*Current Genomics*, 2022, *23*, 275-288

RESEARCH ARTICLE

# Establishment of New Genetic Markers and Methods for Sex Determination of Mouse and Human Cells using Polymerase Chain Reactions and Crude DNA Samples

Keyin Zhang[1,2], Jianglin Yang[3,4], Zhenwei Qin[5], Tianzu Lu[6], Didong Lou[5], Qianchuan Ran[5], Hai Huang[4], Shuqiang Cheng[4], Lucas Zellmer[7], Hong Ma[6,*] and Dezhong J. Liao[1,2,3,*]

*[1]Department of Pathology, School of Clinical Medicine, Guizhou Medical University, 4 Beijing Road, Guiyang 550004, Guizhou Province, P.R. China; [2]Department of Pathology, The Affiliated Hospital of Guizhou Medical University, 4 Beijing Road, Guiyang 550004, Guizhou Province, P.R. China; [3]Key Lab of Endemic and Ethnic Diseases of the Ministry of Education of China in Guizhou Medical University, Guiyang 550004, Guizhou Province, P. R. China; [4]Center for Clinical Laboratories, Guizhou Medical University Hospital, 4 Beijing Rd, Guiyang 550004, Guizhou Province, P.R. China; [5]Forensic Science Section, School of Basic Medical Sciences, Guizhou University of Traditional Chinese Medicine, Dong-Qing-Nan Road, Guiyang 550025, Guizhou Province, P.R. China; [6]Department of Stomatology, School of Stomatology, Guizhou Medical University, 4 Beijing Road, Guiyang 550004, Guizhou Province, P.R. China; [7]Department of Medicine, Hennepin County Medical Center, 730 South 8th St., Minneapolis, MN 55415*

ARTICLE HISTORY

Received: February 11, 2022
Revised: March 20, 2022
Accepted: April 04, 2022

*DOI:*
*10.2174/1389202923666220610121344*

**Abstract:** ***Background*:** The currently available methods for sexing human or mouse cells have weaknesses. Therefore, it is necessary to establish new methods.

***Methods*:** We used bioinformatics approach to identify genes that have alleles on both the X and Y chromosomes of mouse and human genomes and have a region showing a significant difference between the X and Y alleles. We then used polymerase chain reactions (PCR) followed by visualization of the PCR amplicons in agarose gels to establish these genomic regions as genetic sex markers.

***Results*:** Our bioinformatics analyses identified eight mouse sex markers and 56 human sex markers that are new, *i.e.* are previously unreported. Six of the eight mouse markers and 14 of the 56 human markers were verified using PCR and ensuing visualization of the PCR amplicons in agarose gels. Most of the tested and untested sex markers possess significant differences in the molecular weight between the X- and Y-derived PCR amplicons and are thus much better than most, if not all, previously-reported genetic sex markers. We also established several simple and essentially cost-free methods for extraction of crude genomic DNA from cultured cells, blood samples, and tissues that could be used as template for PCR amplification.

***Conclusion*:** We have established new sex genetic markers and methods for extracting genomic DNA and for sexing human and mouse cells. Our work may also lend some methodological strategies to the identification of new genetic sex markers for other organismal species.

**Keywords:** Sex determination, DNA extraction, genetic sex marker, genomic DNA, X chromosome, Y chromosome.

## 1. INTRODUCTION

There are many reasons for determining the genetic sex of human cells or tissues in biomedical, palaeobiological, and archaeological research as well as in forensic practice [1, 2]. For instance, cell lines of sex-specific tumors like breast, ovarian, or prostate cancer often need to be verified [2]. Mouse is the most commonly used animal species in genetic engineering for biomedical research; determination of the genetic sex of mouse cells, tissues, embryos, and young cubs is often essential and typically can only be done using genetic markers in the laboratory. There are some sophisticated methods for sex identification, such as restriction site-associated DNA sequencing (RAD-seq) [3, 4] or even whole genome sequencing. However, for the cells from the human, mouse, or other XX/XY species, the most commonly used methods are those utilizing Polymerase Chain Reactions (PCR) to amplify a DNA region shared by a Y-linked gene and its X-linked counterpart, followed by fractionation and visualization of the PCR products, dubbed as amplicons, in an agarose gel [1, 2, 5-8]. Most, if not all, of these PCR-involving methods, have a common weakness, *i.e.* an insufficient difference in the molecular weight, namely the molecular size, between the X- and Y-derived PCR amplicons. For instance, in a method that is most commonly used for determining the sex of mouse cells, the Y- and X-derived PCR amplicons differ only by 29 base pairs (bps) in size [5]. This size difference is too small, making it difficult to determine whether a PCR amplicon discerned in the agarose gel indicates a male or a female, especially when the whole panel of samples are all derived from single sex, *i.e.* are all males or all females. In another method that is frequently used for determining the sex of human samples, the targeted Y allele, *i.e.*, the AMELY gene, is often lost in male cancerous cells [1, 2, 9].

*Address correspondence to these authors at the Key Lab of Endemic and Ethnic Diseases of the Ministry of Education of China in Guizhou Medical University, Guiyang 550004, Guizhou Province, P.R. China;
Tel/Fax: 86-85186752814; E-mail: djliao@gmc.edu.cn and Department of Oral and Maxillofacial Surgery, School of Stomatology, Guizhou Medical University 9 Beijing Road, Guiyang 550004, Guizhou Province, P.R. China;
Tel/Fax: 86-851-88512238; E-mail: mahong@gmc.edu.cn

1875-5488/22 $65.00+.00 

A drawback in the determination of the genetic sex of human or mouse cells is the lack of thorough bioinformatics analyses to identify additional Y-linked genes that have one or more counterparts on the X chromosome and to identify the regions in these paired genes that differ significantly between the Y and X alleles; thus establishing them as effective sex markers. Therefore, we thoroughly interrogated the relevant literature and the NCBI (National Center for Biotechnology and Information of the United States) database of genes to search for such heterogametologous pairs in the mouse and human genomes, and then identified regions in the pairs that could be used as effective sex markers in a PCR approach. Moreover, because traditional techniques for isolation of DNA from cells or tissues involve phenol and chloroform and are relatively complicated, we also developed several simple and nearly cost-free methods for the isolation of crude genomic DNA from cultured cells, blood samples, and animal tissues that can be used as the template for the PCR amplification.

## 2. MATERIALS AND METHODS

### 2.1. Selection of Target Genes and Targeted Regions of the Genes

We searched the literature [10-18] and the NCBI database to identify genes that have alleles on both the Y and X chromosomes in mouse (Table **1**) and human (Table **2**) genomes. We retrieved the sequences of these paired Y- and X-linked genes from the NCBI database and used the CLUSTALW 2.1 (Multiple Sequence Alignment - CLUSTALW (genome.jp)) online software to align the sequences to identify regions that differ significantly between the X and Y alleles. Some genes were too large and thus were segregated into several parts for the online alignments. Three criteria were used for the selection of a genomic region as a potential sex marker: 1) The region should be about 150-1,100 bps in length so that it can be easily amplified with PCR while the resultant amplicon can still be clearly distinguished from the primer dimer in agarose gel. 2) The Y- and X-derived amplicons should have a large difference in size in an agarose gel. 3) Both the 5'- and 3'-flanks of the selected region should have a short sequence that is identical between the two alleles and is optimal for serving as a PCR primer, in terms of the gene specificity and annealing temperature. Examples of such aligned genomic regions are illustrated in Figs. (**1** and **2**). Unfortunately, most of the paired mouse and human genes we identified did not have a region meeting one or more of these criteria, although many of them still have some regions that show certain differences between the Y and X alleles.

### 2.2. *In silico* PCR

Once a pair of primers were designed, *in silico* PCR was performed using the Insilico PCR Function of the online software UCSC.GENOME.BROWSER (https://genome.soe.sucs.edu) to preliminarily determine the annealing temperature of the primers and determine the sequences and sizes of the PCR amplicons. Some of the primer pairs we designed did not produce results with this online software.

**Table 1. Mouse sex markers.**

| X-Linked | Y-Lined | Primer | Amplicon Sizes |
|---|---|---|---|
| Uba1 | Uba1y | mUba1-F: 5'-GTTGGCCTTGTGTGTCTGG-3' | X=561 bp, |
| | | mUba1-R 5'-TCAAAGCGATCCCACA-3' | Y= 725 and 722 bp |
| Uba1 | Uba1y-ps1 | mUba1-PS-F: 5'-TGGATGGTGTGGCCAATGC-3' | X=317 bp, |
| | | mUba1-PS-R: 5'-ATGGGGATGGATTTCTC-3' | Y=399-403 bp (5 copies) |
| Ddx3x | Ddx3y | mDdx3-F: 5'-GCCCATAACTTTTTACAAAG-3' | X & Chr6 = 332bp, |
| | | mDdx3-R: 5'-CTGTCTTTTGGTCAAGTTTC-3' | Y=489bp |
| Slxl1 | Sly | mSlxy-F1: 5'-CTTAGGAAGATATCATCCAGAT-3' | X=855 bp, |
| | | mSlxy-R1R: 5'-CACATAAAACCAGAGACACTG-3' | Y=384 bp |
| | | mSlxy-F2: 5'-TAGAGAAAACCAATCTA-3' | X=652-656 bp (many) |
| | | mSlxy-R2: 5'-AAAGGAATAGTTTTCCA-3' | Y=248-257 bp |
| Slx | | mSlxy-F8-3: 5'-TTTCATGAATTCATTGTTT-3' | X=807 bp, |
| | | mSlxy-R8-3: 5'-CCAGAGACACTGAAACTTAT-3' | Y=334 bp |
| | | mSlxy-F3: 5'-TCATTGTTTTTAATAGCTGTGTAGT-3' | X=797 bp, |
| | | mSlxy-R3: 5'-CCAGAGACACTGAAACTTATAGAG-3' | Y=324 bp |
| Rbm31x (Gm4916) | Rbm31y | mRbm31xy-F: 5'-GGGAACATACCCAAATGCTTTAGGAG-3' | X=244 bp, |
| | | mRbm31xy-R: 5'-CTGGCTTGTCCTGAAAACATTTGG-3' | Y=328 bp (2 copies) |

**Note:** The primer set for the Ddx3x/Ddx2y also amplifies a region of chromosome 6. The primer set for the Uba1/Uba1y-ps1 amplifies 5 regions of the Uba1y-ps1 with the resulting amplicons (copies) varying between 399 and 403 bps. The primer set for the Rbm31x(Gm4916)/Rbm31y amplifies 2 regions of the Rbm31y.

**Table 2. Human sex markers.**

| X-Linked | Y-Linked | Primer Sequence | Amplicon Size |
|---|---|---|---|
| ZFX | ZFY | ZFXY-F1: 5'-TGATCAAAGCAGTGCTTCA-3' | X=827 bp; Y=703 bp |
| | | ZFXY-R1: 5'-CACTGCACTCCAGCCTGGGC-3' | |
| | | ZFXY-F2: 5'-GCACATAGATGAGTCTGCTGGCC-3' | X=258 bp; Y=198 bp |
| | | ZFXY-R2: 5'-CATTCATGAATATCACTGAATT-3' | |
| | | ZFXY-F3: 5'-AATTCAGTGATATTCATGAATG-3' | X= 1027 bp; Y=664 bp |
| | | ZFXY-R3: 5'-CCATCAGGGCCAATAATTATTGC-3' | |
| NLGN4X | NLGN4Y | hNLGN4XY-F1: 5'-ATTGAAGATGAATCTGTACA-3' | X=354 bp; Y=681 bp |
| | | hNLGN4XY-R1: 5'-ATTTTTATGGATGATATGAC-3' | |
| | | hNLGN4XY-F2: 5'-GGTTGCTGGCAATTGGAAAA-3' | X=500 bp; Y=189 bp |
| | | hNLGN4XY-R2: 5'-GACAGAGCAAGACCCTGTCTCT-3' | |
| | | hNLGN4X-F3: 5'-GCTAAGAAATCTGAAAGCATTTGC-3' | X=196 bp; Y=645 bp |
| | | hNLGN4X-R3: 5'-CACTACCATTCAATCTTGGCC-3' | |
| | | hNLGN4X-F4: 5'-CTGAAGTCACAGAATGTTCTC-3' | X=174 bp; Y=335 bp |
| | | hNLGN4X-R4: 5'-GACTGCCTGGGAAATGCAAAC-3' | |
| DDX3X | DDX3Y | hDDX3-F1: 5'-AAGTAGTTTGGGTGGAAGA-3' | X=816 bp; Y=369 bp |
| | | hDDX3-R: 5'-TGGTATAAGAAATCCTCCAG-3' | |
| | | hDDX3-F3: 5'-CATCACACAGAAAGTAGTTTGGGTGG-3' | X=833 bp; Y=386 bp |
| | | hDDX3-R3: 5'-CCTTCATGGTATAAGAAATCCTCCAG-3' | |
| | | hDDX3-F4: 5'-CAGCAGAGGATTTGGTGGAGG-3' | X=277 bp; Y=589 bp |
| | | hDDX3-R4: 5'-ACCAGTCAACCCCCTGGGAGT-3' | |
| USP9X | USP9Y | hUSP9XY-F1: 5'-CTCTGCCTCCTGGGTTCA-3' | X=313 bp; Y=219 bp |
| | | hUSP9XY-R1: 5'-GGCAGGAGAATCGCTTGAA-3' | |
| | | hUSP9XY-F2: 5'-TTTGTCTGAAATGATTA-3' | X=449bp; Y=342 bp |
| | | hUSP9XY-R2: 5'-AATTAGCCAGGCATGGT-3' | |
| | | hUSP9-F3: 5'-TGAAAAAACTGCTTACATGTAATGG-3' | X=649 bp; Y=477 bp |
| | | hUSP9-R3: 5'-ATTTCTTCTGCTTCAGGCAACA-3' | |
| UTX | UTY | UTXY-F1: 5'-GAGGTTGCGGTGAGC-3' | X=563 bp; Y=297 bp |
| | | hUTXY-R1: 5'-AGCCTGGGCAACAGAG-3' | |
| TBL1X | TBL1Y | hTBL1-F1: 5'-AGTAGCTGGGACTACAGGCGC-3' | X=650 bp; Y=161 bp |
| | | hTBL1-R: 5'-TAATCCCAGCACTTTGGGAGG-3' | |
| TBL1X | TBL1YP1 | hTBL1XYP1-F: 5'-CCTGAGATTTGGGTGGGG-3' | X=620 bp; Y=512 bp |
| | | hTBL1XP1-R: 5'-TTGGACACTAGCCCAGGAGGTC-3' | |
| GYG2 | GYG2P1 | hGYG2-F: 5'-GACTTAAGAGGCCTAGATATGTG-3' | Y=738 bp; X=236 bp |
| | | hGYG2-R: 5'-GACTGCTTGAGCCTAGAAGTTC-3' | |
| | | hGYG2-F2: 5'-GGTAGATGATAGGTGGATAGATGG-3' | X=582 bp; Y=476 bp |
| | | hGYG2-R2: 5'-TCATCAGTCCATATCTATCCATCT-3' | |

**(Table 2) contd….**

| X-Linked | Y-Linked | Primer Sequence | Amplicon Size |
|---|---|---|---|
| ARSE (ARSL) | ARSLP1 | hARSL-F1: 5'-AGGCAGGAGAATCACTTGAACC-3' | F1/R: X=644 bp; Y=239 bp<br>F2/R: X=609 bp; Y=188 bp |
| | | hARSL-R: 5'-GAACACAGGCGTCACAAAGTGGA-3' | |
| | | hARSL-F2: 5'-TTGCACCACTGCACTCTAGCCTG-3' | |
| MXRA5 (ADLICAN) | MXRA5Y (ADLICANP) | hMXRA5-F1: 5'-AGGGGGTACTCAAGTGATTCCC-3' | X=565 bp; Y=238 bp |
| | | hMXRA5-R1: 5'-CTGGTTGATTTCCCTTCACATCTG-3' | |
| | | hMXRA5-F2: 5'-TTGACATTCAGCTATCCATGGCTT-3' | X=760 bp; Y=532 bp |
| | | hMXRA5-R2: 5'-AGGGGCAAACTCATCCATTGTTG-3' | |
| | | hMXRA5-F4: 5'-CACCTTGGCAGGTCCCATGCT-3' | X=922 bp; Y=217 bp |
| | | hMXRA5-R4: 5'-CCACCCTCATTTCCAGCCTG-3' | |
| | | hMXRA5-F5: 5'-TCTCACATCTGGCACCCTTGG-3' | X=509 bp; Y=174 bp |
| | | hMXRA5-R5: 5'-CTGGTTGATTTCCCTTCACATCTG-3' | |
| PRKX | PRKY | hPRK-F1: 5'-GGTTTGCTTTAGTATGAAGGACCCC-3' | Y=853 bp; X=539 bp |
| | | hPRK-R1: 5'-ACCAGCCTGGCCAACATGG-3' | |
| | | hPRK-F2: 5'-GTGACTCAGTATGCCCGAGCTCT-3' | X=207 bp; Y=486 bp |
| | | hPRK-R2: 5'-ATGACTGGGTGTGGTGGCTCCT-3' | |
| | | hPRK-F3: 5'-GATATTGGCAACTTTGGATGCACA-3' | X=522 bp; Y=222 pb |
| | | hPRK-R3: 5'-TTAGCTCTAGTTTGTGCCCAGGG-3' | |
| | | hPRK-F4: 5'-AGGATCACGCTGCAGTACTCCAG-3' | X=307 bp; Y=494 pb |
| | | hPRK-R4: 5'-CACTCCAGCCTGGGTGACAGAG-3' | |
| | | hPRK-F5: 5'-CACTGCAGTGTATTGGACAGTTCA-3' | X=603 bp; Y=261 bp |
| | | hPRK-R5: 5'-ACATGGGGACAGAGGCAGAGAC-3' | |
| | | hPRK-F6: 5'-GAGACAGGGTCTTGTTCTGTGTCC-3' | X=244 bp; Y=515 bp |
| | | hPRK-R6: 5'-AGGAGTTCAAGACCAGCCTGGAC-3' | |
| | | hPRK-F7: 5'-CTGGATGAACAGCTGGCAAGGA-3' | X=307 bp; Y= 589 bp |
| | | hPRK-R7: 5'-TTGCCCACCACTTTCCCGGA-3' | |
| | | hPRK-F7b: 5'-AGCCTGGGTGATAGAGCGAGA-3' | X=302 bp; Y=692 bp |
| | | hPRK-R7b: 5'-TGAATTACCTCCTTCCTGACAGG-3' | |
| | | hPRK-F8: 5'-CTGATGCCAAGTACTGGACGGG-3' | X=432 bp; Y=730 bp |
| | | hPRK-R8: 5'-GCATCCCTTCTCTCAGACCCTC-3' | |
| STS | STSP1 | hSTS-F1: 5'-TCATTCAGCAGCCCATGTCCTATG-3' | X=600 bp; Y=271 bp |
| | | hSTS-R1: 5'-GAGAGCACACTGCATTGCAATACC-3' | |
| ANOS1 (KAL1) | ANOS2P (KALP) | hANOS1-F1: 5'-AATAATTAGAAGGAAGATGGGCC-3' | X=437 bp; Y=571 bp and 253 bp |
| | | hANOS1-R1: 5'-TGGGCTCATTTGCAGGAGTTGG-3' | |
| | | hANOS1-F2: 5'-CCCCAAGCTTCCCTGTTCCCC-3' | X=499 bp; Y=641 bp |
| | | hANOS1-R1: 5'-TGGGCTCATTTGCAGGAGTTGG-3' | |
| TMSB4X | TMSB4Y | hTMSB4-F: 5'-CCCCCATCTTTAGAAAGGCTGG-3' | X=644 bp; Y=1013 bp |
| | | hTMSB4-R: 5'-TTTCCTTCCCTGCCAGCCAGA-3' | |

**(Table 2) contd….**

| X-Linked | Y-Linked | Primer Sequence | Amplicon Size |
|---|---|---|---|
| GPR143 | GPR143P | hGPR143-F: 5’-GGGTGGAGGGCTTTGGAGGCT-3’ | X=1105 bp; Y=497 bp |
| | | hGPR143-R: 5’-TCACATTTTTCTCTTTACTCCCC-3’ | |
| ARSF | ARSP1 | hARSF-F1: 5'-TATGTCTATGAATATATATTCA-3' | X= 836 bp and 87 bp; Y=366 bp |
| | | hARSF-R1: 5'-AGGAAACAGAATGAATTCTGCTG-3' | |
| | | hARSF-F2: 5'-TGTGCAGTTACCTGCTCTGGG-3' | X=324 bp; Y=256 bp |
| | | hARSF-R2: 5'-CAGAGAGAGGGAGAGAGAGAGAC-3' | |
| EIF1AX | EIF1AY | hEIF1A-F: 5'-CCGCCATGCCCAAGAATAAAG-3' | X=228 bp; Y=187 bp |
| | | hEIF1A-R: 5'-CCACCCCAATGTGACCAAA-3' | |
| SHROOM2 (APXL) | SHROOM2P1 (APXLP) | hAPXL-F: 5'-AGTAGGCTGCAGGCCTCTCT-3' | X=446 bp; Y=282 bp |
| | | hAPXL-R: 5'-CTGGGAGGCTCCTGGGCTCC-3' | |
| OFD1 | OFD1P6Y | hOFD1-F1: 5’-GCAAACTTTAGTGTTTGAAA-3’ | X=650 bp; Y=351 bp (2 copies) |
| | | hOFD1-R1: 5’-AAGGCACATGTTCACTGA-3’ | |
| | | hOFD1-F2: 5’-CAGTCAACTGTATGAATATTCATGC-3’ | X=284 bp; Y=605 bp (2 copies) |
| | | hOFD1-R2: 5’-ACAGTTTCTGAAGCCATGGTA-3’ | |
| OFD1 | OFD1P15Y | hOFD15Y-F1: 5'-ATGGATTTGCTAAGAGGAAGAGAAG-3' | X=434 bp; Y=515 bp |
| | | hOFD15Y-R1: 5’-AGCTGGTAAAAGCTGCAGG-3’ | |
| | | hOFD15Y-F2: 5'-GCTTTTAGCAAATAAGTGGATAATGC-3' | X=408 bp; Y=742 bp |
| | | hOFD15Y-R2: 5’-CTCTGCTGTGGGGCCCA-3’ | |
| OFD1 | OFD1P10Y | hORD10Y-F1: 5’-CAGATGGATTACTCTGTCCACTTAAC-3’ | X=594 bp; Y=320-321 bp (6 copies) |
| | | hORD10Y-R1: 5’-CTTTTCACCTTTTTGTCTTCAA-3’ | |
| PCDH11X | PCDH11Y | hPCDH11-F1: 5’-GAAAGGTGTGAGGCTCTAGATG-3’ | X=234 bp; Y=670 bp |
| | | hPCDH11-R1: 5’-CGAACTCCTGGCCTCAAGCGAT-3’ | |
| | | hPCDH11-F2: 5’-GTTGTCATTCATCGACTACGGGA-3’ | X=215 bp; Y=541 bp |
| | | hPCDH11-R2: 5’-TGAGTGCTTGCAGGGGAGAAAGT-3’ | |
| | | hPCDH11-F3: 5’-CTTCTTGGCACTTACATTGAAAGC-3’ | X=256 bp; Y=571 bp |
| | | hPCDH11-R3: 5’-GGAAATGTCTTATCCTAAAGCTGG-3’ | |
| | | hPCDH11-F4: 5’-TGGGTCTATTCTGATGACTGAGGT-3’ | X=438 bp; Y=781 bp |
| | | hPCDH11-R4: 5’-TGCCTGGCTGGAGAACATTGTA-3’ | |
| | | hPCDH11-F5: 5’-CCACTCGTTGAGTGATGGGC-3’ | X=382 bp; Y=232 bp |
| | | hPCDH11-R5: 5’-AGCACTAATGATCAGGGATTCC-3’ | |
| | | hPCDH11-F6: 5’-CCTGCACCTGCATTTCTTTTTGC-3’ | X=239 bp; Y=536 bp |
| | | hPCDH11-R6: 5’-CTCAAACTGTTCATCTGCAGTGTC-3’ | |
| RBMX | RBMY-pseudo[#] | hRBM-F1: 5'-GGAGGCAGAGGTTGCAGTGAG-3' | X=319 bp; Y=714 bp; 710 bp (many) |
| | | hRBM-R: 5’-TGGGAGGCTGAGGCAGGA-3’ | |
| PUDP | PUDPP1 | hPUDP-F: 5’-CCTCCATTGTATCTTGAACAAATCC-3’ | X=1003 bp; Y=161 bp |
| | | hPUDP-R: 5’-CTCACACCTCTGGAGGCTGGG-3’ | |
| MED14 | MED14P1 | hMED-F: 5’-TCCCAGCTACTTGGGAGGCTGA-3’ | X=1088 bp; X=431 bp |
| | | hMED-R: 5’-CTGCACTCCAGCCTGGG-3’ | |

**Note:** Some primer sets can amplify multiple regions of the X or Y chromosome, resulting in multiple amplicons (copies). [#]: RBMY-pseudogenes include the Y-linked RBMY1A1, RBMY1B, RBMY1D, RBMY1E, RBMY1F, RBMY1K (RBMY1KP), RBMY1HP, RBMY2DP, RBMY2CP, RBMY1J, and RBMY2VP, as well as a few other unannotated regions on the Y chromosome.


**Fig. (1).** Several examples of alignment of PCR-amplified regions of mouse gene pairs to illustrate the primer locations and the differences between the X- and Y-derived amplicons. Primer sequences are shaded. Dashes and dashed lines within a sequence indicate the nucleotides that are lacked by the allele. Stars below a sequence indicate the nucleotides that are identical between the X and Y alleles, whereas dots and other markers below a sequence indicate the nucleotides that differ between the two alleles.

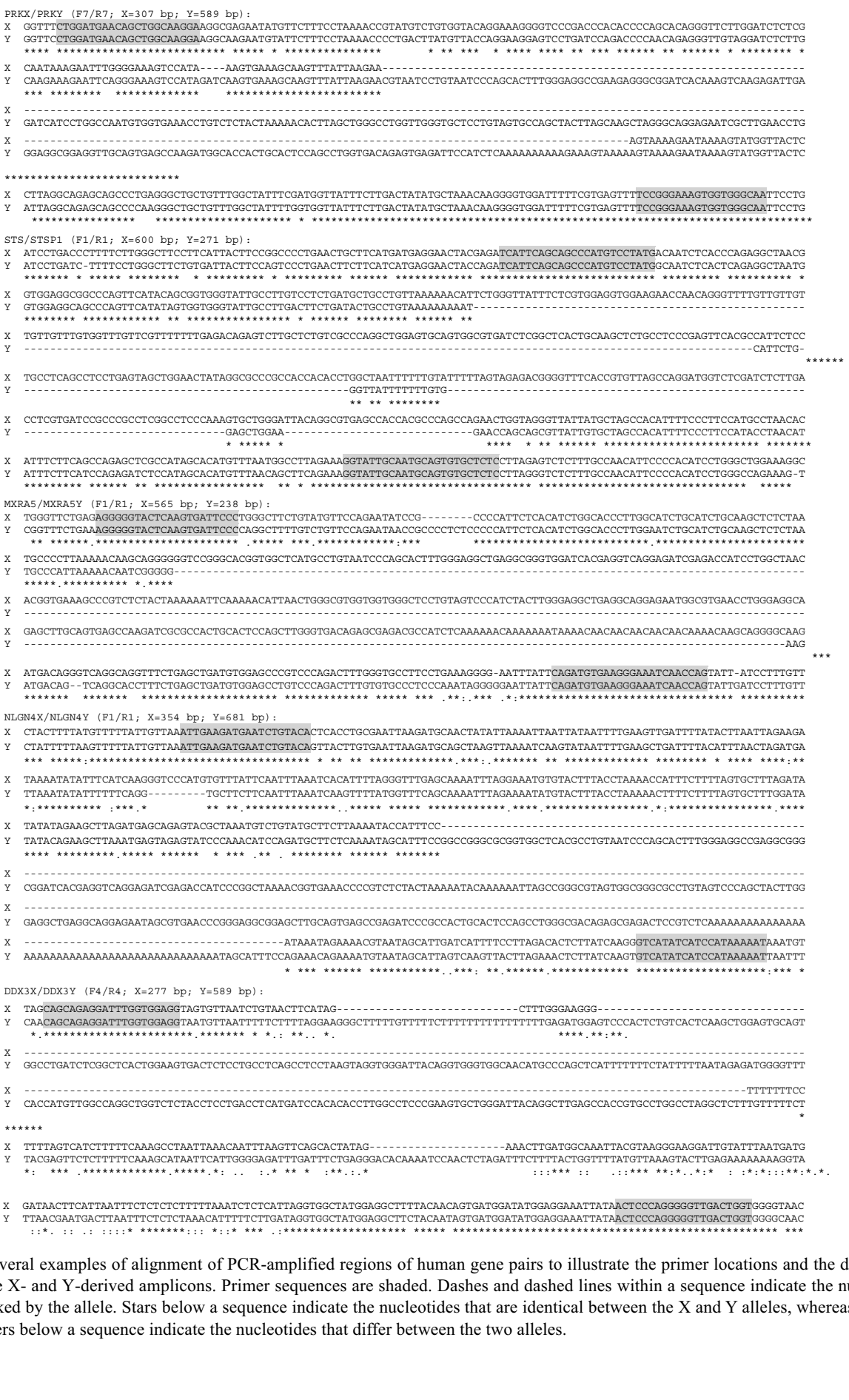


**Fig. (2).** Several examples of alignment of PCR-amplified regions of human gene pairs to illustrate the primer locations and the differences between the X- and Y-derived amplicons. Primer sequences are shaded. Dashes and dashed lines within a sequence indicate the nucleotides that are lacked by the allele. Stars below a sequence indicate the nucleotides that are identical between the X and Y alleles, whereas dots and other markers below a sequence indicate the nucleotides that differ between the two alleles.

### 2.3. Culture and Collection of Cells

This study included one male-origin cell line and one female-origin cell line, both of which were recently established by us from mouse skin tissues for other purposes irrelevant to the present study. The cells were cultured in flasks at 37°C in a Dulbecco's modified eagle medium (DMEM) containing 10% fetal bovine serum as routine. Cells reaching about 80% confluence were briefly treated with DMEM containing 0.25% trypsin-EDTA (Ethylene Diamine Tetraacetic Acid) so that they detached from the flask. The cells were then collected into a 15-ml tube and centrifuged at 5,000 rpm for 5 minutes at room temperature. After the supernatant was discarded, the cell pellet was suspended in 150 µl of phosphate-buffered saline (PBS) and transferred into a 1.5-ml Eppendorf tube. The tube was then centrifuged at the top speed (about 5,000 rpm) of a mini-centrifuge for 5 minutes at room temperature and the PBS was removed as much as possible with a pipette so that it would not significantly dilute the to-be-added cell lysis solution. The cell pellet was used for isolation of the nuclear DNA, as described below.

### 2.4. Isolation of DNA from Cultured Cells and Blood Samples

Inspired by several methods reported in the literature [19-22], we developed two simple solutions, dubbed as "the NaCl-SDS working solution" and "the NaCl-NP40 working solution", and developed a simple procedure for using these solutions to isolate DNA from cultured cells. To make the former solution, we made 1) a 10% SDS (sodium dodecyl sulfate) solution (5 ml) and 2) a saturated (about 6 M) NaCl solution by adding 2.5 grams of NaCl into 5 ml of $H_2O$ [6, 22]. Both solutions can be stored at room temperature. We then sequentially added 730 µl of $H_2O$, 250 µl of the saturated NaCl, and 20 µl of the 10% SDS into a 1.5-ml Eppendorf tube to make 1 ml of the NaCl-SDS working solution, which contained 0.2% SDS and about 1.5 M NaCl. It is important to make the working solution and use it immediately, as SDS may form cloudy precipitates in a high concentration of NaCl. Should precipitates appear, one could use a 200-µl tip to stir the solution and mix it well. To make the latter solution, we first made 1 ml of 10% Nonidet ® P40 (Cat # A100109-0100, Diamond Sangon Biotech (Shanghai) Co. Ltd., China; www.sangon.com) by adding 900 µl of $H_2O$ and 100 µl of the NP40 into a 1.5-ml tube. Since NP40 is very sticky, it is best to cut off part of a 200-µl tip to facilitate pipetting it and to vortex the tube to dissolve the NP40 into the water. A NaCl-NP40 working solution was made by sequentially adding 730 µl of $H_2O$, 250 µl of the above-described saturated NaCl, and 20 µl of the 10% NP40 into a 1.5-ml tube. The solution carried 0.2% NP40 and about 1.5 M NaCl.

The above-described pellet of cultured cells in a 1.5-ml tube was added to 100 µl of the NaCl-SDS or the NaCl-NP40 working solution. A pipette was used to mix the cells with the solution, followed by vortexing for tens of seconds to break the cells and their nuclei. The tube was then put into boiling water for five minutes or into a heating block preset at 95°C for 10 minutes and, during which, was vortexed once or twice to release DNA from nuclear proteins and to denature the DNA. The tube was then spun in a mini-centrifuge at the top speed (about 5,000 rpm) for two minutes to precipitate proteins. The supernatant, which contained DNA, was transferred into a new 1.5-ml tube, followed by the addition of 1.4 ml of 70% ethanol to precipitate the DNA and remove salts from the DNA. The DNA was precipitated by centrifuging the tube at 13,000 rpm for 10 minutes at 4°C. After the ethanol was discarded as much as possible, the tube was spun in a mini-centrifuge for a few seconds and the extant ethanol in the tube was aspirated using a pipette with a 10-µl tip. The DNA pellet that was now basically free of ethanol was dissolved by adding 20-25 µl of $H_2O$ with subsequent vortexing. A volume of 2 to 3 µl of this DNA solution should be sufficient as the template for a 20-µl volume of PCR system.

We also used the above-described methods to extract DNA from human blood samples. We collected 100 µl of anticoagulated blood samples that were provided by individuals during their regular health checkups and were discarded leftovers from all clinical uses, with notification to and permission from the individuals about this study prior to the use of the blood samples. We added 200 µl of distilled water into the 100 µl blood in a 1.5-ml tube and stored the tube at room temperature for 5 minutes so that the erythrocytes were lysed in the hypotonic condition, whereas the leucocytes should largely remain intact as they have been shown to be more refractory to hypotonic shock than erythrocytes [23-25]. The tube was then centrifuged with a mini-centrifuge at the top speed (about 5,000 rpm) for 2 minutes at room temperature to precipitate the leucocytes, along with a small amount of erythrocyte debris. After the supernatant (which mostly contained lysed erythrocytes) was discarded, 100 µl of the above-described NaCl-SDS or NaCl-NP40 working solution was added into the tube, followed by vortexing the tube to break the leucocytes and release their DNA from nuclear proteins. The remaining procedure was the same as described above for cultured cells.

### 2.5. Isolation of DNA from Tissues

The above-described procedure entailing the NaCl-SDS or NaCl-NP40 working solution could also be used to isolate DNA from frozen or fresh tissues; however, the yield of DNA was low if the tissue was not well homogenized. Since tissue homogenization is relatively complicated and involves additional reagents or equipment such as a polytron, we chose to use a proteinase K (PK) approach modified from previously described methods [7, 19]. Several milligrams of mouse tissue, such as roughly 5 mm of the tail tip of a weaning mouse or a piece of liver tissue collected from mouse carcasses that were discarded from terminated animal studies, were collected and fractioned into small pieces using a surgical blade. The tissue shreds were put into a 1.5-ml tube containing 50 µl of a PK solution that contained 10 mM Tris (pH 8.0), 1 mM EDTA (pH 8.0), and 0.4 µg /µl of PK (Cat. # P1120, the Solarbio Life Sciences, Beijing, China; http://www.solarbio.net). The tube was incubated for 4 to 6 hours at 56°C in a heating block for the PK to digest proteins, followed by incubation of the tube at 95°C for additional 15 minutes to inactivate the PK. A volume of 2 to 3 µl of this crude DNA solution was directly used as the template in a 20-µl volume of PCR system.

### 2.6. PCR and Electrophoresis of PCR Products in Agarose Gels

PCR was performed with a volume of 20 µl containing 2 or 3 µl of crude DNA as the template. The DNA was initially denatured at 95˚C for 5 minutes, followed by 40 cycles of 30 seconds at 95˚C (for DNA denaturation), 20 seconds at 52-58˚C (for primer annealing, varying among different sex markers as shown in Table **3**), and 30 seconds at 72˚C (for elongation). A final extension at 72˚C was set for 3 minutes. PCR products were fractionated by electrophoresis in a 1% agarose gel containing 0.01% of ExRed, a fluorescent dye (Beijing Zoma Biotechnology Co., Ltd, Beijing, China; Catalog# ZS203-1; www.zomabio.com) that visualized the DNA during electrophoresis.

### 2.7. Sequencing of PCR Amplicons

We selected some PCR amplicons, especially those that appeared at unanticipated positions on agarose gels and thus might be off-targets, for DNA sequencing to determine their identities. The DNA band of interest in an agarose gel was excised and purified using a simple method we described in detail before, in which the DNA-containing gel slice was centrifuged to separate the DNA from the agarose [26, 27]. The purified DNA was sent (Diamond Sangon Biotech (Shanghai) Co. Ltd., China; www.sangon.com) for direct sequencing from both ends using the PCR primers. The resulting sequences were analyzed using the BLAT function of the online software UCSC.GENOME.BROWSER (https://genome.soe.sucs.edu) to determine which gene the sequence belonged to. If the amplified gene was an off-target, the sequence would be aligned with the sequences of the targeted Y- and X-linked genes using the abovementioned Multiple Sequence Alignment function of the online CLUSTALW software. This allowed for further analysis of whether the gene was mis-amplified due to certain similarities of our primers to the gene's sequence, and whether modification of the primer sequences or the PCR conditions would be able to solve the off-target issue.

## 3. RESULTS

### 3.1. Identification of Paired Genes with a Regional Difference Between the X and Y Alleles

Through analysis of the literature and gene sequences in the NCBI database, we found a total of 25 pairs of mouse genes (Table **1**) and 52 pairs of human genes (Table **2**) that have regional difference(s) between the Y and X alleles. Besides these genes, there are many genes that are identical, or differ only by single nucleotides (*i.e.*, single nucleotide polymorphism; SNP), between the X and Y alleles, however, these genes could not be used as sex markers and thus were not further analyzed. Some genes, such as the mouse Slx, Sly, and Rbmy as well as the human RBMX and UPS9X, have multiple copies, mostly as pseudogenes, on the same (X or Y) chromosome and thus can be paired differently. On the other hand, several genes that had previously been reported in the literature were not included because we could not find them in the NCBI database; such as the human UBE1Y, which was reported to be the male counterpart of the X-linked UBE1X (UBA1), and the mouse Sstx, which was reported to be the female counterpart of the Y-linked Ssty1 and Ssty2. We used the BLAT function of the UCSC.GENOME.BROWSER software to search for sequences similar to these genes on the opposite chromosome and were unable to find their counterpart allele.

**Table 3. PCR conditions for the tested genetic sex makers.**

| Species | Marker | Denature | Annealing | Elongation |
|---|---|---|---|---|
| Mouse | Uba1/Ubaly | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| | Ddx3x/Ddx3y | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| | Slxl1/Sly | 95˚C, 30 sec | [#]52-57˚C, 20 sec | 72˚C, 30 sec |
| | Slx/Sly | 95˚C, 30 sec | 62˚C, 20 sec | 72˚C, 30 sec |
| | Rbm31x/Rbm31y | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| Human | NLGN4X/NLGN4Y | 95˚C, 30 sec | [#]56-58˚C, 20 sec | 72˚C, 30 sec |
| | DDX3X/DDX3Y | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| | ARSL/ARSLP1 | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| | MXRA5/MXRA5Y | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| | PRKX/PRKY | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| | STS/STSP1 | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| | ANOS1/ANOS2P | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| | TMSB4X/TMSB4Y | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |
| | OFD1/OFD1P10Y | 95˚C, 30 sec | 58˚C, 20 sec | 72˚C, 30 sec |

[#]: The annealing temperature for these primer sets can be changed with the range without affecting the efficacy and specificity, allowing the marker to be performed together with other marker(s) in the same PCR cycler.

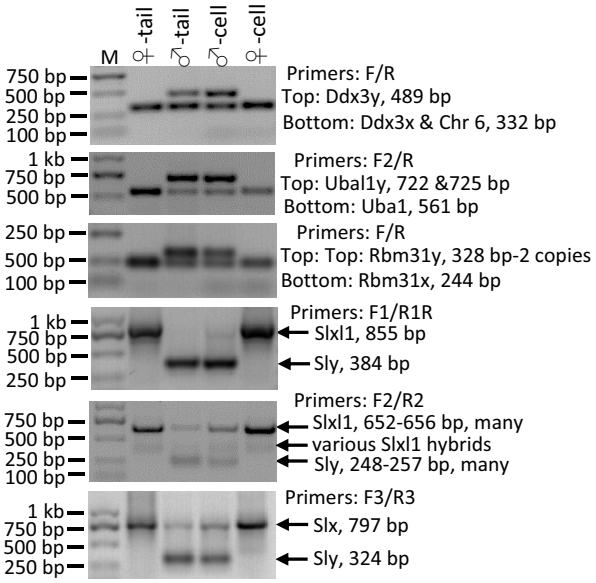


**Fig. (3).** PCR verification of six mouse sex markers, with detailed information of the primer sequences and amplicon sizes given in Table **1** and PCR conditions given in Table **3**. "Tail" and "Cell" mean that the template DNA was extracted from mouse tail tissues or mouse cell lines, respectively. "M" indicates marker of molecular weights.

### 3.2. Identification of Genomic Regions as Sex Markers

Of the 25 pairs of mouse genes identified, we were able to identify six with one or more regions allowing us to design primers for PCR amplification to produce X- and Y-derived amplicons with significant differences in molecular weight (size). Because the gene pairs Slxl1/Sly and Slx/Sly have several such genomic regions, a total of eight sex markers were identified. The sizes of the X and Y amplicons of these markers are listed in Table **1**, with their sequences and primer locations shown in the Supplementary Word Document. We performed PCR to verify six of the eight sex markers, with the results shown in Fig. (**3**). Of these tested sex markers, the Rbm31x-derived amplicon (244 bps) and the Rbm31y-derived amplicon (328 bps) differed by only 84 bps (Table **1** and Figs. **1** and **3**); this size difference is too small and requires known male and female samples to be included as references to indicate the positions of the X- and Y-derived amplicons on the agarose gel.

Similarly, we were able to identify regions in 25 of the 52 pairs of human genes, with regions showing significant differences between the X and Y alleles. Because some of the gene pairs had several such regions, a total of 56 human sex markers were identified, with their PCR amplicon sizes and primer sequences listed in Table **2** and their amplicon sequences and primer locations shown in the Supplementary Word Document. Some of the markers were less optimal than the others with regards to the difference in the sizes between the X- and Y-derived amplicons or to the PCR conditions, particularly the annealing temperature associated with the AT to GC ratio in the primer sequence. We performed PCR to verify 14 of the 56 markers, with the results shown in Fig. (**4**). One of the sex markers, F3/R3 of the DDX3-DDX3Y, also amplified the X-linked gene FTX as an off-target (Fig. **4**).

Some of the genes have multiple alleles on the X or Y chromosome or have multiple sequence repeats within the same allele. We were able to design a single primer pair that

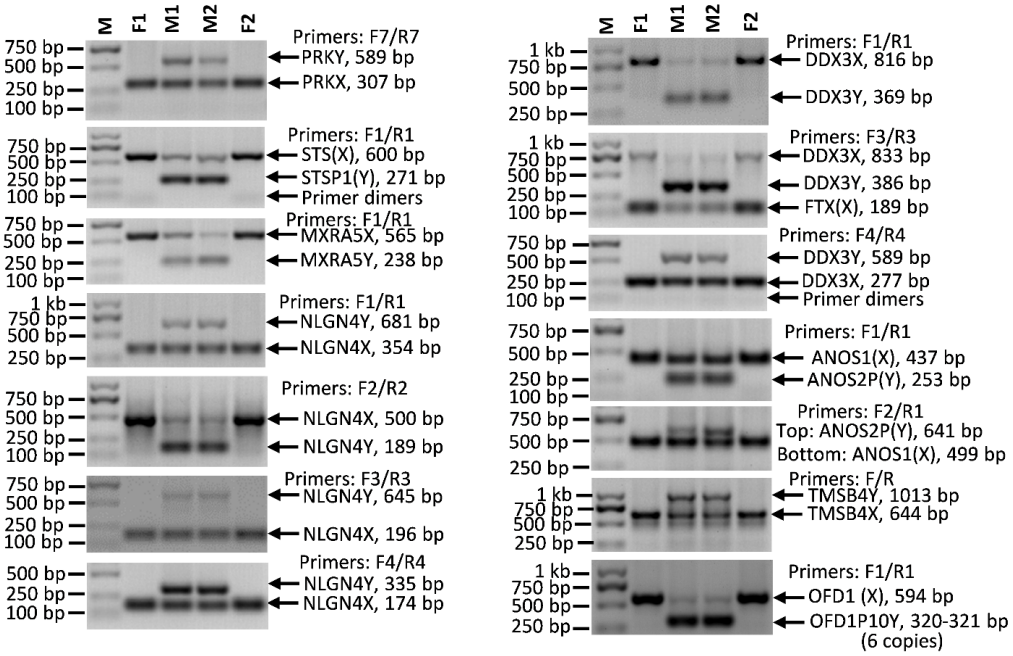


**Fig. (4).** PCR verification of 14 human sex markers, with detailed information of the primer sequences and amplicon sizes given in Table **2** and PCR conditions given in Table **3**. One of the markers, the F3/R3 of DDX3X-DDX3Y, also amplifies the X-linked gene FTX as an off-target. "F1" and "F2" indicate human blood samples from two different females, whereas "M1" and "M2" indicate blood samples from two different males. "M" indicates marker of molecular weights.

could amplify two or more of these alleles or these sequence repeats. For instance, the primer pair hORD10Y-F1/ hORD10Y-R1 could amplify six Y-linked copies (Table **2**), whereas the primer pair F2/R2 for the mSlxl1/mSly could amplify multiple X alleles and multiple Y alleles (Supplementary Word Document).

The remaining 19 pairs of the mouse genes and the remaining 27 pairs of the human genes from which we were unable to select a good sex marker are listed in the Supplementary Tables **1** and **2**, respectively. Many of these genes still encompass regions that differ between the X and the Y alleles and, in certain situations, may still be used as sex markers although they are less-optimal than those listed in Tables **1** and **2**.

## 4. DISCUSSION

Positive PCR results shown in Figs. (**3** and **4**) showcase that our simple methods for extraction of crude DNA from cells, blood samples, or tissues work sufficiently for PCR amplification. Our NaCl-SDS and NaCl-NP40 methods for cultured cells or blood samples are more simple and more cost-effective compared with many methods of this kind. SDS is cheaper and more assessable compared to NP40, whereas the NaCl-NP40 working solution seems easier to make and to use as the NaCl-SDS may form cloudy precipitates.

The sex markers we identified, especially those that have been verified and shown in Figs. (**3** and **4**), are preferred over most, if not all, of those reported in the literature due to their much larger size differences between the X- and Y-derived PCR amplicons. Most of our sex markers have a large enough difference in the amplicon sizes to allow omission of positive controls, making the performance simpler and less expensive. The sizes of the amplicons are less than 1 kb in most cases, which allows the targeted genomic regions to be amplified easily in a short time. The primers (Table **3**) we designed allow researchers to use the same PCR conditions, mainly the same annealing temperature, to amplify several sex markers in a single performance of PCR. Some of the genes still have some nucleotides at the 5' and/or the 3' flanks of a primer that are homologous between the X and Y alleles, as shown in Figs. (**1** and **2**) and the Supplementary Word Document, which provides researchers with opportunities to modify the primers for different reasons as needed, such as adjusting the annealing temperature or the length of the primer.

For two reasons we recommend that several markers should be simultaneously used for sex determination, with those markers that amplify two or more Y-linked alleles as the first choices since identification of the Y chromosome is the key point of sex determination. First, the efficiency of a PCR amplification depends not only on the primer sequence but also on the size and sequence of the to-be-amplified DNA template. Therefore, a given set of PCR conditions that work for both X and Y alleles may be optimal for, and thus may preferentially amplify, only one of the two. For example, the marker F1/R1 of the ANOS1/ANOS2P should, theoretically, also yield a Y-derived 571-pb band (Table **2** and Supplementary Word Document), but this band is often missing (Fig. **4**). Second, it is well known that some Y-linked genes or even the whole Y chromosome may have been lost in various pathological conditions [14, 16, 28], especially in some cancers [13]. Actually, about 20% of older men have lost the Y chromosome in certain cell types [29], and quite a few cell lines in the depository of the American Type Cell Culture are known to be derived from male patients but test negative for some male marker genes, such as the amelogenin [30]. Therefore, use of only one marker may result in misleading data.

In forensic applications and in palaeobiological or archaeological research, there are often situations wherein the DNA template is so degraded that the sex markers shown in Figs. (**3** and **4**) may not result in positive PCR data. In these scenarios, some of the untested sex markers may be tried, and new markers may be selected from the gene pairs listed in Supplementary Tables **1** and **2**. Since DNA degradation occurs randomly, some of these untested markers, unselected genomic regions, or unselected gene pairs may be less degraded. PCR can splice fragmented DNA templates [31-33], and thus less-degraded genomic regions have higher chances of being amplified than more severely degraded ones. However, the PCR conditions for amplifying intact DNA may differ from those for fragmented DNA. For instance, it is likely to require a longer time for a long, intact DNA double helix to be denatured at 95°C. Therefore, researchers who want to use some of the untested markers from Tables **1** and **2**, or even select new markers from the gene pairs in these tables or in the Supplementary Tables **1** and **2**, need to optimize the PCR conditions according to their DNA template.

Determination of the genetic sex of cells, which usually requires identification of new sex markers [3, 34-36], is needful in a great variety of areas of not only biology research but also agriculture, industry, *etc*., and involves a large range of species of animals and dioecious plants. For example, dairy farms favor female sperms in cow breeding because only the females produce milk, and hop farmers try hard to eliminate the male plants during hop breeding because beer brewing industry needs only the female cones [34]. Like what we present herein, the determination of genetic sex in all of these organismal species should entail huge bioinformatic or computational work: Sex-linked gametologous pairs need to be identified first, followed by the identification of optimal sequence gap(s) as potential sex markers between the two sex alleles of each pair and identification of sequences as optimal forward and reverse primers. PCR conditions will then be designed based on the amplicon size and primer sequences; optimal methods and relevant conditions to visualize the PCR products, such as the concentration of the agarose gel, need to be estimated. Probably, biology experts and bioinformatics wizards should collaborate to establish certain computational tools, like those machine learning tools that have been developed for chemical docking simulation [37, 38], to facilitate each of these lines of bioinformatic work. More importantly, computational work is needed to integrate 1) establishment of gametologous pairs, 2) identification of new sex markers, 3) primer design, 4) PCR condition estimation, and 5) visualization of the PCR products into a single bioinformatic tool. For instance, the human TSPY has 30-60 tandem repeats on

the Y [14, 18]; each of these repeats need to be aligned with its X counterpart TSPX (TSPYL2) to identify optimal pairs, which may be aided by computer. Moreover, different species differ greatly from one another in sexing strategy because sex chromosomes evolve quickly [39, 40]. Some organisms are male-heterogametic (XX/XY) species like humans, mice and plants, whereas some others are female-heterogametic (ZZ/ZW) species like birds [39, 40]. More complicatedly, there are also species of which the haploid females have the U chromosome whereas the haploid males have the V chromosome [39, 40]. Such a huge organismal variation in sex chromosomes may require specialization of our work on sex marker identification, including establishment of computational tools, for each particular species.

## CONCLUSION

In summary, our bioinformatics analyses have led to the identification of new sex markers for the mouse and human cells, with some of them verified using PCR. We have also established several simple methods for extracting crude DNA from cultured cells, blood samples, and animal tissues to serve as the templates for PCR amplification. These sex markers and DNA extraction methods possess certain advantages when compared with those reported in the literature. Our work may lend some methodological strategies to other researchers for identification of new genetic sex markers and ensuing the establishment of new methods for sex determination of other organismal species.

## AUTHORS' CONTRIBUTIONS

KYZ performed most of the study and drafted the manuscript. JLY, ZWQ and TZL performed parts of the study, prepared the figures and tables, and participated in the discussion. LZ edited the manuscript and contributed to the discussion. HY participated in the discussion and helped draw a conclusion. HM and DJL conceptualized the manuscript. DZL finalized the manuscript.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was approved by the Ethics Committee of School of Stomatology, Guizhou Medical University of China (no.2021-26).

## HUMAN AND ANIMAL RIGHTS

Not applicable.

## CONSENT FOR PUBLICATION

Not applicable.

## AVAILABILITY OF DATA AND MATERIALS

Not applicable.

## FUNDING

This work was supported by a grant (No. 81660501 and 82060489) to Dezhong Joshua Liao.

## CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

## ACKNOWLEDGEMENTS

We would like to thank Dr. Fred Bogott at the Austin Medical Center, Mayo Clinic in Austin, Minnesota, USA, for his excellent English editing of this manuscript.

## SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

## REFERENCES

[1] Shah, K.; McCormack, C.E.; Bradbury, N.A. Do you know the sex of your cells? *Am. J. Physiol. Cell Physiol.,* **2014**, *306*(1), C3-C18. http://dx.doi.org/10.1152/ajpcell.00281.2013 PMID: 24196532

[2] Durkin, A.S.; Cedrone, E.; Sykes, G.; Boles, D.; Reid, Y.A. Utility of gender determination in cell line identity. *In Vitro Cell. Dev. Biol. Anim.,* **2000**, *36*(6), 344-347. http://dx.doi.org/10.1290/1071-2690(2000)036<0344:UOGDIC>2.0.CO;2 PMID: 10949990

[3] Gamble, T.; Zarkower, D. Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Mol. Ecol. Resour.,* **2014**, *14*(5), 902-913. http://dx.doi.org/10.1111/1755-0998.12237 PMID: 24506574

[4] Feron, R.; Pan, Q.; Wen, M.; Imarazene, B.; Jouanno, E.; Anderson, J.; Herpin, A.; Journot, L.; Parrinello, H.; Klopp, C.; Kottler, V.A.; Roco, A.S.; Du, K.; Kneitz, S.; Adolfi, M.; Wilson, C.A.; McCluskey, B.; Amores, A.; Desvignes, T.; Goetz, F.W.; Takanashi, A.; Kawaguchi, M.; Detrich, H.W., III; Oliveira, M.A.; Nóbrega, R.H.; Sakamoto, T.; Nakamoto, M.; Wargelius, A.; Karlsen, Ø.; Wang, Z.; Stöck, M.; Waterhouse, R.M.; Braasch, I.; Postlethwait, J.H.; Schartl, M.; Guiguen, Y. RADSex: A computational workflow to study sex determination using restriction site-associated DNA sequencing data. *Mol. Ecol. Resour.,* **2021**, *21*(5), 1715-1731. http://dx.doi.org/10.1111/1755-0998.13360 PMID: 33590960

[5] Clapcote, S.J.; Roder, J.C. Simplex PCR assay for sex determination in mice. *Biotechniques,* **2005**, *38*(5), 702-706, 704, 706. http://dx.doi.org/10.2144/05385BM05 PMID: 15945368

[6] Lambert, J.F.; Benoit, B.O.; Colvin, G.A.; Carlson, J.; Delville, Y.; Quesenberry, P.J. Quick sex determination of mouse fetuses. *J. Neurosci. Methods,* **2000**, *95*(2), 127-132. http://dx.doi.org/10.1016/S0165-0270(99)00157-0 PMID: 10752483

[7] McFarlane, L.; Truong, V.; Palmer, J.S.; Wilhelm, D. Novel PCR assay for determining the genetic sex of mice. *Sex Dev.,* **2013**, *7*(4), 207-211. http://dx.doi.org/10.1159/000348677 PMID: 23571295

[8] Navarro-Romero, M.T.; Muñoz, M.L.; Alcala-Castañeda, E.; Terreros-Espinosa, E.; Domínguez-de-la-Cruz, E.; García-Hernández, N.; Moreno-Galeana, M.Á. A novel method of male sex identification of human ancient skeletal remains. *Chromosome Res.,* **2020**, *28*(3-4), 277-291. http://dx.doi.org/10.1007/s10577-020-09634-1 PMID: 32621020

[9] von Wurmb-Schwark, N.; Bosinski, H.; Ritz-Timme, S. What do the X and Y chromosomes tell us about sex and gender in forensic case analysis? *J. Forensic Leg. Med.,* **2007**, *14*(1), 27-30. http://dx.doi.org/10.1016/j.jcfm.2005.09.003 PMID: 16931101

[10] Murata, C.; Kuroki, Y.; Imoto, I.; Kuroiwa, A. Ancestral Y-linked genes were maintained by translocation to the X and Y chromosomes fused to an autosomal pair in the Okinawa spiny rat Tokudaia muenninki. *Chromosome Res.,* **2016**, *24*(3), 407-419. http://dx.doi.org/10.1007/s10577-016-9531-y PMID: 27333765

[11] Burgoyne, P. S.; Arnold, A. P. A primer on the use of mouse models for identifying direct sex chromosome effects that cause sex differences in non-gonadal tissues. *Biol. Sex Differ,* **2016**, *7*, 68. http://dx.doi.org/10.1186/s13293-016-0115-5

[12] Skaletsky, H.; Kuroda-Kawaguchi, T.; Minx, P.J.; Cordum, H.S.; Hillier, L.; Brown, L.G.; Repping, S.; Pyntikova, T.; Ali, J.; Bieri, T.; Chinwalla, A.; Delehaunty, A.; Delehaunty, K.; Du, H.; Fewell, G.; Fulton, L.; Fulton, R.; Graves, T.; Hou, S.F.; Latrielle, P.; Leonard, S.; Mardis, E.; Maupin, R.; McPherson, J.; Miner, T.; Nash, W.; Nguyen, C.; Ozersky, P.; Pepin, K.; Rock, S.; Rohlfing, T.; Scott, K.; Schultz, B.; Strong, C.; Tin-Wollam, A.; Yang, S.P.; Waterston, R.H.; Wilson, R.K.; Rozen, S.; Page, D.C. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature,* **2003**, *423*(6942), 825-837. http://dx.doi.org/10.1038/nature01722 PMID: 12815422

[13] Kido, T.; Lau, Y.F. Roles of the Y chromosome genes in human cancers. *Asian J. Androl.,* **2015**, *17*(3), 373-380. PMID: 25814157

[14] Lau, Y.C. Y chromosome in health and diseases. *Cell Biosci.,* **2020**, *10*, 97. http://dx.doi.org/10.1186/s13578-020-00452-w

[15] O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C.M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V.S.; Kodali, V.K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K.M.; Murphy, M.R.; O'Neill, K.; Pujar, S.; Rangwala, S.H.; Rausch, D.; Riddick, L.D.; Schoch, C.; Shkeda, A.; Storz, S.S.; Sun, H.; Thibaud-Nissen, F.; Tolstoy, I.; Tully, R.E.; Vatsan, A.R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M.J.; Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T.D.; Pruitt, K.D. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.,* **2016**, *44*(D1), D733-D745. http://dx.doi.org/10.1093/nar/gkv1189 PMID: 26553804

[16] Jobling, M.A.; Tyler-Smith, C. Human Y-chromosome variation in the genome-sequencing era. *Nat. Rev. Genet.,* **2017**, *18*(8), 485-497. http://dx.doi.org/10.1038/nrg.2017.36 PMID: 28555659

[17] Jangravi, Z.; Alikhani, M.; Arefnezhad, B.; Sharifi Tabar, M.; Taleahmad, S.; Karamzadeh, R.; Jadaliha, M.; Mousavi, S.A.; Ahmadi Rastegar, D.; Parsamatin, P.; Vakilian, H.; Mirshahvaladi, S.; Sabbaghian, M.; Mohseni Meybodi, A.; Mirzaei, M.; Shahhoseini, M.; Ebrahimi, M.; Piryaei, A.; Moosavi-Movahedi, A.A.; Haynes, P.A.; Goodchild, A.K.; Nasr-Esfahani, M.H.; Jabbari, E.; Baharvand, H.; Sedighi Gilani, M.A.; Gourabi, H.; Salekdeh, G.H. A fresh look at the male-specific region of the human Y chromosome. *J. Proteome Res.,* **2013**, *12*(1), 6-22. http://dx.doi.org/10.1021/pr300864k PMID: 23253012

[18] Heydari, R.; Jangravi, Z.; Maleknia, S.; Seresht-Ahmadi, M.; Bahari, Z.; Salekdeh, G.H.; Meyfour, A. Y chromosome is moving out of sex determination shadow. *Cell Biosci.,* **2022**, *12*(1), 4. http://dx.doi.org/10.1186/s13578-021-00741-y PMID: 34983649

[19] Hofstetter, J.R.; Zhang, A.; Mayeda, A.R.; Guscar, T.; Nurnberger, J.I., Jr; Lahiri, D.K. Genomic DNA from mice: A comparison of recovery methods and tissue sources. *Biochem. Mol. Med.,* **1997**, *62*(2), 197-202. http://dx.doi.org/10.1006/bmme.1997.2637 PMID: 9441873

[20] Lahiri, D.K.; Nurnberger, J.I., Jr A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res.,* **1991**, *19*(19), 5444. http://dx.doi.org/10.1093/nar/19.19.5444 PMID: 1681511

[21] Lahiri, D.K.; Bye, S.; Nurnberger, J.I., Jr; Hodes, M.E.; Crisp, M. A non-organic and non-enzymatic extraction method gives higher yields of genomic DNA from whole-blood samples than do nine other methods tested. *J. Biochem. Biophys. Methods,* **1992**, *25*(4), 193-205. http://dx.doi.org/10.1016/0165-022X(92)90014-2 PMID: 1494032

[22] Lahiri, D.K.; Schnabel, B. DNA isolation by a rapid method from human blood samples: Effects of MgCl2, EDTA, storage time, and temperature on DNA yield and quality. *Biochem. Genet.,* **1993**, *31*(7-8), 321-328. http://dx.doi.org/10.1007/BF00553174 PMID: 8274138

[23] Fallon, H.J.; Frei, E., III; Davidson, J.D.; Trier, J.S.; Burk, D. Leukocyte preparations from human blood: Evaluation of their morphologic and metabolic state. *J. Lab. Clin. Med.,* **1962**, *59*, 779-791. PMID: 13891473

[24] Thomson, A.E.; Bull, J.M.; Robinson, M.A. A procedure for separating viable lymphocytes from human blood and some studies on their susceptibility to hypotonic shocks. *Br. J. Haematol.,* **1966**, *12*(4), 433-446. http://dx.doi.org/10.1111/j.1365-2141.1966.tb05652.x PMID: 5944490

[25] Dain, A.R.; Hall, J.G. A method for the isolation of white cells from the blood of sheep by differential lysis with hypotonic saline solutions. *Vox Sang.,* **1967**, *13*(3), 281-284. http://dx.doi.org/10.1111/j.1423-0410.1967.tb03397.x PMID: 6033403

[26] Gao, X.; Zhang, KY.; Lu, TZ.; Zhao, Y.; Zhou, HY.; Yu, YQ.; Zellmer, L.; He, Y.; Huang, H.; Liao, D. A reassessment of several erstwhile methods for isolating DNA fragments from agarose gels. *3 Biotech,* **2021**, *138*, 2691. http://dx.doi.org/10.1007/s13205-021-02691-1

[27] Sun, Y.; Sriramajayam, K.; Luo, D.; Liao, D.J. A quick, cost-free method of purification of DNA fragments from agarose gel. *J. Cancer,* **2012**, *3*, 93-95. http://dx.doi.org/10.7150/jca.4163 PMID: 22359530

[28] Rogers, M.J. Y chromosome copy number variation and its effects on fertility and other health factors: A review. *Transl. Androl. Urol.,* **2021**, *10*(3), 1373-1382. http://dx.doi.org/10.21037/tau.2020.04.06 PMID: 33850773

[29] Zhou, W.; Machiela, M.J.; Freedman, N.D.; Rothman, N.; Malats, N.; Dagnall, C.; Caporaso, N.; Teras, L.T.; Gaudet, M.M.; Gapstur, S.M.; Stevens, V.L.; Jacobs, K.B.; Sampson, J.; Albanes, D.; Weinstein, S.; Virtamo, J.; Berndt, S.; Hoover, R.N.; Black, A.; Silverman, D.; Figueroa, J.; Garcia-Closas, M.; Real, F.X.; Earl, J.; Marenne, G.; Rodriguez-Santiago, B.; Karagas, M.; Johnson, A.; Schwenn, M.; Wu, X.; Gu, J.; Ye, Y.; Hutchinson, A.; Tucker, M.; Perez-Jurado, L.A.; Dean, M.; Yeager, M.; Chanock, S.J. Mosaic loss of chromosome Y is associated with common variation near TCL1A. *Nat. Genet.,* **2016**, *48*(5), 563-568. http://dx.doi.org/10.1038/ng.3545 PMID: 27064253

[30] Brinkmann, B. Is the amelogenin sex test valid? *Int. J. Legal Med.,* **2002**, *63*, 263. http://dx.doi.org/10.1007/s00414-001-0263-x

[31] Peng, Z.; Yuan, C.; Zellmer, L.; Liu, S.; Xu, N.; Liao, D.J. Hypothesis: Artifacts, including spurious chimeric RNAs with a short homologous sequence, caused by consecutive reverse transcriptions and endogenous random primers. *J. Cancer,* **2015**, *6*(6), 555-567. http://dx.doi.org/10.7150/jca.11997 PMID: 26000048

[32] Xie, B.; Yang, W.; Ouyang, Y.; Chen, L.; Jiang, H.; Liao, Y.; Liao, D. J. Two Rnas or dnas may artificially fuse together at a Short Homologous Sequence (SHS) during reverse transcription or polymerase chain reactions, and thus reporting an shs-containing chimeric RNA requires extra caution. *PLoS. One,* **2016**, *11*(5), e0154855. http://dx.doi.org/10.1371/journal.pone.0154855

[33] Yuan, C.; Han, Y.; Zellmer, L.; Yang, W.; Guan, Z.; Yu, W.; Huang, H.; Liao, D.J. It is imperative to establish a pellucid definition of chimeric RNA and to clear up a lot of confusion in the relevant research. *Int. J. Mol. Sci.,* **2017**, *18*(4), E714. http://dx.doi.org/10.3390/ijms18040714 PMID: 28350330

[34] Čerenak, A.; Kolenc, Z.; Sehur, P.; Whittock, S. P.; Koutoulis, A.; Beatson, R.; Buck, E.; Javornik, B. New male specific markers for hop and application in breeding program. *Sci. Rep,* **2019**, *9*(1), 14223. http://dx.doi.org/10.1038/s41598-019-50400-z

[35] Kroczak, A.; Wołoszyńska, M.; Wierzbicki, H.; Kurkowski, M.; Grabowski, K. A.; Piasecki, T.; Galosi, L.; Urantówka, A. D. New bird sexing strategy developed in the order psittaciformes involves multiple markers to avoid sex misidentification: Debunked myth of the universal DNA marker. *Genes (Basel),* **2021**, *12*(6), 878. http://dx.doi.org/10.3390/genes12060878

[36] Adolfi, M. C.; Du, K.; Kneitz, S.; Cabau, C.; Zahm, M.; Klopp, C.; Feron, R.; Paixão, R. V.; Varela, E. S.; de Almeida, F. L.; de Oliveira, M. A.; Nóbrega, R. H.; Lopez-Roques, C.; Iampietro, C.; Lluch, J.; Kloas, W.; Wuertz, S.; Schaefer, F.; Stöck, M.; Guiguen, Y.; Schartl, M. A duplicated copy of id2b is an unusual sex-determining candidate gene on the Y chromosome of arapaima (Arapaima gigas). *Sci. Rep,* **2021**, *11*(1), 21544. http://dx.doi.org/10.1038/s41598-021-01066-z

[37] Zhong, S.; Zhang, K.; Bagheri, M.; Burken, J.G.; Gu, A.; Li, B.; Ma, X.; Marrone, B.L.; Ren, Z.J.; Schrier, J.; Shi, W.; Tan, H.; Wang, T.; Wang, X.; Wong, B.M.; Xiao, X.; Yu, X.; Zhu, J.J.; Zhang, H. Machine Learning: New ideas and tools in environmental science and engineering. *Environ. Sci. Technol.,* **2021**, *55*(19), 12741-12754.
http://dx.doi.org/10.1021/acs.est.1c01339 PMID: 34403250

[38] Raza, A.; Bardhan, S.; Xu, L.H.; Yamijala, S.S.R.K.C.; Lian, C.; Kwon, H.; Wong, B.M. A machine learning approach for predicting Defluorination of Per- and Polyfluoroalkyl Substances (PFAS) for their efficient treatment and removal. *Environ. Sci. Technol. Lett.,* **2019**, *6*(10), 624-629.
http://dx.doi.org/10.1021/acs.estlett.9b00476

[39] Wilson Sayres, M.A. Genetic diversity on the sex chromosomes. *Genome Biol. Evol.,* **2018**, *10*(4), 1064-1078.
http://dx.doi.org/10.1093/gbe/evy039 PMID: 29635328

[40] Furman, B.L.S.; Metzger, D.C.H.; Darolti, I.; Wright, A.E.; Sandkam, B.A.; Almeida, P.; Shu, J.J.; Mank, J.E. Sex chromosome evolution: So many exceptions to the rules. *Genome Biol. Evol.,* **2020**, *12*(6), 750-763.
http://dx.doi.org/10.1093/gbe/evaa081 PMID: 32315410