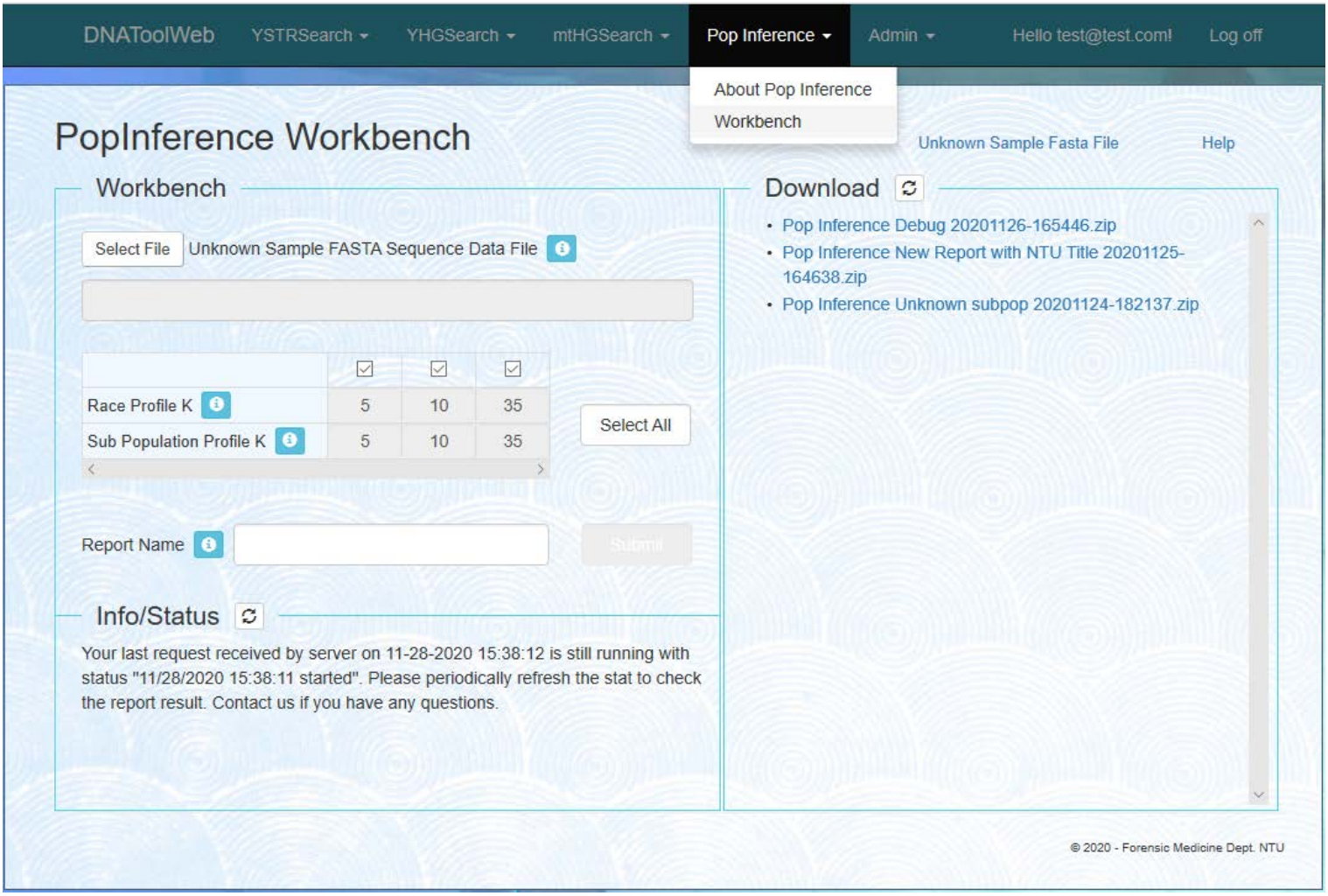
Introduction

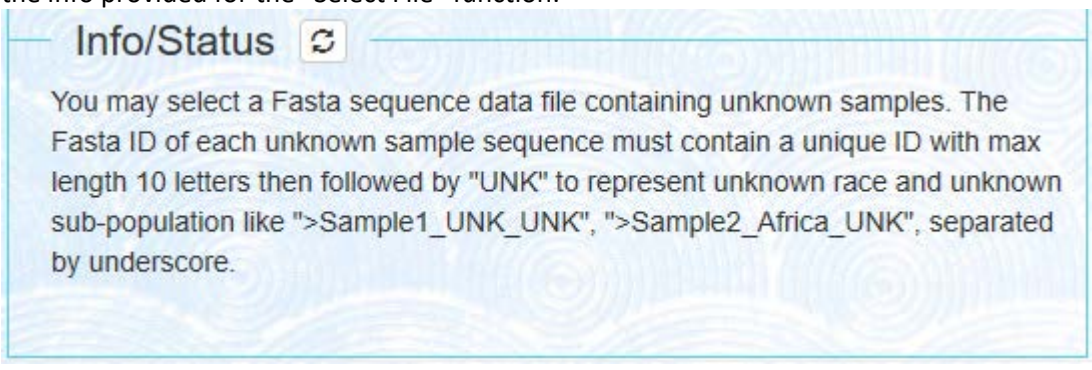
Pop Inference is designed to allow users to submit DNA sequences for unknown individuals in FASTA format. The results of ethnic inference for continental populations and subpopulations accuracy profiles can be provided by this webtool.

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



Select Unknown Sample FASTA File

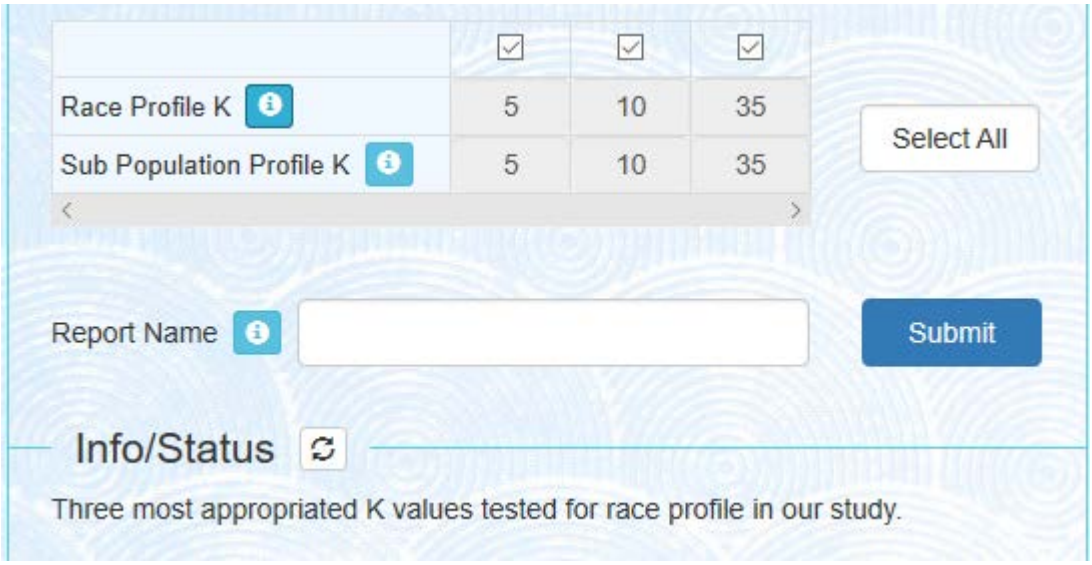
A FASTA file containing unknown sample individual is required. From the above app main screen, a “Select File” button can assist the user to select the FASTA file for upload. Notice a blue information icon is the context sensitive help button. Below is the info provided for the “Select File” function:



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
TAAGATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGT
ATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAA
CCCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGAT
ACCAACAAACCCACCCACCCTCAACAGTACATAGTACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATG
GATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCAT
AACACTTGGGGGTAGCTAAAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAA
TAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACGGGAGCTCTCCATGCATTTGGTATTTTCGTCTGGGGGGTGTGCAC
GCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCCTATGTCGCAGTATCTGTCTTTGATTCCTGCCTCATTCTATTATTTATCGCACCTACG
TTCAATATTACAGGCGAACATACTTACTAAAGTGTGTTAATTAATTAATGCTTGTAGGACATAATAATAACAATTGATGTCTGCACAGCCGCT
TTCCACACAGACATCATAACAAAAAATTTCCACCAAACCCCCCCTCCCCCCGCTTCTGGCCACAGCACTTAAACACATCTCTGCCAAACCCCAA
AAACAAAGAACCCTAACACCAGCCTAACCAGATTTCAAATTTTATCTTT

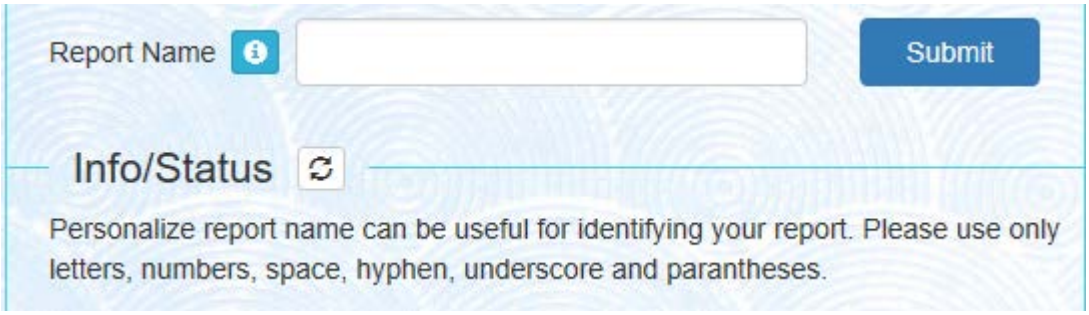
Race/Sub Population Profile K



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

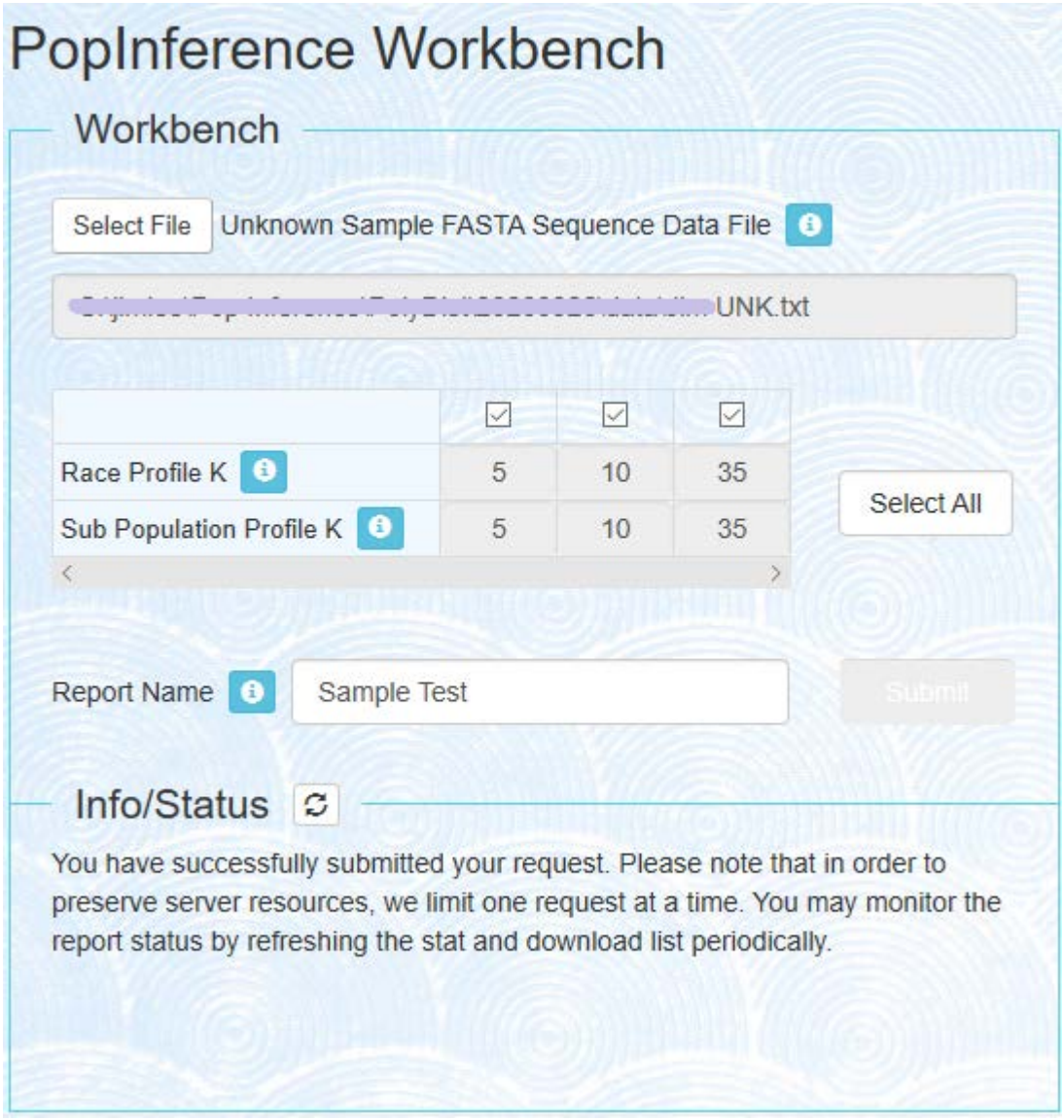
In our study we have identified and tested three most appropriated K values for the respective race profile and sub population profiling. The Pop Inference app allows users to select up to three of them.

Custom Report Name



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

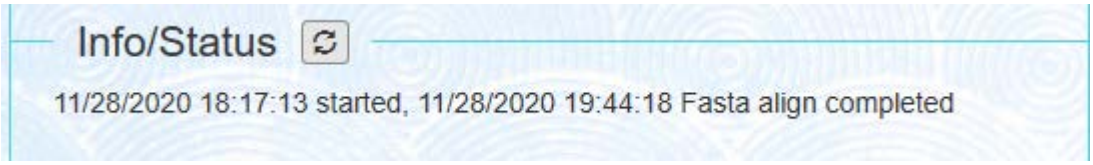
A Sample Test

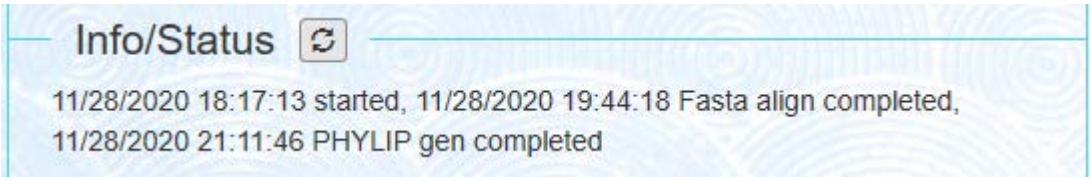


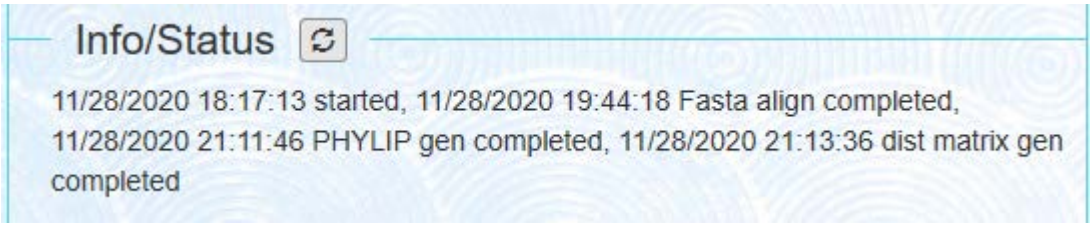
Notice the Submit button is disabled once a request was submitted.

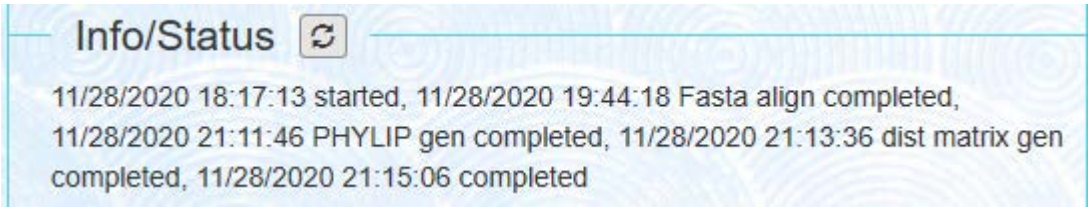
Also notice a refresh icon in the Info/Status area, clicking on it will show the report generation status. The Submit button will be enabled when the report gen is completed or if errors occurred.

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







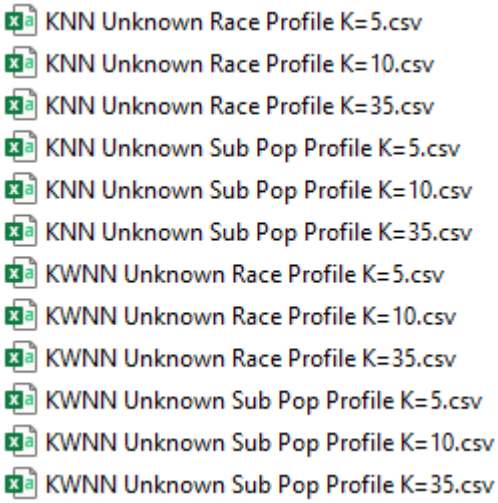


This report took roughly 3 hours to complete

Report Download

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

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.



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	