



OPEN

The Poses for Equine Research Dataset (PFERD)

DATA DESCRIPTOR

Ci Li¹, Ylva Mellbin², Johanna Krogager², Senya Polikovskiy³, Martin Holmberg⁴, Nima Ghorbani^{5,7}, Michael J. Black³, Hedvig Kjellström^{1,2}, Silvia Zuffi⁶ & Elin Hernlund²✉

Studies of quadruped animal motion help us to identify diseases, understand behavior and unravel the mechanics behind gaits in animals. The horse is likely the best-studied animal in this aspect, but data capture is challenging and time-consuming. Computer vision techniques improve animal motion extraction, but the development relies on reference datasets, which are scarce, not open-access and often provide data from only a few anatomical landmarks. Addressing this data gap, we introduce PFERD, a video and 3D marker motion dataset from horses using a full-body set-up of densely placed over 100 skin-attached markers and synchronized videos from ten camera angles. Five horses of diverse conformations provide data for various motions from basic poses (eg. walking, trotting) to advanced motions (eg. rearing, kicking). We further express the 3D motions with current techniques and a 3D parameterized model, the hSMAL model, establishing a baseline for 3D horse markerless motion capture. PFERD enables advanced biomechanical studies and provides a resource of ground truth data for the methodological development of markerless motion capture.

Background & Summary

Over the years, capturing and modeling the articulated motion of humans and animals has been a research topic across different disciplines, ranging from medicine¹ to robotics² to computer graphics³. Humans and animals integrate verbal communication with body posture and movements, and understanding body language would greatly advance the design of intelligent interactive artificial systems. Detecting anomalies in articulated motion⁴ can serve as a crucial tool for early health intervention, helping to mitigate potential long-term injuries or disease, hence improving the subject's quality and length of life. Systems able to synthesize realistic motion can be a useful tool for artists to create characters for games and virtual worlds⁵. Recently, generative AI methods have been wildly trained to synthesize human motion⁵⁻⁷.

We focus our study on horses. Horses have historical and cultural significance in most human societies and are one of the oldest domesticated mammals. They have played a significant role in human history, from transportation to warfare, from agriculture to culture, from work to equestrian sports. Their significance is related to their unique bodily strength and speed, made possible by an efficient quadruped locomotor system. This body system has evolved to have bulky muscles close to the upper body, proportionally long limbs that act like pogo-sticks thanks to specialized tendons and a mass reduction of the lower limb and foot that effectively reduces inertia. This optimized locomotor apparatus is however susceptible to injuries as it operates under loading conditions close to its point of failure⁸. Thus, studying the motions of horses has been the focus of researchers in different fields such as robotics^{9,10}, biology^{11,12} and veterinary medicine^{13,14}.

Marker-based motion capture systems are widely used to capture complex human and animal motions by recording the positions of wearable markers placed on the body in an indoor environment. Marker-based motion capture has demonstrated its significance for motion study¹⁵⁻¹⁷ and utility across a wide array of applications^{14,18}. While marker-based solutions are in need of physical contact with the animal and not scalable to in-the-wild scenarios, computer vision techniques have been developed to implement markerless motion capture, where the articulated motion of a skeleton is inferred from visual data¹⁹⁻²⁹. With images and video as the input, straightforward systems provide the image coordinates of skeleton joints as solutions¹⁹⁻²¹. This is not sufficient for many high-quality downstream applications, where a solution independent from the capture geometry

¹KTH Royal Institute of Technology, Stockholm, Sweden. ²Swedish University of Agricultural Sciences, Uppsala, Sweden. ³Max Planck Institute for Intelligent Systems, Tübingen, Germany. ⁴Qualisys, Göteborg, Sweden. ⁵Sporttotal.tv, Immersive Technologies, Cologne, Germany. ⁶CNR Institute for Applied Mathematics and Information Technologies, Milan, Italy. ⁷This work was done while the author was at MPI-IS. ✉e-mail: elin.hernlund@slu.se

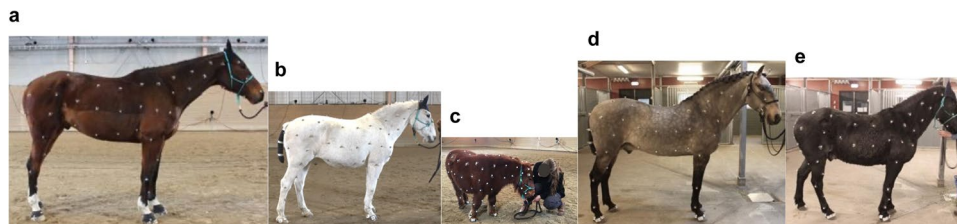


Fig. 1 Five horses with different sizes and breeds. (a) Horse No. 1. (b) Horse No. 2. (c) Horse No. 3. (d) Horse No. 4. (e) Horse No. 5.

is often required. In these cases, a 3D articulated pose, given as the set of 3D rotation angles of skeleton joints, is preferable.

3D markerless motion capture for humans is a novel technology, accurate enough to be used in many applications. In particular, monocular markerless capture, where the 3D pose of the subject is inferred from just one camera, allows designing applications that can exploit low-cost capture devices like smartphones. The basis of these achievements is data-driven methods that leverage large amounts of captured human data. While 2D methods use data in the form of large image datasets with body joint annotations, which are easy to obtain, the data capture task for learning systems outputting 3D poses is significantly more challenging. Moreover, estimating 3D pose from a single view is ambiguous. The problem can be approached by learning explicit 3D pose priors, that constrain the ambiguous solutions to the more likely ones, or learning implicit 3D priors from annotated datasets. The first solution makes use of decoupled data, usually image or video datasets of humans and 2D joints, which are also used for 2D pose estimation, and datasets of human motions. The second solution requires large datasets of images and corresponding 3D poses. These datasets can be obtained at a large scale but only synthetically. Notably, the synthetic dataset generation still requires human articulated motion data. However, relying solely on the priors of human articulated motion is insufficient to achieve markerless motion capture. Seen from the camera, a long leg pointed to the camera can have the same appearance as a shorter leg bent with a different angle. To deal with these ambiguities, a 3D shape prior, encoding the correlation between the body segment proportions, is required.

In the last few years, we have seen a tremendous advance in 3D markerless motion capture for humans. This has been facilitated by the availability of the SMPL (Skinned Multi Person Linear) model³⁰, a 3D parametric model, learned from thousands of 3D scans of people, encoding articulated human shape, and the AMASS dataset³¹, a large dataset of human articulated motion, captured with mocap systems, expressed in the parameters space of the SMPL model. Together, SMPL and AMASS incorporate knowledge about how people appear in shape and how they move^{7,32–37}.

Animal motion capture has been making strides in recent years but is still behind human motion studies. Analogous to the SMPL model, the SMAL (Skinned Multi Animal Linear) model³⁸, learned from 41 toy scans, encodes articulated shapes of quadruped animals. Versions of the SMAL model have been made specifically for dogs^{39,40} and horses⁴¹. Nonetheless, the data collection on animals is more challenging compared to humans, since it is more difficult to instruct them to perform specific motions and to keep them in fragile indoor environments. Existing animal mocap datasets prefer docile and small animals^{42–45}, but are constrained by pose variability and lack the motion diversity that AMASS offers. This results in a paucity of comprehensive animal motion datasets for data-driven motion study, particularly for larger animals, like horses. In the equine veterinary field, motion capture has demonstrated its potential in clinic applications for lameness diagnostics^{46,47}. However, within this field, the focus is often limited to capture of locomotion data from a limited number of anatomical landmarks^{48–50} due to difficulty and time constrain in placing markers on the horse's body. This may lead to less analysis and a lesser understanding of full-body motions.

To bridge the gap, we introduce PFERD⁵¹, a dense motion capture dataset of horses of diverse conformation and poses with rich 3D horse articulated motion data. Recorded in an indoor riding arena in Sweden, using an optical motion capture system from the company Qualisys, the dataset includes five horses of different sizes and breeds, to ensure shape diversity (Fig. 1). Over 100 reflective markers were placed on each horse, covering both skeletal structures and soft tissues, to accurately capture motions. The dataset covers a wide variation of horse motions, guided by human instructors, ranging from basic activities like standing, walking, and trotting, to complex motions like the piaffe, the passage, the pirouett, jumping, sitting as shown in Fig. 3. Two highly trained horses perform these advanced motions while the rest of the subjects provide common gaits and motions encountered in every-day horses. Furthermore, to promote the study of markerless motion capture, we provide multiple data types. In line with AMASS, we express 3D horse articulated motion with the hSMAL model⁴¹. The dataset further enriches its data diversity by including synchronized videos from ten camera views and corresponding 2D joints.

The PFERD dataset serves as an open resource for equine motion research and for the scientific development of computer vision and modeling applications that can benefit horse health and welfare and strengthen our understanding of horse behavior. It provides synchronized 3D data from skin-placed markers and multi-view 2D RGB video streams. The dataset is small in terms of subject numbers, but unique thanks to the wealth of markers placed on the horses' bodies as well as the subject variation in size, shape, and the rare motions that some of the horses perform. The dataset can be expanded given the detailed descriptions of data capture and model estimation procedures. With this data, we invite researchers to develop both statistical analysis and data-driven methods. We suggest the following tasks:

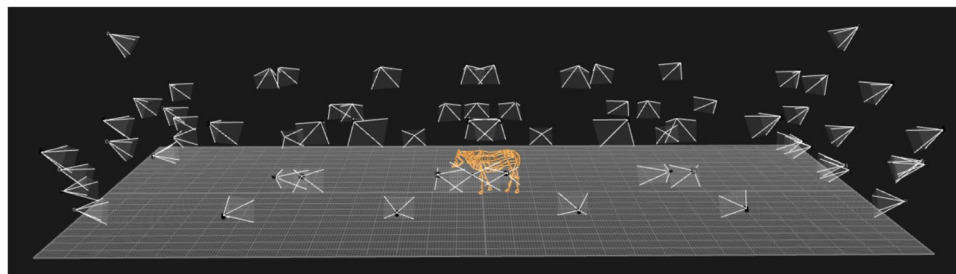


Fig. 2 Recording scenario with 56 mocap cameras and ten color cameras using Qualisys system. The system recognizes the markers and the orange lines are connections between the markers.

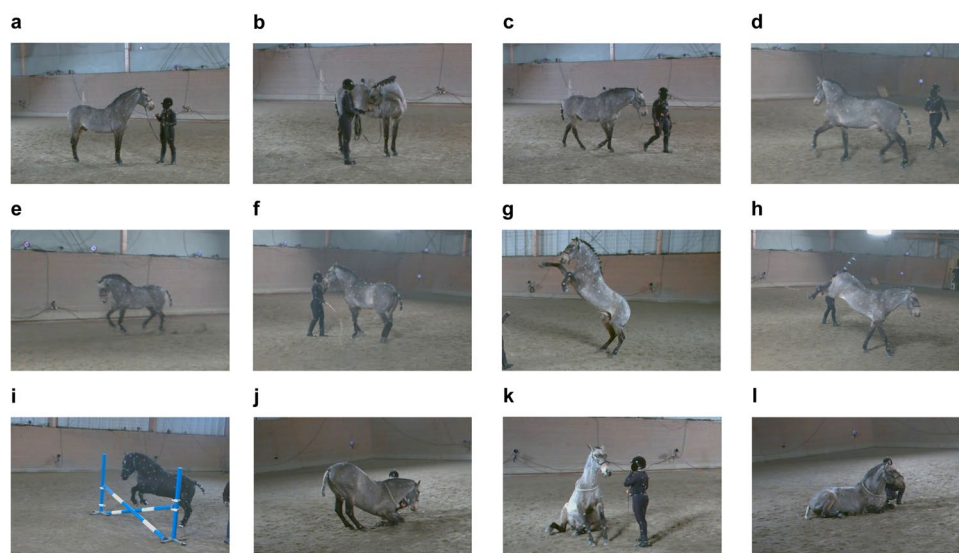


Fig. 3 Motion examples. (a) Standing. (b) Neck bending. (c) Walking. (d) Trotting. (e) Cantering. (f) Piaffing. (g) Rearing. (h) Kicking. (i) Jumping. (j) Lying down. (k) Sitting. (l) Lying.

- 1) Quantitative motion analysis: The diverse motions and full body marker setup allow for detailed biomechanical studies. The mocap data are, at the time of publication, the most marker-dense horse motion dataset available for research in horses. This data can help veterinary researchers to understand full body motions at a more detailed level than was earlier possible, and it contains unique movements that only highly trained horses can perform.
- 2) Addressing animal-related computer vision problems: Researchers can utilize this dataset to develop novel computer vision models or refine existing algorithms, promoting the development of markerless motion capture. The 3D motion data can be further used for many graphic tasks, like the development of motion generative models and the improvement of 3D animatable models.
- 3) Benchmarking in method evaluation through provided groundtruth data: The dataset contains precise 3D mocap data and multi-view RGB video, and can provide both 3D and 2D groundtruth for method evaluation, helping the development of new methods for 3D pose and shape estimation. Furthermore, it enables benchmarking of differences against the state-of-the-art methods we apply here^{31,52}.

Methods

In this section, the procedures used to record the data are explained. In addition, the processing of mocap data and 3D pose data are presented.

Study subjects. The dataset has a diversity in terms of body shape and motion. We selected five horses of different breeds to provide variable information on shape and size. In terms of diverse motions, all horses performed some basic movements, such as standing (with the head moving from side to side, up and down), moving forward/backward, walking, trotting, and cantering. Two of the horses (Horse No.4 and No.5) performed advanced movements based on signaling cues from their owners, such as pirouetting, rearing, piaffing, kicking, jumping, etc. Table 1 shows the detailed characteristics of the five horses and some motions are listed in Fig. 3.

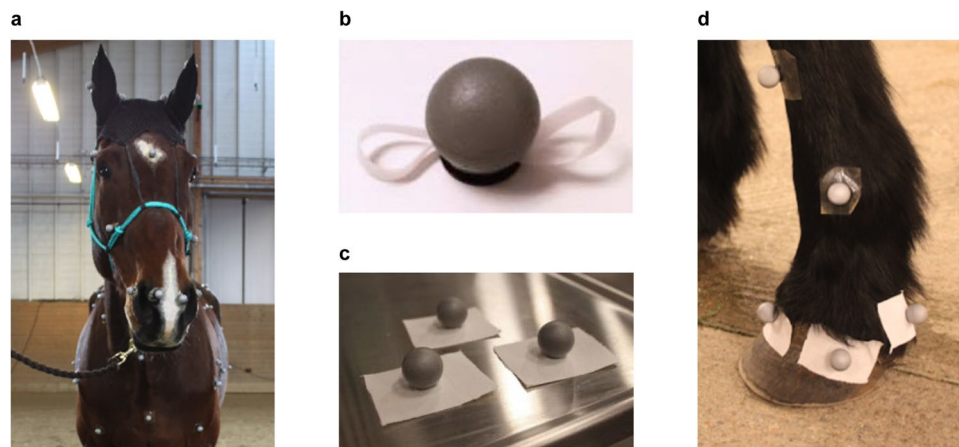


Fig. 4 Modified marker attachment methods. **(a)** Placement on the head: Markers on a soft halter and cap. **(b)** Placement on the mane: Markers with incorporated tassels. **(c)** Placement on the hooves: Markers with a piece of fabric in between the basis and marker itself. **(d)** Example of marker placements on the hoof.

Before each subject was selected, the horse owners were introduced to the aim of data collection and informed about the procedures. Written informed consent was obtained from the owners, permitting the use of the horses' data for research purposes. The study was non-invasive and the procedure was covered by an animal ethical permission No. 5.8.18-15533/2018. Written consent was provided by all humans appearing in the video recordings.

Experimental design. In this subsection, the description of the mocap system and marker setup are presented.

Motion capture system. The data were collected using Qualisys optical motion capture system on November 26–29, 2020. The system was set up in a riding arena of approximately $19 \times 30\text{m}$ at the Equine clinic of the University Animal Hospital (UDS) of the Swedish University of Agricultural Sciences (SLU) in Uppsala, Sweden. In total 56 mocap cameras from the Qualisys system (35 Oqus_700 + cameras and 21 Arqus_A12 cameras) and ten RGB full HD video cameras (Miquis_Video cameras) were mounted to the walls of the arena shown in Fig. 2. All cameras were synchronized creating an approximate $16 \times 20\text{m}$ effective recording volume in the center of the arena. The capture rate of the mocap cameras was 240 Hz and the RGB videos captured by cameras were 20,30,60 Hz, respectively, depending on the data recording.

Marker placement and attachment methods. Reflective spherical markers with a diameter of 19 mm were attached to the horses' skin with double-coated adhesive tape cut in pieces of around $2 \times 3\text{cm}$. Different methods were empirically tested for the more challenging attachment markers to body parts such as the ears and the hooves, shown in Fig. 4.

The marker setup aimed to maximize both captures of body shape and motion and included 132 markers on both skeletal structure and soft tissues. Based on expert anatomical knowledge, markers were separated into three groups. The first group with 50 markers focuses on the precise palpation of anatomical skeletal structures that mark out the most important skeletal segments related to locomotion. Connecting these markers provides a “stick figure”, roughly representing skeletal movement from landmarks on the skin surface. We call these markers the “skeletal model”, see Fig. 5a. The second group of around 70 markers, were dispersed over the horse's soft tissues, mainly covering the area of the neck, the thoraco-abdominal, and hindquarter segments. The third group of 12 markers were placed in groups of three on each hoof, to allow tracing of rotational motion of the hooves. The final full body marker setup is shown in Fig. 5b. Detailed descriptions are reported in Table 3.

Data acquisition. In this subsection, the whole procedure of data recording is explained, including mocap system calibration, subject preparation, and type of motion performances recorded.

Qualisys calibration. Calibration of the motion capture system was done with wand calibration, according to the manufacturer's instructions. The video cameras were calibrated along with the marker cameras. For the first calibration, an L-shaped frame with static markers was placed in the approximate center of the capture volume to define the coordinate system. Then a calibration wand with two markers at a fixed distance was moved through the volume to present it to all cameras at different angles. Subsequent calibrations were done with only the calibration wand. The system was recalibrated before recording the first, second and fourth subject.

Study subject preparation. The fur and the hooves of horses were washed with soap water before attaching markers. Markers were cleaned between different trials if needed. Markers in the skeletal model were placed by palpating specific skeletal structures by two people with anatomical knowledge for precise positioning. The

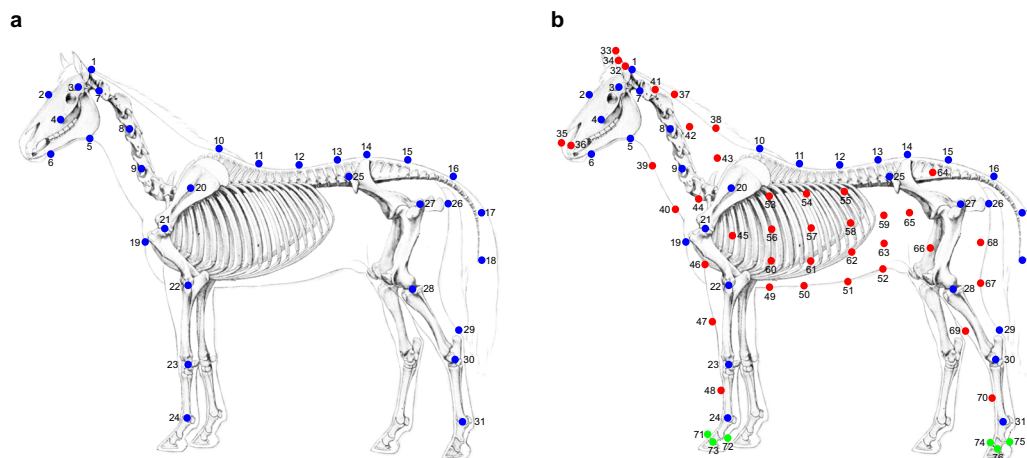


Fig. 5 Design marker setup (Modified figure from⁵⁵). **(a)** Skeletal model (in blue), focusing on crucial skeletal segments for locomotion. **(b)** Final model, including skeletal model and markers on soft tissues (in red) and hooves (in green) over the surface.

HorseID	Breed	Ages (years)	Gender	Weight (kg)	Wither height (cm)	Marker Num
1	Swedish Warmblood	9	Gelding	702	178	131
2	Irish Pony	22	Gelding	378	130	132
3	Miniaure Shetland Pony	4	Gelding	110	76	120
4	Lusitano	17	Stallion	492	157	132
5	Connemara	17	Gelding	407	142	117

Table 1. Information about the five horses captured in PFERD.

Horse ID	SeqNum	Total Length (unit: s)	Average Distance (unit: m)	FrameNum of Silhouette Subset	Average IOU
1	12	887	0.034	17610	0.86
2	9	720	0.032	34878	0.85
3	5	373	0.028	4478	0.79
4	13	1068	0.030	13718	0.86
5	7	420	0.029	6124	0.85
Total	46	3468	0.031	76808	0.85

Table 2. Detailed data and evaluation metrics for the five horses: includes sequence counts, total data lengths, and frame numbers of silhouette subsets for each horse; evaluations report average 3D distances and average IOU.

remaining markers were located on the body segment's proportions using nearby skeletal markers and tape measure as references shown in Fig. 6. Each horse took 2-3 hours to finish all the preparations. The number of markers per horse (shown in Table 1) was a bit different since certain markers had to be excluded for various reasons. For example, Horse No.3 was a small horse, and we had to reduce the number of markers to avoid marker merging, or label-swapping due to them being too close on the small body. Horse No.5 was sweating, resulting in markers not being attached properly, especially markers on the lower belly.

Recording sessions. During the recording session for each horse, the horse was first led to the recording arena by the owner and was familiarized with the recording environment. Then, the owner engaged with the horse, using methods such as whistling, waving the whip, and offering treats to guide the horse into performing specific motions. The movements began with fundamental actions, like standing, neck bending, moving forward or backward, as well as, walking, trotting and cantering. The complexity of the movements varied depending on the horse's ability. For instance, Horse No.4 demonstrated more advanced movements like pirouetting, rearing, piaffing, passage, spanish walk and kicking, while Horse No.5 was rearing and jumping over an obstacle. The dataset comprised different numbers of data sequences for each of the five horses, ranging from 5 to 13, shown in Table 4. While most data recordings lasted approximately one minute, there were exceptions. Each data recording captured more than one motion, allowing for a diverse range of horse movements.

Data processing. In this subsection, we describe the data process from Qualisys and the procedure of learning the 3D model from the mocap data using the hSMAL model and MoSh++.

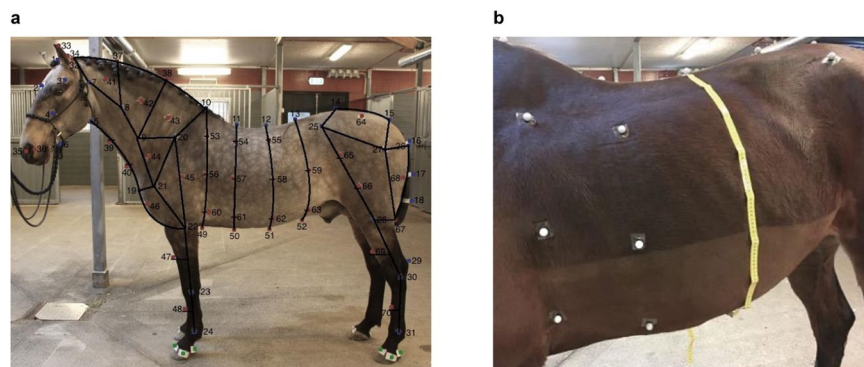


Fig. 6 Marker placement measurement on real horses. **(a)** Marker placement and corresponding proportion on the horse's body. **(b)** A measuring tape was used for marker placement on the barrel.

Number	Description	Labels	Landmarks For Marker Placement
1	Poll	Centre	Nuchal crest.
2	Forehead	Centre	Frontal bone between the eyes.
3	Temple	Left/Right	Zygomatic arch.
4	Cheek	Left/Right	Rostral part of facial crest.
5	Jaw	Left/Right	Ventral part at the angle of mandibula.
6	Chin	Left/Right	Ventral part of mandibula at the location of the tooth P2/at the chin groove.
7	Neck_1	Left/Right	Wing of Atlas vertebrae.
8	Neck_2	Left/Right	Transverse process of cervical vertebrae of C3.
9	Neck_3	Left/Right	Transverse process of cervical vertebrae of C5.
10	Back_1	Centre	Dorsal spinal process of thoracic vertebrae of TH5(6) (withers).
11	Back_2	Centre	Dorsal spinal process of thoracic vertebrae of TH12.
12	Back_3	Centre	Dorsal spinal process of thoracic vertebrae of TH18.
13	Back_4	Centre	Dorsal spinal process of lumbar vertebrae of L4.
14	Back_5	Centre	In between the two tuber sacrale.
15	Back_6	Centre	Dorsal spinal process of sacral vertebrae of S5.
16	Tail_1	Centre	Placed at the dorsal spinal process of the caudal vertebrae at the first third part of the distance from Marker15 to Marker18.
17	Tail_2	Centre	Placed at the dorsal spinal process of the caudal vertebrae at the second third part of the distance from Marker15 to Marker18.
18	Tail_3	Centre	Dorsal spinal process of the last of the caudal vertebrae (C15-C21).
19	Chestpoint	Centre	Dorsal part of cartilage of manubrium.
20	ShoulderBlade	Left/Right	Tubercle of spine of scapula.
21	ShoulderJoint	Left/Right	Caudal part of greater tubercle of humerus.
22	ElbowJoint	Left/Right	Lateral collateral ligament over the joint space of the elbow joint.
23	CarpalJoint	Left/Right	Ulnar carpal bone.
24	Foreleg_FetlockJoint	Left/Right	Lateral collateral ligament over the joint space of the fetlock joint on the fore leg.
25	Pelvis_1	Left/Right	Dorsal cranial part of the coxial tuberosity of pelvis.
26	Pelvis_2	Left/Right	Caudal part of the ischiatic tuberosity.
27	HipJoint	Left/Right	Greater trochanter of the femur.
28	StifleJoint	Left/Right	Lateral collateral ligament over the joint space of the stifle joint.
29	PointOfHock	Left/Right	Dorsal part of tuber calcaneus.
30	HockJoint	Left/Right	Tarsal bone IV.
31	Hindleg_FetlockJoint	Left/Right	Lateral collateral ligament over the joint space of the fetlock joint on the hind leg.
32	Ear_Base	Left/Right	Base of the ear.
33	Ear_Top	Left/Right	Top of the ear.
34	Ear_Side	Left/Right	Lateral side of the middle part of the ear.
35	Nostril_In	Left/Right	Medial wing of the nostril.
36	Nostril_Out	Left/Right	Lateral wing of the nostril.
37	Neck_Top_1	Centre	In the main dorsally on the neck of the first third part from the distance of Marker1 to Marker10.
38	Neck_Top_2	Centre	In the main dorsally on the neck of the second third part from the distance of Marker1 to Marker10.

Continued

Number	Description	Labels	Landmarks For Marker Placement
39	Neck_Bottom_1	Centre	Ventrally on the neck of the first third part from the distance of Marker5 to Marker19.
40	Neck_Bottom_2	Centre	Ventrally on the neck of the second third part from the distance of Marker5 to Marker19.
41	Neck_Up_1	Left/Right	An approximate site in the central part of the area shaped by the markers of 1, 7, 8 and 37.
42	Neck_Up_2	Left/Right	An approximate site in the central part of the area shaped by the markers of 8, 9, 37 and 38.
43	Neck_Up_3	Left/Right	An approximate site in the central part of the area shaped by the markers of 9, 10, 20 and 38.
44	Shoulder_1	Left/Right	An approximate site between the markers of 9, 20 and 21.
45	Shoulder_2	Left/Right	In between the markers of 20 and 22.
46	Chest	Left/Right	At an approximate site in the central part of the area shaped by the markers of 19, 21 and 22.
47	Foreleg_Front_1	Left/Right	Cranial side of the tibia bone in between the markers of 22 and 23.
48	Foreleg_Front_2	Left/Right	Dorsal side of the metacarpal bone III of the foreleg in between the markers of 23 and 24.
49	Belly_1	Centre	Ventral part of the belly at the opposite location of Marker10.
50	Belly_2	Centre	Ventral part of the belly at the opposite location of Marker11.
51	Belly_3	Centre	Ventral part of the belly at the opposite location of Marker12.
52	Belly_4	Centre	Ventral part of the belly at the opposite location of Marker13.
53	Barrel_1	Left/Right	At the first fourth part of the distance from Marker10 to Marker49.
54	Barrel_2	Left/Right	At the first fourth part of the distance from Marker11 to Marker50.
55	Barrel_3	Left/Right	At the first fourth part of the distance from Marker12 to Marker51.
56	Barrel_4	Left/Right	At the second fourth part of the distance from Marker10 to Marker49.
57	Barrel_5	Left/Right	At the second fourth part of the distance from Marker11 to Marker50.
58	Barrel_6	Left/Right	At the second fourth part of the distance from Marker12 to Marker51.
59	Barrel_7	Left/Right	At the second fourth part of the distance from Marker13 to Marker52.
60	Barrel_8	Left/Right	At the third fourth part of the distance from Marker10 to Marker49.
61	Barrel_9	Left/Right	At the third fourth part of the distance from Marker11 to Marker50.
62	Barrel_10	Left/Right	At the third fourth part of the distance from Marker12 to Marker51.
63	Barrel_11	Left/Right	At the third fourth part of the distance from Marker13 to Marker52.
64	Croup_1	Left/Right	At an approximate site in the central part of the area shaped by the markers of 14, 15, 25 and 27.
65	Thigh_1	Left/Right	At the first third part of the distance from Marker25 to Marker28.
66	Thigh_2	Left/Right	At the second third part of the distance from Marker25 to Marker28.
67	Stifle_Back	Left/Right	Caudal side of the leg at the height level of Marker28.
68	Croup_2	Left/Right	In between the markers of 26, 27 and 67.
69	Hindleg_Front_1	Left/Right	Cranial side of the tibia bone in between the markers of 28 and 30.
70	Hindleg_Front_2	Left/Right	Dorsal side of the metacarpal bone III of the hind leg in between the markers of 30 and 31.
71	Hoof_Front_ForeLeg	Left/Right	Dorsal and proximal part of the hoof wall on the front leg.
72	Hoof_Back_ForeLeg	Left/Right	Palmar and proximal part of the hoof bulb on the front leg.
73	Hoof_Side_ForeLeg	Left/Right	Lateral and proximal part of the hoof wall on the front leg.
74	Hoof_Front_Hindleg	Left/Right	Dorsal and proximal part of the hoof wall on the hind leg.
75	Hoof_Back_Hindleg	Left/Right	Plantar and proximal part of the hoof bulb on the hind leg.
76	Hoof_Side_Hindleg	Left/Right	Lateral and proximal part of the hoof wall on the hind leg.

Table 3. Description and placement of marker setup on the horse's body.

Mocap data. The motion capture data was collected using Qualisys Track Manager (QTM) version 2020.3. The collected 2D data were combined into 3D trajectories using the tracking algorithms in QTM, and the trajectories were then labeled. The labeled data has been exported to c3d and fbx format. Since markers might fall off or be occluded during a capture, the number of labeled trajectories might vary slightly for the same horse. In the No.8 trial capture of Horse No.5, some miscommunication with the camera system resulted in a lot of short gaps in the data. Therefore, linear interpolation of the marker position has been used to fill single frame gaps in this one measurement. In the other measurements, no gap fill has been used. The camera calibration information was exported from QTM, including the extrinsic and intrinsic parameters.

2D keypoints and silhouette extraction. For 2D joint extraction, the 3D mocap data was projected onto each image frame from every camera view. This process utilizes the corresponding camera parameters, which are exported from Qualisys, and aligned the first frame of the mocap files to the initial video frame. Considering the differing framerates between the c3d files and videos, a downsample of the c3d files was performed to synchronize the c3d and video frames.

For silhouette extraction, Track Anything⁵³, one of the state-of-the-art segmentation models, was employed to extract the 2D silhouettes of each horse in each video frame. The method operated as follows: every five frames within each video, Segment Anything (SAM)⁵⁴ extracted the horse's mask for the first frame, using the bounding box calculated from 2D key points. Track Anything then applied the results from SAM as a template mask to guide the five-frame segmentation. To ensure the quality of the segmentation, we selected 130 video sequences and manually excluded instances with occlusion or incomplete body visibility.

Name	Frame Num	Detected Marker	Motion Description
20201128_ID_1_0001	19200	131	Walking a few steps backward, forward, and side to side.
20201128_ID_1_0002	19200	131	Standing still. Moving head side to side.
20201128_ID_1_0003 ^{1,3}	19200	131	Standing still. Moving head side to side.
20201128_ID_1_0004	19200	131	Walking on straight lines. Walking in curves.
20201128_ID_1_0005	19200	131	Eating from the ground.
20201128_ID_1_0006	19200	131	Walking and trotting in small circles. Walking a few steps backward.
20201128_ID_1_0007	19200	131	Walking, trotting, and canter in small circles.
20201128_ID_1_0008 ³	19200	131	Trotting in straight lines and in curves.
20201128_ID_1_0009	1680	131	Walking, trotting, and cantering in circles.
20201128_ID_1_0010	19200	131	Cantering in circles. Gait transitions.
20201128_ID_1_0011 ³	19200	131	Totting in straight lines. Gait transitions. Walking a few steps backward.
20201128_ID_1_0012 ³	19200	131	Cantering in straight lines. Walking a few steps backward.
20201128_ID_2_0005	19200	132	Walking a few steps backward, forward, and side to side. Moving head side to side and down.
20201128_ID_2_0006	19200	132	Walking in a straight line and with turns.
20201128_ID_2_0007	19200	132	Eating from the ground.
20201128_ID_2_0008	19200	132	Walking with many turns.
20201128_ID_2_0009 ³	19200	132	Trotting and cantering in big circles.
20201128_ID_2_0010	19200	132	Walking and trotting in curves and in big circles.
20201128_ID_2_0011	19200	131	Trotting and cantering in big circles.
20201128_ID_2_0012	19200	132	Trotting and cantering in big circles. Gait transitions. Walking a few steps backward and forwards.
20201128_ID_2_0013	19200	132	Walking and trotting in curves.
20201128_ID_3_0001 ²	32116	116	Walking and trotting in big circles.
20201128_ID_3_0002	14400	116	Walking a few steps backward, forward, and side to side. Moving head up.
20201128_ID_3_0003	14307	117	Moving head side to side. Walking and trotting in circles.
20201128_ID_3_0004	14400	117	Trotting and cantering in circles. Walking a few steps backward.
20201128_ID_3_0005 ³	14400	117	Trotting and cantering in straight lines and circles. Stop from cantering.
20201129_ID_4_0002	19200	130	Walking a few steps backward, forward, and side to side. Moving head side to side.
20201129_ID_4_0003	19200	131	Moving head down and up. Walking.
20201129_ID_4_0005	19200	130	Standing still. Eating from the ground.
20201129_ID_4_0006	55845	132	Standing still. Eating from the ground. Walking and trotting in small circles. Pirouetting.
20201129_ID_4_0007	20663	132	Trotting in small circles around a human. Walking a few steps backward. Pirouetting and rearing.
20201129_ID_4_0008	19200	132	Piaffing/passage (collected trotting). Rearing. Standing still.
20201129_ID_4_0009	14400	131	Piaffing. Rearing.
20201129_ID_4_0010	14400	131	Rearing. Pirouetting. Standing still.
20201129_ID_4_0011	14400	132	Trotting. Kicking with both hindlegs in trotting. Canter. Rearing. Pirouetting.
20201129_ID_4_0012	14400	129	Spanish walk. Passage. Standing still. Trotting.
20201129_ID_4_0013	14400	130	Pirouetting. Trotting. Kicking with both hindlegs in trotting. Standing still. Canter.
20201129_ID_4_0019 ³	14400	111	Lying. Standing up. Walking. Lying down.
20201129_ID_4_0020	16799	109	Sitting. Standing up.
20201129_ID_5_0003	14400	116	Walking, trotting and cantering in small circles.
20201129_ID_5_0004 ³	14400	117	Rearing. Walking. Trotting.
20201129_ID_5_0007 ³	14400	116	Rearing. Walking. Trotting.
20201129_ID_5_0008 ³	14400	115	Standing. Moving head side to side, down and up. Walking.
20201129_ID_5_0009 ³	14400	115	Trotting and cantering in circles and in straight lines.
20201129_ID_5_0011 ³	14400	115	Trotting in circles and in straight lines. Walking over the obstacle.
20201129_ID_5_0012 ³	14400	115	Walking. Jumping over an obstacle in trotting.

Table 4. File lists and motions in the files. ¹Later parts of all videos are missing. ²Later part of the video from Camera “23348” is broken. ³Sometimes markers are out of detection.

3D shape and pose modeling. The body model. The 3D shape and pose of the horse are modeled and represented through the parameters of the hSMAL body model. As a horse-specific version of SMAL³⁸, the hSMAL model⁴¹ defines a 3D horse mesh, consisting of 1,497 vertices, 2,990 faces, and 36 body segments. hSMAL can be described as a function $\xi(\beta, \theta, \gamma)$, where β is the shape parameter; θ is the 3D pose parameter; γ is the model translation. The model is learned from 37 horse toys using the procedure described in³⁸. More specifically, a purchased 3D mesh of a horse, created by an artist, is used to create a Global/Local Stitched Shape model (GLoSS)³⁸ for horses. The GLoSS model is fitted to each toy scan such that scans have the same mesh topology. To de-correlate body and tail shapes, tails among different toy horses are interchanged to generate a broader

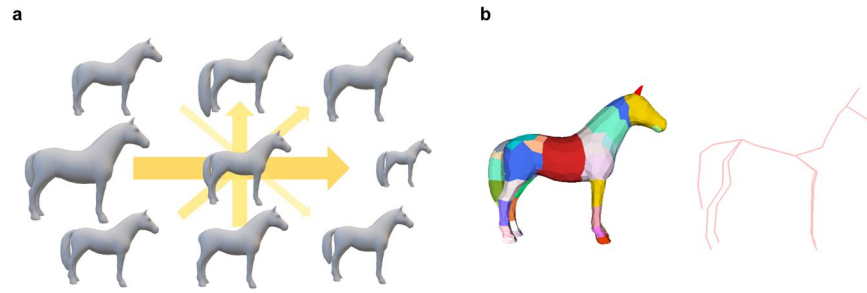


Fig. 7 The hSMAL model. **(a)** The hSMAL model and its skeleton. **(b)** The first four principal components in the hSMAL shape space. The arrow width shows from the first to the fourth component. All components are shown with $\pm 2std$.

range of data. After a process of pose-normalization, the mean template V_{mean} is computed by averaging the data. The vertex-based residuals between the data and the mean template are modeled by principle component analysis (PCA). β represents the coefficients of the learned low dimensional linear space, while B_s defines the shape deformations. More specifically, under the template pose, the shape is given as:

$$V = V_{mean} + B_s \beta^T. \quad (1)$$

The learned shape space of the model is shown in Fig. 7a, where the first three components mainly capture the model's sizes of the body, the tail, and the neck, respectively. θ represents the relative rotation of each joint with respect to its parent joint in the axis-angle representation according to the skeleton tree defined in the model. The θ parameter is a vector of dimension $3 \times 36 = 108$. The skeleton joint positions are manually defined (Fig. 7b) to better represent the animal anatomy similar to⁴⁵. The final mesh is then posed with Linear Blend Skinning (LBS)^{30,38} and shifted with translation parameter γ . More details are in the original papers^{38,41}. We used the first 10 PCA coefficients of the shape space as the shape parameters.

Model fitting. The parameters of a 3D articulated shape model can be estimated from mocap markers using the MoSh⁵² and MoSh++³¹ methods. These methods consider the fact that markers cannot be attached to fixed positions on the human body, especially when the human is moving or the markers are on soft tissues and solve not only for model parameters, but also for the marker position on the body surface. MoSh++, the updated version of MoSh, has been applied to different human mocap datasets to create a unified dataset, AMASS, which includes markers and aligned SMPL model parameters.

Following MoSh++³¹, we use two stages to capture the 3D shape and pose of the horse using the hSMAL model from mocap data. We use a similar notation as in MoSh++. Please check for more details in the original paper^{31,52}.

Stage I: Stage I focuses on estimating the shape and marker positions. A marker parameterization denoted as $m(\tilde{m}_i, \beta, \theta_i, \gamma_i)$ is utilized to estimate marker positions considering the body's shape, pose and location. More specifically, the latent markers \tilde{m}_i are mapped to the world by accounting for the model parameters ($\beta, \theta_i, \gamma_i$) at a particular frame t for marker i . To do this, F frames are randomly selected from subject-specific mocap sequences. The goal is to optimize the model parameters ($\beta, \Theta = \{\theta_{1:F}\}, \Gamma = \{\gamma_{1:F}\}$) and latent marker positions $\tilde{M} = \{\tilde{m}_i\}$, based on the observed marker positions $M = \{m_{i,t} \in M_i\}_{1:F}$. More specifically, an objective function is defined as:

$$E(\tilde{M}, \beta, \Theta, \Gamma) = \lambda_D E_D(\tilde{M}, \beta, \Theta, \Gamma) + \lambda_R E_R(\tilde{M}, \beta) + \lambda_I E_I(\tilde{M}, \beta) + \lambda_\beta E_\beta(\beta) + \lambda_\theta E_\theta(\theta) \quad (2)$$

where E_D measures the distance between the parameterized markers $m(\tilde{m}_i, \beta, \theta_i, \gamma_i)$ and the observed markers $m_{i,t}$, E_R ensures the markers are at an appropriate distance from the model surface (here we set 10 mm). E_I maintains the parameterized markers close to their initial positions. E_β and E_θ are regularizers related to shape and pose prior to the hSMAL model defined in⁴¹. Finally, a four-staged approach is performed to help avoid getting stuck in local optima. Here we randomly selected $F = 12$ frames from sequences where the horse was in more static poses, as it allowed for better optimization. It was worth noting that some markers may not be visible in these selected sequences. We used specific values for $\lambda_D = 105.0 \times d$, $\lambda_R = 10300.0$, $\lambda_I = 250.0$, $\lambda_\beta = 14.5$, $\lambda_\theta = 7.5$ and a scaling factor $d = 50/n$, to deal with varying numbers of markers, where 50 was the marker number of the skeletal model and n was the observed mocap marker number in a frame.

Stage II: Stage II focuses on optimizing the 3D poses from all subject-based sequences. The body shape β and latent marker positions of each subject in Stage I are determined and kept fixed during Stage II. More specifically, we minimize:

$$E(\theta, \gamma) = \lambda_D E_D(\theta, \gamma) + \lambda_\theta E_\theta(\theta) + \lambda_u E_u(\theta) \quad (3)$$

where E_D and E_θ are the same as in Stage I, measuring the alignment of the model with the observed data and maintaining specific constraints. E_u is a temporal smooth term, ensuring the 3D poses changed over time are

natural and smooth. Here we set $\lambda_D = 480.0 \times d$, $\lambda_\theta = 2.3 \times q$, $\lambda_u = 2.5$. A variable $q = 1 + \left(\frac{x}{|M|} * 2.5\right)$ was introduced based on the number of missing markers x . The more markers were missing, the higher the pose constrained in every given frame.

Hyper-parameter search. Certain hyperparameters λ during the two-stage optimizations were determined by line search on a synthetic dataset, as inspired by MoSh++. The synthetic dataset was first created using the toy shapes from the training data of the hSMAL model. We placed 38 synthetic markers on the model with toy shapes and animated the model. We divided this data into a training set, consisting of 32 toys and five animations and a validation set with five toys and two animations.

The searching process was adjusting one parameter while keeping other parameters fixed, both in Stage I and Stage II fitting, using different random seeds on training and validation datasets. In Stage I, we initialized the marker positions by randomly placing them near the true position. The goal was to find a better combination of $(\lambda_D, \lambda_R, \lambda_f, \lambda_\beta, \lambda_\theta)$ in Eq. 2. We aimed to find the values that provide the relatively lower distance between the estimated markers and the synthetic true markers within the selected 12 frames in both training and validation sets. In Stage II, the process was to find a better combination of $(\lambda_D, \lambda_\theta, \lambda_u)$ in Eq. 3. Here we aimed to minimize the error on all vertices of the model between the estimated results and the true results in both the training and validation set.

Data Records

The datasets are available at Harvard Dataverse⁵¹. The data is organized by subject folder with [Subject ID]. The name of files for each trial starts with [Record Date]_[Subject ID]_[Trial Number], indicated as [Trial Name]. Each subject folder stores six sub-folders with complete trials:

- C3D_DATA: One C3D file per trial, exported from the QTM software, referenced as [Trial Name].c3d.
- FBX_DATA: One FBX per trial, with the whole scenario including all information of cameras and 3D position of the subject, exported from the QTM software, referenced as [Trial Name].fbx.
- VIDEO_DATA: Each folder per trial named as [Trial Name], with videos from ten camera views, references as [Trial Name]_[Camera Code].avi.
- SEGMENT_DATA: Selected segmentation subsets for evaluating the fitting results, referenced as [Trial Name]_[Camera Code]_seg.mp4.
- MODEL_DATA: NPZ files, referenced as [Trial Name]_hsmal.npz includes the hSMAL parameters per trial. Another NPZ file, reference as [User ID]_stagei.npz contains the latent representation of the markers.
- KP2D_DATA: Each folder per trial named as [Trial Name], with 2D keypoints projected from 3D mocap data into video frames with ten camera views, referenced as [Trial Name]_[Camera Code]_2Dkp.npz.
- CAM_DATA: The camera parameters from ten camera views, referenced as Camera_Miquis_Video_[Cam ID].npz.

The correspondences are as follows:

[Record Date]: Recording date (e.g. 20201128 or 20201129)

[Subject ID]: Subject ID (e.g. ID_1, ID_2, ID_3, ID_4, ID_5)

[Trial Number]: Trial number (e.g. 0001, shown in Table 4)

[Camera Code]: Camera code (e.g. Miquis_65_20715)

[Camera ID]: Camera ID (e.g. 20715, 21386, 23348, 23350, 23414, 23415, 23416, 23417, 23603, 23604)

Technical Validation

In this section, we first provide calibration errors from QTM software and quantitative and qualitative results of the reconstructed 3D model.

Mocap data. The motion capture system was calibrated three times over the two days of data capture. The standard deviation of the wand length varies from 1.0–1.9 mm for the calibrations, with an average camera residual between 2.6–2.9 mm during calibration. For the different captures, the average camera residual varied between 1.5–3.1, with higher residuals when the horse moves close to the edge of the covered volume.

3D model evaluation. *Qualitative Results.* We show the optimized shape and marker positions in Fig. 8. The positions of the markers (in red) have been optimized to fit different shapes, starting from their initial guess positions (in blue) on the template shape. Additionally, some markers (in gray) represent markers not visible during Stage I, which could be either the exclusion of attaching certain markers or fail marker detection during the data recording process.

Examples of the captured 3D model are displayed in Fig. 9 with corresponding motions illustrated in Fig. 3. The 3D shape and pose representations, derived from mocap data, effectively capture the horse's real motions, even in challenging poses, like prancing and kicking. However, room for improvement remains in capturing more complex poses, such as sitting and lying down, especially when the limbs are in unusual positions.

Figure 10 provides results for the five horses. The images on the left side display the fitting results and the mocap frame, while the images on the right side provide a view of the reconstructed model as seen from ten different camera views. This visual comparison highlights the precision of the captured 3D shape and pose of the horses.

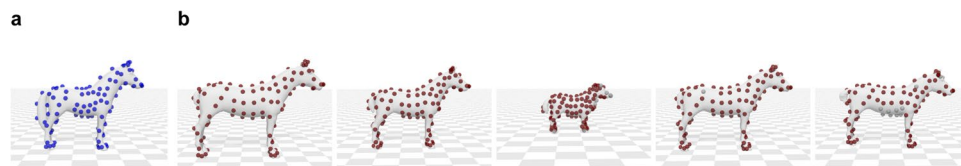


Fig. 8 Optimized shape and markers. (a) Initial guess of markers (in blue) on the template shape. (b) Optimized shape and optimized marker locations (in red) of five horses. Markers (in gray) are non-observable labels in observed frames in Stage I.

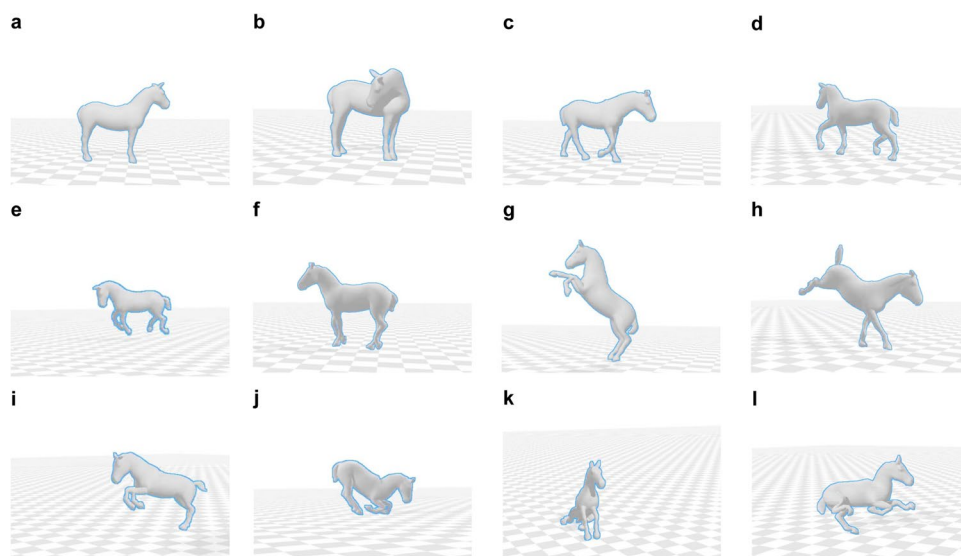


Fig. 9 Visualization of different motions. (a) Standing. (b) Neck bending. (c) Walking. (d) Trotting. (e) Cantering. (f) Piaffing. (g) Rearing. (h) Kicking. (i) Jumping. (j) Lying down. (k) Sitting. (l) Lying.

Shape Visualization. We provide the visualization of the estimated 3D horse shapes. Figure 12b shows the UMAP visualization of the components of the shape parameters from our five subjects, together with those of all the hSMAL toy training data. We can observe that Horse No. 3, the smallest pony, is quite distinct from the other four horses (Fig. 12a).

Quantitative Results. **3D Mocap Error.** To evaluate the accuracy of the model in capturing the shape and pose information from the mocap data, we analyze the Euclidean distance between the observed markers and the estimated virtual markers in each frame. As failed markers and noisy labels are inevitable, we focus on frames where more than 23 markers are detected. The results, detailed in Table 2, show results per horse across all trials and the average between different horse subjects is 0.031 meters.

2D Silhouette Error. We measure the accuracy of the captured 3D shape and pose by calculating the intersection over union (IOU) (Eq. 4) to measure the overlapping (Fig. 11c) between the extracted segmentation S (Fig. 11a) and the silhouette \tilde{S} (Fig. 11b), obtained by projecting the corresponding 3D model with the camera. We calculate the IOU of frames that are inside the selected silhouette subsets with more than 23 markers detected. Table 2 shows results over the selected silhouette subset per horse, with an average IOU of 0.85 across different horse subjects. Intersection over union is defined as:

$$IOU = \frac{|S \cap \tilde{S}|}{|S \cup \tilde{S}|} \quad (4)$$

Current limitations. Our current dataset has some limitations, as in Fig. 13, where the motion is not well captured. While the mocap system is precise, occasional irregularities may appear. For example, Fig. 13a shows a case where some of the markers are undetected, leading to misalignment of the model with the silhouette. Another limitation is the hSMAL model itself, which is learned from toy scans, and cannot perfectly represent real-world horses. As seen in Fig. 13b,c, certain body parts like the shape of the cheek and the back cannot be accurately captured, indicating the need for a more precise model.



Fig. 10 Visualization of example frames in different horses. Left: The 3D model and the corresponding mocap data (purple). Right: Project the 3D model in the same image frame in ten views. Left body (in yellow) and right body (in lightgray). (a) Horse No. 1. (b) Horse No. 2. (c) Horse No. 3. (d) Horse No. 4. (e) Horse No. 5.

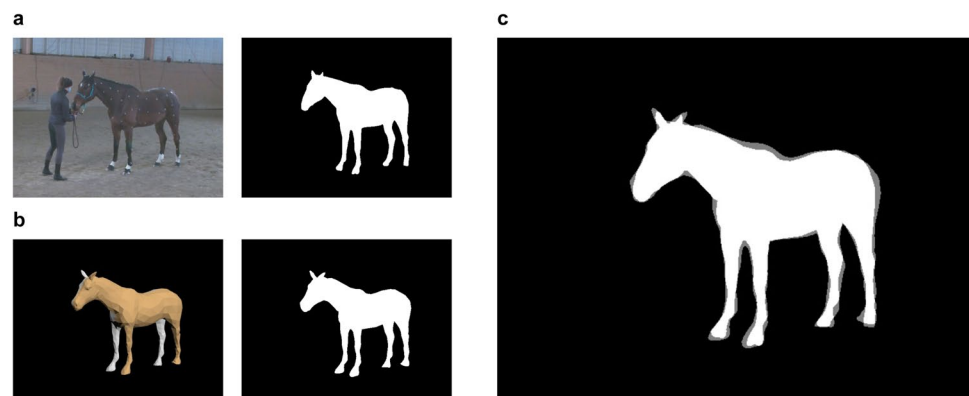


Fig. 11 3D model evaluation with intersection over union (IOU). (a) The video frame and the silhouette extracted from video. (b) The image and the silhouette obtained from projecting the model into the image plane. (c) The overlap of the two silhouettes.

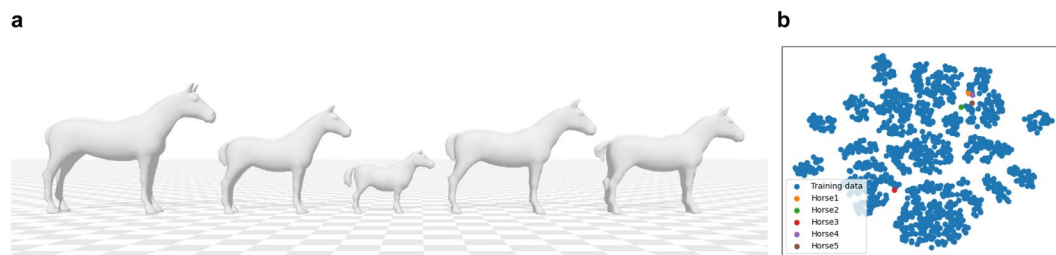


Fig. 12 Shape visualization of the five PFERD horses and UMAP analysis with the first two components of the shape parameters of these five horses and all the training toy data. **(a)** The shape of the PFERD horses as reconstructed in hSMAL. **(b)** UMAP visualization.

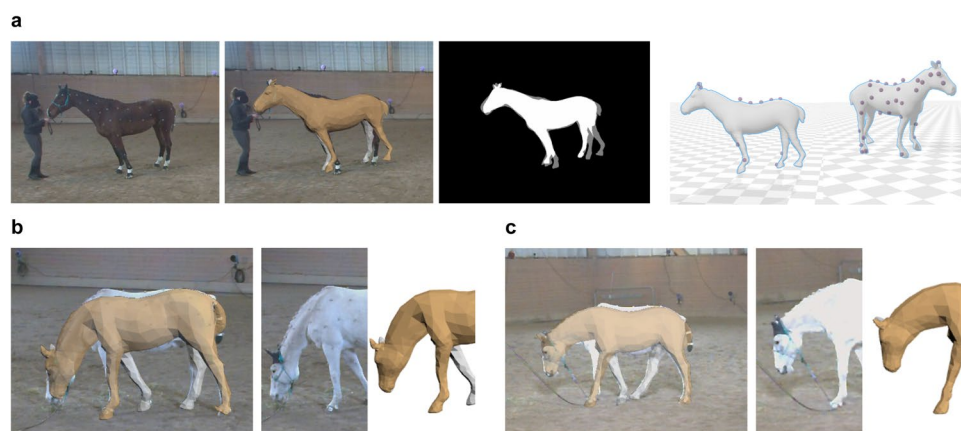


Fig. 13 Imperfect capture. **(a)** Missing mocap markers. From left to right: original video frame, the overlapping model with the image, overlapping silhouette, visualization of the model and mocap markers in two different views. **(b)** and **(c)** are two examples of model limitations in real-world representation. Example 1 **(b)**, where the model did not capture the shape of the neck and back well. Example 2 **(c)**, where the model did not capture the shape of the neck and the shoulder well.

Usage Notes

The provided FBX files can be imported into different animation applications, including Blender and any other application that supports loading fbx files. The provided c3d files can be processed using Python.

Researchers should be aware of certain anomalies in the data. Despite our meticulous data collection, a small number of irregularities may appear as we mentioned in the previous section, including missing markers, swapped marker labels, and noisy marker positions. This results in weird poses of the reconstructed model when too few correct markers are visible within a given frame. To mitigate this, we set a threshold requiring more than 23 visible markers for pose evaluation. However, these potential issues may influence the final results.

The silhouettes generated through deep learning methods are not fully infallible but can be considered as pseudo-ground truth. Even when we carefully select a subset manually comprised of relatively complete silhouettes, it's important to note that potential errors remain.

Additionally, discrepancies in response times among different camera sensors exist, despite all cameras being synchronized. In our current setup, we align the first frame of the mocap files with the first video frame of the corresponding videos, which may still result in minor inaccuracies. Users should be cognizant of this small margin of error during data analysis and interpretation.

Code availability

We make available functions for users to use our datasets:

- Loading c3d files and the hSMAL model with the captured parameters to visualize the mocap data and the fitted results.
- Projecting the reconstructed model in image planes with provided camera information.
- Quantitative evaluation using the mocap data and silhouette subsets. Further detail about environment settings and code usage can be found in <https://github.com/Celiali/PFERD.git>.

Received: 3 October 2023; Accepted: 25 April 2024;

Published online: 15 May 2024

References

- Louis, N. *et al.* Temporally guided articulated hand pose tracking in surgical videos. *International Journal of Computer Assisted Radiology and Surgery* **18**, 117–125, <https://doi.org/10.1007/s11548-022-02761-6> (2023).
- Zhang, J. Z. *et al.* Slomo: A general system for legged robot motion imitation from casual videos. *IEEE Robotics and Automation Letters* **8**, 7154–7161, <https://doi.org/10.1109/LRA.2023.3313937> (2023).
- Luo, H. *et al.* Artemis: Articulated neural pets with appearance and motion synthesis. *ACM Transactions on Graphics* **41**, <https://doi.org/10.1145/3528223.3530086> (2022).
- Khokhlova, M., Migniot, C., Morozov, A., Sushkova, O. & Dipanda, A. Normal and pathological gait classification lstm model. *Artificial Intelligence in Medicine* **94**, 54–66, <https://doi.org/10.1016/j.artmed.2018.12.007> (2019).
- Raab, S. *et al.* Single motion diffusion. In *The Twelfth International Conference on Learning Representations*, <https://doi.org/10.48550/arXiv.2302.05905> (2024).
- Guo, C. *et al.* Generating diverse and natural 3d human motions from text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5152–5161, <https://doi.org/10.1109/CVPR52688.2022.00509> (2022).
- Mir, A., Puig, X., Kanazawa, A. & Pons-Moll, G. Generating continual human motion in diverse 3d scenes. In *International Conference on 3D Vision*, <https://doi.org/10.48550/arXiv.2304.02061> (2024).
- Wilson, A. & Weller, R. The biomechanics of the equine limb and its effect on lameness. In *Diagnosis and Management of Lameness in the Horse*, 270–281, <https://doi.org/10.1016/B978-1-4160-6069-7.00026-2> (Elsevier, 2011).
- Makita, S., Murakami, N., Sakaguchi, M. & Furusho, J. Development of horse-type quadruped robot. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, vol. 6, 930–935, <https://doi.org/10.1109/ICSMC.1999.816677> (IEEE, 1999).
- Moro, F. L. *et al.* Horse-like walking, trotting, and galloping derived from kinematic motion primitives (kmps) and their application to walk/trot transitions in a compliant quadruped robot. *Biological cybernetics* **107**, 309–320, <https://doi.org/10.1007/s00422-013-0551-9> (2013).
- Hoyt, D. F. & Taylor, C. R. Gait and the energetics of locomotion in horses. *Nature* **292**, 239–240, <https://doi.org/10.1038/292239a0> (1981).
- Park, H. O., Dibazar, A. A. & Berger, T. W. Cadence analysis of temporal gait patterns for seismic discrimination between human and quadruped footsteps. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1749–1752, <https://doi.org/10.1109/ICASSP.2009.4959942> (IEEE, 2009).
- Buchner, H., Obermüller, S. & Scheidl, M. Body centre of mass movement in the lame horse. *Equine Veterinary Journal* **33**, 122–127, <https://doi.org/10.1111/j.2042-3306.2001.tb05374.x> (2001).
- Rhodin, M. *et al.* Vertical movement symmetry of the withers in horses with induced forelimb and hindlimb lameness at trot. *Equine veterinary journal* **50**, 818–824, <https://doi.org/10.1111/evj.12844> (2018).
- Ionescu, C., Papava, D., Olaru, V. & Sminchiescu, C. Human3.6 m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**, 1325–1339, <https://doi.org/10.1109/TPAMI.2013.248> (2013).
- Sigal, L., Balan, A. & Black, M. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision* **87**, 4–27, <https://doi.org/10.1007/s11263-009-0273-6> (2010).
- Mandery, C., Terlemeç, Ö., Do, M., Vahrenkamp, N. & Asfour, T. The kit whole-body human motion database. In *International Conference on Advanced Robotics*, 329–336, <https://doi.org/10.1109/ICAR.2015.7251476> (IEEE, 2015).
- Santos, G., Wanderley, M., Tavares, T. & Rocha, A. A multi-sensor human gait dataset captured through an optical system and inertial measurement units. *Scientific Data* **9**, 545, <https://doi.org/10.1038/s41597-022-01638-2> (2022).
- Cao, Z., Simon, T., Wei, S.-E. & Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7291–7299, <https://doi.org/10.1109/CVPR.2017.143> (2017).
- Mathis, A. *et al.* Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* <https://doi.org/10.1038/s41593-018-0209-y> (2018).
- Cao, J. *et al.* Cross-domain adaptation for animal pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9498–9507, <https://doi.org/10.1109/ICCV.2019.00959> (2019).
- Kocabas, M., Karagoz, S. & Akbas, E. Self-supervised learning of 3d human pose using multi-view geometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1077–1086, <https://doi.org/10.1109/CVPR.2019.00117> (2019).
- Li, X., Fan, Z., Liu, Y., Li, Y. & Dai, Q. 3d pose detection of closely interactive humans using multi-view cameras. *Sensors* **19**, 2831, <https://doi.org/10.3390/s19122831> (2019).
- Joska, D. *et al.* Acinonet: a 3d pose estimation dataset and baseline models for cheetahs in the wild. In *IEEE International Conference on Robotics and Automation*, 13901–13908, <https://doi.org/10.1109/ICRA48506.2021.9561338> (IEEE, 2021).
- Günel, S. *et al.* Deepfly3d, a deep learning-based approach for 3d limb and appendage tracking in tethered, adult drosophila. *Elife* **8**, e48571, <https://doi.org/10.7554/eLife.48571> (2019).
- Bala, P. C. *et al.* Automated markerless pose estimation in freely moving macaques with openmonkeystudio. *Nature communications* **11**, 4560, <https://doi.org/10.1038/s41467-020-18441-5> (2020).
- Patel, M., Gu, Y., Carstensen, L. C., Hasselmo, M. E. & Betke, M. Animal pose tracking: 3d multimodal dataset and token-based pose optimization. *International Journal of Computer Vision* **131**, 514–530, <https://doi.org/10.1007/s11263-022-01714-5> (2023).
- Zimmermann, C., Schneider, A., Alyahyay, M., Brox, T. & Diester, I. Freipose: a deep learning framework for precise animal motion capture in 3d spaces. *BioRxiv* 2020–02, <https://doi.org/10.1101/2020.02.27.967620> (2020).
- Gosztolai, A. *et al.* Liftpose3d, a deep learning-based approach for transforming two-dimensional to three-dimensional poses in laboratory animals. *Nature methods* **18**, 975–981, <https://doi.org/10.1038/s41592-021-01226-z> (2021).
- Loper, M., Mahmood, N., Romero, J., Pons-Moll, G. & Black, M. J. Smpl: A skinned multi-person linear model. *ACM Transactions on Graphics* **34**, 1–16, <https://doi.org/10.1145/2816795.2818013> (2015).
- Mahmood, N., Ghorbani, N., Troje, N. F., Pons-Moll, G. & Black, M. J. Amass: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5442–5451, <https://doi.org/10.1109/ICCV.2019.00554> (2019).
- Kocabas, M., Athanasiou, N. & Black, M. J. Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5253–5263, <https://doi.org/10.1109/CVPR42600.2020.00530> (2020).
- Ghorbani, N. & Black, M. J. Soma: Solving optical marker-based mocap automatically. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11097–11106, <https://doi.org/10.1109/ICCV48922.2021.01093> (2021).
- Yuan, Y., Iqbal, U., Molchanov, P., Kitani, K. & Kautz, J. Glamr: Global occlusion-aware human mesh recovery with dynamic cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11028–11039, <https://doi.org/10.1109/CVPR52688.2022.01076> (2022).
- Li, J. *et al.* Task-generic hierarchical human motion prior using vaes. In *International Conference on 3D Vision*, 771–781, <https://doi.org/10.1109/3DV53792.2021.00086> (IEEE, 2021).
- Rempe, D. *et al.* Humor: 3d human motion model for robust pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, <https://doi.org/10.1109/ICCV48922.2021.01129> (2021).
- Voleti, V. *et al.* Smpl-ik: Learned morphology-aware inverse kinematics for ai driven artistic workflows. In *SIGGRAPH Asia 2022 Technical Communications*, SA '22, <https://doi.org/10.1145/3550340.3564227> (Association for Computing Machinery, New York, NY, USA, 2022).

38. Zuffi, S., Kanazawa, A., Jacobs, D. W. & Black, M. J. 3d menagerie: Modeling the 3d shape and pose of animals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6365–6373, <https://doi.org/10.1109/CVPR.2017.586> (2017).
39. Biggs, B., Boyne, O., Charles, J., Fitzgibbon, A. & Cipolla, R. Who left the dogs out? 3D animal reconstruction with expectation maximization in the loop. In *European Conference on Computer Vision*, https://doi.org/10.1007/978-3-030-58621-8_12 (2020).
40. Rüegg, N., Zuffi, S., Schindler, K. & Black, M. J. Barc: Learning to regress 3d dog shape from images by exploiting breed information. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3876–3884, <https://doi.org/10.1109/CVPR52688.2022.00385> (2022).
41. Li, C. *et al.* hsmal: Detailed horse shape and pose reconstruction for motion pattern recognition. In *First CV4Animals Workshop, IEEE/CVF Conference on Computer Vision and Pattern Recognition*, <https://doi.org/10.48550/arXiv.2106.10102> (2021).
42. Luo, H. *et al.* Artemis: articulated neural pets with appearance and motion synthesis. *ACM Transactions on Graphics* **41**, 1–19, <https://doi.org/10.1145/3528223.3530086> (2022).
43. Dunn, T. W. *et al.* Geometric deep learning enables 3d kinematic profiling across species and environments. *Nature methods* **18**, 564–573, <https://doi.org/10.1038/s41592-021-01106-6> (2021).
44. Zhang, H., Starke, S., Komura, T. & Saito, J. Mode-adaptive neural networks for quadruped motion control. *ACM Transactions on Graphics* **37**, 1–11, <https://doi.org/10.1145/3197517.3201366> (2018).
45. Kearney, S., Li, W., Parsons, M., Kim, K. I. & Cosker, D. Rgb-dog: Predicting canine pose from rgb-d sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8336–8345, <https://doi.org/10.1109/CVPR42600.2020.00836> (2020).
46. Warner, S., Koch, T. & Pfau, T. Inertial sensors for assessment of back movement in horses during locomotion over ground. *Equine Veterinary Journal* **42**, 417–424, <https://doi.org/10.1111/j.2042-3306.2010.00200.x> (2010).
47. Bragança, F. S., Rhodin, M. & Van Weeren, P. On the brink of daily clinical application of objective gait analysis: What evidence do we have so far from studies using an induced lameness model. *The Veterinary Journal* **234**, 11–23, <https://doi.org/10.1016/j.tvjl.2018.01.006> (2018).
48. Unt, V., Evans, J., Reed, S., Pfau, T. & Weller, R. Variation in frontal plane joint angles in horses. *Equine Veterinary Journal* **42**, 444–450, <https://doi.org/10.1111/j.2042-3306.2010.00192.x> (2010).
49. Bosch, S. *et al.* Equimoves: A wireless networked inertial measurement system for objective examination of horse gait. *Sensors* **18**, 850, <https://doi.org/10.3390/s18030850> (2018).
50. Ericson, C., Stenfeldt, P., Hardeman, A. & Jacobson, I. The effect of kinesiotape on flexion-extension of the thoracolumbar back in horses at trot. *Animals* **10**, 301, <https://doi.org/10.3390/ani10020301> (2020).
51. Li, C. *et al.* The Poses for Equine Research Dataset (PFERD). *Harvard Dataverse* <https://doi.org/10.7910/DVN/2EXONE> (2024).
52. Loper, M., Mahmood, N. & Black, M. J. Mosh: motion and shape capture from sparse markers. *ACM Transactions on Graphics* **33**, 220–1, <https://doi.org/10.1145/2661229.2661273> (2014).
53. Yang, J. *et al.* Track anything: Segment anything meets videos. *arXiv:2304.11968* <https://doi.org/10.48550/arXiv.2304.11968> (2023).
54. Kirillov, A. *et al.* Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4015–4026, <https://doi.org/10.1109/ICCV51070.2023.00371> (2023).
55. Sisson, S. A *Textbook of Veterinary Anatomy* (Saunders, 1911).

Acknowledgements

The project was supported by a career grant from SLU No. 57122058, received by Elin Hernlund. Silvia Zuffi was in part supported by the European Commission's NextGeneration EU Programme, PNRR grant PE0000013 Future Artificial Intelligence Research–FAIR CUP B53C22003630006. The authors sincerely thank Tove Kjellmark for her assistance during the data collection and all horses and owners for their collaboration in the data collection experiments. Thank you to Zala Zgank for helping with the labeling of motion capture data.

Author contributions

C.L. processed the data, conducted the computer science experiments and wrote the manuscript. Y.M. and J.K. labeled and processed the motion capture data. E.H., S.P., and M.H. planned, designed and contributed the resources for the data collection and performed the data collection together with J.K.. N.G. provided the framework for supporting the 3D modeling experiments. M.J.B. provided helpful insight, discussion and resources in the project. H.K., S.Z. and E.H. supervised the project, revised the manuscripts, and E.H. provided overall project management. All authors reviewed the manuscript.

Funding

Open access funding provided by Swedish University of Agricultural Sciences.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.H.

Reprints and permissions information is available at 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 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 Author(s) 2024