ovirg_F1 + Ccap_R1	Ovirg_F2 + Ovirg_R1
CWD	Ovirg_F1 + Ccap_R1	Ovirg_F2 + Ovirg_R1
WAP	Ovirg_F1 + Ccap_R1	Ovirg_F2 + Ovirg_R1

GE, Gel Excision.

c

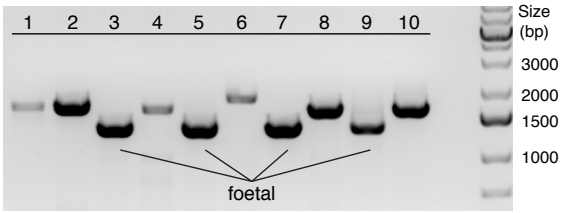


d



1-SD; 2-SIKA; 3-PDD; 4-BFD; 5-REIN; 6-ELK; 7-RED; 8-WLD; 9-BACT

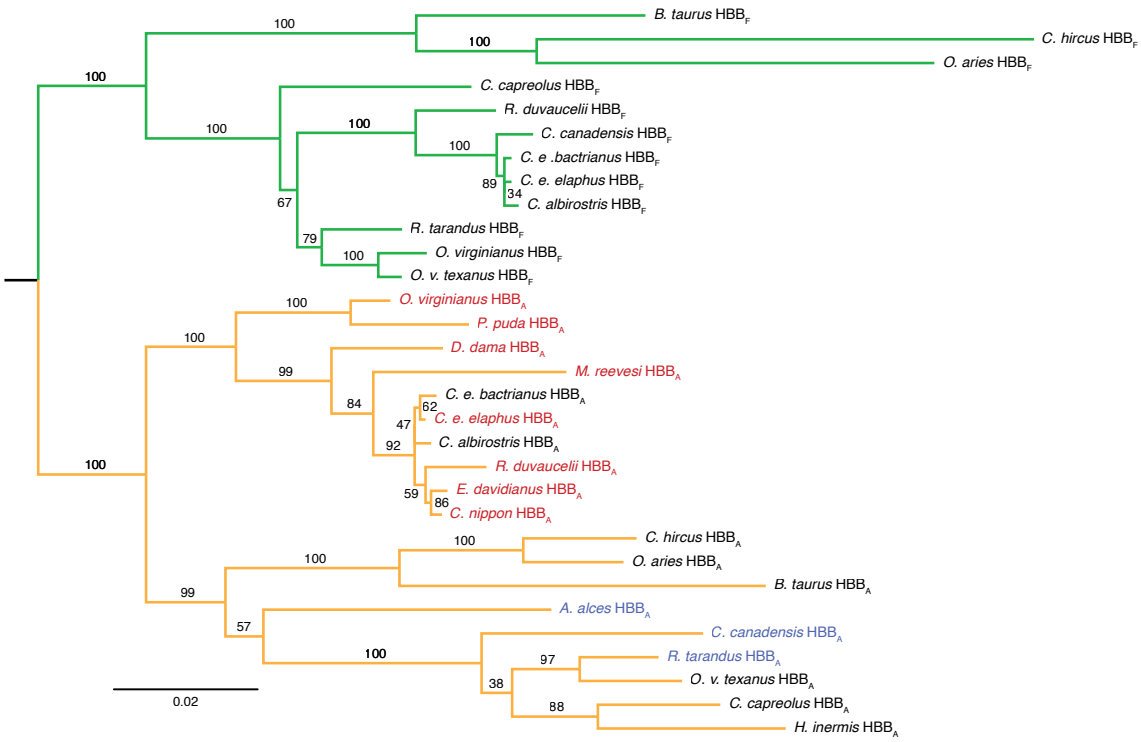
e



1-WLD; 2-SD; 3,4-WAP; 5,6-REIN; 7,8-ROE; 9,10-CWD

Supplementary Figure 2

Amplification of deer β-globin genes. a, Primers used in this study. **b**, Primer combinations used to amplify adult and foetal β-globin genes in different species. Where possible, gel excision was used to isolate the co-amplified foetal β-globin band. In certain cases, the adult gene could also be selectively amplified using primers Ovirg_F1/Ovirg_R2 (see panel e, lanes 1,2). **c**, Agarose gel showing co-amplification of adult and foetal β-globin genes in white-tailed deer with Ovirg_F1/Ovirg_R1; the two lanes show two different individuals. **d**, Agarose gel showing heterogeneity of amplification products in different deer using Ovirg_F1/Ovirg_R1. **e**, Agarose gel showing selective amplification of adult and foetal β-globin genes. Lanes 1 & 2: adult β-globin using Ovirg_F1/Ovirg_R2; all other lanes with primer combinations described in panel b. See Supplementary Table 1 for sample abbreviations.



Supplementary Figure 3

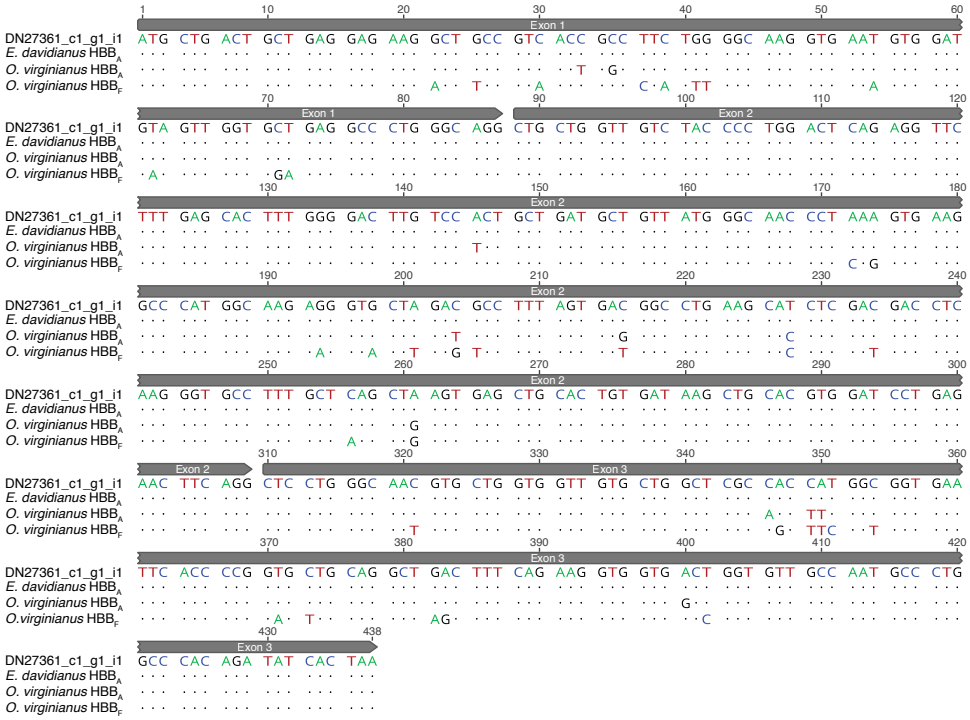
Putative HBB_A and HBB_F genes cluster separately on an HBB_A/HBB_F gene tree. The tree is a maximum likelihood reconstruction based on a nucleotide alignment of genic sequences (coding exons and intervening introns; see Methods). The sequence of *H. inermis* HBB_F was only partially resolved and is hence omitted. Branches are coloured as adult (orange) or foetal (green). Tip labels for HBB_A are coloured according to the species' propensity to sickle (red = majority sickling, blue = majority non-sickling, black=undetermined). Bootstrap values (%) are derived from 1000 bootstrap replicates (see Methods for further information on tree reconstruction). The scale bar shows the number of nucleotide substitutions per site.

a

ID of reconstructed transcript	Transcript length	Predicted product	Genbank ID (species)*	Abundance (TPM)
DN5482_c3_g1_i1	731	α-globin chain	JF811751.1 (*Pantholops hodgsonii*)	325146
DN1442_c2_g1_i1	294	β-globin mRNA	NM_001014902.3 (*Bos taurus*)	285287
DN27361_c1_g1_i1	747	β-globin mRNA	XM_006061581.1 (*Bubalus bubalis*)	224345
DN10562_c0_g1_i1	651	ubiquitin A-52 residue ribosomal protein fusion product 1 (UBA52)	XM_019963827.1 (*Bos indicus*)	2255
DN9617_c0_g5_i1	322	16S bacterial ribosomal RNA gene	CP019213.1, nucleotides 680747-681068 (*Escherichia coli*)	2219
DN27542_c0_g1_i1	2033	5'-aminolevulinate synthase 2 (ALAS2)	XM_005894452.1 (*Bos mutus*)	2010
DN20291_c0_g1_i1	621	S100 calcium binding protein A12 (S100A12)	XM_005909570.2 (*Bos mutus*)	1850
DN52287_c4_g1_i1	2893	18S ribosomal RNAgene	JN412502.1 (*Bubalus bubalis*)	1478
DN5641_c0_g1_i1	979	ferritin heavy chain 1 (FTH1)	NM_174062.3 (*Bos taurus*)	1404
DN9617_c0_g2_i1	1480	16S bacterial ribosomal RNA gene	CP018801.1 (4464861-4466340) (*Escherichia coli*)	1348

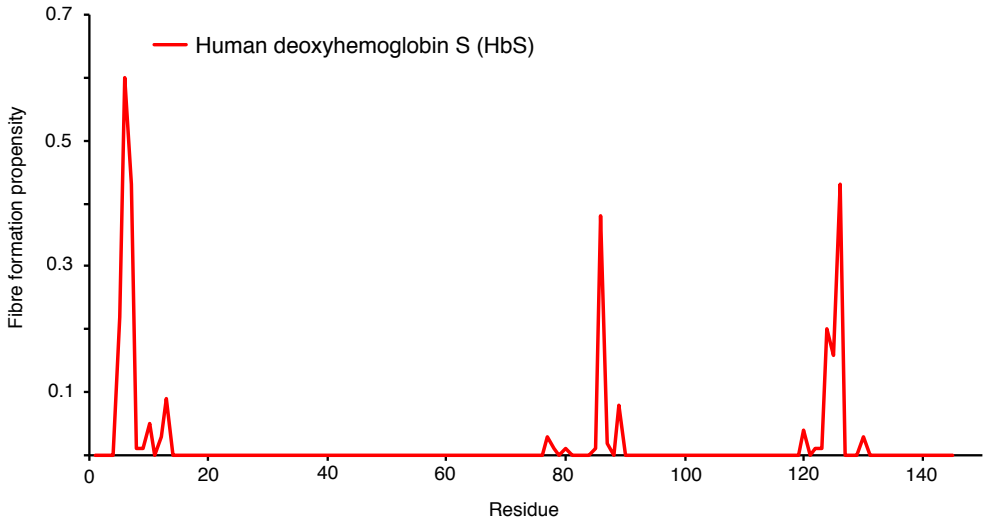
*Best match in the non-redundant nucleotide database (queried using MegaBLAST); TPM, transcripts per million.

b



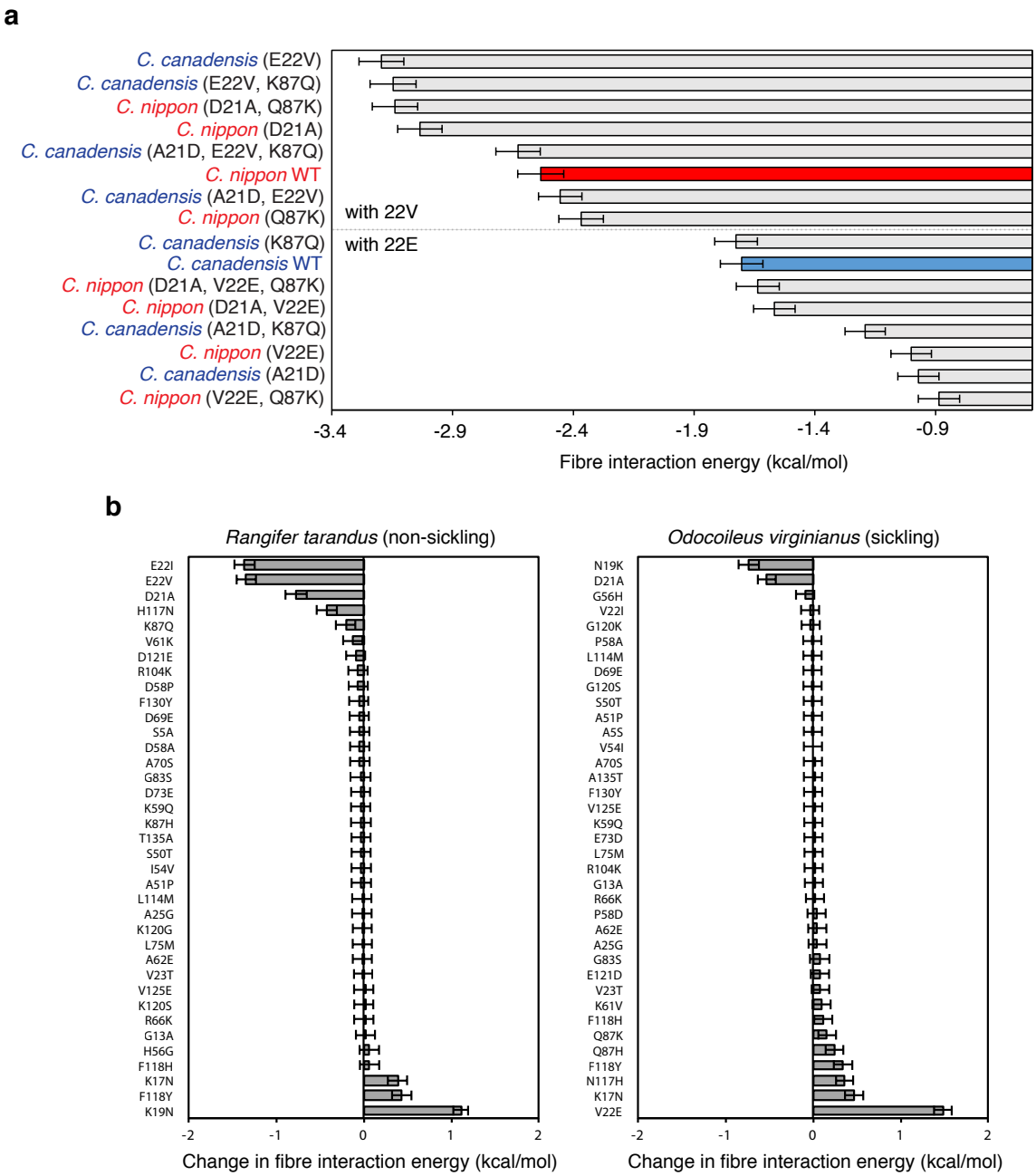
Supplementary Figure 4

Reconstructing the adult β-globin sequence from RNA sequencing data. a, The ten most abundant transcripts in the *de novo*-assembled *E. davidianus* red blood cell transcriptome. **b**, Nucleotide alignment of the *E. davidianus* β-globin CDS derived from the *de novo* transcriptome assembly, the putative *E. davidianus* adult β-globin CDS derived from amplification, and the foetal and adult β-globin CDSs from *O. virginianus*.



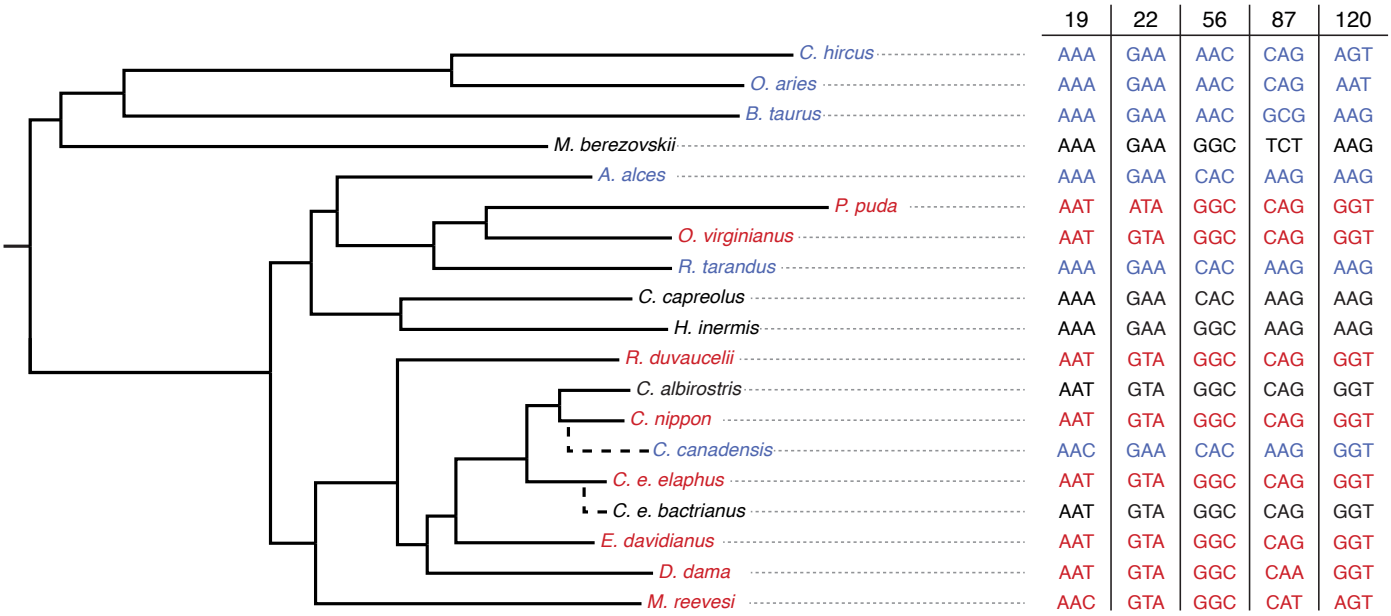
Supplementary Figure 5

Fibre formation propensity for human HbS. Fibre formation is assumed to occur via an interaction between the EF pocket on one β-globin molecule and a given focal residue in a β-globin molecule in the other strand of the fibre, essentially as in Fig. 2d, but using the human HbS sequence. Fibre formation propensity represents the fraction of the 100 β-globin dimer models built for each position that can form HbS-like fibres. The highest fibre formation propensity is observed for the valine at position 6.



Supplementary Figure 6

Comparative structural investigation of key sickling residues. **a,** Effects on fibre interaction energy of replacing defined single amino acids in HBB_A of *C. nippon* (sickling) with variant amino acids found in the closely related but non-sickling HBB_A sequence from *C. canadensis*, and vice versa. Fibre interaction energy is consistently higher for sequences that include 22V, regardless of the genetic background. **b,** Effects on fibre interaction energy of replacing defined single amino acids in the primary sequence of either *R. tarandus* or *O. virginianus*. All amino acids that differ between sickling and non-sickling sequences are considered. Negative values indicate stronger interactions and thus an increased likelihood of fibre formation. These values show the mean over all 270 22V-87Q docking models compatible with fibre formation, and error bars represent standard error of the mean.



Supplementary Figure 7

Codons specifying sickling-associated amino acids in HBB_A genes of different deer species. Tip labels and associated codons are coloured and arranged along the species phylogeny as in Fig. 3a to illustrate the presence of the same codons in distantly related species.

a

Recombination event	Sequence(s) with recombination signal	Recombinant source	Breakpoints*		Significance of detection by method						
			Start	End	RDP	GENECONV	BootScan	Maxchi	Chimaera	SiScan	3Seq
1	*H. inermis* (A)	Foetal	172	344	4.25E-10	2.65E-09	2.69E-10	5.26E-04	1.05E-03	1.40E-04	2.12E-07
2	*C. canadensis* (A)	Foetal	308	502	3.96E-06	3.01E-05	3.17E-06	2.60E-03	4.27E-06	1.64E-04	4.29E-06
3	*C. capreolus* (F)	Adult	344	591	4.29E-04	NS	2.04E-03	8.53E-03	1.55E-03	NS	1.34E-03
4	*D. dama* (A)	Foetal	420	580	NS	NS	NS	1.99E-03	NS	NS	NS
5	*R. tarandus* (A)	Foetal	344	534	NS	NS	NS	5.29E-03	NS	NS	NS
6	*R. tarandus* (F) *O. virginianus* (F)	Adult	355	684	NS	NS	NS	1.19E-02	6.44E-03	NS	NS
7	*R. duvaucelii* (A)	Foetal	96	170	NS	1.24E-02	NS	NS	NS	NS	NS
8	*O. virginianus* (A) *P. puda* (A)	Foetal	772	894	NS	NS	NS	2.21E-02	NS	NS	NS
9	*P. puda* (A)	Foetal	317	344	NS	NS	1.44E-02	NS	NS	NS	NS
10	*C. canadensis* (A)	Adult	1336	1466	NS	NS	4.64E-02	NS	NS	NS	NS
11	*O. virginianus* (F) *R. tarandus* (F)	Adult	156	304	NS	NS	4.84E-02	NS	NS	NS	NS

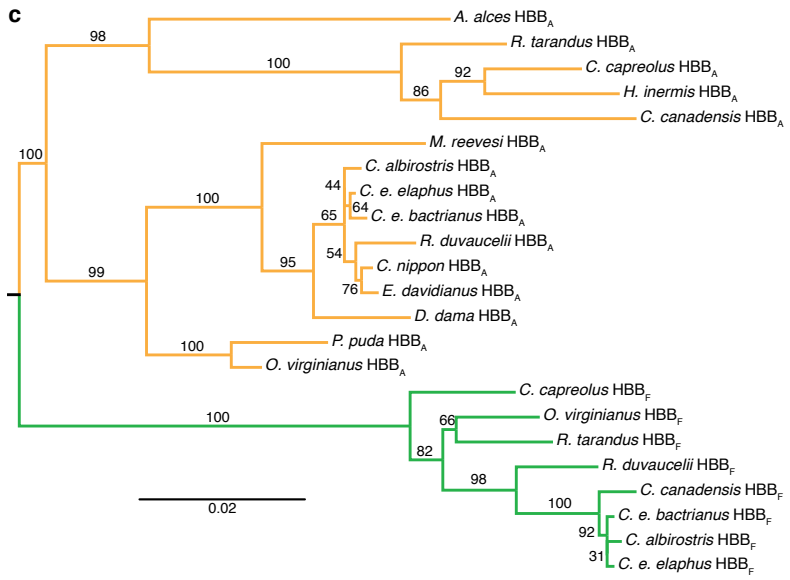
*due to high local conservation, the exact breakpoint position can be uncertain. F: HBB_F; A: HBB_A; NS: Not significant.

b

RDP option tab	RDP options different from default			
	Window size	Step size	Model	Variable sites per window
BootScan*	20	5	JN90	-
SiScan	20	5	-	-
PhylPro	40	-	-	-
VisRD	100	-	-	-
DSS(TOPAL)	40	5	JN90	-
Distance Plots	40	5	JN90	-
MaxChi	-	-	-	50
Chimaera	-	-	-	50

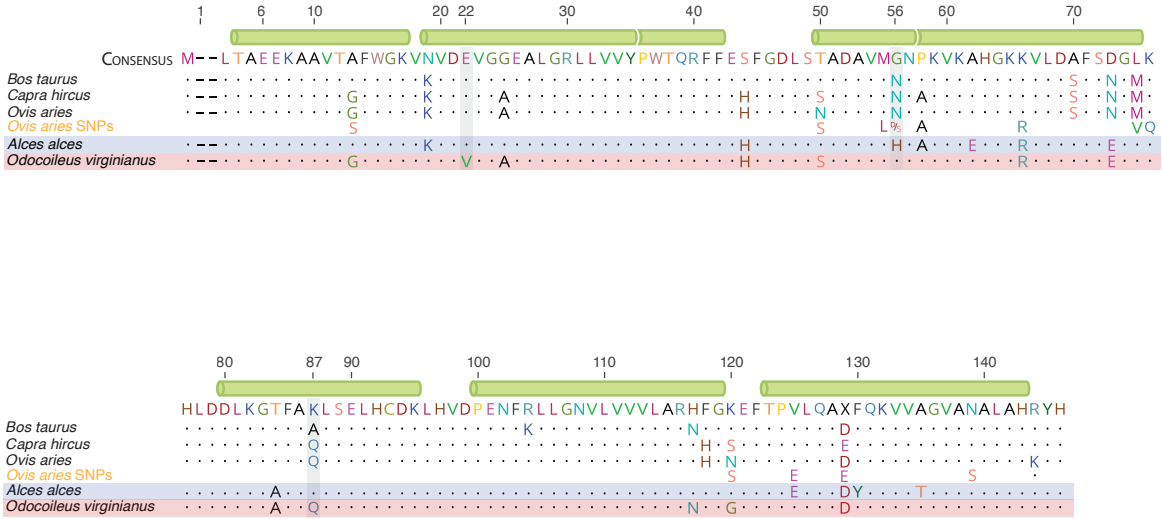
*In addition for BootScan: Relationship measure = UPGMA, Bootstrap replicates = 300

c



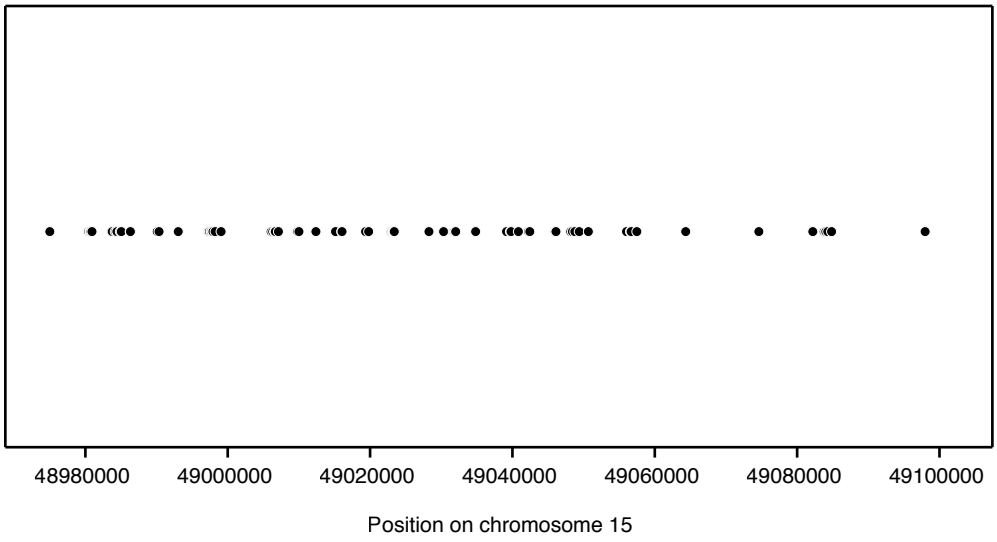
Supplementary Figure 8

Detection of gene conversion and introgression events in deer β-globin genes. a, Recombination events predicted by different methods from an alignment of deer HBB_F and HBB_A genes. Breakpoint positions are given relative to each focal sequence. Where two sequences are affected (events 6,8,11) positions refer to the top sequence. **b**, Non-default parameters used for detecting recombination events with RDP. **c**, Maximum likelihood tree derived from the alignment of adult (orange) and foetal (green) β-globin genes after predicted recombinant regions have been removed. Branch support values (%) are derived from 1000 bootstrap replicates (see Methods for further information on tree reconstruction).



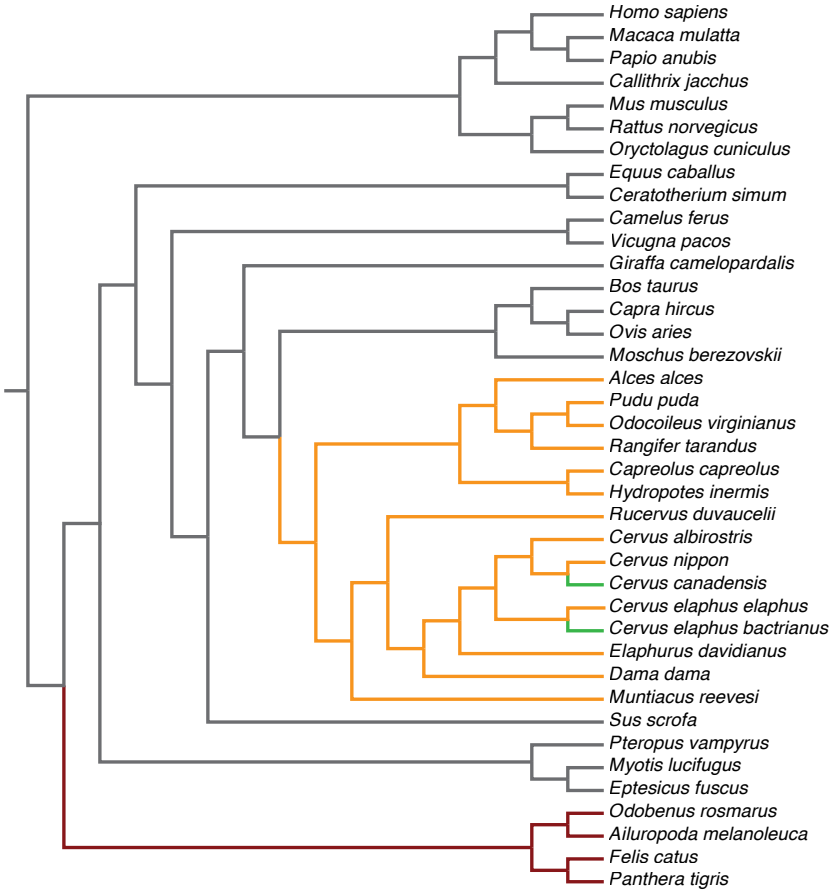
Supplementary Figure 9

Sheep HBB_A diversity. An excerpt from the alignment in Fig. 1 is shown alongside amino acid variants, found across 75 breeds of sheep, that differ from the *O. aries* reference sequence. All seven residues previously found to differentiate HbA and HbB (see main text) are recovered (50S, 58A, 75V, 76Q, 120S, 129E, 144R, where the letter indicates the amino acid found in HbA).



Supplementary Figure 10

Mapping locations in the *B. taurus* genome of reads used to seed *O. virginianus* HBB_A local assembly.



Supplementary Figure 11

Cladogram of the mammalian species phylogeny used in this study. Coloured branches indicate deviations from and additions to the Timetree of Life phylogeny (see Methods); red: re-grafted Carnivora; orange: 10kTrees deer phylogeny; green: manually added branches absent from the 10kTrees phylogeny.

Species	Common Name	Sample	Sample type	Source	Sickling state	References
Odocoileus virginianus	White-tailed deer	WTD	Blood*	Penn State Deer Research Centre, Pennsylvania PA 16802, USA	Sickling†	8,10,15,18, 20,23,86
Dama dama	Fallow deer	BFD	Blood*	ZSL Whipsnade Zoo, Whipsnade, Dunstable, LU6 2LF, UK (ZSL)	Sickling	9-11,42
Rucervus duvaucelii	Swamp deer	SD	Blood*	ZSL	Sickling†	11,42,86
Elaphurus davidianus	Père David's deer	PDD	Blood*	ZSL	Sickling	9-11,42
Cervus nippon	Sika deer	SIKA	Blood*	ZSL	Sickling†	9-11,22,42, 86,87
Cervus elaphus elaphus	Red deer	RED	Blood	RZSS Highland Wildlife Park, Kincraig, Kingussie, PH21 1NL, UK (RZSS)	Sickling	9-11,17,42
Cervus elaphus bactrianus	Bactrian deer	BACT	Blood	RZSS	Indeterminate	
Cervus albirostris	Whitelipped deer	WLD	Blood	RZSS	Indeterminate	
Alces alces	European elk (Moose)	ELK	Blood / Muscle tissue	RZSS / Kezie Foods, Duns, TD11 3TT, UK	Does not sickle	9,11,42
Rangifer tarandus	Reindeer	REIN	Blood / Muscle tissue	RZSS / Kezie Foods, Duns, TD11 3TT, UK	Does not sickle	9-11,42
Muntiacus reevesi	Reeve's muntjac	MUNT	Muscle tissue	The Wild Meat Company, Woodbridge, IP12 2DY, UK	Sickling	8,11,42
Pudu puda	Pudu	PUDU	Blood	Bristol Zoo, Bristol Zoo Gardens, Clifton, Bristol, BS8 3HA, UK	Sickling	42,85
Capreolus capreolus	Roe deer	ROE	Tissue	V. Savolainen	Indeterminate	
Hydropotes inermis	Chinese water deer	CWD	Tissue	V. Savolainen	Indeterminate‡	10,42
Cervus canadensis	Wapiti (Elk)	WAP	Genomic DNA	East Stroudsburg University, 200 Prospect St, East Stroudsburg, PA 18301, USA	Does not sickle†	9-12,42

*Fresh blood samples processed with the PAXgene Blood DNA kit.
†Both sickling and non-sickling adult individuals previously recorded.
‡Only one individual tested (non-sickling). Conservatively listed as indeterminate.

Supplementary Table 1

Species considered in this study, previous evidence for sickling and sample origins. For each sample, species identity was confirmed by sequencing the mitochondrial *cytB* gene (see Methods, Supplementary Table 2). For species in which both sickling and non-sickling individuals have been previously identified, the more common phenotype (as found in the associated references) is listed.

Supplementary Discussion

Why is sickling in deer better tolerated than in humans? – a brief note

Although patterns of evolution of HBB_A suggest that sickling is associated with fitness costs and benefits, it is striking, especially in comparison to HbS, that sickling homozygotes do not exhibit an easily measurable phenotype. As pointed out in previous literature on the subject, two features of deer erythrocytes in particular might contribute to reduced sickling cost: they are small and they are pliable.

First, the cause of haemolytic anaemia in human sickle cell patients can be traced to the increased mechanical fragility of their red blood cells (RBCs). In deer, on the other hand, sickled and non-sickled cells have the same mechanical fragility[23]. Why sickled deer erythrocytes remain highly pliable despite extensive internal polymer formation remains unknown. Future comparative studies of cytoskeletal architecture and membrane composition will be interesting in this regard.

Second, as noted by Gulliver himself, RBCs in deer appear unusually small compared to other mammals[88]. This might reduce the probability of vaso-occlusions, under the assumption that capillary diameter is constant across mammals and has not decreased proportionally in deer. We will revisit this assumption below. First, however, we want to revisit the claim that deer RBCs are unusually small in the context of allometric scaling. To do so, we intersected measurements of RBC diameters made by Gulliver for more than 200 mammals, with more recent data on body mass from the PanTHERIA database[89]. We manually matched old and new taxonomic names and re-assigned the correct Order where necessary. We only considered taxa that could be unambiguously matched to an extant taxon, excluded camelids (which have atypical oval-shaped RBCs) and added data on five additional deer species from http://www.genomesize.com/cellsize/mammals.htm.

We then considered the relationship between body mass and RBC diameter. There is a persistent claim in the literature, prominently advocated by Schmidt-Nielsen[90] and subsequently West and colleagues[91], that RBC size is body mass invariant. This indeed appears to be the case when all mammals are considered together (left-hand panel in Figure A below). However, we find strong scaling relationships within different mammalian Orders. Notably, scaling coefficients are indistinguishable for Rodentia, Primates, and Carnivora (middle panel in Figure A), whereas Artiodactyla – which include deer – exhibit deviant scaling behaviour, with RBC size dropping faster with decreasing body mass than in other mammalian orders. It therefore appears that the relationship between RBC size and body mass in artiodactyls is governed by a different design principle than in other mammals. However, this constitutes insufficient evidence to claim that small RBC size in deer mitigates sickling effects. As highlighted above, this logic hinges on the assumption that – while RBC size drops – terminal capillary size (e.g. in lung alveoli) remains constant. Is this true? The same authors that argued for body mass-invariant RBC size assume that terminal capillary diameter is body size invariant[92] but there is no concrete empirical data to support this claim, which is extrapolated from a cross-species study of nephron diameters in the kidney. In light of the fact that RBC size does, in fact, vary with body mass, as we demonstrate here, we think assertions about invariant capillary diameters across mammals are premature. Thus, although artiodactyls do show unusually small

RBCs and a different scaling relationship with body mass, in the absence of corresponding data on capillary diameters, we cannot establish whether this would impact the probability of vaso-occlusion. What we can say, however, is that RBC size of sickling deer does not scale any differently from that of non-sickling deer (right-hand panel in Figure A).

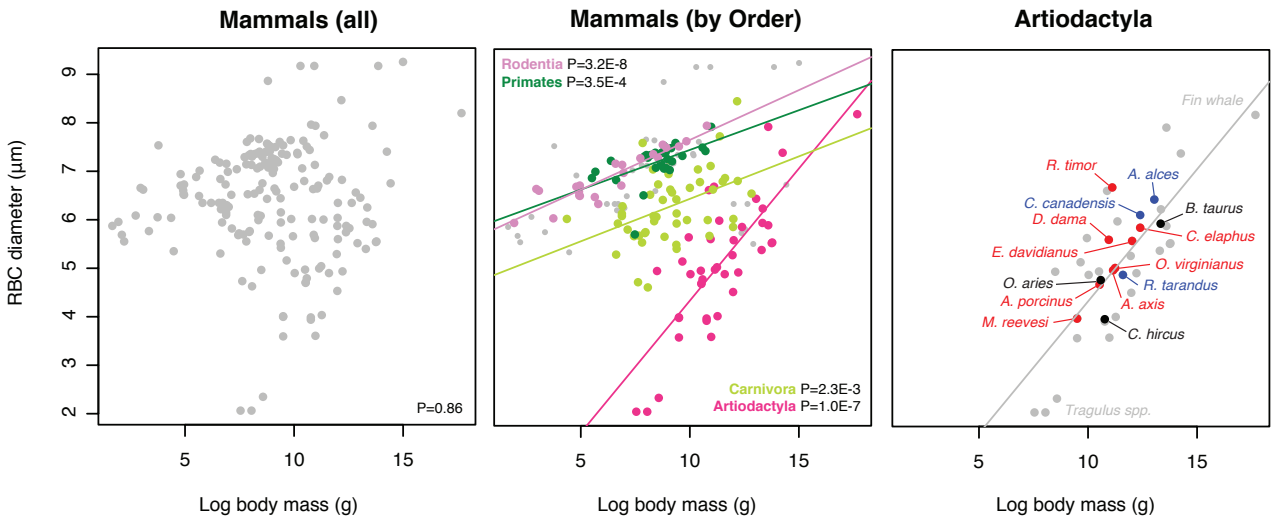


Figure A

Supplementary References

85. Junge, R. E., Duncan, M. C., Miller, R. E., Gregg, D. & Kombert, M. Clinical presentation and antiviral therapy for poxvirus infection in pudu (Pudu puda). *Journal of Zoo and Wildlife Medicine* **31,** 412–418 (2000).
86. Weisberger, A. S. The Sickling Phenomenon and Heterogeneity of Deer Hemoglobin. *Proceedings of the Society for Experimental Biology and Medicine* **117,** 276–280 (1964).
87. Ogawa, E., Hasegawa, M., Fujise, H. & Kobayashi, K. Erythrocyte Sickling in Sika Deer (Cervus nippon). *J. Vet. Med. Sci.* **53,** 1075–1077 (1991)
88. Gulliver, G. Observations on the sizes and shapes of the red corpuscles of the blood of vertebrates, with drawings of them to a uniform scale, and extended and revised tables of measurements. *Proceedings of the Zoological Society of London* **1,** 474–495 (1875).
89. Jones, K. E. *et al*. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90,** 2648–2648 (2009).
90. Schmidt-Nielsen, K. *Scaling*. (Cambridge University Press, 1984).
91. Savage, V. M. *et al*. Scaling of number, size, and metabolic rate of cells with body size in mammals. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 4718–4723 (2007).
92. West, G. B., Brown, J. H. & Enquist, B. J. A General Model for the Origin of Allometric Scaling Laws in Biology. *Science* **276,** 122–126 (1997).