

Deep Sequencing of Fosmid Clones Indicates Gene Conversion in the Male-Specific Region of the Giant Panda Y Chromosome

Fei-Ran Han, Xuan-Min Guang, Qiu-Hong Wan, and Sheng-Guo Fang*

The Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, State Conservation Center for Gene Resources of Endangered Wildlife, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, China

*Corresponding author: E-mail: sgfanglab@zju.edu.cn.

Accepted: August 10, 2018

Data deposition: This project has been deposited at NCBI BioProject (<http://www.ncbi.nlm.nih.gov/bioproject>; last accessed August 9, 2018) under the accession PRJNA483520.

Abstract

The giant panda (*Ailuropoda melanoleuca*) is popular around the world and is widely recognized as a symbol of nature conservation. A draft genome of the giant panda is now available, but its Y chromosome has not been sequenced. Y chromosome data are necessary for study of sex chromosome evolution, male development, and spermatogenesis. Thus, in the present study, we sequenced two parts of the giant panda Y chromosome utilizing a male giant panda fosmid library. The sequencing data were assembled into two contigs, each ~100 kb in length with no gaps, providing high-quality resources for studying the giant panda Y chromosome. Annotation and transposable element comparison indicates varied evolutionary pressure in different regions of the Y chromosome. Two genes, zinc finger protein, Y-linked (*ZFY*) and lysine demethylase 5D (*KDM5D*), were annotated and gene conversion was observed for *ZFY* exon 7. Phylogenetic analysis also revealed that this gene conversion event happened independently in multiple mammalian lineages, indicating a putative mechanism to maintain the function of this particular gene on the Y chromosome. Furthermore, a transposition event, discovered through comparative alignment with the giant panda X chromosome sequence, may be involved in the process of gaining new genes on the Y chromosome. Thus, these newly obtained Y chromosome sequences provide valuable insights into the genomic patterns of the giant panda.

Key words: fosmid library, sex chromosome, transposable elements, gene conversion, transposition.

Introduction

The giant panda (*Ailuropoda melanoleuca*) has attracted significant attention from researchers and the public over the past decades, not only because of its unique appearance, but also for its special evolutionary history and current “vulnerable” status. The ancient species once lived throughout southern and eastern China, but now its habitat is extremely limited and fragmented (Fang et al. 2002). Considerable efforts have been made to prevent extinction, yet the giant panda is still in danger, mainly due to continued habitat loss (Xu et al. 2017) and low birthrate, both in the wild and in captivity (Zhang and Wei 2006).

Although numerous studies have been carried out to provide insights into this threatened and enigmatic species, our understanding of the giant panda is still insufficient. Research concerning the phylogeny and unique biological traits of the

giant panda has relied heavily on genetic data, obtained primarily by genome and transcriptome sequencing. Indeed, the genome of a female giant panda was previously sequenced and assembled de novo (Li et al. 2010). However, the current literature lacks an analysis of the Y chromosome sequence. Thus, our understanding of the giant panda genome is still incomplete.

The sex chromosomes of therian (marsupial and eutherian) species, including the X and Y chromosomes, originated from a homologous pair of autosomes (Graves 2006). The Y chromosome is essential for the process of male development and spermatogenesis, and genes on the Y chromosome usually bear important functions related to sex (Bachtrog 2013). However, because of its complex and repetitive nature, the Y chromosome is difficult to sequence. This has led to a preference for female sequencing targets in mammalian whole genome sequencing projects (Hughes and Rozen 2012).

In research focused on the Y chromosome, sequencing has been largely aimed at the X-degenerate region. The X-degenerate region is one of the major sequence classes in the male specific region of the Y chromosome (MSY) and carries most of the single-copy X ancestral genes on the Y chromosome (Li et al. 2013). Although this region is easier to assemble than the ampliconic region or the heterochromatic region, few studies have analyzed its sequence, and our understanding of the X-degenerate region is still limited.

In this study, we deeply sequenced two contigs (~100 kb each) on the giant panda MSY utilizing a male giant panda fosmid library. Annotation of the two contigs provided detailed information of protein coding sequences and transposable elements (TEs). Using these sequences, we analyzed the TEs to identify those that are evolutionarily conserved and revealed the bias of evolution pressure between regions of the Y chromosome. This analysis also includes comparative alignment of the sequences to evaluate transposition events as well as phylogenetic analysis of the zinc finger protein, Y-linked (*ZFY*) and lysine demethylase 5D (*KDM5D*, also known as *SMCY*) genes, both of which play a role in Y chromosome-related functions. The phylogenetic analysis revealed gene conversion in the *ZFY* gene in giant panda lineage as well as other mammalian lineages. As the analysis utilized full length exon sequences of giant panda generated by the sequencing of the fosmid clones, the gene conversion event was located in exon 7 of the *ZFY* gene. To our knowledge, this is the first time these analyses have been conducted at such a depth on the Y chromosome of the giant panda.

Materials and Methods

Construction and Arrangement of the Fosmid Library

We used a skeletal muscle-derived cell line derived from a male giant panda to yield enough genomic DNA material to construct the fosmid library. Genomic DNA from the cell line was isolated using a Genomic DNA Miniprep Kit (Generay Biotech, Shanghai, China) following the manufacturer's instructions.

The genomic DNA was sheared randomly by pipetting up and down and then separated using low melting point agarose gel electrophoresis. The gel containing 40-kb DNA fragments was excised, melted, and digested using GELase enzyme (Epicentre), and the DNA fragments were restored by ethanol precipitation. The restored DNA fragments were end-repaired to blunt, 5'-phosphorylated ends, and ligated into the CopyControl pCC1FOS vector. The ligation reaction was packaged into phage particles and transformed into the EPI300-T1^R phage T1-resistant strain of *Escherichia coli*. The library was then plated on LB plates with chloramphenicol and incubated overnight. To reduce the storage space needed,

every 100–200 clones were mixed and stored in a well of a 96-well plate.

We established a PCR-screening system based on the 4D-PCR system (Asakawa et al. 1997) with modification. The 96-well plates were numbered, and for each plate, all clones in 96 wells were mixed as a superpool. Then, the clones in the same columns (designated column-subpools 1–12) and rows (designated row-subpools 1–8) were mixed. For each superpool, there were a corresponding set of column- and row-subpools. Fosmid DNA of all superpools and subpools were isolated for PCR-screening. We traced the target clone to its specific plate by PCR-screening the superpools, and then traced back to the specific well by PCR-screening the corresponding subpools. Then the clone mixture was streaked out to isolate the target clone by colony-PCR.

Primer Designing and Library Screening

To obtain Y chromosome clones, we designed two primer pairs based on the published sequence of the final intron of the giant panda *ZFY* gene, and fourth intron of the giant panda *KDM5D* (*SMCY*) gene. These two primer pairs were named ZFYi and KDM5Di, respectively. We also used a primer pair (designated 322) for the polar bear Y chromosome scaffold (Kutschera et al. 2014). These three primer pairs were validated by PCR amplifying male giant panda DNA. Female giant panda DNA was used as a control to ensure male-specificity. After acquiring the positive clones using these three Y-specific primer pairs, end-sequencing of the clones was performed using the pCC1FOS forward and reverse vector primers, and end-primers were designed based upon the end sequences. Then, the end sequences of the PCR-positive clones were aligned to the female giant panda genome using the Basic Local Alignment Search Tool (BLAST), to eliminate autosomal or X-linked clones. Using these new end-primers, male-specificity validation and library screening was repeated to get tilling path clones. The end-primers were also used to validate overlapping clones. Detailed information concerning the primers used in this study is shown in [supplementary table S1, Supplementary Material](#) online.

Clone Sequencing and Assembly

Fosmid DNA from eight clones in two minimal tiling paths was sequenced separately using an Illumina 2000 system (Illumina, San Diego, CA, USA) at the Beijing Genomics Institute (Shenzhen, China). Insert libraries with lengths of 450 bp were constructed, and the pair-end read length was set to 250 bp. Paired-end reads were assembled using SPAdes 3.11.1 (Nurk et al. 2013), with two k-mer sets (set 1: 31, 41, 51, 61, 71, and 81; set 2: 91, 111, and 127). These two assemblies were aligned with each other to fill the gaps. The clone sequences were then vector-removed and

Table 1

Giant Panda Y-Linked Fosmid Clones Used for Sequencing

Contig Name	Clone Number	Insert Size (bp)	Screening Primer	Average Sequencing Depth
Contig 1	18-7-3-1	42408	<i>KDM5Di</i>	20,462.13
	8-3-5-1	40136	<i>KDM5Di</i> , <i>ZFYi</i>	21,325.73
	11-9-6-1	31922	<i>ZFYi</i>	25,549.47
Contig 2	13-3-4-1	36448	11961R3	23,164.59
	26-6-5-2	35440	18941F	24,420.92
	18-9-4-1	33823	322	25,498.14
	25-2-4-1	36987	18941R	24,356.44
	6-6-3-1	33094	25241R4	25,313.50

assembled into two complete contigs in accordance with the minimal tiling path.

Gene Annotation, Sequence Content, and Comparative Alignment

Homology-based methods were used to predict genes in both contigs. We downloaded the exon sequences of known Y-linked genes in the giant panda and other mammals from Genbank. These exon sequences were then mapped to the contigs using TblastN (E -value $<1E-5$). The gene models were defined by aligning the contig sequences showing homology with the exon sequences against the corresponding protein utilizing Genewise. Additional similarity searches against the nucleotide collection of NCBI was carried out using BLAST. TEs were identified using RepeatMasker 4.0.7 with the Ailuropoda library (version 2017-01-27). Sequence divergence rates of the TEs were calculated as the percentage of different bases relative to the TE library. Guanine-cytosine content of the sequences was estimated using Emboss Isochore (Rice et al. 2000) with 200-bp window size and 5-bp shift. Comparative alignments between the giant panda Y chromosome contigs and the polar bear Y-linked genomic scaffolds as well as between Contig 2 and giant panda genomic scaffold 1583 were performed using LASTZ (Harris 2007).

Sequencing Depth

Using the assembled clone insert sequences as a reference, the pair-end reads were mapped using BWA 0.7.17 (Li and Durbin 2009) with the BWA-MEM algorithm. The sequencing depth of each site was exported using SAMtools 0.1.19 (Li et al. 2009). Average sequencing depth was determined with 200-bp window size.

Phylogenetic Analysis

The coding sequences of the Y-linked genes *ZFY* and *KDM5D* (*SMCY*) of the giant panda were obtained from the assembly. Furthermore, the coding sequences of their X-linked paralogous genes *ZFX* and *KDM5C* (*SMCX*) were downloaded from Genbank. The coding sequences of the orthologous genes

from other mammalian species and chicken were also downloaded from NCBI and Vega Genome Browser. The *ZFY/ZFX* exons 1–6, *ZFY/ZFX* exon 7 (last exon), and *KDM5D/KDM5C* coding sequences were, respectively, aligned using the ClustalW program. After model testing using jModelTest 2.1.7 according to AICc, Bayesian trees were built using MrBayes 3.2.6 (Ronquist et al. 2012) with the following parameters: Six different substitution rates, over 10,000,000 generations, a sampling frequency of 1000 generations, and a gamma-distributed rate with a proportion of invariable sites. These settings correspond to the GTR model. Neighbor-Joining (NJ) trees were then built using MEGA 7.0 (Kumar et al. 2016). Tree reliability was calculated using the bootstrapping method with 1000 replications. The chicken sequences were used as outgroups.

Results

Fosmid Library Construction and Screening

The whole fosmid library of the male giant panda contained about 370,000 clones in total. The library was stored in 28 96-well plates, and the average number of clones in each well was 138.2. The average insert size of the library was about 40 kb. Library screening using Y-specific primers (*ZFYi*, *KDM5Di*, and 322) was performed, and eight Y-linked clones were identified, four of which were derived from the *ZFYi* and *KDM5Di* primer pairs (designated Contig 1), whereas the other four were derived from the 322 primer pair (designated Contig 2) (table 1).

Fosmid Clone Sequencing

The eight fosmid clones, which had the minimal tiling path, were then sequenced individually, and each sequencing reaction generated ~ 1.1 Gb of data, with an average sequencing depth over 20,000 \times . The sequencing data was assembled clone by clone, and then combined into the two complete contigs, each of which contains four clones (fig. 1). The average length of the clone insert sequences is 36,282.25 bp, and the lengths of Contigs 1 and 2 are 106 952 and 103 648 bp



Fig. 1.—Sequence content and annotation of Contig 1 and Contig 2 of the giant panda Y chromosome. (A, D) GC content plot of the two contigs. (B, E) Gene organization and repeat distribution of the contigs. Genes are shown in pink and exon sequences are highlighted in red. Pseudogenes are shown in sage green, and exon sequences of the pseudogenes are shown in gray. Blue bars and green bars indicate the distribution of long and short interspersed nuclear elements (LINEs and SINEs), respectively. (C, F) The overlapping clones are shown as yellow bars with plots of the sequencing depth. Dotted lines indicate the overlapping regions of the clones.

respectively. Furthermore, the two contigs had similar guanine-cytosine ratios (36.63% and 38.09%, respectively).

Comparison of TEs

Overall, TEs cover 39.2% of Contig 1 and 63.8% of Contig 2 (table 2). As a comparison, TEs cover 34.7% of the female giant panda genome (Li et al. 2010). The majority of the TEs are long interspersed nuclear elements (LINEs, mostly

LINE-L1), covering 21.5% of Contig 1 and 49.1% of Contig 2. LINEs also comprise 18.2% of the female giant panda genome. Moreover, short interspersed nuclear elements (SINEs) cover 12% and 6.3% of Contigs 1 and 2, respectively. The percentage of long terminal repeats (LTRs)/DNA transposons in Contig 1 are slightly lower than that in Contig 2. The distribution of LINEs and SINEs are shown in figure 1. We also analyzed the divergence rate of the TEs (supplementary fig. S1, Supplementary Material online). Notably, most of the

sequence divergence in the two contigs occurs in the LINES, with those of Contig 1 having a higher divergence rate than those of Contig 2.

Gene Annotation

Two Y-linked genes, *ZFY* and *KDM5D* (*SMCY*), were found on Contig 1 (fig. 1). No genes were found on Contig 2 during annotation, but there is a high level of similarity between this contig and the giant panda neuronal membrane glycoprotein M6-B (*GPM6B*) gene. However, alignment of Contig 2 and the giant panda *GPM6B* gene sequence revealed that two small exons are missing in Contig 2, indicating that the “*GPM6B*” on Contig 2 is likely a pseudogene present on the giant panda Y chromosome (fig. 1).

Table 2

Statistics of the Transposable Elements (TEs) in Contig 1, Contig 2, and Giant Panda Genome

Type	% Contig 1	% Contig 2	% Giant Panda Genome
LINES	21.5	49.1	18.2
SINEs	12	6.3	7.9
LTRs	3.6	5.6	5.6
DNA _s	2.1	2.8	3.2
Unknown	0	0	0.1
Total	39.2	63.8	34.7

LINES, long interspersed nuclear elements; SINEs, short interspersed nuclear elements; LTRs, long terminal repeats; DNA_s, DNA transposons.

Comparative Alignment with Polar Bear Y-Linked Scaffolds

We aligned the two contigs with all the polar bear genomic scaffolds that were previously identified to be Y-linked (Bidon et al. 2015). The majority of Contig 1 showed synteny with polar bear scaffold 318, with only several breakpoints being found in the sequence (supplementary fig. S2, Supplementary Material online). Notably, the syntenic region covered the whole *ZFY* gene and ended within the *KDM5D* gene. The rest of Contig 1 shared identity with polar bear scaffold 297 (supplementary fig. S3, Supplementary Material online). The gene order of *ZFY* and *KDM5D* was largely the same between the giant panda Y chromosome and polar bear scaffolds, but *KDM5D* was separated into two incomplete parts in the polar bear assembly. Contig 2 showed synteny with polar bear scaffold 322 and 393 (supplementary figs. S4 and S5, Supplementary Material online), although no genes have been annotated on either of these scaffolds.

Synteny Between Contig 2 and the Giant Panda X Chromosome Scaffold

Owing to the presence of the putative *GPM6B* pseudogene on Contig 2, additional alignment was performed between Contig 2 and giant panda genomic scaffold 1,583, which bears the *GPM6B* gene (fig. 2). This alignment clearly shows synteny with the *GPM6B* sequences (highlighted in red). The syntenic region also appears to be interrupted by the insertion of LINES.

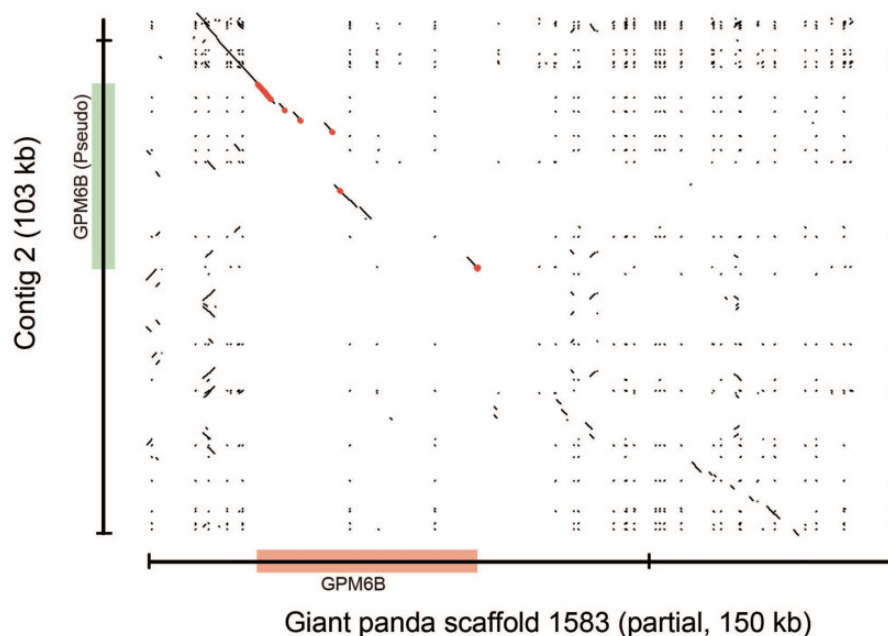


FIG. 2.—Comparative alignment between Contig 2 and giant panda scaffold 1,583. The red dots indicate similarity between Contig 2 and the giant panda neuronal membrane glycoprotein M6-B (*GPM6B*) exon sequences. Gene positions are shown in pink.

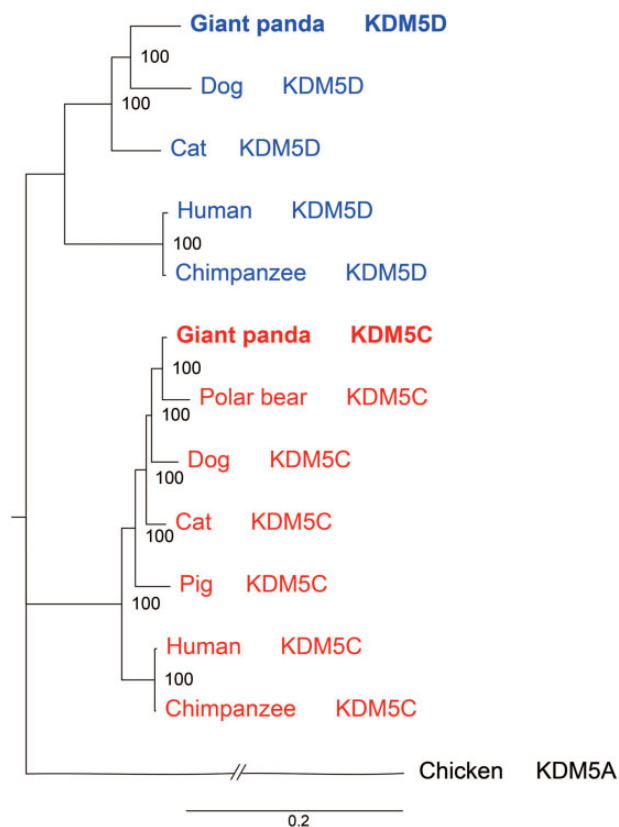


FIG. 3.—Bayesian tree of the mammalian lysine demethylase 5D (*KDM5D*) gene. Bayesian posterior probabilities (>75%) are shown beside the nodes. X- and Y-linked copies are marked in red and blue, respectively. The *KDM5A* sequence from chicken was used as an outgroup.

Phylogenetic Analyses of *KDM5D* and *ZFY*

The Bayesian and NJ trees for *KDM5D/KDM5C* show the same topology (fig. 3), and reciprocally monophyletic clusters of X/Y-linked genes are formed. As for *ZFY/ZFX*, the trees for exons 1–6 and those for exon 7 show distinctly different topologies, whereas there is only subtle distinction between Bayesian and NJ trees (figs. 4 and 5). Similar to the topology of the *KDM5D/KDM5C* trees, the X/Y-linked copies of *ZFY/ZFX* exon 1–6 formed reciprocally monophyletic clusters. However, in trees constructed for exon 7, the sequences for the giant panda, cat, pig, and cattle clustered together without any other gene copy, whereas all copies of Ursidae formed a separate clade.

To further investigate *ZFY/ZFX* exon 7 of the giant panda, a three-way alignment was performed using the exon sequences of giant panda *ZFX*, giant panda *ZFY*, and polar bear *ZFY*. Our data indicate that a total of 118 sites were variable in at least 2 sequences (fig. 5). Interestingly, while exon 7 makes up more than half (50.8%) of the *ZFY/ZFX* exon sequence, only 17 of these variable sites (14.4%) were located in this exon. For giant panda *ZFY*, 88 of the variable sites (87.1%) present in exons 1–6 are consistent with those in polar bear *ZFY*,

whereas only 6 sites (5.9%) in exons 1–6 are consistent with those in giant panda *ZFX*. Moreover, four of the variable sites (23.5%) in giant panda *ZFY* exon 7 are consistent with those in polar bear *ZFY*, whereas seven sites (41.2%) are consistent with those in giant panda *ZFX*.

Discussion

The phylogeny and unique biological traits of the giant panda, as well as their threatened status, have propelled this species to the forefront of multiple fields, including genomics. Unfortunately, our understanding of giant panda genetics is largely limited to the female sex. This bias is a consequence of the Y chromosome being difficult to sequence and assemble using standard shotgun sequencing. As a result, Y chromosome sequencing is usually performed using genomic libraries, such as bacterial artificial chromosome and fosmid libraries. By sequencing short clones separately, assembly is easier, particularly for regions with high abundance of repeat sequences. However, these methods have not yet been applied to the Y chromosome of the giant panda. In this study, using an Illumina HiSeq 2500, we sequenced selected fosmid clones from the male giant panda MSY at a depth of over 20,000 \times .

Theoretically, this extremely high depth ensures full coverage of the target sequence, thus generating a perfect assembly. Assuming that no sequence gaps were generated due to lack of coverage, we aligned two fragmented versions of the assembled sequence to form a complete assembly with no gaps. Moreover, such high-quality sequence assembly also revealed some new features of the sequencing itself. For example, by comparing the sequencing depth curves of overlapping clones, we found that the curve of a specific region maintains the same shape. The relative sequencing depth also appears to be fixed for each specific sequence, a phenomenon that is likely due to the unequal distribution of double strand breakage.

After sequencing, the TEs were analyzed in the two contigs to identify those that are evolutionarily conserved. Notably, TE coverage was vastly different in Contigs 1 (39.2%) and 2 (63.8%), and both are higher than that of the female giant panda genome (34.7%). These differences in repeat content between the two contigs lie mostly in the abundance of LINES. In fact, we observed a lower LINE-1 divergence rate between the annotated library and Contig 2 than that between the annotated library and Contig 1, indicating that there are more recent insertions in the nongenic region of the giant panda Y chromosome. This difference in repeat distribution also highlights unequal evolutionary pressure on these different regions of the Y chromosome, as the genic region is much more conserved compared with the nongenic region. As LINES are usually considered deleterious (Boissinot et al. 2001), our data might also reflect the biased accumulation of deleterious mutation on the Y chromosome, again showing different selection pressure between the two regions.

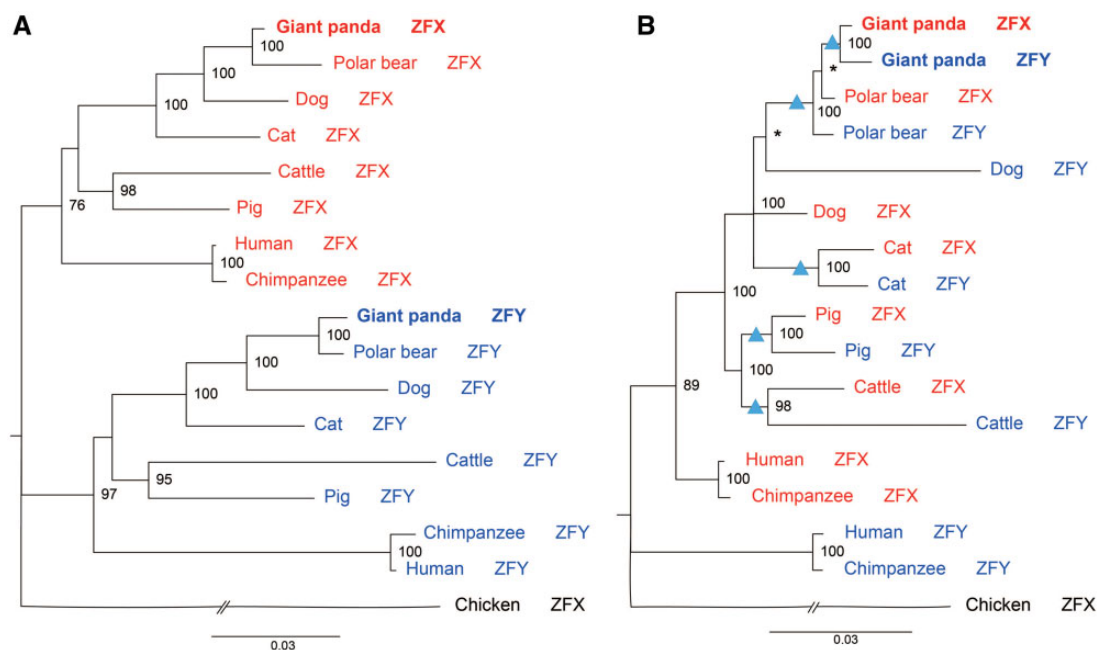


FIG. 4.—Bayesian trees of the mammalian zinc finger protein, Y-linked (*ZFY*) gene. Individual trees were constructed for exons 1–6 (A) and exon 7 (B). Bayesian posterior probabilities (>75%) are shown beside the nodes. *Different Neighbor-Joining (NJ) clade from that of the Bayesian tree. X- and Y-linked copies are marked in red and blue, respectively. Each blue triangle indicates an X–Y gene conversion event. The *ZFX* sequence from the chicken was used as an outgroup.

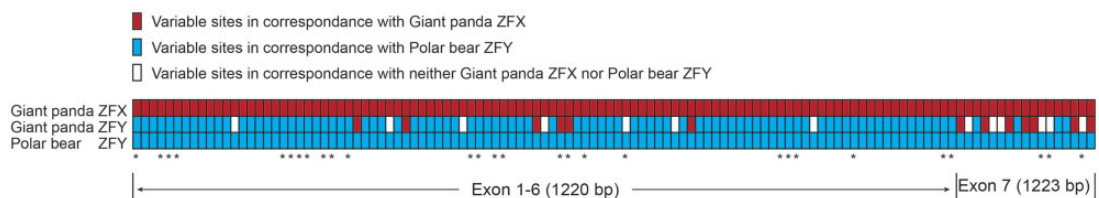


FIG. 5.—Three-way alignment between the coding sequences of giant panda zinc finger protein, X-linked (*ZFX*), giant panda zinc finger protein, Y-linked (*ZFY*), and polar bear *ZFY*. Distribution of sites that vary in at least two of the sequences are shown. The identical sites between giant panda *ZFY* and *ZFX* are shown in red, whereas those between giant panda *ZFY* and polar bear *ZFY* are shown in blue. *Nonsynonymous substitution between giant panda *ZFX* and polar bear *ZFY*.

It is important to note that assembly of the Y chromosome is easily disturbed by its repetitive nature, yielding small or chimeric scaffolds, and some sequences with repetitive elements might be completely missing (Bidon et al. 2015). To obtain further insight into the TEs of the Y chromosome, higher accuracy Y-linked sequencing is needed. Still, using our current deep sequencing technique, we have obtained the highest quality assembly of the giant panda Y chromosome to date.

Each contig was also annotated. Contig 1, containing the X-degenerate genes *ZFY* and *KDM5D*, is predicted to be located in the single copy X-degenerate region of the giant panda Y chromosome, whereas the nongenic Contig 2 has yet to be classified. The *ZFY* gene encodes a putative transcription factor which contains zinc fingers and is responsible

for important functions in male development and spermatogenesis (Decarpentrie et al. 2012). *KDM5D*, also known as *SMCY*, is involved in the process of spermatogenesis (Akimoto et al. 2008). These two genes belong to the core set of eutherian MSY genes that have been constrained by negative selection and retained in single-copy through evolution (Li et al. 2013). To better understand the evolution and conservation of these sequences, we performed a comparative analysis with sequences from the polar bear Y chromosome.

The two giant panda Y chromosome contigs showed conserved synteny with polar bear Y chromosome scaffolds, both in their genic and nongenic regions. This was unexpected as Y chromosome organization is usually highly variable between species because of rapid evolution (Skinner et al. 2016), and

gene order can be totally different even when comparing closely related species (Li et al. 2013). The sequence structure stability observed between the giant panda and polar bear indicates a close relationship. In fact, while many mammalian Y chromosomes are known to contain an evolutionarily conserved region spanning 2–3 Mb that carries important X ancestral genes, most of the sequences outside of this area are less conserved. Therefore, considering the synteny between Contigs 1 and 2 of the giant panda and the Y chromosome of the polar bear, it is possible that these Contigs 1 are located in the conserved region. Notably, the breakpoint generated by the separation of *KDM5D* on the polar bear scaffold could be a result of chromosomal variation or may have simply been caused by misassembly of the chimeric scaffold.

In addition to comparing the contigs to the Y chromosome of other species, it was also important to consider their relationship to other genomic regions in the giant panda. This was particularly important in the evaluation of Contig 2. In our analysis, we observed synteny between Contig 2 and giant panda genomic scaffold 1,583. The scaffold has yet to be positioned to a particular chromosome, but it has been predicted to be X-linked as the genes on scaffold 1,583 are usually X-linked in other mammalian genomes. It can be inferred that the synteny is most likely to be caused by transposition from the X chromosome. The corresponding region on scaffold 1,583 contains the *GPM6B* gene. However, the sequence in Contig 2 lacks two small exons, suggesting that this gene is pseudogenized on the Y chromosome. The *GPM6B* gene encodes a membrane glycoprotein. Although its position in the giant panda genome remains unclear, it has been shown to be X-linked in other mammal genomes.

A similar *GPM6B* X–Y transposition was observed on the Y chromosome in pigs (Skinner et al. 2016). In fact, an entire 90.25 kb region containing three genes, trafficking protein particle complex 2 (*TRAPPC2*), oral-facial-digital syndrome 1 (*OFD1*), and *GPM6B*, was transposed. While both *TRAPPC2* and *GPM6B* were pseudogenized, the *OFD1* gene, which is involved in cilia formation, remained functional. The dog Y chromosome contains a similar transposition of the *OFD1* gene and a *TRAPPC2* pseudogene relic (Li et al. 2013). This previous study also showed that *OFD1* expression was restricted to the testis in this species, and they speculated that this gene is the main driver of the transposition event (Li et al. 2013). Furthermore, two *GPM6B* pseudogenes have been found on the human Y chromosome, both with an *OFD1* pseudogene by its side, which further supports the theory that independent transposition events occurred carrying the same genes. Moreover, *OFD1* is thought to repeatedly acquire function in the process of spermatogenesis in multiple mammalian lineages (Skinner et al. 2016). Thus, it can be inferred that *OFD1* also exists on the giant panda Y chromosome through the process of gaining new genes. The absence of *OFD1* on Contig 2 can likely be attributed to the limited sequencing region.

Through the reconstruction of mammalian sex chromosome evolution (Bellott et al. 2014), we know that X–Y gene pairs formed at least four strata based on when recombination was suppressed. *ZFY* and *KDM5D* both belong to stratum 2/3, which means the differentiation of homologous X–Y gene pairs of *ZFY* and *KDM5D* should have happened at nearly the same time. Our analysis and tree construction for *KDM5D* and *ZFY* exons 1–6 had the same topology, which supports the conclusion that *KDM5D* and *ZFY* both diverged before the split between Laurasiatheria and Eurchontoglires. However, our phylogenetic analysis shows that the clustering pattern of *ZFY/ZFX* exon 7 differs from other exons and genes. This distinctive clustering likely reflects sequence homogenization caused by frequent gene conversion in this exon, indicating that a gene conversion event happened independently in the ancestral giant panda lineage as well as in other mammal lineages, including cat, pig, cattle, and ancestral Ursidae. From these data, we can infer that exon 7 of *ZFY/ZFX* is more evolutionarily conserved than other exons in this gene. Furthermore, three-way alignment of *ZFY* exon 7 also shows that an X–Y gene conversion event occurred in the giant panda. Gene conversion in the giant panda lineage has been reported before (Bidon et al. 2015), but the previous study utilized only partial exon sequences (397 bp). Our analysis using full length coding sequences of the *ZFY/ZFX* genes validates the gene conversion event in the giant panda lineage, and the range of the gene conversion event is made clear.

Gene conversion occurs when a gene sequence is overwritten by another sequence from a different region or chromosome. Gene conversion usually happens between homologous gene pairs during chromosome recombination and is sometimes considered deleterious when pseudogenes are the donors (Chen et al. 2007). Outside of the pseudoautosomal region, X–Y gene conversion is rare due to lack of recombination between the X and Y chromosomes, but intrachromosome gene conversion happens frequently on Y chromosome palindromes (Rozen et al. 2003). Intrachromosomal gene conversion events which preferentially use the ancestral state as the template may alleviate Muller's Ratchet (Skov et al. 2017). For homologous genes on the X and Y chromosomes, the X-linked copy carries less novel mutations compared with the Y-linked copy, as mutations tend to accumulate faster on the Y chromosome (Hurst and Ellegren 1998), and the same pattern was found in the *ZFY* gene (Erlandsson et al. 2000). Therefore, X–Y gene conversion resulted in conservation of the exon 7 sequence, where the copy of *ZFY* exon 7 was overwritten by the sequence of *ZFX* exon 7. Most studies of gene conversion, as well as our research, have focused on exonic sequences which bear sufficient homology to detect gene conversion events (Wright et al. 2014). Study of sheep and goat found no evidence for gene conversion in *ZFX* and *ZFY* introns (Lawson and Hewitt 2002), though it is hard to detect gene conversion in

noncoding regions due to the difficulties in aligning introns and noncoding sequences.

ZFY is among the core male-specific genes constrained by negative selection in eutherian genomes (Li et al. 2013). *ZFY* gene plays an important role in the process of spermatogenesis. Mouse *Zfy2* (one of the two copies of *ZFY* gene in mouse genome) was found to be the gene responsible for sperm formation and function, and mice without Y chromosome can initiate spermatogenesis with the presence of *Zfy2*, *Sry*, and *Eif2s3y* transgene (Yamauchi et al. 2015). Sheep with interrupted *ZFY* showed abnormal sperm morphology and biased sex ratio in offspring (Zhang et al. 2018), and *Zfy1* and *Zfy2* double knockout mice were infertile (Nakasuji et al. 2017). The important role of the *ZFY* gene indicates the need of maintaining its function. Exon 7, which is the last exon of *ZFY/ZFX* gene, makes up the 3' half of the gene, and encodes the zinc finger domain of the protein (Pamilo and Bianchi 1993). According to an exon by exon analysis of the Y-linked genes in multiple mammal species (Wilson and Makova 2009), only *ZFY/ZFX* exon 7 was found to undergo gene conversion. Interestingly, in cats, *ZFY* exon 7 gene conversion may cover nearly the whole exon or may only involve a small portion of the exon (Slattery et al. 2000). However, as gene conversion is a mechanism serves to maintain gene function and/or alleviate Y chromosome damage/degeneration and *ZFY* gene conversion only occurs in a fixed direction in a specific exon, we can infer that the zinc finger domain of *ZFY* that is encoded by exon 7 plays a special function related to sex.

Conclusions

In this study, deep sequencing of giant panda fosmid clones allowed the high-quality assembly of two giant panda Y chromosome contigs. Analysis revealed the different patterns of TE distribution between the two contigs and the giant panda genome, suggesting different evolutionary pressure between the genic and nongenic regions of the MSY. Through the alignment of Y chromosome sequences, we also found evidence of X–Y transposition and gene conversion events, which may act as a protection mechanism to maintain gene function and gain new genes. As sequencing techniques become more advanced and more Y chromosomes are sequenced, the processes involved in sex chromosome evolution can be further elucidated. While additional study is required, this study provides the first insight into the sequence and biology of the giant panda Y chromosome and highlights the position of this species in sex chromosome evolution.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Funding

This work was supported by a National Key Program (2016YFC0503200) from the Ministry of Science and Technology of China, a special grant for the giant panda from the State Forestry Administration of the People's Republic of China, and the Fundamental Research Funds for the Central Universities of the People's Republic of China.

Acknowledgments

We would like to thank Prof. Chang-Jun Zeng from Sichuan Agricultural University for assistance with experiments.

Literature Cited

- Asakawa S, et al. 1997. Human BAC library: construction and rapid screening. *Gene* 191(1):69–79.
- Akimoto C, et al. 2008. Spermatogenesis-specific association of *SMCY* and *MSH5*. *Genes Cells* 13(6):623–633.
- Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* 14(2):113–124.
- Bellott DW, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* 508(7497):494–499.
- Bidon T, et al. 2015. Genome-wide search identifies 1.9Mb from the polar bear Y chromosome for evolutionary analyses. *Genome Biol Evol* 7(7):2010–2022.
- Boissinot S, et al. 2001. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol Biol Evol* 18(6):926–935.
- Chen JM, et al. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet* 8(10):762–775.
- Decarpentrie F, et al. 2012. Human and mouse *ZFY* genes produce a conserved testis-specific transcript encoding a zinc finger protein with a short acidic domain and modified transactivation potential. *Hum Mol Genet* 21(12):2631–2645.
- Erlandsson R, et al. 2000. Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. *Mol Biol Evol* 17(5):804–812.
- Fang SG, et al. 2002. Genetic diversity of the giant panda (*Ailuropoda melanoleuca*) between big and small populations. *J Appl Anim Res* 21(1):65–74.
- Graves JAM. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* 124(5):901–914.
- Harris RS. 2007. Improved pairwise alignment of genomic DNA [PhD thesis]. The Pennsylvania State University.
- Hughes JF, Rozen S. 2012. Genomics and genetics of human and primate Y chromosomes. *Annu Rev Genomics Hum Genet* 13:83–108.
- Hurst LD, Ellegren H. 1998. Sex biases in the mutation rate. *Trends Genet* 14(11):446–452.
- Kumar S, et al. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7):1870–1874.
- Kutschera VE, et al. 2014. Bears in a forest of gene trees: phylogenetic inference is complicated by incomplete lineage sorting and gene flow. *Mol Biol Evol* 31(8):2004–2017.
- Lawson LJ, Hewitt GM. 2002. Comparison of substitution rates in *ZFX* and *ZFY* introns of sheep and goat related species support the hypothesis of male-biased mutation rates. *J Mol Evol* 54(1):54–61.
- Li G, et al. 2013. Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Res* 23(9):1486–1495.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li H, et al. 2009. The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics* 25:2078–2079.

- Li RQ, et al. 2010. The sequence and de novo assembly of the giant panda genome. *Nature* 463(7279):311–317.
- Nakasuji T, et al. 2017. Complementary critical functions of *Zfy1* and *Zfy2* in mouse spermatogenesis and reproduction. *PLoS Genet.* 13(1):e1006578.
- Nurk S, et al. 2013. Assembling genomes and Mini-metagenomes from highly chimeric reads. *International Conference on Research in Computational Molecular Biology*. Berlin, Heidelberg: Springer.
- Pamilo P, Bianchi NO. 1993. Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. *Mol Biol Evol.* 10(2):271–281.
- Rice P, et al. 2000. Emboss: the European molecular biology open software suite. *Trends Genet.* 16(6):276–277.
- Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3):539–542.
- Rozen S, et al. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423(6942):873–876.
- Skinner BM, et al. 2016. The pig X and Y Chromosomes: structure, sequence, and evolution. *Genome Res.* 26(1):130–139.
- Skov L, et al. 2017. Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion. *PLoS Genet.* 13(8):e1006834.
- Slattery JP, et al. 2000. Novel gene conversion between X-Y homologues located in the nonrecombining region of the Y chromosome in Felidae (Mammalia). *Proc Natl Acad Sci U S A.* 97(10):5307–5312.
- Wilson MA, Makova KD. 2009. Evolution and survival on eutherian sex chromosomes. *PLoS Genet.* 5(7):e1000568.
- Wright AE, et al. 2014. Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. *Evolution* 68(11):3281–3295.
- Xu W, et al. 2017. Reassessing the conservation status of the giant panda using remote sensing. *Nat Ecol Evol.* 1(11):1635–1638.
- Yamauchi Y, et al. 2015. Mouse Y-encoded transcription factor *Zfy2* is essential for sperm formation and function in assisted fertilization. *PLoS Genet.* 11(12):e1005476.
- Zhang YS, et al. 2018. A genetic method for sex determination in *Ovis* spp. by interruption of the zinc finger protein, Y-linked (*ZFY*) gene on the Y chromosome. *Reprod Fert Dev.* 30(9): 1161–1168.
- Zhang Z, Wei F. 2006. *Giant panda: ex-situ conservation theory and practice*: Beijing (China): Science Press.

Associate editor: Judith Mank