Balancing selection at a premature stop mutation in the *myostatin* gene underlies a recessive leg weakness syndrome in pigs

Oswald Matika*, Diego Robledo, Ricardo Pong-Wong, Stephen C. Bishop†, Valentina Riggio, Heather Finlayson, Natalie R. Lowe, Annabelle E. Hoste, Grant A. Walling, Jorge del Pozo, Alan L. Archibald, John A. Woolliams, Ross D. Houston

1 The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, United Kingdom, 2 JSR Genetics, Southburn, Driffield, East Yorkshire, United Kingdom

* These authors contributed equally to this work.
† Deceased.
* oswald.matika@roslin.ed.ac.uk

Abstract

Balancing selection provides a plausible explanation for the maintenance of deleterious alleles at moderate frequency in livestock, including lethal recessives exhibiting heterozygous advantage in carriers. In the current study, a leg weakness syndrome causing mortality of piglets in a commercial line showed monogenic recessive inheritance, and a region on chromosome 15 associated with the syndrome was identified by homozygosity mapping. Whole genome resequencing of cases and controls identified a mutation causing a premature stop codon within exon 3 of the porcine *Myostatin* (*MSTN*) gene, similar to those causing a double-musc cling phenotype observed in several mammalian species. The *MSTN* mutation was in Hardy-Weinberg equilibrium in the population at birth, but significantly distorted amongst animals still in the herd at 110 kg, due to an absence of homozygous mutant genotypes. In heterozygous form, the *MSTN* mutation was associated with a major increase in muscle depth and decrease in fat depth, suggesting that the deleterious allele was maintained at moderate frequency due to heterozygous advantage (allele frequency, \(q = 0.22\)). Knockout of the porcine *MSTN* by gene editing has previously been linked to problems of low piglet survival and lameness. This *MSTN* mutation is an example of putative balancing selection in livestock, providing a plausible explanation for the lack of disrupting *MSTN* mutations in pigs despite many generations of selection for lean growth.

Author summary

Lameness is an important problem in livestock production for both animal welfare and economic reasons. A severe piglet lameness syndrome was observed in a commercial pig population. The incidence of the condition was low (6.3%), but was higher in affected families (~25%), which suggested a genetic basis and a recessive mode of inheritance. We discovered a region on Chromosome 15 where cases shared the same alleles that were
different to healthy piglets. In this region, we discovered a mutation that causes a premature stop codon in the myostatin gene. Myostatin causes ‘double-muscle’ phenotype in several mammalian species. Piglets with two copies of this mutant allele suffer the lameness syndrome and do not survive post 40 kg live weight. However, those that carry a single copy have higher muscle depth and lower fat depth compared to wild type. We suggest that despite the negative consequences of the mutant allele in homozygous form, the mutation was maintained in the herd due to positive selection for this allele in heterozygous form. This is an interesting example of so-called ‘balancing selection’ and may explain why naturally occurring myostatin mutations have not previously been reported in pigs despite centuries of selection for lean growth.

Introduction

Leg weakness is a heterogeneous condition causing lameness in pigs, and has negative impacts on both animal welfare and productivity [1, 2]. Significant heritability estimates have been reported for leg weakness traits [reviewed in 3], with moderate to high estimates in certain pig breeds, e.g. $h^2 = 0.45$ in Landrace [1]. Several quantitative trait loci (QTL) have been identified for these traits, albeit they are generally not consistent across studies and breeds [5–8], which may be partly due to the heterogeneity of this condition. Interestingly, significant genetic correlations between leg weakness and other production traits (such as growth and muscle depth) have been detected [4]. Further, in a divergent selection experiment in Duroc lines, selection for high leg weakness was associated with a significant increase in muscle length and weight [9]. Taken together, these results suggest a degree of antagonistic genetic relationship between leg weakness and muscle growth traits in pigs, potentially explaining increases in the syndrome observed with intense selection for lean growth in recent decades.

Deleterious alleles can be maintained at relatively high frequency in commercial livestock populations due to heterozygous advantage for traits under selection [10]. Examples of such balancing selection in cattle include a frame-shift mutation in the mannose receptor C type 2 gene (MRC2) responsible for crooked tail syndrome and also associated with increased muscle mass in Belgian Blue [11], and a large deletion with antagonistic effects on fertility and milk production traits in Nordic Red breeds [12]. In pigs, balancing selection at c.C1843T mutation in the ryanodine receptor 1 (RYR1) gene [13, 14] is likely to have caused an increase in incidence of porcine stress syndrome (also known as malignant hyperthermia) in the 1970s and 1980s, due to the association of the causative missense mutation with reduced backfat—a trait under selection. The ryanodine receptor 1 protein RYR1 acts as a calcium release channel in skeletal muscle. The c.C1843T mutation interferes with the proper function of the calcium release channel rendering homozygotes susceptible to stress induced malignant hyperthermia and death. Involuntary muscle contraction by the leaky channel may contribute to reduced fat levels in carriers and homozygotes [15]. More recently, a study showed an unique example of allelic pleiotropy in which one allele (deletion) is responsible for both increased growth and late foetal mortality by affecting two different genes [16].

We investigated genetic parameters and mode of inheritance for a leg weakness syndrome causing piglet mortality in a commercial Large White terminal sire line historically marketed by JSR Genetics under the product name “Yorker”. A monogenic recessive inheritance was observed, and homozygosity mapping was used to identify a genomic region on Sus scrofa chromosome (SSC) 15 associated with the trait. A mutation causing a premature stop codon in exon 3 of the Myostatin (MSTN) gene (similar to mutations causing the ‘double-muscling’
phenotype in cattle [17] was the outstanding functional candidate in the region. Comparison of MSTN genotype frequencies at birth and 110 kg supported this hypothesis, and carriers were shown to have significantly higher muscle mass and reduced fat depth than wild type homozygotes. Therefore, we propose that the MSTN mutant allele is highly deleterious in this population in homozygous form, but was maintained at moderate frequency due to heterozygous advantage.

Results

The piglet leg weakness trait shows a recessive mode of inheritance

The overall prevalence of leg weakness in the commercial cohort was 6.3% (Table 1). When only affected litters were considered, the mean proportion of affected piglets was 23% ± 0.7. This within-litter prevalence is consistent with the expectation under the hypothesis of a single recessive locus (i.e. 25%). Complex Bayesian segregation analysis [18] suggested that almost all the variation was explained by a single locus with almost no environmental variation. The estimate of the additive effect was 0.50 ± 0.001 and dominance effect was -0.50 ± 0.001, which is in precise agreement with a recessive locus model. Estimates of heritability for the leg weakness syndrome (analysed as a binary trait on the underlying liability scale) was high (0.57 ± 0.10 in the sire and dam model) with low (0.17 ± 0.02 and 0.11 ± 0.02) but significant effects observed for permanent environmental effects due to the dam and litter, respectively (Table 2, S1 Table).

Causative variant maps to chromosome 15

Homozygosity mapping was used to map the underlying recessive variant, and the longest shared homozygous segment was a region of ~ 8.3 Mbp on SSC15. This unique region, contained 55 informative single nucleotide polymorphisms (SNPs) on the Illumina PorcineSNP60 SNP chip [19]. These 55 SNPs were homozygous in affected animals, but contained SNPs that were heterozygous or homozygous for the alternative allele in the unaffected animals, albeit there was one unaffected animal also sharing the homozygous segment. The segment started with SNP ALGA0110636 (rs81338938) at position 86,745,668 in the current Sscrofa11.1 reference genome assembly (Genbank assembly accession GCA_000003025.6) and finished with SNP H3GA0044732 (rs80936849) at position 95,062,143 (Fig 1A and 1B). The MSTN gene was located within this homozygous segment, from position 94,620,269–94,628,630. The 8.3 Mbp segment was assumed to represent an identical-by-descent (IBD) region likely to contain the

Table 1. Number of animals in pedigree and records used for variance components analyses.

<table>
<thead>
<tr>
<th>Description</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of records</td>
<td>19,006</td>
</tr>
<tr>
<td>Number of pedigree records</td>
<td>27,501</td>
</tr>
<tr>
<td>Number of generations</td>
<td>9</td>
</tr>
<tr>
<td>Number of litter</td>
<td>1,903</td>
</tr>
<tr>
<td>Number of Sires</td>
<td>346</td>
</tr>
<tr>
<td>Number of Sires of Sire</td>
<td>175</td>
</tr>
<tr>
<td>Number of Dams of Sire</td>
<td>239</td>
</tr>
<tr>
<td>Number of Dams</td>
<td>1,929</td>
</tr>
<tr>
<td>Number of Sires of Dam</td>
<td>240</td>
</tr>
<tr>
<td>Number of Dams of Dam</td>
<td>882</td>
</tr>
<tr>
<td>Leg weakness prevalence %</td>
<td>6.3</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pgen.1007759.t001
underlying causative mutation, and became the focus of further analyses to discover and characterise this mutation. The SNP genotype used in the homozygosity mapping at SSC15 is given in S3 Table.

Sequence analysis reveals MSTN mutation as causative candidate

To identify candidates for the causative variant, whole genome sequence data from ten cases, six presumed heterozygous carrier dams, and 22 controls were analysed. A total of 40 SNPs identified within the homozygous segment fitted the pattern of a potential causative variant assuming a recessive mode of inheritance. Functional annotation of these SNPs revealed that 19 were intergenic, 19 were intronic, 1 was in a pseudogene, and 1 caused a premature stop codon. There were also 10 InDels identified, 3 of which were intergenic and 7 of which were intronic (S4 Table). The outstanding functional candidate was a mutation in the third exon of the MSTN locus that results in the replacement of a codon for glutamic acid with a stop codon in exon 3 at position 274 (c.820G>T; p.E274X) (Fig 2A). The mutation is located in a region that is highly conserved across multiple species, and is predicted to result in truncation of the protein (Fig 2B). Functional annotation of all other variants detected in the IBD region did not reveal any other obvious causative candidates. This stop gain mutation was not present on the Ensembl variation database, accessed 25th July 2018. The whole genome sequencing raw reads are available in the NCBI, BioProject accession PRJNA506339.

Changes in mutant allele frequency from piglets to adults

The MSTN c.820G>T mutation showed no statistically significant deviation from Hardy Weinberg equilibrium (HWE) in the 486 piglets sampled at birth (q = 0.22, α = 0.019, χ² = 0.18, P > 0.05). Random mating of the dioecious population would be expected to result in a value of α that is slightly negative [20], and the value observed does not differ significantly from this value. However, the mutation deviated significantly from HWE at 40 kg (q = 0.17 α = -0.180, χ² = 12.2, P > 0.001) and at 110 kg (q = 0.17 α = -0.210, χ² = 11.45, P > 0.001). This was due to the loss of homozygous mutant piglets, with all but one dying (or being euthanized) shortly after birth, and the remaining piglet being euthanized due to poor health before it reached 110 kg. For detailed information, see S2 Table. The large change in α in a negative

Table 2. Genetic parameter estimates and standard errors (in parentheses) for the trait of leg weakness on the liability scale using the logit transformation showing the outcomes of fitting sire or dam models with or without maternal environment (σ²w). All models have litter variance (σ²v) fitted.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sire</th>
<th>Sire and Dam</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>σ²w &gt; 0</td>
<td>σ²w &gt; 0</td>
</tr>
<tr>
<td>σ²s</td>
<td>1.069 (0.283)</td>
<td>1.100 (0.290)</td>
</tr>
<tr>
<td>σ²d</td>
<td>0.664 (0.231)</td>
<td>1.110 (0.186)</td>
</tr>
<tr>
<td>σ²v</td>
<td>0.681 (0.123)</td>
<td>0.681 (0.123)</td>
</tr>
<tr>
<td>σ²p</td>
<td>0.999 (0.148)</td>
<td>0.437 (0.203)</td>
</tr>
<tr>
<td>σ²w</td>
<td>6.049 (0.307)</td>
<td>6.182 (0.322)</td>
</tr>
<tr>
<td>h²u</td>
<td>0.71 (0.16)</td>
<td>0.57 (0.10)</td>
</tr>
<tr>
<td>σ²w/σ²p</td>
<td>0.11 (0.02)</td>
<td>0.11 (0.02)</td>
</tr>
<tr>
<td>σ²v/σ²p</td>
<td>0.17 (0.2)</td>
<td>0.07 (0.03)</td>
</tr>
</tbody>
</table>

- σ²s, σ²d, σ²v, and σ²w are variances due to the sire and dam genetic effects, litter and maternal environment effects. σ²p is phenotypic variance. h²u is estimated heritability on observed scale, and σ²w/σ²p and σ²v/σ²p are the proportion of phenotypic variance explained by the litter and maternal environment effects respectively.

https://doi.org/10.1371/journal.pgen.1007759.002

https://doi.org/10.1371/journal.pgen.1007759.t002

Balancing selection at MSTN locus in pigs

PLOS Genetics | https://doi.org/10.1371/journal.pgen.1007759 January 30, 2019 4 / 15
A:

B:
direction is quantitative evidence of the selective disappearance of homozygote genotypes, as opposed to disappearance as a result of selection against the allele itself. There were no significant changes in the relative genotype frequencies (GG, GT) over the total period or any sub-period from birth to the end of the test, confirming all changes in \( q \) and \( a \) are due to the selective loss of homozygotes, and that any other mortality or culling was at random with respect to MSTN genotype.

**Association of the MSTN mutation with performance traits**

The association of the porcine MSTN c.820G>T mutation with performance traits was assessed on 384 pigs which had completed a commercial performance test. Given the loss of the homozygous mutant animals the effect of the MSTN c.820G>T mutation was only estimated by the difference between the heterozygotes and the wild type pigs. The genotype means and differences are shown in Table 3, with the most notable of these being a major increase in muscle depth and a reduction in fat depth in the carriers (\( p < 0.001 \)), with no evidence of a difference in live weight at 110 kg. Approximately 31% of the genetic variation in muscle depth and 18% of the genetic variation in fat depth was explained by this single variant (Table 3). The heterozygous animals had on average 5 mm increased muscle depth, and 1.7 mm decreased backfat depth when compared with wild type homozygous animals.

**Histological comparison of genotypes**

Histological comparison of biceps femoris muscle between a piglet carrying the TT genotype (putative double MSTN knockout) and a heterozygous piglet revealed had significantly larger myofibre sizes in the homozygote (13.4 \( \mu m^2 \) vs 11.5 \( \mu m^2 \), \( p < 0.001 \)), suggesting comparative myofibre hypertrophy. However, there was no difference in myofibre number between the two animals, which does not suggest hyperplasia (S1 Fig).

**Discussion**

A piglet leg weakness syndrome identified in a commercial line of Large White pigs showed moderate to high heritability, consistent with the upper range of estimates reported in the literature [2, 21, 22]. The within-litter incidence of the syndrome and the complex Bayesian segregation analysis both pointed to a monogenic recessive condition. Homozygosity mapping revealed a 8.3 Mbp segment on SSC15 as a putative IBD region likely to contain the causative variant. A single control animal appeared to also carry the putative IBD segment in this region, although it is possible that this animal was an affected animal misclassified as a control. The leg weakness syndrome is not lethal *per se*, but was debilitating under farm conditions and affected animals were typically crushed or failed to suckle effectively and were therefore typically euthanized. The outstanding functional candidate variant identified by whole genome
**A:**

**Myostatin** Chromosome 15: 94,620,269-94,628,630

SNP - 94,623,834

---

**B:**

---

Fig 2. A: Position of the premature stop causing mutation within the porcine myostatin locus; B: Conservation of the amino acid sequence surrounding the mutation, with the consequences of the premature stop mutation highlighted in red.

https://doi.org/10.1371/journal.pgen.1007759.g002
sequencing of cases and controls caused a premature stop codon within exon 3 of the MSTN gene (Fig 2), and this variant was in HWE at birth but significantly distorted by 110 kg due to an absence of homozygous mutant animals. Knockout of MSTN by gene editing in previous studies has been associated with poor health and mortality of knockout piglets [23–26], including observations of a piglet leg weakness syndrome with an inability to stand or walk [24], strikingly similar to the syndrome described in the current study.

MSTN is a member of the transforming growth factor beta (TGF-β) superfamily, which is highly conserved across species, and is typically expressed in developing and mature skeletal muscle as a key regulator of muscle growth [27]. The MSTN gene has been a gene of interest to animal breeders for over twenty years since the discovery of loss-of-function mutations in the cattle MSTN gene, which cause muscle hypertrophy leading to double muscling phenotypes [17, 27, 28]. Interestingly, these loss-of-function mutations causing double muscling are frequently due to premature stop mutations in the highly conserved exon 3 of MSTN [28], (Fig 2B). Further, in the case of Marchigiana beef cattle, the MSTN variant causing double muscling results is due to replacement of Glutamic Acid with a stop codon (the same change observed in the current study) in the equivalent position in the protein as observed in the current study [29]. This observation provides additional indirect evidence that the porcine MSTN mutation described herein is likely to cause the increase in muscle depth and decrease in fat depth associated with carrier animals. A histological analysis of muscle showed larger myofibre sizes in a homozygous TT (putative double MSTN knockout) than a heterozygous piglet suggesting comparative myofibre hypertrophy. A similar phenotype was observed in MSTN gene knockout piglets from a genome editing study [24]. However, unlike in another study of MSTN knockout piglets [25], where a greater number of myofibre nuclei were observed in mutants compared with the WT group, the density of myofibres in our study was comparable between the two animals which does not suggest hyperplasia (S1 Fig).

Table 3. Estimates and statistical significance of the effect of the MSTN c.820G>T locus on the growth and carcass traits of pigs obtained from a commercial performance test. Estimates are shown in absolute units and standardised by phenotypic standard deviations (σp). Standard errors are in parentheses. The traits are categorised into: live weights and live weight gain; muscle and fat depths measured by ultrasound at the end of the test either conditional on age or on live weight; and periods to achieve growth targets.

<table>
<thead>
<tr>
<th>Trait</th>
<th>MSTN c.820G&gt;T genotypes</th>
<th>σp</th>
<th>TG-GG</th>
<th>σp</th>
<th>Wald test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Animal Numbers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 151</td>
<td>GG 227</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Live weights:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At start (kg)</td>
<td>36.83 (0.84)</td>
<td>-1.36 (0.44)</td>
<td>3.81 (0.18)</td>
<td>-0.36</td>
<td>P&lt;0.01</td>
</tr>
<tr>
<td>At end (kg)</td>
<td>84.89 (1.59)</td>
<td>-2.24 (0.94)</td>
<td>7.96 (0.33)</td>
<td>-0.28</td>
<td>P&lt;0.05</td>
</tr>
<tr>
<td>Daily gain (kg/d)</td>
<td>1.02 (0.02)</td>
<td>0.01 (0.01)</td>
<td>0.11 (0.01)</td>
<td>-0.09</td>
<td>NS</td>
</tr>
<tr>
<td>Depths for age:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fat depth (mm)</td>
<td>8.42 (0.62)</td>
<td>-1.76 (0.31)</td>
<td>2.72 (0.13)</td>
<td>-0.65</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Muscle depth (mm)</td>
<td>53.21 (1.11)</td>
<td>4.83 (0.68)</td>
<td>5.95 (0.25)</td>
<td>0.81</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Depths for weight:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fat depth (mm)</td>
<td>9.25 (0.57)</td>
<td>-1.49 (0.28)</td>
<td>2.49 (0.12)</td>
<td>-0.60</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Muscle depth (mm)</td>
<td>54.89 (1.05)</td>
<td>5.40 (0.63)</td>
<td>5.54 (0.24)</td>
<td>0.97</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Periods:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To 40kg (d)</td>
<td>91.66 (2.12)</td>
<td>3.04 (1.01)</td>
<td>9.01 (0.45)</td>
<td>0.36</td>
<td>P&lt;0.01</td>
</tr>
<tr>
<td>From 40 to 100kg (d)</td>
<td>66.33 (1.67)</td>
<td>-0.29 (1.81)</td>
<td>6.75 (0.36)</td>
<td>-0.04</td>
<td>NS</td>
</tr>
</tbody>
</table>

*The most salient results presented in the manuscript

https://doi.org/10.1371/journal.pgen.1007759.t003
In livestock breeding schemes, MSTN-inactivating mutations have been retained due to selection for increased lean growth associated with meat production, particularly for double-muscled cattle. However, despite many generations of selection for lean growth in pigs no such mutations have been reported previously. Associations between polymorphisms within the porcine MSTN gene and production traits have been shown in small-scale studies [30–32]. Further, a genome-wide association study in a sire line Large White population (related to the animals in the current study) detected SNPs on SSC15 significantly associated with rib fat between 81.1 and 87.8 Mbp [33]. While these SNPs are between 7 and 13 Mbp closer to the centromere than MSTN, two of these SNPs overlap with the region on homozygosity indicative of an IBD region in the current study. Therefore, it is plausible that this association may be due to linkage disequilibrium (LD) with the MSTN variant, or that other variants impacting performance traits exist on SSC15. The selection index applied in this line explicitly benefited animals with positive muscle depth and negative fat depth estimated breeding values, with analysis of the index showing that heterozygotes had a slight but significant advantage ($p < 0.05$). Therefore, it is likely that the MSTN c.820G>T mutation was maintained at moderate frequency in this line despite its deleterious impact on piglet mortality due to heterozygous advantage for muscle and fat traits. Similar examples of balancing selection have observed to explain the maintenance of deleterious alleles at moderate frequency in commercial cattle [10, 12] and pig [16] populations. The results described herein have major implications for the targeted ablation of MSTN via gene editing to increase lean growth in pigs, and provide a plausible explanation of why MSTN loss-of-function mutations have not been previously reported in pigs despite decades of selection for lean growth.

Materials and methods

Ethics statement

All samples were collected on a commercial nucleus farm as part of standard husbandry and management procedures in the nucleus herd, which complied with conventional UK red tractor farm assurance standards (https://assurance.redtractor.org.uk/) where sick or injured livestock that do not respond to treatment are promptly and humanely euthanized by a trained and competent stockperson.

Animals

A piglet leg weakness syndrome was characterised in a Large White terminal sire line, historically marketed by JSR Genetics under the product name “Yorker”, reared in a nucleus herd, under standard conditions but with additional data recording. Leg weakness trait observations were collected on 19,006 piglets from 2007 to 2010, during which time a high incidence of the syndrome was detected. In addition, DNA was sampled on a further 119 piglets in 2011 and 486 piglets from the same population in 2012, of these 384 also had weight and carcass phenotypes available. The cohort born in 2007 to 2010 will be referred to as the commercial cohort and those in 2011 and 2012 as the survey cohort. Pedigree was available for all animals and spanned 9 generations. Details of the data and pedigree structure are presented in Table 1.

The leg weakness in the phenotyped animals was visually classified as normal or affected (0 / 1 respectively). The leg defect is characterised by the piglet not being able to straighten its legs to stand, this being most apparent for the front legs, and being slow to suckle. Detailed post mortems were conducted on two affected individuals but the results were ineffective in providing additional diagnostic aids. The Online Mendelian Inheritance in Animals database (OMIA: http://omia.org/OMIA000585/9823/) was searched for previous reports of leg weakness in pigs but the syndrome observed here did not appear. These problems frequently
resulted in death from either starvation or being crushed by the sow. Where the outcome resulted in poor welfare the piglets were euthanized in accordance with the Ethics Statement. Additional farrowing data were collected on the females in the commercial cohort including numbers born alive, dead or mummified, parity and year of birth. Body weights, ultrasound muscle and fat depths were obtained for individuals in the survey cohort that were retained in the herd until commercial slaughter age (when average weight is 110 kg). The ultrasonic measurements taken were the average depth of the *m. longissimus dorsi* and overlaying subcutaneous fat layer across the last four ribs.

**Statistical analyses of genetic parameters**

Initial studies to establish the genetic basis of the leg weakness syndrome were undertaken by fitting linear mixed models to the commercial cohort. The binary record of the syndrome was modelled on the observed 0/1 scale and the full model fitted was:

\[
y = X\beta + Z_u\mu + Z_v\nu + Z_w\omega + e
\]

where \( y \) is the vector of leg weakness phenotypes; \( \beta \), a vector of fixed effects for month of observation (43 df), sex (1 df), parity (7 df), numbers in litter born dead (1 df) or alive (1 df), with design matrix \( X \); \( u \), additive polygenic effects, assumed to be distributed MVN(0, \( \sigma_u^2 \)) with design matrix \( Z_u \); \( v \), random litter effects assumed to be distributed MVN(0, \( \sigma_v^2 \)), with design matrix \( Z_v \); \( w \), maternal environment effects across litters assumed to be distributed MVN(0, \( \sigma_w^2 \)) with design matrix \( Z_w \); and \( e \) residuals assumed to be distributed MVN(0, \( \sigma_e^2 \)). Variations in this model were fitted replacing the individual polygenic effects with sire and/or dam effects, assumed to be distributed MVN(0, \( \sigma_s^2 \)) and MVN(0, \( \sigma_d^2 \)) respectively. All models were fitted using the ASREML software [34]. Likelihood ratio tests were used to assess the random effects. In addition, an analogous threshold model with an underlying continuous liability was fitted with a logit link function and sire and dam effects associated with the pedigree, but not with individual polygenic effects, following recommendation of Gilmour et al. [34]. For the full model, the phenotypic variance was calculated as \( \sigma_p^2 = \sigma_u^2 + \sigma_v^2 + \sigma_w^2 + \sigma_e^2 \). Where sire and dam models were used, \( \sigma_u^2 \) was replaced by \( \sigma_s^2 + \sigma_d^2 \). Heritability (\( h^2 \)) was calculated as \( \sigma_u^2/\sigma_p^2 \) or \( 2(\sigma_s^2 + \sigma_d^2)/\sigma_p^2 \) depending on the model. The proportion of variance explained by the litter and maternal environmental effects were estimated as \( \sigma_v^2/\sigma_p^2 \) and \( \sigma_w^2/\sigma_p^2 \) respectively. Heritabilities on the observed scale (0/1) was transformed to an underlying liability scale following Dempster and Lerner [35] using the observed prevalence of the syndrome in the commercial population which was 6.3%.

Inspection of the data suggested that the syndrome may be due to a single gene with the predisposing deleterious allele showing a recessive mode of inheritance, and this hypothesis was tested using chi square tests and segregation analyses. An initial test of a monogenic recessive mode of inheritance was carried out by pooling all affected litters, estimating the probability of being affected conditional on being born in an affected litter, and using chi-squared to test the null hypothesis that the probability of being affected was 0.25. A weakness of this approach is that some litters by chance will have no affected offspring, so a more complex segregation model was fitted. This model included all known phenotypes and pedigree data, and assumed a monogenic inheritance with environmental variation fitted by Gibbs sampling [18].

**Homozygosity mapping of the recessive mutation**

Ten affected animals from different litters and 10 unaffected full-sib controls from the commercial cohort were genotyped using the Illumina PorcineSNP60 SNP chip [19]. Only those SNPs that mapped to known positions on autosomal chromosomes and were not fixed nor
completely heterozygous were retained. This left 38,570 segregating autosomal SNPs for the use in homozygosity mapping. Homozygous regions were assessed by alignment with the Sscrofa11.1 reference genome assembly sequence (Genbank assembly accession GCA_000003025.6).

**Whole genome resequencing**

The genomes of the ten cases used for homozygosity mapping and six separate dams with affected offspring, assumed heterozygotes, were whole genome shotgun sequenced on an Illumina HiSeq 2500 platform as 125 bp paired-end reads. The dams were individually sequenced with a 10x genome coverage. The piglets were barcoded and individually sequenced at 3x coverage to achieve 30x coverage for the pool. The full sequencing output resulted in ~1.3 billion paired-end reads with an average of 48 million paired-end reads/sample for the piglets and 157 million paired-end reads/sample for the dams. Quality filtering and removal of residual adapter sequences was conducted on read pairs using Trimmomatic v.0.32 [36]. Only reads where both pairs had a length greater than 32 bp post-filtering were retained, leaving a total of ~1.2 bn paired-end reads.

Whole genome resequencing was followed by alignment to the Sscrofa11.1 assembly; using the Burrows-Wheeler Aligner with default parameters [37]. The average alignment rate of properly paired reads was of 92%. PCR duplicates were marked using Picard Tools (http://broadinstitute.github.io/picard). Variant calling was performed using the Genome Analysis Toolkit (GATK) HaplotypeCaller after read recalibration [38]. The parameter setting for the hard filters that were applied to the raw genotypes were: QualByDepth < 2.0, FisherStrand > 60.0, RMSMappingQuality < 40.0, MappingQualitySumTest < -12.5, ReadPosRankSumTest < -8.0.

Candidate loci were identified from the sequences of the 16 animals from this study plus 22 additional Sus scrofa control sequences obtained from a public database [39], comprising 7 domesticated breeds (Duroc, Hampshire, Jiangquhai, Landrace, Large White, Meishan and Pietrain) and wild boar. The following criteria were used to identify candidate SNPs:

i. Homozygous for the same allele in all the affected piglets

ii. Heterozygous in parents of affected piglets (i.e. putative carriers)

iii. Homozygous for the alternative allele (i.e. the one not observed in the affected offspring) in the control (unaffected) animals

It is worth noting that the limited sequencing depth means that both alleles will not be detected for all bases in all individuals. This limitation is particularly relevant to the reliable detection of heterozygous SNPs.

**Genotyping**

A ‘kompetitive allele specific PCR’ (KASP) assay was designed by LGC Genomics (Teddington, UK) to enable genotyping of the mutation in the MSTN stop codon in large numbers of animals. The survey cohort of 486 piglets sampled at birth were genotyped by LGC. Of these, 265 remained as candidates for the final selection at 110 kg with a complete record of their performance test, together with another 119 pigs phenotyped at slaughter age (total n = 374). In both the 486 piglets and the subsequent subsets surviving to 40 and 110 kg, the frequency of the MSTN mutation (q) was calculated by counting, and the departure from Hardy Weinberg equilibrium the genotypes was estimated as $\alpha = 1 - H_{\text{obs}}/H_{\text{exp}}$ [20] where $H_{\text{obs}}$ is the observed heterozygosity and $H_{\text{exp}}$ is the expected heterozygosity calculated as $2q(1-q)$. The significance
of departure from true random mating genotype frequencies ($\alpha = 0$) was tested using a chi-squared test.

**Association analysis**

Associations between the MSTN c.820G>T locus and variation in performance traits in commercial testing conditions were examined in the survey cohort (n = 384). The performance test was started at a target weights of 40 kg, at an average age of 85 days, and continued for 54 (s.d. 12) days. The performance tests were performed over two distinct periods. The traits available were live weights and ages at the start and the end of the test, ultrasonic muscle and fat depths measured at the end of the test, days from birth to 40 kg and days from 40 to 110 kg. Univariate mixed models were fitted to these data in ASReml-R4 using the following model:

$$y = 1\mu + X_1\beta + X_2b + Zu + e$$

where $y$ is the vector of phenotypes; $\mu$, a fitted mean, and $1$, a vector of 1's; $\beta$, a vector of fixed nuisance effects with design matrix $X_1$; $b$, a scalar fixed effect for the effect of SNP genotype with design matrix $X_2$, this has only 1 df due to the absence of homozygotes completing the test; $u$, additive polygenic effects assumed to be distributed $MVN(0, \sigma_u^2)$, with design matrix $Z$; and $e$, residuals assumed to be distributed $MVN(0, \sigma_e^2)$. For all traits, the sex of the piglet (1 df), parity of dam (4 df) and period of testing (1 df) were fitted as nuisance factors, together with cubic smoothing splines for the start date of the test fitted separately within each period [40]. The age at the time of measurement was fitted as covariate (1 df) for all traits other than days to 40 kg and days from 40 to 110 kg. The significance of fixed effects was assessed using Wald tests [ASReml Manual].

**Histological analysis**

Samples of biceps femoris were taken from one myostatin homozygous knockout piglet (TT) and one heterozygous piglet (TG), and fixed in 10% neutral buffered formalin. After 24h of fixation, the tissues were processed to paraffin embedded blocks, and 4 $\mu$m sections were taken and stained with haematoxylin & eosin. One Tiff image was taken from each animal in an area where myofibres had been sectioned transversally using a Zeiss Axiocam 105 colour camera using the same light settings, and measurements were calibrated using a microscope graticle. For image analysis, images were loaded in FIJI (https://fiji.sc/), and for each image blood vessels were manually removed. The green channel was chosen, and the threshold adjusted to identify myofibres. This was followed by two binary functions (‘fill holes’ and ‘watershed’ consecutively), and the resulting mask was measured using the particle analyser, filtering particles below 5 $\mu$m² in size. The myofibre size was compared between the two samples using a Mann-Whitney test to assess hypertrophy. For assessment of hyperplasia, the total number of myofibres was expressed as the number of myofibres per 100 $\mu$m².

**Supporting information**

S1 Fig. Results of comparative morphometric analysis of skeletal muscle histology of heterozygous TG (id: S1313) vs. homozygous TT (id: S1322) animal. Panel A illustrates the histological presentation (raw), featuring histologically normal myofibres for both genotypes, and the results of the morphometric measurement on the same area of the sections (post analysis). Panel B, contains a dotplot chart of the comparative measurements for each myofibre taken from an identical area for each animal. The white dot indicates the median value for each group. The homozygous myofibres were significantly larger that the heterozygous (P<0.001), with a median of 13.4 $\mu$m² and 11.5 $\mu$m² respectively. This is suggestive of significant
hypertrophy of myofibres in the homozygous group. Conversely, the density of myofibres is comparable in both groups (i.e. 2.2 myofibres/100 μm² for both groups), suggesting no hyperplasia is present in the homozygous group. Stain: Haematoxylin and Eosin, scale bar = 50 μm (applies to all figures).

(TIF)

S1 Table. Genetic parameter estimates and standard errors for the trait of leg weakness on the observed scale 0/1 showing the outcomes of fitting animal or sire/dam models with or without maternal environment (σ²w). All models have litter variance (σ²v) fitted.

(DOCX)

S2 Table. Summary statistics for the stop codon: number of genotypes, allele frequencies (p,q), Chi-square (χ²) and alpha values (α) of piglets from birth to adults.

(DOCX)

S3 Table. Case/control genotypes for all SNPs on chromosome 15 together with the positions according to the Sscrofa11.1 reference genome assembly. Also included is the longest segment of 55 SNPs identified in the homozygosity mapping analysis.

(XLSX)

S4 Table. SNP and INDEL variants found in the whole-genome sequencing data that are consistent with a monogenic recessive disease. Variants are considered consistent if they are: 1) homozygous for the same allele in all the affected piglets (allowing one heterozygote); 2) heterozygous in parents of affected piglets (i.e. putative carriers, allowing one homozygous for the reference allele); 3) homozygous for the alternative allele (i.e. the one not observed in the affected offspring) in the control (unaffected) animals.

(XLSX)

Author Contributions

Conceptualization: Oswald Matika, Ricardo Pong-Wong, Stephen C. Bishop, Grant A. Walling, Alan L. Archibald, John A. Woolliams, Ross D. Houston.

Data curation: Oswald Matika, Ricardo Pong-Wong, Annabelle E. Hoste.


Funding acquisition: Stephen C. Bishop, John A. Woolliams.


Methodology: Oswald Matika, Stephen C. Bishop, Jorge del Pozo.

Resources: Annabelle E. Hoste, Grant A. Walling.

Software: Ricardo Pong-Wong.

Writing – original draft: Oswald Matika, Ross D. Houston.

Writing – review & editing: Oswald Matika, Diego Robledo, Valentina Riggio, Grant A. Walling, Jorge del Pozo, Alan L. Archibald, John A. Woolliams, Ross D. Houston.

References


