Clustering of Swiss HIV Patients Not Enrolled in the Swiss HIV Cohort Study (SHCS)

Mohaned Shilaih[1,6], Alex Marzel[1,6], Wan Lin Yang[1], Alexandra U Scherrer[1], Jörg Schüpbach[2], Jürg Böni[2], Sabine Yerly[3], Thomas Klimkait[4], Vincent Aubert[5], Matthias Cavassini[7], Hans H. Hirsch[8], Pietro L. Vernazza[9], Enos Bernasconi[10], Hansjakob Furrer[11], Huldrych F Günthard[1], Roger Kouyos[1,6], and the Swiss HIV Cohort Study

1- Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, 2- Swiss National Center for Retroviruses, Institute of Medical Virology, University of Zurich, 3- Laboratory of Virology, Geneva University Hospital, 4- Department of Biomedicine–Petersplatz, University of Basel, 5- Division of Immunology and Allergy, University Hospital Lausanne, 6-Institute of Medical Virology, Swiss National Center for Retroviruses, University of Zurich, Zurich, Switzerland, 7- Division of Immunology, University Hospital Lausanne, 8- Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, 9- Division of Infectious Diseases, Cantonal Hospital St. Gallen, 10- Division of Infectious Diseases, Regional Hospital Lugano, 11- Department of Infectious Diseases, Bern University Hospital and University of Bern.

Questions

- Could one utilize pathogens genomic data to extend the scope of classical infectious diseases epidemiological data?
- How good is the coverage of the SHCS and are their marginalized populations that fall off the health care radar?

Data

- A unique database of HIV genotypic resistance tests in both SHCS enrolled and non-SHCS patients allows for the assessment of the presence of marginalized HIV-infected populations in Switzerland and test the utility of demographics inference based on pathogens sequence data

- 11130 SHCS and 2875 Swiss non-SHCS sequences
- 27803 non Swiss background sequences from the Los Alamos database

Methods

- A maximum likelihood phylogenetic tree was built using all 14005 Swiss sequences and the background sequences

- As non-SHCS patients are lacking demographics (transmission group and ethnicity), that information was inferred using phylogenetic proximity (Figure 1)

- Demographics inference predictive power was validated on a subset of the SHCS after removing their demographics

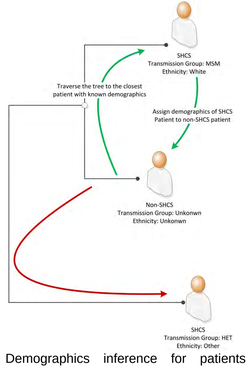

Figure 1. Demographics inference for patients lacking such information. Starting from a given non-SHCS tip the tree was traversed up to the nearest SHCS-sequence, where the corresponding SHCS-patient demographic was assigned to the non-SHCS patient.

Results

- Demographics inference proved to be accurate (up to 84%) (Table 1) yet dependent on the demographic feature inferred, the genetic distance on the phylogeny, and the sample date.

Direct Agreement of True Versus Predicted Demographics	Random SHCS Sample (2875)	Late Enrollers (1116)
Risk Group	69%	66%
Risk Group - sampling date > 2003	73%	68%
Risk Group - Genetic distance <1.5%	76%	72%
Risk Group - Genetic distance <1.5% and sample date >2003	81%	78%
Ethnicity	81%	82%
Ethnicity - sampling date > 2003	77%	79%
Ethnicity - Genetic distance <1.5%	84%	81%
Ethnicity - Genetic distance <1.5% and sample date >2003	80%	79%

Non-SHCS patients demographics differed from the SHCS'. More precisely the following groups were overrepresented:

- Non-B subtype (odds ratio 3.0; 95%C.I 2.8-3.3) – Figure 2
- Female gender (OR 1.4; 95%C.I 1.3-1.6)
- Non-white ethnicity (*inferred)* (OR 2.8; 95%C.I 2.5-3.1)
- Heterosexual transmission group (*inferred)* (OR 1.7; 95%C.I 1.5-1.8)

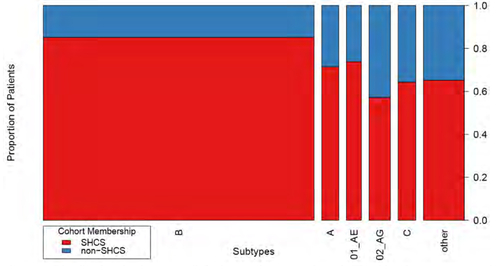

Figure 2. Subtypes Distribution in the overall Swiss patients analyzed. The X-axis represents the proportion of the patients, while the breadth of the column reflects the raw number of Swiss patients per subtype.

- We found 344 purely non-SHCS Swiss transmission clusters. However, these outbreaks were small (Median 2, IQR 2-2, Maximum 7 patients, Figure 3).

- Non-SHCS and SHCS outbreaks overlap strongly, as 53% and 70% of SHCS and non-SHCS sequences, respectively, were parts of mixed clusters.

- Non-SHCS patients showed the same clustering preference as the SHCS 0.98 (0.88-1.09)

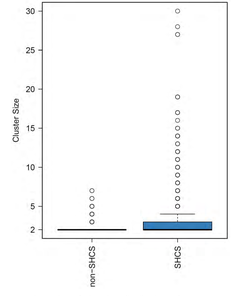

Figure 3. A box plot of the size distribution of pure SHCS clusters (transmission clusters consisting only of SHCS sequences) and pure non-SHCS clusters SHCS (transmission clusters consisting only of Swiss non-SHCS sequences).

Conclusions

- A novel method for using sequence data to fill in the gaps arising in classical epidemiological is demonstrated in the framework of the SHCS and non-SHCS Swiss patients.
- As sequence data become more available, this method could be utilized to evaluate the extent epidemiological data captures the propagation of infectious diseases and to better target care to marginalized populations.
- Even in the remarkably comprehensive SHCS, marginalized populations were underrepresented, which confirms the previously reported reluctance of said populations to participate in any form of surveillance maybe due to the lack of reach or fearing persecution.
- No large outbreaks were observed that the SHCS did not capture, thus ascertaining the SHCS exceptional coverage.