

Applied Research of Distinguishing Similar Y-STR Haplotypes by Y Chromosome Complete Sequencing

Xingyi Yang

School of Forensic Medicine, Southern Medical University

Zhonghao Yu

School of Forensic Medicine, Southern Medical University

Shuanglin Li

School of Forensic Medicine, Southern Medical University

Manling Huang

School of Forensic Medicine, Southern Medical University

Song Cai

Department of Anatomy, Histology and Development Biology, School of Basic Medical Sciences, Shenzhen University Health Science Centre

Ling Chen

School of Forensic Medicine, Southern Medical University

Changhui Liu

Guangzhou Forensic Science Institute, Guangdong Province Key Laboratory of Forensic Genetics

Huijun Wang

School of Forensic Medicine, Southern Medical University

Chao Liu (✉ liuchaogzf@163.com)

Guangzhou Forensic Science Institute <https://orcid.org/0000-0001-5633-3929>

Primary research

Keywords: Y-STR haplotype, Y-SNP haplogroup, Y chromosome, Liquid phase probe capture next generation sequencing technology, Coancestry

Posted Date: March 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-262197/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: To explore a technical method to distinguish similar Y-STR haplotypes and its value in deducing the differentiation the males in paternal line, we used a complete genome sequence of Y chromosome using streptavidin–biotin magnetic particle-based capture methodology (Y chromosome liquid phase probe capture next generation sequencing technique (NGS)) to detect male individuals with similar Y-STR haplotypes. Based on our independently developed mathematical model and the new topological structure of Y chromosome mutation sites as well as haplogroups and pedigree trees updated by the International Society of Genealogy (ISOGG) every year, we distinguished the coancestry of male individuals with similar Y-STR haplotypes and **Results:** Identifying differences between the judgment results of Y full sequencing and the pedigree survey results allowed for the estimation whether the individuals have close relation within 3~5 generations or not. Y chromosome liquid phase probe capture NGS technique could capture the 16M region and effectively analyze tens of thousands of Y-SNP loci. Among them, the coancestry obtained by analysis of 8 sample cases was consistent with the actual total case time obtained by family investigation. **Conclusions:** Detecting the Y-STR haplotype similarity between male individuals and conducting the previously reported mathematical model analysis by using Y chromosome liquid phase probe capture NGS technology can uncover the coancestry of the different male individuals. These results provide the foundation for further investigation of the similar Y-STR haplotype males in the Y-STR database.

Background

Y chromosome is stably inherited by male offspring and transmitted from the father. Only cells belonging to the biologically male sex will express it. Y-STR typing technology has unique advantages and is an effective supplement to autosomal STR test technology. The Y-STR haplotype can be used to divide several family lines of the paternal clan, to detect the Y-STR of representative male members in each family, and to delineate the families which were screened out by suspects' Y-STR haplotype. Finally, minimizing the screening scope and even directly screening out the suspect [1]. The typical cases involving Y-STR reported in China include the *Baiyin case* in Gansu [2](Y-STR detection technology helps break "Baiyin Case")(Y-STR detection technology helps break "Baiyin Case")(Y-STR detection technology helps break "Baiyin Case") and the *410 case* in Guangzhou [3]. However, in practical application, inconsistencies are commonly found in Y-STR haplotype comparisons within 1 ~ 3 Y-STR loci. It is especially hard to consider it as an effective clue when one Y-STR haplotype matches many male individuals with multiple surnames, multiple regions or families to which the individual belongs. Even so, whether a similar Y-STR haplotype can effectively locate the same paternal line is still controversial and guarding access to this highly personal data remains in question [4].

NGS technology boasts advantages including high sequencing depth, high parallelism and high accuracy, which enable simultaneous detection of multiple samples and analysis of a large number of SNP sites in a single experiment [5]. Here, Y chromosome liquid phase probe capture NGS technology [6] was used to distinguish similar Y-STR haplotypes, and the detailed operations were shown in Fig. 1. Combining

application of multiple genetic markers on Y chromosome, we have harnessed a mathematical model to analyze the coancestry of male individuals with similar Y-STR haplotypes. This model is dependent on regular annual updates to the International Society of Genealogy (ISOGG), which provides important contents such as the new topological structure of Y chromosome mutation sites, haplogroups and pedigree trees. The sequencing length of the present invention is close to the limit of the regions available on the Y chromosome compared with past research, and can test for more mutations, so that the differentiation time between different branches can be calculated more accurately. Roughly, mutations can be detected every two to three generations. This technique is used in sporadic cases [7], or to solve cases for the purpose of verifying the accuracy of this technique in inferring paternal ancestry. Through the application cases of our unit and Guangdong Provincial Public Security Department, the value of this technology in practical application is evaluated to provide reference for future use of this technology.

Results

In sample case 1, 53 of the 60 Y-STR loci were identical in typing, while the other 7 Y-STR loci were deleted. Sequencing results showed that there were large deletions in the range of 18.45 M ~ 26.41 M in the target sample SX19G0022. In the judgment of lineage relationship, the loci with numerical deletions, except DYF387S1, were all located in Yq11.222-Yq11.223 bands, which were indeed adjacent in position. According to the Y-captured sequencing data, the sample has a high probability of deletion in the Y chromosome 1845562-26426189 region, while the comparative human sample SX19G0023 has read values in these regions. The two belonged to the same subdivided Y haplogroup. It is inferred that the coancestry was within 50 years. The results showed that there was a father-son relationship between the individuals of the physical evidence on site and the personnel in the warehousing ratio.

In sample case 2, 60 of the 60 Y-STR loci had the same typing among V300041222B, V300062806B, SX19G006208, SX19G006207 and SX19G006206 and there were 1-2 private mutations of Y-SNP loci between case samples(V300041222B) and personnel samples(V300062806B), which belonged to the same subdivided Y haplogroup, and 177 ± 35 years might be the upper limit of differentiation. There were 5-7 private mutations of Y-SNP loci between case samples(V300041222B) and personnel samples (SX19G006208, SX19G006207 and SX19G006206), and 500-1000 years might be the upper limit of differentiation (Fig.S1). The final results proved that the man in the warehouse was the suspect's uncle, with a difference of 4-5 generations.

In sample case 3, 60 of the 60 Y-STR loci have the same typing. Physical evidence and personnel samples belonged to a very rare branch. In one example, two male individuals were subdivided into nearly identical haplogroups, with an inferred that between 50 and 100 years was probably the most period to form this distinction. The final results indicated that one individual was the other's nephew.

Among the 60 Y-STR loci in sample case 4, 59 loci possessed the same typing and belonged to the same subdivided Y haplogroup, with a coancestry within 50 years. The final results proved that the person whose information was in database was the suspect's brother.

In sample case 5, 57 of the 60 Y-STR loci were identical in typing, belonging to the same subdivided rare Y haplogroup O2a1b2-FGC3750 and having biological and recent paternal kinship. According to the number of mutations shared by the two samples and the number of mutations unique to each other, it was inferred that the coancestry was about 600 years. The final results proved that one branch of the people in the warehouse was distributed in Sichuan Province, China while the male families of the suspects were distributed in a village in Hunan Province, China, where was far apart.

In sample case 6, 44 of 45 Y-STR loci had the same typing, both belonged to the same subdivided Y haplogroup, with 4-5 differentiated SNP loci, and it was inferred that the coancestry was over 300 years. The final results proved that the relationship between the two male individuals could not be determined through genealogy and did not belong to a small family (within 5 generations).

Sample case 7 had 32 of 36 Y-STR loci with the same typing (GZS-A and GZS-B in Table S1), which did not belong to the same subdivided Y haplogroup and had no paternal kinship in biological sense. The genetic relationship could be completely excluded, and the number of differentiated SNP loci was 8-9. It was inferred that the coancestry was more than 1000 years. There was no biological paternal kinship, and the kinship between them could be completely excluded. The final results proved that the two male individuals came from far apart areas and did not belong to a small family.

Among the 60 Y-STR loci in sample case 8, 43 loci were consistent in typing, and sequencing results showed that the sample SX19G0019 belonged to O1a1a1a1a1a11-F492-A12442, while sample SX19G0020 belonged to O1a1a2-F4084. There was no paternal homology between the two, and it was inferred that the coancestry was 3200 years, which was consistent with the system query results. And the corresponding comparison of Y-STR and Y full sequencing results of above sample cases was shown in Table 1.

Discussion

Y-chromosome DNA analysis can be performed with either Y-STRs or Y-SNPs. The Y-STR genetic marker is the main genetic marker in our Y chromosome database. Y-STRs with mutation rate approximately 10^{-3} nt per generation which are greater used in forensic applications while Y-SNPs with mutation rate approximately 10^{-9} nt per generation which are greater used in ancestry studies [8, 9].

With an increasing Y-STR database capacity, Y haplotype diversity of 27 Y-STRs (AmpFISTR® Yfiler™ plus) are unappeasable to locate a criminal suspect' family, because we would find so many male individuals own the same Y-STR haplotype. Therefore, the number of Y-STRs input Y database in Guangzhou, China has gradually increased from 27 with AmpFISTR® Yfiler™ plus kit to Y37 and Y plus 41. At present, the number of warehousing Y-STRs has increased to 60 with AGCU Y Supp plus kit. When there are not enough loci in the warehousing ratio, the strategy of adding Y-STR loci is adopted.

However, it tends to consider that males with 3 Y-STR loci different in typing are not from the same paternal lineage when 27 Y-STRs but 60 Y-STRs was used. Especially if the step size is different by more

than 3, usually the analysis of mutation rate will be used as irrelevant individual treatment. Male families that still match completely after adding Y-STR loci require the police officer from investigation department to conduct in-depth investigation to investigate the relevant families. However, when it is found that multiple Y-STR are the same and only 1~3 loci are different, especially in the case of one-step mutation, 2 or less mismatched loci or 2 or less step length, it will be difficult to determine whether they come from the same paternal line, which will cause great troubles to forensic practice. It is more likely that this comes from the same pedigree instead of unrelated pedigree [10]. The mechanism that stepwise mutation model of STR contributed to the problem of Y-STRs to trace paternal lineage [11]. With the widespread use of rapidly mutated Y-STR loci, further relationships between the same paternal line increase the possibility of mutation. When there are several loci mismatches in the two Y-STR haplotypes, we can neither deny the paternal line because the typing results of several Y-STR are different [12] when judging whether they come from the same paternal line. It is reported that the cumulative mutation step size of Y-STR haplotypes in two different male individuals is 5 steps, and both share the same Y-SNP haplotype group [13].

Based on AmpFISTR®Yfiler™ plus and precise identification PrecisionID system (Ion Chef and Ion S5XL), 165 Y-SNP haplotype groups were constructed and detected [14], and likelihood was calculated by family search index (FSindex) to try to solve this problem. This study showed that the sample case 5 and case 6 had the same Y-STR typing 57/60 and 44/45 respectively. The Y full sequencing results suggested that the coancestry in sample case 5 and case 6 was more than 300-500 years, and only about 9 generations or more may be a family, suggesting that priority should not be given to investigating the male family, thus saving manpower. The Y full sequencing results of sample case 2-4 were consistent with those suggested by Y-STR results. The 60 Y-STRs were identical or have one-step tolerance, basically being set under the same subdivided haplogroup, and the differentiated SNP sites were between 0-2. However, it can be seen in sample case 6 and 7 that even if multiple Y-STRs were the same, the coancestry of the two males exceeded 300 years under the number of Y-STR mutation sites allowed by the same paternal line and the same subdivided haplogroup. It is worth noting that in sample case 1, microdeletions on Y chromosome lead to a large amount of Y-SNP loss in this region, but this was a unique mutation occurring in individuals, which did not affect the judgment of paternal kinship for men in this case.

In this study, the full length of Y chromosome (16M region) was captured by liquid probe, allowing millions of Y-SNP to be detected. Based on population genetics method and genetic statistics calculation of large-scale population, the difference between the target personnel and the personnel in the warehousing ratio was calculated by Y-SNP difference at the cut-off point of Y chromosome evolutionary tree. Generally, 25 years is defined as a generation, which depends on the evolution speed of different haplotypes. If the calculation result is within 100 years, it means that the male to whom the DNA on the physical evidence belongs has a close relationship with the male in the database ratio, with a difference of about 3-4 generations (Fig.S1). If the calculation result is 100-300 years, it means that the male to which the DNA belongs on the physical evidence has a cousin or more distant relationship with other

men, within 4-10 generations. If the calculation result is beyond 1000 years, it can be directly excluded and family investigation is not needed.

In statistics, the larger the population, the more the estimation of statistical parameters fits the real value. The smaller the sample amount, the greater the fluctuation and uncertainty of the estimated value is. Usually, there will be a confidence interval. For a long period, this interval is relatively small, and the inferred conclusion will tend to be more accurate. At the same time, the Y chromosome liquid phase probe capture next generation sequencing technology also has two conditions affecting the accuracy of distance age judgment. Due to the poor quality of the on-site physical evidence template and the poor quality of the on-site physical evidence template, the first case results in the low coverage of the target sample on the Y chromosome and the poor quality of Y capture and sequencing, resulting in the failure to detect the family private locus, and the given estimation is close to the age range. The next case produces false positive SNP loci, resulting in families with closer coancestry being judged to be far away. The first case occurred in the physical evidence examination of sample case 5, and the quality of the database was subsequently improved to obtain a narrower chronological range.

Conclusions

In summary, Y-chromosome testing are not as meaningful as autosomal STR matches from a random match probability point-of view. However, Y-chromosome DNA testing results can aid forensic investigations, a match between the who contributed to the forensic stain and individual in question only means they may from the same paternal lineage. This Y chromosome liquid phase probe capture NGS technology can help to infer the generations apart and obviously reduce the scope of polices' investigation. Our mathematical model can compensate and overcome the shortage of full sequencing target samples on Y chromosome and give relatively accurate upper limitation of paternal line differentiation. Preliminary results suggest that Y chromosome liquid phase probe capture NGS technology can be expected to predict the geographical origin of male lineage, distinguish the inbreeding groups in Y-STR ratio, and have good application prospects.

Materials And Methods

Sample preparation

Seven pieces of male Y-STR profiles were placed into the China National Y-STR Database and found the individual male with 0-3 mismatch loci (tolerance no more than 3 loci). These similar Y-STR haplotypes were respectively compiled into sample cases 1-7; Men with tolerance equal to 7 were numbered as sample case 8.(the detail of the case in Table 1). After obtaining the informed consent of all members of all above eight cases and the approval of the biomedical ethical committee of Southern Medical University, we collected oral swabs of the subjects with disposable sterile cotton swabs and immediately stored the samples in a -80°C refrigerator for the purpose of DNA extraction and sequencing uniformly.

DNA extraction

The DNA was extracted using the Eltbio Mag Saliva & Swab DNA Extraction Kit (Enlighten Biotech, Shanghai, China) according to the manufacturer's protocol.

DNA library preparation

DNA libraries were prepared using the DeepReads DNA Library Prep Kit for Fiseq (DeepReads Genomics Technology, Guangzhou, China) according to the manufacturer's protocol.

Y-chromosome bait design and Y-chromosome capture

Y-chromosome bait was designed by DeepReads Genomics Technology. Capture reactions were performed according to DeepReads WHO-Y Kit protocol (DeepReads Genomics Technology, Guangzhou, China). A series of bait libraries were designed to capture the sequences of a region of ~ 16 M bp on the Y chromosome on the basis of previous ~ 11 M [15].

Fiseq sequencing

The libraries were sequenced on an Fiseq-2000 platform (DeepReads Genomics Technology, Guangzhou, China) and generated 2×100 -base pair (bp) paired-end reads.

A proper amount of DNA was collected for the construction of the whole genome library, and the whole genome library was used as a template for the construction of liquid phase probe capture library using liquid phase probe capture library construction kits of Deepreads Biotech Co., Ltd. Compared with the imbrined probe design, we adopt the staggered design. For some areas with high GC, high repetition and palindromic structure, we focus on adding some probes to improve the capture efficiency (Schematic design of primer is shown in Fig.2). In liquid phase hybridization, the probe existed in the liquid phase and carried biotin labels. When the probe hybridized with the target region, the probe was absorbed by streptavidin modified magnetic beads. Target region fragments were captured, and the uncaptured fragments were discarded. The probe and the target region were separated by re-denaturation, and all excess probe materials were discarded. Magnetic beads were ultimately used to extract the DNA and thereby obtain the target region library.

Using AMPure XP Beads to screen the fragment length of the target region library, the fragment length was concentrated in the range of 300-400bp, and the target fragment capture was completed [7]. DNA concentration of the library was measured by quality test through Qubit Fluorometer 4.0, with quality control for only concentrations greater than 1.0 ng/ μ L. The DNA length distribution of the library was detected via Agilent 2100. Qualifying samples possessed the following traits: concentrated at about 400bp, single peak, containing no obvious linker peaks and a large fragment peak. Subsequently, the

library was cyclized to prepare single-stranded circular DNAs. These were amplified by 2-3 orders of magnitude according to the principle of rolling loop amplification to prepare DNB nanospheres that can be sequenced on the computer.

Finally, the Fiseq sequencing setting of Beijing Genomics Institute (BGI) United Deepreads Biotech Co., Ltd Platform was used for high-throughput parallel sequencing. For specific operations, please refer to the operating instructions.

Data analysis

We followed the DeepReads Genomics Technology standard procedure to analyze the next-generation sequencing data [16]. The sequences of low quality, and very short reads are discarded to filter out areas of poor quality that affect data quality and subsequent analysis. Reference of genome alignment and correction are required to detect variations. Clean Data should be compared to the reference hg38 genome, however if it fails to completely match the reference genome, such as a mismatch or gap, it can also serve as a basis for subsequent variation detection.

The splitting time of the phylogenetic tree were calculated with Bayesian Evolutionary Analysis Sampling Trees (BEAST (v2.4.3)) software as our previous study [17]. The number of SNP sites differentiated between male individuals was analyzed with reference to the new topological structure of Y chromosome, haplogroup and pedigree tree updated by the ISOGG every year [18]. According to the mutation rate of different Y-SNP sites in the population, the age of distance from pedigree tree nodes was inferred.

Coalescence dating

Recent descent clusters (DCs) are characterised by a high frequency Y-microsatellite haplotype and a set of close mutational neighbors, which means the signals of continued transmission of success over generations. The time to the most recent common ancestor (TMRCA) of each DC was determined by using the average squared distance (ASD) estimator as described previously [19]. A generation time of 25 years was used to produce time estimates in years [20].

Abbreviations

NGS: Next Generation Sequencing Technique; ISOGG: International Society of Genealogy; GZS: Guangzhou Samples; BGI: Beijing Genomics Institute; BEAST: Bayesian Evolutionary Analysis Sampling Trees; DCs: Descent Clusters; TMRCA: The Most Recent Common Ancestor; ASD: Average Squared Distance

Declarations

Ethics approval and consent to participate

All the experiments in this study were performed in accordance with relevant guidelines and regulations of the biomedical ethical committee of the Southern Medical University (NO.2020-001). All experiments were conducted after obtaining the informed consent of the participants.

Consent for publication

I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and agree with the publication of this article.

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the “National Natural Science Foundation of China NSFC to YYL” (No.81801046) and “Science and Technology Program of Guangzhou, China” (No.2019030016).

Authors’ contributions

Xingyi Yang, Zhonghao Yu, Shuanglin Li and Manling Huang analyzed the data and wrote the manuscript. Song Cai and Ling Chen conceived the experimental methodology and performed the experiments. Changhui Liu, Huijun Wang and Chao Liu conceived and designed the experiment and methodology. All authors participated in the analysis and interpretation of the results. All authors read and approved the final manuscript.

Acknowledgements

The authors sincerely thank all the volunteers for providing samples, thank Lindsey Chew for assistance in manuscript revision, thank Guangdong Provincial Public Security Department for providing these cases and thank for the technical support provided by Bo Lei and Xuding Xu of Deepreads Biotech Co., Ltd.

Author details

¹School of Forensic Medicine, Southern Medical University, 510515, Guangzhou, China.

²Guangzhou Forensic Science Institute, Guangdong Province Key Laboratory of Forensic Genetics, 510030, Guangzhou, China.

References

1. Liu C, Han X, Min Y, Liu H, Xu Q, Yang X, Huang S, Chen Z, Liu C: **Genetic polymorphism analysis of 40 Y-chromosomal STR loci in seven populations from South China.** *Forensic science international* 2018, **291**:109-114.
2. **Y-STR detection technology helps break "Baiyin Case".** *Invention and Innovation (Big Science and Technology)*, **2016(10)**:53.
3. Yang X, Xu Q, Liu Q, Liu C, Hu H, Han X, Chen L, Liu C: **Combined application of multiple autosomal and Y-chromosomal STR loci in solving a homicide case in 2009.** *International journal of legal medicine* 2020.
4. Wei W, Ayub Q, Chen Y, McCarthy S, Hou Y, Carbone I, Xue Y, Tyler-Smith C: **A calibrated human Y-chromosomal phylogeny based on resequencing.** *Genome research* 2013, **23(2)**:388-395.
5. Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, Snyder M, Quintana-Murci L, Kidd JM, Underhill PA *et al*: **Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females.** *Science* 2013, **341(6145)**:562-565.
6. Alvarez-Cubero MJ, Santiago O, Martínez-Labarga C, Martínez-García B, Marrero-Díaz R, Rubio-Roldan A, Pérez-Gutiérrez AM, Carmona-Saez P, Lorente JA, Martínez-Gonzalez LJ: **Methodology for Y Chromosome Capture: A complete genome sequence of Y chromosome using flow cytometry, laser microdissection and magnetic streptavidin-beads.** *Scientific reports* 2018, **8(1)**:9436.
7. Yang X, Liu H, Liu C, Xu Q, Yang D, Han X, Chen L, Lei B, Liu C, Du W: **Application of Y-chromosomal microdeletions in a homicide case.** *Forensic science international* 2020, **314**:110370.
8. Balanovsky O: **Toward a consensus on SNP and STR mutation rates on the human Y-chromosome.** *Human genetics* 2017, **136(5)**:575-590.
9. Butler JM: **Chapter 13 - Y-Chromosome DNA Testing.** In: *Advanced Topics in Forensic DNA Typing: Methodology.* Edited by Butler JM. San Diego: Academic Press; 2012: 371-403.
10. Liu H, Li X, Mulero J, Carbonaro A, Short M, Ge J: **A convenient guideline to determine if two Y-STR profiles are from the same lineage.** *Electrophoresis* 2016, **37(12)**:1659-1668.
11. Fan H, Chu JY: **A brief review of short tandem repeat mutation.** *Genomics, proteomics & bioinformatics* 2007, **5(1)**:7-14.
12. Li L, Liu Y, Lin Y, *al e*: **To explore the limitation of Y-STR marker in paternal relationship identification from actual cases.** *China Judicial Authentication* 2015, **2015(03)**:112-113.
13. Qian E: **Construction of Y-SNP composite detection system and its application in ethnic region inference and family survey[D].** Guizhou Medical University; 2019.
14. Wang M, Wang Z, He G, Liu J, Wang S, Qian X, Lang M, Li J, Xie M, Li C *et al*: **Developmental validation of a custom panel including 165 Y-SNPs for Chinese Y-chromosomal haplogroups**

- dissection using the ion S5 XL system.** *Forensic science international Genetics* 2019, **38**:70-76.
15. Huang YZ, Wei LH, Yan S, Wen SQ, Wang CC, Yang YJ, Wang LX, Lu Y, Zhang C, Xu SH *et al*: **Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407.** *Molecular genetics and genomics : MGG* 2018, **293**(3):657-663.
 16. Yan S, Wang CC, Zheng HX, Wang W, Qin ZD, Wei LH, Wang Y, Pan XD, Fu WQ, He YG *et al*: **Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers.** *PloS one* 2014, **9**(8):e105691.
 17. Wang CC, Wang LX, Shrestha R, Wen S, Zhang M, Tong X, Jin L, Li H: **Convergence of Y Chromosome STR Haplotypes from Different SNP Haplogroups Compromises Accuracy of Haplogroup Prediction.** *Journal of genetics and genomics = Yi chuan xue bao* 2015, **42**(7):403-407.
 18. Naidoo T, Xu J, Vicente M, Malmström H, Soodyall H, Jakobsson M, Schlebusch CM: **Y-Chromosome Variation in Southern African Khoe-San Populations Based on Whole-Genome Sequences.** *Genome biology and evolution* 2020, **12**(7):1031-1039.
 19. Wang CC, Wang LX, Shrestha R, Zhang M, Huang XY, Hu K, Jin L, Li H: **Genetic structure of Qiangic populations residing in the western Sichuan corridor.** *PloS one* 2014, **9**(8):e103772.
 20. Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G *et al*: **The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time.** *American journal of human genetics* 2004, **74**(1):50-61.

Tables

Due to technical limitations, table 1 is only available as a download in the Supplemental Files section.

Figures

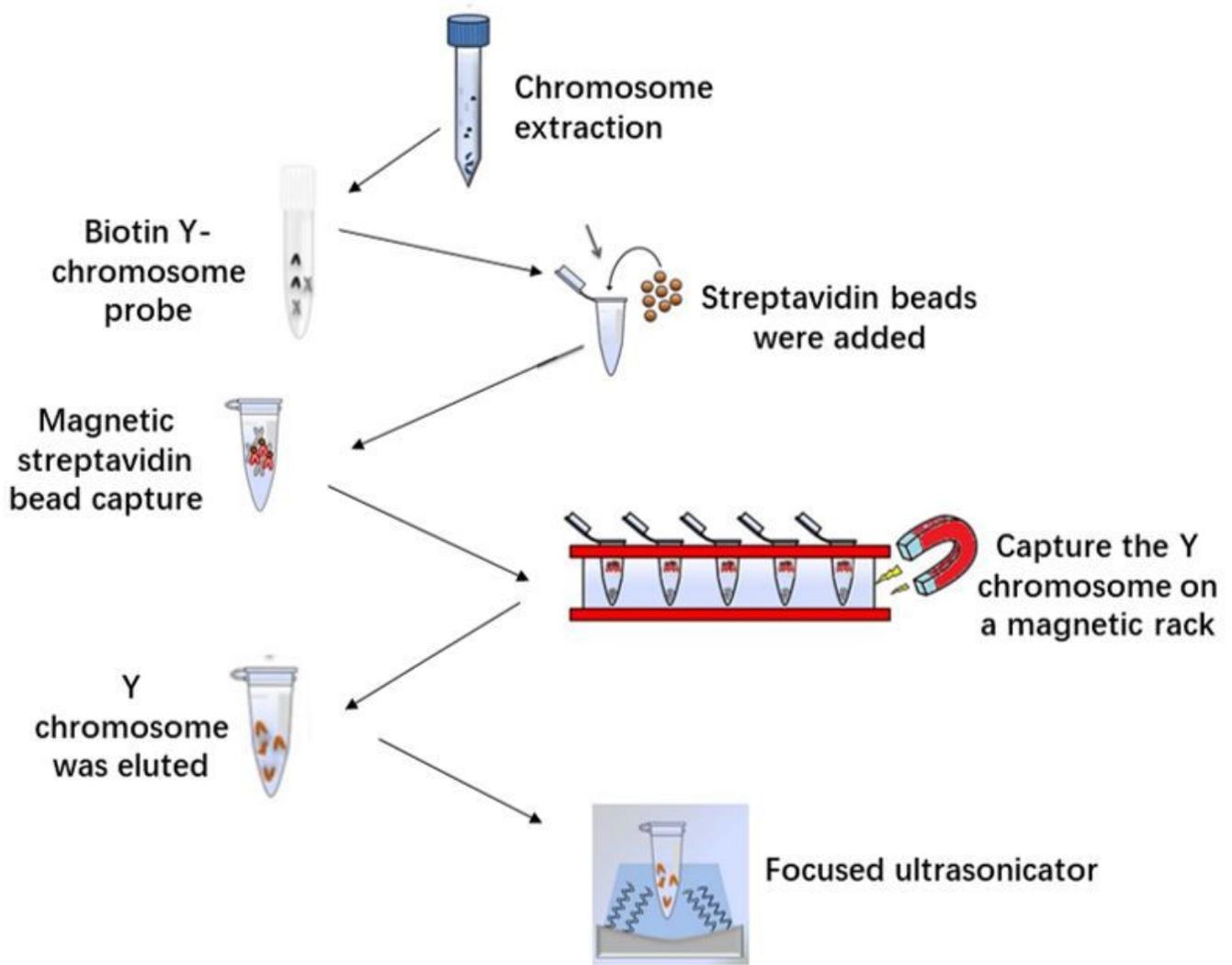
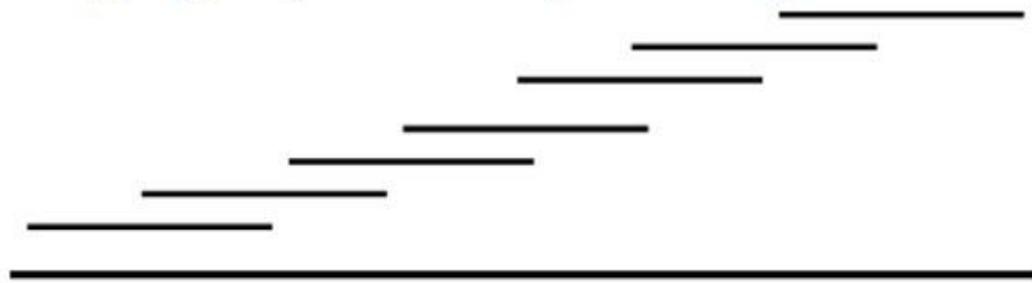


Figure 1

The Overview of the magnetic streptavidin-bead capture technology

overlapping regions of targeted amplicons



DNA sequencing

modified local dense design



DNA sequencing

Figure 2

Schematic design of primer

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S1.docx](#)
- [Table1.xlsx](#)
- [TableS1.xlsx](#)