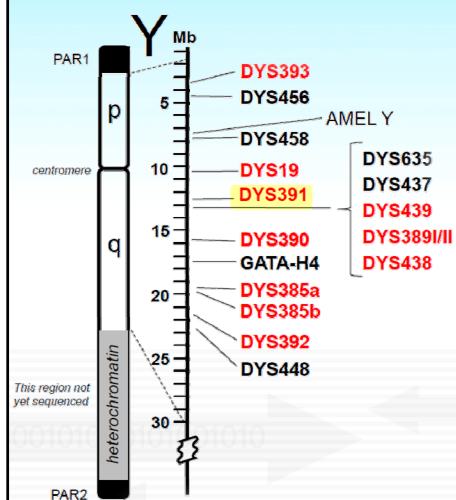


## Web-based Y-STR database for haplotype frequency estimation and kinship index calculation

In Seok Yang  
Dept. of Forensic Medicine  
Yonsei University College of Medicine

## Y chromosome short tandem repeat (Y-STR)



- The Y-STR loci are located on the NRY part of the Y chromosome and are inherited unchanged (barring mutation) as a block of linked haplotypes from generation to generation.
- An estimate of the frequency of occurrence of a particular haplotype requires the counting method which is based upon how many times a particular Y-STR haplotype is observed in a population.
- [Therefore, Y-STR database is required to estimate the frequency of haplotype.](#)

## Y-STR databases

- Current representative Y-STR databases on the web
  1. Y chromosome Haplotype Reference Database (YHRD) : 101,055 haplotypes
  2. US Y-STR Database : 18,719 haplotypes
- [Limitations of the databases](#)
  1. YHRD
    - Restricts the number of searches in a day
    - Shows some of the most frequent haplotypes in search result for the matched haplotype
  2. US Y-STR Database
    - Established with samples of only U.S. peoples → limited usage of haplotype frequency estimates from this database

## Kinship index

- Y-STR haplotype data have been used to test relationship among paternal relatives including father-son pairs.
- Kinship index (KI) is an important statistical value for explaining their relationship.
- When perfectly matching between two haplotypes, KI can be calculated from haplotype frequency.
- [In non-matching cases due to mutation](#), Rolf et al. presented calculation method of KI with average value of mutation rates of Y-STR loci.  
→ [It is limited to reflect different effect of mutation for each locus.](#)

## In this study

- Goal  
Y-STR database suitable in practice of forensic genetics
- 1. Estimation of haplotype frequency using search function in various conditions
- 2. Kinship indices calculation function for various relationship levels
- 3. User database configuration

## ySTRmanager

<http://ystrmanager.yonsei.ac.kr>

Y chromosome STR manager

E-mail:   
 Password:

About ySTRmanager
Search
Kinship

**1. Introduction**

- Y-STR haplotyping of male DNA is a powerful tool for identifying paternal lineage in both forensic and genealogical case work, and in anthropological and evolutionary studies.
- In practice, between nine and 17 Y-STRs are used for haplotyping, which normally provides enough information for reliable lineage identification.
- ySTRmanager is a database system for managing Y-STR data, which provides three functions for Y-STR haplotype search, kinship calculation, and user data depository.

**2. Three functions of ySTRmanager**

- Y-STR haplotype search function using database system
  - (1) Basically, a haplotype provided manually can be searched in open database to obtain haplotype frequency without membership login.
  - (2) Registered users are able to use totally 4 types of searches with combinations of two search targets (open or user DBs) and two types of input queries (by providing manually or selecting a data in stored user DB).
  - (3) Three types of results will be displayed at the bottom of search options.
    - Summary of search results that comprise number of matches and haplotype frequency obtained from selected population
    - List of matched haplotypes that are sorted by number of haplotypes in descending order
    - List of neighbor haplotypes including single-step mutations that are sorted by Y-STR locus-specific mutation rates in descending order
- Kinship calculation function using Y-STR locus-specific mutation rates
  - (1) Kinship index (KI) and kinship probability can be calculated from two haplotypes of male relatives (e.g., father-son pair) when a single-step mutation is observed in a locus as a difference between the haplotypes.
  - (2) To obtain more accurate KI value, Y-STR locus specific mutation rates are used instead of average value of them.
    - Previously published worldwide mutation rates are basically provided to calculate kinship index.
    - Registered users are able to store their own mutation rate data into ySTRmanager and to use the data for this calculation function.
- Depository function of user data
  - (1) ySTRmanager supports storage and management of Y-STR data and mutation data.
  - (2) Stored user's Y-STR data can be used as input queries to estimate their frequencies in selected population. Moreover, each group of user's Y-STR data can be used as a search target.
  - (3) User's mutation rate data can also be used in kinship data calculation.

**3. Contact us**

For comments, bug reports, suggestions for improvement, please contact through e-mail [kjshin@yuhs.ac](mailto:kjshin@yuhs.ac).  
 When reporting a bug or any apparent malfunction, please try to include as much information as possible about the problem.  
 In many cases, the data used help a lot too.

About us, please visit [Yonsei DNA Profiling Group](#).



YONSEI UNIVERSITY  
COLLEGE OF MEDICINE

Yonsei DNA Profiling Group  
Department of Forensic Medicine, Yonsei University College of Medicine  
50 Yonsei-ro, Seodaemun-gu, Seoul, 120-752, Korea

Metapopulation	Population	No. of samples	No. of loci
African	African American	258	17
East Asian	Korean (3)	2,253	17
	Chinese Han (7)	1,104	11, 12, or 17
	Chinese minor populations (8)	1,337	11, 12, or 17
	Japanese (2)	2,245	17
	Taiwanese Han	200	17
	Taiwanese Paiwan	208	17
	Malay (Malaysian, Singaporean)	520	12 or 17
West Eurasian	Austrian	135	17
	Danish	185	12
	German	279	11
	Hungarian	215	12
	Italian	155	17
	Polish	255	17
	Portuguese (2)	425	17
	Resident Basques	197	17
	Russian	545	17
	Serbian	185	17
	Spanish (2)	395	14 or 17
	Swiss	150	12
	UK Caucasian	250	12
	US Caucasian	260	17
Admixed	Argentine	224	12
	Brazilian	500	17
	Colombian	950	9 or 12
	Ecuadorian	120	12
	Mexican-Mestizo	357	9
	US Hispanic	139	17
	Venezuelan	173	12
<b>Total</b>		<b>14,219</b>	

Metapopulation	Population	No. of samples	No. of loci
African	African American	258	17
East Asian	Korean (3)	2,253	17
	Chinese Han (7)	1,104	11, 12, or 17
	Chinese minor populations (8)	1,337	11, 12, or 17
	Japanese (2)	2,245	17
	Taiwanese Han	200	17
	Taiwanese Paiwan	208	17
	Malay (Malaysian, Singaporean)	520	12 or 17
West Eurasian	Austrian	135	17
	Danish	185	12
	German	279	11
	Hungarian	215	12
	Italian	155	17
	Polish	255	17
	Portuguese (2)	425	17
	Resident Basques	197	17
	Russian	545	17
	Serbian	185	17
	Spanish (2)	395	14 or 17
	Swiss	150	12
	UK Caucasian	250	12
	US Caucasian	260	17
Admixed	Argentine	224	12
	Brazilian	500	17
	Colombian	950	9 or 12
	Ecuadorian	120	12
	Mexican-Mestizo	357	9
	US Hispanic	139	17
	Venezuelan	173	12
<b>Total</b>		<b>14,219</b>	

These Y-STR data were stored into open database and are used as targets for search function of ySTRmanager.

## (1) Y-STR search

### 1. Various search conditions

- Y-STR haplotype
  - ✓ Standard allele
  - ✓ Microvariant allele
- Sample information
- Y-haplogroup

### 2. Search results

- Matched haplotypes
- Neighbor haplotypes

### 3. Estimation of haplotype frequency

- Clopper & Pearson method

$$\sum_{k=0}^x \binom{n}{k} p_0^k (1-p_0)^{n-k} = 0.05 \quad (x > 0)$$

$$p_0 = 1 - 0.05^{1/n} \quad (x = 0)$$

Clopper CJ, Pearson ES. Biometrika 1934;26(4):404-13.

Buckleton JS, Krawczak M, Weir BS. Forensic Sci Int Genet 2011;5(2):78-83.

## An example of Y-STR search

### Y-STR search

#### 1. Search setting

##### ① Y-STR haplotype information

A. Direct input    B. Select from user DB

Haplotype range:

DYS19	14	DYS389I	Others	12	DYS389II	28	DYS390	25
DYS391	10	DYS392	14	DYS393	Others	12	DYS395	13,19
DYS438		DYS439		DYS437		DYS448		
DYS456		DYS458		DYS635		YGATAH4		

Sample information:

Haplogroup:      Exact match     Include the keyword

##### ② Target population

A. Open DB    B. User DB

Metapopulation	East Asian	Group name	Korean Yonsei
Haplotype information	Yfiler haplotype	No. of samples	706
Group information	Conducted by Yonsei DNA Profiling Group	Reference	Korean J Leg Med. 2007;31:162-70.

### Y-STR search using wildcard(\*)

**Y-STR search**

**1. Search setting**

① Y-STR haplotype information

A. Direct input B. Select from user DB

Haplotype range: Yfiler

DYS19	14	DYS389I	Others	12	DYS389II	28	DYS390	25
DYS391	10	DYS392	14	DYS393	Others	12	DYS385	13,19
DYS438		DYS439		DYS437			DYS448	
DYS456							YGATAH4	

Sample information: Seoul  
Haplogroup: O3a2c1a

Exact match  Include the keyword

② Target population

A. Open DB B. User DB

Metapopulation: East Asian Group name: Korean Yonsei

Haplotype information	Yfiler haplotype	No. of samples	706	Group information	Conducted by Yonsei DNA Profiling Group	Reference	Korean J Leg Med. 2007;31:162-70.
-----------------------	------------------	----------------	-----	-------------------	---	-----------	-----------------------------------

Buttons: Search, Worldwide frequency, Reset

**Annotations:**  
12 or 12.1 for exact match (pointing to '12' in the haplotype range)  
12.\* for ignoring microvariant alleles → 12, 12.1, and 12.2 in search result (pointing to '12' in the haplotype range)

### An example of search result

**A. Matched haplotypes**

(1) Statistics

No. of matches	5 / 706	Haplotype frequency estimator*	0.01484 (approximately 1 in every 67 individuals)
----------------	---------	--------------------------------	---

\*Haplotype frequency was estimated by calculating upper 95% one-sided confidence limit. [Help](#)

(2) Matched haplotypes Export

No.	Sample ID	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS43
1	KorYH-229	14	12	28	25	10	14	12	13,19	11	12	15
2	KorYH-356	14	12	28	25	10	14	12	13,19	11	12	14
3	KorYH-503	14	12	28	25	10	14	12	13,19	11	12	15
4	KorYH-588	14	12	28	25	10	14	12	13,19	11	11	15
5	KorYH-678	14	12	28	25	10	14	12	13,19	11	13	15

**B. Neighbor haplotypes**

(1) Options for neighbor haplotypes

No. of meioses: 1 Y-STR mutation rates: Worldwide 1 (NIST) [View data](#) Sorting order: No. of matches [Help](#)

(2) Neighbor haplotypes

No.	No. of matches*	Kinship Index#	Kinship probability#	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385
1	1 / 706 <a href="#">View</a>	0.25761	20.48%	14	12	28	25	10	14	12	13,19
2	6 / 706 <a href="#">View</a>	0.07374	6.87%	14	12	28	24	10	14	12	13,19

**Annotations:**  
+1 repeat gain (pointing to '28' in row 1, column 6)  
-1 repeat loss (pointing to '24' in row 2, column 6)

## (2) Kinship index (KI) calculation

- Usage of loci-specific mutation rates instead of average value
  - To provide more exact kinship index value
  - To reflect different effect of mutation for each locus
- Perfectly matched case between two haplotypes

$$KI = \frac{\prod_{l=1}^N (1 - \mu_l)^m}{f}$$

- Non-matched case between two haplotypes
  - Single-step mutation in each locus based on stepwise mutation model

$$KI \approx \frac{\prod_{l=1, l \neq k}^N (1 - \mu_l)^m \mu_k^{x \rightarrow y} (1 - \mu_k)^{m-1}}{f} \approx \frac{\prod_{l=1, l \neq k}^N (1 - \mu_l)^m \mu_k (1 - \mu_k)^{m-1}}{2f}$$

Buckleton JS, Triggs CM, Walsh SJ. Forensic DNA evidence interpretation. 1st ed. Boca Raton: CRC press; 2005. p. 388-9.

## An example of kinship index calculation

### Kinship calculation

#### 1. Calculation setting

##### ① Two Y-STR haplotypes

	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATA4
Haplotype 1	14	12	28	23	10	14	12	13,20								
Haplotype 2	14	12	27	23	10	14	12	13,20								

- Single-step mutation (expansion or contraction) is only considered in this calculation.
- The haplotype including mutation should be provided as Haplotype 2 (not Haplotype 1).
- In a Y-STR locus, two or more allele values should be separated by commas (e.g., "10,21" or "10,11,14")
- Blank (no input allele value) is excluded from calculation of kinship index.

##### ② Target population

A. Open DB    B. User DB

Metapopulation	East Asian	Group name	Korean Yonsei
Haplotype information	Yfiler haplotype	No. of samples	706
Group information	Conducted by Yonsei DNA Profiling Group	Reference	Korean J Leg Med. 2007;31:162-70.

##### ③ No. of meioses

= 1

##### ④ Y-STR mutation rates

Worldwide 1 (NIST) [View data](#)

Calculate    Reset

An example of kinship test among alleged father and two sons

Loci	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385
Mutation rates	0.0025	0.0024	0.0035	0.0025	0.0028	0.0007	0.0008	0.0021
Alleged father	14	12	28	23	10	14	12	13,20
Son 1	14	12	28	23	10	14	12	13,20
Son 2	14	12	27	23	10	14	12	13,20

	Alleged father and son 1	Alleged father and son 2
Matched count for son's haplotype in a population (M / N)	1 / 706	1 / 706
Frequency estimate for son's haplotype	0.00670	0.00670
Kinship index	146.38209	0.25707
Kinship probability (prior probability: 0.5)	98.32%	20.45%

### (3) User database configuration

- ySTRmanager supports storage and management of Y-STR data and mutation data.
- Stored user's Y-STR data can be used directly to estimate its haplotype frequency in a selected population.
- Moreover, each group of user's Y-STR data can be used as a target population.
- User's mutation data can also be used in kinship index calculation.



### An example of stored user's Y-STR data

**A. Group**

Metapopulation: East Asian

Group name	Haplotype range	Number of samples	Group information	Reference
Korean	Yfiler haplotype	706	대한민국, 연세대학교 의과대학 법의학과	

---

**B. Sample**

Metapopulation: East Asian | Group name: Korean

Haplotype range	Yfiler haplotype	No. of samples	706	Group information	대한민국, 연세대학교 의과대학 법의학과	Reference
-----------------	------------------	----------------	-----	-------------------	-----------------------	-----------

[ Current page : 1/48 ]

No.	Sample ID	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439
1	KorYH-001	14	12	28	24	10	14	12	13,20	11	11
2	KorYH-002	15	12	28	23	10	12	12	12,17	10	11
3	KorYH-003	15	13	29	24	11	13	13	12,18	10	11

Analysis of "Korean" data

**① Summary of Y-STR haplotypes**

No. of haplotypes	Frequencies	Sample info.	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439
11	0.0156	<input type="button" value="View"/>	16	14	29	23	10	13	13	10,1		
8	0.0113	<input type="button" value="View"/>	15	12	29	23	10	12	12	11,1		
3	0.0042	<input type="button" value="View"/>	15	12	28	23	11	14	13	13,1		
3	0.0042	<input type="button" value="View"/>	16	13	28	23	10	13	13	10,1		
3	0.0042	<input type="button" value="View"/>	15	14	30	22	10	13	13	10,1		
3	0.0042	<input type="button" value="View"/>	15	12	28	23	10	12	12	12,1		
2	0.0028	<input type="button" value="View"/>	14	12	28	24	10	14	12	13,2		
2	0.0028	<input type="button" value="View"/>	16	13	28	23	10	13	13	10,1		
2	0.0028	<input type="button" value="View"/>	15	12	29	23	10	12	13	12,1		
2	0.0028	<input type="button" value="View"/>	16	14	29	23	10	13	13	10,1		

**② Haplotype information**

Number of samples	706
Number of haplotypes	657
Number of unique haplotypes	627
Discriminatory capacity (%)	93.1
Haplotype diversity	0.9995
Standard error	0.0001

**③ Allele information**

Loci	Alleles	Frequency	Gene diversity	Standard error
DYS19	18	0.0014	0.7115	0.0058
	17	0.1091		
	16	0.3286		
	16,17	0.0014		
	15	0.3810		
DYS389I	14	0.1544	0.6675	0.0040
	13	0.0241		
	16	0.0014		
	15	0.0071		
	14	0.3739		

## Conclusion

1. Search function with various search options based on approximately 14,200 Y-STR haplotypes
  2. Kinship index calculation function in various level (Matched and non-matched cases)
  3. Storing and management of user's own Y-STR and mutation data
- [On the basis of the above three functions, the ySTRmanager will be a useful system to analyze and manage Y-STR data in practice of forensic genetics.](#)