

RESEARCH ARTICLE

Open Access



# Comparative genomic analysis of a naturally born serpentized pig reveals putative mutations related to limb and bone development

Yankai Jiang<sup>1</sup>, Xinyue Cao<sup>2</sup> and Haibin Wang<sup>1\*</sup>

## Abstract

**Background:** It is believed that natural selection acts on the phenotypical changes caused by mutations. Phenotypically, from fishes to amphibians to reptiles, the emergence of limbs greatly facilitates the landing of ancient vertebrates, but the causal mutations and evolutionary trajectory of this process remain unclear.

**Results:** We serendipitously obtained a pig of limbless phenotype. Mutations specific to this handicapped pig were identified using genome re-sequencing and comparative genomic analysis. We narrowed down the causal mutations to particular chromosomes and even several candidate genes and sites, such like a mutation-containing codon in gene *BMP7* (bone morphogenetic protein) which was conserved in mammals but variable in lower vertebrates.

**Conclusions:** We parsed the limbless-related mutations in the light of evolution. The limbless pig shows phenocopy of the clades before legs were evolved. Our findings might help deduce the emergence of limbs during vertebrate evolution and should be appealing to the broad community of human genetics and evolutionary biology.

**Keywords:** Serpentized pig, Hindlimb, Bone, Genome sequencing, Mutations

## Background

Both mutations and environmental factors shape the phenotype of living organisms. When the environmental effects are controlled, mutations become the most plausible trigger of phenotypical changes. Indeed, during the long-term evolution, mutations are the source of natural selection [1, 2]. Based on the neutral theory of evolution [3, 4], most novel mutations are deleterious and already eliminated by purifying selection. The more essential a gene is, the stronger the purifying selection would be. As

a consequence, in the inter-species comparison of orthologous genes, the essential genes tend to show less divergence (higher sequence similarity) than the unimportant genes [5, 6]. From another aspect, a novel mutation on an essential gene, if deleterious, would severely reduce the fitness and even cause the death of an individual, which maintains the high conservation level of sequence among populations.

Given the strong connection between the cause (mutation), effect (fitness change), and consequence (individual eliminated), one could infer the effect of a mutation from the mutation spectrum within a species or a population. A population consists of numerous individuals. When a novel mutation (in an individual) is beneficial,

\* Correspondence: [sdeygj1182@163.com](mailto:sdeygj1182@163.com)

<sup>1</sup>Department of Joint Surgery, The Second Hospital of Shandong University, Jinan 250033, Shandong, China

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



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the carrier would have higher chance of survival or reproduction and therefore this mutation would increase its allele frequency (AF) during evolution [7, 8]. If a deleterious mutation occurs, the carrier is less viable so that the AF of this mutation would be constantly low or sometimes this mutation would be directly eliminated from the population. At species level (which is beyond population), beneficial mutations might be eventually fixed in all individuals of a species, leading to the fixation of divergent sites between two species [9, 10]. The deleterious and beneficial mutations show completely different evolutionary traces. Therefore, analyzing the mutations in the light of evolution is helpful in distinguishing which sets of mutations are beneficial or deleterious, and might facilitate the identification of causal mutations (for a phenotype) from the sea of randomly occurred mutations.

Undoubtedly, the emergence of limbs is a milestone during the evolution of vertebrates. From the very basal water-living fishes to the most recently diverged mammalian branch (also the most successful branch), landing should be the most significant event among all the evolutionary steps, which brings about numerous advantages. Thus, understanding the evolutionary traces and molecular mechanisms of limb development is of grave importance. Experimental biologists indeed find a series of candidate genes (proteins) responsible for that. To date, the broadly conserved regulators related to limb or bone development include the T-box family transcription factor (TBX4) [11, 12], the bone morphogenetic proteins (BMP) [13, 14] (especially two members BMP7 [15] and BMP2 [16]), downstream effector SMAD [17, 18], PTX1 [19], R-spondin secreted ligand RSPO [20], forehead transcription factor FOXF [21], and HOX [22].

The above discoveries are either found in humans or some well-established model organisms. However, one still wonders if the defects in limbs could take place in natural environment or in a less-studied organism. Although genetic manipulation is not feasible in non-model organisms, the evolutionary and comparative genomic analyses across multiple species still make sense in the absence of functional verification [23]. In this study, we report a naturally born farmed pig (*Sus scrofa*) without hindlimbs, similar to the serpented phenotype [24]. Although finding deformities in farmed pigs are not rare, few cases have been thoroughly investigated by bioinformatic analysis. By considering the Mendelian independent assortment law, we also narrowed down the specific mutations to particular chromosomes and even several candidate genes and sites, such like a mutation-containing codon in gene *BMP7* which was conserved in mammals but variable in birds, reptiles, amphibians, and fishes. Our study shows that the limbless phenotype is not confined to a few model organisms but could be

more general across mammals, we also parse the specific mutations in the light of evolution. We propose the functional role of many candidate genes and sites in hindlimb development. Our work should be valuable to the human genetics as well as evolutionary biology community.

## Results

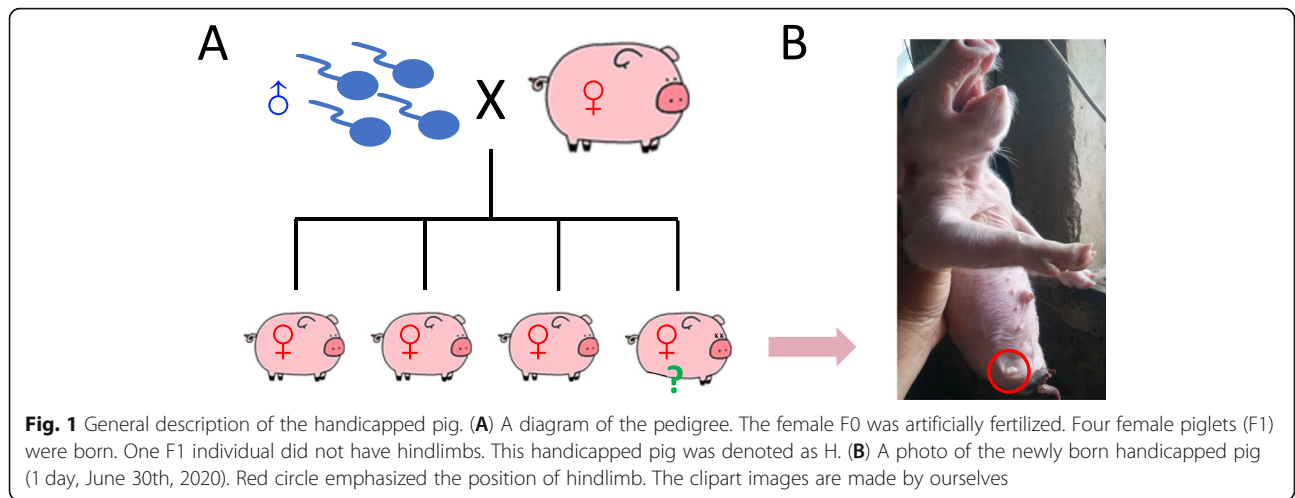
### General description of a naturally born limbless pig

We obtained a naturally born handicapped pig from a farm near Yantai city, China (see [Methods](#) for details). This handicapped pig without hindlimbs is one of the four female piglets (F1) delivered by a female pig (F0) who has previously been artificially fertilized (Fig. 1A and B). The morphology of this handicapped pig resembles the well-known serpented phenotype (Fig. 1B). Further X-ray photos confirmed that the leg bones are undeveloped (Additional file 1: Fig. S1).

The handicapped pig lived for 30 days and died of natural causes or causes due to its deformities. The pig's hind leg deformities made it harder to function like normal pigs, including walking and eating (Additional file 2: supplementary data). The failure of the hindlimb development was our main focus in this study and not the cause of the early death to investigate the if there was a link between phenotype and genotype.

### Genomic sequencing reveals numerous mutations

Intuitively, the phenotype should be governed by the genotype as well as the external factor (environment). Indeed, there are a wide range of teratogenic agents that can affect limb morphogenesis and cause vascular disruption on limbs. However, since the pigs were subjected to the same environmental conditions, it is hard for us to tell how this particular pig could be the only handicapped one among all the offspring. Therefore, we could only try to analyze the contribution of genotype to the phenotype. We sequenced the genomic DNA of the female F0 individual (denoted as M, standing for mother), the handicapped individual (denoted as H), and another F1 individual (denoted as S, standing for sister) (Fig. 2A). After mapping the DNA-sequencing reads to the reference genome, numerous bona fide mutations are found (see [Methods](#)). Briefly, M has 13,212,738 variation sites, S has 12,503,665, and H has 12,355,632. The variation types include both SNPs (point mutations) and Indels (potential frame-shift variations), covering all the chromosomes of the pig genome (Fig. 2B). Notably, when talking about variations in diploid species, it should be clarified whether a variation is homozygous or heterozygous. For deleterious mutations, homozygosity is possibly lethal, so the commonly observed deleterious mutations are heterozygous. One could conceive that Indels are more harmful than point mutations (SNPs),

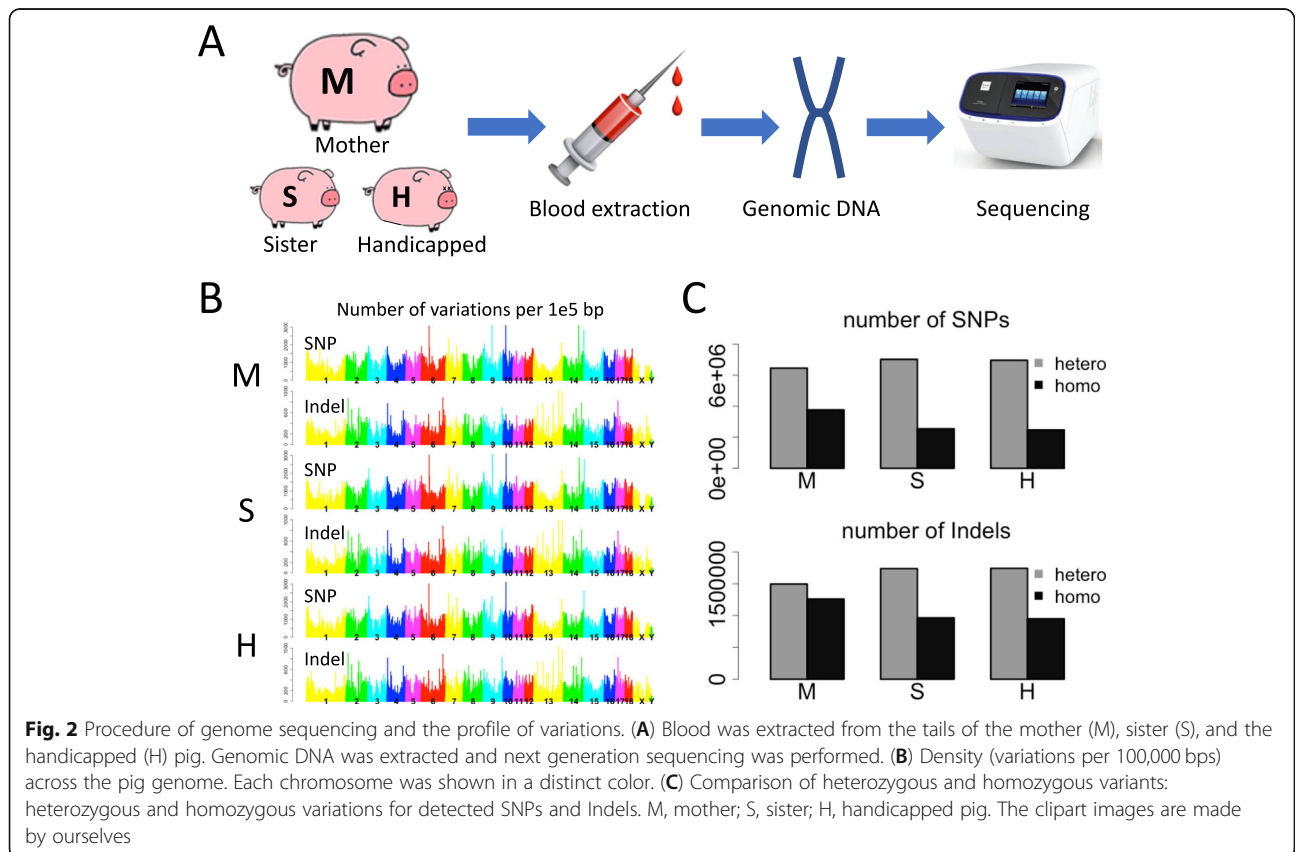


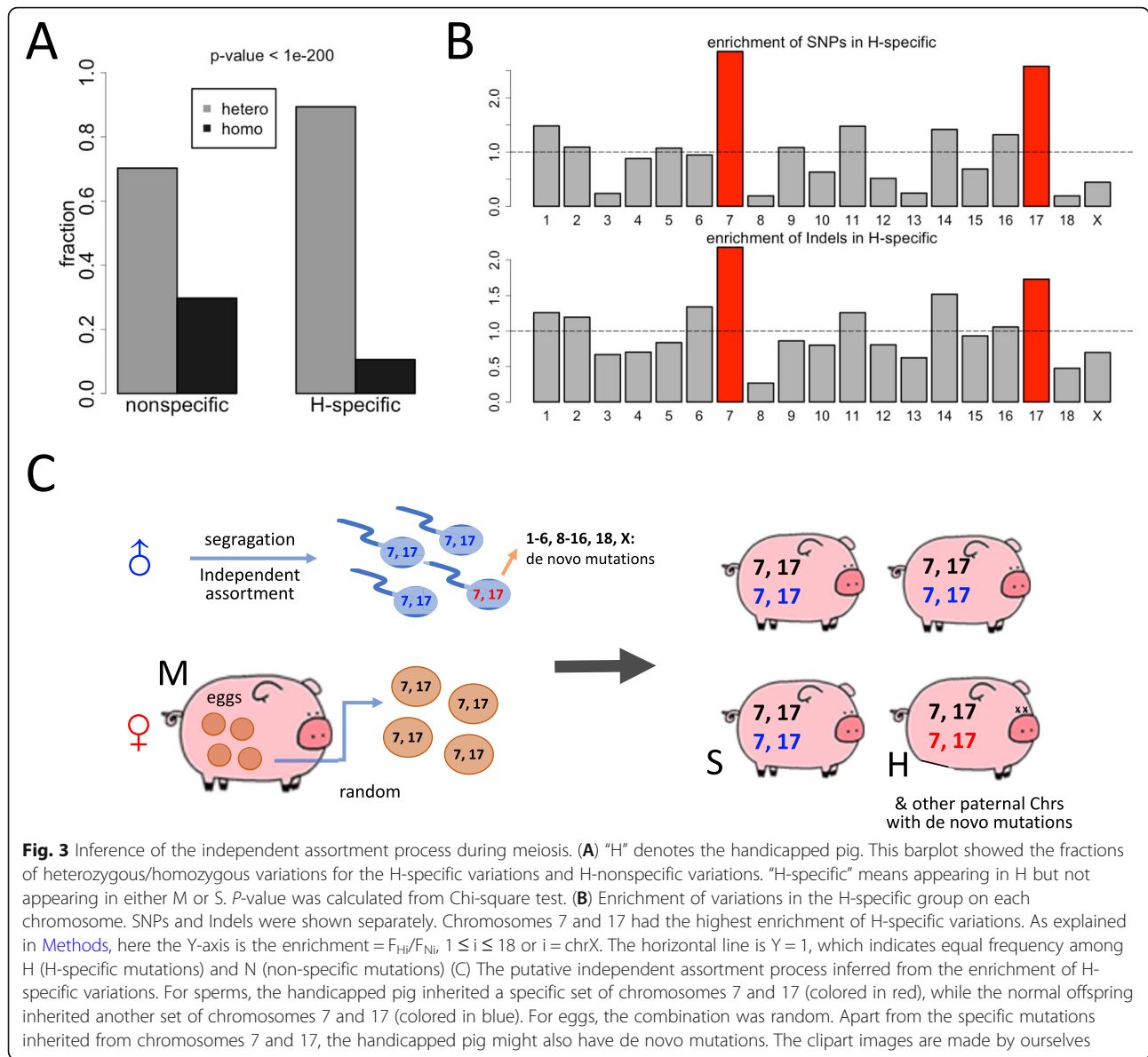
and accordingly, in our data, Indels are more likely to be heterozygous than SNPs (Fig. 2C). This result is expected and conform with known theories.

**Profiling the variations specific to the handicapped pig**

To find putative variations accountable for the handicapped phenotype, we should not focus on the variant sites shared by M, S, and H. Those sites are fixed in this pedigree and should not cause handicap in only one

individual. Instead, the variant sites specific to the handicapped pig are more likely to explain this unique phenotype. We found that 8.7% (1,079,634 out of 12,355,632) variations in H belong to H-specific category (Additional file 3: Table S1) and the remaining 91.3% is shared with at least one of other two sequenced individuals. Interestingly, H-specific variations had a significantly higher fraction of heterozygous sites (89.4%) compared to non-specific variations (70.2%) (Fig. 3A).





Again, this is plausible as the non-specific variations are shared in this pedigree so that they might already be fixed during evolution.

According to the Mendelian segregation and independent assortment laws, each F1 individual has a 0.5 chance to inherit a particular paternal chromosome. In our case, the H-specific variations could only be acquired from the independent assortment of paternal alleles or de novo germline mutations. The number of de novo mutations (under a rate of  $\sim 10^{-8}$  per nucleotide per generation, in mammals) [1] was negligible compared to the specific mutations contained in a particular paternal allele. For example, the de novo mutations carried by each individual is usually less than 100 (under a rate of  $\sim 10^{-8}$ ), while the paternal allele versus maternal

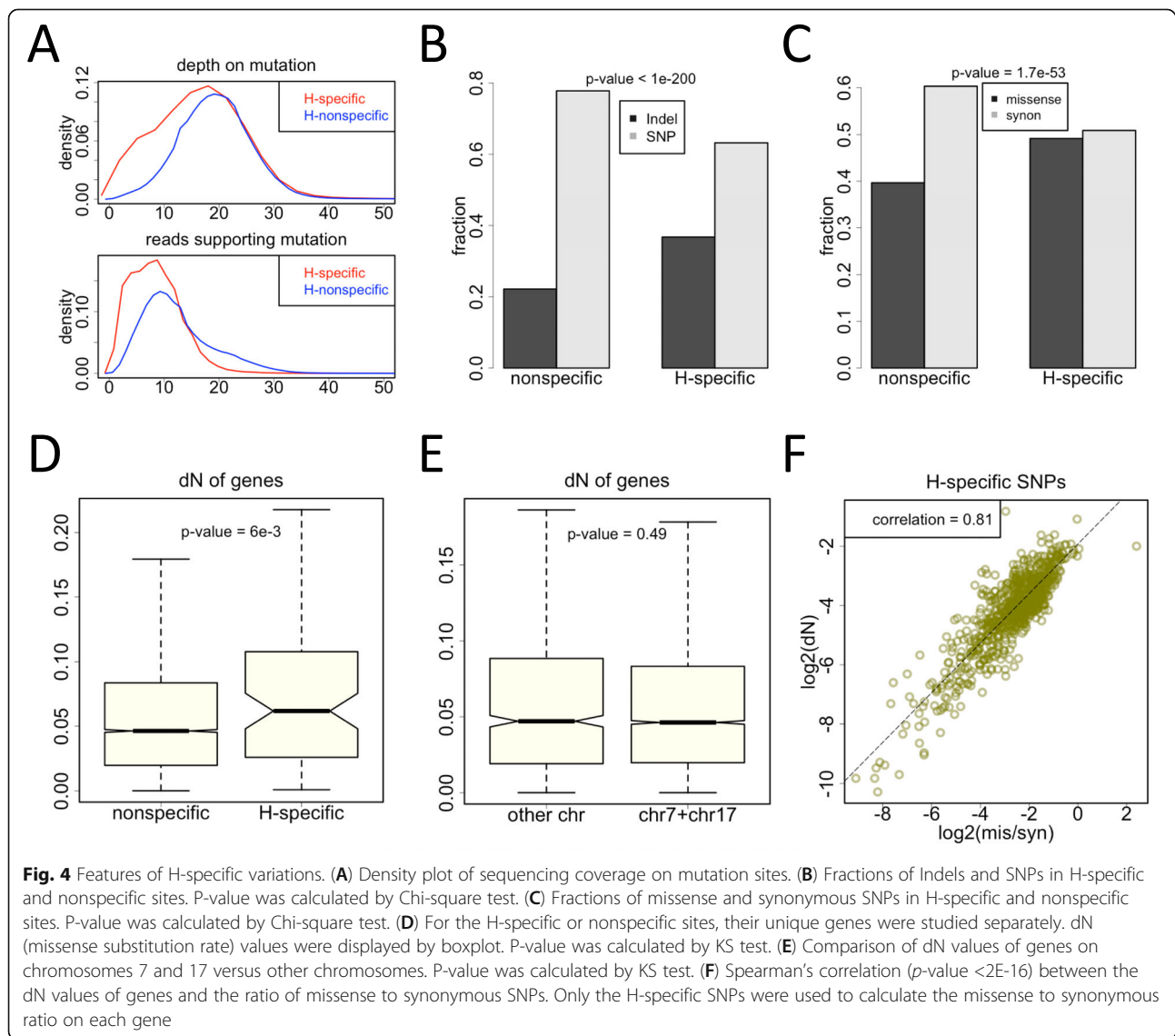
allele could have millions of nucleotides to be different, just like the millions of different sites between two randomly picked human individuals. Therefore, the density profile of H-specific variations could tell us which paternal chromosomes were uniquely transmitted to this handicapped individual. We calculated the frequency of variations appearing in each chromosome, and compared the frequencies of H-specific and non-specific sites (see Methods). We found that the H-specific variations are significantly enriched in chromosome 7 and chromosome 17 (Fig. 3B). This is a strong indication that individual H inherited a set of chromosomes 7 and 17 from paternal alleles while the other sister analyzed inherited another set of chromosomes 7 and 17 from the father (Fig. 3C). Indeed, de novo mutations were also contained

in the variation profile, however, as we have stated, the number of de novo mutations was too small to affect the global distribution of variations. Therefore, there is high confidence to claim that the chromosomes 7 and 17 of this handicapped pig were “differentially” inherited from the paternal side due to independent assortment.

**Evolutionary analysis on the H-specific variations**

To have a better knowledge of the functional impact of the H-specific variations, we intend to analyze these mutations in the light of evolution [23]. First, we show that these mutation sites are covered by adequate reads and that the numbers of reads supporting the variation (alternative allele) are sufficiently high (Fig. 4A). This ensures that the variations especially the H-specific ones were not produced from artefacts. The mean depth of mutation sites is ~ 20 and the mean reads supporting

the mutated allele is ~ 10. Given a sequencing error rate of 1%, the observed mutation is not likely to be the artefact of sequencing error (binomial test  $p$ -value  $< 1E-20$  under depth = 20 and mutated allele = 10). Therefore, the observed variants should be genuine mutation sites. Next, we found that the H-specific variations have higher fractions of Indels compared to non-specific variations (Fig. 4B). Indels are generally more deleterious than SNPs, and therefore the Indels specific to H might play a role in shaping the handicapped phenotype. Similarly, among the SNPs in coding regions, the H-specific SNPs have higher fraction of missense mutations compared to non-specific SNPs (Fig. 4C). In coding regions, missense mutations have stronger functional consequence than synonymous mutations due to the change in amino acids. Again, the higher fraction of missense SNPs for the H-specific mutations indicates that these



impactful variations might contribute to the handicapped phenotype.

We next focus on the conservation level of each gene. The dN (nonsynonymous substitution rate) values of genes are used to measure the divergence between the pig and ox orthologous genes (see [Methods](#)). Lower dN value represents higher conservation level for a gene. If a gene only has H-specific variations but not the non-specific variations, then this gene is defined as H-specific gene (Additional file 4: Table S2), and the other genes with variations are defined as non-specific genes. We found that H-specific genes are significantly less conserved than the non-specific genes (Fig. 4D). This difference in conservation level is not caused by the genes on chromosomes 7 and 17 (Fig. 4E). Intuitively, the genes with the highest conservation level are usually those constitutionally expressed genes with great importance, and mutations on these most essential genes might be lethal and eliminated by natural selection. Now that we observed the H-specific variations in this viable (but with lower fitness) individual, suggesting that these H-specific variations must not be so deleterious to cause a lethal phenotype. This assumption well confirms our observation that the H-specific genes are less conserved (Fig. 4D).

Nevertheless, we still found an interesting pattern for the H-specific SNPs. We focused on the genes with both missense and synonymous SNPs (H-specific ones), and calculated the missense to synonymous ratio (denoted as “mis/syn”) for each gene. Strikingly, the mis/syn ratio of genes (the number of missense mutations in a gene divided by the number of synonymous mutations in this gene) is significantly correlated with the conservation level of genes (Fig. 4F). This is to say, conserved genes tend to have fewer H-specific missense mutations (compared to the synonymous mutations, which are regarded as neutral). As we mentioned above, conserved genes are susceptible to the deleterious missense mutations, so these missense mutations are suppressed in these genes. From another aspect, these observed H-specific missense mutations should be paid special attention to. These H-specific mutations are non-lethal so that they escaped the strongest purifying selection and are maintained to the present, but they might be associated with the limbless phenotype (which is of lower fitness) observed in the handicapped pig.

#### Putative candidate genes related to limb development

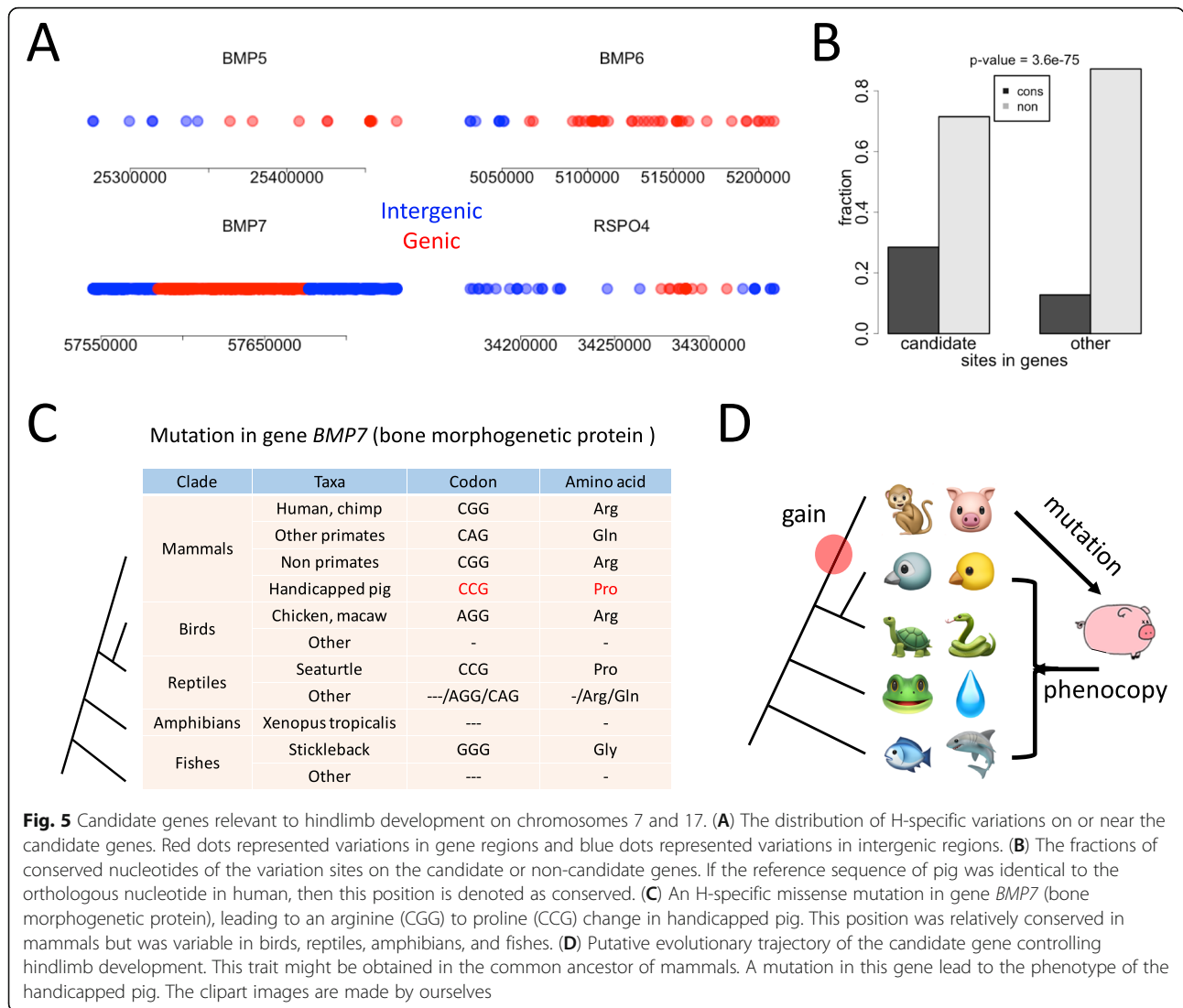
The H-specific variations are composed of two parts. Most of the H-specific variations are supposed to be inherited from particular paternal alleles, and only a minor fraction of H-specific variations belongs to de novo mutations that are randomly dispersed across the genome. Here we would mainly discuss the candidate genes (the genes previously described to be involved in

limb development) which are located on chromosomes 7 and 17. These genes contribute most to the H-specific variations. Some variations that could not be explained by independent assortment are attributed to the possible de novo mutations.

Among the various hindlimb-related genes we introduced, *BMP5*, *BMP6* and *FOXF2* are located in chromosome 7, *BMP2*, *BMP7* and *RSPO4* are located in chromosome 17. Note that the most “promising” gene *TBX4* is located in chromosome 12, and very few H-specific variations are found in *TBX4*. Thus, we regard the six genes (*BMP2*, *BMP5*, *BMP6*, *BMP7*, *FOXF2*, and *RSPO4*) as candidate genes. We show that many of the variations related to candidate genes are intergenic (Fig. 5A), and for the genic variations, most of them are intronic or located in UTRs. Our assumption is, if these six candidate genes really play essential roles in driving the limbless phenotype, then they might exhibit different mutation profiles (differences in the fractions of conserved sites) when compared to other non-candidate genes. This is intuitive because the essential residues are usually conserved and constrained (maintained) by purifying selection. If the essential sites are mutated, the consequence might be more severe than the mutations on non-essential sites. Since we want to attribute the handicapped phenotype to the particular genotype, we should find any possible patterns on these H-specific sites and candidate genes.

We first compared the six candidate genes with other genes on chromosome 7 and 17. The mutation sites could be classified into two categories regarding whether the genomic nucleotide is conserved between pig and human, that is, the conserved group and non-conserved group. By this way, we looked at the fractions of conserved nucleotides. We found that variations in candidate genes have a higher fraction of conserved nucleotides than the variations in non-candidate genes (Fig. 5B). This result is meaningful because the mutations on more conserved sites are more likely to cause deleterious effects.

However, although the mutations on candidate genes exhibit higher conservation level, we only found missense mutations in gene *BMP7*. There are two H-specific missense mutations in gene *BMP7*. One caused an arginine to proline change at protein position 73 (Arg73Pro, chr17:57676401, 18 reads covered, 10 of which supported the mutated allele), another caused an alanine to threonine change at protein position 148 (Ala148Thr, chr17:57676177, 28 reads covered, 11 of which supported the mutated allele). From the reads covering the sites, it is easy to tell that these H-specific SNPs are heterozygous: given a sequencing error rate of 1%, the observed mutation is not likely to be the artefact of sequencing error (binomial test  $p$ -value  $< 1E-20$  under



**Fig. 5** Candidate genes relevant to hindlimb development on chromosomes 7 and 17. **(A)** The distribution of H-specific variations on or near the candidate genes. Red dots represented variations in gene regions and blue dots represented variations in intergenic regions. **(B)** The fractions of conserved nucleotides of the variation sites on the candidate or non-candidate genes. If the reference sequence of pig was identical to the orthologous nucleotide in human, then this position is denoted as conserved. **(C)** An H-specific missense mutation in gene *BMP7* (bone morphogenetic protein), leading to an arginine (CGG) to proline (CCG) change in handicapped pig. This position was relatively conserved in mammals but was variable in birds, reptiles, amphibians, and fishes. **(D)** Putative evolutionary trajectory of the candidate gene controlling hindlimb development. This trait might be obtained in the common ancestor of mammals. A mutation in this gene lead to the phenotype of the handicapped pig. The clipart images are made by ourselves

depth = 20 and mutated allele = 10). Therefore, it should be a reliable heterogenic SNP. We then looked at the genome alignment in vertebrate species, and found an intriguing pattern on site Arg73Pro (CGG to CCG) (Fig. 5C). In most mammals including human, chimpanzee, and non-primates (including pig), the reference genome is CGG that encodes Arg. This site is mutated to CCG and encodes Pro in the handicapped pig. In contrast, this site is highly variable in birds, reptiles, amphibians, and fishes (Fig. 5C). This site could be summarized/described as “conserved in mammals but not conserved in lower vertebrates”. Specially, seaturtle has the same sequence as the handicapped pig, which is a CCG codon encoding Pro. It is plausible that the development of normal mammalian hindlimbs was gained in the common ancestor of mammals but reversed in the handicapped pig, leading to the phenotype mimicking some of the lower vertebrates (Fig. 5D).

### Discussion

During long-term evolution, highly deleterious mutations have been eliminated and could no longer be seen at present time [25]. Beneficial mutations increase their frequency in natural populations [26] but the net gain of fitness is so subtle to be observed at phenotypical level [27]. This dilemma adds difficulty to the functional verification of mutations and makes the evolutionary theories hard to guide experimental biologists [28]. Our work proposes putative solution to this dilemma by obtaining the genetic information from (otherwise) extinct nodes in the tree of life, providing precious information (or direct observation) of the fitness cost of the deleterious mutations.

In most phylogenetic trees, the topology is determined by the sequences of the extant species at the end of each branch. The ancestral nodes have already gone extinct so that the information is missing. However, the genome

sequences of extinct species, if available, will be extremely helpful in the inference of adaptation and function, and this information could even calibrate the phylogeny. For instance, the sequences of extinct Neanderthals (*Homo neanderthalensis*) and Denisova help researchers define the mutations related to cognition that might cause the extinction of the ancient human species [29]. Intriguingly, the extinction process not only took place in ancient times, it takes place all the time.

One could regard this handicapped pig as an extinct individual and the living pigs as extant individuals. Our study is to perform comparative genomics between the sequences of an extinct individual and extant individuals, which is similar to the comparison between *H. sapiens* and *Homo neanderthalensis* or *Homo erectus*. We found H-specific missense mutations that might be accountable for the disadvantageous phenotype. Here, we should admit that we have presumed that the synonymous mutations are neutral, as many studies did [30, 31]. However, this might not always be the case since the synonymous codon usage is also subjected to weak selection as numerous papers have already demonstrated in all kingdoms of organisms [32–43].

Another fortune of this study is that we found particular candidate genes and sites that potentially related to the limbless phenotype, such as the Arg73Pro site on gene *BMP7*. Among the numerous H-specific mutations, Arg73Pro site is one of limited mutations in coding region and caused amino acid change. More importantly, the sequence on this site exhibits clear clade-specific pattern during vertebrate evolution (relatively conserved in mammals but variable in lower vertebrates). This made it the best candidate to account for the limbless phenotype. We also found other candidate mutations (not missense mutations) on the same chromosome (chr17) like those in genes *BMP2* and *RSPO4*. These mutations are provided together with the H-specific variations in Additional file 3: Table S1. These mutations, although not missense mutations, could also be relevant. But the detailed mechanism is uninvestigated at this stage, possibly interacting with other regulators and epigenetic or post-transcriptional pathways [44–48].

Interestingly, while genes *BMP2*, *BMP7*, and *RSPO4* are located in chromosome 17 in pig, these three genes are also located in the same chromosome (chr20) in human. That means that the linkage of these three genes is highly conserved during evolution. Since genes linked within the same chromosome are more likely to have interaction and co-regulation, the linkage of *BMP2*, *BMP7*, and *RSPO4* in mammals might be associated with the epistasis and regulatory network among these genes. During evolution, the disruption of this linkage might break the regulatory chain in bone and limb development and therefore is purged by purifying selection.

In addition, there is a minor but interesting point that might explain the importance of *BMP7* in mammals. The phylogenetic tree of gene *BMP7* in vertebrates shows an extremely low evolution rate in mammals compared with other vertebrate clades like fish, amphibian, reptile, and bird (Fig. 6). This indicates that *BMP7* is highly conserved in mammals but less conserved in lower vertebrates. The selection constraint on mammalian *BMP7* makes the missense mutations in *BMP7* deleterious.

We admit that the deformed phenotype is not necessarily caused by mutations. On one hand, we are fully aware that the phenotype is caused by the combinatory effect of genotype and environment. On the other hand, we do not wish to overstate the rarity of this handicapped pig as we know that these kinds of animals are frequently seen in worldwide farms and that pig is already used as a model animal. However, in our study, since all pigs are subjected to the same environmental conditions, it is hard for us to tell how this particular pig could be the only handicapped one among all the piglets. Therefore, we could only surmise this handicapped pig might have inherited specific alleles from her parents (due to independent assortment) or have suffered from de novo germline mutations. This is the most intuitive and plausible explanation. In addition, we found relevant mutations with evolutionary significance by whole genome sequencing. Our efforts should of great biological significance.

## Conclusion

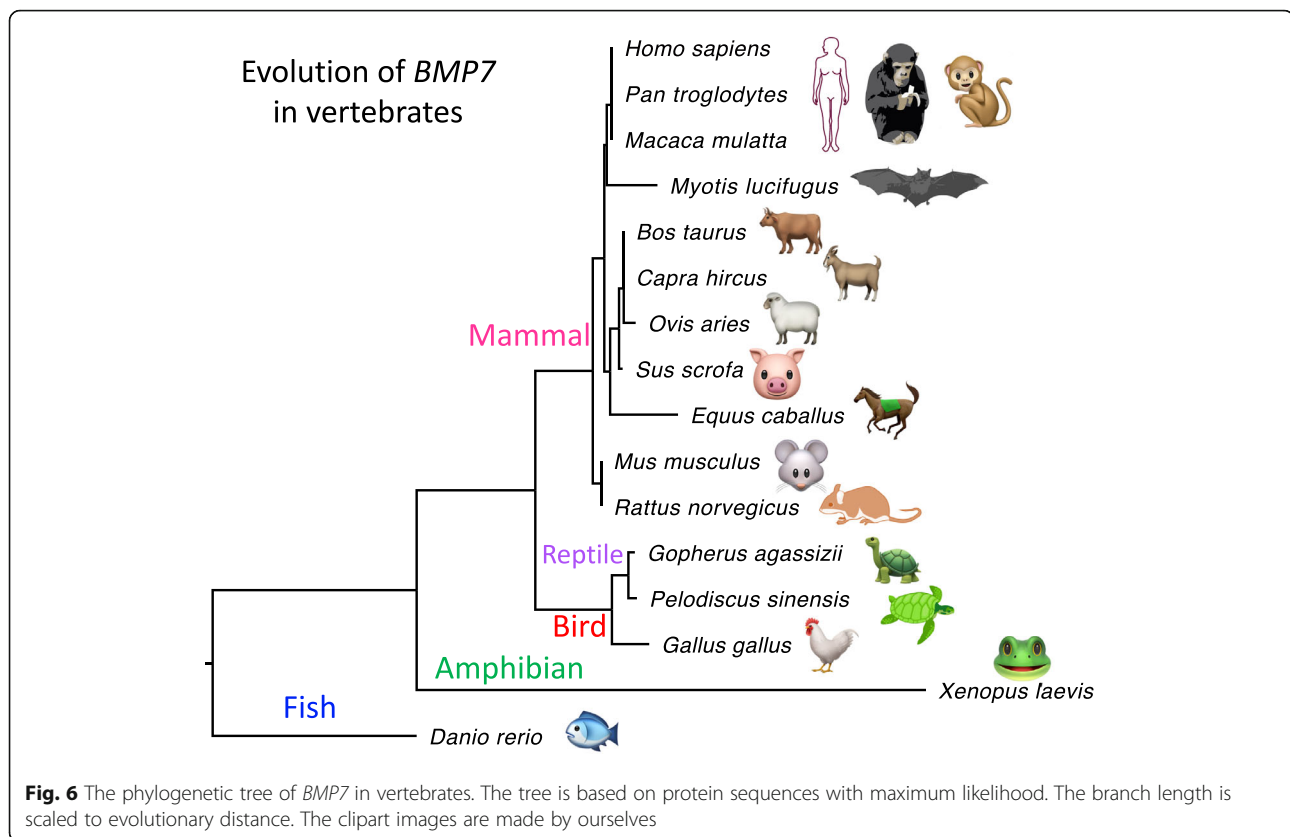
We parsed the limbless-related mutations in the light of evolution. The limbless pig shows phenocopy of the clades before legs were evolved. Our findings might help deduce the emergence of limbs during vertebrate evolution and should be appealing to the broad community of human genetics and evolutionary biology.

## Methods

### Sample collection

The pigs raised in a farm in Yantai are subjected to intensive farming. The feeding trough is filled with feed all the time. Feed would be automatically added to the trough when it goes empty. The new-born pigs are vaccinated with foot-and-mouth disease (FMD) vaccine. Usually, one female pig would deliver five to ten offspring one time. The female F0 pig was artificially fertilized. Four female piglets were born on 30th June, one of which did not have hindlimbs. We did not perform a necropsy. As we have surmised, this pig has deficiency in movement and has disadvantage in competing for food. The lower fitness might be the cause of early death. The other pigs were living well.





**Fig. 6** The phylogenetic tree of *BMP7* in vertebrates. The tree is based on protein sequences with maximum likelihood. The branch length is scaled to evolutionary distance. The clipart images are made by ourselves

DNA was extracted from the tail blood. Next generation sequencing was accomplished by “Geneseq” in Nanjing, China. The platform was Illumina Hiseq 2500, 150 bp, pair-ended.

**Reference genome**

The direct links of the database are given as follows.

Pig (*Sus scrofa*) genome sequence: [ftp://ftp.ensembl.org/pub/release-101/fasta/sus\\_scrofa/dna/Sus\\_scrofa.Ssrofa11.1.dna\\_sm.toplevel.fa](ftp://ftp.ensembl.org/pub/release-101/fasta/sus_scrofa/dna/Sus_scrofa.Ssrofa11.1.dna_sm.toplevel.fa).

Pig (*Sus scrofa*) genome annotation file: [ftp://ftp.ensembl.org/pub/release-101/gtf/sus\\_scrofa/Sus\\_scrofa.Ssrofa11.1.101.gtf.gz](ftp://ftp.ensembl.org/pub/release-101/gtf/sus_scrofa/Sus_scrofa.Ssrofa11.1.101.gtf.gz).

Ox (*B. taurus*) genome sequence: [ftp://ftp.ensembl.org/pub/release-101/fasta/bos\\_taurus/dna/Bos\\_taurus.ARS-UCD1.2.dna\\_sm.toplevel.fa.gz](ftp://ftp.ensembl.org/pub/release-101/fasta/bos_taurus/dna/Bos_taurus.ARS-UCD1.2.dna_sm.toplevel.fa.gz).

Ox (*B. taurus*) genome annotation file: [ftp://ftp.ensembl.org/pub/release-101/gtf/bos\\_taurus/Bos\\_taurus.ARS-UCD1.2.101.gtf.gz](ftp://ftp.ensembl.org/pub/release-101/gtf/bos_taurus/Bos_taurus.ARS-UCD1.2.101.gtf.gz).

Human (*H. sapiens*) genome sequence: [ftp://ftp.ensembl.org/pub/release-101/fasta/homo\\_sapiens/dna/Homo\\_sapiens.GRCh38.dna\\_sm.toplevel.fa.gz](ftp://ftp.ensembl.org/pub/release-101/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna_sm.toplevel.fa.gz).

Human (*H. sapiens*) genome annotation file: [ftp://ftp.ensembl.org/pub/release-101/gtf/homo\\_sapiens/Homo\\_sapiens.GRCh38.101.gtf.gz](ftp://ftp.ensembl.org/pub/release-101/gtf/homo_sapiens/Homo_sapiens.GRCh38.101.gtf.gz).

**Mapping and variant calling**

Reads mapping was accomplished by software BWA mem [49] (version 0.7.17-r1188). Unique mappers were maintained. Variant calling was done by GATK [50] (version 3.7). Reads depth at a given position was called by Samtools [51] (version 1.12). Given a variant in the genome, software SnpEff [52] (version 4.2) was used to annotate the effect of the variant, including whether it is a SNP or Indel, whether it is located in gene region, whether it hit coding region, or whether it changes amino acid. The information of heterozygosity was provided in the variant calling result. For bi-allelic sites, “0/1” represents heterozygous variants and “1/1” represents homozygous sites. The reads count of two alleles could also be extracted from the output file of variant calling. We further require a mutation site to have reads coverage > 10 and alternative reads count > 3. Under these criteria, the variations obtained were not likely to be artefacts like sequencing errors (*p*-value < 1e-10).

**Conservation and divergence**

A tool named liftOver [53] was able to convert the orthologous sites across species. For example, if species-A is the anchor species of interest (pig), by converting the variant coordinate in species-A to the genomic coordinate in species-B, one could know the orthologous nucleotide in species-B by using Bedtools [54]. The

conservation pattern between species-A and species-B could be told from the nucleotides on orthologous sites.

Next, the divergence between two species is measured by dN and dS, named missense and synonymous substitution rates. Each pair of orthologous gene (between two species) has one set of dN and dS values. The exact values were calculated by PAML [5]. Usually, conserved and essential genes have lower dN and dS values.

The phylogenetic tree of BMP7 is based on protein sequences from UniProt website [55]. Sequences are aligned [56–58] and constructed with maximum likelihood [59].

### Enrichment of H-specific mutations

Let “H” denote H-specific mutations and “N” denote non-specific mutations.

Let “F” represent the frequency, which is calculated as  $F_{xi} = x_i / (x_1 + x_2 + \dots + x_n)$ , where  $1 \leq i \leq n$ .

For simplicity, here we presume there are 18 chromosomes numbered 1 ~ 18:

For H-specific mutations,  $F_{Hi} = H_i / (H_1 + \dots + H_{18})$ ,  $1 \leq i \leq 18$ .

For non-specific mutations,  $F_{Ni} = N_i / (N_1 + \dots + N_{18})$ ,  $1 \leq i \leq 18$ .

The enrichment of chromosome  $i = F_{Hi} / F_{Ni}$ ,  $1 \leq i \leq 18$ .

The significance is calculated by Chi-square test:

Let “H” denote H-specific mutations and “N” denote non-specific mutations. For example, to test the significance of chromosome 7, we need four numbers:  $H_7$ ,  $N_7$ ,  $H_{[other\_chr]}$ , and  $N_{[other\_chr]}$ . The p-value was calculated by Chi-square test using these four numbers. P-value < 0.05 was regarded as significant.

### Data deposition

The DNA-sequencing data were deposited to the Genome Sequence Archive (GSA) under accession number PRJCA003678.

### Abbreviations

BMP: bone morphogenetic protein; *PTX1*: Paired Like Homeodomain 1; *RSPO*: R-Spondin; *FOXF*: Forkhead Box F; *HOX*: Homeobox; AF: allele frequency; SNP: single nucleotide polymorphism; dN: nonsynonymous substitution rate; M: mother; S: sister; H: handicapped pig

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07925-3>.

**Additional file 1: Fig. S1.** X-ray (side view and vertical view) of the handicapped pig (14 days, July 13th, 2020). Human hands were inevitably included in this graph.

**Additional file 2.** Video showing the clumsiness of the handicapped pig.

**Additional file 3: Table S1.** List of H-specific variations.

**Additional file 4: Table S2.** List of H-specific genes.

### Acknowledgements

We thank all people who have given help or suggestions to this project. We thank the farm owner that allowed us to sequence the animals. We thank

our institutional ethics committee in The Second Hospital of Shandong University to approve us to conduct the research.

### Authors' contributions

YJ and XC analyzed the data and drafted this manuscript. HW supervised this work and revised the manuscript. All authors including YJ, XC, and HW approved the final submission of this manuscript.

### Funding

This work was not supported by funding.

### Availability of data and materials

Pig (*Sus scrofa*) genome sequence: [ftp://ftp.ensembl.org/pub/release-101/fasta/sus\\_scrofa/dna/Sus\\_scrofa.Sscrofa11.1.dna\\_sm.toplevel.fa](ftp://ftp.ensembl.org/pub/release-101/fasta/sus_scrofa/dna/Sus_scrofa.Sscrofa11.1.dna_sm.toplevel.fa).

Pig (*Sus scrofa*) genome annotation file: [ftp://ftp.ensembl.org/pub/release-101/gtf/sus\\_scrofa/Sus\\_scrofa.Sscrofa11.1.101.gtf.gz](ftp://ftp.ensembl.org/pub/release-101/gtf/sus_scrofa/Sus_scrofa.Sscrofa11.1.101.gtf.gz).

Ox (*B. taurus*) genome sequence: [ftp://ftp.ensembl.org/pub/release-101/fasta/bos\\_taurus/dna/Bos\\_taurus.ARS-UCD1.2.dna\\_sm.toplevel.fa.gz](ftp://ftp.ensembl.org/pub/release-101/fasta/bos_taurus/dna/Bos_taurus.ARS-UCD1.2.dna_sm.toplevel.fa.gz).

Ox (*B. taurus*) genome annotation file: [ftp://ftp.ensembl.org/pub/release-101/gtf/bos\\_taurus/Bos\\_taurus.ARS-UCD1.2.101.gtf.gz](ftp://ftp.ensembl.org/pub/release-101/gtf/bos_taurus/Bos_taurus.ARS-UCD1.2.101.gtf.gz).

Human (*H. sapiens*) genome sequence: [ftp://ftp.ensembl.org/pub/release-101/fasta/homo\\_sapiens/dna/Homo\\_sapiens.GRCh38.dna\\_sm.toplevel.fa.gz](ftp://ftp.ensembl.org/pub/release-101/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna_sm.toplevel.fa.gz).

Human (*H. sapiens*) genome annotation file: [ftp://ftp.ensembl.org/pub/release-101/gtf/homo\\_sapiens/Homo\\_sapiens.GRCh38.101.gtf.gz](ftp://ftp.ensembl.org/pub/release-101/gtf/homo_sapiens/Homo_sapiens.GRCh38.101.gtf.gz).

Genome transfer tools: <http://hgdownload.soe.ucsc.edu/goldenPath/SusScr11/liftOver/>.

The DNA-sequencing data were deposited to Genome Sequence Archive (GSA) under accession number PRJCA003678.

### Declarations

#### Ethics approval and consent to participate

The pigs used in this study were raised in a farm in Yantai area, China. We obtained written informed consent to use the animals in our study from the owner of the animals. Our study was approved by our institutional ethics committee in The Second Hospital of Shandong University.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Joint Surgery, The Second Hospital of Shandong University, Jinan 250033, Shandong, China. <sup>2</sup>School of Medicine and Nursing, Dezhou University, Dezhou 253023, Shandong, China.

Received: 30 October 2020 Accepted: 8 August 2021

Published online: 28 August 2021

### References

- Lynch M. Evolution of the mutation rate. *Trends Genet.* 2010;26(8):345–52.
- Pan Z, Zhang X, Ma Y, Xu S, Shuai Z, Pan F, et al. Genetic variation of rs7958311 in P2X7R gene is associated with the susceptibility and disease activity of ankylosing spondylitis. *Postgrad Med J.* 2019;95(1123):251–7.
- Kumar N, Gross JB Jr. Mutation in the ATP7A gene may not be responsible for hypocupraemia in copper deficiency myelopathy. *Postgrad Med J.* 2006; 82(968):416.
- Kimura M: The neutral theory of molecular evolution. *Sci Am* 1979, 241(5): 98–100, 102, 108 passim.
- Xu B, Yang Z. PAMLX: a graphical user interface for PAML. *Mol Biol Evol.* 2013;30(12):2723–4.
- Yu YY, Li Y, Dong Y, Wang XK, Li CX, Jiang WQ. Natural selection on synonymous mutations in SARS-CoV-2 and the impact on estimating divergence time. *Future Virol.* 2021;16(7):447–50.
- Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KM, et al. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell.* 2016;166(2):481–91.
- Wei L. Selection on synonymous mutations revealed by 1135 genomes of *Arabidopsis thaliana*. *Evol Bioinformatics Online.* 2020;16:1176934320916794.

9. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*. 1991;351(6328):652–4.
10. Li Y, Yang XN, Wang N, Wang HY, Yin B, Yang XP, et al. The divergence between SARS-CoV-2 and RaTG13 might be overestimated due to the extensive RNA modification. *Future Virol*. 2020;15(6):341–7.
11. Naiche LA, Papaioannou VE. Loss of Tbx4 blocks hindlimb development and affects vascularization and fusion of the allantois. *Development*. 2003;130(12):2681–93.
12. Kariminejad A, Szenker-Ravi E, Lekszas C, Tajsharghi H, Moslemi AR, Naert T, et al. Homozygous null TBX4 mutations Lead to posterior Amelia with pelvic and pulmonary hypoplasia. *Am J Hum Genet*. 2019;105(6):1294–301.
13. Lawson KA, Dunn NR, Roelen BA, Zeinstra LM, Davis AM, Wright CV, et al. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev*. 1999;13(4):424–36.
14. Ying Y, Liu XM, Marble A, Lawson KA, Zhao GQ. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol Endocrinol*. 2000;14(7):1053–63.
15. Brunmeir R, Wu J, Peng X, Kim SY, Julien SG, Zhang Q, et al. Comparative Transcriptomic and Epigenomic analyses reveal new regulators of murine Brown Adipogenesis. *PLoS Genet*. 2016;12(12):e1006474.
16. Wozney JM, Rosen V, Celeste AJ, Mitscock LM, Whitters MJ, Kriz RW, et al. Novel regulators of bone formation: molecular clones and activities. *Sci*. 1988;242(4885):1528–34.
17. Tremblay KD, Dunn NR, Robertson EJ. Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. *Development*. 2001;128(18):3609–21.
18. Lechleider RJ, Ryan JL, Garrett L, Eng C, Deng C, Wynshaw-Boris A, et al. Targeted mutagenesis of Smad1 reveals an essential role in chorioallantoic fusion. *Dev Biol*. 2001;240(1):157–67.
19. Lanctot C, Moreau A, Chamberland M, Tremblay ML, Drouin J. Hindlimb patterning and mandible development require the Ptx1 gene. *Development*. 1999;126(9):1805–10.
20. Szenker-Ravi E, Altunoglu U, Leushacke M, Bosso-Lefevre C, Khatoo M, Thi Tran H, et al. RSPO2 inhibition of RNF43 and ZNRF3 governs limb development independently of LGR4/5/6. *Nature*. 2018;557(7706):564–9.
21. Mahlapuu M, Ormestad M, Enerback S, Carlsson P. The forkhead transcription factor Foxf1 is required for differentiation of extra-embryonic and lateral plate mesoderm. *Development*. 2001;128(2):155–66.
22. Meyer A. Hox gene variation and evolution. *Nature*. 1998;391(6664):225–228.
23. Dobzhansky T. Nothing in biology makes sense except in light of evolution. *Am Biol Teach*. 1973;35(3):125–9.
24. Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, et al. Progressive loss of function in a limb enhancer during Snake evolution. *Cell*. 2016;167(3):633–42 e611.
25. Chang S, Li J, Li Q, Yu CP, Xie LL, Wang S. Retrieving the deleterious mutations before extinction: genome-wide comparison of shared derived mutations in liver cancer and normal population. *Postgrad Med J*. 2021. <https://doi.org/10.1136/postgradmedj-2021-139993>.
26. Wang N, Wang D. Genome-wide transcriptome and translome analyses reveal the role of protein extension and domestication in liver cancer oncogenesis. *Mol Gen Genomics*. 2021;296(3):561–9.
27. Zhao S, Song S, Qi Q, Lei W. Cost-efficiency tradeoff is optimized in various cancer types revealed by genome-wide analysis. *Mol Gen Genomics*. 2021;296(2):369–78.
28. Li Y, Yang XN, Wang N, Wang HY, Yin B, Yang XP, et al. Pros and cons of the application of evolutionary theories to the evolution of SARS-CoV-2. *Future Virol*. 2020;15(6):369–72.
29. Paixao-Cortes VR, Viscardi LH, Salzano FM, Hunemeier T, Bortolini MC. *Homo sapiens*, *Homo neanderthalensis* and the Denisova specimen: New insights on their evolutionary histories using whole-genome comparisons. *Genet Mol Biol*. 2012;35(4 (suppl)):904–11.
30. Chu D, Wei L. The chloroplast and mitochondrial C-to-U RNA editing in *Arabidopsis thaliana* shows signals of adaptation. *Plant Direct*. 2019;3(9):e00169.
31. Liu Z, Zhang J. Human C-to-U coding RNA editing is largely nonadaptive. *Mol Biol Evol*. 2018;35(4):963–9.
32. Chu D, Wei L. Characterizing the heat response of *Arabidopsis thaliana* from the perspective of codon usage bias and translational regulation. *J Plant Physiol*. 2019;240:153012.
33. Chu D, Wei L. Context-dependent and -independent selection on synonymous mutations revealed by 1,135 genomes of *Arabidopsis thaliana*. *BMC Ecol Evol*. 2021;21(1):68.
34. Bennetzen JL, Hall BD. Codon selection in yeast. *J Biol Chem*. 1982;257(6):3026–31.
35. Li Y, Yang X, Wang N, Wang H, Yin B, Yang X, et al. GC usage of SARS-CoV-2 genes might adapt to the environment of human lung expressed genes. *Mol Gen Genomics*. 2020;295(6):1537–46.
36. Chu D, Wei L. Parsing the synonymous mutations in the maize genome: isoaccepting mutations are more advantageous in regions with codon co-occurrence bias. *BMC Plant Biol* 2019;19:422. <https://doi.org/10.1186/s12870-019-2050-1>.
37. Chu D, Wei L. Genome-wide analysis on the maize genome reveals weak selection on synonymous mutations. *BMC Genomics*. 2020;21(1):333.
38. Hershberg R, Petrov DA. General rules for optimal codon choice. *PLoS Genet*. 2009;5(7):e1000556.
39. Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet*. 2012;8(3):e1002603.
40. Baker SF, Nogales A, Martinez-Sobrido L. Downregulating viral gene expression: codon usage bias manipulation for the generation of novel influenza A virus vaccines. *Future Virol*. 2015;10(6):715–30.
41. Taghinezhad S, Razavilar V, Keyvani H, Razavi MR, Nejadstattari T. Codon optimization of Iranian human papillomavirus type 16 E6 oncogene for *Lactococcus lactis* subsp *cremoris* MG1363. *Future Virol*. 2017;12(9):499–511.
42. Li Q, Li J, Yu CP, Chang S, Xie LL, Wang S. Synonymous mutations that regulate translation speed might play a non-negligible role in liver cancer development. *BMC Cancer*. 2021;21(1):388.
43. Arella D, Dilucca M, Giansanti A. Codon usage bias and environmental adaptation in microbial organisms. *Mol Gen Genomics*. 2021;296(3):751–62.
44. Deffit SN, Hundley HA. To edit or not to edit: regulation of ADAR editing specificity and efficiency. *Wiley Interdiscip Rev RNA*. 2016;7(1):113–27.
45. Song B, Shiromoto Y, Minakuchi M, Nishikura K. The role of RNA editing enzyme ADAR1 in human disease. *Wiley Interdiscip Rev RNA* 2021:e1665.
46. Li Y, Yang X, Wang N, Wang H, Yin B, Yang X, et al. Mutation profile of over 4500 SARS-CoV-2 isolations reveals prevalent cytosine-to-uridine deamination on viral RNAs. *Future Microbiol*. 2020;15:1343–52.
47. Fung TS, Liu DX. Post-translational modifications of coronavirus proteins: roles and function. *Future Virol*. 2018;13(6):405–30.
48. Chu D, Wei L. Reduced C-to-U RNA editing rates might play a regulatory role in stress response of *Arabidopsis*. *J Plant Physiol*. 2020;244.
49. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
50. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
51. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
52. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012, 6(2):80–92.
53. Gao B, Huang Q, Baudis M. segment\_liftover : a Python tool to convert segments between genome assemblies. *F1000Res*. 2018;7:319.
54. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
55. UniProt C. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43(Database issue):D204–12.
56. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
57. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35(6):1547–9.
58. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007;24(8):1596–9.
59. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74.

## Publisher's Note

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