Supplementary Material

Similar but different: Integrated phylogenetic analysis of Austrian and Swiss HIV-1 sequences reveal differences in transmission patterns of the local HIV-1 epidemics

Kusejko et al.

Contents

S1 Statistical Analysis

We first determined the distribution of the traits of interest among SHCS and AHIVCOS patients in the tree, denoted by

$$d_{SHCS} = (X_1, \ldots, X_n) \text{ and } d_{AHIVCOS} = (Y_1, \ldots, Y_n),$$

with n the number of levels of the trait of interest, e.g., $n = 4$ in the case of the four risk groups MSM, IDU, male HET and female HET, X_i and Y_i the fractions of the i-th level in the SHCS and AHIVCOS, respectively. We then determined whether the distribution of traits is more common in cherries than would be expected by these distributions of traits in the cohort.

1. Domestic cherries: First, we calculated the distribution of a trait when randomly pairing patients of the same cohort, denoted by $e_{i,j}^{SHCS}$ and $e_{i,j}^{AHIVCOS}$ for patients of the i-th and j-th level of the trait in a cherry:

$$e_{i,j}^{SHCS} = 2 * X_i * X_j \text{ for } i \neq j, \text{ and } X_i^2 \text{ for } i = j; i, j = 1, \ldots, n$$

$$e_{i,j}^{AHIVCOS} = 2 * Y_i * Y_j \text{ for } i \neq j, \text{ and } Y_i^2 \text{ for } i = j, i, j = 1, \ldots, n.$$

   Note: The factor 2 indicates that the pair is not ordered, e.g., MSM/IDU-cherry is the same as IDU/MSM-cherry.

   The observed distribution of traits in the domestic cherries is denoted by $d_{SHCS/SHCS} = (x_{1,1}, x_{1,2}, \ldots, x_{n,n})$ and $d_{AHIVCOS/AHIVCOS} = (y_{1,1}, y_{1,2}, \ldots, y_{n,n})$. The ratio of the observed and expected distribution of traits in the SHCS/SHCS-cherries and AHIVOCS/AHIVCOS-cherries, i.e.,

$$AF_{i,j}^{SHCS} = \frac{x_{i,j}}{e_{i,j}^{SHCS}}$$

$$AF_{i,j}^{AHIVCOS} = \frac{y_{i,j}}{e_{i,j}^{AHVICOS}}$$

   was used to assess the assortativenes of the traits in the cherries and termed assortative factor (AF).

2. International cherries: We compared the distributions of traits in SHCS/LA-cherries ($d_{SHCS/LA} = (x_1, \ldots, x_n)$) and AHIVOCS/LA-cherries ($d_{AHIVCOS/LA} = (y_1, \ldots, y_n)$) with the distribution of traits in the cohorts by calculating the ratios

$$r_i^{SHCS/LA} = \frac{x_i}{X_i} \text{ and } r_i^{AHIVCOSS/LA} = \frac{y_i}{Y_i}.$$

   The i-th level of a trait is more common in the SHCS or AHIVCOS than expected if $r_i^{SHCS/LA} > 1$ or $r_i^{AHIVCOS/LA} > 1$, respectively.

3. AHIVCOS/SHCS-cherries: The distribution of a trait when randomly pairing patients of the SHCS and AHIVCOS, denoted by $e_{i,j}^{SHCS/AHIVCOS}$ for SHCS patients of the i-th and AHIVCOS patients of the j-th level of the trait in a cherry, was calculated by:

$$e_{i,j}^{SHCS/AHIVCOS} = X_i * Y_j; i, j = 1, \ldots, n.$$

   The observed distribution of traits in the SHCS/AHIVCOS-cherries is denoted by $d_{SHCS/AHIVCOS} = (z_{1,1}, z_{1,2}, \ldots, z_{n,n})$. The ratio of the expected and observed distribution of traits in the SHCS/AHIVCOS-cherries was then calculated as

$$r_{i,j}^{SHCS/AHIVCOS} = \frac{z_{i,j}}{e_{i,j}^{SHCS/AHIVCOS}}.$$

S2 Patients in the cohorts and the phylogenetic tree

S2.1 SHCS

In the SHCS, we could include sequences of 12902/20802(62%) patients in the phylogenetic tree. The characteristics of the patients in the phylogenetic tree and the whole cohort are as follows:

Table S1: Characteristics of SHCS patients in the phylogeny and the whole cohort

		SHCS phylogeny	Whole SHCS
Total		12902	20802
Cohort Center	Zürich	4'869 (37.7%)	7'660 (36.8%)
	Basel	1'427 (11.1%)	2'213 (10.6%)
	Bern	1'712 (13.3%)	2'688 (12.9%)
	Geneva	1'762 (13.7%)	3'070 (14.8%)
	Lausanne	1'911 (14.8%)	3'298 (15.9%)
	Lugano	388 (3.0%)	636 (3.1%)
	St Gallen	833 (6.5%)	1'237 (5.9%)
Sex	male	9'272 (71.9%)	15'114 (72.7%)
	female	3'630 (28.1%)	5'688 (27.3%)
Birth year	median (IQR)	1965 [1959, 1972]	1963 [1957, 1971]
Registration year	median (IQR)	2001 [1996, 2009]	1999 [1991, 2008]
Risk group	MSM	5'168 (40.1%)	8'138 (39.1%)
	male HET	2'133 (16.5%)	3'241 (15.6%)
	female HET	2'491 (19.3%)	3'633 (17.5%)
	male IDU	1'643 (12.7%)	3'186 (15.3%)
	female IDU	887 (6.9%)	1'656 (8.0%)
	male other	357 (2.8%)	595 (2.9%)
	female other	223 (1.7%)	353 (1.7%)
Ethnicity	white	9'881 (76.6%)	14'099 (67.8%)
	black	1'638 (12.7%)	2'267 (10.9%)
	Hispanic	397 (3.1%)	558 (2.7%)
	Asian	425 (3.3%)	591 (2.8%)

S2.2 AHIVCOS

In the AHIVCOS, squences of 3141/9793(32.1%) patients could be included in the phylogenetic tree. The characteristics of the patients in the tree and the whole cohort are as follows:

Table S2: Characteristics of AHIVCOS patients in the phylogeny and the whole cohort

		AHIVCOS phylogeny	Whole AHIVCOS
Total		3141	9793
Cohort Center	Vienna	1'837 (58.5%)	5'845 (59.7%)
	Linz	487 (15.5%)	1'089 (11.1%)
	Salzburg	228 (7.3%)	458 (4.7%)
	Innsbruck	235 (7.5%)	1'341 (13.7%)
	Graz	354 (11.3%)	718 (7.3%)
	Klagenfurt	0 (0.0%)	264 (2.7%)
	Feldkirch	0 (0.0%)	78 (0.8%)
Sex	male	2'375 (75.6%)	7'490 (76.5%)
	female	766 (24.4%)	2'303 (23.5%)
Birth year	median (IQR)	1972 [1964, 1981]	1968 [1960, 1978]
Registration year	median (IQR)	2009 [2003, 2013]	2005 [1996, 2012]
Risk group	MSM	1'361 (43.3%)	3'809 (38.9%)
	male HET	562 (17.9%)	1'543 (15.8%)
	female HET	559 (17.8%)	1'503 (15.3%)
	male IDU	335 (10.7%)	1'446 (14.8%)
	female IDU	158 (5.0%)	606 (6.2%)
	male other	117 (3.7%)	692 (7.1%)
	female other	49 (1.6%)	194 (2.0%)
Ethnicity	white	2'541 (80.9%)	6'034 (61.6%)
	black	303 (9.6%)	736 (7.5%)
	Hispanic	24 (0.8%)	71 (0.7%)
	Asian	80 (2.5%)	194 (2.0%)

S3 Distance of different cherry types

The following figures shows the distributions of distances among the different cherry types in domestic and international cherries.

We can see that in the SHCS, domestic cherries with both patients belonging to the transmission group MSM have the smallest median distance. Regarding the ethnicity of patients, in both cohorts, cherries with one patient being of white ethnicity and the other patient being of Hispanic ethnicity have the smallest distance threshold (Figure S1).

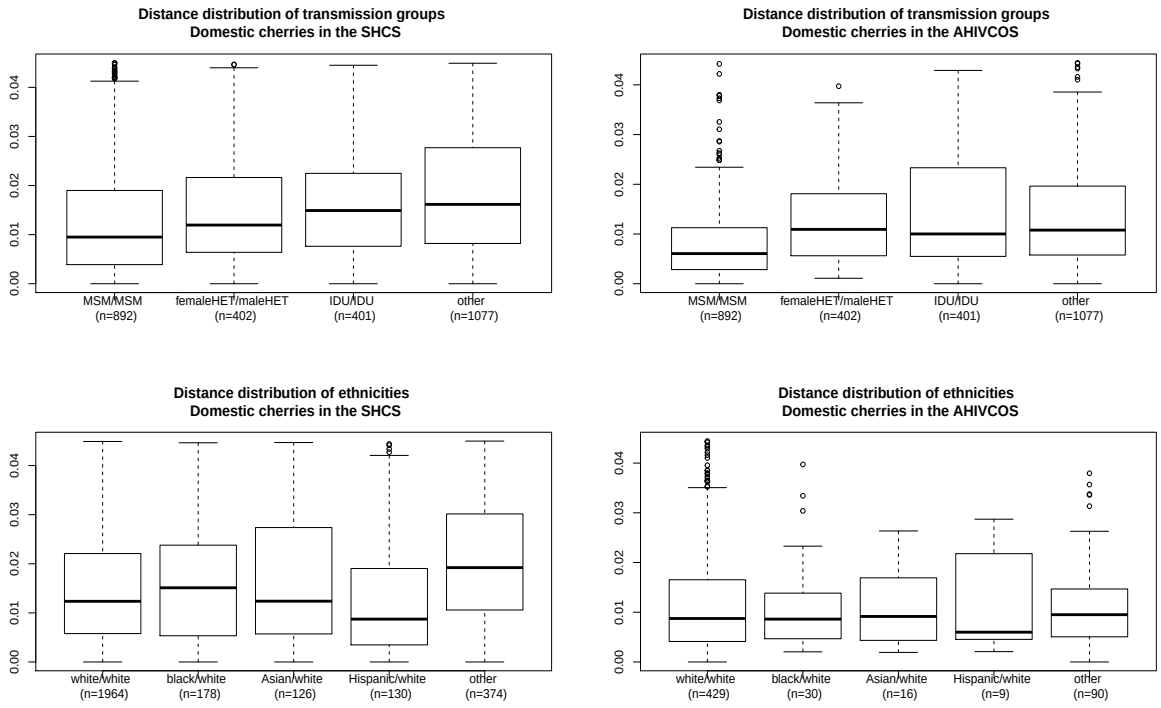


Figure S1: The phylogenetic distance distribution in the different types of domestic cherries.

In the case of international cherries, in both cohorts MSM have the smallest distance, in the SHCS also IDU have a smaller distance as compared to heterosexuals. Concerning ethnicities, black ethnicities show a higher distance threshold in both cohorts, see Figure S2.

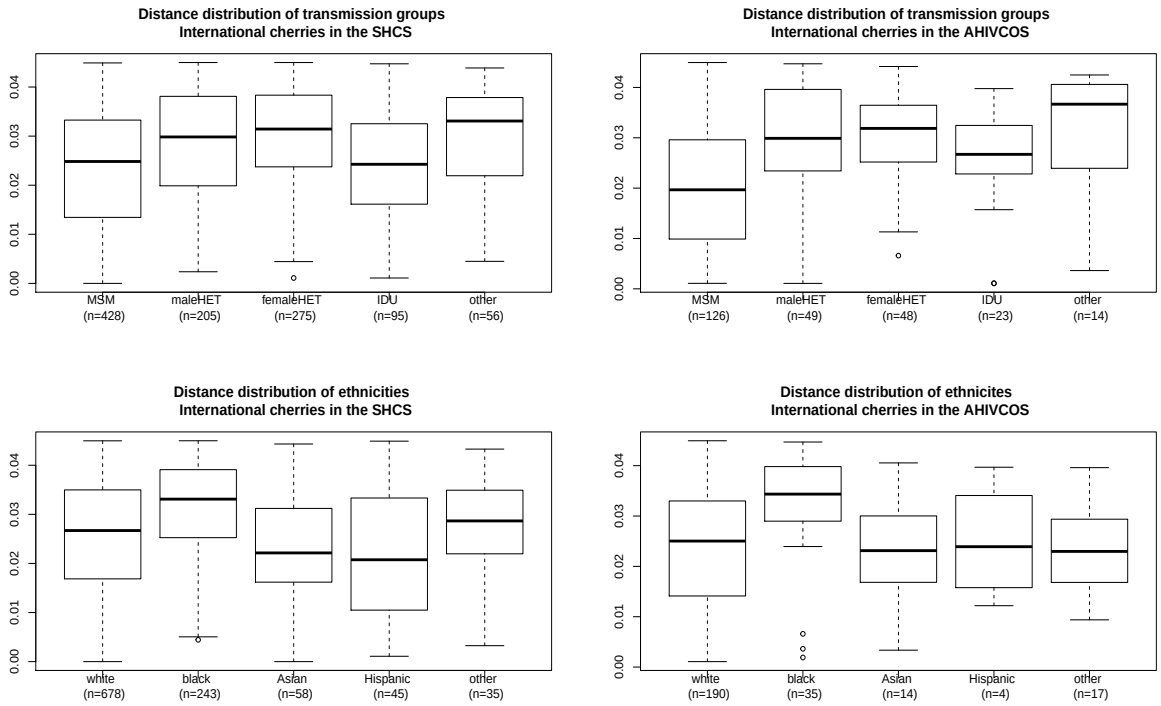


Figure S2: The phylogenetic distance distribution in the different types of international cherries.

S4 Number of cherries

We included 3141 AHIVCOS and 12902 SHCS patients in the combined phylogeny. The SHCS has a higher sample density as compared to the AHIVCOS and as a result, we assume to obtain more SHCS/SHCS-cherries on the tree. We would like to understand the effect of these differences in sample densities by down-sampling the SHCS.

For each chosen down-sampling fraction x, we randomly dropped a fraction of $1 - x$ SHCS tips in the tree. We analyzed the number of cherries and traits on these pruned trees. We repeated this process 100 times for each down-sampling fraction $0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2$ and averaged the results.

S4.1 SHCS down-sampling fraction and number of domestic cherries

The following table shows the number of different cherry types obtained from the trees with different SHCS down-sampling fractions. The fraction of 1 includes all SHCS sequences. For lower fractions, the numbers refer to the median and interquartile range of the 100 pruned trees of a given down-sampling fraction. The cophenetic distance cut-off is 0.045.

Table S3: Number and percentage of AHIVCOS/AHIVCOS-cherries by SHCS-downsampling fraction and distance cut-off

	0.045	0.04	0.035	0.03	0.025	0.02	0.015
1	574, 36.5%	565, 36%	548, 34.9%	527, 33.6%	501, 31.9%	471, 30%	410, 26.1%
0.9	576 [575,577],36.7%	567 [566,568],36.1%	550 [549,550],35%	528 [528,529],33.6%	502 [501,503],32%	472 [471,472],30.1%	410 [410,411],26.1%
0.8	578 [577,580],36.8%	569 [567,570],36.2%	552 [550,552],35.1%	530 [529,531],33.7%	503 [502,504],32%	472 [472,473],30.1%	411 [410,412],26.2%
0.7	581 [580,583],37%	571 [570,573],36.4%	554 [553,555],35.3%	532 [531,533],33.9%	504 [504,505],32.1%	473 [472,474],30.1%	412 [411,412],26.2%
0.6	584 [582,586],37.2%	573 [572,575],36.5%	556 [554,557],35.4%	534 [532,535],34%	506 [505,507],32.2%	474 [473,475],30.2%	412 [411,413],26.2%
0.5	588 [586,590],37.4%	577 [575,579],36.7%	559 [557,561],35.6%	537 [535,538],34.2%	507 [506,509],32.3%	475 [474,477],30.2%	413 [413,414],26.3%
0.4	592 [590,594],37.7%	580 [579,582],36.9%	562 [561,564],35.8%	539 [538,541],34.3%	510 [508,511],32.5%	477 [475,478],30.4%	414 [413,416],26.4%
0.3	596 [593,599],37.9%	584 [582,586],37.2%	565 [564,567],36%	542 [541,544],34.5%	511 [510,513],32.5%	478 [477,479],30.4%	415 [414,416],26.4%
0.2	602 [600,604],38.3%	589 [587,590],37.5%	570 [568,571],36.3%	546 [544,548],34.8%	514 [513,515],32.7%	480 [479,481],30.6%	417 [416,418],26.6%

Table S4: Number and percentage of SHCS/SHCS-cherries by SHCS-downsampling fraction and distance cut-off

	0.045	0.04	0.035	0.03	0.025	0.02	0.015
1	2772, 43%	2659, 41.2%	2528, 39.2%	2357, 36.5%	2155, 33.4%	1887, 29.3%	1535, 23.8%
0.9	2445 [2431,2453],42.1%	2342 [2332,2350],40.3%	2224 [2212,2233],38.3%	2066 [2056,2076],35.6%	1882 [1875,1890],32.4%	1639 [1632,1649],28.2%	1327 [1315,1332],22.9%
0.8	2118 [2109,2129],41%	2030 [2017,2039],39.3%	1924 [1911,1934],37.3%	1784 [1770,1795],34.6%	1617 [1606,1626],31.3%	1397 [1388,1409],27.1%	1122 [1112,1137],21.7%
0.7	1798 [1783,1808],39.8%	1717 [1702,1727],38%	1622 [1610,1632],35.9%	1497 [1484,1508],33.2%	1351 [1338,1362],29.9%	1165 [1152,1173],25.8%	925 [916,936],20.5%
0.6	1481 [1471,1495],38.3%	1414 [1400,1426],36.5%	1333 [1322,1344],34.4%	1228 [1214,1240],31.7%	1099 [1088,1114],28.4%	940 [931,952],24.3%	745 [732,754],19.2%
0.5	1174 [1159,1184],36.4%	1119 [1104,1128],34.7%	1050 [1039,1064],32.6%	967 [954,978],30%	859 [848,868],26.6%	727 [718,738],22.5%	569 [555,577],17.6%
0.4	876 [867,889],33.9%	834 [826,844],32.3%	780 [770,792],30.2%	711 [700,720],27.6%	626 [618,638],24.3%	526 [519,536],20.4%	404 [396,415],15.7%
0.3	602 [592,612],31.1%	568 [559,580],29.3%	529 [517,539],27.3%	481 [468,490],24.9%	417 [408,430],21.5%	346 [335,357],17.9%	260 [253,268],13.4%
0.2	345 [336,352],26.7%	324 [316,332],25.1%	298 [293,308],23.1%	268 [262,275],20.8%	231 [224,238],17.9%	187 [180,194],14.5%	137 [130,143],10.6%

We see that for a down-sampling fraction of around 0.5, we obtain about the same number of AHIVCOS/AHIVCOS- and SHCS/SHCS-cherries.

S4.2 SHCS down-sampling fraction and number of Los Alamos cherries

Similar to the above analysis, we look at the number of AHIVCOS/Los Alamos- and SHCS/Los Alamos-cherries, dependent on the SHCS down-sampling fraction as well as the distance cut-off, ranging from 0.015 to 0.045, for inclusion of the cherries.

Table S5: Number and percentage of AHIVCOS/LA-cherries by SHCS-downsampling fraction and distance cut-off

	0.045	0.04	0.035	0.03	0.025	0.02	0.015
1	260, 8.3%	233, 7.4%	199, 6.3%	163, 5.2%	117, 3.7%	89, 2.8%	61, 1.9%
0.9	264 [262,265],8.4%	236 [235,237],7.5%	201 [200,203],6.4%	165 [164,166],5.3%	118 [117,119],3.8%	90 [89,90],2.9%	61 [61,62],1.9%
0.8	268 [267,270],8.5%	240 [239,241],7.6%	205 [203,206],6.5%	167 [166,168],5.3%	120 [119,121],3.8%	91 [90,92],2.9%	62 [61,63],2%
0.7	272 [271,274],8.7%	243 [242,245],7.7%	207 [206,209],6.6%	169 [167,170],5.4%	121 [120,122],3.9%	92 [90,93],2.9%	63 [62,63],2%
0.6	277 [275,280],8.8%	247 [246,250],7.9%	211 [209,213],6.7%	171 [170,173],5.4%	123 [121,124],3.9%	92 [92,94],2.9%	63 [62,64],2%
0.5	283 [281,286],9%	253 [251,255],8.1%	215 [213,218],6.8%	174 [173,176],5.5%	125 [124,126],4%	94 [93,95],3%	64 [63,65],2%
0.4	289 [287,291],9.2%	257 [255,259],8.2%	218 [216,220],6.9%	176 [175,178],5.6%	127 [125,128],4%	95 [94,96],3%	65 [64,65],2.1%
0.3	298 [295,299],9.5%	264 [262,266],8.4%	224 [222,225],7.1%	180 [179,181],5.7%	129 [128,130],4.1%	97 [96,98],3.1%	65 [64,66],2.1%
0.2	306 [304,309],9.7%	271 [269,273],8.6%	230 [228,232],7.3%	184 [183,186],5.9%	132 [131,133],4.2%	98 [97,99],3.1%	66 [66,67],2.1%

Table S6: Number and percentage of SHCS/LA-cherries by SHCS-downsampling fraction and distance cut-off

1	1061, 8.2%	912, 7.1%	766, 5.9%	606, 4.7%	432, 3.3%	315, 2.4%	190, 1.5%
0.9	1003 [997,1012],8.6%	861 [855,869],7.4%	721 [715,728],6.2%	568 [563,576],4.9%	405 [401,411],3.5%	294 [291,298],2.5%	177 [174,179],1.5%
0.8	941 [930,950],9.1%	806 [796,814],7.8%	673 [662,683],6.5%	530 [521,536],5.1%	377 [371,384],3.7%	272 [266,277],2.6%	161 [157,166],1.6%
0.7	877 [864,888],9.7%	748 [738,758],8.3%	621 [615,632],6.9%	489 [483,496],5.4%	349 [344,354],3.9%	251 [247,256],2.8%	148 [144,152],1.6%
0.6	795 [786,806],10.3%	679 [673,687],8.8%	566 [558,573],7.3%	443 [436,450],5.7%	316 [310,324],4.1%	226 [219,231],2.9%	132 [129,136],1.7%
0.5	717 [701,728],11.1%	607 [594,617],9.4%	504 [492,513],7.8%	394 [384,402],6.1%	281 [275,288],4.4%	199 [192,206],3.1%	116 [108,120],1.8%
0.4	623 [613,640],12.1%	529 [520,544],10.3%	440 [427,449],8.5%	344 [333,353],6.7%	248 [236,252],4.8%	173 [166,178],3.4%	100 [95,104],1.9%
0.3	515 [503,526],13.3%	435 [424,446],11.2%	358 [350,368],9.2%	279 [272,287],7.2%	198 [191,207],5.1%	140 [133,145],3.6%	79 [74,84],2%
0.2	385 [378,398],14.9%	325 [316,336],12.6%	268 [259,276],10.4%	210 [200,215],8.1%	150 [142,155],5.8%	102 [96,107],4%	57 [53,62],2.2%

S4.3 SHCS down-sampling fraction and number of SHCS/AHIVCOS-cherries

Now, we look at the number of AHIVCOS/SHCS-cherries, dependent on the SHCS down-sampling fraction as well as the distance cut-off, ranging from 0.015 to 0.045, for inclusion of the cherries.

Table S7: Percentage of AHIVCOS patients in AHIVCOS/SHCS-cherries - by SHCS-downsampling fraction and distance cut-off

	0.045	0.04	0.035	0.03	0.025	0.02	0.015
1	220, 7%	193, 6.1%	161, 5.1%	125, 4%	98, 3.1%	78, 2.5%	46, 1.5%
0.9	210 [206,212],6.7%	184 [181,186],5.9%	153 [150,155],4.9%	118 [116,121],3.8%	92 [90,94],2.9%	74 [72,75],2.4%	44 [42,45],1.4%
0.8	198 [194,203],6.3%	172 [168,177],5.5%	143 [140,147],4.6%	111 [106,113],3.5%	86 [82,89],2.7%	68 [66,71],2.2%	41 [38,43],1.3%
0.7	183 [178,188],5.8%	160 [156,164],5.1%	133 [129,137],4.2%	102 [97,106],3.2%	79 [76,83],2.5%	63 [60,65],2%	38 [36,40],1.2%
0.6	171 [166,175],5.4%	149 [144,153],4.7%	123 [118,128],3.9%	94 [90,98],3%	72 [69,75],2.3%	57 [54,60],1.8%	34 [32,36],1.1%
0.5	154 [146,159],4.9%	133 [126,138],4.2%	110 [102,114],3.5%	83 [78,88],2.6%	65 [60,68],2.1%	50 [46,54],1.6%	30 [28,33],1%
0.4	132 [127,138],4.2%	112 [110,118],3.6%	94 [91,98],3%	72 [68,75],2.3%	55 [52,58],1.8%	42 [39,45],1.3%	25 [23,28],0.8%
0.3	113 [105,116],3.6%	95 [88,102],3%	79 [73,84],2.5%	60 [56,66],1.9%	46 [42,50],1.5%	35 [32,38],1.1%	22 [19,23],0.7%
0.2	84 [80,89],2.7%	72 [68,77],2.3%	58 [56,64],1.8%	45 [42,49],1.4%	34 [31,37],1.1%	25 [23,28],0.8%	15 [13,17],0.5%

Table S8: Percentage of SHCS patients in AHIVCOS/SHCS-cherries - by SHCS-downsampling fraction and distance cut-off

	0.045	0.04	0.035	0.03	0.025	0.02	0.015
1	220, 1.7%	193, 1.5%	161, 1.2%	125, 1%	98, 0.8%	78, 0.6%	46, 0.4%
0.9	210 [206,212],1.8%	184 [181,186],1.6%	153 [150,155],1.3%	118 [116,121],1%	92 [90,94],0.8%	74 [72,75],0.6%	44 [42,45],0.4%
0.8	198 [194,203],1.9%	172 [168,177],1.7%	143 [140,147],1.4%	111 [106,113],1.1%	86 [82,89],0.8%	68 [66,71],0.7%	41 [38,43],0.4%
0.7	183 [178,188],2%	160 [156,164],1.8%	133 [129,137],1.5%	102 [97,106],1.1%	79 [76,83],0.9%	63 [60,65],0.7%	38 [36,40],0.4%
0.6	171 [166,175],2.2%	149 [144,153],1.9%	123 [118,128],1.6%	94 [90,98],1.2%	72 [69,75],0.9%	57 [54,60],0.7%	34 [32,36],0.4%
0.5	154 [146,159],2.4%	133 [126,138],2.1%	110 [102,114],1.7%	83 [78,88],1.3%	65 [60,68],1%	50 [46,54],0.8%	30 [28,33],0.5%
0.4	132 [127,138],2.6%	112 [110,118],2.2%	94 [91,98],1.8%	72 [68,75],1.4%	55 [52,58],1.1%	42 [39,45],0.8%	25 [23,28],0.5%
0.3	113 [105,116],2.9%	95 [88,102],2.5%	79 [73,84],2%	60 [56,66],1.6%	46 [42,50],1.2%	35 [32,38],0.9%	22 [19,23],0.6%
0.2	84 [80,89],3.3%	72 [68,77],2.8%	58 [56,64],2.2%	45 [42,49],1.7%	34 [31,37],1.3%	25 [23,28],1%	15 [13,17],0.6%

S5 Los Alamos countries

The following table shows the frequency of all countries in AHIVCOS/Los Alamos- and SHCS/Los Alamos-cherries.

Table S9: Los Alamos countries in AHIVCOS/Los Alamos- and SHCS/Los Alamos-cherries

Los Alamos countries	AHIVCOS/Los Alamos-cherries	SHCS/Los Alamos-cherries
United States	39 (10.6%)	189 (13.7%)
Spain	19 (5.2%)	107 (7.7%)
United Kingdom	27 (7.4%)	88 (6.4%)
Cameroon	9 (2.5%)	85 (6.1%)
Germany	43 (11.7%)	84 (6.1%)
Brazil	10 (2.7%)	78 (5.6%)
France	12 (3.3%)	51 (3.7%)
Thailand	8 (2.2%)	43 (3.1%)
Kenya	8 (2.2%)	39 (2.8%)
South Africa	10 (2.7%)	39 (2.8%)
Uganda	6 (1.6%)	34 (2.5%)
Senegal	4 (1.1%)	27 (2%)
Ethiopia	1 (0.3%)	25 (1.8%)
Italy	4 (1.1%)	25 (1.8%)
Russia Federation	10 (2.7%)	19 (1.4%)
Poland	12 (3.3%)	18 (1.3%)
China	7 (1.9%)	17 (1.2%)
Portugal	4 (1.1%)	17 (1.2%)
Sweden	3 (0.8%)	17 (1.2%)
Argentina	1 (0.3%)	15 (1.1%)
Burkina Faso	6 (1.6%)	15 (1.1%)
Cuba	3 (0.8%)	15 (1.1%)
Belgium	1 (0.3%)	14 (1%)
Congo, The Democratic Republic of	3 (0.8%)	14 (1%)
Japan	2 (0.5%)	13 (0.9%)
Botswana	2 (0.5%)	12 (0.9%)
Canada	1 (0.3%)	12 (0.9%)
Nigeria	8 (2.2%)	12 (0.9%)
Denmark	4 (1.1%)	11 (0.8%)
Tanzania, United Republic of	1 (0.3%)	10 (0.7%)
Australia	0 (0%)	9 (0.7%)
Burundi	3 (0.8%)	8 (0.6%)
Vietnam	3 (0.8%)	8 (0.6%)
Cyprus	3 (0.8%)	7 (0.5%)
Malaysia	2 (0.5%)	7 (0.5%)
Mali	5 (1.4%)	7 (0.5%)
Rwanda	1 (0.3%)	7 (0.5%)
Zambia	2 (0.5%)	7 (0.5%)
Mexico	5 (1.4%)	6 (0.4%)
Netherlands	2 (0.5%)	6 (0.4%)
Singapore	1 (0.3%)	6 (0.4%)
Togo	2 (0.5%)	6 (0.4%)
Korea, Republic of	0 (0%)	6 (0.4%)
Indonesia	2 (0.5%)	5 (0.4%)
Slovenia	7 (1.9%)	5 (0.4%)
Tunisia	1 (0.3%)	5 (0.4%)
Angola	0 (0%)	5 (0.4%)

French Guiana	0 (0%)	5 (0.4%)
Finland	1 (0.3%)	4 (0.3%)
India	3 (0.8%)	4 (0.3%)
Malawi	1 (0.3%)	4 (0.3%)
Peru	1 (0.3%)	4 (0.3%)
Dominican Republic	0 (0%)	4 (0.3%)
Venezuela	0 (0%)	4 (0.3%)
Romania	13 (3.5%)	3 (0.2%)
Central African Republic	0 (0%)	3 (0.2%)
Gabon	0 (0%)	3 (0.2%)
Ghana	0 (0%)	3 (0.2%)
Honduras	0 (0%)	3 (0.2%)
Israel	0 (0%)	3 (0.2%)
Norway	0 (0%)	3 (0.2%)
Philippines	0 (0%)	3 (0.2%)
Taiwan, Province of China	0 (0%)	3 (0.2%)
Belarus	4 (1.1%)	2 (0.1%)
Benin	1 (0.3%)	2 (0.1%)
Hong Kong	2 (0.5%)	2 (0.1%)
Luxembourg	1 (0.3%)	2 (0.1%)
Morocco	1 (0.3%)	2 (0.1%)
Sudan	1 (0.3%)	2 (0.1%)
Turkey	3 (0.8%)	2 (0.1%)
Algeria	0 (0%)	2 (0.1%)
Greece	0 (0%)	2 (0.1%)
Mozambique	0 (0%)	2 (0.1%)
Niger	0 (0%)	2 (0.1%)
Panama	0 (0%)	2 (0.1%)
Uruguay	0 (0%)	2 (0.1%)
Afghanistan	1 (0.3%)	1 (0.1%)
Armenia	1 (0.3%)	1 (0.1%)
Cote d'Ivoire	1 (0.3%)	1 (0.1%)
Czechia	1 (0.3%)	1 (0.1%)
Georgia	1 (0.3%)	1 (0.1%)
Guinea-Bissau	2 (0.5%)	1 (0.1%)
Iran, Islamic Republic of	3 (0.8%)	1 (0.1%)
Republic of Serbia	1 (0.3%)	1 (0.1%)
Ukraine	1 (0.3%)	1 (0.1%)
Uzbekistan	1 (0.3%)	1 (0.1%)
Yemen	1 (0.3%)	1 (0.1%)
Zimbabwe	1 (0.3%)	1 (0.1%)
Azerbaijan	0 (0%)	1 (0.1%)
Barbados	0 (0%)	1 (0.1%)
Bulgaria	0 (0%)	1 (0.1%)
Cape Verde	0 (0%)	1 (0.1%)
Chile	0 (0%)	1 (0.1%)
Djibouti	0 (0%)	1 (0.1%)
Greenland	0 (0%)	1 (0.1%)
Guinea	0 (0%)	1 (0.1%)
Hungary	0 (0%)	1 (0.1%)
Jamaica	0 (0%)	1 (0.1%)
Seychelles	0 (0%)	1 (0.1%)
Cambodia	1 (0.3%)	0 (0%)
Kuwait	1 (0.3%)	0 (0%)

Nepal	1 (0.3%)	0 (0%)
Pakistan	1 (0.3%)	0 (0%)
Slovakia	2 (0.5%)	0 (0%)
Tajikistan	1 (0.3%)	0 (0%)
other	6 (1.6%)	23 (1.7%)

S6 Sensitivity Analysis: Sample density and distance threshold

S6.1 Characteristics of domestic cherries

We investigate the impact of the genetic distance threshold as well as the SHCS sample density on the assortativity factor (AF) of certain traits of the cherries. In particular, we analyze the AF of IDU/IDU-cherries, MSM/MSM-cherries and white/white-cherries. We observe that downsampling the SHCS does not change the AF of the AHIVCOS much, indicating a minor role of SHCS/AHIVCOS-cherries.

Interestingly, we observe that decreasing the genetic distance cut-off can lead to two different effects: In the case of IDU/IDU-cherries, Figure S3, the AF increases by increasing distance cut-off, while the AF decreases by increasing distance cut-off in the case of MSM/MSM-cherries, Figure S4.

The AF in the case of white/white-cherries is very stable in both cohorts, see Figure S6.

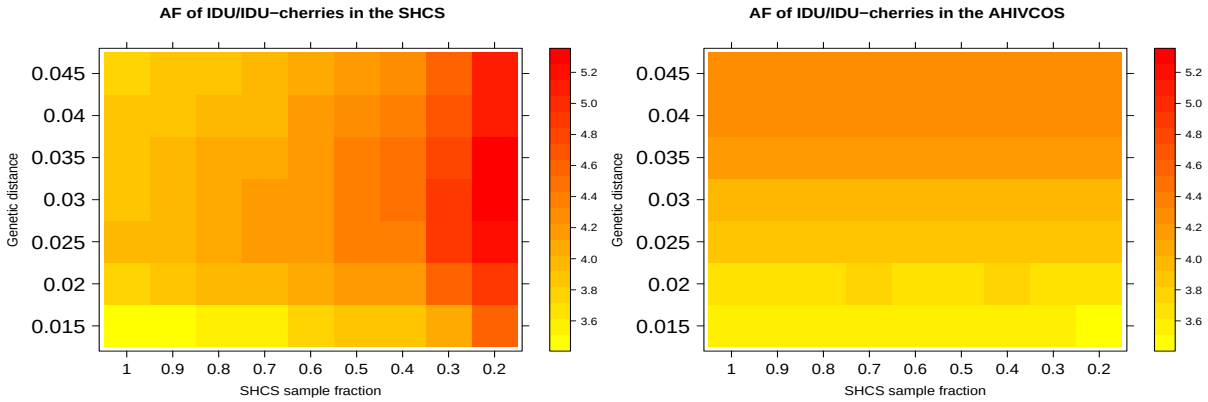


Figure S3: Assortativity factor (AF) of IDU/IDU-cherries in the SHCS and AHIVCOS by varying the genetic distance threshold and the sample density of the SHCS.

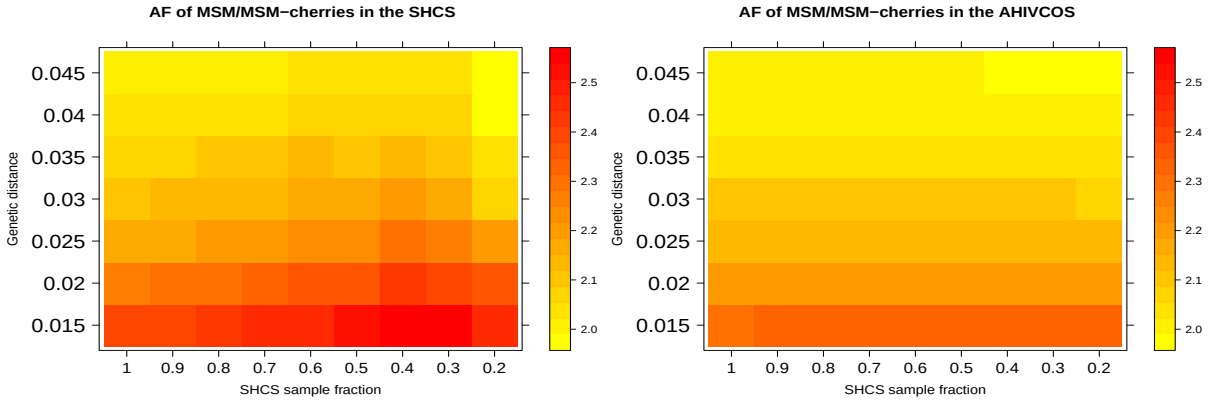


Figure S4: Assortativity factor (AF) of MSM/MSM-cherries in the SHCS and AHIVCOS by varying the genetic distance threshold and the sample density of the SHCS.

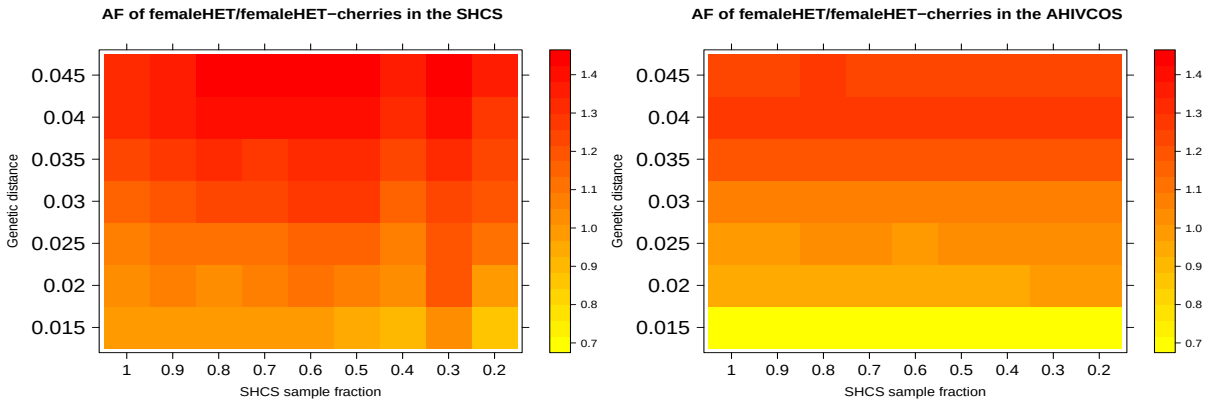


Figure S5: Assortativity factor (AF) of femaleHET/femaleHET-cherries in the SHCS and AHIVCOS by varying the genetic distance threshold and the sample density of the SHCS.

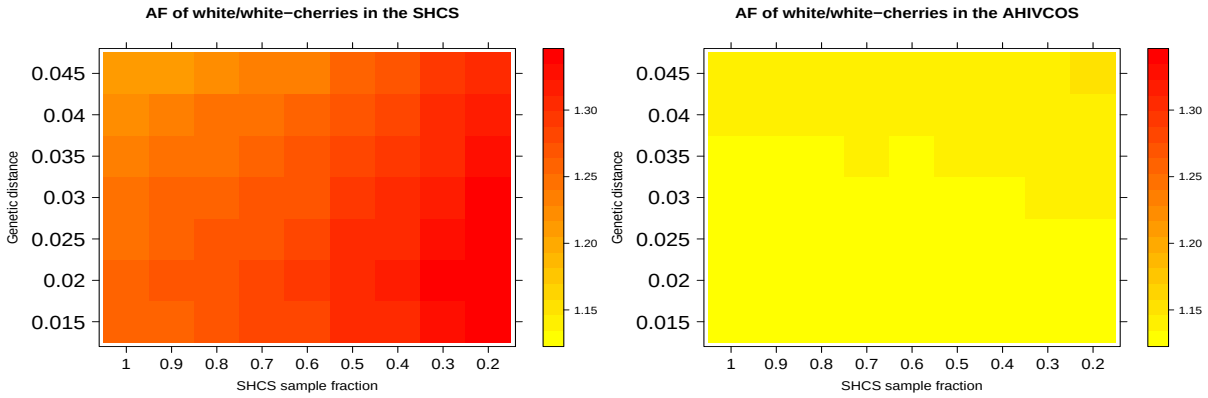


Figure S6: Assortativity factor (AF) of white/white-cherries in the SHCS and AHIVCOS by varying the genetic distance threshold and the sample density of the SHCS.

S6.2 Characteristics of international cherries

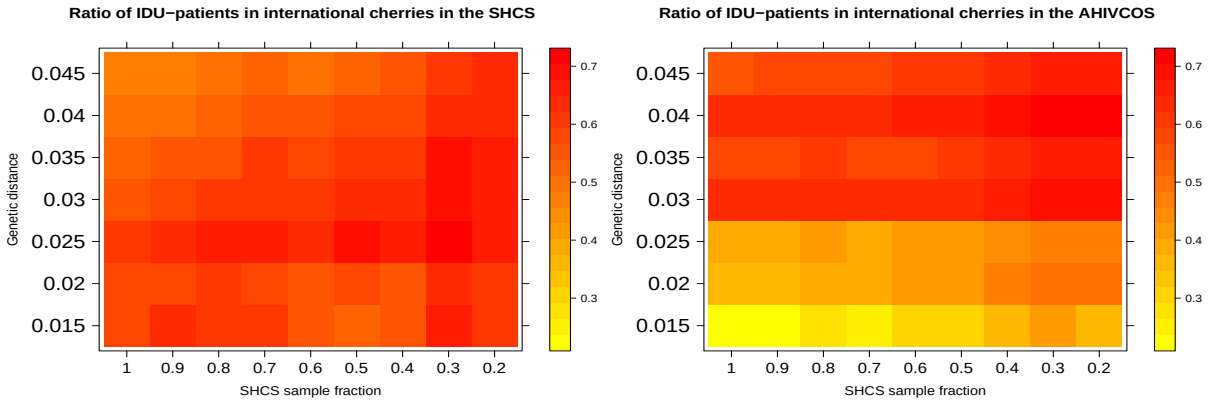


Figure S7: Ratio of IDU-patients in international cherries in the SHCS and AHIVCOS by varying the genetic distance threshold and the sample density of the SHCS.

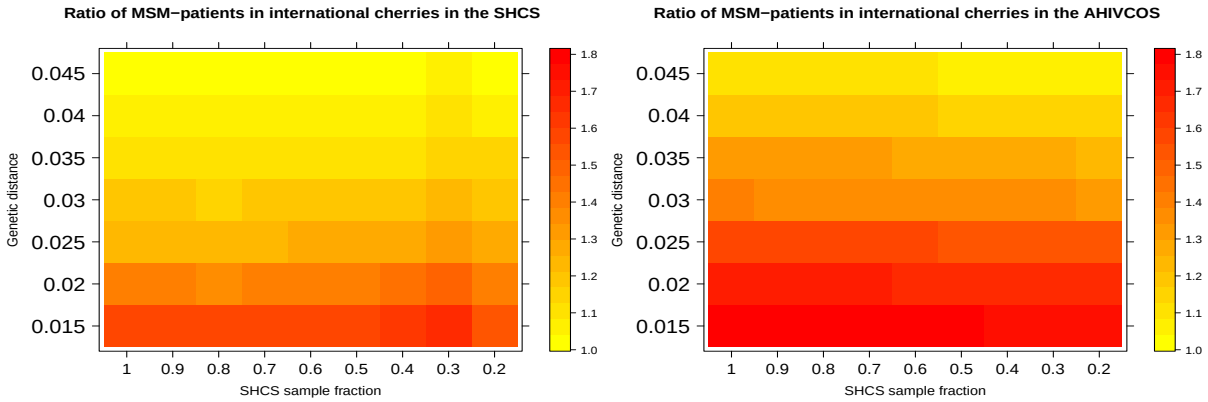


Figure S8: Ratio of MSM-patients in international cherries in the SHCS and AHIVCOS by varying the genetic distance threshold and the sample density of the SHCS.

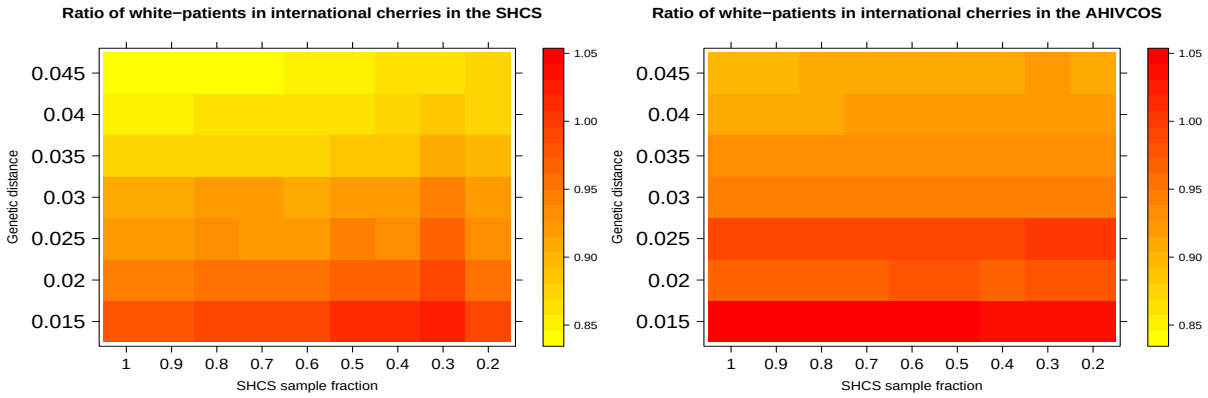


Figure S9: Ratio of white patients in international cherries in the SHCS and AHIVCOS by varying the genetic distance threshold and the sample density of the SHCS.

S7 Sensitivity Analysis: Phylogenetic reconstruction

S7.1 Restriction to subtype B

Building a phylogenetic tree with different subtypes combined can lead to different results as compared to building subtype-specific trees. To analyze the impact of including several subtypes in one phylogenetic tree, we performed a sensitivity analysis by restricting to subtype B and compared:

1. Tree 1: Removing all non-B sequences from the original tree
2. Tree 2: Building a new phylogenetic tree with only subtype B sequences included

In Tree 1, we obtain 3925 cherries, in Tree 2 we obtain 3950 cherries. The distribution of the type of cherries, as well as HIV transmission group and ethnicity distribution are very similar in both trees, suggesting a minor impact of the two approaches, i.e., removing non-B sequences from the original tree, or building a new tree with subtype B sequences only:

TREE 1	AT/AT	CH/CH	AT/CH	AT/LA	CH/LA
Total	404	2438	234	177	672
MSM/MSM	206	885	77		
IDU/IDU	33	404	21		
white/white	338	1922	189		

TREE 2	AT/AT	CH/CH	AT/CH	AT/LA	CH/LA
Total	419	2482	227	172	650
MSM/MSM	211	905	85		
IDU/IDU	38	406	23		
white/white	353	1974	166		

The direct overlap of cherries from Tree 1 and Tree 2 increases with decreasing distance threshold. The following table indicates the fraction of cherries identical in both trees:

	AT/AT	AT/CH	AT/LA	CH/CH	CH/LA	Total	Number
0.045	0.88	0.54	0.79	0.71	0.59	0.71	3472
0.04	0.88	0.58	0.84	0.73	0.59	0.72	3301
0.035	0.88	0.61	0.87	0.74	0.62	0.74	3075
0.03	0.91	0.67	0.88	0.76	0.67	0.77	2773
0.025	0.90	0.70	0.91	0.79	0.72	0.80	2473
0.02	0.91	0.81	0.92	0.82	0.79	0.83	2125
0.015	0.92	0.96	0.91	0.85	0.82	0.87	1703

S7.2 Swiss and Austrian phylogenies separately

Similar to the sensitivity analysis above, we wanted to understand the impact of including Austrian and Swiss sequences in the same phylogeny, or building separate phylogenies. For this, we take the following approach:

1. Cherry-Set 1: Extracting all AT/AT and CH/CH cherries from the original phylogeny including Swiss and Austrian sequences
2. Cherry-Set 2: Extracting AT/AT cherries from a phylogeny built from Austrian sequences (plus LA background sequences) and CH/CH cherries extracted from a phylogeny based on Swiss sequences (and LA background sequences)

Comparing the resulting total numbers of AT/AT and CH/CH cherries, as well as transmission group and ethnicity, reveals a high similarity of the two approaches.

	AT/AT (Cherry-Set 1)	CH/CH (Cherry-Set 1)	AT/AT (Cherry-Set 2)	CH/CH (Cherry-Set 2)
Total	620	3130	780	3341
MSM/MSM	221	959	256	1043
IDU/IDU	62	416	74	423
white/white	455	2128	564	2240

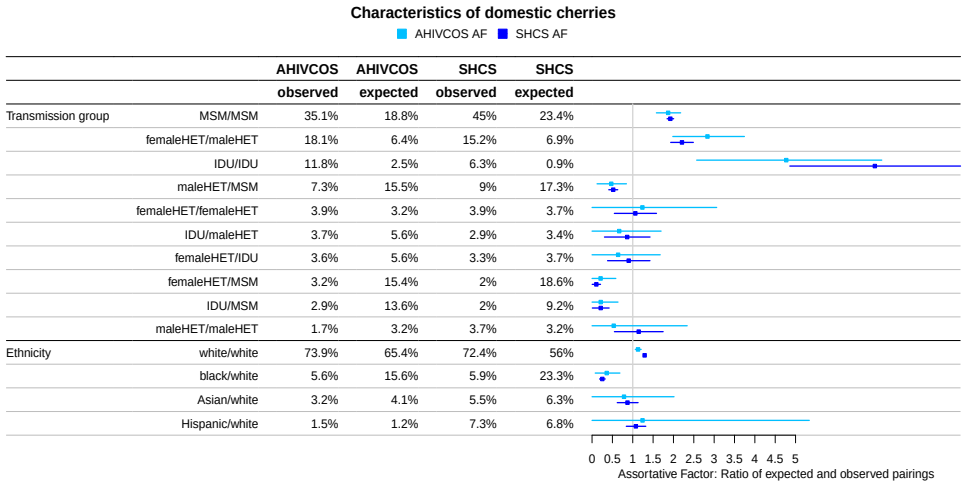
The direct overlap of the cherries obtained in the trees is increasing with decreasing distance threshold. The following tables show the fraction of the overlap of the original phylogeny and the two separate trees:

	CH/CH	CH/LA	Total	Number
0.045	0.73	0.54	0.68	3854
0.04	0.75	0.55	0.70	3567
0.035	0.76	0.56	0.72	3286
0.03	0.78	0.57	0.74	2935
0.025	0.81	0.61	0.78	2544
0.02	0.83	0.64	0.81	2168
0.015	0.85	0.73	0.84	1701

	AT/AT	AT/LA	Total	Number
0.045	0.75	0.51	0.68	902
0.04	0.77	0.53	0.70	850
0.035	0.78	0.55	0.72	798
0.03	0.80	0.58	0.75	726
0.025	0.81	0.61	0.77	638
0.02	0.83	0.60	0.79	569
0.015	0.83	0.71	0.81	469

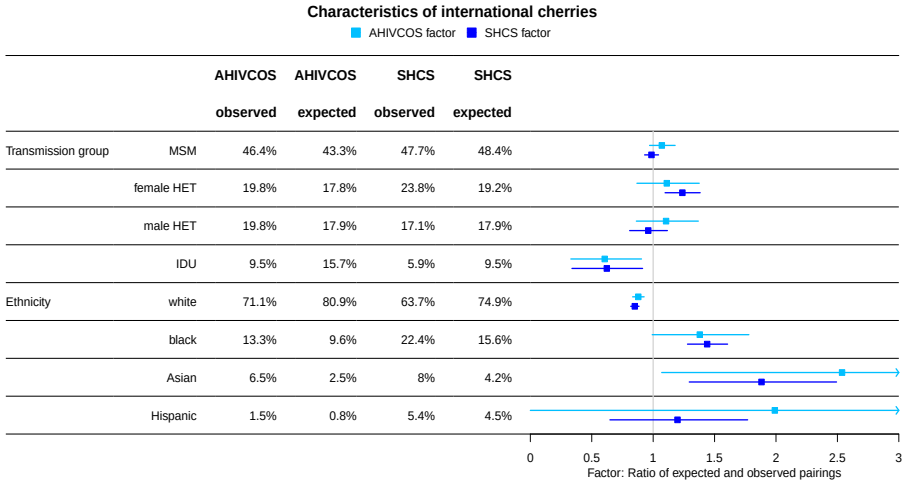
S8 Sensitivity Analysis: Excluding retrospectively sampled sequences

One major difference between the SHCS and AHIVCOS sequence sampling-strategies is, that the SHCS sampled a lot of sequences retrospectively for research purposes. Because of this potential bias, we reran the whole procedure to produce a new phylogenetic tree with retrospectively sequenced SHCS samples removed. We included 6109 sequenced, i.e., only around half of the sample size in this analysis. The results regarding characteristics regarding domestic and international cherries, however, stay robust after removing these samples (see Figure S10 and S11).



		AHIVCOS observed	AHIVCOS expected	SHCS observed	SHCS expected
Transmission group	MSM/MSM	35.1%	18.8%	45%	23.4%
	femaleHET/maleHET	18.1%	6.4%	15.2%	6.9%
	IDU/IDU	11.8%	2.5%	6.3%	0.9%
	maleHET/MSM	7.3%	15.5%	9%	17.3%
	femaleHET/femaleHET	3.9%	3.2%	3.9%	3.7%
	IDU/maleHET	3.7%	5.6%	2.9%	3.4%
	femaleHET/IDU	3.6%	5.6%	3.3%	3.7%
	femaleHET/MSM	3.2%	15.4%	2%	18.6%
	IDU/MSM	2.9%	13.6%	2%	9.2%
	maleHET/maleHET	1.7%	3.2%	3.7%	3.2%
Ethnicity	white/white	73.9%	65.4%	72.4%	56%
	black/white	5.6%	15.6%	5.9%	23.3%
	Asian/white	3.2%	4.1%	5.5%	6.3%
	Hispanic/white	1.5%	1.2%	7.3%	6.8%

Figure S10: Characteristics of domestic cherries: Sensitivity analysis after dropping retrospectively sampled SHCS sequences.



		AHIVCOS observed	AHIVCOS expected	SHCS observed	SHCS expected
Transmission group	MSM	46.4%	43.3%	47.7%	48.4%
	female HET	19.8%	17.8%	23.8%	19.2%
	male HET	19.8%	17.9%	17.1%	17.9%
	IDU	9.5%	15.7%	5.9%	9.5%
Ethnicity	white	71.1%	80.9%	63.7%	74.9%
	black	13.3%	9.6%	22.4%	15.6%
	Asian	6.5%	2.5%	8%	4.2%
	Hispanic	1.5%	0.8%	5.4%	4.5%

Figure S11: Characteristics of international cherries: Sensitivity analysis after dropping retrospectively sampled SHCS sequences.