animals

MDPI

Article

Insights into Nuclear Mitochondrial Sequence Distribution in the Pig Genome Based on the Latest Reference Assembly

Hongtao Li, Cheng Yang, Guiming Zhu, Qin Zhang, Chao Ning * and Dan Wang *

College of Animal Science & Technology, Shandong Agricultural University, Taian 271000, China; 2023110343@sdau.edu.cn (H.L.); 2020010090@sdau.edu.cn (C.Y.); zhuguiming1929@163.com (G.Z.); qzhang@sdau.edu.cn (Q.Z.)

* Correspondence: ningchao@sdau.edu.cn (C.N.); wangd18@sdau.edu.cn (D.W.)

Simple Summary

During evolution, fragments of mitochondrial DNA occasionally integrate into the nuclear genome, forming nuclear mitochondrial sequences (NUMTs). Pigs are among the earliest domesticated livestock species and display extensive breed diversity. This study used the latest pig genome assembly to systematically find and analyze the distribution and characteristics of the NUMTs in the pig genome. We identified a total of 513 high-quality NUMTs distributed across the chromosomes. Our analyses revealed that these NUMTs were not generated randomly; instead, they preferentially originated from specific regions of the mitochondrial genome, and their nuclear insertion sites were associated with particular repetitive sequences. By leveraging a more accurate genome assembly, this study detected substantially more NUMTs than earlier versions of the pig genome, providing a more comprehensive landscape of NUMT integration. These findings enhance our understanding of pig evolutionary history and offer a valuable foundation for the conservation and utilization of pig genetic resources.

Abstract

Horizontal transfer of mitochondrial DNA into the nuclear genome generates nuclear mitochondrial sequences (NUMTs), which serve as molecular fossils reflecting long-term mitochondrial–nuclear interactions and genome evolution. However, the biological mechanisms governing NUMT integration, retention, and evolutionary fate remain incompletely understood in domesticated animals. Here, using the latest pig reference genome assembly (Sscrofa11.1), we present a comprehensive genome-wide characterization of NUMTs in pigs and provide new insights into their genomic distribution and evolutionary constraints. We identified 513 high-confidence NUMTs, of which 460 were chromosomally mapped, accounting for 0.0106% of the nuclear genome. Beyond increased detection, our analyses reveal that pig NUMTs exhibit non-random origins, preferentially integrate into genomic regions under weak selective constraint, and are frequently associated with repetitive elements, consistent with a DNA repair-mediated insertion mechanism. NUMTs predominantly occur as short, fragmented sequences and show signatures of long-term neutral evolution, while insertions disrupting coding sequences are strongly selected against. Synteny-based analyses further identified clustered NUMT regions and duplicated NUMTs, suggesting secondary genomic duplication events following initial integration. Comparative analysis with the earlier Sscrofa10.2 assembly demonstrates that improved genome quality substantially enhances NUMT detection, particularly in repetitive and GC-rich regions, clarifying previously ambiguous sequence-context associations. Together, this high-quality pig NUMT map provides a robust foundation for future functional, evolution-

Check for updates

Received: 17 December 2025
Revised: 11 February 2026
Accepted: 9 March 2026
Published: 14 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

ary, and population-level investigations and contributes to the conservation and utilization of pig genetic resources.

Keywords: NUMTs; pigs; genome; mitochondrial DNA

1. Introduction

Nuclear DNA sequences of mitochondrial origin (NUMTs) represent mitochondrial DNA (mtDNA) sequences that have transferred into the cell nucleus and integrated onto chromosomes [1]. The formation of NUMTs is an indispensable mechanism in the evolution of eukaryotes and their genomes [2–5]. Although NUMTs share high homology with mitochondrial sequences, their evolution proceeds at a significantly slower rate than mtDNA due to the slower evolutionary pace of nuclear DNA (nDNA). NUMTs carry more ancestral mitochondrial genetic information, serving as "molecular fossils." They link maternally inherited mtDNA with Mendelian-inherited nDNA polymorphisms, forming crucial genomic coevolution [6]. Their presence–absence variation (PAV) can reveal overlapping biogeographic distribution histories between populations or varieties. These advantages make NUMTs a powerful tool in species evolution research [7].

In 1994, Lopez et al. first discovered and named NUMTs in their mitochondrial studies of felids [8]. To date, NUMTs have been discovered in hundreds of species, including animals, plants, and fungi, indicating that NUMTs are widely present in the genomes of eukaryotes. The distribution of NUMTs within the genome exhibits certain characteristics. In the human genome, the frequency of transposons within flanking sequences of NUMTs is significantly higher than the genome-wide average, with comparable frequencies upstream and downstream. However, no single transposon class—SINE, LINE, or LTR—exhibits a distinct difference. There is also no significant difference in the frequency of tandem repeats in NUMT flanking sequences and in the genome. The AT oligomer content in the flanking regions of NUMTs is higher than at the genomic level. Fewer NUMTs are generated by mitochondrial sequence translocation from the D-loop region [1].

Research on NUMTs in the pig genome remains limited. Schiavo et al. [9] analyzed genome-wide NUMT sequences in the pig Sscrofa10.2 reference genome, detecting a total of 430 NUMTs resulting from 246 mitochondrial sequence insertion events. The total length of these NUMTs covered 0.0078% of the genome, with their distribution being proportional to chromosome length. The longest NUMT sequence was located on chromosome 2, reaching a length of 111 kb. Detection results for NUMTs may vary across different genome versions [10]. The reference genome Sscrofa10.2 contains errors in completeness, sequence redundancy, order of sequence clusters, and orientation, which may lead to erroneous detection of NUMTs [9,11]. Therefore, this study utilizes the latest version of the pig reference genome from the NCBI to detect NUMTs in the pig genome, analyze their distribution characteristics, and lay the foundation for the conservation and utilization of pig germplasm resources.

2. Materials and Methods

2.1. Pig Reference Genome Sequence

The latest pig reference genome assembly (Sscrofa11.1, GenBank assembly accession: GCA_000003025.6) was obtained from the Ensembl database and used for the identification of NUMTs in pigs.

2.2. Detection of NUMTs in the Pig Genome

2.2.1. Obtain Common Mitochondrial Sequences Shared by *Sus scrofa*

Considering the independent domestication origins of pigs in Europe and Asia [12] and to ensure comparable base dosage of European and Asian pigs in the shared sequences, mitochondrial genome sequences from 14 European and 14 Asian pigs were retrieved from public databases (Table S1). Multiple sequence alignments were generated using MAFFT software (v7.487) [13], and the mitochondrial consensus sequence for modern Eurasian pigs was obtained using CONS software (https://www.ebi.ac.uk/jdispatcher/msa/emboss_cons, accessed on 16 December 2025) [14].

2.2.2. Avoid the Marginal Effects of mtDNA

Because the mitochondrial genome is circular in structure, if it is broken off at the start position of the control region and directly used for homology alignment, the levels of corresponding NUMTs at both ends of the mitochondrial genome may be underestimated. The mitochondrial genome sequence was linearized by joining its 5′ and 3′ ends to generate a linear mtDNA of double the original length. This linear mtDNA was then aligned against nDNA to identify NUMT sequences.

2.2.3. Shielding Repetitive Sequences

When aligning two homologous sequences, simple repetitive sequences tend to overestimate similarity. Therefore, masking repetitive sequences or low-complexity sequences before performing homology alignment can improve the accuracy of NUMT detection. First, construct a repeat sequence library using RepeatModeler software (v2.0.4). Then, combine it with the RepeatMasker.lib library included with RepeatMasker software (v4.1.5) to filter out as many potentially interfering simple repeat sequences as possible.

2.2.4. Homologous Alignment

Adjusting parameters for homology-based alignment can capture "ancient NUMTs" that exhibit lower similarity to modern mtDNA and were inserted into the nuclear genome at an earlier time. To facilitate distant homology detection, i.e., to identify early-arising NUMTs, the scoring system for base substitutions and gaps was revised such that matches were assigned +1, mismatches −1, gaps −7, and gap extensions −1. The NUMT sequences were detected in the pig genome through distant source comparison and their size, number, and density were quantified.

2.2.5. Detection of False Positives

To assess the risk of false positive alignments, a decoy test method was employed. Since completely inverted mitochondrial genome sequences cannot arise through evolution, alignments obtained by comparing inverted mitochondrial sequences with the nuclear genome should yield false positives [9]. This method was used to detect false positives within NUMTs, utilizing the strictest E-value threshold to ensure a low false positive rate.

2.3. Analysis of the Distribution Characteristics of NUMTs

2.3.1. Distribution of NUMTs on Chromosomes

We annotated the physical locations and lengths of NUMTs within the nuclear genome and calculated the number and total length of NUMTs on each chromosome. The proportion of each chromosome occupied by NUMTs was calculated by normalizing the total NUMT length to chromosome size. To assess whether NUMT distribution was associated with chromosome length, Pearson's correlation analyses were performed between chromosome

size and (i) the total length of NUMTs and (ii) the number of NUMTs per chromosome. Distribution plots and correlation analyses were conducted using R software (v4.4.0).

2.3.2. Mitochondrial "Hotspots" for NUMT Insertion

All NUMTs were mapped onto the mitochondrial reference genome using MAFFT software. For each NUMT, the physical location and length within the mitochondrial genome were annotated. We statistically analyzed the frequency of NUMTs occurring within key mitochondrial regions, including the D-loop, the 13 polypeptide-coding genes, the 2 rRNA-coding genes, and the 22 tRNA-coding genes. Specifically, our statistical analyses encompassed NUMT length distribution, repetitive coverage proportions across the mitochondrial genome, coverage differences among genes of varying functional types, and identification of genes with the highest coverage. These analyses were used to identify potential "hotspot" regions for NUMT occurrence within the mitochondrial genome. In addition, the sequence characteristics of these hotspot regions were examined to explore features associated with elevated NUMT formation.

To quantify NUMT coverage over mitochondrial genes, homology mapping information for each NUMT was normalized by converting reverse-strand alignments to forward-strand orientation. Overlapping NUMT fragments were merged to avoid duplicate counting. Subsequently, a custom Python script (Python 3.10) calculated the proportion of genes covered by NUMTs on chromosomes and the proportion of genes covered by NUMT homologous sequences within mitochondria.

2.3.3. Chromosomal Environment Preference

Based on the repetitive element annotations and positional information provided by the UCSC Genome Browser, we calculated the frequency of repetitive elements within 1000 bp flanking regions on both sides of each NUMT. In addition, the GC content of the 1000 bp upstream and downstream sequences of NUMTs was quantified. These metrics were used to examine the association between NUMT occurrence and specific sequence features and to statistically infer the sequence environment preferences of NUMTs.

2.4. NUMT Collinearity Merging

Mitochondrial DNA segments that insert into nuclear chromosomes may subsequently undergo structural variation along with the host chromosome. In this case, LAST (Local Alignment Search Tool), a local alignment software (v2.32.1) that employs an adaptive seed algorithm for sensitive homology search across entire genomes [15], will resolve a single fragmented NUMT insertion event into multiple discrete homologous matching sequences. Therefore, collinear sequence merging was performed based on the physical location of NUMTs in the reference genome to obtain the NUMT region.

Sequences of adjacent NUMTs with physical positions of less than or equal to 30 kb on the nuclear genome and homologous alignment to physical positions of less than or equal to 2.5 kb on the mitochondrial genome are classified as an NUMT region, which is considered to be the only insertion event that has undergone nuclear DNA insertion or deletion mutations. Among them, NUMTs with physical positions less than or equal to 2.5 kb on the mitochondrial genome are considered to be caused by nDNA deletion mutations; NUMTs with physical positions less than or equal to 30 kb on the nuclear genome are considered to be caused by nDNA insertion mutations.

In addition, there may be complex NUMT regions that have undergone structural mutations such as chromosome duplication and rearrangement due to one or more insertion events, i.e., NUMTs with physical positions of less than or equal to 30 kb on the nuclear genome but much greater than 2.5 kb on the mitochondrial genome.

2.5. NUMT Source Prediction

For each NUMT region, the sequences 200 bp upstream and downstream were extracted and concatenated. Pairwise sequence similarity was then assessed among all concatenated sequences. If two sequences exhibited both a similarity greater than 0.8 and coverage greater than 0.8, the corresponding NUMTs were classified as duplicated NUMTs, indicating that they likely underwent duplication events following chromosomal integration. Otherwise, they were considered unique or non-duplicated NUMTs, representing original insertions derived directly from mitochondrial fragments.

3. Result

3.1. NUMTs in the Pig Genome

The mitochondrial consensus sequence of modern Eurasian pigs was homologously aligned with the Sscrofa11.1 nuclear reference genome (613 sequences in total). A total of 530 NUMT sequences were identified with significant homology, of which 477 were located on chromosomes and 53 on scaffolds (Table S2). No alignments were detected when using the reverse mitochondrial sequence, indicating that no false-positive NUMTs were obtained. To ensure stringent quality control, NUMTs with an E-value > 1×10^{-4} were removed, resulting in the exclusion of 17 sequences. Ultimately, 513 high-quality NUMTs (LAST score > 61) were retained, including 460 located on chromosomes and 53 on scaffolds. Based on the total length of the Sscrofa11.1 nuclear genome (2,501,895,775 bp), NUMTs accounted for 0.0106% of the genome (=266,298 bp/2,501,895,775 bp). Figure 1 illustrates their chromosomal distribution. NUMT sequences showed occasional overlap on chromosomes, mostly involving only a few base pairs. The largest overlapping region spanned 52 bp and occurred on the X chromosome. The distances between adjacent NUMTs were generally small, with more than 90% of intervals shorter than 16 Mb. The maximum observed distance between tandem NUMTs reached 61 Mb (Figure 2).

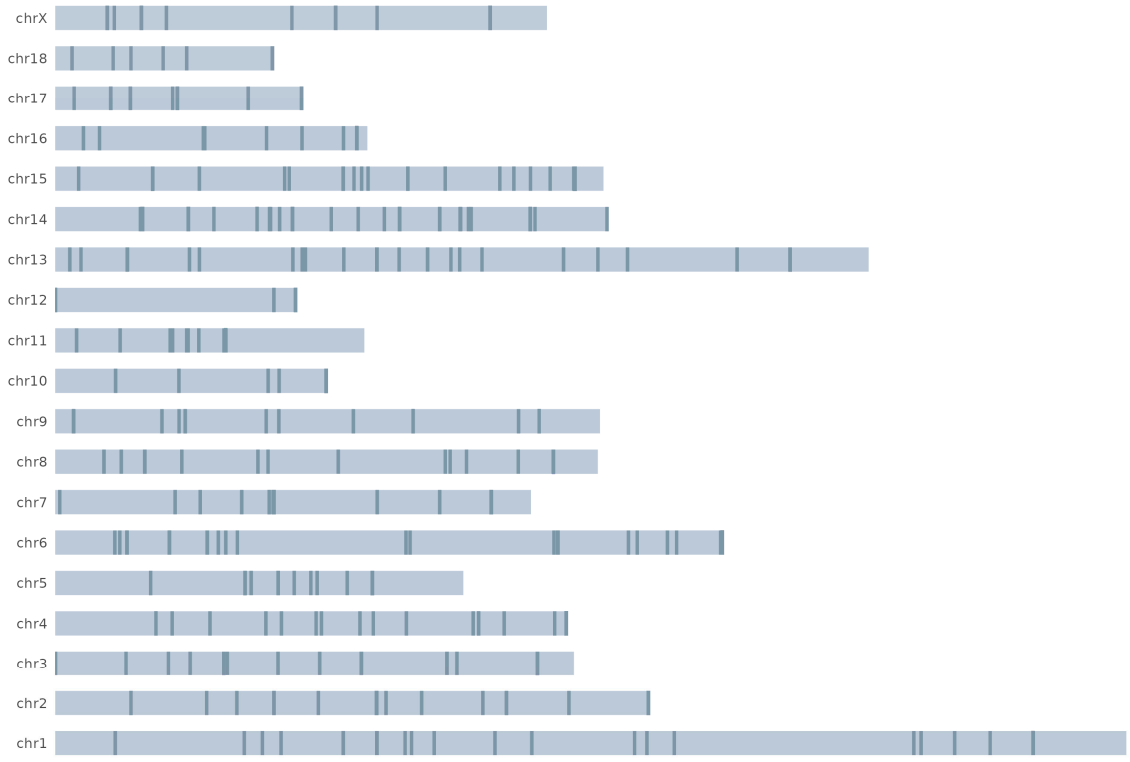

Figure 1. Distribution of NUMTs in the pig nuclear genome. NUMT positions are represented in the Sscrofa11.1 assembled chromosomes.

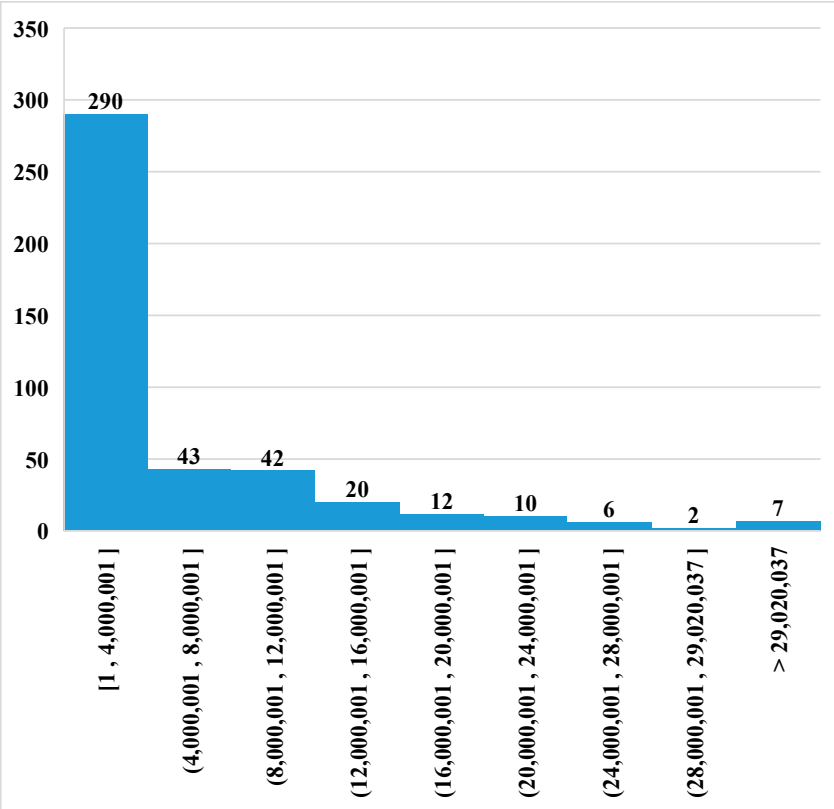

Figure 2. Distance between NUMTs.

The distribution of NUMTs across pig autosomes exhibits significant unevenness, with certain chromosomes (such as chromosome 14) carrying markedly higher numbers and total lengths of NUMTs, while others (such as chromosomes 16 and 18) harbor relatively fewer. A considerable number of NUMTs were also detected on the sex chromosome X (Table S3). There was a moderate strong correlation between the total length of NUMTs on each chromosome and the length of the chromosome (Pearson's r = 0.40, *p*-value = 0.086) and a significant strong correlation between the number of NUMTs on each chromosome and their relative length to the chromosome (Pearson's r = 0.66, *p*-value = 0.002).

The similarity between NUMTs and the mitochondrial consensus sequence of modern Eurasian pigs ranges from 62.28% to 100%, with an average similarity of 76.82%. NUMTs have a wide distribution in length, ranging from 39 to 11,182 bp, with an average length of 519 bp; the vast majority of NUMT sequences are 39 to 4800 bp long, and only one ultra-long NUMT sequence (11,182 bp) appears on chromosome 2, with a similarity of 90.75% to the modern consensus sequence (Table 1).

Table 1. Statistical information on the length of NUMTs and their similarity to mitochondrial homologous sequences.

Number	Size Range (bp)	Similarity Range (%)
66	39–99	75–100
150	100–249	64.62–100
131	250–499	63.24–94.14
68	500–749	62.28–99.81
29	750–999	63.89–100
58	1000–1999	64.02–99.86
10	2000–4999	67.09–91.06
1	>5000	90.75

3.2. Genomic Environment Preferences of NUMTs

The GC content of the flanking sequences of NUMTs was calculated to determine whether there was any difference between the GC content of the entire genome. The results showed that the GC content of the upstream 1000 bp region of NUMTs was 0.4148, and the GC content of the downstream 1000 bp region was 0.4125, which was not much different from the GC content of the genome of 0.4187.

RepeatMasker was used to test whether there was any difference between the frequency of various repetitive elements flanking sequences of NUMTs and the proportion of repetitive elements in the entire genome. The results showed that the repetitive sequence content of the upstream 1000 bp region of NUMTs was 0.4445, and the GC content of the downstream 1000 bp region was 0.4374, which was not much different from the genomic GC content of 0.4404. Counting the number of occurrences of each repeat element, it was found that the three common types of retrotransposons-short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs) and long terminal repeats (LTRs), and simple repeats (Simple_repeat) appeared most frequently on the flank of NUMTs, with 473, 279, 157, and 217, respectively, among which SINE/tRNA, LINE/L1, and Simple_repeat had the most (Table 2).

Table 2. Statistics of the number of repeated elements in upstream and downstream flanking sequences of NUMTs.

Repeat Elements	Number Occurred
DNA/hAT-Blackjack	2
DNA/hAT-Charlie	27
DNA/hAT-Tip100	7
DNA/PiggyBac	1
DNA/TcMar-Tc1	1
DNA/TcMar-Tc2	2
DNA/TcMar-Tigger	7
LINE/CR1	4
LINE/L1	227
LINE/L2	43
LINE/RTE-BovB	2
LINE/RTE-X	3
Low_complexity	24
LTR/ERV1	87
LTR/ERVK	21
LTR/ERVL	25
LTR/ERVL-MaLR	24
None	70
Satellite	2
Simple_repeat	217
SINE/MIR	59
SINE/tRNA	414
snRNA	3
tRNA	47

3.3. NUMT Region

Collinear merging of NUMT sequences yielded 240 NUMT insertion events, i.e., 69 NUMT regions that integrated at least two NUMT sequences (61 NUMT regions were found on chromosomes, 8 on scaffolds, and some NUMT regions are shown in Figure 3), and 171 NUMTs were singleton NUMTs. Among them, 48 NUMT areas are formed by a single insertion event, and 21 are complex insertion events. Most NUMT regions (88.52%) contain fewer than nine NUMT sequences. The number of NUMT regions on

each chromosome was significantly and strongly correlated with the length of NUMTs (Pearson's r = 0.87, *p*-value = 1.72×10^{-6}) and with its relative length (Pearson's r = 0.82, *p*-value = 1.77×10^{-5}). The average distance between adjacent NUMT areas is 28.12 Mb.

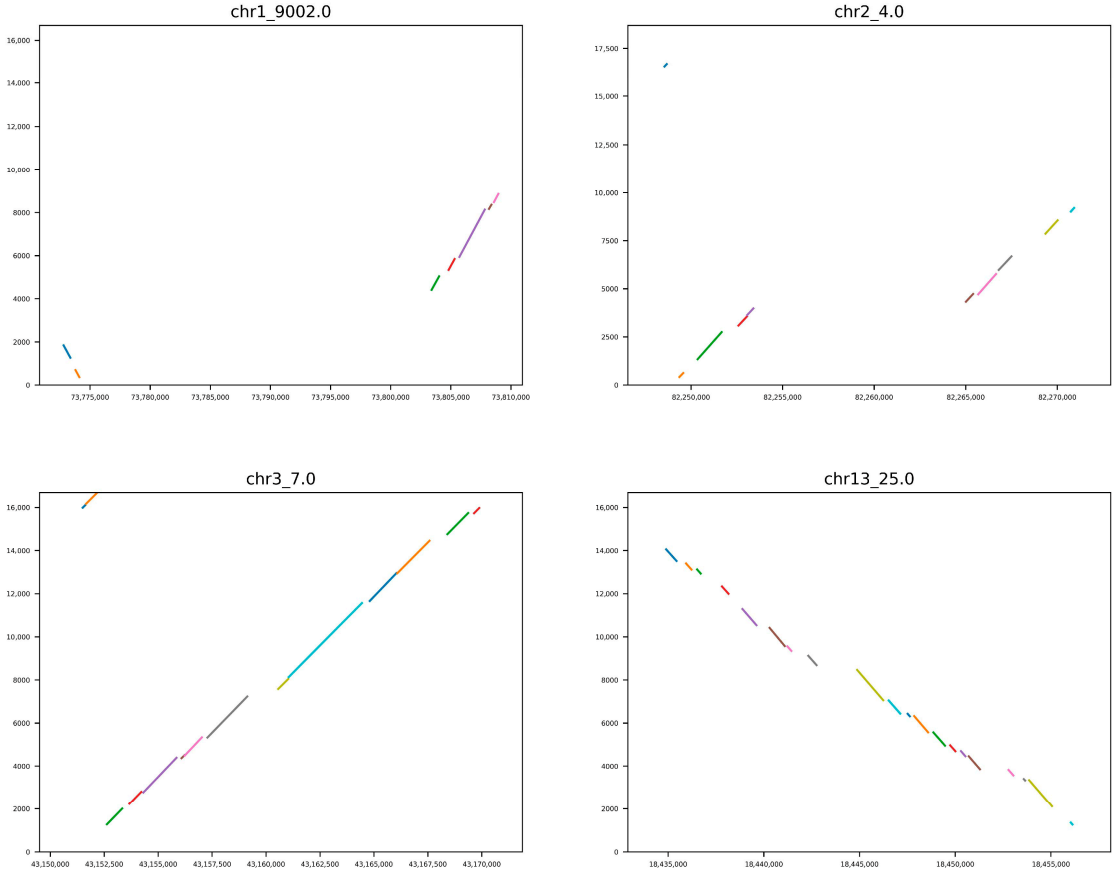

Figure 3. A few NUMT regions. NUMT region ID and numt sequences are reported at the top of each plot. The *x* axis reports the positions in the indicated chromosomes. The *y* axis reports the mtDNA sequence coordinates. Different colored lines in the plots indicate different NUMT sequences. In each panel, the slope of the line reflects the orientation of the NUMT insertion. A positive slope indicates an NUMT region in the same orientation as the mitochondrial reference, while a negative slope (e.g., the NUMT region on chr13) indicates an inverted orientation.

3.4. NUMT Sources

We found that the similarity between five pairs of NUMTs was above 0.8 (considering that the similarity between NUMTs and the mitochondrial consensus sequence of modern Eurasian pigs was at least 62.28%, and the maximum reached 100%, the average similarity was 76.82%), and the coverage was above 0.8 (Table S4).

3.5. Mitochondrial DNA Regions of NUMT Origin

According to the distribution of the mitochondrial homologous sequences of NUMTs in the mitochondrial genome, it can be seen that almost the entire mitochondrial genome is covered by NUMT sequences at least once (Figure 4). The mtDNA regions most covered by NUMTs include parts of the *16SrRNA*, *COX1*, *ND1*, *COX2*, and *ND2* genes, covered more than 50 times.

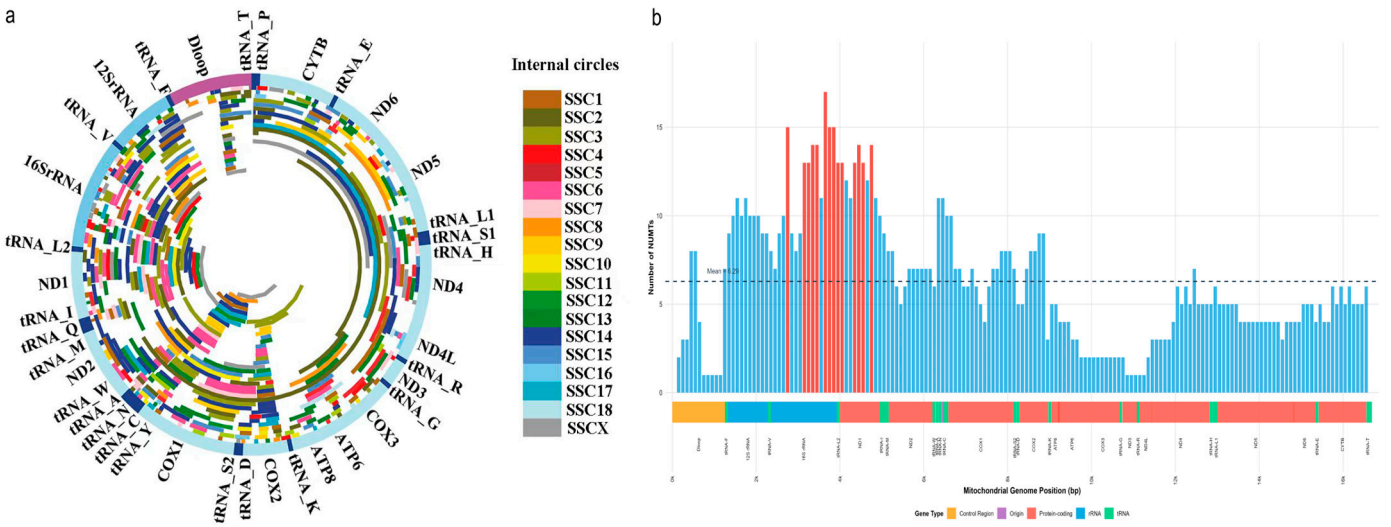

Figure 4. Distribution and insertion frequency of NUMTs in the pig mitochondrial genome. (**a**) Circular map of the pig mitochondrial genome. NUMTs are color-coded according to the pig chromosome into which they are inserted. Mitochondrial genes are displayed in the outer ring region. (**b**) Quantitative distribution of NUMTs along the mitochondrial genome. The *x* axis represents mitochondrial DNA positions. Each bar corresponds to a 100 bp window, with height indicating the number of NUMTs within that window. Red regions denote high-count windows, and dashed lines mark the global mean (Mean = 6.29).

To further analyze the evolutionary constraints on NUMTs after integration into the nuclear genome, we examined their structural features and overlap with functional regions. Structural comparisons of three representative NUMTs (NUMT314, NUMT359, and NUMT445, located at chr14:21808471-21808667, chr14:70688062-70688137, and chrX:15135619-15135851, respectively) located in different nuclear genes revealed that all were fully contained within non-coding regions (introns or UTRs). Compared to their mitochondrial homologs, they retained high structural integrity and sequence similarity (86.7–97.97%) and did not overlap with any exons.

Therefore, we instead compared the distribution proportions of NUMT sequences across two functional genomic regions. Notably, 99.42% of NUMT homologous sequences originated from mitochondrial gene regions, whereas only 30.38% of NUMTs within the nuclear genome were located inside gene regions.

3.6. Comparison with Sscrofa10.2 Genomic NUMTs

Comparing the NUMT sequences identified in the reference genomes of Sscrofa11.1 and Sscrofa10.2, it was found that 513 and 435 NUMT sequences were detected in Sscrofa11.1 and Sscrofa10.2, respectively. The total length of NUMTs is 266,298 bp and 216,323 bp, respectively. Moreover, 69 and 61 NUMT regions containing more than one NUMT sequence were identified, respectively. It can be seen from the distribution map of NUMTs on chromosomes (Figure 5) that the difference in the distribution of NUMTs on the two sets of chromosomes mainly occurs at the ends of chromosomes and other positions.

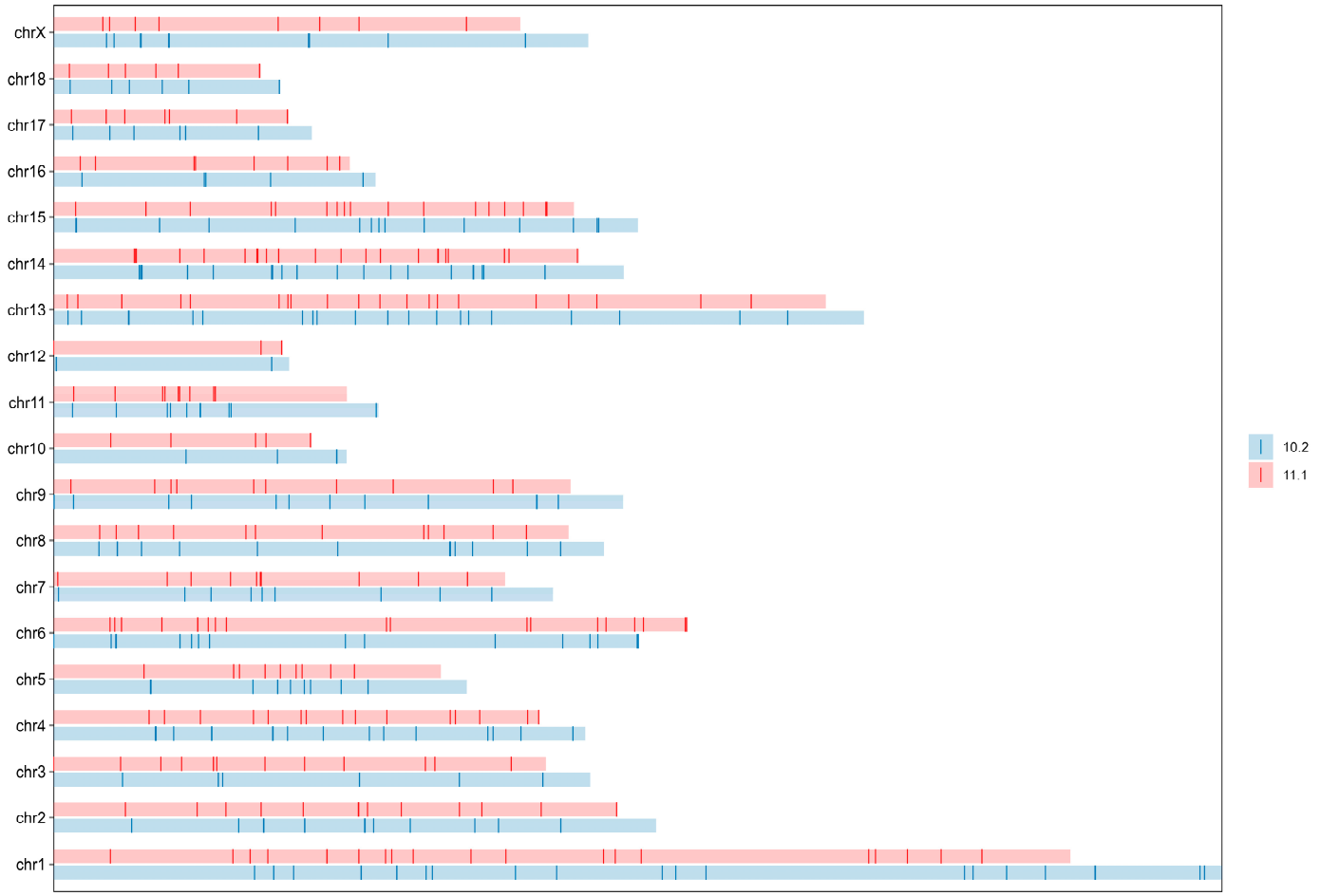

Figure 5. Distribution of NUMTs in Sscrofa10.2 and Sscrofa11.1.

4. Discussion

Nuclear mitochondrial sequences (NUMTs) arise from the transfer of mitochondrial DNA fragments into the nuclear genome, a process widely attributed to DNA double-strand break repair, primarily via non-homologous end joining [16]. In this framework, free mitochondrial DNA released during cellular stress or mitochondrial turnover can be opportunistically integrated into the nuclear genome, a process facilitated by local genomic instability and repetitive or structurally complex regions.

Consistent with this mechanism, the genome-wide NUMT map generated in this study reveals widespread and non-random mitochondrial-to-nuclear DNA transfer in the pig genome. Nearly the entire mitochondrial genome is represented within NUMTs, and the unequal contribution of different mitochondrial regions supports previous observations that NUMT formation frequencies vary among mitochondrial loci [7]. Unlike reports in some primate genomes [1], we did not observe substantial NUMT underestimation associated with the highly mutable D-loop region, suggesting that NUMT detection in pigs is not strongly biased by mitochondrial mutation rate.

Following integration, NUMTs exhibit distinct evolutionary fates depending on their genomic context. The majority of NUMTs inserted into intergenic or non-coding regions experience weak selective constraints and evolve largely neutrally [10], undergoing progressive fragmentation and sequence decay. This is consistent with the predominance of short NUMT fragments and their enrichment in non-coding regions observed in this study. In contrast, NUMT insertions that disrupt coding sequences are strongly selected against, explaining their near absence from exons, while a small subset retained within introns

or untranslated regions may persist and, in rare cases, acquire regulatory or structural roles [17,18]. In repeat-rich regions, secondary duplication events may further amplify certain NUMTs, contributing to local copy number variation [19,20].

Comparative analyses across domestic animals indicate that pigs harbor a relatively high number of NUMTs compared with several other species [21–25]. Differences in NUMT abundance appear to be partially associated with genome size and chromosome length [7,9,26], consistent with previous studies [10], although variation among species with similar genome sizes suggests that additional factors, such as transposon activity and genome turnover rates, also influence NUMT accumulation [27].

Importantly, the use of the latest Sscrofa11.1 reference genome substantially increased NUMT detection compared with earlier analyses based on Sscrofa10.2 [28], highlighting the critical role of reference genome quality in accurately characterizing NUMT landscapes, particularly in repetitive and GC-rich regions [9,11].

Taken together, these observations support a unified conceptual model in which NUMTs originate through DNA repair-mediated integration, are preferentially retained in genomic regions under weak selective constraint, and are progressively shaped by neutral evolution and local duplication events. While evolutionary and domestication-related implications of NUMTs have been proposed in previous studies [9,26,29–31], the present work does not directly test such hypotheses. Interpretations based on NUMT distribution patterns—such as potential signals of interspecific admixture [9] or domestication-related differentiation [30]—should therefore be regarded as literature-informed extrapolations rather than conclusions supported by direct evidence. Accordingly, the primary contribution of this study lies in providing a high-quality, genome-wide NUMT reference map for pigs, together with a conceptual framework for interpretation. Further validation of the proposed evolutionary implications will require population-scale analyses, functional genomic data, and cross-species comparisons in future studies.

5. Conclusions

This study identified genome-wide NUMTs in pigs using the Sscrofa11.1 assembly and characterized their distribution. NUMT length correlated with chromosome length, with hotspots found in retrotransposon-rich regions but not associated with GC content. Structural variation fragmented or duplicated NUMTs after nuclear insertion, and their mitochondrial origins were non-random. These findings provide insights into NUMT evolution and contribute to understanding pig genetic diversity and germplasm conservation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ani16060919/s1, Table S1: Mitochondrial genome information for the construction of a consensus sequence across Eurasian pigs; Table S2: List of identified NUMTs in the pig genome; Table S3: Distribution characteristics of NUMTs in chromosomes; Table S4: Information of duplicated NUMTs.

Author Contributions: Conceptualization, Q.Z., C.N., and D.W.; methodology, Q.Z., C.N., and D.W.; formal analysis, H.L., C.Y., and G.Z.; investigation, H.L., C.Y., and G.Z.; data curation, H.L., C.Y. and G.Z.; writing—original draft preparation, H.L.; writing—review and editing, H.L., C.N., and D.W.; supervision, Q.Z., C.N., and D.W.; funding acquisition, Q.Z., C.N., and D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key R&D Program of China (14th Five-Year Youth Scientist Program, 2022YFF1000900), the National Natural Science Foundation of China (32102526), the Shandong Provincial Natural Science Foundation (ZR2020QC176), the Shandong Provincial Postdoctoral Innovation Talent Support Program, and the Shandong Provincial Key R&D Program (2024LZGC017 and 2022LZGC003).

Institutional Review Board Statement: Not applicable. This study exclusively analyzed publicly available genomic sequence data from online databases and did not involve any direct experiments on live animals or human participants.

Informed Consent Statement: Not applicable. This study did not involve human subjects or individual animals.

Data Availability Statement: All sequence data analyzed in this study are publicly available. The pig reference genome assembly (Sscrofa11.1) was obtained from the Ensembl database. Mitochondrial genome sequences were retrieved from the National Center for Biotechnology Information (NCBI). The specific accession numbers used in this study are provided in Supplementary Table S1.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tsuji, J.; Frith, M.C.; Tomii, K.; Horton, P. Mammalian NUMT insertion is non-random. *Nucleic Acids Res.* **2012**, *40*, 9073–9088. [CrossRef] [PubMed]
2. Gray, M.W.; Burger, G.; Lang, B.F. Mitochondrial evolution. *Science* **1999**, *283*, 1476–1481. [CrossRef]
3. Kleine, T.; Maier, U.G.; Leister, D. DNA Transfer from Organelles to the Nucleus: The Idiosyncratic Genetics of Endosymbiosis. *Annu. Rev. Plant Biol.* **2009**, *60*, 115–138. [CrossRef]
4. Johnston, I.G.; Williams, B.P. Evolutionary Inference across Eukaryotes Identifies Specific Pressures Favoring Mitochondrial Gene Retention. *Cell Syst.* **2016**, *2*, 101–111. [CrossRef]
5. Zhang, G.J.; Dong, R.; Lan, L.N.; Li, S.F.; Gao, W.J.; Niu, H.X. Nuclear Integrants of Organellar DNA Contribute to Genome Structure and Evolution in Plants. *Int. J. Mol. Sci.* **2020**, *21*, 707. [CrossRef]
6. Ladoukakis, E.D.; Zouros, E. Evolution and inheritance of animal mitochondrial DNA: Rules and exceptions. *J. Biol. Res.-Thessalon.* **2017**, *24*, 2. [CrossRef]
7. Uvizl, M.; Puechmaille, S.J.; Power, S.; Pippel, M.; Carthy, S.; Haerty, W.; Myers, E.W.; Teeling, E.C.; Huang, Z.X. Comparative Genome Microsynteny Illuminates the Fast Evolution of Nuclear Mitochondrial Segments (NUMTs) in Mammals. *Mol. Biol. Evol.* **2024**, *41*, msad278. [CrossRef] [PubMed]
8. Lopez, J.V.; Yuhki, N.; Masuda, R.; Modi, W.; O'Brien, S.J. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* **1994**, *39*, 174–190. [CrossRef]
9. Schiavo, G.; Hoffmann, O.I.; Ribani, A.; Utzeri, V.J.; Ghionda, M.C.; Bertolini, F.; Geraci, C.; Bovo, S.; Fontanesi, L. A genomic landscape of mitochondrial DNA insertions in the pig nuclear genome provides evolutionary signatures of interspecies admixture. *DNA Res.* **2017**, *24*, 487–498. [CrossRef] [PubMed]
10. Hazkani-Covo, E.; Zeller, R.M.; Martin, W. Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. *PLoS Genet.* **2010**, *6*, e1000834. [CrossRef]
11. Warr, A.; Affara, N.; Aken, B.; Beiki, H.; Bickhart, D.M.; Billis, K.; Chow, W.; Eory, L.; Finlayson, H.A.; Flicek, P.; et al. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience* **2020**, *9*, giaa051. [CrossRef]
12. Han, Y.; Zhao, Z.H.; Stanton, D.W.G.; Lei, X.L.; Wu, Z.; Liu, Y.T.; Yu, C.; Chen, X.; Li, W.Y.; Wang, J.; et al. Ancient Genomics Reveals the Origin, Dispersal, and Human Management of East Asian Domestic Pigs. *Mol. Biol. Evol.* **2025**, *42*, msaf214. [CrossRef] [PubMed]
13. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef]
14. Madeira, F.; Madhusoodanan, N.; Lee, J.; Eusebi, A.; Niewielska, A.; Tivey, A.R.N.; Lopez, R.; Butcher, S. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res* **2024**, *52*, W521–w525. [CrossRef]
15. Kielbasa, S.M.; Wan, R.; Sato, K.; Horton, P.; Frith, M.C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **2011**, *21*, 487–493. [CrossRef]
16. Ricchetti, M.; Fairhead, C.; Dujon, B. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **1999**, *402*, 96–100. [CrossRef]
17. Golubenko, M.V.; Puzyrev, V.P. Liberties of the genome: Insertions of mitochondrial DNA fragments into nuclear genome. *Vavilovskii Zhurnal Genet. Sel.* **2024**, *28*, 467–475. [CrossRef] [PubMed]
18. Huang, M.; Frith, M.C. Probability-Based Sequence Comparison Finds the Oldest Ever Nuclear Mitochondrial DNA Segments in Mammalian Genomes. In *RECOMB International Workshop on Comparative Genomics*; Springer Nature: Cham, Switzerland, 2026; pp. 247–250.
19. VanBuren, R.; Ming, R. Organelle DNA accumulation in the recently evolved papaya sex chromosomes. *Mol. Genet. Genom.* **2013**, *288*, 277–284. [CrossRef]

20. Lupski, J.R. Charcot-Marie-Tooth polyneuropathy: Duplication, gene dosage, and genetic heterogeneity. *Pediatr. Res.* **1999**, *45*, 159–165. [CrossRef] [PubMed]
21. Liang, B.; Wang, N.; Li, N.; Kimball, R.T.; Braun, E.L. Comparative Genomics Reveals a Burst of Homoplasy-Free Insertions. *Mol. Biol. Evol.* **2018**, *35*, 2060–2064. [CrossRef]
22. Nergadze, S.G.; Lupotto, M.; Pellanda, P.; Santagostino, M.; Vitelli, V.; Giulotto, E. Mitochondrial DNA insertions in the nuclear horse genome. *Anim. Genet.* **2010**, *41*, 176–185. [CrossRef]
23. Ning, F.Y.; Fu, J.; Du, Z.H. Mitochondrial DNA insertions in the nuclear Capra hircus genome. *Genet. Mol. Res.* **2017**, *16*, gmr16018266. [CrossRef] [PubMed]
24. Verscheure, S.; Backeljau, T.; Desmyter, S. In silico discovery of a nearly complete mitochondrial genome Numt in the dog (*Canis lupus* familiaris) nuclear genome. *Genetica* **2015**, *143*, 453–458. [CrossRef] [PubMed]
25. Féménia, M.; Charles, M.; Boulling, A.; Rocha, D. Identification and characterisation of mitochondrial sequences integrated into the ovine nuclear genome. *Anim. Genet.* **2021**, *52*, 556–559. [CrossRef]
26. Yoshida, T.; Furihata, H.Y.; Kawabe, A. Analysis of nuclear mitochondrial DNAs and factors affecting patterns of integration in plant species. *Genes Genet. Syst.* **2017**, *92*, 27–33. [CrossRef] [PubMed]
27. Song, H.J.; Moulton, M.J.; Whiting, M.F. Rampant Nuclear Insertion of mtDNA across Diverse Lineages within Orthoptera (Insecta). *PLoS ONE* **2014**, *9*, e110508. [CrossRef]
28. Groenen, M.A.M.; Archibald, A.L.; Uenishi, H.; Tuggle, C.K.; Takeuchi, Y.; Rothschild, M.F.; Rogel-Gaillard, C.; Park, C.; Milan, D.; Megens, H.J.; et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **2012**, *491*, 393–398. [CrossRef]
29. Marzetti, E.; Di Lorenzo, R.; Calvani, R.; Coelho, H., Jr.; Pesce, V.; Landi, F.; Leeuwenburgh, C.; Picca, A. Nuclear mitochondrial DNA transfer revisited: From genomic noise to hallmark of aging. *Ageing Res. Rev.* **2025**, *112*, 102892. [CrossRef]
30. Bolner, M.; Bovo, S.; Ballan, M.; Schiavo, G.; Taurisano, V.; Ribani, A.; Bertolini, F.; Fontanesi, L. A comprehensive atlas of nuclear sequences of mitochondrial origin (NUMT) inserted into the pig genome. *Genet. Sel. Evol.* **2024**, *56*, 64. [CrossRef]
31. Schiavo, G.; Bovo, S.; Ribani, A.; Kazemi, H.; Fontanesi, L. A comparative genome landscape of mitochondrial DNA insertions into two cattle nuclear genome versions. *Anim. Genet.* **2020**, *51*, 149–151. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.