Animal Biotechnology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/labt20

# Mitochondrial DNA Sequence and Phylogenetic Evaluation of Geographically Disparate Sus scrofa Breeds

M. V. Cannon [a] , T. D. Brandebourg [b] , M. C. Kohn [a] , D. Đikić [c] , M. H. Irwin [a] & C. A. Pinkert [a d]

[a] Department of Pathobiology , Auburn University , Auburn , Alabama , USA
[b] Department of Animal Sciences , Auburn University , Auburn , Alabama , USA
[c] Department of Animal Physiology , University of Zagreb , Croatia
[d] Department of Biological Sciences , The University of Alabama , Tuscaloosa , Alabama , USA

Published online: 25 Aug 2014.

To cite this article: M. V. Cannon , T. D. Brandebourg , M. C. Kohn , D. Đikić , M. H. Irwin & C. A. Pinkert (2015) Mitochondrial DNA Sequence and Phylogenetic Evaluation of Geographically Disparate Sus scrofa Breeds, Animal Biotechnology, 26:1, 17-28, DOI: 10.1080/10495398.2013.875474

To link to this article: http://dx.doi.org/10.1080/10495398.2013.875474

*Animal Biotechnology*, 26:17–28, 2015

ISSN: 1049-5398 print/1532-2378 online
DOI: 10.1080/10495398.2013.875474

Taylor & Francis
Taylor & Francis Group

# Mitochondrial DNA Sequence and Phylogenetic Evaluation of Geographically Disparate *Sus scrofa* Breeds

**M. V. Cannon,[1] T. D. Brandebourg,[2] M. C. Kohn,[1] D. Đikić,[3] M. H. Irwin,[1] and C. A. Pinkert[1,4]**

[1]*Department of Pathobiology, Auburn University, Auburn, Alabama, USA*
[2]*Department of Animal Sciences, Auburn University, Auburn, Alabama, USA*
[3]*Department of Animal Physiology, University of Zagreb, Croatia*
[4]*Department of Biological Sciences, The University of Alabama, Tuscaloosa, Alabama, USA*



**Next generation sequencing of mitochondrial DNA (mtDNA) facilitates studies into the metabolic characteristics of production animals and their relation to production traits. Sequence analysis of mtDNA from pure-bred swine with highly disparate production characteristics (Mangalica Blonde, Mangalica Swallow-bellied, Meishan, Turopolje, and Yorkshire) was initiated to evaluate the influence of mtDNA polymorphisms on mitochondrial function. Herein, we report the complete mtDNA sequences of five *Sus scrofa* breeds and evaluate their position within the phylogeny of domestic swine. Phenotypic traits of Yorkshire, Mangalica Blonde, and Swallow-belly swine are presented to demonstrate their metabolic characteristics. Our data support the division of European and Asian breeds noted previously and confirm European ancestry of Mangalica and Turopolje breeds. Furthermore, mtDNA differences between breeds suggest function-altering changes in proteins involved in oxidative phosphorylation such as ATP synthase 6 (MT-ATP6), cytochrome oxidase I (MT-CO1), cytochrome oxidase III (MT-CO3), and cytochrome b (MT-CYB), supporting the hypothesis that mtDNA polymorphisms contribute to differences in metabolic traits between swine breeds. Our sequence data form the basis for future research into the roles of mtDNA in determining production traits in domestic animals. Additionally, such studies should provide insight into how mtDNA haplotype influences the extreme adiposity observed in Mangalica breeds.**

**Keywords** Mangalica; Meishan; Mitochondrial DNA; *Sus scrofa*; Turopolje

## INTRODUCTION

Swine breeds that demonstrate divergent growth and carcass traits are useful as models to investigate the genetic basis of metabolic traits, feed utilization, and carcass composition or as translational animal models of human metabolic disease (1–3). For example, breeds such as the Mangalica (Blonde and Swallow-bellied) exhibit a slow growth rate and an extreme propensity for fat deposition (4). In contrast, Turopolje swine are unique for their exceptionally high feed efficiency (5). Elucidating the basis for such extreme phenotypes could aid the development of methodologies that improve production efficiency in food animals or prevent obesity in humans.

The Turopolje breed is known for a high feed conversion rate presumably as a result of selective pressures within their historic environment that would preclude efficient breeding and growth of many domestic breeds. The Mangalica breeds are considered primitive lard-type hogs due to their extreme genetic propensity to fatten, which suited their use for high-quality ham and loin products in traditional markets of Eastern Europe. The Meishan breed is also known to have a high propensity toward obesity; however, unlike the Mangalica, which exhibit a delayed onset of puberty and small litter sizes, Meishan sows are precocious and farrow exceptionally large litters. Yorkshire pigs are a common breed of domestic swine with metabolic characteristics comparable to other domestic swine, generally exhibiting lean, well-muscled carcasses and fast growth rates. It is possible that differences in mitochondrial function influence the divergent phenotypes of these swine breeds. Characterization of the mitochondrial sequence of Turopolje, Mangalica, and Meishan swine will allow this hypothesis to be tested and provide preliminary insights into the roles the mitochondrial genome may play in regulating the growth and reproductive phenotypes exhibited by these breeds. Whereas mitochondrial DNA represents only

Address correspondence to Carl A. Pinkert, The University of Alabama, 201 Rose Administration Building, Tuscaloosa, AL 35487, USA. E-mail: cap@ua.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/labt.

a fraction of the total genetic material of an individual that determines phenotypic characteristics, genes encoded by the mtDNA are critical to proper metabolic function, as shown by the devastating effects of mitochondrial mutations (6). Establishing the complete mitochondrial DNA sequence associated with divergent metabolic phenotypes is a necessary first step toward better understanding the regulation of feed efficiency by mitochondrial genes in swine and the ultimate goal of favorably altering feed conversion rates and overall production performance (7–9). Finally, new insights into mitochondrial function will also help to establish lard-type breeds as models of obesity and type 2 diabetes.

## MATERIALS AND METHODS

### Animals

All animals used for sample isolation were handled in accordance with the national animal care and management guidelines.

### Mitochondrial DNA Sequencing

Mitochondrial DNA (mtDNA) sequences were determined by high throughput next generation sequencing using an Illumina Genome Analyzer IIx. To prepare for sequencing, mtDNAs were amplified by conventional PCR using TaKaRa LA Taq for 30 cycles with an annealing temperature of 60°C and a 10 min. extension phase at 68°C. All mtDNAs were amplified in two overlapping fragments using the following primers:

TTCTCCTCGCACACGCTTACATCA, Sus scrofa 2872F
TTAGTTCGGTGGCGGTGAAGGTTA, Sus scrofa 11654R
TCAGCACGCCTCCCATTCTCAATA, Sus scrofa 10714F
TTACGCAATTACCGGTCTCTGCCA, Sus scrofa 3837R

Amplicons were combined in equimolar amounts and sequenced using an Illumina Genome Analyzer IIx at the Genomic Services Lab at the HudsonAlpha Institute for Biotechnology.

### Sequence Data Analysis

Data were analyzed using the public Galaxy server (10, 11). Sequence reads were mapped to a known *Sus scrofa* mtDNA sequence (Large White Acc# NC_012095) using BWA for Illumina, converted to .bam format from .sam format using samtools and a pileup file was created from .bam files to generate consensus sequences (12, 13). Files in .bam format were viewed with the Integrative Genomics Viewer (IGV) to determine coverage and confirm sequence (14).

Mitochondrial protein sequence derived from the new mtDNA sequences and the reference used for mapping (Large White Acc# NC_012095) were aligned using MAFFT (15, 16).

### Phylogenetic Analysis

A phylogeny was constructed using both the new mtDNA sequences and sequences obtained from the NCBI sequence database (Warthog gi|88766360, Lanyu gi|306485876, European wild boar gi|209166268, Duroc gi|209166170, wei10 gi|145315568, Large White gi|223976078, Banna Mini gi|238625868, Taoyuan gi|95116704, taiwanensis gi|312233363, Xiang gi|145315790, Iberian gi|209166100, Malasian wild boar gi|145315776, Bihu gi|145315750, Hampshire gi|45826183, Berkshire gi|45826169), with warthog as an outgroup to root the tree. MAFFT aligned the mitochondrial genomes and changed the alignment into Phylip format. The alignment was then imported into the MEGA program package to construct a bootstrapped (1000 replications) maximum likelihood phylogram (Fig. 1) (17). The phylogram was constructed using nucleotide sequences of NADH dehydrogenase 1 (MT-ND1), NADH dehydrogenase 2 (MT-ND2), cytochrome c oxidase I (MT-CO1), cytochrome c oxidase II (MT-CO2), ATP synthase 8 (MT-ATP8), ATP synthase 6 (MT-ATP6), cytochrome c oxidase III (MT-CO3), NADH dehydrogenase 3 (MT-ND3), NADH dehydrogenase 4 (MT-ND4), NADH dehydrogenase 5 (MT-ND5), and

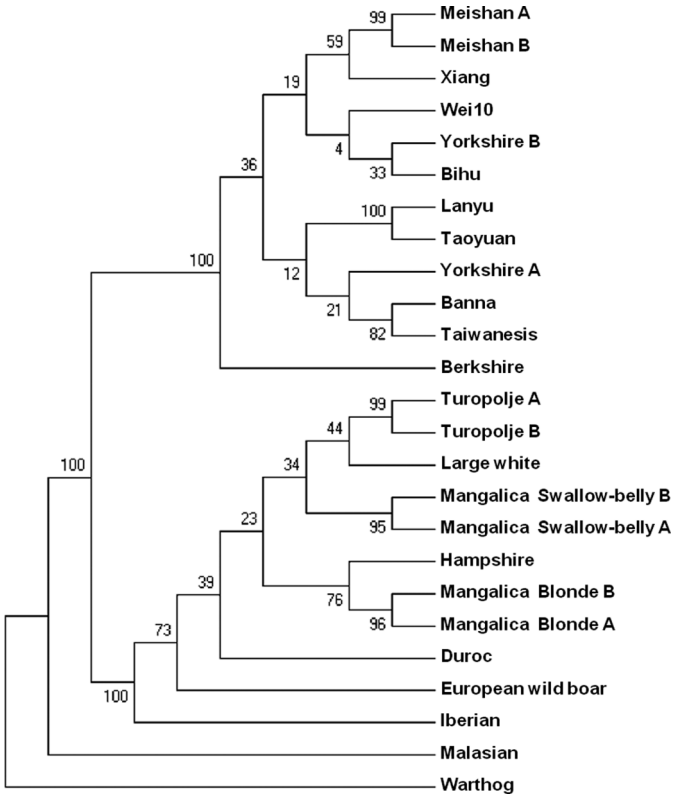



FIG. 1. mtDNA sequence derived phylogram of swine breeds. A phylogram was generated using the MEGA program from both coding and non-coding regions of the mtDNA. 1000 replicates allowed generation of bootstrap values. "A" and "B" samples refer to the two samples sequenced from each breed analyzed in this report.

cytochrome b (MT-CYB) proteins (all three codon positions), noncoding regions including the 12S and 16S rRNAs and non-repetitive regions of the D-loop.

### Protein Structural Prediction

In order to consider functional consequences of amino acid differences between swine breeds, the three-dimensional structure of mitochondrial proteins from the Large White mtDNA sequence was predicted using the protein threading program Phyre2 (18). The three-dimensional information presented in a two dimensional map of protein domains was obtained from the data output and annotated with the location of amino acid differences among swine breeds sequenced in this study. Only high confidence structures based on homologous protein structures were considered.

### Protein Conservation Estimates

To calculate estimates of conservation of mitochondrial proteins, the Large White sequence of each protein was blasted against the Pubmed nonredundant (nr) database using blastp, keeping a maximum of 5000 vertebrate matches (19). Any match shorter than 95% of the length of the Large White sequence was discarded to remove incomplete sequences. A single sequence was kept per species and the sequences were aligned using clustal omega (20). Custom Perl scripts then counted the frequency of each amino acid at all positions in each protein. Percent conservation was calculated by dividing the most abundant amino acid count by the total amino acid count for each position in each protein, with a conservation score of 1 (100%) representing a position where only one amino acid was present.

### Animals and Measurement of Growth Parameters

Purebred Yorkshire pigs and both Blonde and Swallow-bellied lines of Mangalica pigs were housed in individual pens at the Auburn University Swine Research and Education Center (SREC) for the duration of these experiments. Pigs were fed a standard grower ration (17% CP) from 40 to 120 lb body weight (BW) and a finisher ration (15% CP) from 120 to 240 lb BW. Daily feed intakes and weekly body weights were recorded during a period spanning growth from 40 lb live weight until reaching harvest weight of 240 lb to facilitate measurement of average daily gain (ADG), feed efficiency (lb gained/lb feed), and total feed intake. Twelve barrows were randomly selected for each group to determine phenotypic differences inherent to breed. Data were subjected to an analysis of variance, and levels of significance (p-values) were corrected for multiple comparisons by a Bonferroni t-test.

### Carcass Fabrication

Animals were harvested at the Auburn University Lambert-Powell Meats Lab under USDA-FSIS inspection and carcasses were chilled at $2 \pm 1^{\circ}C$ for 24 h. Carcass characteristics of Mangalica (Swallow-belly and Blonde) and Yorkshire animals were then determined by splitting each carcass between the 10th and 11th ribs for evaluation of back fat (BF) and loin eye area (LEA).

### Gene Expression Analysis

Gene expression analysis was performed on samples from Yorkshire, Mangalica Blonde and Swallow-belly swine ($n = 6$/group). Upon exsanguination, subcutaneous adipose tissues were immediately collected, snap frozen in liquid nitrogen, and stored at $-80^{\circ}C$ until mRNA analysis. Total RNA was extracted from adipose tissue using a two-step purification protocol with total RNA first being extracted from whole tissue using RNAzol RT (MRC, Inc, Cincinnati, OH) followed by purification using RNAeasy spin columns (QIAGEN, Inc., Valencia, CA) according to the manufacturers' recommendations. RNA was quantified using a BioTek Synergy 4 plate reader utilizing the Take3 system (BioTek U.S., Winooski, VT) with all samples exhibiting an OD 260/280 between 1.8 and 2.0 and an OD 260/230 value between 1.8 and 2.2. Spectral scans ranging from 200 to 400 nm further verified sample purity as all RNA samples produced smooth curves exhibiting one peak at 260 nm. Total RNA integrity was accessed both visually by resolving 2 μg of RNA on a denaturing formaldehyde gel containing ethidium bromide and by determining an RNA Integrity Number (RIN) using an Agilent 2100 bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA). All samples demonstrated sharp ribosomal bands with a 28S to 18S ratio greater than 1 and RIN values greater than 7.0 and were thus judged intact and nondegraded. Total RNA was then reverse transcribed using Superscript II reverse transcriptase (Promega Inc, Madison, WI) and oligo-dT primers. Real-time PCR was performed on the resultant cDNA using a Roche Lightcycler 480 Real-time PCR machine and LightCycler 480 SYBR Green I Master Mix (Roche Applied Science, Indianapolis, IN) according to manufacturer's recommendations. All PCR reactions were performed using intron-spanning primers for porcine leptin (forward: 5′-TCGACCATCAAGCAGGGTTC-3′; reverse: 5′-CAACTTGTGTTGCGTGGGAG-3′) and porcine S15 (forward: 5′-GGTAGGTGTCTACAATGGCAAGG-3′; reverse: 5′-GGCCGGCCATGCTTC-3′) under optimized conditions. Primer efficiencies were calculated using standard curves and were 93% and 100% for leptin and S15 reactions, respectively. Product purity was assessed by melting curve analysis and expected amplicon sizes were verified on a 2% agarose gel stained with ethidium bromide. Values were normalized to ribosomal protein S15 (S15) mRNA expression. The S15 mRNA levels represent an appropriate control as the efficiency of the S15 primers was 100% and S15 mRNA expression was not different

TABLE 1
Predicted amino acid alterations based upon mtDNA sequence with differences between *Sus scrofa* breeds

| | Gene (amino acid alterations – percentage of total amino acids) | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MT-ATP6 (3/226–1.3%) | | | MT-ATP8 (3/67–4.5%) | | | MT-CO1 (1/514–0.20%) | MT-CO2 (0/229–0%) | MT-CO3 (1/261–0.4%) | MT-CYTB (3/379–0.8%) | | | MT-ND1 (2/318–0.6%) | | MT-ND2 (4/347–1.2%) | | | |
| Amino Acid # | 119 | 136 | 185 | 40 | 63 | 66 | 191 | none | 32 | 198 | 295 | 314 | 246 | 252 | 100 | 211 | 239 | 336 |
| Conservation | 72.34% | 71.11% | 53.76% | 49.60% | 76.16% | 45.19% | 99.97% | | 64.85% | 99.14% | 77.41% | 91.03% | 49.72% | 83.06% | 29.67% | 69.63% | 36.08% | 61.84% |
| Large White | Y | L | N | I | L | P | T | | T | M | V | G | S | P | M | M | V | V |
| Mangalica Blonde | Y | L | N | I | L | P | T | | T | L | V | G | S | P | M | M | V | V |
| Mangalica Swallow-belly | Y | L | N | I | L | P | T | | T | L | V | G | S | P | M | M | V | V |
| Turopolje | Y | L | N | I | L | P | T | | T | L | V | G | S | P | M | M | V | V |
| Meishan | H | P | S | T | S | L | T | | I | L | M | S | F | S | L | T | I | I |
| Yorkshire A | H | P | S | T | S | L | T | | I | L | M | S | F | S | L | T | I | I |
| Yorkshire B | H | P | S | T | S | L | A | | I | L | M | S | F | S | L | T | I | I |

| | Gene (amino acid alterations – percentage of total amino acids changed) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MT-ND3 (4/115–3.5%) | | | | MT-ND4L (2/98–2.0%) | | MT-ND4 (2/459–0.4%) | | MT-ND5 (7/606–1.2%) | | | | | | | MT-ND6 (1/175–0.6%) |
| Amino Acid # | 8 | 29 | 96 | 114 | 13 | 48 | 361 | 413 | 198 | 399 | 434 | 488 | 577 | 584 | 600 | 5 |
| Conservation | 32.36% | 80.64% | 36.02% | 84.08% | 30.38% | 26.19% | 43.96% | 59.39% | 41.51% | 57.97% | 23.60% | 54.14% | 37.77% | 89.25% | 22.29% | 69.08% |
| Large White | L | A | I | A | A | I | V | T | L | V | K | M | V | I | T | I |
| Mangalica Blonde | L | A | I | A | A | I | V | T | L | V | K | M | V | V | T | I |
| Mangalica Swallow-belly | L | A | I | A | A | I | V | T | L | V | K | I | V | I | T | I |
| Turopolje | L | A | I | A | A | I | V | T | L | V | K | M | M | I | T | I |
| Meishan | F | T | T | T | T | V | M | I | F | A | Q | M | V | I | M | I |
| Yorkshire A | F | T | T | T | T | V | M | I | F | A | Q | M | V | I | M | I |
| Yorkshire B | F | T | T | T | T | V | M | I | F | A | Q | M | V | I | M | V |

Amino acid sequences for each mitochondrial protein were compared and differences between breeds are noted below. Values in parenthesis next to gene name are: (“number of amino acid changes”/“amino acids in protein”—“percent amino acids altered”). The number below the gene name denotes the amino acid position of the change and the single letters within the table are the single letter amino acid symbols.

between any groups tested ($P < 0.93$). Data were calculated and expressed as fold change relative to baseline (21).

## RESULTS

### Mitochondrial Sequence Analysis

Sequence coverage of the mitochondrial genomes was relatively consistent across the entire genome (~40,000X) with two notable exceptions. Coverage was lower at the ends of the reference sequence due to the mapping algorithm (~1000X coverage). Since the mtDNA could not be amplified in a single piece, two separate PCRs were required, which affected the sequence data read coverage. Coverage was higher in the region overlapped by the two original amplicons, as would be expected (~80,000X coverage).

mtDNA sequence data were used to predict amino acid sequences. Based upon comparison of the newly sequenced mtDNA from the five breeds examined in the current study, potential differences were observed at 33 amino acids across the mitochondrial genome. Predicted amino acid changes included alterations that would be expected to cause differences in polarity or pH, as noted in Supplemental Figures 1–4. Predicted amino acid sequences, for a select group of electron transport chain proteins based upon newly sequenced mtDNAs, are presented in Table 1.



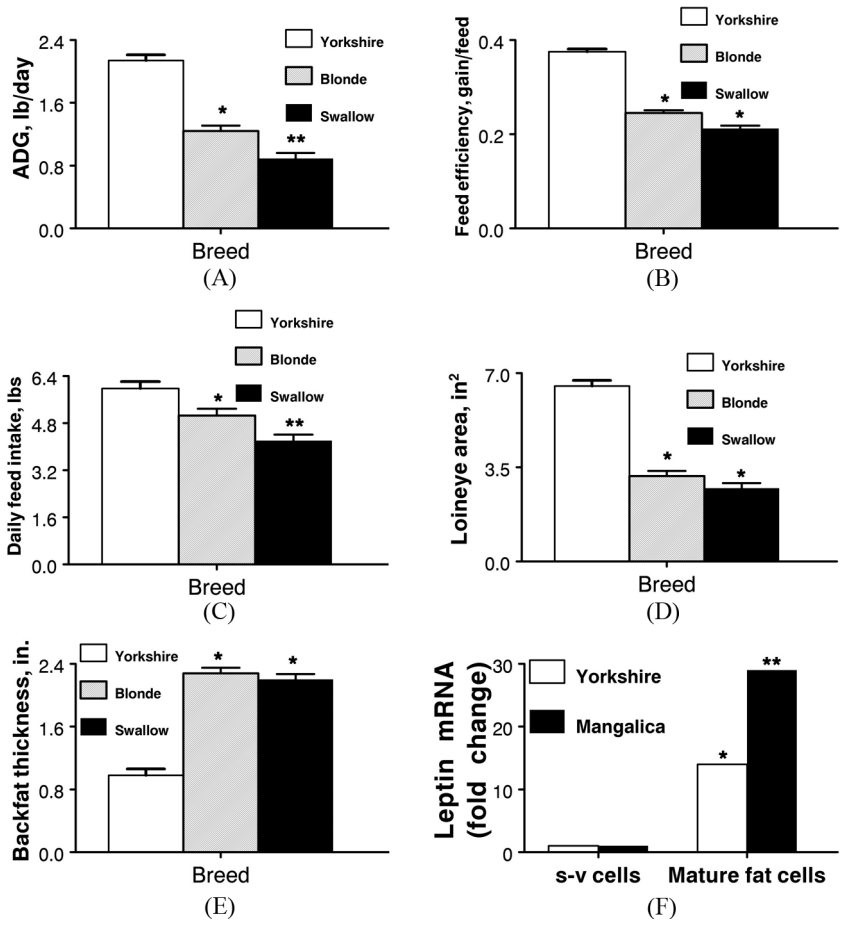



FIG. 2. Phenotypic characterization of Yorkshire and Mangalica breeds. Comparison of growth and carcass parameters between purebred lines of Yorkshire, Blonde Mangalica (Blonde) and Swallow-bellied (Swallow) Mangalica barrows demonstrated divergent phenotypes as judged by A) average daily gain (ADG), B) feed efficiency, C) average daily feed intake, D) LEA as a measure of muscle growth, and E) backfat thickness as a measure of carcass adiposity. Expression of mRNA for the adipocyte marker, leptin, was up-regulated in mature fat cells harvested from adipose tissue relative to cells in the stromal fraction (s-v cells) while Mangalica Blonde pigs express a greater amount of leptin mRNA compared to Yorkshire pigs consistent with Mangalica Blonde pigs having a higher propensity to fatten (F). Pigs were housed individually in pens and fed standard commercial diets matched to growth stage. Individual feed intakes were recorded daily and body weights were recorded weekly. The mRNA levels were measured using real-time PCR. Expression levels were normalized relative to the porcine S15 gene and are presented as fold change relative to expression in s-v cells, with standard error bars; $^{*}p < 0.05$ and $^{**}p < 0.001$.

### Phylogenetic Analysis

The phylogram reveals a general division of swine into European and Asian groups. The Mangalica and Turopolje mtDNA samples were most closely related to European swine breeds, and the Meishan and Yorkshire mtDNA sequences grouped with Asian swine breeds. The fine structure of the phylogram had low bootstrap values, but the base of the tree was supported well. Our data support earlier studies showing a division of swine breeds into European and Asian breeds (18–24).

### Two Dimensional Structures of Mitochondrial Proteins

Structures of mitochondrial proteins were created by protein threading using the Phyre2 server and were considered high confidence if known structures from closely related species were available to facilitate analysis (18). Mitochondrial proteins that did not fit this criterion were excluded. Using this approach, high confidence three dimensional protein structures (MT-ATP6, MT-CO1, MT-CO3, and MT-CYB) were generated and are presented in Supplemental Figures 1–4. Such predicted structures illustrate the potential consequences of divergent mtDNA sequences between breeds. For instance, protein threading analysis suggested several amino acid differences existed in proteins between breeds that would be expected to be associated with a change in polarity or pH of the residue in the primary sequence such as amino acid site 119 of MT-ATP6. This site had a residue predicted as a neutral polar tyrosine in Large White and a basic polar histidine in Meishan and Yorkshire samples. Changes in the polarity of amino acids or the pH of their side chains may alter the folding properties of mitochondrial proteins. Given the functional significance of proteins involved in the electron transport chain, it is tempting to speculate such changes would give rise to subtle differences in protein function that may underlie the divergent phenotypes observed between breeds.

### Protein Conservation Estimates

After blast search and quality control of data; between 498 and 3777 vertebrate protein sequences from different species were kept. Conservation estimates for each protein averaged 80.17% across all amino acids, with regions of high conservation (Supplemental Figures). For amino acids with differences between sequenced breeds and Large White, conservation estimates ranged from 22.29%–99.97% (Table 1).

### Phenotypic Characterization of Yorkshire, Mangalica Blonde and Mangalica Swallow-belly Breeds

Numerous phenotypic differences were noted between the Mangalica Blonde, Swallow-belly and Yorkshire breeds (Fig. 2). Average daily weight gain and daily feed intake were lower in the Mangalica breeds compared to Yorkshire pigs. Differences in body composition between the breeds were striking, as Yorkshire pigs achieved approximately twice the muscle mass of Mangalica breeds, as estimated by loin eye area, while backfat thickness in the Mangalica was roughly twice that of Yorkshire pigs. Despite lower feed intake, Mangalica breeds exhibited poorer feed efficiency than Yorkshire animals; reflecting a higher percentage of body fat per unit gain in the Mangalica breeds. As a consequence of increased adiposity, leptin mRNA levels were higher in mature fat cells of the Mangalica Blonde relative to mature fat cells obtained from Yorkshire pigs.

## DISCUSSION

Analysis of sequencing data revealed numerous differences between the swine breeds. Considering only changes predicted to result in amino acid differences, the observed pattern mirrors the phylogenetic grouping of European and Asian breeds. All polymorphisms seen in MT-ATP6, MT-ATP8, MT-CYB, MT-ND1, MT-ND2, MT-ND3, MT-ND4, and MT-ND4L were identical in the European breeds, but differed from Meishan and Yorkshire breeds (Table 1). Of note is the observation that most of the amino acid variation among the sequenced European breeds occurred in MT-ND5 (3 amino acids), with one amino acid change occurring in MT-CYB. This suggests that mtDNA-related differences among the European breeds occurred in Complexes I and III of the electron transport chain (ETC).

Also of note are amino acid changes in MT-COI and MT-ND6 between the two Yorkshire samples sequenced. These changes highlight how different the two samples were, despite being obtained from a single herd. Such differences could result in variance in metabolic efficiencies within the population.

Several studies in monogastric species, such as the broiler chicken, support the hypothesis that subtle differences in mitochondrial ETC membrane protein characteristics, especially those that potentially effect proton leak, is associated with differences in the phenotypic expression of feed efficiency. For instance, inferior feed efficiency in broilers was associated with higher reactive oxygen species generation in duodenal mitochondria (29). Furthermore, studies examining mitochondrial proton leak kinetics in isolated breast muscle demonstrated differences in mitochondrial proton leak between genetic lines selected for either high or low feed efficiency (7, 30, 31). Observations such as these suggest that it would not be necessary for polymorphisms in proteins of the electron transport chain to cause dramatic changes in protein function, as subtle differences in proton motive force could meaningfully impact feed efficiency.

Despite low bootstrap values for the fine structure of the tree, our data illustrate a phylogenic relationship that is consistent with features of trees reported previously and show that Mangalica swine are most closely related to European breeds (22, 23, 26, 27). The grouping of Turopolje and Mangalica breeds with European swine is unsurprising. The two breeds generally grouped together, though with lower bootstrap values, while European and Duroc breeds clustered. We hypothesized that the Yorkshire breed would be most similar to European breeds. However, our finding that the Yorkshire breed was most similar to Asian breeds is consistent with previous reports, which established that many European breed populations harbor Asian haplotypes (32, 33). The two Yorkshire animals sequenced in this study both possessed an Asian mtDNA haplotype, though it should not be assumed that all Yorkshire swine necessarily share this haplotype. It was noted that haplotypes observed within a single herd indeed vary (32). Previously published swine phylogenies often support the relationship of Yorkshire breeds with other European breeds, and it may be important to consider nuclear loci in any phylogeny of swine in addition to mtDNA sequences in order to obtain an accurate evolutionary lineage (22–28). The observation that both Asian and European haplotypes still exist may suggest that none of the haplotypes studied were associated with more optimal production or metabolic characteristics. Otherwise, specific loci would likely have swept to fixation. An alternate explanation could be that the different mtDNA haplotypes could provide metabolic advantages, but only under specific agricultural conditions (e.g., environmental temperature or nutritional plane associated with quality or quantity of feed). Additionally, there could be specific nuclear-encoded proteins that function more efficiently with mtDNA-encoded proteins from a given haplotype. Thus, genetic variants among swine populations could preserve specific mtDNA haplotypes due to nuclear-mitochondrial interplay.

Three-dimensional computer models of mitochondrial proteins allowed visualization of how amino acid changes could affect protein conformation. The noted amino acid changes tended to occur in alpha helices, where amino acid changes could affect the kinetics of protein folding or the strength of protein: protein interactions. Changes from polar to nonpolar amino acids in MT-CO1, MT-CO3, or from nonpolar to polar in MT-CYB could affect function of these proteins resulting in altered metabolic characteristics between breeds.

An additional method with which to gauge functionality of divergent amino acids between sequenced breeds is to estimate the conservation of the position across many species. Protein conservation was estimated using blast searches identifying similar vertebrate protein sequences to Large White followed by counting each amino acid at each position in the aligned protein sequences. The value calculated by dividing the count of the most common amino acid by the total amino acids at that position provides a rough estimate of protein conservation across vertebrate species. The estimate is not perfect, as it does not consider evolutionary distance between identified species in the calculation, but still provides a useful metric with which to gauge conservation. The conservation scores for amino acids differing among sequenced breeds of swine (Table 1) show that seven of the 23 amino acids are highly conserved (>80%) across vertebrate species, indicating that these differences could affect protein function. These differences largely separate Asian from European haplotypes, though differences in highly conserved (>80%) amino acids in MT-CO1, MT-CYTB, and MT-ND5 indicate that there may be breed specific differences in mitochondrial function.

Phenotypic characterization of Mangalica Blonde, Mangalica Swallow-belly, and Yorkshire animals revealed striking differences in carcass composition. The Mangalica breeds had roughly double the backfat and half of the muscle mass displaying a much greater propensity to fatten relative to Yorkshire animals. This extreme adiposity was present despite a lower overall feed intake and largely explains the lower feed conversion efficiency of this breed. This reveals that the increased fat mass of Mangalica breeds is not the result of mechanisms that increase dietary intake. Rather, their increased adiposity may result from alterations of the genetic programming of tissue development and metabolic pathways that convert dietary energy into fat mass. The observed mtDNA polymorphisms could in part play a role in determining the metabolic efficiency of this breed. In the MT-ND5 sequences of the two Mangalica breeds, polymorphisms exist at amino acids 488 and 584 that are unique to these breeds. If these polymorphisms affect the efficiency of the electron transport chain in even a subtle manner, a lifetime exposed to metabolic differences may lead to altered carcass composition. In order to prove this conclusively, a more thorough population-level analysis correlating phenotype to mitochondrial haplotype will be required. Given the limited number of individuals analyzed herein, it is not possible to exclude the possibility that stochastic events led to the association of these phenotypes with mitochondrial haplotypes.

The mtDNA contributes only 13 proteins to the electron transport chain, with the rest being encoded in the nuclear genome and imported from the cytoplasm. The polymorphisms described here may influence phenotypic characteristics of the breeds evaluated, but the nuclear genes will also influence mitochondrial function. Future studies will aim to provide sequence data for the nuclear genome of these breeds and investigate how altered protein sequences could affect cellular signaling and development. A thorough understanding of these factors will reveal the interplay between the nuclear and mitochondrial genomes

and provide a greater understanding of the cellular mechanisms influencing production traits in these breeds.

The mitochondrial sequences provided could have functional relevance for human health, as the Mangalica pig could prove a useful model for identifying genetic factors that can influence body weight and composition independent of dietary intake. Additionally, the Mangalica breeds could be useful preclinical models for pharmaceuticals targeting obesity.

Our ultimate aim in generating these sequencing data relate to further characterization of differential metabolic capacity in these breeds. Overall, the newly sequenced mtDNAs of Turopolje, Blond Mangalica, Swallow-bellied Mangalica, Yorkshire, and Meishan breeds will assist in studies into developmental and performance characterization of these breeds and as models for human development. Additionally, this knowledge will assist in genetic studies involving mtDNA polymorphisms or maternal lineage tracking.

## ACKNOWLEDGMENTS

We gratefully acknowledge the assistance and consideration provided by Brian Anderson, W. Frank Owsley, F.F. Bartol (Auburn University), and M. Đikić (University of Zagreb).

## FUNDING

Funding for this project was provided by the Auburn Agricultural Experiment Station (AAES), NSF, USDA, and Auburn University.

## ACCESSION NUMBERS

Newly sequenced mtDNA sequences are available through GenBank, accession numbers JN601066 through JN601075.

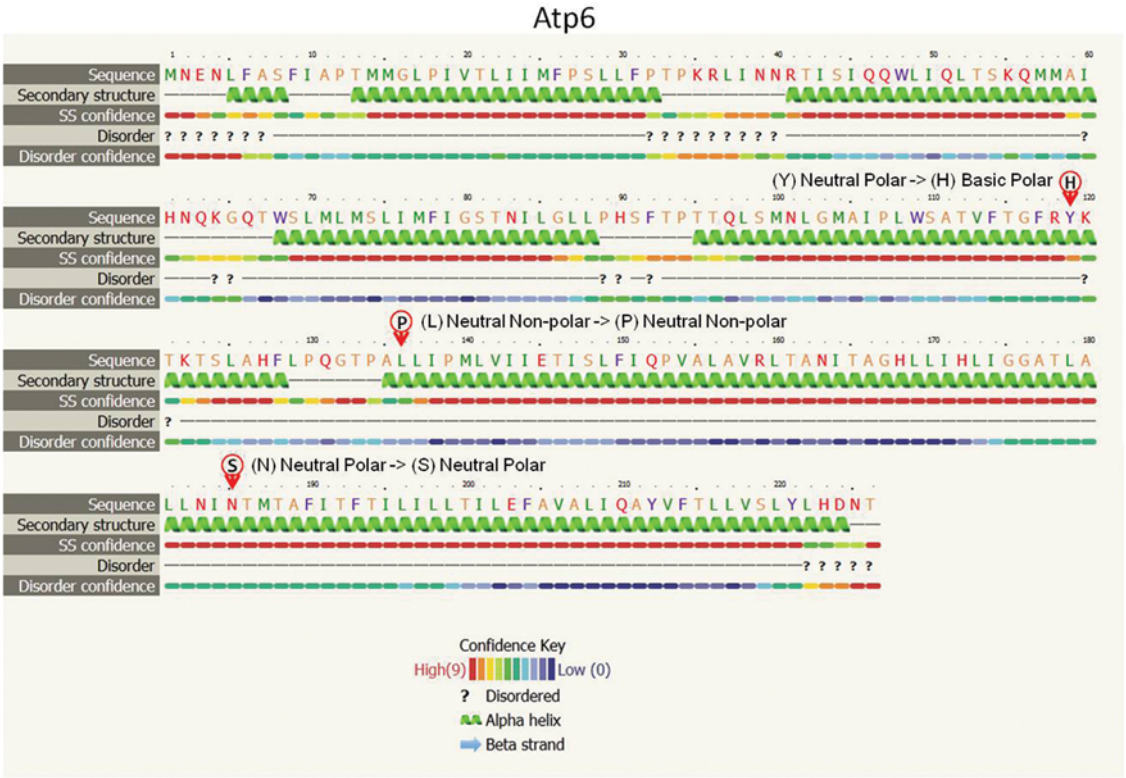
## REFERENCES

1. Pinkert CA. Genetic engineering and competitiveness of livestock production. Agric Conspec Sci 2003; 68:45–54.
2. Pinkert CA, Smith LC, Trounce IA. Transgenic Animals: Mitochondrial Genome Modification. Encyclopedia of Biotechnology in Agriculture and Food. Taylor & Francis, Boca Raton, FL; 2010; 619–621.
3. Cannon MV, Dunn DA, Irwin MH, et al. Xenomitochondrial mice: investigation into mitochondrial compensatory mechanisms. Mitochondrion 2011; 11:33–39.
4. Egerszegi I, Rátky J, Solti L, Brüssow K. Mangalica: an indigenous swine breed from Hungary (Review). Arch Tierz 2003; 46:245–256.
5. Harcet M, Dikic M, Gamulin V. Low genetic diversity of the Turopolje pig breed. Food Technol Biotechnol 2006; 44:105–109.
6. Wallace DC. Mitochondrial diseases in man and mouse. Science 1999; 283:1482–1488.
7. Bottje WG, Carstens GE. Association of mitochondrial function and feed efficiency in poultry and livestock species. J Anim Sci 2009; 87(Suppl):E48–63.
8. Bell BR, McDaniel BT, Robison OW. Effects of cytoplasmic inheritance on production traits of dairy cattle. J Dairy Sci 1985; 68:2038–2051.
9. Tess MW, Reodecha C, Robison OW. Cytoplasmic genetic effects on preweaning growth and milk yield in Hereford cattle. J Anim Sci 1987; 65:675–684.
10. Blankenberg D, Von Kuster G, Coraor N, et al. Galaxy: a web-based genome analysis tool for experimentalists. In: Ausubel FM et al. (ed). *Current Protocols in Molecular Biology*. Hoboken: Wiley; 2010; Ch. 19: Unit 19.10.1–21.
11. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 2010; 11:R86.
12. Li H, Durbin R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. Bioinformatics 2009; 25:1754–1760.
13. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009; 25:2078–2079.
14. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011; 29:24–26.
15. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acid Res 2002; 30:3059–3066.
16. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. brief Bioinformatics 2008; 9:286–298.
17. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 2011; 28:2731–2739.
18. Kelley LA, Sternberg MJ. Protein structure prediction on the web: a case study using the phyre server. Nat Protocols 2009; 4:363–371.
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215: 403–410.
20. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011; 7:539.
21. Pfaffl MW. The ongoing evolution of qPCR. Methods 2010; 50:215–216.
22. Chang WH, Chu HP, Jiang YN, et al. Genetic variation and phylogenetics of Lanyu and exotic pig breeds in Taiwan analyzed by nineteen microsatellite markers. J Anim Sci 2009; 87:1–8.
23. Fang M, Hu X, Jiang T, et al. The phylogeny of Chinese indigenous pig breeds inferred from microsatellite markers. Anim Genet 2005; 36:7–13.
24. Groenen MA, Archibald AL, Uenishi H, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 2012; 491:393–398.

25. Kim TH, Kim KS, Choi BH, et al. Genetic structure of pig breeds from Korea and China using microsatellite loci analysis. J Anim Sci 2005; 83:2255–2263.
26. Liu G, Liu Y, Zhang H, Huang J, Fang M. Genetic variations and sequences analysis of MTATP6 and MTATP8 genes among different Chinese pig breeds. J Anim Breeding Genet (Zeitschrift fur Tierzuchtung und Zuchtungsbiologie). 2010; 127:474–480.
27. Luetkemeier ES, Sodhi M, Schook LB, Malhi RS. Multiple Asian pig origins revealed through genomic analyses. Molec Phylogen Evol 2010; 54:680–686.
28. Tao J, Qin ZQ, Tao Y, et al. Genetic relationships among Chinese pigs and other pig populations from Hunan Province, China. Anim Genet 2007; 38:417–420.
29. Ojano-Dirain C, Tinsley NB, Wing T, Cooper M, Bottje WG. Membrane potential and $H_2O_2$ production in duodenal mitochondria from broiler chickens (Gallus gallus domesticus) with low and high feed efficiency. Comp Biochem Physiol Part A 2007; 147:934–941.
30. Bottje W, Brand MD, Ojano-Dirain C, Lassiter K, Toyomizu M, Wing T. Mitochondrial proton leak kinetics and relationship with feed efficiency within a single genetic line of male broilers. Poultry Sci 2009; 88:1683–1693.
31. Bottje W, Iqbal M, Tang ZX, et al. Association of mitochondrial function with feed efficiency within a single genetic line of male broilers. Poultry Sci 2002; 81:546–555.
32. Fang M, Andersson L. Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. Proc Biol Sci 2006; 273:1803–1810.
33. Curwen EC. Early agriculture in Denmark. Aniquity 1938; 12:135–153.

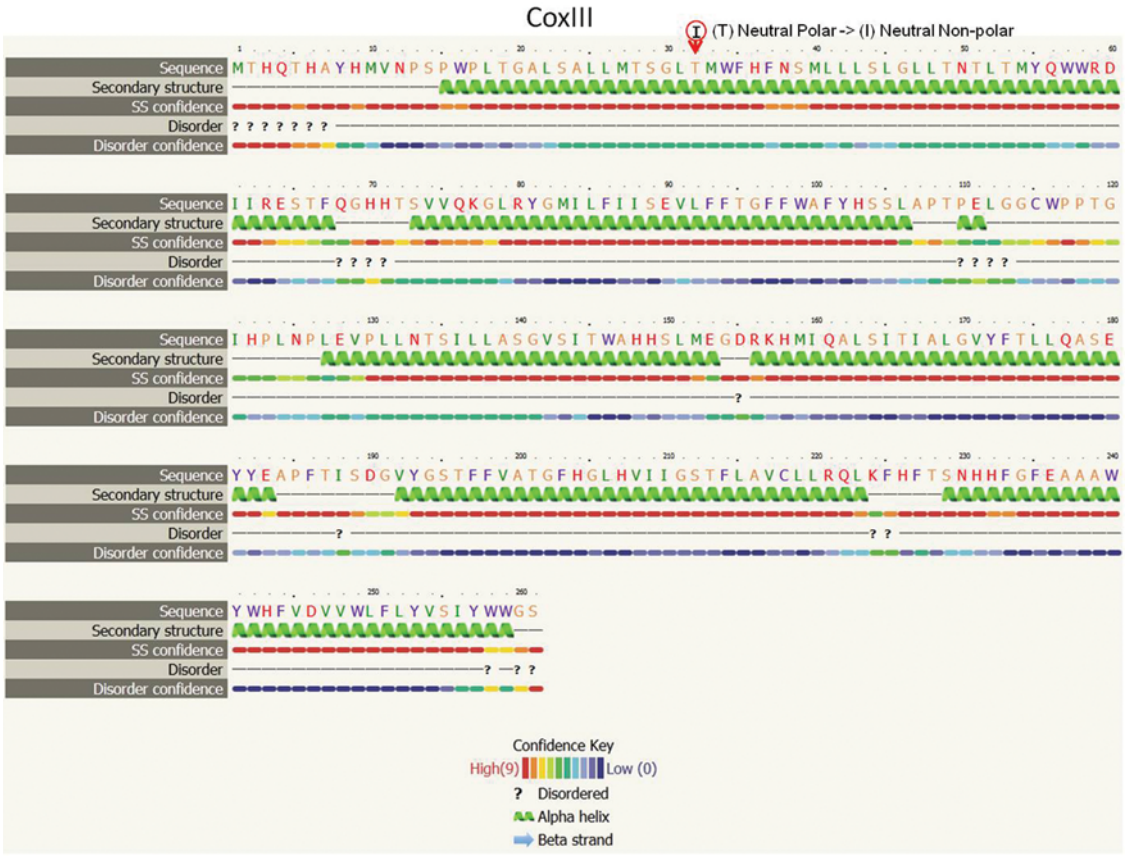


## SUPPLEMENTAL FIGURES




SUPPL. FIG. 1. Secondary structure of Large White MT-ATP6 as determined using the Phyre2 protein threading program. Circled amino acid replacements are present in Meishan and Yorkshire samples.


SUPPL. FIG. 2. Secondary structure of Large White MT-CO1 as determined using the Phyre2 protein threading program. Circled amino acid substitution is present in the Yorkshire sample B.




SUPPL. FIG. 3. Secondary structure of Large White MT-CO3 as determined using the Phyre2 protein threading program. Circled amino acid replacement is present in Meishan and Yorkshire samples.


SUPPL. FIG. 4. Secondary structure of Large White MT-CYTB as determined using the Phyre2 protein threading program. Circled amino acid replacements at 295 and 314 are present in Meishan and Yorkshire samples. Circled amino acid replacement at 198 is present only in Large White.