Remarkably Duplicate Genes as Sources for Rapid Adaptive Evolution of Sperm under Environmental Pollution in Tree Sparrow

Shengnan Wang[1], Yingmei Zhang[1], Wenzhi Yang[1], Yue Shen[1], Zhaocun Lin[1], Sheng Zhang[1], and Gang Song[2]

[1]Lanzhou University School of Life Sciences
[2]Institute of Zoology, Chinese Academy of Sciences

February 22, 2024

Abstract

Environmental pollution can result in poor sperm quality either directly or indirectly. However, adaptive and compensatory sperm morphology change and motility improvement rapidly evolved in tree sparrow (Passer montanus) inhabited the polluted area within the past 65 years. To identify the genetic underpinnings of the rapidly evolved sperm phenotype, both the population genomic and transcriptomic methods were used in our study. We identified a gene encoding serine/threonine protein kinase PIM1 which may drive the rapid phenotypic evolution of sperm. An unprecedent and remarkably expansion of PIM gene family caused by tandem and segmental duplication of PIM1 was subsequently noticed in tree sparrow genome. Most of the PIM1 duplicates showed a testis-specific expression pattern, suggesting their functions related to male reproduction. Furthermore, the elevated expression level of PIM1 was consistent with our earlier findings of longer and faster swimming sperm in polluted site, indicating an important role of duplicated PIM1 in facilitating rapid evolution of sperm. Our results suggested that the duplicated PIM1 provide sources of genetic variation that enable rapid evolution of sperm under environmental heavy metal pollution. The findings in this study verified the duplicated genes can be targets of selection and predominant sources for rapid adaptation to environmental change and shed lights on the sperm evolution under environmental stress.

Remarkably Duplicate Genes as Sources for Rapid Adaptive Evolution of Sperm under Environmental Pollution in Tree Sparrow

Shengnan Wang[1], Yingmei Zhang[1*], Wenzhi Yang[1], Yue Shen[1], Zhaocun Lin[1], Sheng Zhang[1], Gang Song[2]

[1]Gansu Key Laboratory of Biomonitoring and Bioremediation for Environmental Pollution, School of Life Science, Lanzhou University, Lanzhou, 730000, China

[2]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Science, Beijing, 100101, China

*Corresponding author: E-mail:*ymzhang@lzu.edu.cn*

Running title: Duplicate Genes Enables Evolution of Sperm

Abstract

Environmental pollution can result in poor sperm quality either directly or indirectly. However, adaptive and compensatory sperm morphology change and motility improvement rapidly evolved in tree sparrow (*Passer montanus*) inhabited the polluted area within the past 65 years. To identify the genetic underpinnings of the rapidly evolved sperm phenotype, both the population genomic and transcriptomic methods were used in our study. We identified a gene encoding serine/threonine protein kinase PIM1 which may drive the rapid phenotypic evolution of sperm. An unprecedent and remarkably expansion of *PIM* gene family

caused by tandem and segmental duplication of *PIM1* was subsequently noticed in tree sparrow genome. Most of the *PIM1* duplicates showed a testis-specific expression pattern, suggesting their functions related to male reproduction. Furthermore, the elevated expression level of *PIM1* was consistent with our earlier findings of longer and faster swimming sperm in polluted site, indicating an important role of duplicated *PIM1* in facilitating rapid evolution of sperm. Our results suggested that the duplicated *PIM1* provide sources of genetic variation that enable rapid evolution of sperm under environmental heavy metal pollution. The findings in this study verified the duplicated genes can be targets of selection and predominant sources for rapid adaptation to environmental change and shed lights on the sperm evolution under environmental stress.

Keywords: tree sparrow, rapid evolution, sperm, heavy metal pollution, duplicate genes, *PIM1*

Introduction

Human alterations of animal living environment can adversely affect male fertility through affecting spermatogenesis and sperm quality (Aulsebrook et al., 2020; Kumar & Singh, 2022). On the one hand, due to the close connection between male fertility and population growth, if sperm fail to adapt to the ever-changing environment to buffer the detrimental effects caused by environmental change, a decrease in population fitness and subsequent population collapse will happen (Minnameyer et al., 2021; Walsh et al., 2019). On the other hand, the unique evolutionary feature of spermatozoa, which is they undergo direct and strong selection such as sperm competition and cryptic female choice, make sperm have the capacity to evolve rapidly and provide hopes for adaptation to fast-changing environment (Avidor-Reiss, 2018; Birkhead & Pizzari, 2002; Wang & Gunderson, 2022), and a handful of studies have found sperm traits can evolve in a few generations to adapt to environmental stress (Breckels & Neff, 2014; Vasudeva et al., 2019). However, the genetic mechanisms that enable sperm fast adaptation to environmental degradation are poorly known.

Rapid evolution of sperm requires genetic variation within reproductive genes. Gene duplication has long been regarded as an important force speeding up evolution by facilitating increase in genetic and functional diversities and providing new redundant genetic material (Ames et al., 2010; Conant & Wolfe, 2008; Kondrashov, 2012; Lynch & Katju, 2004). It seems that duplicate genes are potential sources which selection can operate with swiftness and without deleterious effects when environment is undergoing rapid changes. However, when most studies focus on the macroevolutionary pattern in gene duplication, little is known of the fate and importance of duplicate genes in the context of rapid adaptation and evolution of sperm in a short time.

Tree sparrow (*Passer montanus*) is an abundant and resident passerine bird in northwest China including Baiyin (BY), a city which is heavily polluted with heavy metals (mainly Cu, Zn, Pb and Cd) resulting from metal mining and smelting (Li et al., 2006; Liu et al., 2016; Nan & Zhao, 2000) for more than 60 years. Sperm traits are known to be sensitive to heavy metal exposure which can result in decreases of sperm count (Akinloye et al., 2006), vability (Chiou et al., 2008; Tabassomi & Alavi-Shoushtari, 2013), motility (Hardneck et al., 2018; Kushawaha et al., 2021) and velocity (Hardneck et al., 2018; Naha & Chowdhury, 2006). Nevertheless, relative to the tree sparrow live in Liujiaxia (LJX), a nearby tourism city selected as a control unpolluted site whose ecological environmental is similar with BY, longer and faster swimming sperm which has been proven to be beneficial for reproductive success have been observed in BY tree sparrow population (Yang et al., 2020a, 2020b). In addition, tree sparrows from the BY population showed significantly higher extra-pair fertilization than LJX population indicating that the sperm competition may be more intense in heavy metal polluted environment (Yang et al., 2021). The unexpected phenotypic variation of sperm and the selective pressures caused by intense sperm competition may indicates that the sperm traits of tree sparrow in BY fast evolved, and this speculation motivates us to explore the molecular basis of sperm phenotype accommodation to the heavy metal pollution through population genomic and transcriptomic. We then identified remarkably duplicate genes are predominant sources driving rapid evolution of sperm in tree sparrow under environmental heavy metal pollution. The results are expected to provide a reference for further exploring the role of duplicate genes in fast adaptation to environmental change and the function of sperm evolution in animal population development under environmental stress.

Materials and Methods

Study sites and sampling

All animal collections and experiments were approved by the Committee on the Ethics of Animal Experiments of School of Life Sciences of Lanzhou University. Adult male tree sparrows were caught by mist nets in our study sites, Baiyin (BY) and Liujiaxia (LJX), during their breeding season (April-July) from 2020 to 2022 (Table S1). The polluted site BY has a long history of mining non-ferrous metals since 1956, in contrast with LJX, a tourism city and important water resource protection area that is relatively unpolluted and chosen as control site. The tree sparrow reaches sexual maturity within a year after hatching and breeds every single year in our study sites. The sexes of individuals were determined based on the presence of an incubation patch (Bailey, 1952). Afterward, females were ringed with a uniquely numbered metal band and released, while males were kept from light and transported alive to the laboratory for further experiments. In the laboratory, males were euthanized and dissected immediately. Their tissues were collected and frozen in liquid nitrogen, and then stored at -80 until subjected to DNA or RNA extraction.

Population genetic structure and differentiation

DNA was extracted from liver sample of 22 male tree sparrow, a half from BY and the other from LJX, using the Qiagen DNeasy Blood and Tissue Kit. DNA concentration (minimum of 80 ng/μL) was measured using Qubit DNA Assay Kit in Qubit 2.0 Flurometer (Life Technologies, CA, USA). After sample quality control, the qualified DNA samples are randomly fragmented by Covaris and the fragments are collected by magnetic beads. Adenine are added to 3'end of end-repaired DNA fragments before adaptor ligation. The ligation products are then cyclized and then amplified by linear isothermal Rolling-Circle Replication and DNA NanoBall technology. Sequencing of these DNA libraries is performed with BGISEQ sequencing platform.

SOAPnuke v2.1.0 (Chen et al., 2018) was used to remove any remaining adapter and trim low quality reads. The clean reads were mapped to reference genome using Burrows-Wheeler Alignment tool (BWA) v0.7.17 (Li & Durbin, 2009) with high mapping rates (>99%). Properly paired ratio of samples varies from 97.90% to 98.45 and the effective mapping depth is between 9.6145 × to 12.2982 ×. After alignment, we used Picard v2.25.0 to remove duplicate reads arisen during sample preparation e.g. library construction. We used Genome Analysis Toolkit (GATK) v4.2.0.0 (McKenna et al., 2010) for variant calling and hard-filtering followed different filtering thresholds for SNPs (QD < 2.0, QUAL < 30.0, SOR > 3.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0) and indels (QD < 2.0, QUAL < 30.0, FS >200.0, ReadPosRankSum < -20.0) separately. The obtained variant calls were annotated by ANNOVAR (Wang et al., 2010) (Table S3).

We used PLINK v1.90 (Purcell et al., 2007) to remove SNPs with a minor allele frequency (MAF) [?] 0.05 and genotyped in less than 80% samples. We then calculated the pairwise linkage disequilibrium (LD) between the left SNPs and removed one SNP from each pair of neighbor markers whose pairwise r^2 is higher than 0.2 using PLINK and the remaining SNPs were used for subsequent analysis. We constructed a maximum-likelihood tree using RAxML-NG v1.0.3 (Kozlov et al. 2019) and the constructed phylogenetic tree was visualized using an online tree visualization tool iTOL v6.4.3 (Letunic & Bork, 2021). Principal component analysis was performed using PLINK and individual ancestries were estimated using ADMIXTURE v1.3.0 (Alexander et al., 2009) with assumed number of populations (K) ranging from 1 to 4.

Identification of genomic regions under selection

We used VCFtools v0.1.16 (Danecek et al., 2011) to calculate Tajima's D in 50 kb non-overlapping windows and nucleotide diversity (π) and population-differentiation statistic (F_{ST}) for the two populations both in 50 kb sliding windows with 10 kb steps. To identify regions of the genome that may have been subject to selection associated with sperm variation of tree sparrow living in BY, we detected the regions with both high divergence and low diversity, characteristic of recent selective sweeps. We only selected the top 1% of outliers with low diversity in BY population estimated by the pairwise differences (π_{LJX}/π_{BY}) converted to

Z scores and high divergence estimated by F_{ST}.

Transcriptome sequencing, alignment and quantification

RNA from three tissues (liver, spleen, testis) was extracted from 7 male tree sparrows from BY and 7 male tree sparrows from LJX. RNA sequencing was performed based on 150 bp paired-end reads with insert size around 350 bp using Illumina NovaSeq 6000 platform. After filtering and quality control procedures, the clean reads were mapped to reference genome using STAR v2.7.9a (Dobin et al., 2013).

The gene-level quantification approach utilizes the files of aligned reads generated during alignment and a gene transfer format (GTF) file containing gene models of tree sparrow to count the number of reads that mapped to each gene using featureCounts v2.8.1 (Liao et al., 2014). We obtained the resulting fragments per kilobase of exon per million fragments mapped (FPKM) and transcripts per million (TPM). We used the trimmed mean of the M-values (TMM) normalization to compare gene expression values among replicates from the same tissue.

Identification of differentially expressed genes in testis

We detected differentially expressed genes (DEGs) of testis using both DESeq2 v1.34.0 (Love et al., 2014) and edgeR v3.36.0 (Robinson et al., 2010) to compare expression levels between the BY and LJX tree sparrows. We used a threshold value of fold change of 2 and adjusted *P*value (Benjamin-Hochberg correction in DESeq2 and FDR correction in edgeR) of 0.05 for differential expression analysis. Only genes identified by both DESeq2 and edgeR were considered to be DEGs and were used for further analysis.

Identification and characterization of PIM gene family in tree sparrow and zebra finch

The PIM protein sequences of human (PIM1, PIM2 and PIM3) and chicken (PIM1 and PIM3) were obtained from UniProt database and then used as query to search against the tree sparrow and zebra finch (downloaded from NCBI database) protein dataset using the BLASTP program in BLAST+ (Camacho et al., 2009) with and e-value of 1e-5 and identity of 50% as the threshold. We used the Hidden Markov Model (HMM) of protein kinase domain (PF00069) to reconfirm the predicted PIM protein sequences using HMMER v3.3.2 (Eddy, 2011). Furthermore, the putative PIM sequences with incomplete protein kinase domain which related with the protein kinase catalytic activity were discarded by searching manually using online HmmerWeb v2.41.2 (Potter et al., 2018). Ultimately, there are 6 PIM sequences left for zebra finch and 143 for tree sparrow. The molecular weight (MW) and isoelectric points (PI) of PIM proteins of tree sparrow were calculated by Peptides v2.4.4 (Osorio et al., 2015). We used FastTree v2.1.10 (Price et al., 2010) to construct a ML tree and MEME v5.4.1 (Bailey et al., 2015) to analyze the conserved motif structures of a total of 154 full-length amino acid sequences of PIM family (2 of chicken, 3 of human, 6 of zebra finch and 143 of tree sparrow). The combination of constructed ML tree and predicted motif structure of PIM family were visualized by iTOL.

Chromosomal distribution, gene duplication, synteny analysis, expression profiles and copy number variation

The chromosome distribution of PIM1 and adjacent genes were obtained from the genome annotation information and visualized by TBtools v1.098685 (Chen et al., 2020). Gene duplication events were identified manually. Segmental duplication events were characterized as copying the whole blocks of genes from one chromosome to another while tandem duplication was defined as paralogous genes located within 50kb in tandem and was separated by fewer than five nonhomologous spacer genes (Cannon et al., 2004). The collinear blocks were detected by JCVI v0.7.5 (Tang et al., 2008) and visualized by Circos v0.69.8 (Krzywinski et al., 2009).

Duplication of PIM gene were explored in other avian species. The*PIM* gene family were identified in 47 avian genomes released by Ensembl covered 12 avian orders (Accipitriformes, Anseriformes, Apterygiformes, Casuariiformes, Charadriiformes, Falconiformes, Galliformes, Passeriformes, Psittaciformes, Strigiformes, Struthioniformes, Tinamiformes) by using the same method as the identification of *PIM* in tree sparrow.

Then according to the completeness of pkinase domain, the identified *PIM* genes in all avian species were divided into two categories: complete and incomplete.

The expression patterns of the chicken and tree sparrow *PIM1* in eight tissues (brain, heart, kidney, liver, lung, muscle, spleen, testis) were analyzed following the method described in transcriptome analysis. Except for the liver, spleen and testis of tree sparrow were sequenced in our study, the other transcriptome data were downloaded from NCBI sequence read archive (SRA) database. The expression heatmap was generated by ComplexHeatmap v2.10.0 (Gu et al., 2016) package in R.

The expression level and copy number of the tree sparrow *PIM1* between LJX and BY were compared. We calculated the read depth per *PIM1* gene and per individual by using bedtools v2.30.0 (Quinlan & Hall, 2010). The read depth was normalized according the gene length and the total reads aligned to the assembly genome per sample. Large changes in read depth might indicate copy number variation between two population.

Inhibition of PIM1 kinase

Four male tree sparrows captured from LJX in 2022 were dosed daily by oral gavage at 100 mg/kg of SGI-1776, an experimental PIM1 kinase small molecule inhibitor, and the same numbers of LJX males were used as control group. SGI-1776 at 100 mg/kg was previously shown to be effective in abrogating tumor growth in mice (Chen et al., 2011). After 1 week of drug treatment was completed, all birds were euthanized and dissected. Total RNA was extracted from testis and used for sequencing. Their left seminal glomera were extracted and placed in 500 μL pre-warmed (40) Hank's Balanced Salt Solution with an incision to obtain sperm suspensions (Yang et al, 2019). After 3 minutes dilution, the sperm suspensions were used to assess curvilinear velocity as an index of sperm velocity using a computer-assisted sperm analysis (CASA) system. Then the sperm suspensions were placed on a glass slide and smeared using another slide. The air-dried smeared slides were fixed with 5% formalin for 10 min and stained using Wright-Giemsa staining solution. The slides were used to assess sperm length at 400 × magnification.

Results

Identification of candidate gene associated with tree sparrow sperm evolution under environmental pollution

22 male individuals, a half from BY and the other from LJX, with an average coverage of ˜11× (Table S2) were sequenced. As the genome-wide differentiation between the two populations is low (F_{ST} = 0.0058; Figure 1A; Figure S1), the BY population shows reduced nucleotide diversity and more compactly clustered in biplot of principal component analysis (Figure 1B; Figure S2), possibly reflecting some genomic regions are under strong recent selection in BY population. We search for genome-wide signatures of selection related to sperm variation in BY tree sparrows through pairwise genetic differentiation (F_{ST}) and differences in nucleotide diversity (π) between BY and LJX populations in 50 kb sliding windows with 10 kb steps. We define the 1% most differentiated regions as outliers for BY population (Figure 2C; Figure S3). A total of 45 outlier regions distributed across 17 chromosomes are detected and the genes contained in these regions are defined as positively selected genes (PSGs) (Table S4).

We subsequently compare gene expression levels of testis between BY (n = 7) and LJX (n = 7) tree sparrows (Table S5) and find there are only 2 genes differentially expressed between the two populations (Figure S4; Table S6). Although so few differentially expressed genes (DEGs) are detected, one of the two DEGs, the gene encoding serine/threonine protein kinase PIM1, do overlap with the PSGs (Figure 1C and 1D). At the same time, we find 5 *PIM1* genes encoded PIM1 are centered in the outlier regions spread across 4 chromosomes (chromosome 25, 27, 30 and 35) (Table S4), 2 of which have complete kinase domain (Figure 1C). Signatures of selection in multiple *PIM1* genes coupled with upregulated *PIM1* gene expression in testis indicate that *PIM1* may play an important role in the sperm evolution of BY tree sparrows.

Expansion of PIM family by tandem and segmental duplication of PIM1

PIM1 , a member of the *PIM* gene family composed of three single-copy genes (*PIM1, PIM2, PIM3*) and highly evolutionarily conserved in multicellular organism, was originally identified as a common proviral

insertion site for the Moloney murine leukemia virus (Selten et al., 1985) and has been known as a proto-oncogene and implicated in the control of cancer cell proliferation, migration and apoptosis (Narlik-Grassow et al., 2014). In contrast to other amniotic genomes (including chicken) that harbor only single copy of *PIM1* gene, large expansion of *PIM1* which have been reported in zebra finch (*Taeniopygia guttata*) (Kong et al., 2010), another species of songbird, is also found in tree sparrow when *PIM3* keeps one single copy. Out of a total of 449 initially predicted *PIM1* genes, only 142 are found to be complete for the conserved protein kinase domain which related with the protein kinase catalytic activity. To deduce the evolutionary relationship of the *PIM* family, we construct a phylogenetic tree containing 154 PIM (2 of chicken, 3 of human, 6 of zebra finch and 143 of tree sparrow) proteins and divided the expanded *PIM1* genes of tree sparrow into 2 subgroups based on the constructed phylogenetic tree (Figure 2A). All subgroup I PIM1 proteins are found to have similar conserved motifs (motif 6-10) with the chicken and human PIM1 when subgroup II PIM1 have at most 5 excess motifs (motif 1-5) which also appear in the zebra finch (Figure 2A; Figure S5). The structural variation of subgroup II PIM1 proteins may be derived from an insertion event during duplication.

Notably, besides the structural difference, the two subgroups show different chromosomal distribution pattern and duplication mechanism. All subgroup I *PIM1* are centered in chromosome 7 and 24 when subgroup II have more decentralized distribution (Table S7). The centered distribution pattern of subgroup I is mainly owing to the tandem duplication events involved in the expansion of subgroup I *PIM1* genes and the more dispersed subgroup II *PIM1* genes are derived from segmental duplication (Figure S6). The dispersed duplicates of *PIM1* are noticed to be always adjacent to genes of several other families including C_2H_2 zinc finger (*C_2H_2ZNF*) protein, olfactory receptor (*OR*), p21-activated kinase (*PAK*), maestro heat-like repeat containing protein family member (*MROH*), hydrocephalus-inducing protein homolog (*HYDIN*) and inositol 1,4,5-trisphosphate receptor-interacting protein-like (*ITPRIPL*) which result in these 6 gene families expand simultaneously but at different degree in tree sparrow genome (Figure 2B; Figure S6). Furthermore, we find that the genomic regions contained these 7 families mainly distribute across chromosomes Z and 18-36, and these regions also contain high density of long terminal repeat (LTR) retrotransposons (Figure 2C; Figure S7), the same distribution pattern of these 7 families members and LTR retrotransposons may indicate that the expansion of these 7 families and LTR retrotransposons happened in tree sparrow genome simultaneously. Based on these results, we assume that large, interspersed segmental duplication of genomic regions contained random members of these 7 gene families occurred multiple times at different time points during evolution, and these replication events may be the strong driving force of the evolution of tree sparrow genome.

As so far, the expansion events of *PIM1* are only found in passerine (zebra finch and tree sparrow). To find out if the expansion of *PIM1* is songbird-specific, the *PIM* gene family members are identified and counted for all available avian genome in Ensembl, including 47 species covered 12 avian orders (Table S8). By comparing the number of *PIM1* in the 47 genomes, the expansion of *PIM1* is detected in 18 species of 3 orders (Figure 2C; Table S9). This result show that the expansion of *PIM* gene family caused by duplicated *PIM1* genes occurred not only in the order Passeriformes, but also in Psittaciformes and Strigiformes (Figure 2C) which may indicate the *PIM1* genes duplicate independently in several avian lineages.

Testis-specific expression patterns of the PIM1 genes and elevated expression level of tree sparrow from polluted site

Even though we have figured out the duplication events of *PIM1* in tree sparrow genome, there is still a gap between the *PIM1* , known as a proto-oncogene, and rapid evolution of sperm in BY. To further explore the physiological functions of *PIM1* and speculate their role in the rapid evolution of sperm under environmental pollution, we measured the expression level of *PIM* in different tissues (brain, heart, kidney, liver, lung, muscle, spleen, testis) and different sites (BY and LJX). Relative to the chicken *PIM1* widely expressed in the eight tissues, we surprisingly find that most of the *PIM1* genes show testis-specific expression patterns (Figure 3A). In addition to tissue expression profile, the average expression levels of *PIM1* between BY and LJX testis samples are compared. The results turn out that the average expression level of *PIM1*

is significantly higher in the testis of BY tree sparrow than of LJX while the expression of all transcripts and the copy number of *PIM1* show no difference (Figure 3B). Besides the average expression, the expression value of almost every single *PIM1* duplicate in testis is found to be higher in BY when *PIM3* is found to be unchanged (Figure S8).

Functional verification of PIM1

To validate the function of PIM1, 4 male tree sparrows from LJX were dosed daily by oral gavage at 100 mg/kg of SGI-1776, an experimental PIM1 kinase small molecule inhibitor, and the same numbers of LJX males were used as control group. After 1 week's treatment, we analyzed the sperm traits of two groups. However, we find there is no observed sperm traits variation between two groups. The sperm length and velocity of PIM1 inhibitor treatment group show no difference with those of the control group and the natural LJX population (Figure 3C). We then use the transcriptome data to quantify *PIM1* expression level, and we find that the *PIM1* expression level of treatment group is similar to control group.

Discussion

Based on the phenotypic variation of sperm in tree sparrow lived in BY which has been detected in our previous researches, we uncover a dramatically expanded *PIM* gene family may play an important role in shaping the phenotypic change of sperm under environmental pollution. As the only gene overlapped between the PSGs and DEGs, *PIM1* is identified as candidate gene for sperm evolution and targeted as the focus of our research. Although there is no documented example of *PIM1* enables rapid evolution of sperm, the genetic variation and differentially expression of cAMP-dependent protein kinase (*PKA*), another serine/threonine protein kinase which have high sequence homology with *PIM1* (Jacobs et al., 2005), is associated with rapid evolution of sperm in deer mouse (*Peromyscus maniculatus*) has been reported (Fisher et al., 2016).

As soon as our focus are narrowed down, the unusual and large-scale expansion of *PIM* gene family mainly caused by the duplication of *PIM1* are noticed in the genome of tree sparrow. 449 *PIM1* genes were predicted from the tree sparrow genome assembly and 142 of them are complete for the kinase domain. Gene duplications provide a source of genetic material for mutation, drift and selection to act upon, making new evolutionary opportunities for rapid adaptation (Crow & Wagner 2006; Hu et al., 2022; Magadum et al., 2013). Duplicate genes have been proven to be important targets of positive selection in Arabidopsis (Moore & Purugganan, 2003). As a result, the large number of duplicate *PIM1* gene copies may provide raw materials that facilitate rapid evolution of sperm when selective pressures shift in BY, which is demonstrated from the side by the fact that multiple *PIM1* duplicates appear repeatedly in outlier regions.

Our analysis also finds that both tandem duplication and segmental duplication events are involved in the expansion of *PIM* gene family, and the segmental duplications cause simultaneous expansion of 6 other gene families and LTR retrotransposons adjacent to *PIM* on chromosomes. The adjacent chromosomal distribution and simultaneously expansion suggest that the 7 families co-evolved in genome, and it is noteworthy that the expansion of these 7 gene family is so dramatical that make it may become a strong driving force of the evolution of tree sparrow genome.

Among the 7 expanded gene families in tree sparrow, the duplications of the most two prolifically expanded gene families C_2H_2ZNF and *OR* are widespread and prevalent in other species (Freitag et al., 1998; Seetharam & Stuart 2013) relative to the other 5 gene families. The duplication of *PIM1* has only been reported in zebra finch (Kong et al., 2010) whereas *PIM1* is believed to be single-copy and conserved in other amniotic genomes, even in chicken. Our analysis of *PIM1* duplication events in other avian genome indicate that *PIM1* expansion is not peculiar to passerine but arose in several lineages during the evolution of bird. Fixation of duplicate genes occurs when natural selection favors function or expression of duplicates (Cardoso-Moreira et al., 2016; Chain et al., 2011). Although the protein activity of duplicate PIM1 has not confirmed yet, compared with the species with single-copy *PIM1* , the retention of duplicates in genome and the increased gene dosage may suggest a more important and beneficial function of *PIM1* in the evolution and adaptation of the species with *PIM1* expansion. At the same time, there still much less *PIM1* duplicates are found in any other avian genome than in tree sparrow which may due to the further expansion of *PIM1* in tree

sparrow than other species or the incomplete genome assemblies.

The expression pattern of *PIM1* seems to be different in different species. *PIM1* has been reported to be expressed mainly in testis and hematopoietic tissues including thymus, spleen, bone marrow and fetal liver in both mouse and man (Bachmann & Möröy, 2005; Eichmann et al., 2000) when *PIM1* expressed in the brain and testis of zebra finch (Kong et al., 2010). Based on the current evidence, it seems that testis is only tissue in which *PIM1* of mammal and passerine both expressed, which indicates that *PIM1* may play an important role in male reproduction. But besides a few early researches imply *PIM1* is implicated in the normal development of male germ cells (Sorrentino et al., 1988; Wingett et al., 1992), the functions related to male reproduction of *PIM1* are fleetingly addressed in studies and people pay more attention to its role in cancer (Narlik-Grassow et al., 2014) and brain function of songbird (Kong et al., 2010). A large number of duplicates and the testis-specific expression pattern, as revealed by our analysis, as well as the involvement of mammal *PIM1* in normal germ cell maturation published before (Sorrentino et al., 1988; Wingett et al., 1992) may indicate *PIM1* play a strong and essential role in spermatogenesis of tree sparrow. We thus consider the role of *PIM1* in male reproduction, an indispensable and inseparable part of *PIM1* function, should never be neglected not only in the species with multiple *PIM1* but in species with conserved single-copy.

Higher expression levels of *PIM1* , wholly or individually, are detected in testis sampled from BY tree sparrow. Interestingly, the expression level of *PIM3* , a member of *PIM* family as well as *PIM1* but without duplication and testis-specific expression pattern, shows no difference between BY and LJX. The elevated expression level may be the consequence of shifts in selective pressures on duplicate *PIM1* . Furthermore, we test if the elevated expression level of *PIM1* in polluted site is due to larger scale of duplication and higher copy numbers of *PIM1* . However, the results show that there is no increase in copy number of *PIM1* in BY population. Therefore, the molecular mechanism of change in expression level of *PIM1* still need to be further researched in the future.

When we take into account the link between transcriptional changes to sperm phenotypic effects, the higher expression levels of *PIM1* seem to be consistent with our earlier findings of longer and faster swimming sperm in BY (Yang et al., 2020a, 2020b). No research is found to explore the relationship between expression level of *PIM1* and sperm traits variation, nonetheless, several studies have proven that *PIM1* can accelerate cancer cell motility and overexpression of *PIM1* result in increased cancer cell motility (Mou et al., 2016; Santio et al., 2020; Tanaka et al., 2009). PIM1 kinase normally accelerates cell motility through phosphorylating and activating target proteins. We thus speculate that PIM1 also accelerate sperm motility through phosphorylating downstream proteins in tree sparrow. However, there are so many proteins that can be phosphorylated by PIM1. In prostate cancer cells, PIM1 can accelerate cell motility through phosphorylating actin capping proteins (Sanito et al., 2020). PIM1 accelerate human oral squamous cell carcinoma cell motility through activating Ras-related C3 botulinum toxin substrate 1 (RAC1) instead (Tanaka et al., 2009). Therefore, it is hard to confirm the target protein and signaling pathway of PIM1 in tree sparrow. We attempt to use PIM1 inhibitor to verify the molecular function and preliminary explore the downstream protein of PIM1 kinase. However, there is no observed sperm trait change are found in the inhibitor treatment group. The unchanged sperm trait may due to the limitations of our inhibition experiment. We cannot ensure whether the inhibitor can pass through the blood-testis barrier and reach the target tissue, or whether the mammal inhibitor can work equally on tree sparrow.

In conclusion, our work utilizes population genomic and transcriptomic studies, exploring the genetic underpinning of phenotypic variation of sperm in tree sparrow lived in BY which is polluted by heavy metal. On the one hand, our results find *PIM1* , a single-copy gene in most of amniotic genomes, remarkably expanded in tree sparrow and other avian species of three lineages including Passeriformes, Psittaciformes and Strigiformes. Duplication of *PIM1* may be an important force driving speciation and phenotypic diversity in avian evolution. On the other hand, while the molecular functions and pathways of the expanded *PIM1* remain to be defined by experimental verification, the signatures of selection and upregulated gene expression in polluted sites, indicating the expanded *PIM1* may be important sources for rapid evolution of sperm in tree sparrow following the environmental change.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 31572216) and the Science and Technology Program of Gansu Province (No. 21JR7RA479). We received support for computational work from Supercomputing Center of the Lanzhou University.

References

Akinloye, O., Arowojolu, A. O., Shittu, O. B., & Anetor, J. I. (2006). Cadmium toxicity: a possible cause of male infertility in Nigeria.*Reproductive Biology,* 6(1), 17-30.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research,*19(9), 1655-1664.

Ames, R. M., Rash, B. M., Hentges, K. E., Robertson, D. L., Delneri, D., & Lovell, S. C. (2010). Gene duplication and environmental adaptation within yeast populations. *Genome Biology and Evolution,* 2(1), 591-601.

Aulsebrook, L. C., Bertram, M. G., Martin, J. M., Aulsebrook, A. E., Brodin, T., Evans, J. P., Hall, M. D., O'Bryan, M. K., Pask, A. J., Tyler, C. R., & Wong, B. B. M. (2020). Reproduction in a polluted world: implications for wildlife. *Reproduction,* 160(2), R13-R23.

Avidor-Reiss, T. (2018). Rapid evolution of sperm produces diverse centriole structures that reveal the most rudimentary structure needed for function. *Cells,* 7(7), 67.

Bachmann, M., & Möröy, T. (2005). The serine/threonine kinase Pim-1.*The International Journal of Biochemistry & Cell Biology,* 37(4), 726-730.

Bailey, R. E. (1952). The incubation patch of passerine birds. The Condor, 54(3), 121-136.

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME suite. *Nucleic Acids Research,* 43(W1), W39-W49.

Birkhead, T. R., & Pizzari, T. (2002). Postcopulatory sexual selection.*Nature Reviews Genetics,* 3(4), 262-273.

Breckels, R. D., & Neff, B. D. (2014). Rapid evolution of sperm length in response to increased temperature in an ectothermic fish.*Evolutionary Ecology,* 28(3), 521-533.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics,* 10(1), 421.

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N.D., & May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana .BMC Plant Biology,* 4(1), 10.

Cardoso-Moreira, M., Arguello, J. R., Gottipati, S., Harshman, L. G., Grenier, J. K., & Clark, A. G. (2016). Evidence for the fixation of gene duplications by positive selection in Drosophila. *Genome Research,* 26(6), 787-798.

Chain, F. J. J., Dushoff, J., & Evans, B. J. (2011). The odds of duplicate gene persistence after polyploidization. *BMC Genomics,*12(1), 599.

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant,*13(8), 1194-1202.

Chen, L. S., Redkar, S., Taverna, P., Cortes, J. E., & Gandhi, V. (2011). Mechanisms of cytotoxicity to Pim kinase inhibitor, SGI-1776, in acute myeloid leukemia. *Blood,* 118(3), 693-702.

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., Zhang, X., Wang, J., Yang, H., Fang, L., & Chen, Q. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience,* 7(1), 1-6.

Chiou, T. J., Chu, S. T., Tzeng, W. F., Huang, Y. C., & Liao, C. J. (2008). Arsenic trioxide impairs spermatogenesis via reducing gene expression levels in testosterone synthesis pathway. *Chemical Research in Toxicology,*21(8), 1562-1569.

Conant, G. C., & Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics,*9(12), 938-950.

Crow, K. D., & Wagner, G. P. (2006). What is the role of genome duplication in the evolution of complexity and diversity.*Molecular Biology and Evolution,* 23(5), 887-892.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics,*27(15), 2156-2158.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29(1), 15-21.

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology,* 7(10), e1002195.

Eichmann, A., Yuan, L., Bréant, C., Alitalo, K., & Koskinen, P. J. (2000). Developmental expression of Pim kinases suggests functions also outside of the hematopoietic system. *Oncogene,* 19(9), 1215-1224.

Fisher, H. S., Jacobs-Palmer, E., Lassance, J. M., & Hoekstra, H. E. (2016). The genetic basis and fitness consequences of sperm midpiece size in deer mice. *Nature Communications,* 7(1), 13652.

Freitag, J., Ludwig, G., Andreini, I., Rössler, P., & Breer, H. (1998). Olfactory receptors in aquatic and terrestrial vertebrates.*Journal of Comparative Physiology A,* 183(5), 635-650.

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data.*Bioinformatics,* 32(18), 2847-2849.

Hardneck, F., Israel, G., Pool, E., & Maree, L. (2018). Quantitative assessment of heavy metal effects on sperm function using computer-aided sperm analysis and cytotoxicity assays. *Andrologia,* 50(10), e13141.

Hu, Y., Wu, X., Jin, G., Peng, J., Leng, R., Li, L., Gui, D., Fan, C., & Zhang, C. (2022). Rapid genome evolution and adaptation of*Thlaspi arvense* mediated by recurrent RNA-based and tandem gene duplications. *Frontiers in Plant Science,* 12(1), 772655.

Jacobs, M. D., Black, J., Futer, O., Swenson, L., Hare, B., Fleming, M., & Saxena, K. (2005). Pim-1 ligand-bound structures reveal the mechanism of serine/threonine kinase inhibition by LY294002. *Journal of Biological Chemistry,* 280(14), 13728-13734.

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B,* 279(1749), 5048-5057.

Kong, L., Lovell, P. V., Heger, A., Mello, C. V., & Ponting, C. P. (2010). Accelerated evolution of *PAK3*- and *PIM1*- like kinase gene families in the zebra finch, *Taeniopygia guttata* .*Molecular Biology and Evolution,* 27(8), 1923-1934.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics,* 35(21), 4453-4455.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research,*19(9), 1639-1645.

Kumar, N., & Singh, A. K. (2022). Impact of environmental factors on human semen quality and male fertility: a narrative review.*Environmental Science Europe,* 34(1), 6.

Kushawaha, B., Yadav, R. S., Swain, D. K., Kumari, P., Kumar, A., Yadav, B., Anand, M., Yadav, S., Singh, D., & Garg, S. K. (2021). Collapsed mitochondrial cristae in goat spermatozoa due to mercury result in lethality and compromised motility along with altered kinematic patterns. *Scientific Reports,* 11(1), 646.

Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research,* 49(W1), W293-W296.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25(14), 1754-1760.

Li, Y., Wang, Y., Gou, X., Su, Y., & Wang, G. (2006). Risk assessment of heavy metals in soils and vegetables around non-ferrous metals mining and smelting sites, Baiyin, China. *Journal of Environmental Sciences,* 18(6), 1124-1134.

Liao, Y., Smyth, G. K., & Shi. W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30(7), 923-930.

Liu, B., Ai, S., Guo, R., Ren, L., Zhang, W., & Zhang, Y. (2016). Chemical fractions and mobility of Cd, Cu, Pb and Zn in soil profile of a sewage irrigation area, northwest China. *Journal of Residuals Science & Technology,* 13(S1), S13-S20.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15(12), 550.

Lynch, M., & Katju, V. (2004). The altered evolutionary trajectories of gene duplicates. *Trends in Genetics,* 20(11), 544-549.

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of Genetics,* 92(1), 155-161.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.*Genome Research,* 20(9), 1297-1303.

Minnameyer, A., Strobl, V., Bruckner, S., Camenzind, D. W., Oystaeyen, A. V., Wäckers, F., Williams, G. R., Yañez, O., Neumann, P., & Straub, L. (2021). Eusocial insect declines: insecticide impairs sperm and feeding glands in bumblebees. *Science of the Total Environment,*785(1), 146955.

Moore, R. C., & Purugganan, M. D. (2003). The early stages of duplicate gene evolution. *Proceedings of the National Academy of Sciences of the United States of America,* 100(26), 15682-15687.

Mou, S., Wang, G., Ding, D., Yu, D., Pei, Y., Teng, S., & Fu, Q. (2016). Expression and function of PIM kinase in osteosarcoma.*International Journal of Oncology,* 49(5), 2116-2126.

Naha, N., & Chowdhury, A. R. (2006). Inorganic lead exposure in battery and paint factory: effect on human sperm structure and functional activity. *Journal of UOEH,* 28(2), 157-171.

Nan, Z., & Zhao, C. (2000). Heavy metal concentration in gray calcareous soils of Baiyin region, Gansu Province, P.R. China. *Water, Air, and Soil Pollution,* 118(1), 131-142.

Narlik-Grassow, M., Blanco-Aparicio, C., & Carnero, A. (2014). The PIM family of serine/threonine kinase in cancer. *Medicinal Research Reviews,* 34(1), 136-159.

Osorio, D., Rondón-Villarreal, P., & Torres, R. (2015). Peptides: a package for data mining of antimicrobial peptides. *The R Journal,*7(1), 4-14.

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research,* 46(W1), W200-W204.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE,* 5(3), e9490.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics,* 81(3), 559-575.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26(6), 841-842.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26(1), 139-140.

Santio, N. M., Vainio, V., Hoikkala, T., Mung, K. L., Lång, M., Vahakoski, R., Zdrojewska, J., Coffey, E. T., Kremneva, E., Rainio, E. M., & Koskinen, P. J. (2020). PIM1 accelerates prostate cancer cell motility by phosphorylating actin capping proteins. *Cell Communication and Signaling,* 18(1), 121.

Seetharam, A., & Stuart, G. W. (2013). A study on the distribution of 37 well conserved families of C2H2 zinc finger genes in eukaryotes.*BMC Genomics,* 14(1), 420.

Selten, G., Cuypers, H. T., & Berns, A. (1985). Proviral activation of the putative oncogene *Pim-1* in MuLV induced T-cell lymphomas.*The EMBO Journal,* 4(7), 1793-1798.

Sorrentino, V., McKinney, M. D., Giorgi, M., Geremia, R., & Fleissner, E. (1988). Expression of cellular protooncogenes in the mouse male germ line: a distinctive 2.4-kilobase pim-1 transcript is expressed in haploid postmeiotic cells. *Proceedings of the National Academy of Sciences of the United States of America,* 85(7), 2191-2195.

Tabassomi, M., & Alavi-Shoushtari, S. M. (2013). Effects of *in vitro* copper sulphate supplementation on the ejaculated sperm characteristics in water buffaloes (*Bubalus bubalis*).*Veterinary Research Forum,* 4(1), 31-36.

Tanaka, S., Kitamura, T., Higashino, F., Hida, K., Ohiro, Y., Ono, M., Kobayashi, M., Totsuka, Y., & Shindoh, M. (2009). Pim-1 activation of cell motility induces the malignant phenotype of tongue carcinoma.*Molecular Medicine Reports,* 2(2), 313-318.

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., & Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science,*320(5875), 486-488.

Vasudeva, R., Sutter, A., Sales, K., Dickinson, M. E., Lumley, A. J., & Gage, M. J. G. (2019). Adaptive thermal plasticity enhances sperm and egg performance in a model insect. *eLife,* 8(1), e49452.

Walsh, B. S., Parratt, S. R., Hoffmann, A. A., Atkinson, D., Snook, R. R., Bretman, A., & Price, T. A. R. (2019). The impact of climate change on fertility. *Trends in Ecology & Evolution,* 34(3), 249-259.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.*Nucleic Acids Research,* 38(16), e164.

Wang, W. W. Y, & Gunderson, A. R. (2022). The physiological and evolutionary ecology of sperm thermal performance. *Frontiers in Physiology,* 13(1), 754830.

Wingett, D., Reeves, R., & Magnuson, N. S. (1992). Characterization of the testes-specific pim-1 transcript in rat. *Nucleic Acids Research,* 20(12), 3183-3189.

Yang, W., Wang, S., Yang, Y., Shen, Y., & Zhang, Y. (2021). Improvement of sperm traits related to the high level of extra-pair fertilization in tree sparrow population under long-term environmental heavy metal pollution. *Science of the Total Environment,* 790(1), 148109.

Yang, Y., Zhang, Y., Ding, J., Ai, S., Guo, R., Bai, X., & Yang, W. (2019). Optimal analysis conditions for sperm motility parameters with a CASA system in a passerine bird, *Passer montanus* . *Avian Research,* 10(1), 35.

Yang, Y., Zhang, H., Wang, S., Yang, W., Ding, J., & Zhang, Y. (2020a). Variation in sperm morphology and performance in tree sparrow (*Passer montanus*) under long-term environmental heavy metal pollution. *Ecotoxicology and Environmental Safety,* 197(1), 110622.

Yang, Y., Zhang, W., Wang, S., Zhang, H., & Zhang, Y. (2020b). Response of male reproductive function to environmental heavy metal pollution in a free-living passerine bird, *Passer montanus* . *Science of the Total Environment,* 747(1), 141402.

Data Accessibility Statement

All raw sequence data have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (BioProject: PRJNA867105).

Author contributions

Y. Z. and S.W. conceived the project and designed the research. W.Y., S.W., Y.S., Z.L. and S.Z. collected samples in the field. S.W. performed the bioinformatic analysis and drafted the original manuscript. Y.Z. and G.S. revised and edited the manuscript.

Figures

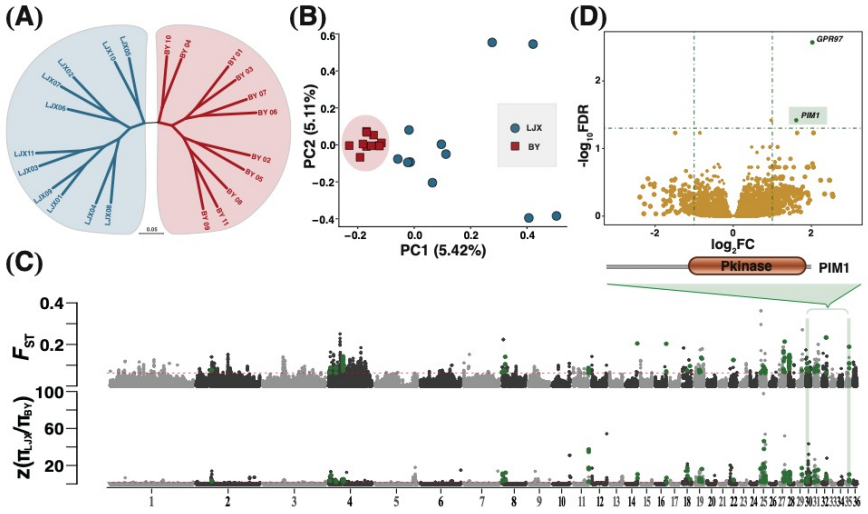

Figure 1. Genetic variation between the two populations. (A) Maximum-likelihood tree from genome-wide single nucleotide polymorphisms (SNPs). The samples from BY and LJX are shadowed in red and blue respectively. (B) Principal component analysis plot of the first two components. The fraction of the variance explained is 5.42% for PC1 and 5.11% for PC2. The samples from BY are clustered more compactly than those from LJX. (C) Manhattan plot of genetic differentiation between the BY and LJX population. Green dots indicate outlier regions (1% most divergent) with $F_{ST} > 0.07$ (top panel) and $\pi_{LJX}/\pi_{BY} > 1.64$ (bottom panel). The genes encoding serine/threonine protein kinase PIM1 with complete kinase domain are highlighted by a shaded green column and the schematic representation of PIM1 domain structure is shown. (D) Volcano plot indicating the results of differential expression analysis for testis. The green dots represent the significantly upregulated genes in BY population and the overlapped gene (*PIM1*) between PSGs and DEGs is shadowed in green.

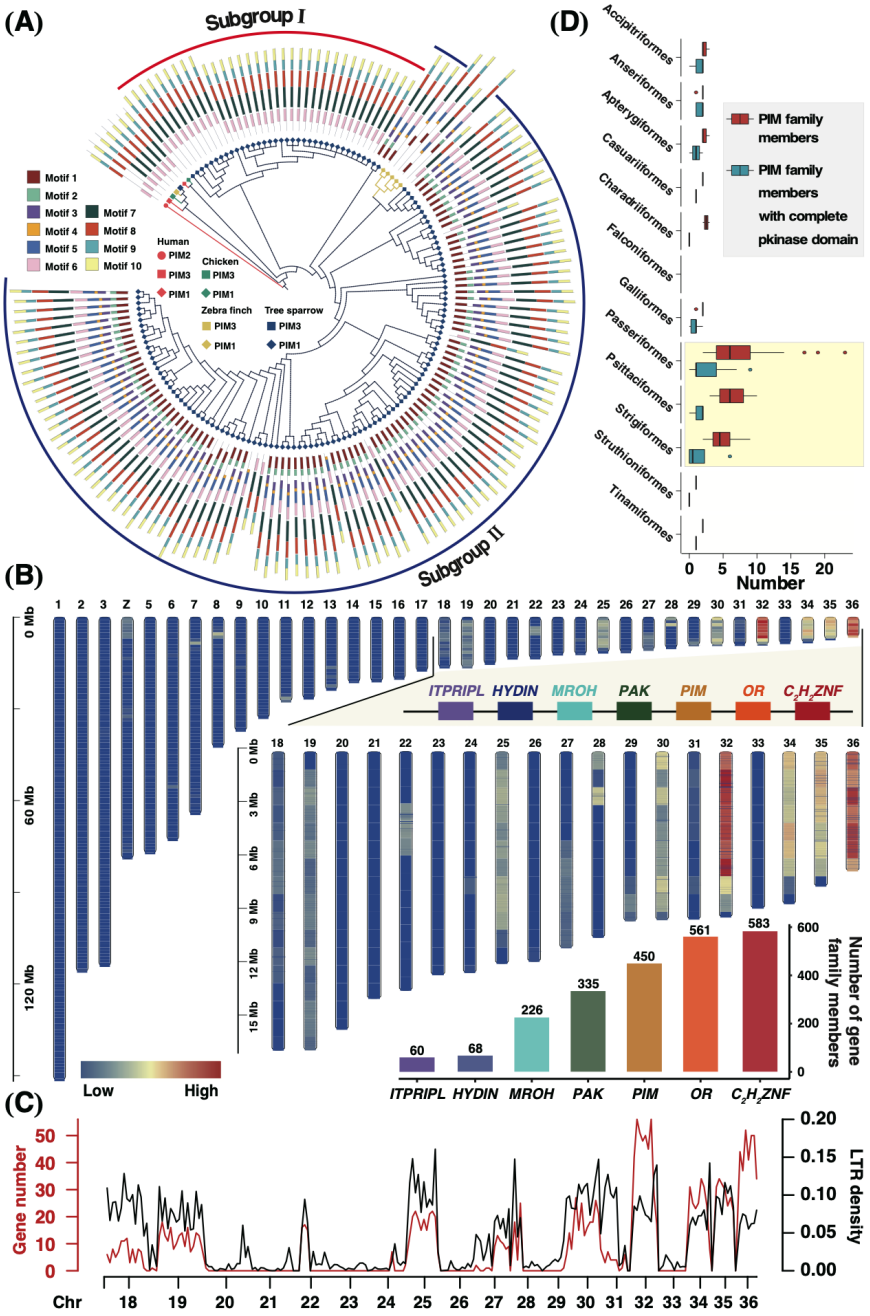

Figure 2. Expansion of *PIM1* through segmental duplication. (A) Phylogenetic relationship and motif composition of 143 duplicated PIM proteins identified with complete protein kinase domain in tree sparrow genome and PIM family members from other species (2 of chicken, 3 of human and 6 of zebra finch). The inner layer is a maximum-likelihood tree of the total 154 PIM proteins. All PIM1 of tree sparrow are divided into 2 subgroups based on the phylogenetic tree. The outer layer is the distribution of the conserved motifs in PIM proteins and the 10 motifs are represented by boxes with different colors. The PIM1 from same subgroup have more similar motifs. (B) Chromosomal distribution and segmental duplication events of 7 expanded gene families in tree sparrow genome. The expansion of *PIM* is derived from segmental duplication of genomic regions contained *PIM* and other 6 families includingC_2H_2ZNF , *OR* , *PAK* ,*MROH* , *HYDIN*

and *ITPRIPL* . The distribution density of these 7 families on each chromosome is shown in the form of heatmap and the scale on the left represented the length of chromosomes. The smaller chromosomes 18-36 contained most of the duplicated regions are zoomed in and the different degree of expansion of 7 families are shown in the bottom bar chart. (C) The distribution of 7 expanded gene families and LTR retrotransposons in chromosomes 18-36. The duplicate gene number is represented by the red line when LTR retrotransposons density is represented by the black line. The distribution of duplicate gene is consistent with the density of LTR retrotransposons. (D) Expansion of *PIM* family in other avian lineages. Box plot for the numbers of *PIM* family members in 12 avian orders. Expansion of*PIM* occurred in three orders which are shadowed in yellow.

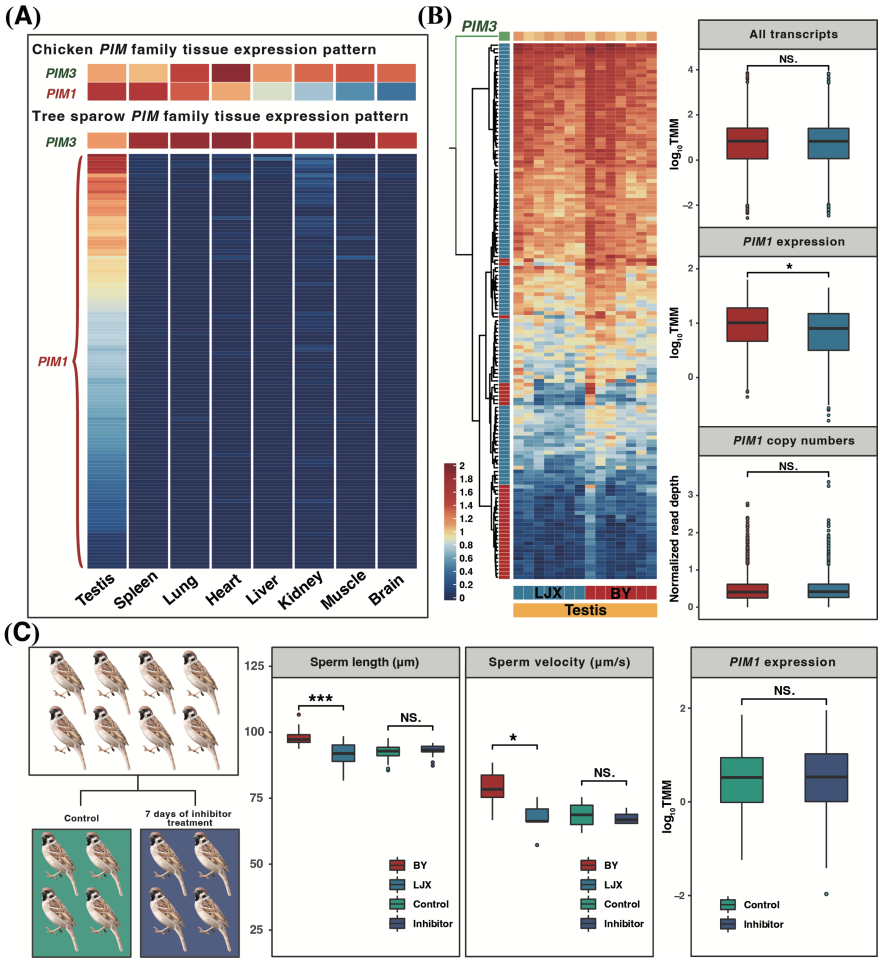

Figure 3. Expression patterns of *PIM1* and other 6 gene families. (A) Heatmap of expression profiles of chicken and tree sparrow *PIM* genes in different tissues. The scale bar on the right represents the $\log_{10}$-transformed TMM values. (B) Heatmap of expression profiles of BY and LJX tree sparrow *PIM*genes in testis (left panel). The colored squares on the left represent the *PIM3* (green) and two subgroups of *PIM1* (red for subgroup I and blue for subgroup II). All transcripts (right top panel) in testis and *PIM1* copy number (right bottom panel) show no difference between BY and LJX population, when *PIM1* expression level is significantly higher in BY tree sparrow testis (right middle panel). (C) Inhibition experiment protocol for the PIM1 inhibitor treatment group and control group (left panel). Both sperm traits (middle panel) and *PIM1* expression level (right panel) show no difference between control and inhibitor group. *P* values are calculated with Student's t-test. "NS." denotes $P > 0.05$, "*" denotes $0.01 < P < 0.05$, "**" denotes

$0.001 < P < 0.01$ and "***" denotes $P < 0.001$.