In-silico investigation of porcine X-chromosome reveals deleterious non-synonymous SNPS

1. Utsav Surati, 1, 2Ravina, 1, 2Dharamshaw CA, 1, 2Gaurav Patel, 1, 2Chandana Sree Chinnareddyvari, 1, 2Pallavi Rathi, 1, 2Rangasai Chandra Goli, 1, 2Kiyevi G Chishi, 3Ayushi Singh and 1, 2Kanaka KK

1Animal Genetics and Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, India
2ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana, India
3Animal Genetics and Breeding, Indian Veterinary Research Institute, Izzatnagar Bareilly, Uttar Pradesh, India

Abstract
Deleterious mutations in genomes, particularly those occurring in germ cells, can have heritable consequences. Next Generation Sequencing (NGS) technology has facilitated the discovery and prediction of novel mutations. Computational algorithms and robust bioinformatics pipelines play a crucial role in identifying genetic mutations, especially in scenarios lacking animal records and in vivo testing. In this study, genomic sequence data from twenty Sus scrofa (pig) samples obtained from the European Variation Archive repository was annotated using VEP-Ensembl. A total of 178,745 variants were processed, including 530 novel variants. Subsequently, only sex chromosomes were considered for further analysis to identify sex-linked mutations, resulting in the processing of 2300 variants, including 28 non-synonymous SNPs (nsSNPs). Using various predictive tools such as SIFT, PANTHER, Predict SNP, Polyphen-1, Polyphen-2, and SNAP, four nsSNPs were identified as deleterious, with one nsSNP found deleterious across all tools. Additionally, the I-mutant tool confirmed a decrease in protein stability for all deleterious nsSNPs. This study identified four deleterious mutations (S261F, D664N, A704, and I76N) mapped to two genes (HUWE1 and NKRF) in pigs. One mutation name D664N (HUWE1) was found to be deleterious by all tools. While most deleterious mutations in pigs have been reported in exotic breeds, limited information exists for Indian native pig breeds. The in-silico identification of deleterious SNPs in this study provides a foundation for further investigation into the biological consequences of these mutations in Indian pig breeds through in vivo studies.

Keywords: Next generation sequencing, bioinformatics, Sus scrofa, deleterious mutations, in silico analysis.

1. Introduction
A single nucleotide polymorphism (SNP) is a source variance in a genome. SNPs account for 90% of all genetic variations in the genome and are the most prevalent and basic type of polymorphism. The nonsynonymous SNPs can decrease the protein stability and have an impact on how protein functions, which can lead to deleterious effect [1, 2]. Deleterious nsSNPs in genomes, particularly those occurring in germ cells, can have heritable consequences. Next Generation Sequencing (NGS) technology has facilitated the discovery and prediction of novel mutations. Computational algorithms and robust bioinformatics pipelines play a crucial role in identifying genetic mutations, especially in scenarios lacking animal records and in vivo testing. Utilizing in silico tools for identifying mutational changes represents a novel and increasingly embraced approach, offering enhanced ease and reliability. These computational tools are gaining widespread adoption for pinpointing deleterious mutations across the genome in both animals and humans. They prove particularly valuable in detecting genome-wide deleterious nucleotide mutations, notably missense SNPs, and in predicting their impact on protein structure and function [3]. Previously, in silico analysis was instrumental in identifying deleterious mutations by scrutinizing non-synonymous alterations in key candidate genes such as BRAF, TNF-α, BARD-1, and IGF1R in humans, and SLC11A2 in animals [4-7]. While most deleterious mutations in pigs have been reported in exotic breeds, limited information exists for Indian native pig breeds. The in-silico identification of deleterious SNPs in this study provides a foundation for further investigation into the biological consequences of these mutations in Indian pig breeds through in vivo studies.

2. Materials and Methods
2.1 Data Set
Genomic sequence data from twenty Sus scrofa (pig) samples were obtained from the European Variation Archive repository. The data were retrieved and processed using VEP-Ensembl for annotation, encompassing a total of 178,745 variants, which included 530 novel variants.

2.2 Identification of Deleterious nsSNPs
Deleterious non-synonymous SNPs (nsSNPs) were identified to understand their impact on protein function and
phenotypic changes. The Sorting Intolerant from Tolerant (SIFT) score from the Ensembl database was utilized for predicting the deleteriousness of SNPs. SIFT assigns a tolerance index (TI) score ranging from 0.0 to 1.0, with a TI score of 0.05 or less indicating intolerance or deleteriousness [8, 9]. To validate the identified deleterious SNPs from SIFT, additional predictive tools such as PANTHER, Predict SNP, PolyPhen-1, PolyPhen-2, and SNAP were employed. Results from PolyPhen-1 and SNAP were obtained through the consensus classifier of Predict SNP.

2.3 Prediction of structural and functional effect of nsSNP

For prediction of protein stability change, I-Mutant 2.0 was used. This was a Support Vector Machine (SVM)-based web server for the automatic prediction of protein stability changes upon single-site mutations. The input protein FASTA sequence along with the residue changes were provided. I-Mutant 2.0 predicts free energy change and RI value (reliability index). If the DDG value is negative, then the mutated protein will have less stability and vice versa for high stability [10]. HOPE version 1.1.1 (https://www3.cmbi.umcn.nl/hope/) was used to recognise the structural effect of nonsynonymous change in the ANPEP protein sequence. It also provided 3D structure visualization of altered protein and superimposition of wild and mutant after providing altered protein sequence as input [11].

3. Results and Discussion

A total of 178,745 variants were processed, including 530 novel variants. Subsequently, only sex chromosomes were considered for further analysis to identify sex-linked mutations, resulting in the processing of 2300 variants, including 28 non-synonymous SNPs (nsSNPs) (Table 1). Among them, 45 percent belonged to intronic variation and 30 percent was intergenic variants. The non-coding transcript variant came up to 12 percent of the total variants. The variants in upstream and downstream regions were 5 percent and 4 percent respectively. There were no synonymous variants in the coding region and 1 percent of total variants were missense in nature (Figure 1). Within the coding sequence, 64 percent were missense variants, 18 percent were synonymous. Stop retained variant was 9 percent while start loss variant constituted 2 percent of the total coding sequence variants. A 7 percent frame shift variant was also observed (Figure 2). The missense variants were analysed further.

Table 1: Details of processed variants in sex chromosome of Pig

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variants processed</td>
<td>2300</td>
</tr>
<tr>
<td>Variants filtered out</td>
<td>0</td>
</tr>
<tr>
<td>Novel/ existing variants</td>
<td>2298 (99.9)/2 (0.1)</td>
</tr>
<tr>
<td>Overlapped genes</td>
<td>171</td>
</tr>
<tr>
<td>Overlapped transcripts</td>
<td>379</td>
</tr>
<tr>
<td>Overlapped regulatory features</td>
<td>-</td>
</tr>
</tbody>
</table>

Fig 1: Details of variants in sex chromosome of Pig
3.1 Identification of deleterious nsSNPs in sex chromosome of Pig

Many research efforts have looked at potential genes, including RASSF5, MECP2, CSN3, BRAF, TNF-κ, BARD-1, IGF1R, and SLC11A2, in both humans and animals in order to find harmful SNPs [12-16]. Studies have been carried out in porcine TERT gene [17] and MYH1 gene [18] using in silico tools. About 30% of false negative outcomes and 20% of false positive results are provided by the SIFT [9]. To increase the accuracy of prediction by SIFT, we used other in silico tools. Although all in silico programs have certain limitations, it was imperative to use the combination of the tools to assess the deleteriousness. Reports indicate that these tools may also have a considerable error rate [19]. In this study, PANTHER, Predict SNP, Polyphen-1, Polyphen-2 and SNAP were used to analyse identified missense SNPs for deleterious effects.

A total of 28 non-synonymous variants were identified on sex chromosome (X-chromosome) of pig. Based on SIFT score obtained from VEP, 4 nsSNPs named S261F, D664N, A704T and I760N were found to be deleterious. These four nsSNPs were further analysed and three were found deleterious using PANTHER, one using Predict SNP, four using Polyphen-1, three using Polyphen-2, and one using SNAP for further confirmation. Only one nsSNP was found deleterious in common using all tools (Table 2).

Table 2: Common deleterious nsSNPs in sex chromosome of Pig

<table>
<thead>
<tr>
<th>Position</th>
<th>Gene symbol</th>
<th>Variant name</th>
<th>PANTHER</th>
<th>Predict SNP</th>
<th>Polyphen-1</th>
<th>Polyphen-2</th>
<th>SNAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>X:10435436</td>
<td>-</td>
<td>S261F</td>
<td>Probably damaging</td>
<td>-</td>
<td>Deleterious</td>
<td>Deleterious</td>
<td>Neutral</td>
</tr>
<tr>
<td>X:46366236</td>
<td>HUWE1</td>
<td>D664N</td>
<td>Probably damaging</td>
<td>Deleterious</td>
<td>Deleterious</td>
<td>Deleterious</td>
<td>Deleterious</td>
</tr>
<tr>
<td>X:98011469</td>
<td>NKRF</td>
<td>A704T</td>
<td>possibly damaging</td>
<td>Neutral</td>
<td>Deleterious</td>
<td>-</td>
<td>Neutral</td>
</tr>
<tr>
<td>X:100773108</td>
<td>-</td>
<td>I760N</td>
<td>-</td>
<td>Neutral</td>
<td>Deleterious</td>
<td>Deleterious</td>
<td>Neutral</td>
</tr>
</tbody>
</table>

3.2 Structural and functional effects on protein

The I-mutant 2.0 tool showed a decrease in protein stability in three deleterious nsSNPs, thus adding another layer of confirmation. Furthermore, the one common SNP (D664N) from all the tools was analysed using HOPE server. Differences in the size, affinity towards water, flexibility and preferred secondary structure have been observed between wild and mutant residues which also led to structural and functional effects in the new mutant protein. Project HOPE identified the deleterious effect of point mutation on protein structure. This analysis revealed the effect of D664N (HUWE1) mutation; that is the mutation is damaging to the protein. A difference in charge between the wild-type and mutant amino acid (Table 3) was also identified. The net charge of amino acid sequence has a dynamic effect on interaction between other proteins as well as cell membrane phospholipids [20]. Surface charge can also influence protein localisation based on the motifs present on it [21]. A change in the net charge due to the SNP may deleteriously affect the normal functioning of the coded protein.

HUWE1 is a ubiquitin ligase-containing HECT domain linked to the spermatogenesis, oncogenesis and neurogenesis. Mouse models have demonstrated the critical role HUWE1 gene plays in the N-Myc pathway-mediated regulation of neurogenesis in the cerebral cortex. It is now widely acknowledged that X-linked intellectual impairment (XLID) is connected to HUWE1 mutations or rearrangements [22]. Genetic variants of HUWE1 gene were found to cause severe intellectual disabilities, disfigured...
limbs, absence of speech, dysmorphic features and craniosynostosis in several human patients \cite{23-27}. Mouse embryo study shows that HUWE1 knockdown could lead to increased apoptosis in blastocysts \cite{30}. Thus, a deleterious SNP may lead to cognitive as well as reproductive impairment of the affected pigs.

<table>
<thead>
<tr>
<th>Position</th>
<th>Allele</th>
<th>Gene symbol</th>
<th>Variant name</th>
<th>ddg</th>
<th>1 Mutant 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>X:10435436</td>
<td>A</td>
<td>-</td>
<td>S261F</td>
<td>0.21</td>
<td>Increase Stability</td>
</tr>
<tr>
<td>X:46366236</td>
<td>T</td>
<td>HUWE1</td>
<td>D664N</td>
<td>-1.1</td>
<td>Decrease Stability</td>
</tr>
<tr>
<td>X:98011469</td>
<td>T</td>
<td>NKRF</td>
<td>A704T</td>
<td>-1.1</td>
<td>Decrease Stability</td>
</tr>
<tr>
<td>X:100773108</td>
<td>A</td>
<td>-</td>
<td>I760N</td>
<td>-0.21</td>
<td>Decrease Stability</td>
</tr>
</tbody>
</table>

4. Conclusion
This study highlights the utility of computational algorithms and bioinformatics pipelines in identifying deleterious mutations, particularly in scenarios lacking extensive animal records. The identified deleterious SNPs in pigs, mapped to genes HUWE1 and NKRF, provide valuable insights into potential genetic risks. Further research, particularly in Indian native pig breeds, is warranted to elucidate the biological consequences of these mutations through in vivo studies.

5. Acknowledgements
The authors acknowledge the support of Co-authors and Institute.

6. Conflict of Interest
The authors declare no conflict of interest.

7. References


