# Introduction

# Methods

## Sample collection and sequencing

## Typing methods

For genotyping scheme not all levels are given by the program. Every missing level was designated 0. Note all previous levels are designated by the software is a more granular level is provided.

## Group size determination

## Statistical analysis

# Results

## Level of granularity

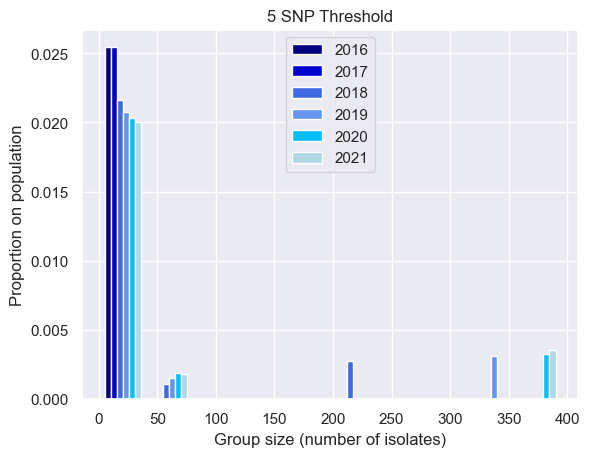
A picture containing screenshot, square, rectangle, line

Description automatically generated Figure . Comparison of the number of groups at each level for each typing scheme.

Add description.

## Group size through time

Method 1 = model of actual surveillance – real life

 A picture containing text, screenshot, plot, diagram

Description automatically generated

Full dataset

Non-MSM dataset

A picture containing text, screenshot, plot, diagram

Description automatically generated A picture containing text, screenshot, plot, diagram

Description automatically generated

A picture containing text, screenshot, plot, diagram

Description automatically generated A picture containing text, screenshot, plot, number

Description automatically generated

A picture containing text, screenshot, diagram, plot

Description automatically generated A picture containing text, screenshot, plot, diagram

Description automatically generated

Figure . Group size increases with time during surveillance, with very large group size driven by MSM clades.

Very large groups, present in the full dataset (left), are missing in the non-MSM dataset.

A picture containing text, screenshot, diagram, plot

Description automatically generated A picture containing text, screenshot, plot, diagram

Description automatically generatedA picture containing text, screenshot, plot, diagram

Description automatically generated A picture containing text, screenshot, diagram, plot

Description automatically generated

Figure . MSM-only dataset group size through time, by SNP address level (left) and cgMLST level (right).

Group size increases with time, with some groups becoming very large. For SNP address, the majority of the dataset belongs to large groups (> 200 isolates) by 2018. At the more granular level (HC2), the cgMLST typing groups the MSM-only datatset into predominantly small groups (< 50 isolates) across all years.

Method 2 = removes the cumulative effect of surveillance through time which leads to increasing group sizes with time.

A picture containing text, screenshot, plot, diagram

Description automatically generated A picture containing text, screenshot, plot, diagram

Description automatically generated

A picture containing text, screenshot, plot, diagram

Description automatically generated A picture containing text, screenshot, plot, diagram

Description automatically generated

A picture containing text, screenshot, plot, diagram

Description automatically generated A picture containing text, screenshot, plot, number

Description automatically generated

A picture containing text, screenshot, diagram, plot

Description automatically generated A picture containing text, screenshot, plot, number

Description automatically generated

Figure .Group size increases with time during surveillance, even when accounting for the effect of increasing sample size with time.

Very large groups, present in the full dataset (left), are missing in the non-MSM dataset, showing that very large groups size remains exclusively MSM clade related.

A picture containing text, screenshot, plot, diagram

Description automatically generated A picture containing text, screenshot, diagram, plot

Description automatically generated

A picture containing text, screenshot, plot, diagram

Description automatically generated A picture containing text, screenshot, diagram, plot

Description automatically generated

Figure . MSM-only dataset group size remains constant or decreases through time, by SNP address level (left) and cgMLST level (right).

Majority of MSM clades are grouped into large groups (SNP address and HC5 cgMLST level), however, the HC2 cgMLST level (top right) groups the MSM clades predominantly into small group sizes.

## Association between variables

A close-up of a grid

Description automatically generated with low confidence

Figure . Full dataset Theil’s U comparison of all variables

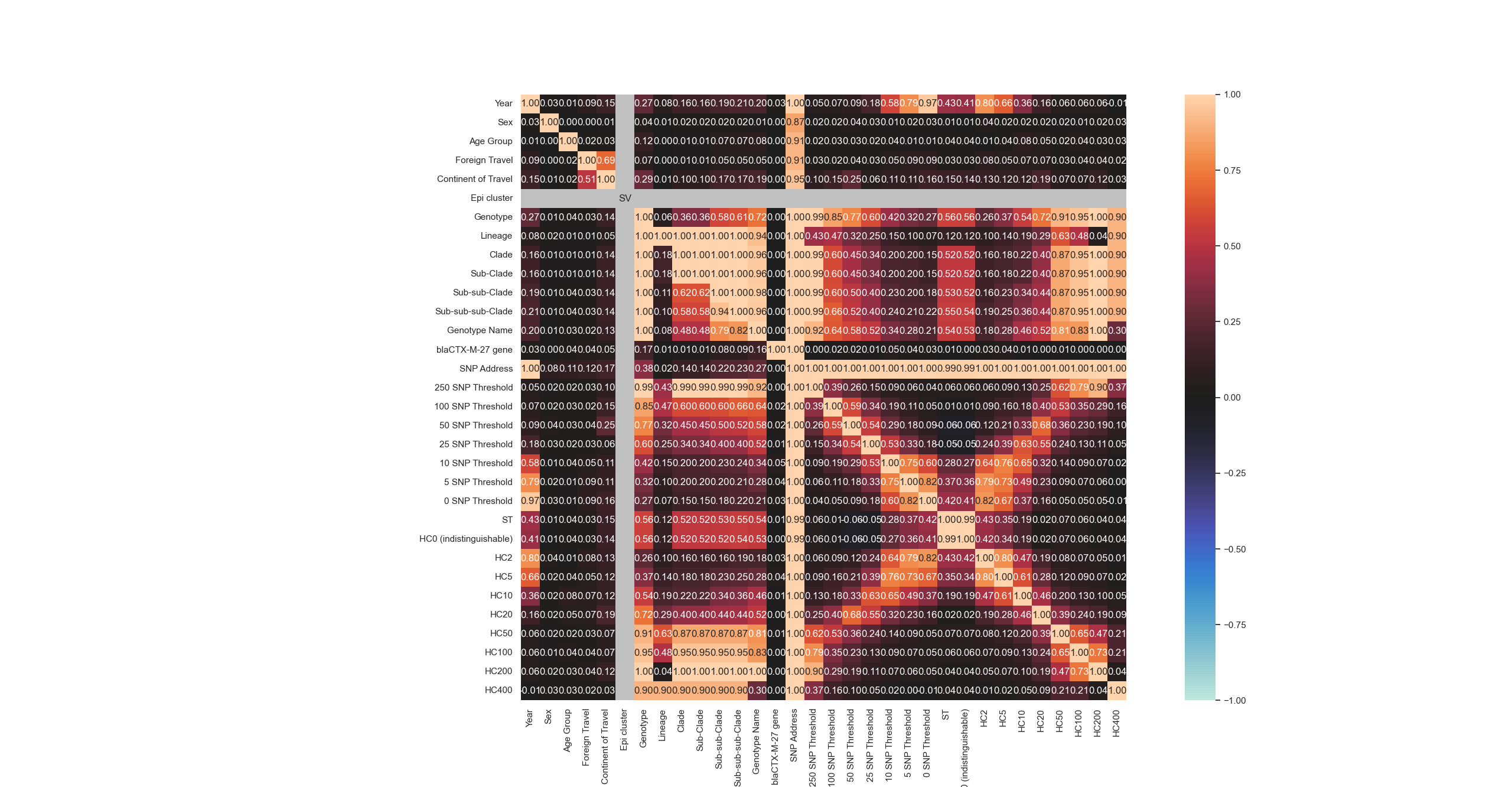


Figure . Non-MSM dataset Theil’s U comparison of all variables

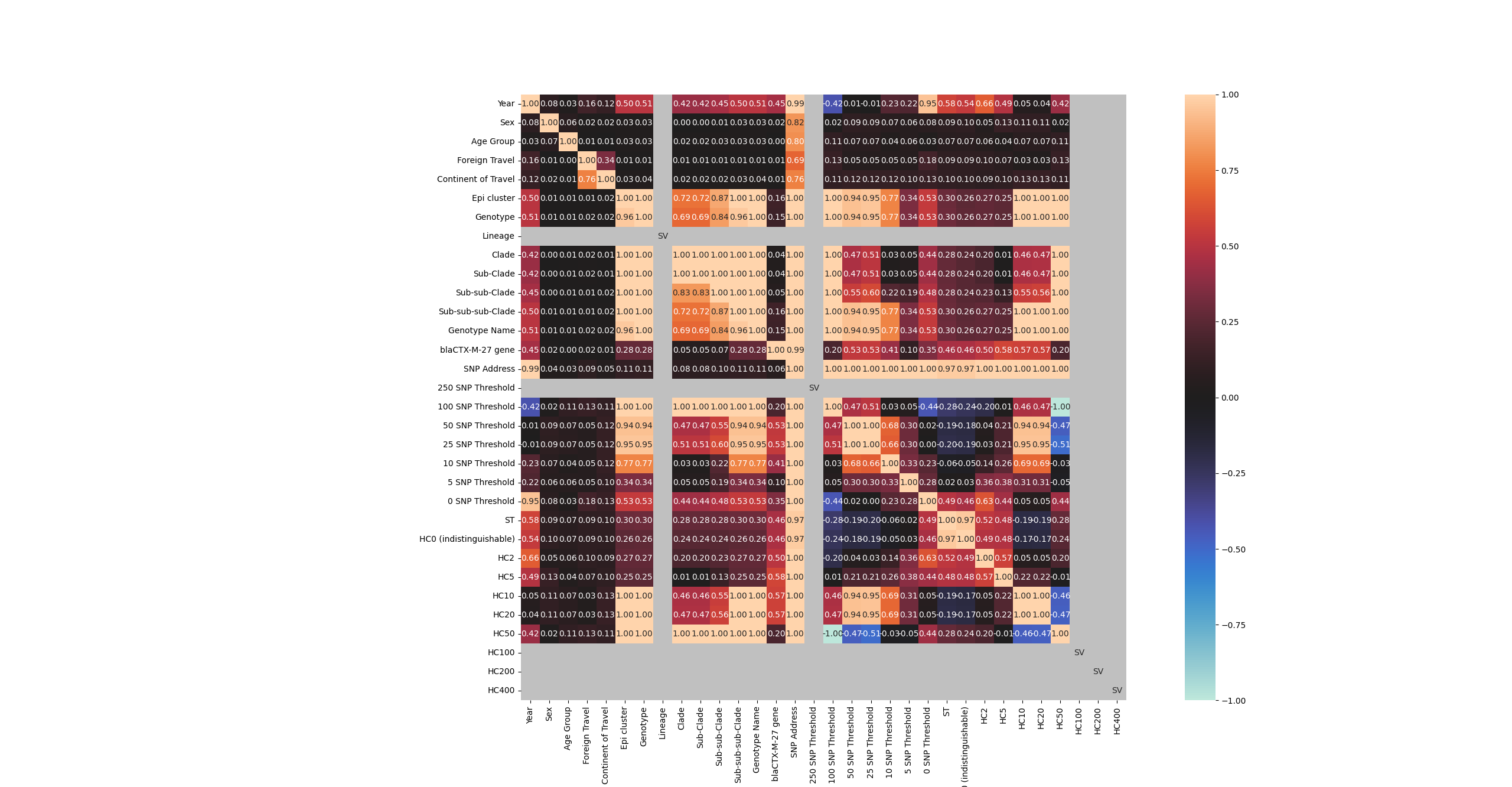


Figure . MSM dataset Theil’s U comparison of all variables

# Discussion

# References