

## Introduction

SEQ Mapper was designed to perform sequence mapping between reference data and generated sample reads. Three types of STR loci were used: the STR repeat region only, the STR region plus the two primer sequences, and the entire STR locus spanning the two primers and flanking DNA. SEQ Mapper identifies STR alleles from complex DNA data obtained from MPS using these three parameters. Sequences of STR alleles are used as references of search. Four levels of search in different stringency were used to detect alleles matching generated reads:

1. Allele Search: is the strictest search requires a full match on the entire reference allele sequence trimmed by user specified primer indices against the specific FASTA/FASTQ read.
2. STR & Primers Search: is the next level search requires a match on STR, user specified 5' and 3' flanking sequences individually.
3. STR Search: is the next level search only requires a match on STR.
4. Primers Search: is the lowest level search requires match on 5' and 3'. Two Primers Search reports will be generated: one excludes the matched allele and another excludes the matched STR & primers. In Primers Search reports, the total number of different bases comparing allele and FASTA/Q read will be calculated. This Difference value is called Levenshtein Distance. Lower Levenshtein Distance suggests a sample read is similar to a reference allele.

The SEQ Mapper app is accessible through below URL:

<http://forensic.mc.ntu.edu.tw:9000/SEQMapperWeb/Default.aspx>

Upon successful logon, user will have access to SEQ Mapper application below:

**SEQ MAPPER WEB APPLICATION** Welcome [ Log Out ]

Home About SEQ Mapper SEQ Mapper Change Password

### SEQ MAPPER MAIN FORM

[Help](#)

**SEQ Mapper Workbench**

Select SEQ File Format  FASTA  FASTQ

5' (Left) Flanking Sequence Begin Index and Length

3' (Right) Flanking Sequence Begin Index and Length

SEQ FASTA Input File

Reference Sequence Input Files

**SEQ Mapper Message**

Please fill in all required parameters then press Submit button to submit the request.

Available reports to download:

- [SEQ Mapper Report 2015-10-03-63579477226 - STR 36 pool Life.zip](#)
- [SEQ Mapper Report 2015-09-27-63578972030 - STR 36 pool Life.zip](#)

Two data sources required:

1. A set of locus/allele reference data files having one locus in one file in CSV format.

2. A set of sample reads in one file in FASTA/FASTQ format .

### Locus/Allele Reference Data

A locus CSV file contains its allele records and each allele record must have an allele, forward sequence and forward STR.

Allele	Forward Sequence	Forward STR
TPOX_5	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTCGACTGGCACAGAACAGGCACTT AGGGAACCCTCACTGAATGAATGAATGAATGAATGTTTGGGCAAATAAACGCTGACAAGGA CAGAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	CACTGAATGAATGAATGAAT GAATGTT
TPOX_6	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTCGACTGGCACAGAACAGGCACTT AGGGAACCCTCACTGAATGAATGAATGAATGAATGAATGTTTGGGCAAATAAACGCTGACA AGGACAGAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	CACTGAATGAATGAATGAAT GAATGAATGTT

Notice the STR is part of the sequence.

### FASTA/FASTQ Read Sample Data

A read in a FASTA/FASTQ sample file must follow a valid FASTA/FASTQ header. The sample FASTA/FASTQ data file may have an optional group header in the first line and this group header line is like a FASTA header. If this group header line is omitted by user, SEQ Mapper will use the header of the first FASTA/FASTQ read as the group header. The group header is for report purpose.

In the sample reads data below the group header is provided.

>Sample reads data file contains 2 reads

>CRVEA:00355:00699

CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTCGACTGGCACAGAACAGGCACTTAGGGAACCCTCACTGAATGAATGAATGAA  
TGAATGTTTGGGCAAATAAACGCTGACAAGGACAGAAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGCC

>CRVEA:02084:02753

CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTCGACTGGCACAGAACAGGCACTTAGGGAACCCTCACTGAATGAATGAATGAA  
TGAATGAATGTTTGGGCAAATAAACGCTGACAAGGACAGAAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC

Notice that the first read contains the forward sequence of allele TPOX\_5 and the second read contains the forward sequence of allele TPOX\_6. SEQ Mapper will find these two matches for Allele Search. In Primers & STR Search and the STR Search both matches will also be found as well, because STR and 5' and 3' are part of the allele in each reference allele.

### The Reports

At the end of the search process multiple reports will be generated and saved in CSV.

1. A Summary Report shows the number of FASTA/Q reads found in the respective Allele Search Report, STR & Primers Search Report, and STR Search Report, and the number of reads found in the Primer Search Report excluding the respective STR & Primers and Allele.

SEQ Mapper: A DNA sequence searching tool for massive parallel sequencing

James Chun-I Lee; Bill Tseng; Liang-Kai Chang; Adrian Linacre

Department of Forensic Medicine; College of Medicine; National Taiwan University

FASTA/Q Group ID: CRVEA:02079:03350

Mapping Options: 5' (Left) Begin Index/Length = 1 / 10; 3' (Right) Begin Index/Length = 1 / 10

4634 read(s) found in Allele Search Report and 440 read(s) found in Primer Search Report - exclude matched allele

5745 read(s) found in STR & Primers Search Report and 76 read(s) found in Primer Search Report - exclude matched STR & Primers

5899 read(s) found in STR Search Report and 1111 read(s) found in Flanking Variation Report

STR Search		STR & Primers Search		Allele Search	
Ameloger	4329	Ameloger	4226	Ameloger	3492
TPOX	1570	TPOX	1519	TPOX	1142
Total	5899	Total	5745	Total	4634

Total 14620 FASTA/FASTQ read(s) entered by user and 8626 read(s) did not match!

8666 reads handled with shortest read length 103 and longest read length 284

Distribution of FASTA/Q read length:

- Number of reads having length less than or equal to 100: 0
- Number of reads having length between 101 and 110: 3
- Number of reads having length between 111 and 120: 19
- Number of reads having length between 121 and 130: 1148
- Number of reads having length between 131 and 140: 181
- Number of reads having length between 141 and 150: 230
- Number of reads having length between 151 and 160: 100
- Number of reads having length between 161 and 170: 304
- Number of reads having length between 171 and 180: 2111
- Number of reads having length between 181 and 190: 141
- Number of reads having length between 191 and 200: 2173
- Number of reads having length between 201 and 210: 2170
- Number of reads having length between 211 and 220: 47
- Number of reads having length between 221 and 230: 6
- Number of reads having length between 231 and 240: 2
- Number of reads having length between 241 and 250: 22
- Number of reads having length between 251 and 260: 2
- Number of reads having length between 261 and 270: 2
- Number of reads having length between 271 and 280: 4
- Number of reads having length between 281 and 290: 1
- Number of reads having length between 291 and 300: 0
- Number of reads having length between 301 and 310: 0
- Number of reads having length between 311 and 320: 0
- Number of reads having length between 321 and 330: 0
- Number of reads having length between 331 and 340: 0
- Number of reads having length between 341 and 350: 0
- Number of reads having length between 351 and 360: 0
- Number of reads having length between 361 and 370: 0
- Number of reads having length between 371 and 380: 0
- Number of reads having length between 381 and 390: 0
- Number of reads having length between 391 and 400: 0
- Number of reads having length greater than 400: 0

15 Allele sequences handled having Amelogenin_X with shortest length 106 and TPOX_16 with longest length 209			
Allele count for each gene:			
Amelogenin contains 2 allele(s)			
TPOX contains 13 allele(s)			
5954 FASTA/Q read(s) skipped from matching! User may be able to re-run using smaller primer begin indices.			

- Allele Search Report shows the number of FASTA/Q reads found matching the entire allele of a specific reference locus. The matched allele sequence will be trimmed according to the begin index of the primers user specified and included in the report.

	A	B	C	D
1	Allele	Sequence	Is Reverse	Count
2	Amelogenin_X	ATCAGAGCTTAAACTGGGAAGCTGGTGGTAGGAACTGTAAAATCAGGACCACTTGAGAAACATCT GGGATAAAGAATCAACACACTATTCTTTACAGAGCCCAGGG	TRUE	1234
3	Amelogenin_X	CCCTGGGCTCTGTAAAGAATAGTGTGTTGATTCTTTATCCAGATGTTTCTCAAGTGGTCTGATTTT ACAGTTCCTACCACCAGCTTCCCAGTTAAGCTCTGAT	FALSE	329
4	Amelogenin_Y	ATCAGAGCTTAAACTGGGAAGCTGATGGTAGGAACTGTAAAATTGGGACCACTTGAGAAACCACT TTATTTGGGATGAAGAATCCACCCACTATTCTTTACAGAGCCCAGGG	TRUE	1164
5	Amelogenin_Y	CCCTGGGCTCTGTAAAGAATAGTGGGTGGATTCTTCATCCCAAATAAAGTGGTTTCTCAAGTGGTCC CAATTTACAGTTCCTACCATCAGCTTCCCAGTTAAGCTCTGAT	FALSE	765
6	TPOX_4	GCGCTGGTCTTACTCCTGTTCCCTTCCCGTAGGCCCTTCTGTCCTTGTCAGCGTTTATTTGCCCAAAC ATTCATTCATTCATTCAGTGAGGGTTCCTAAGTGCCTGTTCTGTGCCAGTCGACGGTCTGGGTGCT AGTGATCACAGCCCTGAAACACAG	TRUE	1
7	TPOX_5	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTCGACTGGCACAGAACAGGCACCTTAGGG AACCCTCACTGAATGAATGAATGAATGAATGTTTGGGCAAATAAACGCTGACAAGGACAGAAGGG CCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	FALSE	1
8	TPOX_6	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTCGACTGGCACAGAACAGGCACCTTAGGG AACCCTCACTGAATGAATGAATGAATGAATGAATGTTTGGGCAAATAAACGCTGACAAGGACAGA AGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	FALSE	2
9	TPOX_6	GCGCTGGTCTTACTCCTGTTCCCTTCCCGTAGGCCCTTCTGTCCTTGTCAGCGTTTATTTGCCCAAAC ATTCATTCATTCATTCATTCATTCAGTGAGGGTTCCTAAGTGCCTGTTCTGTGCCAGTCGACGGTCT GGGTGCTAGTGATCACAGCCCTGAAACACAG	TRUE	1
10	TPOX_7	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTCGACTGGCACAGAACAGGCACCTTAGGG AACCCTCACTGAATGAATGAATGAATGAATGAATGAATGTTTGGGCAAATAAACGCTGACAAGGA CAGAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	FALSE	13
11	TPOX_7	GCGCTGGTCTTACTCCTGTTCCCTTCCCGTAGGCCCTTCTGTCCTTGTCAGCGTTTATTTGCCCAAAC ATTCATTCATTCATTCATTCATTCATTCAGTGAGGGTTCCTAAGTGCCTGTTCTGTGCCAGTCGACG GTTCTGGGTGCTAGTGATCACAGCCCTGAAACACAG	TRUE	2
12	TPOX_8	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTCGACTGGCACAGAACAGGCACCTTAGGG AACCCTCACTGAATGAATGAATGAATGAATGAATGAATGAATGTTTGGGCAAATAAACGCTGACA AGGACAGAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	FALSE	955
13	TPOX_8	GCGCTGGTCTTACTCCTGTTCCCTTCCCGTAGGCCCTTCTGTCCTTGTCAGCGTTTATTTGCCCAAAC ATTCATTCATTCATTCATTCATTCATTCAGTGAGGGTTCCTAAGTGCCTGTTCTGTGCCAGTCG ACGGTCTGGGTGCTAGTGATCACAGCCCTGAAACACAG	TRUE	167

- STR & Primers Search Report shows the number of FASTA/Q reads found matching the STR and 5 prime and 3 prime individually of an allele of a specific reference locus. The matched allele sequence will be trimmed according to the begin index of the primers user specified and included in the report. The matched STR and allele sequences will be included in the report.

	A	B	C	D	E
1	Allele	STR	Sequence	Is Reverse	Count
2	Amelogenin_X	GAAACATCTG	ATCAGAGCTTAAACTGGGAAGCTGGTGGTAGGAACTGTAAAATCAGGACCACTTGAGAA ACATCTGGGATAAAGAATCAACACACTATTCCTTACAGAGCCCAGGG	TRUE	1299
3	Amelogenin_X	CAGATGTTTC	CCCTGGGCTCTGTAAAGAATAGTGTGTGATTCTTTATCCCAGATGTTTCTCAAGTGGTCT GATTTACAGTTCCTACCACCAGCTCCAGTTAAGCTCTGAT	FALSE	766
4	Amelogenin_Y	ACCACTTTAT	ATCAGAGCTTAAACTGGGAAGCTGATGGTAGGAACTGTAAAATTGGGACCACTTGAGAA ACCACTTTATTTGGGATGAAGAATCCACCCACTATTCCTTACAGAGCCCAGGG	TRUE	1242
5	Amelogenin_Y	ATAAAGTGGT	CCCTGGGCTCTGTAAAGAATAGTGGTGGATTCTTCATCCCAAATAAAGTGGTTTCTCAAG TGGTCCCAATTTACAGTTCCTACCATCAGTTCACAGTTAAGCTCTGAT	FALSE	919
6	TPOX_4	AACATTCATTCATTCAT TCAGTG	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCTAGGCCCTTCTGTCCTTGTGACGCTTTATTTGC CCAAACATTCATTCATTCATTCAGTGAGGGTCCCTAAGTGCCTGTTCTGTGCCAGTCGAC GGTCTGGGTGCTAGTGATCACAGCCCTGAAACACAG	TRUE	1
7	TPOX_5	CACTGAATGAATGAAT GAATGAATGTT	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTGCGACTGGCACAGAACAGGCAC TTAGGGAACCCCTACTGAATGAATGAATGAATGAATGTTGGGCAAATAAACGCTGACA AGGACAGAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	FALSE	1
8	TPOX_6	CACTGAATGAATGAAT GAATGAATGAATGTT	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTGCGACTGGCACAGAACAGGCAC TTAGGGAACCCCTACTGAATGAATGAATGAATGAATGTTGGGCAAATAAACGCTG ACAAGGACAGAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	FALSE	2
9	TPOX_6	AACATTCATTCATTCAT TCATTCATTCAGTG	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCTAGGCCCTTCTGTCCTTGTGACGCTTTATTTGC CCAAACATTCATTCATTCATTCATTCATTCAGTGAGGGTCCCTAAGTGCCTGTTCTGTGCC AGTCGACGTTCTGGGTGCTAGTGATCACAGCCCTGAAACACAG	TRUE	1
10	TPOX_7	CACTGAATGAATGAAT GAATGAATGAATGAAT GTT	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTGCGACTGGCACAGAACAGGCAC TTAGGGAACCCCTACTGAATGAATGAATGAATGAATGAATGTTGGGCAAATAAAA CGCTGACAAGGACAGAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	FALSE	17
11	TPOX_7	AACATTCATTCATTCAT TCATTCATTCATTCAGT G	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCTAGGCCCTTCTGTCCTTGTGACGCTTTATTTGC CCAAACATTCATTCATTCATTCATTCATTCATTCAGTGAGGGTCCCTAAGTGCCTGTTCTG TGCCAGTCGACGGTCTGGGTGCTAGTGATCACAGCCCTGAAACACAG	TRUE	5
12	TPOX_8	CACTGAATGAATGAAT GAATGAATGAATGAAT GAATGTT	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGAACCCTGCGACTGGCACAGAACAGGCAC TTAGGGAACCCCTACTGAATGAATGAATGAATGAATGAATGAATGTTGGGCAAAT AAACGCTGACAAGGACAGAAGGGCCTAGCGGGAAGGGAACAGGAGTAAGACCAGCGC	FALSE	1079
13	TPOX_8	AACATTCATTCATTCAT TCATTCATTCATTCATT CAGTG	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCTAGGCCCTTCTGTCCTTGTGACGCTTTATTTGC CCAAACATTCATTCATTCATTCATTCATTCATTCATTCAGTGAGGGTCCCTAAGTGCCTGT TCTGTGCCAGTCGACGGTCTGGGTGCTAGTGATCACAGCCCTGAAACACAG	TRUE	413

4. STR Search Report shows the number of FASTA/Q reads found matching the STR of an allele of a specific reference locus. The matched STR sequence will be included in the report.

	A	B	C	D
1	Allele	STR	Is Reverse	Count
2	Amelogenin_X	CAGATGTTTC	FALSE	768
3	Amelogenin_X	GAAACATCTG	TRUE	1352
4	Amelogenin_Y	ACCACTTTAT	TRUE	1284
5	Amelogenin_Y	ATAAAGTGGT	FALSE	925
6	TPOX_4	AACATTCATTCATTCATT CAGTG	TRUE	1
7	TPOX_5	CACTGAATGAATGAATGAAT GAATGTT	FALSE	2
8	TPOX_6	AACATTCATTCATTCATT CATTTCATTCAGTG	TRUE	1
9	TPOX_6	CACTGAATGAATGAATGAAT GAATGTT	FALSE	2
10	TPOX_7	AACATTCATTCATTCATT CATTTCATTCAGTG	TRUE	5
11	TPOX_7	CACTGAATGAATGAATGAAT GAATGTT	FALSE	18
12	TPOX_8	AACATTCATTCATTCATT CATTTCATTCAGTG	TRUE	432
13	TPOX_8	CACTGAATGAATGAATGAAT GAATGTT	FALSE	1109

5. Primer Search Report - exclude matched Allele: this report shows the number of FASTA/Q read matching the primers only excluding all reads matched the entire allele. Both trimmed and untrimmed read for the matching primers are

included in the report. The trimmed read is aligned with the matching primers from both ends, including the primers. In addition, the total number of different bases comparing allele and FASTA/Q read will be calculated. Lower Difference value suggests a potential match despite the specific matching allele and FASTA/Q read are not 100% identical.

The Difference value is called the Levenshtein Distance. The Levenshtein Distance between two sequences is the minimum number of single-character edits (i.e. insertions, deletions or substitutions) required to change one sequence into the other. Therefore the Levenshtein Distance is ideal to measure how similar both sequences are – zero difference means identical.

Below shows partial Primer Search Report - exclude matched Allele. Notice the report is sorted by Difference then Allele.

Allele Used in Comparison	Trimmed FASTA/Q Read Sequence	Original FASTA/Q Read Sequence	Is Reverse	Difference
TPOX_8	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCT AGGCCCTTCTGTCCTTGTGACGCGTTTATTG CCCAAACATTCAATTCATTCAATTCATT ATTCATTCAAGTGGGTTCTAAGTGCCTGT TCTGTGCCAGTCGACGTTCTGGGTGCTAGT GATCACAGCCCTGAAACACAG	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCTAGG CCCTTCTGTCCTTGTGACGCGTTTATTGCCCAAAC ATTCATTCAATTCATTCAATTCATTCAATTCAGT GAGGGTTCTAAGTGCCTGTTCTGTGCCAGTCGA CGGTTCTGGGTGCTAGTGATCACAGCCCTGAAAC ACAG	TRUE	1
TPOX_9	CTGTGTTTCAGGGCTGTGATCACTAGCACCC AGAACCGTGCAGTGGCACAGAACAGGCAC TTAGGGAACCCTCACTGAATGAATGAATGA ATGAATGAATGAATGAATGAATGATTTGGG CAAATAAACGCTGACAAGGACAGAAGGGC CTAGCGGGAAGGGAACAGGAGTAAGACCA GCGC	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGA ACCGTCGACTGGCACAGAACAGGCACCTTAGGGA ACCCTCACTGAATGAATGAATGAATGAATGAAT GAATGAATGAATGATTTGGGCAAATAAACGCTG ACAAGGACAGAAGGGCCTAGCGGGAAGGGAAC AGGAGTAAGACCAGCGC	FALSE	1
Amelogenin_X	CCCTGGGCTCTGTAAAGAATAGTGTGTTGA TTCTTTATCCAGATGTTTCTCAAGTGGTCCT GATTTTACAGTTCCTACCACCAGCTTCCCA GTTTAAAGCTCTGAT	AACCTCATCCTGGGCACCCTGGTTATATCAACTT CAGCTATGAGGTAATTTTCTTTACTAATTTT ACCATTGTTTGCCTTAAACAATGCCCTGGGCTCTG TAAAGAATAGTGTGTTGATTCTTTATCCAGATG TTTCTCAAGTGGTCCTGATTTTACAGTTCCTACCA CCAGCTTCCCAAGTTTAAAGCTCTGAT	FALSE	2
Amelogenin_X	CCCTGGGCTCTGTAAAGAATAGTGTGTTGA TTCTTTATCCAGATGTTCCCAAGTGGTCCT GATTTTACAGTTCCTACCACCAGCTTCCAG TTTAAAGCTCTGAT	ACCTCATCCTGGGCACCCTGGTTATATCAACTTC AGCTATGAGGTAATTTTCTTTACTAATTTT CATTGTTTGCCTTAAACAATGCCCTGGGCTCTGTA AAGAATAGTGTGTTGATTCTTTATCCAGATGTT CCCAAGTGGTCCTGATTTTACAGTTCCTACCACC AGCTTCCCAAGTTTAAAGCTCTGAT	FALSE	2
Amelogenin_X	CCCTGGGCTCTGTAAAGAATAGTGTGTCGA TTCTTTATCCAGATGTTTCTCAAGTGGTCCT GATTTTACAGTTCCTACCACCAGCTTCCCA GTTTAAAGCTCTGAT	ACCTCATCCTGGGCACCCTGGTTATATCAACTTC AGCTATGAGGTAATTTTCTTTACTAATTTTGA CCATTGTTTGCCTTAAACAATGCCCTGGGCTCTGT AAAGAATAGTGTGTCGATTCTTTATCCAGATGT TTCTCAAGTGGTCCTGATTTTACAGTTCCTACCAC CATGCTTCCCAAGTTTAAAGCTCTGAT	FALSE	2
Amelogenin_X	CCCTGGGCTCTGTAAAGAATAGTGTGTTGA TTCTTTATCCAGATGCTTCTCAAGTGGTCC TGATTTTACAGTTCCTACCACCAGTCTTCCC AGTTTAAAGCTCTGAT	ACCTCATCCTGGGCACCCTGGTTATATCAACTTC AGCTATGAGGTAATTTTCTTTACTAATTTTGA CCATTGTTTGCCTTAAACAATGCCCTGGGCTCTGT AAAGAATAGTGTGTTGATTCTTTATCCAGATGC TTCTCAAGTGGTCCTGATTTTACAGTTCCTACCAC CAGTCTTCCCAAGTTTAAAGCTCTGAT	FALSE	2

6. Primer Search Report - exclude matched STR & Primers: similar to the above report, this report shows the number of FASTA/Q read matching the primers only excluding all reads matched STR.

Below shows partial Primer Search Report - exclude matched STR & Primers. Notice the report is sorted by Difference.

Allele Used in Comparison	Trimmed FASTA/Q Read Sequence	Original FASTA/Q Read Sequence	Is Reverse	Difference
TPOX_8	CTGTGTTTCAGGGCTGTGATCACTAGCACC CAGAACCCTGCTACTGGCACAGAACAGGC ACTTAGGGAACCCTCACTGAATGAATGACT GAATGAATGAATGAATGAATTGTTGGGCA AATAAACGCTGACAAGGACAGAAGGGCCT AGCGGGAAGGGAACAGGAGTAAGACCAG CGC	CTGTGTTTCAGGGCTGTGATCACTAGCACCCAGA ACCGTCTACTGGCACAGAACAGGCACTTAGGG AACCCCTCACTGAATGAATGACTGAATGAATGAA TGAATGAATTGTTGGGCAAATAAACGCTGACA AGGACAGAAGGGCCTAGCGGGAAGGGAACAG GAGTAAGACCAGCGC	FALSE	3
TPOX_8	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCT AGGCCCTTCTGTCCTTGTGACGTTTATTTG CCCAAACATTCTATTCAATTCATTCAATTCATT CATTCAATTCAGTGAGGTTCTAAGTGCCTGT TCTGTGCCAGTCGACGGTTCTGGGTGCTAG TGATCACAGCCCTGAAACACAG	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCTAGGC CCTTCTGTCCTTGTGACGTTTATTTGCCCAAACA TTCTATTCAATTCATTCAATTCATTCAATTCAGTG AGGTTCTAAGTGCCTGTTCTGTGCCAGTCGACG GTTCTGGGTGCTAGTGATCACAGCCCTGAAACA CAG	TRUE	3
TPOX_8	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCT AGGCCCTTCTGTCCTTGTGACGTTTATTTG CCCAACATTCTATTCAATTCATTCAATTCATT ATTCATTCAATTCAGTGAGGTTCTAAGTGCCTGT TCTGTGCCAGTCGACGGTTCTGGGTGCTAG TGATCACAGCCCTGAAACACAG	GCGCTGGTCTTACTCCTGTTCCCTTCCCGCTAGGC CCTTCTGTCCTTGTGACGTTTATTTGCCCAACAT TCTATTCAATTCATTCAATTCATTCAATTCAGTG AGGGTTCTAAGTGCCTGTTCTGTGCCAGTCGAC GGTTCTGGGTGCTAGTGATCACAGCCCTGAAAC ACAG	TRUE	3
Amelogenin_X	CCCTGGGCTCTGTAAAGAATAGTGTGTTGA TTCTTTATCCCAGATGTTCTCAAGTGGTCCT GATTACAGTTCCTACCATCAGTCTCCAG TTAAGCTCTGAT	ACCTCATCCTGGGCACCCTGGTTATATCAACTTC AGCTATGAGGTAATTTTCTTTACTAATTTTGAC CATTGTTTGCGTTAACAATGCCCTGGGCTCTGTA AAGAATAGTGTGTTGATTCTTTATCCCAGATGTT CTCAAGTGGTCTGATTTACAGTTCCTACCATCA GTCTCCCAGTTAAGCTCTGAT	FALSE	4
Amelogenin_X	CCCTGGGCTCTGTAAAGAATAGTGTGTTGA TTCTTTATCCCAGATGTTCTCAAGTAGTCCT GATTACAGTTCCTACCATCAGTCTCCAGT TTAAGCTCTGAT	ACCTCATCCTGGGCACCCTGGTTATATCAACTTC AGCTATGAGGTAATTTTCTTTACTAATTTTGA CCATTGTTTGCGTTAACAATGCCCTGGGCTCTGT AAAGAATAGTGTGTTGATTCTTTATCCCAGATGT TCTCAAGTAGTCCTGATTTACAGTTCCTACCATC GCTCCCAGTTAAGCTCTGAT	FALSE	4
Amelogenin_X	CCCTGGGCTCTGTAAAGAATAGTGTGTTGA TTCTTTATCCCAGATGTTCTCAAGTGGTCCT GATTTTACAGTTCCTACCATCAGTCTCC CAGTTAAGCTCTGAT	ACCTCATCCTGGGCACCCTGGTTATATCAACTTC AGCTATGAGGTAATTTTCTTTACTAATTTTGA CCATTGTTTGCGTTAACAATGCCCTGGGCTCTGT AAAGAATAGTGTGTTGATTCTTTATCCCAGATGT TCTCAAGTGGTCTGATTTTACAGTTCCTACCTAC CATCGCTCCCAGTTAAGCTCTGAT	FALSE	4

7. Allele Sequence Record: all reference alleles, including allele name, forward sequence, forward STR, forward 3', forward 5', reverse sequence, reverse STR, reverse 3' and reverse 5', are collected in this report.

User is required to provide the forward allele and forward STR sequences for each reference allele, SEQ Mapper will automatically decide the primers based on user entered flanking sequences indices and lengths then prepare the reverse primers, STR and allele sequences.

Allele	Forward Sequence	Forward 5'	Forward STR	Forward 3'	Reverse Sequence	Reverse 3'	Reverse STR	Reverse 5'
Amelogenin_X	CCCTGGGCTCTGTAAAGAATA GTGTGTTGATTCTTTATCCAG ATGTTTCTCAAGTGGTCTGA TTTTACAGTTCCTACCACCAG CTCCCAAGTTAAGCTCTGAT	CCCTGGGCTC	CAGATGTTTC	AAGCTCTGAT	ATCAGAGCTTAAACTGGGAA GCTGGTGGTAGGAACTGTAA AATCAGGACCACTTGAGAAA CATCTGGGATAAAGAATCAA CACACTATTCTTTACAGAGCC CAGGG	ATCAGAGCTT	GAAACATCTG	GAGCCCAGGG
Amelogenin_Y	CCCTGGGCTCTGTAAAGAATA GTGGGTGGATTCTTCATCCCA AATAAAGTGGTTTCTCAAGTG GTCCCAATTTTACAGTTCCTAC CATCAGCTTCCCAGTTAAGC TCTGAT	CCCTGGGCTC	ATAAAGTGGT	AAGCTCTGAT	ATCAGAGCTTAAACTGGGAA GCTGATGGTAGGAACTGTAA AATTGGGACCACTTGAGAAA CCACTTTATTTGGGATGAAGA ATCCACCACTATTCTTTACA GAGCCCAGGG	ATCAGAGCTT	ACCACCTTAT	GAGCCCAGGG
TPOX_4	CTGTGTTTCAGGGCTGTGATC ACTAGCACCCAGAACCGTCG ACTGGCACAGAACAGGCACT TAGGGAACCCTCACTGAATG AATGAATGAATGTTTGGGCA AATAAACGCTGACAAGGACA GAAGGGCTAGCGGGAAGGG AACAGGAGTAAGACCAGCGC	CTGTGTTTCA	CACTGAATGAA TGAATGAATGT T	AGACCAGCGC	GCGCTGGTCTTACTCCTGTTC CTTCCCCTAGGCCCTTCTGTC CTTGTGAGCGTTTATTTGCCCA AACATTCATTCATTCATTCAGT GAGGGTCCCTAAGTGCCTGT TCTGTGCCAGTCGACGTTCT GGGTGCTAGTGATCACAGCCC TGAAACACAG	GCGCTGGTCT	AACATTCATTC ATTCATTCAGTG	TGAAACACAG
TPOX_5	CTGTGTTTCAGGGCTGTGATC ACTAGCACCCAGAACCGTCG ACTGGCACAGAACAGGCACT TAGGGAACCCTCACTGAATG AATGAATGAATGAATGTTTGG GCAAATAAACGCTGACAAGG ACAGAAGGGCCTAGCGGAA GGGAACAGGAGTAAGACCAG CGC	CTGTGTTTCA	CACTGAATGAA TGAATGAATGA ATGTT	AGACCAGCGC	GCGCTGGTCTTACTCCTGTTC CTTCCCCTAGGCCCTTCTGTC CTTGTGAGCGTTTATTTGCCCA AACATTCATTCATTCATTCATT CAGTGAGGGTCCCTAAGTGC CTGTTCTGTGCCAGTCGACGG TTCTGGGTGCTAGTGATCACA GCCTGAAACACAG	GCGCTGGTCT	AACATTCATTC ATTCATTCATTC AGTG	TGAAACACAG

8. Skipped FASTA Read Sequence: all FASTA/Q reads excluded from matching are collected in this report. A possible cause is that the read is too short to satisfy the begin index and length of the primers specified by the user. In this case, user is advised to review the data source.

```
>CRVEA:02079:03350: FASTA data skipped
>CRVEA:00004:01405
GATAGATAGATAGGAT
>CRVEA:00005:00787
ATCAGAGCTTAAA
```

9. No Match FASTA Read Sequence: as a last resort, all FASTA/Q reads matching no reference allele are collected in this report.

```
>CRVEA:00200:00987
GCATGAAGCTGCACCGAAAG
>CRVEA:00207:02694
ATCAGAGCTTAAACTGGGA
>CRVEA:00223:00484
GCATGAAGCTGCACCGAAAG
```

The begin index and length of primers can help the user to determine how primers can be extracted from the reference alleles. It depends of how the reference data is prepared in the lab, sometimes the result may not completely meet user's expectation. Using the No Match FASTA Read Sequence can provide the user a second chance to "tune" or better organize the reference data for more ideal results.