July. 2022

Quick Start Guide

Visual Microhap

http://forensic.yonsei.ac.kr/VisualMH/index.html

Yonsei DNA profiling Group

Department of Forensic Medicine
Yonsei University College of Medicine
Seoul, Republic of Korea

TABLE OF CONTENTS

1. Visual Microhap

For the practical application of massively parallel sequencing (MPS) to forensics, we developed a custom haplotype caller, ***Visual Microhap***, that presents a simple workflow schemes to analyze sequence-based microhap data. The Visual Microhap works on a web browser and generates SNP-based haplotype data with four analysis options from sequence-based haplotype data obtained by STRait Razor 3.0.

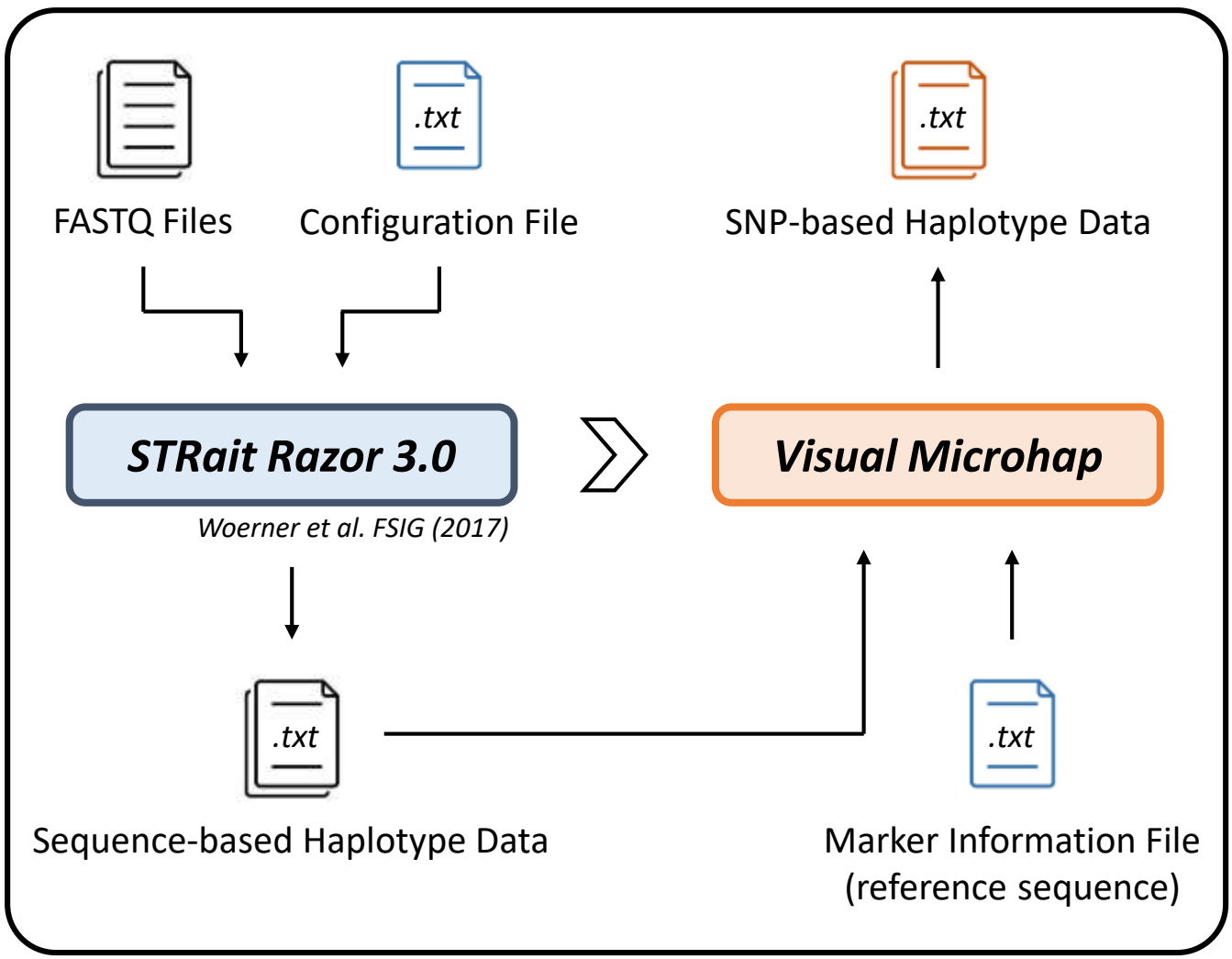


Figure 1. An open-source workflow to analyze sequence-based data of microhaps.

1. Visual Microhap

Visual Microhap was coded in Vanilla JavaScript with Bootstrap v5.0 for user interface; and visualizes the sequence data resulting from STRait Razor 3.0. The sequence-based data generated as a result of the STRait Razor 3.0 was compared with the reference sequence using Visual Microhap and the variations between both sequences were listed in *Result panel* of Visual Microhap.

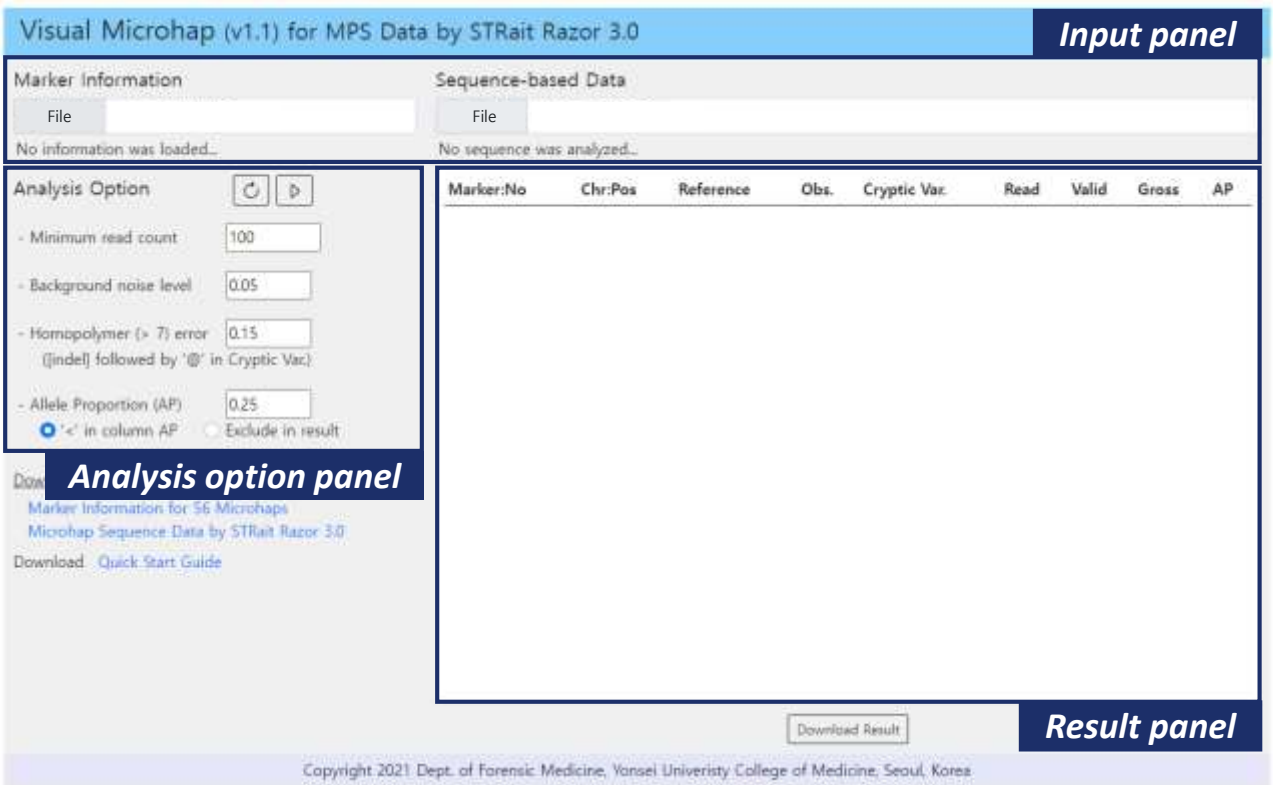


Figure 2. Visual Microhap for MPS. Analysis options and demo files are presented in the left panel and the analysis results are listed in the right panel.

2. How to run

2.1. Input file

You need two types of input files: a *Marker information* and a *Sequence-based haplotype data*. You can download demo-files here.

- ❖ Marker Information for 56 Microhaps
- ❖ Sequence-based Haplotype Data File by STRait Razor 3.0

2.1.1. Marker information file (*.txt*)

Marker information file consists of information about the reference sequence. Marker information file is in a *tab-delimited* text format. The structure of the marker information file is as follows.

- 1 col: [CHARACTER] Marker name
- 2 col: [CHARACTER] Position for the target SNP #1
- 3 col: [CHARACTER] Position for the target SNP #2
  ⋮
  } rs number:position
- N-1 col: [NUMERIC] Start position of the reference sequence
- N col: [CHARACTER] Reference sequence corresponding to the result of STRait Razor 3.0 (case-insensitive)

Example of marker information file

mh01KK_002 chr1 rs4528199:216461086 rs6604596:216461103 216461033
tttcaactattattattattacccagttagaaagtgaataaatgacctaaatgGgaaacctgacataggtAgacatattggct

2. How to run

2.1.2. Sequence-based haplotype data (*.txt*)

Sequence-based haplotype data is an output file of STRait Razor 3.0 and contains unique phase-known haplotype sequences and their length and coverage observed for each microhap. It is recommended to use the analytical options for STRait Razor 3.0 as default except for 10 × minimum read count of each marker.

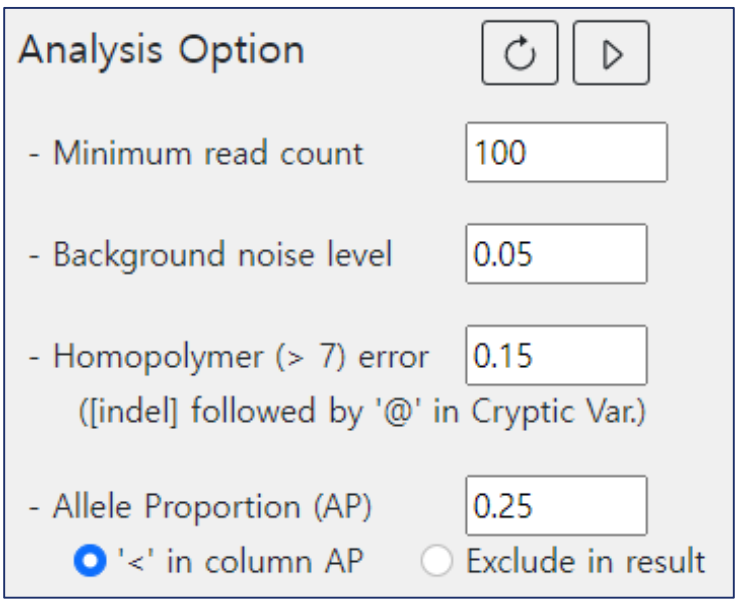
- 1 col: Marker name
- 2 col: Size of the allele
- 3 col: Allelic sequence
- 4 col: Coverage of read #1
- 5 col: Coverage of read #2

Example of sequence data file

mh01KK_002:0	83 bases	TTTCAACTA … ATGGGAAA … GGTAGACATATTGGCT	1514	1453
mh01KK_002:0	83 bases	TTTCAACTA … ATGAGAAA … GGTAGACATATTGGCT	1191	1147
mh01KK_002:-3	80 bases	TTTCAACTA … ATGGGAAA … GGTAGACATATTGGCT	23	25
mh01KK_002:-3	80 bases	TTTCAACTA … ATGAGAAA … GGTAGACATATTGGCT	23	23
mh01KK_002:0	83 bases	TTTCAACTA … ATGGGAAA … GGTAGACATATTGGCT	34	1
mh01KK_002:0	83 bases	TTTCAACTA … ATGAGAAA … GGTAGACATATTGGCT	28	2
mh01KK_002:0	83 bases	TTTCAACTA … ATGAGAAA … GGTAGACATATTGGCT	8	6
mh01KK_002:0	83 bases	TTTCAACTA … ATGGGAAA … GGTAGACATATTGGCT	6	6
mh01KK_002:0	83 bases	TTTCAACTA … ATGGGAAA … GGTAGACATATTGGCT	5	5
mh01KK_002:0.0	0 bases	SumBelowThreshold	310	418

2. How to run

2.2. Analysis option



1) Minimum read count

'Minimum read count' removes reads with below designated number, which can be set to 10 minimally for reliability of read.

2) Background noise level

'Background noise level' excludes reads with under the assigned proportion over gross read.

(continued)

2. How to run

2.2. Analysis option

3) Homopolymer (> 7) error

'Homopolymer (> 7) error' indicates '@' following indel sequence in the column *Cryptic Var.* when a proportion of reads with homopolymer indels which are observed in the region exceeding seven consecutive bases is above the threshold value.

4) Allele Proportion (AP)

'Allele proportion (AP)' marks '<' in the column AP if a proportion of reads is below the threshold value among observed haplotypes in each marker. This haplotype can be excluded from the result by clicking 'Exclude in result' in the analysis option.

3. Interpretation

3.1. Output file

The result of Visual Microhap can be downloaded as a *tab-delimited* text file by clicking the '*Download Result*' button.

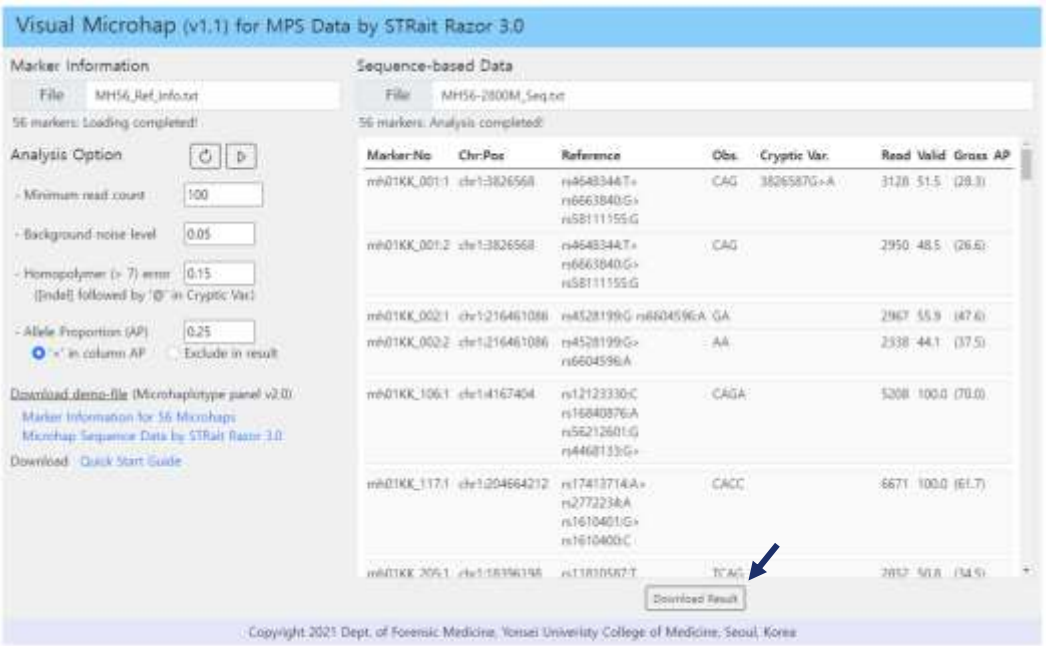


Figure 3. Visual Microhap screenshot for 2800M reference sample

The structure of the output file is as follows.

- *Marker:No*: marker name with allele count
  → Homozygote: only one allele is observed in a marker.
  → Heterozygote: two or more alleles are observed in a marker.
- *Chr:Pos*: the position of the first target SNP in each microhap

(continued)

3. Interpretation

- *Reference.*: rs number of target SNPs and genotype of reference sequence
- *Obs.*: genotype for target SNPs

  In the first allele of mh01KK-001 in Figure 3, the first target SNP (rs4648344) had T > ***C*** variation, the second target SNP (rs6663840) had G > ***A*** variation, and the third target SNP (rs58111155) has no variation (same as the reference sequence, ***G***).
- *Cryptic Var.*: additional variations out of target SNPs including indels (+[sequences] or –[sequences]) with the position
- *Read*: read count
- *Valid*: allele coverage ratio for all reads except noise
- *Gross*: allele coverage ratio for all reads including noise
- *AP*: marked with '<' when an allele is below the threshold proportion.

4. Appendix

Table 1. 2800M genotype *(continued)*

Microhap	rs number of target SNPs:Reference	Genotype	
mh01KK-001	rs4648344:T rs6663840:G rs58111155:G	CAG	CAG
mh01KK-002	rs4528199:G rs6604596:A	GA	AA
mh01KK-106	rs12123330:C rs16840876:A rs56212601:G rs4468133:G	CAGA	CAGA
mh01KK-117	rs17413714:A rs2772234:A rs1610401:G rs1610400:C	CACC	CACC
mh01KK-205	rs11810587:T rs1336130:T rs1533623:G rs1533622:G	TCAG	TTAA
mh01NK-001	rs2296796:G rs2296797:G rs2296798:G	GGG	GGG
mh02KK-003	rs260694:G rs11123719:T rs11691107:C	TTC	TTC
mh02KK-134	rs12469721:A rs3101043:T rs3111398:C rs72623112:G	ACTG	TTTG
mh02KK-136	rs6714835:T rs6756898:C rs12617010:C	TCC	GTC
mh03KK-150	rs1225051:G rs1225050:G rs1225049:C rs1225048:C	AACA	AATA
mh04KK-010	rs3135123:G rs495367:A	AA	GA
mh04KK-013	rs13131164:C rs3775866:G rs11725922:G rs3775867:G rs17088476:T	CGGAT	CGGAT
mh04KK-017	rs4699748:G rs2584461:T rs1442492:A	GCA	ACA
mh05KK-170	rs74865590:C rs438055:A rs370672:G rs6555108:A	CAAG	CAGA
mh09KK-033	rs10815466:G rs9408671:C rs17431629:G	GCG	GCG
mh09KK-035	rs3118582:T rs10776839:G	TG	TG
mh09KK-152	rs10867949:A rs4282648:G rs10780576:C rs7046769:A	AGCA	GTCG
mh09KK-153	rs10125791:T rs2987741:A rs7047561:A	TAC	TAA
mh09KK-157	rs606141:G rs8193001:C rs56256724:C rs2073578:A rs633153:C	GCCCT	GTCAC
mh10KK-163	rs3814588:T rs6602026:G rs3814589:G rs3814590:C	AAGC	AAGC
mh10KK-169	rs10796164:A rs10796165:T rs17154765:T rs10796166:A	GCTG	ACTG
mh11KK-036	rs10500616:A rs2499936:A	CG	CG
mh11KK-103	rs1107162:G rs2075654:C rs1079727:T	GCT	GCT
mh11KK-180	rs12802112:A rs28631755:A rs7112918:T rs4752777:C	ACTG	GCCG
mh11KK-187	rs493442:G rs17137917:C rs551850:G rs17137926:G	GCGG	CCCA
mh11KK-191	rs12421109:T rs12289401:A rs12420819:A rs770566:T	TAAC	TAAC
mh12KK-046	rs1503767:T rs11068953:G	TA	TA
mh12KK-202	rs10506052:A rs4931233:G rs10506053:T rs4931234:T	AATC	CATT

4. Appendix

Table 1. 2800M genotype *(continued)*

Microhap	rs number of target SNPs:Reference	Genotype	
mh13KK-213	rs8181845:C rs679482:C rs9510616:G	TCA	TCA
mh13KK-217	rs9562648:G rs9562649:C rs2765614:A	GCG	GCG
mh13KK-218	rs1927847:T rs9536429:T rs7492234:T rs9536430:C	CCCT	TTTT
mh13KK-223	rs1192204:T rs1192205:C rs3825483:C rs3825481:T	CGCT	CGTT
mh13KK-225	rs4884651:G rs9529023:A rs7329287:G	ACG	GAA
mh14KK-048	rs12717560:A rs12878166:C	GT	GT
mh14KK-101	rs28529526:G rs10134526:T	GT	GC
mh15KK-066	rs1063902:A rs4219:G	CT	AT
mh15KK-067	rs701463:T rs701464:C	GT	TT
mh15KK-104	rs11631544:T rs10152453:C rs80047978:G	CAG	TAG
mh16KK-049	rs9937467:A rs17670098:C rs17670111:A rs12929083:A rs9926495:G	ACAAA	ACAAG
mh16KK-255	rs16956011:G rs3934955:A rs3934956:C rs4073828:A	GACA	ACCG
mh16KK-302	rs1395579:G rs1395580:C rs1395582:T rs9939248:T	ACTT	ACTT
mh17KK-052	rs1059504:A rs8327:A	AA	AA
mh17KK-054	rs2233362:A rs634370:G	AA	AA
mh17KK-272	rs2934897:T rs7207239:C rs16955257:C rs7212184:T	TCCC	CCCT
mh18KK-285	rs16940823:C rs17187688:G rs17187695:C rs1945150:G	CGCT	CGCT
mh18KK-293	rs621320:A rs621340:T rs678179:G rs621766:A	AGAA	AGAA
mh19KK-299	rs12985452:G rs4932999:C rs4932769:A rs2361019:T rs2860462:G	ATGAA	ATGAA
mh19KK-301	rs10408594:G rs11084040:G rs10408037:A rs8104441:T	GGAT	GGAT
mh20KK-058	rs6122890:T rs6095836:A rs6012881:T	CAC	TAC
mh20KK-307	rs6044080:T rs17674942:T rs6044081:A rs16997830:A	CTGA	CTGA
mh21KK-315	rs8126597:G rs8131148:C rs6517971:T	GCT	GTC
mh21KK-316	rs961302:A rs17002090:C rs961301:A rs2830208:C	GCGC	ACGT
mh21KK-320	rs2838081:G rs2838082:A rs78902658:C rs2838083:A	AACG	AACG
mh21KK-324	rs6518223:T rs2838868:C rs7279250:T rs8133697:G	TCAG	TCTG
mh22KK-061	rs763040:G rs5764924:A rs763041:G	GAA	GAA
mh22KK-069	rs8137373:G rs2235845:T	GT	GT