

<https://doi.org/10.1038/s42003-024-07065-w>

A dominant missense variant within *LMBR1* related to equine polydactyly



Yue Luan^{1,3}, Ling Zhong^{1,3}, Cao Li¹, Xiaoyu Yue¹, Mengyan Ye¹, Jianpeng Wang¹, Yiping Zhu² & Qin Wang¹

Polydactyly was recorded before 100 BCE and attracted widespread interest because of its relationship to limb health and ancestral traits in horses. However, the underlying reasons for the development of polydactyly remain unclear. To search for polydactyly-related genes, we utilize a paternal half-sib family and screen for variants that match the mode of inheritance. Through this screening process, 77 variants in 65 genes are filtered. A missense variant (EqCab3.0 chr4: <107353368> A > G) (rs1138485164) in the 3rd exon of *LMBR1* is identified as a source of amino acid sequence variation. Gene editing confirms that the variant down-regulates *LMBR1* expression, increases the proliferative viability of mutant cells, and inhibits apoptosis. This study suggests that *LMBR1* might play a role in the development of polydactyly and that the variant detected in this study is related to polydactyly in horses. However, further research is needed to determine whether a direct relationship exists.

Equine polydactyly has received much attention and has a long history of research. The earliest description of a polydactyly horse dates back to 100 BC, when Julius Caesar's mount was considered a polydactyly¹. In 1663, an eight-digit horse was exhibited in Germany², and later described by George Simon in his book "Horses", published in 1703. In 1827, Geoffroy recorded a case of bilateral polydactyly in the forelimbs of a horse, with three equally sized malformed digits on the left forelimb and double digits deformity on the right forelimb². In China, a photographic record of a polydactyl horse was kept at the Beijing Zoo in the late Qing Dynasty. The great interest in the polydactyl horse has led to its appearance on postcards such as a polydactyl Shire horse. In addition, previous studies suggest that equine polydactyly may be a reversion to an ancestral state³. Therefore, the study of polydactyly is beneficial in analyzing limb and digit development in the horse and in understanding the evolution of the horse.

Polydactyly is an inherited limb malformation that manifests as the presence of extra digits. It is found in many species, including humans, dogs, cats, chickens, guinea pigs, cows, kangaroos, pigs, mice, and toads⁴⁻¹⁵, and has been shown in genetic studies to conform to different modes of inheritance in other species¹⁶⁻²³. In addition, polydactyly is epigenetically regulated in mice²⁴. This indicates the complexity of genetic mechanisms and the difficulty of conducting research for polydactyly. It was found that in all species with a reduced number of digits, the reduction started with five or four digits, suggesting a convergence in the evolution of digit development²⁵. In vertebrates, the signaling pathways involved in limb development are

highly conserved, and three major signaling pathways control limb development, namely the Shh (Sonic hedgehog) signaling pathway, the FGFs (fibroblast growth factor) signaling pathway, and the Wnt (Wnt/ β -catenin) signaling pathway²⁶⁻³⁰. FGFs promote *Shh* expression in the hindlimb bud, and several studies have shown that changes in the expression of FGFs lead to the formation of polydactyly³¹; the Shh signaling pathway is essential for regulating digit development during embryogenesis, and loss of *Shh* expression in mice results in the disappearance of the remaining digit and the formation of only the first digit at the palmar end, which is not Shh-dependent for development³². In the Shh pathway, *Shh* regulates digit development together with genes such as *LMBR1*³³, *Smo*³⁴, *Ptch1*³⁵, *Gli1*^{36,37}, *Gli2*³⁸, and *Gli3*³⁹ through *cis*-regulation or binding. Shh is also a key gene responsible for polydactyly in chicken, human, and mouse⁴⁰⁻⁴². In addition, *IQCE*, *Wnt7a*, *ZNF141*, *FAM92A*, *KIAA0825*, and *DACH1* are associated with the development of polydactyly⁴³⁻⁴⁸.

In most studies of equine polydactyly, researchers have analyzed the external appearance morphology, skeleton, soft tissue, and mobility of polydactyly horses, but there is still a gap in the study of its causal genes. In 1986, Stanek et al. found that both a mare and her foal had polydactyly in the medial forelimbs, and this case supports the fact that polydactyly in horses can also be inherited⁴⁹. However, the mode of inheritance cannot be determined by individual cases alone and relies on the mapping of a population. The discovery of equine polydactyly cases is scattered and not conducive to summary observations. Because horses are singleton breeders

¹State Key Laboratory of Animal Biotech Breeding, National Engineering Laboratory for Animal Breeding, Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, China Agricultural University, Beijing, China. ²College of Veterinary Medicine, China Agricultural University, Beijing, China. ³These authors contributed equally: Yue Luan, Ling Zhong. e-mail: wangqin@cau.edu.cn

with long intergenerational intervals, it is difficult to conduct genetic studies by observing a sufficient number of family phenotypes, causing the study of equine polydactyly to lag behind other common species. Most researchers believe that the mode of inheritance of equine polydactyly is autosomal dominant rather than recessive⁵⁰. However, since there is more than one mode of polydactyly inheritance in other species, it is not certain that equine polydactyly has only one mode of inheritance and the elaborate mode of inheritance and the determinant genes remain obscure.

Therefore, in order to solve the difficulties in the genetic study of equine polydactyly, we were fortunate to obtain a resource of polydactylous stallions, on which we constructed a half-sib family to analyze and validate the inheritance mode of equine polydactyly and screened for candidate genes that might affect polydactyly. Our findings pointed to a missense variant in *LMBRI* being related to the occurrence of polydactyly in horses.

Results

Phenotype observation of horse polydactyly

Horses typically have only one digit, but the polydactyly phenotype in horses is characterized by an extra digit on one or both feet. In a trial Yili horse population, a single individual with polydactyly was found, with additional digits on all four limbs. The additional digits were perpendicular to the ground, as shown in Fig. 1a. The radiographs revealed that the extra digits were positioned medially to the middle digit and were fully formed, with well-developed second metacarpal and phalangeal bones (Fig. 1b). From a mobility standpoint, polydactyly does not affect the horse's mobility, allowing it to walk and canter on various surfaces such as grass, concrete, and dirt. The additional digit aids in maintaining balance while walking on dirt. In this case, the foal exhibited polydactyly from birth, indicating the heritability of polydactyly in Fig. S1a. However, the mode of inheritance has not been fully confirmed, and the determinant genes have not yet been successfully identified.

Establishing candidate genes screening methods based on a half-sib family with seven individuals

A half-sib family of seven individuals was applied to polydactyly candidate gene detection. The family was formed by crossing one male polydactyl Yili horse with three non-Yili mares with normal digit phenotype and the F₁ generation contained three horses. The male parent was the eight-digit mutant individual found in the Yili horse with additional digits in all four limbs, as shown in Fig. 2. The three mares had a normal phenotype (none were Yili horses) and were not related to the polydactyl stallions, originating from a herd without polydactyls. Among the offspring, two males showed polydactyly and one mare showed normal hooves, as shown in Fig. 2a and Supplementary Fig. 1.

Seven individuals were identified as related using the KING-robust kinship estimator. When the KINSHIP value was 0.25, the two individuals were considered to be related or full siblings; when the KINSHIP value was

0.125, the two individuals were considered to be half siblings⁵¹ (Table 1). The parents are not related to each other (see data in Supplementary Table 1). The half-sib family for screening polydactyly-related genes was formed with four parents and three offspring for subsequent studies.

The overall framework of the polydactyly candidate genes screening approach is as follows: first, variant calling was conducted on whole-genome resequencing data obtained from the half-sib family and compared with the horse reference genome to obtain variants. The polydactyly-related variants were screened under three conditions, including (i) the genotype of the variant sites conforms to the inheritance mode, (ii) the genotype of the variant sites is consistent with larger population validation, (iii) functional annotation of the genes in which the variants reside are associated with polydactyly. The mode of inheritance was first predicted based on the phenotype and then screened for the variants that confirmed the mode of inheritance. The screening was carried out next for variants that were closely related to polydactyly and had different genotypes between polydactyl and normal-type horses by expanding the population size. As a result, a gene list was then obtained according to the screened variants and annotated. Finally, polydactyly-related genes were screened based on gene function annotation (Fig. 2b and Supplementary Data 2).

Polydactyly-related gene detection

Whole-genome high-throughput sequencing (30×) was performed on seven individuals. Through variant calling, about seven million single-nucleotide polymorphisms (SNPs) and 700,000 InDels per individual were identified. Polydactyly can be passed on to offspring, but its mode of inheritance has not been thoroughly confirmed. Phenotypic observation suggests four modes of inheritance, including autosomal dominant, autosomal recessive, X-linked recessive, and Y-linked inheritance (Table 2). Genetic markers were utilized to investigate the mode of inheritance and the genes causing the observed phenotype in this population of horses. A total of 25,921 variants conformed to the four predicted inheritance modes (Fig. 2b and Supplementary Data 1). Among these variants, 13,177 conformed to dominant inheritance, 12,126 to autosomal recessive and X-linked recessive inheritance, and 376 to Y-linked inheritance. Of these, 12,120 variants were genotypically probable for autosomal recessive inheritance and were located on 1583 genes, with 19 variants with missense mutations located in the exon regions of 19 genes. Six SNPs were probable for X-linked recessive inheritance and were located on four genes. Additionally, 13,177 SNPs were probable for autosomal dominant inheritance and were located on 1756 genes, with 1137 non-synonymous mutations located in exons on 264 genes (Supplementary Data 1).

Larger populations containing 200 individuals and 12 breeds were used to remove false-positive loci caused by differences in genetic background. A total of 76 SNPs located on 65 genes were identified, of which 52 have functional annotations (Supplementary Data 2). We found that of the 52 genes, *ENSECAG00000024396*, encoding limb development membrane

Fig. 1 | Phenotype observation for polydactyly horse. a Hoof phenotypic differences between normal and polydactyly individuals. **b** X-ray of four hooves of a polydactyly horse.

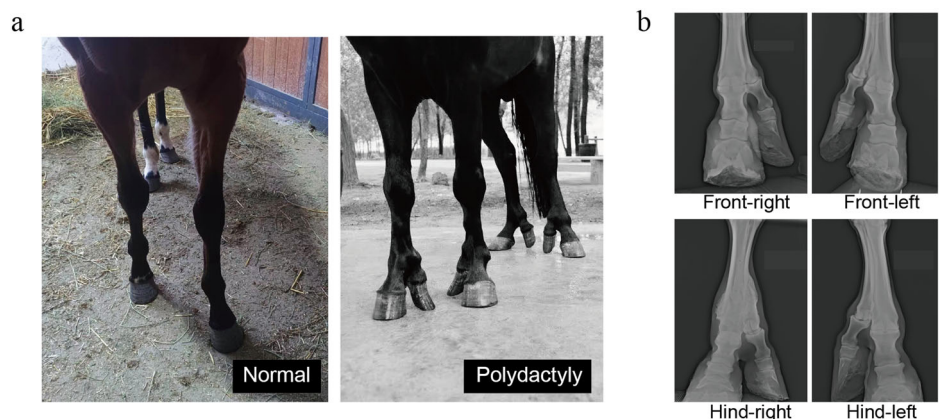


Fig. 2 | Screening protocol for polydactyly candidate genes based on half-sib family. **a** Three offspring were obtained by crossing an eight-digit stallion on the leftmost side of the figure with three mares in the center of the figure. Two colts were mutants. **b** The variant sites obtained from resequencing were screened in one round according to the inheritance model, and the variants obtained from the screening were then intersected with genes related to polydactyly with high homology in other animals to finally obtain equine polydactyly candidate genes. Diagram of gene screening, with gray circles representing the number of genes that conformed to the inheritance modes, and the lake blue color indicates the number of variants finally screened.

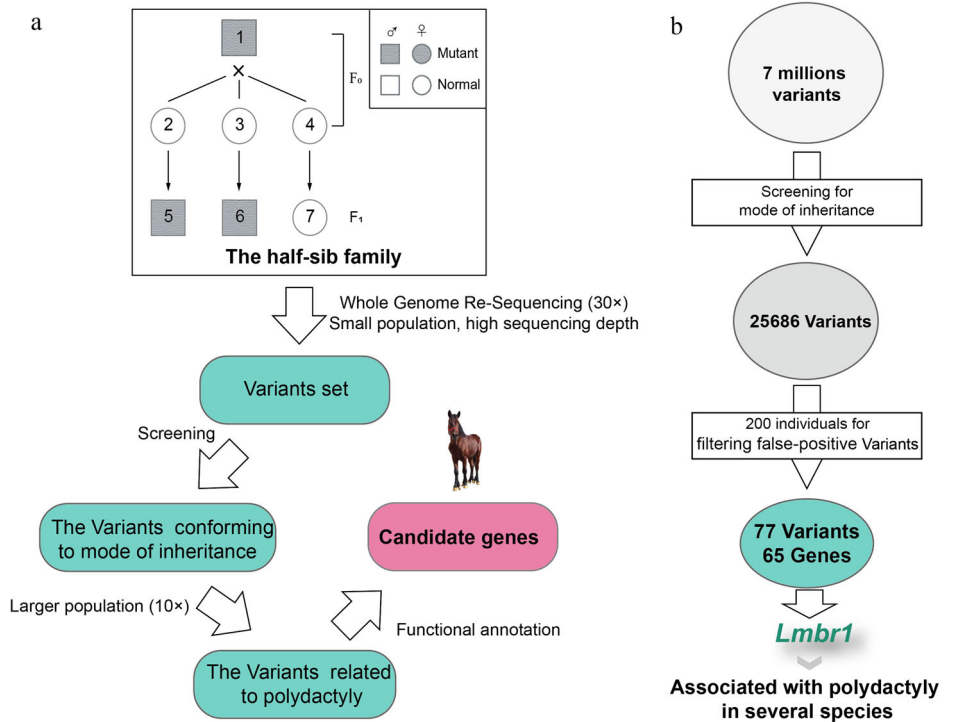


Table 1 | Half-sib family kinship testing

Name	Name	NSNP	HETHET	IBS0	KINSHIP	Relationship
1*	5*	31848318	0.0689483	0.000311225	0.255473	Parent-child
1*	6*	31850737	0.0687427	0.000274028	0.253515	Parent-child
1*	7	31848477	0.0673241	0.000300831	0.253424	Parent-child
2	6*	31848887	0.0709802	0.000250998	0.260332	Parent-child
3	5	31839442	0.068641	0.000372934	0.253747	Parent-child
4	7	31836999	0.0673009	0.000349499	0.252909	Parent-child
6*	5*	31856862	0.0586251	0.0077475	0.159752	Half-sibling
7	6*	31855063	0.0577571	0.00801803	0.152962	Half-sibling
7	5*	31853600	0.0575341	0.00840247	0.150649	Half-sibling

* is a polydactyly individual.

Table 2 | Individual genotype prediction

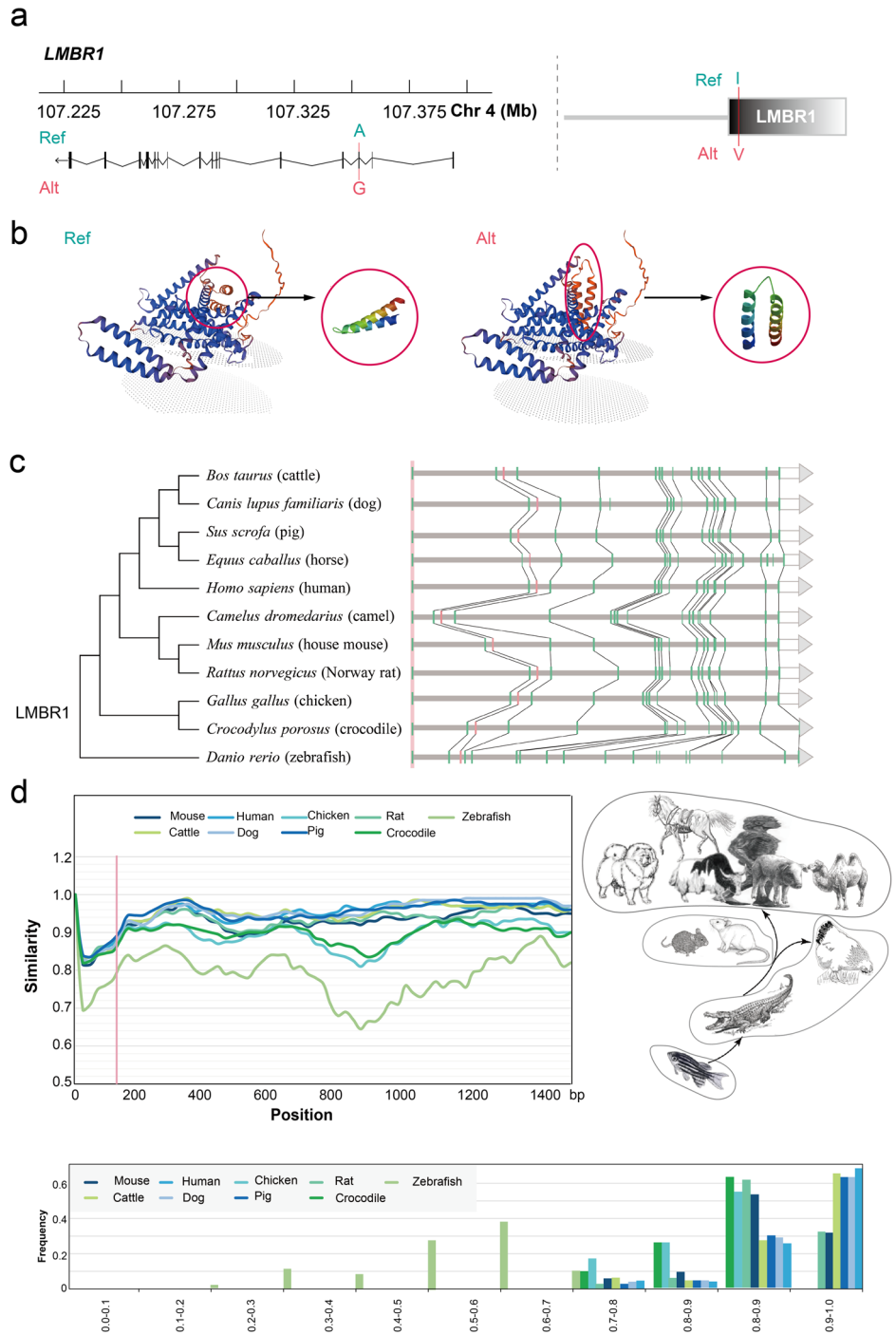
Name	Phenotype	Autosomal dominant	Autosomal recessive	X-linked recessive	Y-linked
1	Mutant	Aa	aa	X _a Y	X _{NA} Y
2	Normal	aa	Aa	X _a X _A	X _{NA} X _{NA}
3	Normal	aa	Aa	X _a X _A	X _{NA} X _{NA}
4	Normal	aa	Aa/AA	X _a X _A /X _A X _A	X _{NA} X _{NA}
5	Mutant	Aa	aa	X _a Y	X _{NA} Y
6	Mutant	Aa	aa	X _a Y	X _{NA} Y
7	Normal	aa	Aa	X _a X _A	X _{NA} X _{NA}

protein 1 (*LMBR1*), is associated with limb development and has been reported to be closely related to the polydactyly in mouse⁵¹, humans⁵², and chickens⁵³. None of the other 64 genes were involved in the Shh signaling pathway, the FGFs signaling pathway, or the Wnt signaling pathway (Supplementary Data 2). Therefore, we selected *LMBR1* as a candidate gene of equine polydactyly for subsequent functional validation (Fig. 2b).

The gene *LMBR1*, which encodes the limb development membrane protein, is located on horse chromosome 4 (Chr4) and is 156830 bp in length, containing 18 exons with an amino acid sequence of 436 aa. It encodes a member of the LMBR family of membrane proteins. Members of this protein family are lipocalin transmembrane receptors. The variant site (EqCab3.0 chr4: <107353368> A > G) (rs1138485164) in the third exon has been found in the study, at amino acid 55 of the protein sequence, where the mutation changes the amino acid sequence from isoleucine (I) to valine (V), which is located at the *LMBR1* functional domain (Fig. 3a). Analysis of the protein structure showed that mutation of this site minor affects the three-dimensional structure of the protein (Fig. 3b). Betts and Russell indicated that the substitution of isoleucine with valine may not significantly impact protein structure⁵². However, whether this minor alteration influences the function of the binding site remains unclear. Furthermore, our analysis of the 10 Kb regions upstream and downstream of the mutation site revealed no additional variations consistent with the genetic model. Consequently, we focused our further analysis on this specific mutation.

A comparative genomic analysis was conducted on the *LMBR1* locus. The covariance analysis showed that the gene structure of *LMBR1* is highly conserved across multiple species (Fig. 3c). The gene exhibits a remarkably

Fig. 3 | Sequence analysis of *LMBR1*. **a** Sequence variation analysis of *LMBR1*. The structure of the gene with the black boxes for exons and lines between the boxes for introns. Ref represents the reference genotype of the variant site, and Alt represents the mutant genotype. Structural domain variants of *LMBR1*, where the standard position of the red line is the amino acid variation site; **b** 3D structure comparison between mutant and wild type protein; **c** multi-species covariance analysis. On the left is the phylogenetic tree of *LMBR1*, and on the right is the covariance analysis of the gene. Green and pink boxes represent exons, black lines connect exons that are homozygous between species, and pink boxes represent exons that are mutated. **d** Multi-species sequence similarity analysis for *LMBR1*. The upper part is a sequence similarity analysis and a demonstration of the degree of sequence similarity between species. The lower part is the similarity frequency of the sequence similarity of each site for each species compared to the horse.

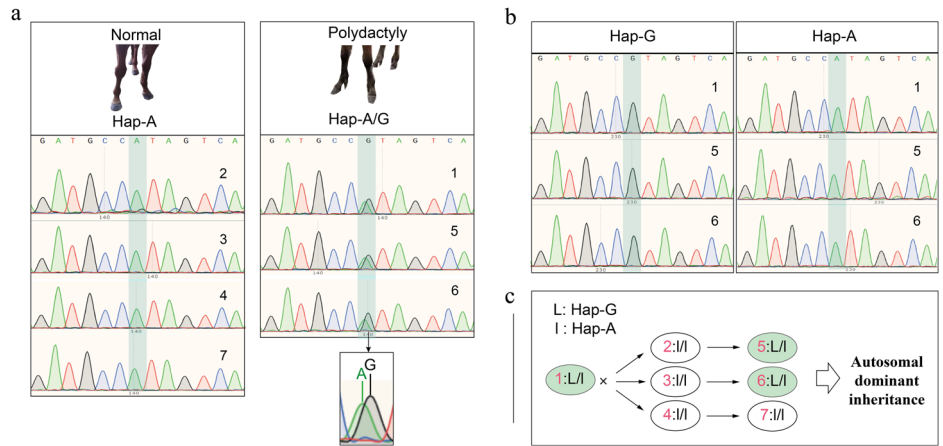


high amino acid sequence similarity of 88.2% among human, cattle, chicken, mouse, rat, pig, dog, camel, crocodile, and horse (Fig. 3c). Furthermore, we observed that the gene sequence was more similar in larger mammals, up to 90%. The gene’s degree of similarity with small mammals, such as mice was 85%. The similarity in birds and reptiles was 78%. Sequence similarity with fish, such as zebrafish, was the lowest, at 58% (Fig. 3d, see data in Supplementary Table 2). These results implied that *LMBR1* was subject to selection during evolution and was more strongly selected when reptiles evolved into mammals. The mutant locus exhibits a high degree of similarity among mammals, birds, and reptiles (Fig. 3d), while differences exist with fish, suggesting that the locus may be a key locus for limb development.

The variant site validation and analysis

We used Sanger sequencing to verify the correctness of the variant in *LMBR1*, and it was found that *LMBR1* was heterozygous in the missense variant and homozygous in the normal individuals. In mutant individuals, the sequencing peak map of this gene showed that the variant was sequenced with hetero-peaks, including two peaks for A and G bases and single peaks for all non-variant sites. In the normal individuals, the variant was shown as a single peak at the A base (Fig. 4a). At the missense variant, we genotyped seven individuals from the half-sib family. The genotypes of three mutant horses were all heterozygous (L/l) (Fig. 4b). Based on the results of the *LMBR1* variant analysis in this study, the genotype of this locus was found to conform to autosomal dominant inheritance. When the allele at the locus

Fig. 4 | The variant validation. **a** The variant in polydactyly individuals and normal individuals, with details of heterozygous peaks in small squares. **b** Heterozygosity validation of polydactyly individuals, including two different monoclonal clones. **c** The genotype of *LMBR1* in the semi-sibling family confirms the autosomal dominant inheritance. The green circles refer to polydactyly individuals and the white circles refer to normal individuals.



was the dominant heterozygous genotype *LMBR1*^{L/I}, the horse exhibited polydactyly and when it was the homozygous recessive genotype *LMBR1*^{I/I}, the horses had normal digit numbers (Fig. 4c).

To verify that this missense variant of *LMBR1* was unique in polydactyly horses, sequence analysis against this variant was performed in a sequencing population with a total of 100 normal-type individuals and consisted of 14 breeds including the Yili horse. Multi-species sequencing populations with widely varying genetic backgrounds were constructed, the utilization of which could avoid false-positive results at this locus due to breed specificity. The sequencing results showed that the 100 normal-type individuals were genotypically identical to *LMBR1*^{I/I} in both introduced and local varieties, rather than heterozygote *LMBR1*^{L/I} in the polydactyly horses in Table 3 and Supplementary Fig. 2. The above results indicated that the missense variant for *LMBR1* identified in this study was unique among individuals with polydactyly and was not affected by differences in genetic background across breeds. This result indicated that the variant was highly correlated with polydactyly.

Functional validation of the *LMBR1* variant

Site-directed mutation vectors were constructed based on the location of the discovered variable site of *LMBR1*. Primary bone marrow mesenchymal stem cells (MSCs) were isolated using horse bone marrow tissues and purified. The site-directed mutation vector and control vector were

transfected into horse bone marrow MSCs using Lipo 3000 transfection reagent, and the *LMBR1* gene was edited using CRISPR/Cas9 and homologous recombination (Fig. 5a and Supplementary Fig. 3a–c). No significant changes in cell morphology were observed in the mutant group compared with the control group after transfection, and green fluorescence was seen uniformly distributed in the stem cells at 24 h (Fig. 5b). Cells were sorted using flow cytometry, and those with green fluorescence were collected and continued to be cultured (Supplementary Fig. 3d). Sequence verification results from first-generation sequencing showed that the base of the target site was successfully replaced from A to G (Fig. 5a). The expression level of *LMBR1* was significantly downregulated ($p = 0.0001$, *t*-test, one-tailed, two-sample equal variance hypothesis) after the site-directed mutagenesis (A to G) (Fig. 5c, see data in Supplementary Table 3).

Research has demonstrated that digit development in horses is regulated by apoptosis and proliferation⁵⁴. Therefore, we analyzed the impact of the variable site on cell proliferation and apoptosis. To determine whether the missense variant of *LMBR1* affects apoptosis, *GAPDH* was used as the reference gene to detect the expression of *LMBR1* in this study. The results showed that the expression of the apoptosis-promoting gene *Casp3* was significantly decreased after *LMBR1* site mutation compared with the control group ($p = 0.0036$, *t*-test, one-tailed, two-sample equal variance hypothesis), among which the expression of *Casp7* was extremely significantly decreased ($p = 5.82e - 05$, *t*-test, one-tailed, two-sample equal variance hypothesis). The expression of the apoptosis-inhibiting gene *Bcl-2* was significantly increased compared to the control group ($p = 2.65e - 05$, *t*-test, one-tailed, two-sample equal variance hypothesis) (Fig. 5d, see data in Supplementary Table 3). To analyze whether the migration ability of cells changed after the site mutation, the cell scratch assay analysis was performed. Based on the results of the cell scratch assay, after 72 h, cells in the *LMBR1* site mutation group filled the wound scratch area faster than cells in the control group (Fig. 5e). To explore the effect of variant on cell proliferation, the cells at 0, 12, 24, and 36 h after transfection were assayed for proliferation viability, and it was found that the cell viability increased at 12 h after transfection. The proliferation viability of the control and mutation groups differed significantly at 24 h (Fig. 5f, see data in Supplementary Table 4). The above results indicated that the proliferation and migration ability of cells were enhanced after *LMBR1* single-site mutation, while cell apoptosis was inhibited.

The Shh pathway plays a crucial role in the development of digits across multiple species, and its protein concentration gradient correlates with digit development⁵⁵. In a previous study, a higher concentration of Shh protein binds to Ptch1, leading to the derepression of Smo, inhibition of full-length Gli3 processing into Gli3R, and an increase in full-length Gli3A protein, thereby activating genes such as *Ptch1* and *Hhip*⁵⁶. *LMBR1* in this pathway affects the Shh pathway exercise function by regulating the expression level of *Shh*. In this study, whether mutation of exons can affect the expression level of Shh is the key to affecting the pathway, so we examined the changes

Table 3 | Genotypes of 100 normal phenotype individuals from 14 breeds

Phenotype	Breeds	Number	Genotype
Polydactyly			A/G
Normal	Thoroughbred horse	11	A/A
Normal	Arabian horse	11	A/A
Normal	Akhal-Teke horse	3	A/A
Normal	Shetland Pony	4	A/A
Normal	American Saddlebred	8	A/A
Normal	American Quarter	21	A/A
Normal	Kazakh horse	2	A/A
Normal	Chakouyi horse	1	A/A
Normal	Dali horse	9	A/A
Normal	Lichuan horse	3	A/A
Normal	Ningqiang horse	3	A/A
Normal	Wenshan horse	9	A/A
Normal	Wumeng horse	5	A/A
Normal	Yili horse	10	A/A

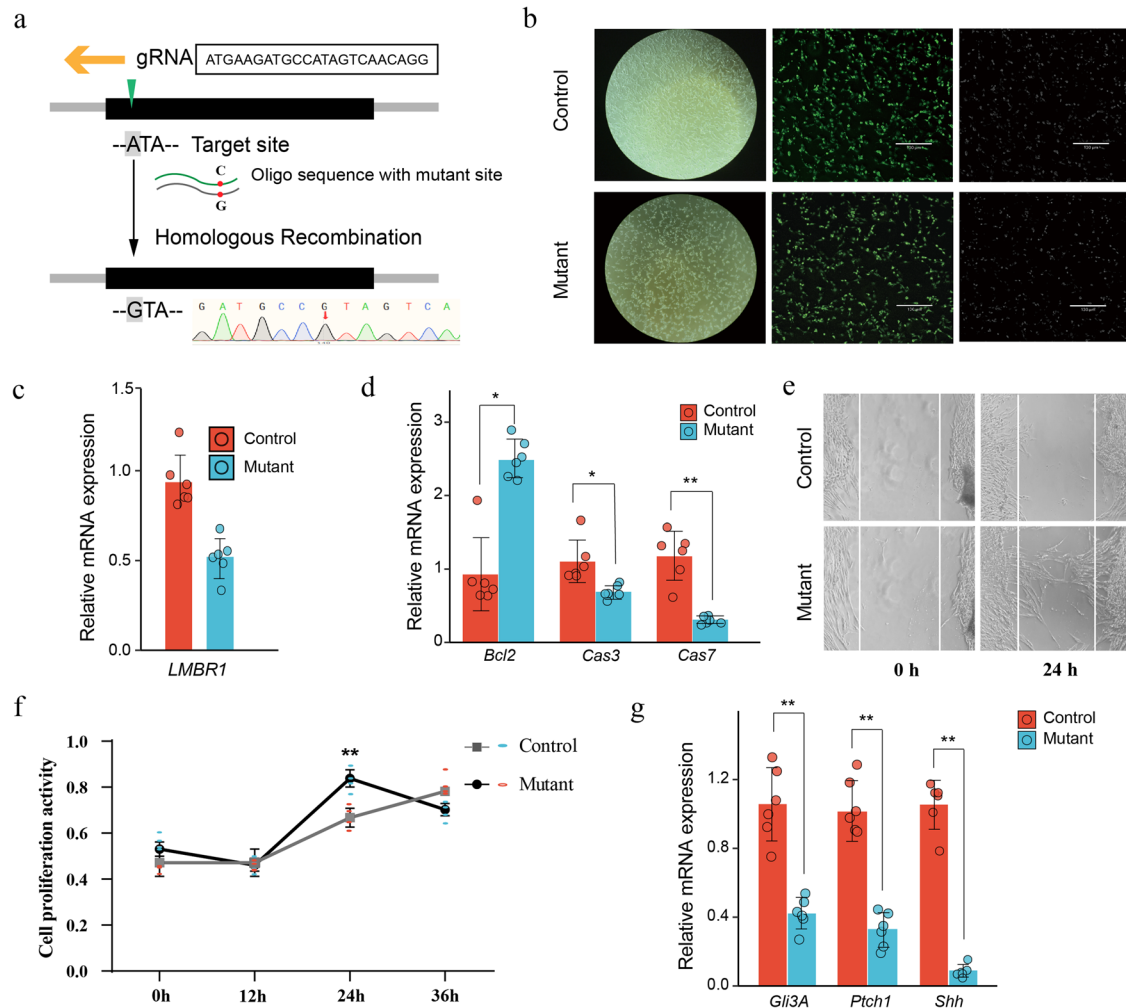


Fig. 5 | Cytological function verification of *LMBR1*. **a** Construction and PCR validation of the *LMBR1* site mutation cells. The gRNA is designed at the PAM sequence close to the target site and guided by Cas9 nucleic acid endonuclease for cleavage to break the DNA double strand. Then homologous recombination repair (HDR) is performed by exogenous oligo carrying the target site mutation as a template, and the target site mutation is recombined into the genomic target site to complete the *LMBR1* gene site mutation vector construction. The control vector consisted of a segment of nonsense gRNA (Supplementary Fig. 3a, b). **b** Cells after

24 h transfection. The upper part is the control group, and the lower part is the mutant group. The dark field cell picture is in the middle and the bright field cell picture is on the right. **c** The mRNA expression of *LMBR1* in the site mutant and control cells ($n = 6$). **d** Changes in apoptotic gene expression levels after *LMBR1* site mutation ($n = 6$). **e** Cell scratch test; **f** CCK-8 cell proliferation assay ($n = 4$); **g** changes in expression levels of pathway key genes after *LMBR1* site mutation. Error bars represent the standard deviation (SD) of the mean for each data point, and error bars denote the range of the data. * p -value < 0.05, ** p -value < 0.01.

in the expression level of other key genes of the pathway after *LMBR1* site mutation. The results showed that the site mutation on the third exon of *LMBR1* significantly downregulated *Shh* gene expression ($p = 1.06e - 08$, t -test, one-tailed, two-sample equal variance hypothesis), and both *Ptch1* and *GliA* expression levels were downregulated ($p = 2.41e - 06$ and $p = 2.72e - 05$, t -test, one-tailed, two-sample equal variance hypothesis) (Fig. 5g, see data in Supplementary Table 3). The above results suggest that the missense variant on the third exon of *LMBR1* is likely to regulate digit development by affecting *Shh* expression levels.

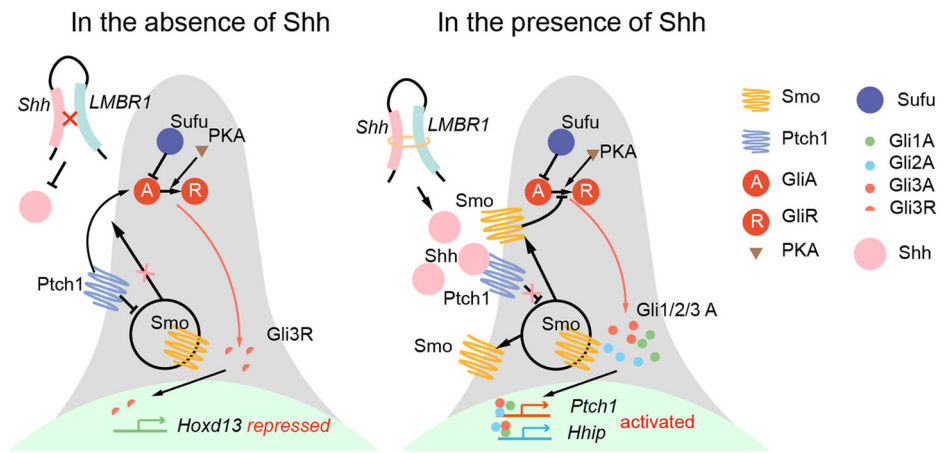
Discussion

The low fertility and dispersed breeding of horses, compared to other livestock species, have hindered the study of the molecular mechanisms responsible for their unique traits. To address this challenge, the present study utilizes a small-scale half-sibling family and integrates genetic mode inference with comparative genomic analysis to identify variants that may be related to the development of polydactyly. In this study, the kinship between individuals in the paternal half-sib family ensured the similarity of the individuals' genetic backgrounds and facilitated the screening of trait-related variants. Other rare species face challenges in establishing large

populations and advancing the study of rare traits, particularly rare diseases. Therefore, this study proposes a candidate gene screening method that could serve as a reference for genetic studies of rare and endangered animals.

The evolution of the horse is characterized by a reduction in the number of digits. This reduction in digits ultimately led to the modern true horse. Fossil evidence shows that the ancestral horse had primitive pentadactylid limbs and forelimbs with four digits and hindlimbs with three digits, with the disappearance of the first digit in the forelimbs and degenerate remnants of the first and fifth digits in the hindlimbs⁵⁷. In the Oligocene, the fore and hind limbs became three digits, and the remnants of the fifth digit can be seen in some specimens. Later, the horse evolved into a true horse with a well-developed middle digit and degenerated other digits. The middle digit, which was used for walking, gradually became stronger, and the tip of the hoof was widened and covered with a cuticle to make standing more stable, eventually evolving into the modern horse with a complete third metacarpal in the distal forelimb⁵⁸. Some vertebrates, such as chickens, camels, and pigs, have evolved with a reduction in the number of digits over a long period of time^{11,59,60}. Some genes that control limb development are more conserved across species, so for genes controlling polydactyly traits,

Fig. 6 | *LMBR1* affects Sonic hedgehog (*Shh*) signal transduction in vertebrate cells. This figure depicts the process of altered pathway signaling resulting from *LMBR1* affecting *Shh* expression levels.



the same selective process may have been selected in the same way during the evolution of multiple species, making polydactyly a model trait for evolutionary studies.

This study identified a missense variant in *LMBR1* related to the polydactyly trait in horses. However, due to the challenges of conducting individual-level experiments in horses and the divergence of this site from that observed in mice, utilizing mice as a model organism for functional studies is also problematic. As a result, validating the function of this variant at the organismal level remains difficult. Studies at the cellular level can provide insights into the variant's potential role, but are insufficient to establish its functional impact conclusively. Consequently, the current findings can only demonstrate a between the variant and equine polydactyly, without confirming a causal relationship. Further validation and investigation are necessary to clarify this link.

In the present study, the gene *LMBR1* was identified as a possible key gene in equine polydactyly. It encodes a member of the LMBR-like membrane protein family and is a lipocalin transmembrane receptor. In other studies, *LMBR1* was found to be associated with preaxial polydactyly development by affecting the expression level of *Shh* and regulating limb development^{61,62}. It has been shown that over-expression and knockdown of this gene resulted in changes in digit number in house mouse⁵¹. Combined with the site mutation of *LMBR1* in equine polydactyly, it is hypothesized that it may be due to a site variation resulting in a functional change that leads to polydactyly, requiring an intensive study to verify the role of the variate site for polydactyly. *LMBR1* has not been shown to directly regulate polydactyly by expression changes, but the variant site caused a reduction in the expression levels of *LMBR1* and *Shh* in our study. When *Shh* expression levels are reduced to lower protein concentrations and prevent the activation of *Smo*, it is unable to contain the conversion of Gli3A to Gli3R, thereby inhibiting the expression of genes such as *Ptch1* and *Hox13*⁶³ (Fig. 6). In the present study and according to OMIM and OMIA, changes in the expression levels of key genes in the Shh pathway aligned with the signaling process of this pathway. This suggests that the mutation may regulate digit development by affecting the Shh pathway. However, previous studies have shown that the ZRS region of *LMBR1* binds to *Shh*, and alterations in the CTTF motif of *LMBR1* affect the formation of a hairpin structure between *LMBR1* and *Shh*, thereby influencing the expression of *Shh*⁶⁴. The variant site identified in this study is not located at the binding site, and there is no evidence yet that its mutation affects the binding to *Shh*. The results of cellular experiments showed that mutation of the site also caused the downregulation of the *Shh* expression level, and the mechanism causing its downregulation needs to be further investigated. In the current study, *LMBR1* tended to have pleiotropism, was associated with endocytic cups, and played a role in cell migration⁶⁵. This gene has also been shown to be associated with body size in chicken⁶⁶. In horses, the embryonic period is a critical stage for the development of polydactyly⁶⁷, and its extraction is

difficult, so targeted experiments in the future are needed to explore the molecular mechanism of *LMBR1* on equine polydactyly. In addition, 32 SNPs out of 76 SNPs screened in this study were located in the exon regions of 25 genes (Supplementary Data 2). Supplementary Data 2 has added OMIM and OMIA entries for all genes harboring the 76 SNPs to aid in understanding their known functions in genetic diseases. These genes may inform in the future in equine polydactyly studies, where there may be undiscovered genes regulating polydactyly.

In addition, the structure of the *LMBR1* gene has changed during evolution, although it is largely similar. Interestingly, the presence of an extra exon at the 3' end of *LMBR1* in horses compared to other animals is a feature that will be investigated to see if it is specific to odd-toed ungulates. The gene diverged during the evolution of reptiles into mammals and birds. In mammals, the gene was under selection and showed similar sequences, whereas in birds the gene sequence is more similar to that of reptiles and was not strongly selected during this evolutionary process. The results confirm the evolution of vertebrates.

Methods

Samples

The horses used for this study were Yili horses, a breed originating from the Xinjiang Uygur Autonomous Region in northwestern China. In this study, the 200 horses used for filtering false-positive sites comprised 12 breeds sourced from farms in Hebei, Inner Mongolia, Qinghai, Xinjiang, and Hubei, including Arabian, Chakouyi, Thoroughbred, Dali, Kazakh, Hequ, Lichuan, American Saddlebred, Ningqiang, Shetland, Wenshan, and Wumeng horses. The large-scale validation population used in the study consisted of 14 breeds from Beijing, Hebei, Xinjiang, Inner Mongolia, Gansu, Yunnan, Shaanxi, and Hubei. This included 11 Arabian horses, 3 Akhal-Teke horses, 11 Thoroughbreds, 4 Shetland ponies, 8 American Saddlebreds, 21 Quarter Horses, 9 Dali horses, 2 Kazakh horses, 3 Lichuan horses, 1 Chakouyi horse, 3 Ningqiang horses, 9 Wenshan horses, 5 Wumeng horses, and 10 Yili horses.

They were fed adequate food and water and given sufficient space to move around to ensure that the horses were in a healthy condition. Blood was collected using blood collection tubes and stored at -80°C . Blood was digested overnight at 56°C (erythrocyte lysate, TianGen Blood Genomic DNA Extraction Kit, Beijing, China) and DNA was extracted using the blood genome extraction Kit (Tiangen Biotech Co., Ltd., Beijing, China) according to the manufacturer's instructions. The DNA was stored at -20°C .

Genome resequencing

Venous blood was aseptically collected from seven polydactylous families using EDTA anticoagulated vacuum blood collection tubes. Concurrently, 5 ml of venous blood was collected in batches from a total of 200 normal

horses across 12 breeds, including Yili, Wumeng, Wenshan, Ningqiang, Lichuan, Dali, Chakouyi, Kazakh, Quart, American Saddlebred, Shetland, Akhal-Teke, Arabian, and Thoroughbred. The samples were transported to the laboratory in an ice box, ensuring collection-to-lab time was within 6 h.

Extraction of genomic DNA from blood. Whole blood DNA was extracted using the Whole Blood DNA Extraction Kit (Tengen, #DP304) following the manufacturer's instructions. Use 800 μ L of fresh blood for DNA extraction, utilizing the TianGen Blood Genomic DNA Extraction Kit. Add a volume of red blood cell lysis buffer equal to twice the blood volume, mix by inverting, and leave at 56 °C overnight to ensure complete lysis. Centrifuge at 12,000 rpm for 1 min, discard the supernatant, and use absorbent paper to remove any residual liquid from the tube walls. Add 200 μ L GS, 30 μ L Proteinase K, and 200 μ L GB solution, and incubate overnight at 56 °C, 1500 rpm in a metal bath for digestion. Allow the mixture to stand at room temperature for 5 min, then add 350 μ L buffer BD and mix by inverting. Transfer the resulting solution and flocculent precipitate into the adsorption column of a collection tube. Centrifuge at 12,000 rpm for 1 min and discard the waste liquid in the collection tube. Add 500 μ L buffer GBD to the adsorption column and centrifuge at 12,000 rpm for 1 min. Discard the waste liquid in the collection tube. Add 600 μ L PWB washing buffer to the adsorption column and let it stand at room temperature for 5 min. Centrifuge at 12,000 rpm for 1 min and discard the waste liquid. Repeat the washing step twice. Then centrifuge the empty column at 12,000 rpm for 2 min to remove residual liquid. Discard the waste and allow the column to air dry at room temperature for 5 min to ensure complete removal of any residual washing buffer. Finally, add 50 μ L TE elution buffer, let it stand at room temperature for 5 min, and centrifuge at 12,000 rpm for 1 min. Reapply the eluted liquid to the adsorption column for a second elution.

Genomic band integrity was assessed by agarose gel electrophoresis, and DNA concentration and OD values were measured using a NanoDrop2000 (Thermo, USA). DNA samples that passed quality control underwent high-throughput conventional second-generation small fragment library construction and sequencing using the UW DNBSEQ-T7 platform. Both library construction and sequencing were conducted by UW Genetics. The data output aimed for an average sequencing depth of \sim 30 \times for the genomes of the seven polydactylous families and an average sequencing depth of \sim 10 \times for the genomes of the 200 normal horse samples, with all reads being 150 bp in length.

Data processing methods

To screen the final variants dataset of the half-sib family, we established a three-step locus verification strategy, which includes screening based on genetic models, expanding the negative cohort validation, and functional gene annotation. Considering the high proportion of false-positive loci in resequencing data, we adopted a manual rule-based secondary filtering method to prioritize loci that best fit the genetic rules. The specific methods are as follows:

Quality control. The raw short reads obtained from sequencing were quality controlled using the fastp software (version 0.21.0) with default parameters, resulting in clean data.

Alignment and file processing. The clean data were aligned to the horse reference genome (EquCab3.0) using the speedseq software (version 1.1). The Samtools software (version 1.18) was used to convert file formats (from .sam to .bam) and to sort the bam files according to genomic coordinates, resulting in sorted bam files.

Variant calculation and secondary filtering. For each sample, the sorted .bam files were processed with the speedseq software (speedseq var command)⁶⁸ using the built-in FreeBayes algorithm to calculate autosomal variants, generating a comprehensive variants table for all samples (.vcf file). To reduce false positives, we performed a secondary threshold

screening on the mutation files with the following criteria: (i) exclude data points with an average coverage of $<5\times$. (ii) Remove loci with an average quality value (qv) of <20 . This resulted in quality-controlled mutation .vcf files.

Genetic model screening and negative cohort validation. Based on the phenotypic prediction of polydactyly inheritance patterns, we used R to screen the .vcf files for variants that fit the genetic models (including SNPs and InDels). Autosomal dominant inheritance screening criteria: set polydactyly individuals as 0/1 and normal individuals as 0/0, excluding loci on the X chromosome. Autosomal recessive inheritance screening criteria: set polydactyly individuals as 0/0 and normal individuals as 0/1. Y-linked inheritance screening criteria: set polydactyly individuals as 0/1 and normal individuals as 0/0.

Additionally, we included 200 normal horses from 12 breeds, including Yili horses, as known negative controls. Secondary filtering was performed to exclude false-positive mutation loci present in normal individuals.

Functional gene annotation. Finally, we screened the variants related to polydactyly based on functional gene annotation.

Primer

The physical variant positions were retrieved from NCBI and Ensembl (<https://asia.ensembl.org/>). Primers were designed using Primer 5.0, as shown in Supplementary Table 5.

PCR amplification system and program

Using the blood genome as the template, polymerase chain reaction (PCR) was used to amplify the region in *LMBR1* containing the SNP. The 25 μ L PCR reaction system included: 21.5 μ L T3 super PCR mix (Tsingke, China), 1.5 μ L template, 1.0 μ L 20 mM forward primer, and 1.0 μ L 20 mM reverse primer. PCR was run at 98 °C for 2 min, followed by 30 cycles of 98 °C for 10 s, 56 °C for 10 s, 72 °C for 15 s, and 72 °C extension for 5 min, then reduced to 4 °C to terminate the reaction. The reaction products were detected by 2% gel agarose electrophoresis.

Sequence analysis

The PCR products were sent to SinoGenoMax (SinoGenoMax Co., Ltd. Beijing, China) for purification and sequencing. SnapGene software was used for the comparison and BioXM 5.0 was used for the alignment analysis. To identify *LMBR1*, the amino acid sequences of *LMBR1* of *Mus musculus* (mouse), *Sus scrofa* (Pig), *Gallus gallus* (chicken), *Bos taurus* (cattle), *Equus caballus* (horse), and *Homo sapiens* (human) from NCBI were subjected to analysis using blastp (expect threshold = 10) in Ensembl (<https://asia.ensembl.org/>).

The comparative genomic analyses in this study included multi-species covariance analysis and sequence similarity analysis. The covariance analysis was performed by matching the exon sequences of each species and linking exons with similar sequences between species. The phylogenetic tree was constructed using MEGA 6 software^{69–72} based on the protein sequences of the species, with bootstrap replications set at 500. The sequence similarity analysis was performed using Simplot 3.5.1, and the *LMBR1* genomic sequence from the equine reference genome EquCab3.0 (GCA_002863925.1) was used as the reference sequence.

Loci screening

A total of 200 normal horses from 12 breeds, including Yili, Wumeng, Wenshan, Ningqiang, Lichuan, Dali, Chakouyi, Kazakh, Akhal-teke, American Saddlebred, Shetland, Arabian, and Thoroughbred. The DNA samples were sent to BGI Genomics Co., Ltd. for high-throughput sequencing using the DNBSEQ-T7, with a sequencing depth of 30 \times in the 7 polydactyl individuals and 10 \times in the 200 normal horses. On completion of sequencing, the raw resequencing data from the seven polydactyl individuals were mapped to the equine reference genome (EquCab3.0) and combined with call variants to generate a vcf file, which was then subjected

to quality control. Data were filtered based on quality scores, and data with <5-fold coverage were excluded. The genetic pattern of equine polydactyly was predicted based on phenotype, and variants that matched the genetic pattern were filtered, and then 200 normal horses, including 12 breeds were used to filter out false-positive sites, and loci associated with polydactyly were filtered based on gene function annotations.

LMBR1 cytological validation

Monoclonal sequencing. Take a clean 1.5 mL centrifuge tube, add 4 μ L of the purified PCR product and 1 μ L of the pEASY-Blunt Cloning Vector. Incubate at 37 °C for 5 min to ligate the PCR product into the vector, then place the centrifuge tube on ice. Take 50 μ L of competent cells from the -80 °C freezer, thaw on ice, and immediately add 5 μ L of the ligation product. Gently pipette to mix and incubate on ice for 30 min. After incubation, heat shock the cells at 42 °C in a metal bath for 30 s, then place them back on ice for 2 min. Add 250 μ L of LB medium to the tube, shake at 200 rpm, and incubate at 37 °C for 1 h. Centrifuge at 5000 rpm for 1 min, remove 150 μ L of the supernatant, and resuspend the remaining liquid. Plate onto LB solid medium containing ampicillin and incubate overnight at 37 °C. Once colonies appear on the plate, pick five colonies from each plate, place them in 10 μ L of sterile water, vortex to mix, and perform colony PCR.

Isolation and culture of equine progenitor bone marrow mesenchymal stem cells. The leg bone of a 3-year-old mare from a domestic horse farm was taken after slaughter, placed in a sterile self-sealing bag, and returned to the laboratory within 4 h for cell isolation. The bone marrow was collected from the bone cavity and transferred to a 50 mL centrifuge tube. Then the bone marrow was washed three times with a washing solution (3% triple antibody (Gibco, Rockville, MD, USA) + 1% fetal bovine serum (FBS; Gibco, Rockville, MD, USA), supplemented with PBS (Gibco, Rockville, MD, USA) to 500 mL). To ensure thorough cutting of the bone marrow, clean and sterilized surgical scissors, and ophthalmic forceps were used. The appropriate amount of washing solution was added and mixed thoroughly. The mixture was then centrifuged at 800 \times g for 5 min. The supernatant was discarded and the process was repeated four times. The precipitate was collected, washed with an appropriate amount of growth medium (20% FBS + 1% sodium pyruvate + 1% dual antibody + 1% hFGF + 1% NEAA, supplemented with Alpha-MEM to 50 mL), and centrifuged to remove the supernatant. Cells were cultured in the Minimum Essential Medium α (MEM α) medium (Gibco, Rockville, MD, USA) containing 1% FBS, and 1% double antibiotics (Gibco, Rockville, MD, USA). The collected precipitate was resuspended and then seeded at a density of 1×10^5 cells with 15 mL of growth medium (20% FBS + 1% sodium pyruvate + 1% dual antibody + 1% hFGF + 1% NEAA, supplemented with Alpha-MEM to 50 mL) in a T75 cell culture flask. The cells were incubated at 37 °C with 5% CO₂ in a cell culture incubator. Cell status was observed at 24, 48, and 72 h. The culture medium was replaced every 48 h.

Purification of equine primary bone marrow mesenchymal stem cells. When the cell confluence reached ~90% for passaging, the following steps were taken: the culture dish was tilted and the cell culture fluid was aspirated along the side wall with a pipette gun. Then, the culture dish was rinsed with PBS (Gibco, Rockville, MD, USA) and discarded. Next, 2 mL of pre-warmed 0.1% trypsin (TE) was added, and the dish was shaken gently to ensure the reagent completely covered the bottom of the dish. Finally, the dish was incubated in the incubator at 37 °C with 5% CO₂ for 2 min. It was observed that the culture dish had a frosted texture and that the cells did not adhere to the wall under the microscope. The cells were dispersed by gently flicking the wall of the culture dish several times by hand. To terminate digestion, two times the volume of trypsin was added to complete the medium to inactivate it. The liquid was aspirated from the bottom of the dish using a pipette gun and the cells were gently blown down from the top. The cell suspension was transferred to a centrifuge tube and centrifuged at

1000 rpm for 5 min. The supernatant was carefully removed with a pipette, 2 mL of PBS was added and mixed to wash off any residual trypsin, and centrifuged again. PBS was discarded, and the appropriate amount of complete medium was added to resuspend the cells. The cells were then seeded into the cell culture dish or culture flask according to the desired cell density (1×10^5). The appropriate amount of medium was added. The cell culture dish or culture flask was returned to the incubator for continuation of the culture. Because bone marrow MSCs are adherent cells, while erythrocytes and lymphocytes are non-adherent cells, cell purification can be performed by fluid exchange passaging.

Site-directed mutation vector construction obtaining gRNA annealing products. The reaction system and procedure are as follows: 10X Buffer 3 μ L, gRNA 3 μ L, ddH₂O 24 μ L, total 30 μ L. Set the reaction to 95 °C for 5 min, then cool down at a rate of 8 s/0.1 °C until it reaches 25 °C, which takes about 90 min. Finally, hold at 4 °C. Digest the vector with restriction enzymes. The digestion system includes 2 μ g of plasmid, 5 μ L of 10X Reaction Buffer, and 1 μ L of each restriction enzyme. Add sterile water to a final volume of 50 μ L and incubate in a metal bath at 37 °C for 2 h. Use agarose gel electrophoresis to confirm digestion efficiency, then excise the target band from the gel and use Qiagen's QIAquick Gel Extraction Kit to recover the gel. Using the QIAquick Gel Extraction Kit, mix the annealed product and linearized vector in a certain ratio for the ligation reaction. Ligation system: solution I 2 μ L, annealed product 1 μ L, linearized vector 1 μ L, ddH₂O 5 μ L. After mixing, incubate in a metal bath at 16 °C for 1 h, then proceed to transformation. Transform the ligation product into Stbl3 competent cells: add 5 μ L of the ligation product to 100 μ L of Stbl3 chemically competent *Escherichia coli* and incubate on ice for 30 min. Heat shock the cells at 42 °C for 30 s. Immediately transfer to ice and incubate for 2 min. Add 250 μ L of S.O.C. medium and incubate at 37 °C, 225 rpm for 1 h. Plate 350 μ L of the transformation product onto an LB plate containing 100 μ g/mL ampicillin and incubate overnight at 37 °C. Perform colony PCR to identify positive transformants. Pick the colonies from the plate, resuspend in 50 μ L of LB medium, and use 2 μ L as the template for colony PCR. The reaction system and PCR cycling conditions are as follows: forward primer (10 μ M) 1 μ L, reverse primer (10 μ M) 1 μ L, bacterial suspension 2 μ L, Taq Premix (2X) 7.5 μ L, ddH₂O up to 15 μ L. The cycling conditions are: 95 °C for 3 min, 95 °C for 15 s, followed by annealing for 15 s, extension at 72 °C for 1 min, for 30 cycles, then a final extension at 72 °C for 5 min. Extract plasmids from positive clones using the Tiangen Endo-free Plasmid Maxi Kit: add 2.5 mL of equilibration buffer (BL) to the adsorption column CP6 (placed in a 50 mL collection tube), centrifuge at 8000 rpm for 2 min, discard the waste liquid, and place the column back into the collection tube. Add 100 mL of overnight-cultured bacterial suspension into a centrifuge tube, centrifuge at 8000 rpm for 3 min at room temperature to collect the bacteria, and carefully remove the supernatant. Repeat until all bacterial suspensions are collected. Remove the supernatant as much as possible and use absorbent paper to remove any water droplets on the bottle wall. Add 8 mL of Solution P1 to the bacterial pellet, and vortex to completely resuspend the pellet. Add 8 mL of Solution P2 to the tube, and gently invert 6–8 times to fully lyse the cells, and leave at room temperature for 5 min. Add 8 mL of Solution P4, gently invert 6–8 times to mix thoroughly until white flocculent precipitate appears. Let stand at room temperature for about 10 min, then centrifuge at 8000 rpm for 10 min. Carefully pour the entire solution into filter CS1, slowly push the plunger to filter, and collect the filtrate in a clean 50 mL tube. Add 0.3 times the volume of isopropanol to the filtrate, invert to mix, then transfer to the adsorption column CP6 (placed in a 50 mL collection tube). Centrifuge at 8000 rpm at room temperature for 2 min, discard the waste liquid, and place the adsorption column CP6 back into the collection tube. Repeat the step for the remaining liquid. Add 10 mL of wash buffer (PW) to the adsorption column CP6, centrifuge at 8000 rpm for 2 min, discard the waste liquid, and place the column back into the collection tube. Add 3 mL of absolute ethanol to the

adsorption column CP6, centrifuge at 8000 rpm for 2 min, and discard the waste liquid. Place the adsorption column CP6 back into the collection tube and centrifuge at 8000 rpm for 5 min to remove any residual wash buffer. Open the lid of the adsorption column CP6 and let it air dry at room temperature for a few minutes. Place the adsorption column CP6 into a clean 50 mL collection tube, add 1 mL of elution buffer (TB) to the center of the membrane, let it stand at room temperature for 5 min, then centrifuge at 8000 rpm for 2 min. To ensure higher yield, pass the eluate through the column again, then transfer it to a clean 1.5 mL tube and store it at -20°C . The relevant and oligo primer sequence information can be found in Supplementary Table 5.

Real-time quantitative PCR (RT-PCR). Amplification reaction system (15 μL): sterilized water up to 15 μL , SYBR Green Master Mix 7.5 μL , Primer (F/R) 0.3/0.3 μL , Template 1 μL . PCR amplification conditions: pre-denaturation at 95°C for 30 s. Denaturation at 95°C for 5 s. Annealing at 60°C for 30 s. Repeat steps ②–③ for 40 cycles. Melting curve: start at 65°C for 5 s, increasing the temperature by 0.5°C increments each cycle until reaching 95°C ; END.

Cell transfection and flow cytometric sorting. The cells were harvested from the culture dish at 90% confluence, following the same steps as before. The cell count was determined, and $0.25\text{--}1 \times 10^6$ cells were seeded per well in a six-well plate. The cells were evenly distributed by blowing and mixing the plate thoroughly. Infection was initiated when the cell confluence reached 70%–90%. Transfection was carried out using the appropriate amount for each well of the six-well plate, and three replicates were established for each group. (i) Three centrifuge tubes were taken, labeled as tube 1, tube 2, and tube 3, respectively. (ii) In tube 1, 250 μL Opti-MEM medium and 10 μL Lipo 3000 were added. In tube 2, 125 μL Opti-MEM medium, 2.5 μg plasmid, and 2.5 μL P3000 reagent were added to dilute the plasmid. (iii) Tube 2 was filled with 125 μL Opti-MEM medium, 2.5 μg plasmid, and 2.5 μL P3000 reagent for the experimental group. Tube 3 was filled with 125 μL Opti-MEM medium, 2.5 μg control plasmid, and 2.5 μL P3000 reagent. (iv) Added 25–150 pmol Oligo to tube 2 and tube 3, respectively, and mixed gently. (v) In total, 125 μL of diluted plasmid was taken from tube 2 and tube 3, respectively, and mixed with the liquid in tube 1 in equal volume. The mixture was gently blown and mixed, and then incubated for 15 min at room temperature. (vi) Added 250 μL of the mixture to each well and gently shook the six-well plate to evenly distribute the reagent over the cells. (vii) The culture was continued by replacing the cell culture medium with a complete medium containing 20% FBS 12 h after transfection. (viii) After 24 h, the successfully transfected cells were observed under a fluorescence microscope emitting green fluorescence. (ix) The cells were digested, resuspended in 100 μL of medium, and then centrifuged. The cells that were successfully transfected and had green fluorescence were sorted out using flow cytometry and then cultured further.

Total RNA extraction and reverse transcription. Add 1 mL of Trizol to the cell culture dish, pipette to mix, and incubate at room temperature for 15 min. Transfer the lysed cell suspension to a 2 mL centrifuge tube, add 0.2 volumes of chloroform, vortex for 1 min, and incubate at room temperature for 15 min. Centrifuge at 12,000 rpm, 4°C , for 15 min. The liquid will separate into three layers: the upper colorless layer contains RNA, the middle white layer contains DNA, and the bottom red layer contains proteins. Transfer the upper colorless layer to a new pre-chilled centrifuge tube. Add an equal volume of pre-chilled isopropanol, invert gently to mix, and incubate at room temperature for 10 min. Centrifuge at 12,000 rpm, 4°C , for 10 min. Discard the supernatant, add 75% ethanol (prepared with DEPC water) to the pellet, and gently pipette to mix. Centrifuge at 7500 rpm, 4°C , for 5 min, and discard the supernatant.

Repeat the washing step 2 more times. Air dry in a biosafety cabinet for 5 min to completely evaporate residual ethanol. Add 20 μL of DEPC water and place it in a 4°C refrigerator for 2 h to fully dissolve the RNA pellet.

Measure the RNA concentration and purity using a UV spectrophotometer. Use the RNA if it meets quality requirements. For reverse transcription, adjust the extracted total RNA concentration to 1000 ng/ μL , and perform reverse transcription according to the instructions of the Thermo reverse transcription kit. The reverse transcription system is as follows: random primer 1 μL , total RNA 3 μg , and ddH₂O to a final volume of 12 μL . Then add 4 μL of 5X Reaction Buffer, 1 μL of Ribolock RNase Inhibitor (20 U/ μL), 2 μL of 10 mM dNTP Mix, and RevertAid M-MuLV RT (200 U/ μL) to make a total volume of 20 μL . Incubate at 25°C for 5 min, 42°C for 60 min, and 70°C for 5 min. After reverse transcription, store the cDNA at -20°C . During the reverse transcription procedure, prepare all components on ice. To minimize errors during pipetting, slightly exceed the actual required volume when preparing the reaction mix. Add the pre-extracted RNA sample using a pipette. Dilute the resulting cDNA fivefold.

Real-time quantitative PCR (RT-PCR) for detection of mRNA expression levels of proliferation and apoptosis genes. *GAPDH* was used as the internal reference gene. The expression levels of apoptosis genes (*Caspase3*, *Caspase7*) and the proliferation gene (*Bcl-2*) were detected. Primers for these genes were designed using NCBI (Supplementary Table 5) and synthesized by Tsingke Biotechnology Co., Ltd.

Real-time quantitative PCR (RT-PCR) to detect mRNA expression levels of limb development pathway genes and *LMBR1*. To investigate the role of *LMBR1* in the pathway, we selected genes from the SHH pathway, including *SHH*, *PITCH1*, and *GLI3A*, for expression analysis. The pathway genes were sourced from KEGG, and primers were designed using the NCBI online tool and synthesized by Tsingke Biotechnology Co., Ltd. The primer sequences used are shown in Supplementary Table 5.

CCK-8 analysis. The cells were harvested from the culture dish at 90% confluence, following the previously described procedure. The cell suspension was then quantified and seeded at a density of 4×10^4 cells per well in a 96-well plate, with four replicates set up for each plasmid. The transfection steps were performed as previously described. In tube 1, 10 μL Opti-MEM medium and 0.2 μL Lipo 3000 were added. In tube 2 and tube 3, 5 μL Opti-MEM, 0.1 μg plasmid, 0.2 μL P3000, and 1–60 pmol Oligo were added. The plasmid was diluted and mixed with an equal volume of Lipo 3000 reagents. Next, 10 μL of the mixture was added to each well of a 96-well plate, which was then gently shaken to ensure even distribution of the reagents. Cell viability was measured at five time points (0, 12, 24, 36, and 48 h) using an enzyme marker. Prior to cell viability measurement, 10 μL of CCK-8 reagent (Solarbio, Beijing, China) was added to each well and incubated for 1.5 h at 37°C in the incubator.

Cell scratch test. The successfully transfected cells were sorted out and spread in a 24-well plate. After the cells were fully grown, two blank scratches were made in a cross pattern using a white gun with a sterilized straightedge. The cells were washed twice with PBS to remove the residual cells and then observed and photographed at 0, 24, 48, and 72 h. The left side of the cross was used for observation.

Statistics and reproducibility

Statistical analysis. All statistical analyses were performed using Excel. Descriptive statistics were calculated for all variables, and results are presented as means \pm standard deviations (SD) unless otherwise stated. For comparisons between groups, we used *t*-tests and ANOVA with a significance level set at $p < 0.05$.

Reproducibility. To ensure reproducibility, all experiments were conducted with at least four biological replicates. Data collection and analysis were performed independently by three researchers to minimize bias. Detailed protocols for sample preparation, data collection, and analysis are provided in the “Methods” section. All datasets generated and analyzed during the current study are available from the corresponding

author on reasonable request. Resequencing data in this study are available at NCBI with accession number PRJNA1099313. The ss ID of the variant is ss11846200000.

Randomization. Samples were randomly collected from the study area using a simple random sampling technique, where each individual had an equal chance of being selected.

Data availability

Resequencing data in this study are available at NCBI with accession number PRJNA1099313. The genome sequences of *LMBRI* for human⁵³, mouse⁵⁵, chicken⁵³, rat⁷³, zebrafish⁷⁴, pig⁷⁵, cattle⁷⁶, dog⁷⁷, camel⁷⁸, and crocodile⁷⁹ were obtained from the NCBI Gene database, including NC_000007.14, NC_037331.1, NC_000071.7, NC_052533.1, NC_086022.1, NC_010460.4, NC_051820.1, NC_087442., NW_017728886.1.

Code availability

The software used in this study includes SpeedSeq (version 1.1) and Samtools (version 1.18). Both are publicly available and can be accessed as follows: SpeedSeq: <https://github.com/hall-lab/speedseq>. Samtools: <http://www.htslib.org/> and <https://github.com/samtools/samtools>. We used default parameters for SpeedSeq and Samtools unless otherwise specified. Custom scripts or specific configurations used for data processing are available upon request from the corresponding author. Additionally, any relevant variables or parameters used to generate, test, or process the current dataset are documented in the “Methods” section of the manuscript.

Abbreviations

Abbreviation	Full name
SNP	Single nucleotide polymorphism
Ref	Reference genome
Alt	Alternative genome
PCR	Polymerase chain reaction
Shh	Sonic hedgehog
LMBRI	Limb development membrane protein 1
FGFs	Fibroblast growth factor
MSCs	Mesenchymal stem cells

Received: 9 April 2024; Accepted: 14 October 2024;

Published online: 31 October 2024

References

- Hall, B. K. Atavisms and atavistic mutations. *Nat. Genet.* **10**, 126–127 (1995).
- Marsh, O. C. Polydactyl horses, recent and extinct. *Sci. Am.* **8**, 2983–2985 (1879).
- Gadow, H., Gegenbaur, C. & Garson, J. G. Critical remarks on polydactyly as atavism. *J. Anat. Physiol.* **16**, 615 (1882).
- Hamelin, A. et al. Genetic heterogeneity of polydactyly in Maine Coon cats. *J. Feline Med. Surg.* **22**, 1103–1113 (2020).
- Zhang, Z. et al. Parallel evolution of polydactyly traits in Chinese and European chickens. *PLoS ONE* **11**, e0149010 (2016).
- Foster, M. Wright’s studies of effects of maternal age on spotting and polydactyly in the guinea pig. *Ann. N. Y. Acad. Sci.* **57**, 491–493 (1954).
- Leipold, H. W., Dennis, S. M. & Huston, K. Polydactyly in cattle. *Cornell Vet.* **62**, 337–345 (1972).
- Warburton, N. M., Cake, M. A. & Kelman, K. R. Extreme bilateral polydactyly in a wild-caught western grey kangaroo. *Anat. Rec.* **304**, 1361–1374 (2021).
- Whittington, R. J., Unger, D. B. & Wilson, J. M. Multiple congenital malformations of the face, nervous system and musculoskeletal system of pigs. *Aust. Vet. J.* **63**, 48–50 (1986).
- Qu, S. et al. Mutations in mouse *Aristaless-like4* cause Strong’s luxoid polydactyly. *Development* **125**, 2711–2721 (1998).
- Bani-Ismail, Z., Hawkins, J. F. & Siems, J. J. Surgical correction of polydactyly in a camel (*Camelus dromedarius*). *J. Zoo Wildl. Med.* **30**, 301–304 (1999).
- Villagómez, D. A. & Alonso, R. A. A distinct Mendelian autosomal recessive syndrome involving the association of anotia, palate agenesis, bifid tongue, and polydactyly in the dog. *Can. Vet. J.* **39**, 642–643 (1998).
- Kelly, D. M., Mahmoud, K. & Mauck, B. M. Polydactyly of the foot: a review. *J. Am. Acad. Orthop. Surg.* **29**, 361–369 (2021).
- Sananta, P., Sahdiniar, F. F. & Siahaan, L. D. The management of mirror foot polydactyly: a case report. *Int. J. Surg. Case Rep.* **91**, 106780 (2022).
- He, C. et al. Genetic pattern and gene localization of polydactyly in Beijing fatty chicken. *PLoS ONE* **12**, e0176113 (2017).
- Pitel, F. et al. Mapping the naked neck (NA) and polydactyly (PO) mutants of the chicken with microsatellite molecular markers. *Genet. Sel. Evol.* **32**, 73–86 (2000).
- Danforth, C. H. Heredity of polydactyly in the cat. *J. Hered.* **38**, 107–112 (1947).
- Browning, R. J., Hayes, E. G. & Lear, A. S. Spontaneous appearance and transmission of polydactyly in dexter cattle. *Case Rep. Vet. Med.* **2020**, 6407847 (2020).
- Gorbach, D., Mote, B., Totir, L., Fernando, R. & Rothschild, M. Polydactyly inheritance in the pig. *J. Hered.* **101**, 469–475 (2010).
- Wettker-Schäfer, R. & Kantner, G. X-linked dominant inherited diseases with lethality in hemizygous males. *Hum. Genet.* **64**, 1–23 (1983).
- Kang, S. et al. Linkage mapping and phenotypic analysis of autosomal dominant Pallister-Hall syndrome. *J. Med. Genet.* **34**, 441–446 (1997).
- Castilla, E. E., Paz, J. E. & Orioli-Parreiras, I. M. Syndactyly: frequency of specific types. *Am. J. Med. Genet.* **5**, 357–364 (1980).
- Budny, B. et al. A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome. *Hum. Genet.* **120**, 171–178 (2006).
- Kano, H., Kurahashi, H. & Toda, T. Genetically regulated epigenetic transcriptional activation of retrotransposon insertion confers mouse dactylaplasia phenotype. *Proc. Natl. Acad. Sci. USA* **104**, 19034–19039 (2007).
- Cooper, K. L. et al. Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature* **511**, 41–45 (2014).
- Brewer, J. R., Mazot, P. & Soriano, P. Genetic insights into the mechanisms of Fgf signaling. *Genes Dev.* **30**, 751–771 (2016).
- Dahn, R. D., Davis, M. C., Pappano, W. N. & Shubin, N. H. Sonic hedgehog function in chondrichthyan fins and the evolution of appendage patterning. *Nature* **445**, 311–314 (2007).
- Skoda, A. M. et al. The role of the Hedgehog signaling pathway in cancer: A comprehensive review. *Bosn J Basic Med Sci.* **18**, 8–20 (2018).
- Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol.* **20**, 781–810 (2004).
- Amerongen, R. & Berns A. Knockout mouse models to study Wnt signal transduction. *Trends Genet.* **22**, 678–689 (2006).
- Zhang, Z., Verheyden, J. M., Hassell, J. A. & Sun, X. FGF-regulated ETV genes are essential for repressing Shh expression in mouse limb buds. *Dev. Cell* **16**, 607–613 (2009).
- Zhu, J., Patel, R., Trofka, A., Harfe, B. D. & Mackem, S. Sonic hedgehog is not a limb morphogen but acts as a trigger to specify all digits in mice. *Dev. Cell* **57**, 2048–2062.e4 (2022).
- Potuijt, J. W. P. et al. The pZRS non-coding regulatory mutation resulting in triphalangeal thumb-polysyndactyly syndrome changes the pattern of local interactions. *Mol. Genet. Genomics* **297**, 1343–1352 (2022).

34. Huangfu, D. & Anderson, K. V. Signaling from Smo to Ci/Gli: conservation and divergence of hedgehog pathways from *Drosophila* to vertebrates. *Development* **133**, 3–14 (2006).
35. Lopez-Rios, J. et al. Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. *Nature* **511**, 46–51 (2014).
36. Ullah, A. et al. A novel homozygous sequence variant in GLI1 underlies first case of autosomal recessive pre-axial polydactyly. *Clin. Genet.* **95**, 540–541 (2019).
37. Umair, M. et al. A novel homozygous missense mutation in the zinc finger DNA binding domain of GLI1 causes recessive post-axial polydactyly. *Front. Genet.* **12**, 746949 (2021).
38. Lopez-Rios, J. The many lives of SHH in limb development and evolution. *Semin. Cell Dev. Biol.* **49**, 116–124 (2016).
39. Alvarez-Medina, R., Cayuso, J., Okubo, T., Takada, S. & Martí, E. Wnt canonical pathway restricts graded Shh/Gli patterning activity through the regulation of Gli3 expression. *Development* **135**, 237–247 (2008).
40. Sagai, T., Hosoya, M., Mizushima, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797–803 (2005).
41. Hars, GofflotF. et al. C. Molecular mechanisms underlying limb anomalies associated with cholesterol deficiency during gestation: implications of hedgehog signaling. *Hum. Mol. Genet.* **12**, 1187–1198 (2003).
42. Dunn, I. C. et al. The chicken polydactyly (Po) locus causes allelic imbalance and ectopic expression of Shh during limb development. *Dev. Dyn.* **240**, 1163–1172 (2011).
43. Schrauwen, I. et al. De novo variants in GREB1L are associated with non-syndromic inner ear malformations and deafness. *Hum. Genet.* **137**, 459–470 (2018).
44. Kalsoom, U. E. et al. Whole exome sequencing identified a novel zinc-finger gene ZNF141 associated with autosomal recessive postaxial polydactyly type A. *J. Med. Genet.* **50**, 47–53 (2013).
45. Umair, M. et al. Biallelic variant in DACH1, encoding Dachshund Homolog 1, defines a novel candidate locus for recessive postaxial polydactyly type A. *Genomics* **113**, 2495–2502 (2021).
46. Umair, M. et al. Exome sequencing revealed a splice site variant in the IQCE gene underlying post-axial polydactyly type A restricted to lower limb. *Eur. J. Hum. Genet.* **25**, 960–965 (2017).
47. Schrauwen, I. et al. FAM92A underlies nonsyndromic postaxial polydactyly in humans and an abnormal limb and digit skeletal phenotype in mice. *J. Bone Min. Res.* **34**, 375–386 (2019).
48. Ullah, I. et al. Variants in KIAA0825 underlie autosomal recessive postaxial polydactyly. *Hum. Genet.* **138**, 593–600 (2019).
49. Stanek, C. & Hantak, E. Bilateral atavistic polydactyly in a colt and its dam. *Equine Vet. J.* **18**, 76–79 (1986).
50. Evans, L. H., Jenny, J. & Raker, C. W. Surgical correction of polydactylism in the horse. *J. Am. Vet. Med. Assoc.* **146**, 1405–1408 (1965).
51. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
52. Betts, M. J. & Russell R. B. Amino-acid properties and consequences of substitutions. <https://doi.org/10.1002/9780470059180.ch13> (2007).
53. Clark, R. M. et al. Reciprocal mouse and human limb phenotypes caused by gain- and loss-of-function mutations affecting Lmbr1. *Genetics* **159**, 715–726 (2001).
54. Laurell, T. et al. A novel 13 base pair insertion in the Sonic hedgehog ZRS limb enhancer (ZRS/LMBR1) causes preaxial polydactyly with triphalangeal thumb. *Hum. Mutat.* **33**, 1063–1066 (2012).
55. Huang, Y. Q. et al. Single nucleotide polymorphisms in the chicken Lmbr1 gene are associated with chicken polydactyly. *Gene* **374**, 10–18 (2006).
56. Kavanagh, K. D., Bailey, C. S. & Sears, K. E. Evidence of five digits in embryonic horses and developmental stabilization of tetrapod digit number. *Proc. Biol. Sci.* **287**, 20192756 (2020).
57. Tickle, C. & Barker, H. The Sonic hedgehog gradient in the developing limb. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 275–290 (2013).
58. Dessaud, E., McMahon, A. P. & Briscoe, J. Pattern formation in the vertebrate neural tube: a Sonic hedgehog morphogen-regulated transcriptional network. *Development* **135**, 2489–2503 (2008).
59. Gan, S. Equine science. *Agric. Press* **1990**, 22–27 (1990).
60. Solounias, N. et al. The evolution and anatomy of the horse manus with an emphasis on digit reduction. *R. Soc. Open Sci.* **5**, 171782 (2018).
61. Towers, M. Evolution of antero-posterior patterning of the limb: Insights from the chick. *Genesis* **56**, e23047 (2018).
62. Sears, K. E. et al. Developmental basis of mammalian digit reduction: a case study in pigs. *Evol. Dev.* **13**, 533–541 (2011).
63. Williamson, I. et al. Developmentally regulated Shh expression is robust to TAD perturbations. *Development* **146**, dev179523 (2019).
64. Lettice, L. A. et al. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc. Natl. Acad. Sci. USA* **99**, 7548–7553 (2002).
65. Tickle, C. & Towers, M. Sonic hedgehog signaling in limb development. *Front. Cell Dev. Biol.* **5**, 14 (2017).
66. Ushiki, A. et al. Deletion of CTCF sites in the SHH locus alters enhancer-promoter interactions and leads to acheiropodia. *Nat. Commun.* **12**, 2282 (2021).
67. Kelsey, J. S., Fastman, N. M. & Blumberg, D. D. Evidence of an evolutionarily conserved LMBR1 domain-containing protein that associates with endocytic cups and plays a role in cell migration in dictyostelium discoideum. *Eukaryot. Cell* **11**, 401–416 (2012).
68. Chiang, C. et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* **12**, 966–968 (2015).
69. Xu, J. et al. Large duplication in LMBR1 gene in a large Chinese pedigree with triphalangeal thumb polysyndactyly syndrome. *Am. J. Med. Genet. A* **182**, 2117–2123 (2020).
70. Huang, Y. et al. Single nucleotide polymorphisms in chicken Lmbr1 gene were associated with chicken growth and carcass traits. *Sci. China C Life Sci.* **50**, 62–69 (2007).
71. Tamura, K., Yonei-Tamura, S., Yano, T., Yokoyama, H. & Ide, H. The autopod: its formation during limb development. *Dev. Growth Differ.* **50**, S177–S187 (2008). Suppl 1.
72. Tamura, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
73. Pasquier, J. et al. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics* **17**, 368 (2016).
74. Zimin, A. V. et al. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* **10**, R42 (2009).
75. Uenishi, H. et al. PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. *Nucleic Acids Res.* **32**, D484–D488 (2004).
76. Potuijt, J. W. P. et al. A point mutation in the pre-ZRS disrupts Sonic hedgehog expression in the limb bud and results in triphalangeal thumb-polysyndactyly syndrome. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **20**, 1405–1413 (2018).
77. Kim, S. et al. Transposable element-mediated structural variation analysis in dog breeds using whole-genome sequencing. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* **30**, 289–300 (2019).
78. Khalkhali-Evrigh, R., Hedayat-Evrigh, N., Hafezian, S. H., Farhadi, A. & Bakhtiarzadeh, M. R. Genome-wide identification of microsatellites and transposable elements in the dromedary camel genome using whole-genome sequencing data. *Front. Genet.* **10**, 692 (2019).

79. Gao, J. et al. Sequencing, de novo assembling, and annotating the genome of the endangered Chinese crocodile lizard *Shinisaurus crocodilurus*. *GigaScience* **6**, 1–6 (2017).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (202303010411070), the Chinese Universities Scientific Fund (2022RC030), the Introduction of Talent Research Start-up Fund (31051204), and Xinjiang Science and Technology Major Program (2023A02005-1).

Author contributions

Conceptualization: Qin Wang and Yue Luan; formal analysis: Yue Luan, Ling Zhong, and Cao Li; sample collection: Jianpeng Wang; funding acquisition: Qin Wang; methodology: Yue Luan, Ling Zhong, and Xiaoyu Yue; supervision: Yiping Zhu, Ling Zhong, and Mengyan Ye; writing—original draft: Yue Luan and Ling Zhong; writing—review and editing: Yue Luan and Qin Wang.

Competing interests

The authors declare no competing interests.

Ethics approval

Samples genotyped in this study were collected with informed owner's consent for commercial genetic testing and approved for use in research. As such, institutional animal research ethics was not required. Approval for the collection and the movement of genetic material for the whole-genome sequencing analyses was granted by the China Agricultural University Animal Research Ethics Committee (AW72203202-1-1). We have complied with all relevant ethical regulations for animal use.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-024-07065-w>.

Correspondence and requests for materials should be addressed to Qin Wang.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Aylin Bircan and Mengtan Xing. A peer review file is available.

Reprints and permissions 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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2024