School of Biological Sciences

QUANTITATIVE AND COMPUTATIONAL METHODS

FIELDCOURSE REPORT

A DUAL ANALYSIS OF WILD AND DOMESTIC PIGS IN EUROPEAN AND EAST-ASIAN POPULATIONS​

1876793

WORD COUNT: 2392

**Abstract**

The events and trends surrounding the domestication of *Sus scrofa* breeds in select European and Asian countries was investigated in a dual analysis with the genetic variance and homogeneity of modern pig populations. *Sus scrofa* genome data from Yang, Bin et al. (2018) was filtered through Blue Pebble supercomputing and calculated statistically in R Studio to create five principal component analysis plots for visual analysis. The four protein-coding genes ABLIM1, EXOC2, TRUB1, and ADGRL2 were investigated for homogeneity and displayed various trends corresponding to socio-economic and geographical parameters surrounding the pig populations. Each gene was a key regulator and component in mammalian developmental processes. Consequently, ABLIM1 and EXOC2 suggested the highest natural selective pressure, while TRUB1 and ADGRL2 suggest artificial selection pressures instead, for livestock production.

**Introduction**

The wild boar *Sus scrofa* is indigenous to Eurasia and North Africa. They were domesticated in Southwest Asia approximately 8,500 BC, over a complex process spanning multiple millennia and involving various Eurasian wild boar populations, according to zooarchaeological evidence. Domestication occurred independently in Eastern and Western Eurasia throughout the agricultural revolution about 10,000 years ago. During the Neolithic period, farmers arrived in Europe with their domestic crops and livestock in tow, such as pigs, *Sus scrofa domesticus* (Ottoni*et al*., 2013). Traditions such as pig transhumance and the genetic introgression occurring with the introduction of wild boars to domestic populations points to human-mediation as a driving factor in genetic flow patterns (Xiang*et al*., 2017). European pigs of 7,100-6,000 ya did possess Near Eastern and European mtDNA, however, modern pigs possess less than 4% of that Near Eastern ancestry. Their haplotypes have been replaced by those of European breeds (Frantz*et al*., 2019). Events corresponding to the domestication, dispersal, and genetic mixing of wild boar and pig breeds across these two regions can be determined with single nucleotide polymorphism (SNP) data (Yang*et al*., 2017). The aim of this investigation was to identify trends of genetic flow among European and Asian wild boar and pig populations, with a focus on specific genes that may have attributed to variance in these breeds.

**Methods**

The genome data gathered by Yang, Bin et al. (2018) of global domesticated pigs and wild boars was utilized. A dual analysis was conducted on the genetic differences between East Asian and European populations, as well as the domestic and wild breed populations within these two regions. There was a focus on data collected from China, Korea and Thailand as representatives of East Asia, and Italy, Portugal, Finland and Sweden as representatives of Europe. Finland and Sweden were both selected, as there was an absence of data for the Finnish domestic pigs and the Swedish wild boars. To filter out this custom dataset, the data was transferred to the Blue Pebble supercomputer. The file was trimmed to only contain the selected breeds. The file was also converted from a .ped file to .bed, .bim and .fam files, using the module plink 1.9, for compatibility with the R package pcadapt for data analysis. The data was loaded into R studio with pcadapt. Three forms of analysis were carried out beginning with a Scree plot, followed by a principal component analysis, and concluding with a Manhattan plot. The Scree plot was created in R to determine the number of significant factors to retain for analysis. It was composed of six principal components with a clear dip between PC1 and PC2. Therefore, two principal components were used for the principal component analysis (PCA) plot axes. PCA plots were created to reduce dimensionality of the dataset for simpler graph visualization. One plot was created to visually compare individuals between domesticated and wild breeds, and between East Asian and European breeds. Four plots were created to visually compare single nucleotide polymorphisms (SNPs) hetero-/homozygosity in selected genes. The Manhattan plot was created to identify statistically significant SNPs. Value 15 was chosen as the statistically significant cut off. R Studio was used to identify how many statistically significant SNPS were present and indicated their chromosome and base position coordinates on the genome. The resultant twenty significant SNP coordinates were searched on NCBI Genome Data Viewer version 10.2. The genome browsing feature was used to find genes that the SNPs were inside. The genes were documented in Table 1, as these would likely be those undergoing selection. Thus, the eight genes, ABLIM1, EXOC2, TRUB1, 2NF84, ADGRL2, MY016, MAST4/NLN, and RNF180 were identified from the SNP coordinates. The remaining twelve SNPs were located in exon regions or near a gene. The four genes, ABLIM1, EXOC2, TRUB1, and ADGRL2 were finally selected for analysis in PCA plots based on the corresponding SNPs having been located inside a gene and having the highest statistical significance as recorded in Table 1.

**Results**

The score plot produced via PCA in Figure 1 shows six defined clusters of *Sus scrofa* and *Sus scrofa domesticus* individuals based on their similarity. There is an even divide between the European breeds and Asian breeds, except for the Korean pigs occurring far closer to the European breeds. The most notable result is the scattered nature of the Thai pigs. They are widely spread with some individuals located among the tightly clustered Asian breeds, scattering down to the Korean pigs. The difference in variation between the Chinese breeds and the most distant Thai pigs is far larger compared to the variation between the European breeds and Korean pigs due to their wide reach across the PC1 axis. Four clusters consist of individuals belonging to the same breed entirely: Italian pigs, Portuguese pigs, Korean pigs, and Swedish pigs. The remaining two clusters consist of a multiple breeds: an admixture of Portuguese, Finnish, and Italian wild boards and an admixture of Thai pigs and wild boars with Chinese pigs and wild boars. Chinese wild and domesticated breeds have similar expression profiles to each other. Throughout the plot, individuals were clustered distinctly in domesticated or wild groups, especially tight among wild boar breeds. An anomaly was a stray Swedish pig among the Portuguese pigs. Another was the Finnish wild boar located between the Italian pigs and wild boars.

Figure 2 shows a panel of four PCA plots corresponding to the genes: ABLIM1, EXOC2, TRUB1, and ADGRL2 respectively. The plots show the distribution of heterozygous (red) and homozygous (black and green) derived alleles among the pig and wild boar individuals, also projected onto PC1 and PC2. Observing gene ABLIM1, all the individuals were homozygous, except for the heterozygous stray Swedish pig among the Portuguese pigs. The Swedish pig community was denoted black for possessing one homozygous version of the allele, while all other communities were denoted green possessing the other version of the homozygous allele. Observing gene EXOC2, again all individuals were homozygous, now with a split in Asian breeds possessing one homozygous allele version (black), and the European breeds possessing the other (green). The Thai pigs were the exception possessing heterozygosity. The vast geographic separation of the ancestral wild boar populations between Europe and Asia for 1.2 million years accounts for their high genetic divergence (Yang*et al*., 2017). The most mixed results were presented upon observing gene TRUB1. Most European breeds were homozygous with one version of the allele (green), except for the Swedish community which shared homozygosity (black) with Asian breeds. Heterozygosity was possessed by two out of seven Thai pigs, the stray Swedish pig among the Portuguese pigs, and a sole Italian pig within its community. Observing gene ADGRL2, all breeds shared homozygosity with one version of the allele (green), except for the Chinese pig community. There was an admixture of heterozygous individuals and two homozygous individuals of the other homozygous allele version (black).

**Discussion**

The results from the PCA plots demonstrate various trends, surrounding European and Asian pigs and wild boars, attributed to socio-economic and geographic parameters, corresponding to Figure 1. Among the six core clusters, both Asian and European wild boar populations cluster tightly compared to their domesticated counterparts which are likely to have been traded in the past allowing genetic mixing. Portuguese, Italian and Finnish wild boar are closely linked by one Finnish boar whose closeness to the Italian pig cluster could indicate mixed ancestry. The Korean pig cluster appears dramatically close to European breeds, unfortunately a lack of literature indicates as to why. In comparison, the stray heterozygous Swedish pig found in in the Portuguese pig cluster suggests some outbreeding (Frantz*et al*., 2019).

By investigating the biomechanical and physiological functions of the four genes, a better understanding of the observed trends in Figure 2 can be distinguished. The gene ABLIM1 codes for a cytoskeletal actin-binding protein used for interactions between actin filaments and cytoplasmic targets. ABLIM1 interacts with F-actin in the retina and muscles. It plays a critical role in regulating cell migration and aggregation by negatively controlling osteoclast differentiation and enhancing cell-cell fusion (Narahara*et al*., 2018). Although this gene’s physiological functions require further research, these known functions likely operate in muscle production and regulation of developmental pathways, attributing to the shared homogeneity across each breed as a highly relevant gene. The heterozygous Swedish pig appears to be an anomaly. Thus, ABLIM1 may be favoured for artificial selection in livestock production. The second gene EXOC2, codes for proteins that are a component of the exocyst complex and assist exocytosis and vesicle tethering plasma membranes. This gene may be a trait of skin pigmentation in *Sus scrofa domesticus,* and it is attributed to skin pigmentation and tanning in humans, especially the Caucasian population, though the biological mechanisms are not yet thoroughly understood (Li*et al*., 2018). The even divide between the homogeneity of the European breeds and Asian breeds could be a result of the differing geographical and climatic conditions putting selective pressure on this gene. In contrast, the heterozygosity of the Thai pigs is interesting in terms of demonstrating the genetic flow among their population. It may be attributed on one account to Thailand’s economic boom in the 1960s where pigs became their major industrial livestock sector. Production was driven by backyard raising of crossbreeds and by a new demand of the tourism influx at the time. Additionally, disease, such as the foot and mouth epidemic, drove fluctuations in pig populations around 1991-1996, whereby European pigs were grossly imported allowing for this genetic flow (Tisdell, 1997). Third, TRUB1 is a protein-coding gene involved in pseudouridine synthase activity, mRNA synthesis, and tRNA modification (Safra*et al*., 2018). The Swedish individuals share homogeneity with all Asian breeds in panel 3 Figure 2, and this raises questions to the historic genetic flow between the Scandinavian and Near Eastern regions, and if the gene has been conserved because of its foundational role in developmental processes. The mixed results corresponding to this gene across all the breeds highlight its vital role and therefore it being under the most selective pressure out of the four genes. Fourth, ADGRL2 is a protein-coding gene regulating cellular development of the external and internal cell environments (Kang*et al*., 2018). Its shared homogeneity (green) with all breeds except for Chinese pigs demonstrates evidence of imports of European pigs to China in panel 4 Figure 2. A majority of heterozygous Chinese pigs and a minority of homozygous Chinese pigs of the second allele version (black) is observed, indicating imports of domesticated European pigs for livestock and breeding with local domestic populations to create these trends. In conclusion, the pigs’ globalization has thus arrived full circle whereby Chinese breeders are selecting European breeds to improve local livestock, where once they were indigenously domesticated (Yang*et al*., 2017).

**Figure and Tables**

**Table 1.** Recorded features of the 20 statistically significant SNPs identified using the Manhattan plot, R Studio and NCBI genome browser, including their statistical significance value, genome coordinates, location around the gene, and name of the gene or region identified.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Statistical Significance Value | Chromosome | Base Position | Inside/Outside the Gene | Gene Identified |
| 5.474845e-82 | 14 | 136174025 | in | ABLIM1 |
| 2.025736e-47 | 14 | 135517916 | in | ABLIM1 |
| 1.598619e-37 | 2 | 76667109 | out | ABLIM1 |
| 2.666206e-36 | 15 | 56937444 | out | LOC102160701 |
| 5.881120e-28 | 7 | 69146 | in | EXOC2 |
| 4.318276e-23 | 14 | 23888647 | out | LOC100739427 |
| 3.375896e-23 | 14 | 136424376 | in | TRUB1 |
| 9.658542e-22 | 14 | 32226387 | out | ZNF84 |
| 3.915625e-21 | 4 | 132253036 | in | ADGRL2 |
| 1.324704e-20 | 10 | 75737576 | out | MY016 |
| 8.326346e-19 | 16 | 44694093 | out | MAST4/NLN |
| 1.271417e-19 | 7 | 18009952 | in | LOCI106504342 |
| 8.491327e-17 | 16 | 44763969 | out | Close to   LOCI06506511 - Exon region around MAST4 |
| 8.491327e-17 | 16 | 45251439 | out | Close to  LOC106506511 - Exon region around MAST4 |
| 8.491327e-17 | 16 | 45350052 | out | Close to   LOC106506511 – Exon region around MAST4 |
| 8.491327e-17 | 16 | 45829750 | in | RNF180 |
| 2.028737e-17 | 14 | 136363034 | In x2 | LOC102160539  (Close to ABLIM) |
| 1.431113e-17 | 4 | 48758272 | out | Close to  LOC102159390 |
| 1.071782e-17 | 3 | 126417350 | in | LOC102160263 |
| 2.308251e-16 | 14 | 137086896 | in | TRUB1 |

Chart, scatter chart

Description automatically generated

**Figure 1.** PCA plot projected onto PC1 and PC2 where K=2, depicting six defined clusters of *Sus scrofa* and *Sus scrofa domesticus* coloured red, pink, magenta, or brown to represent East Asian breeds and blue and green to represent European breeds. Lighter shades distinguish domesticated pigs from darker shaded wild pigs in this dual comparison.

Chart, scatter chart

Description automatically generated

**Figure 2.** Panel of four PCA plots projected onto PC1 and PC2 where K=2, depicting the distribution of heterozygous (red) and homozygous (green and black) derived alleles in individuals of *Sus scrofa* and *Sus scrofa domesticus* within the selected genes ABLIM1, EXOC2, TRUB1, and ADGRL2.

**References**

Frantz, L.A., Haile, J., Lin, A.T., Scheu, A., Geörg, C., Benecke, N., Alexander, M., Linderholm, A.,

Mullin, V.E., Daly, K.G. and Battista, V.M., 2019. Ancient pigs reveal a near-complete

genomic turnover following their introduction to Europe. *Proceedings of the National Academy*

*of Sciences*, *116*(35), pp.17231-17238.

Kang, H.C., Kim, H.S., Lee, J.B., Yoo, C.K., Choi, T.J. and Lim, H.T., 2018. Analysis of the

relationship between the end weight trait and the gene ADGRL2 in purebred landrace pigs

using a Genome-wide association study. Korean Journal of Agricultural Science, 45(2),

pp.238-247.

Li, Z., Wei, S., Li, H., Wu, K., Cai, Z., Li, D., Wei, W., Li, Q., Chen, J., Liu, H. and Zhang, L., 2017.

Genome-wide genetic structure and differentially selected regions among Landrace, Erhualian,

and Meishan pigs using specific-locus amplified fragment sequencing. Scientific reports, 7(1),

pp.1-12.

Narahara, H., Sakai, E., Yamaguchi, Y., Narahara, S., Iwatake, M., Okamoto, K., Yoshida, N. and

Tsukuba, T., 2019. Actin binding LIM 1 (abLIM1) negatively controls osteoclastogenesis by

regulating cell migration and fusion. *Journal of cellular physiology*, *234*(1), pp.486-499.

Ottoni, C., Girdland Flink, L., Evin, A., Geörg, C., De Cupere, B., Van Neer, W., Bartosiewicz, L.,

Linderholm, A., Barnett, R., Peters, J. and Decorte, R., 2013. Pig domestication and human-

mediated dispersal in western Eurasia revealed through ancient DNA and geometric

morphometrics. *Molecular biology and evolution*, *30*(4), pp.824-832.

R Core Team (2020). ‘R: A language and environment for statistical computing’, R Foundation for

Statistical Computing

Safra, M., Nir, R., Farouq, D., Slutskin, I.V. and Schwartz, S., 2017. TRUB1 is the predominant

pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code.

Genome research, 27(3), pp.393-406.

Tisdell, C., Murphy, T. and Kehren, T., 1997. Characteristics of Thailand's commercial pig and

poultry industries, with international comparisons. *World Animal Review*, (89), pp.2-11.

Xiang, H., Gao, J., Cai, D., Luo, Y., Yu, B., Liu, L., Liu, R., Zhou, H., Chen, X., Dun, W. and Wang,

X., 2017. Origin and dispersal of early domestic pigs in northern China. *Scientific*

*reports*, *7*(1), pp.1-9.

Yang, B., Cui, L., Perez-Enciso, M., Traspov, A., Crooijmans, R.P., Zinovieva, N., Schook, L.B.,

Archibald, A., Gatphayak, K., Knorr, C. and Triantafyllidis, A., 2017. Genome-wide SNP data

unveils the globalization of domesticated pigs. *Genetics Selection Evolution*, *49*(1), pp.1-15.