

Introduction

Pop Inference is designed to allow users to submit DNA sequences for unknown individuals in FASTA format. The results of ethnic inference for continental populations and subpopulations accuracy profiles can be provided by this webtool.

Pop Inference app requires users to create an access account at <http://forensic.mc.ntu.edu.tw:9000/DNAToolWeb>. Upon successful login, user will have access to Pop Inference:

DNAToolWeb YSTRSearch YHGSearch mtHGSearch **Pop Inference** Admin Hello test@test.com! Log off

About Pop Inference
Workbench

PopInference Workbench

Unknown Sample Fasta File Help

Workbench

Select File Unknown Sample FASTA Sequence Data File ⓘ

	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Race Profile K ⓘ	5	10	35
Sub Population Profile K ⓘ	5	10	35

Select All

Report Name ⓘ Submit

Download ⓘ

- Pop Inference Debug 20201126-165446.zip
- Pop Inference New Report with NTU Title 20201125-164638.zip
- Pop Inference Unknown subpop 20201124-182137.zip

Info/Status ⓘ

Your last request received by server on 11-28-2020 15:38:12 is still running with status "11/28/2020 15:38:11 started". Please periodically refresh the stat to check the report result. Contact us if you have any questions.

© 2020 - Forensic Medicine Dept. NTU

Select Unknown Sample FASTA File

A FASTA file containing unknown sample individual is required. From the above app main screen, a "Select File" button can assist the user to select the FASTA file for upload. Notice a blue information icon is the context sensitive help button. Below is the info provided for the "Select File" function:

Info/Status ⓘ

You may select a Fasta sequence data file containing unknown samples. The Fasta ID of each unknown sample sequence must contain a unique ID with max length 10 letters then followed by "UNK" to represent unknown race and unknown sub-population like ">Sample1_UNK_UNK", ">Sample2_Africa_UNK", separated by underscore.

Below is a sample data file which can be downloaded by clicking on the “Unknown Sample Fasta File” link in the main screen:

```
>SampleID_UNK_UNK
TAAGATTCTAATTTAAACTATTCTCTGTTCTTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGT
ATTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAAACCAATCCACATCAAAA
CCCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCAAAGCCACCCCTCACCCACTAGGAT
ACCAACAAACCCACCCACCCTCAACAGTACATAGTACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATG
GATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCAT
AACACTTGGGGGTAGCTAAAGTGAAGTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAA
TAAGACATCACGATGGATCACAGGTCTATCACCTATTAACCACTCACGGGAGCTCTCCATGCATTTGGTATTTTCGTCTGGGGGGTGTGCAC
GCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCTATGTCGCAGTATCTGTCTTTGATTCTGCCTCATTCTATTATTTATCGCACCTACG
TTCAATATTACAGGCGAACATACTTACTAAAGTGTGTTAATTAATTAATGCTTGTAGGACATAATAATAACAATTGATGTCTGCACAGCCGCT
TTCCACACAGACATCATAACAAAAAATTTCCACCAAACCCCCCTCCCCCGCTTCTGGCCACAGCACTTAACACATCTCTGCCAAACCCCAA
AAACAAAGAACCCTAACACCAGCCTAACAGATTTCAAATTTTATCTTT
```

Race/Sub Population Profile K

☒

☒

☒

Race Profile K

5

10

35

Sub Population Profile K

5

10

35

Select All

Report Name

Submit

Info/Status

Three most appropriated K values tested for race profile in our study.

The next user input is to select the Race Profile K and Sub Population Profile K.

In our study we have identified and tested three most appropriated K values for the respective race profile and sub population profiling. The Pop Inference app allows users to select up to three of them.

Custom Report Name

Report Name

Submit

Info/Status


Personalize report name can be useful for identifying your report. Please use only letters, numbers, space, hyphen, underscore and parantheses.

Before users click on the Submit button, Pop Inference app allows users to customize their reports.



A Sample Test

PopInference Workbench


Workbench

Select File Unknown Sample FASTA Sequence Data File 

Original File Path: /home/.../UNK.txt

	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Race Profile K 	5	10	35
Sub Population Profile K 	5	10	35

Select All

Report Name  Sample Test

Submit

Info/Status

You have successfully submitted your request. Please note that in order to preserve server resources, we limit one request at a time. You may monitor the report status by refreshing the stat and download list periodically.

Notice the Submit button is disabled once a request was submitted.

Also notice a refresh icon in the Info/Status area, clicking on it will show the report generation status. The Submit button will be enabled when the report gen is completed or if errors occurred.


Below is a complete sequence of operations for the sample test captured in the Info/Status area:

Info/Status


11/28/2020 18:17:13 started

Info/Status


11/28/2020 18:17:13 started, 11/28/2020 19:44:18 Fasta align completed

Info/Status


11/28/2020 18:17:13 started, 11/28/2020 19:44:18 Fasta align completed,
11/28/2020 21:11:46 PHYLIP gen completed

Info/Status


11/28/2020 18:17:13 started, 11/28/2020 19:44:18 Fasta align completed,
11/28/2020 21:11:46 PHYLIP gen completed, 11/28/2020 21:13:36 dist matrix gen completed

Info/Status


11/28/2020 18:17:13 started, 11/28/2020 19:44:18 Fasta align completed,
11/28/2020 21:11:46 PHYLIP gen completed, 11/28/2020 21:13:36 dist matrix gen completed, 11/28/2020 21:15:06 completed

This report took roughly 3 hours to complete













Report Download

Click on the refresh icon will list all reports generated for the user.

Download


- Pop Inference Sample Test 20201128-211507.zip

The report is a downloadable ZIP container file which contains the race and sub population profile for each K values selected based on KNN and KWNN algorithm, respectively.

-  KNN Unknown Race Profile K=5.csv
-  KNN Unknown Race Profile K=10.csv
-  KNN Unknown Race Profile K=35.csv
-  KNN Unknown Sub Pop Profile K=5.csv
-  KNN Unknown Sub Pop Profile K=10.csv
-  KNN Unknown Sub Pop Profile K=35.csv
-  KWNN Unknown Race Profile K=5.csv
-  KWNN Unknown Race Profile K=10.csv
-  KWNN Unknown Race Profile K=35.csv
-  KWNN Unknown Sub Pop Profile K=5.csv
-  KWNN Unknown Sub Pop Profile K=10.csv
-  KWNN Unknown Sub Pop Profile K=35.csv

Below is the sample result for race profile using KNN method with K = 5:

Report of PopInference					
Report Time: 2020-11-28 21:15:07 (UTC+08:00) Taipei					
Forensic Medicine Dept. of NTU					
Distance of K Nearest Samples					
K=5	THanA11_Asia_Han	THanA12_Asia_Han	PAI35ko_Asia_Paiwan	PIN47ko_Asia_Pingpu	HAK21ko_Asia_Hakka
SampleID_UNK_UNK	0.003066	0.003066	0.003076	0.003076	0.003076
Probability of Population					
K=5	Asia	Africa	Europe		
SampleID_UNK_UNK	1	0	0		

Below is the sample result for sub population profile using KNN method with K = 5:

Report of PopInference					
Report Time: 2020-11-28 21:15:07 (UTC+08:00) Taipei					
Forensic Medicine Dept. of NTU					
Distance of K Nearest Samples					
K=5	THanA11_Asia_Han	THanA12_Asia_Han	PAI35ko_Asia_Paiwan	PIN47ko_Asia_Pingpu	HAK21ko_Asia_Hakka
SampleID_UNK_UNK	0.003066	0.003066	0.003076	0.003076	0.003076
Probability of Population					
K=5	Asia Han	Asia Paiwan	Asia Pingpu	Asia Hakka	
SampleID_UNK_UNK	0.4	0.2	0.2	0.2	