Validation of phylogenetic SNP analysis for *Candida auris*

Fredrik Dyrkell[1], Susanne Staaf[1]

[1]1928 Diagnostics AB – Gothenburg (Sweden)

Background

Candida auris is a globally emerging multidrug-resistant pathogen that represents a serious public health threat with increase of healthcare-associated outbreaks and high rates of antifungal resistance. Genomic surveillance through whole-genome sequencing (WGS) can be used to detect and prevent outbreaks.

The 1928 platform (1928 Diagnostics) is a cloud-based service for analysing WGS data, tracing outbreaks and performing prospective genomic surveillance. The platform supports all major sequencing platforms and includes a SNP pipeline for high-resolution strain typing.

The aim of this project was to validate the 1928 SNP pipeline for the pathogen *Candida auris* using a benchmarking dataset.

Method

Twenty-two Illumina samples from the benchmarking article [1] were downloaded from ENA and subsequently uploaded to the 1928 platform. The sequences were trimmed and checked for sufficient quality (>=30x coverage). Reads were aligned against a reference genome (Genbank GCA_016772135.1) and variants were called using methods previously described in [2].

Results

The resulting UPGMA cluster tree, generated based on the pairwise distance matrix of variants, is shown in Figure 1.

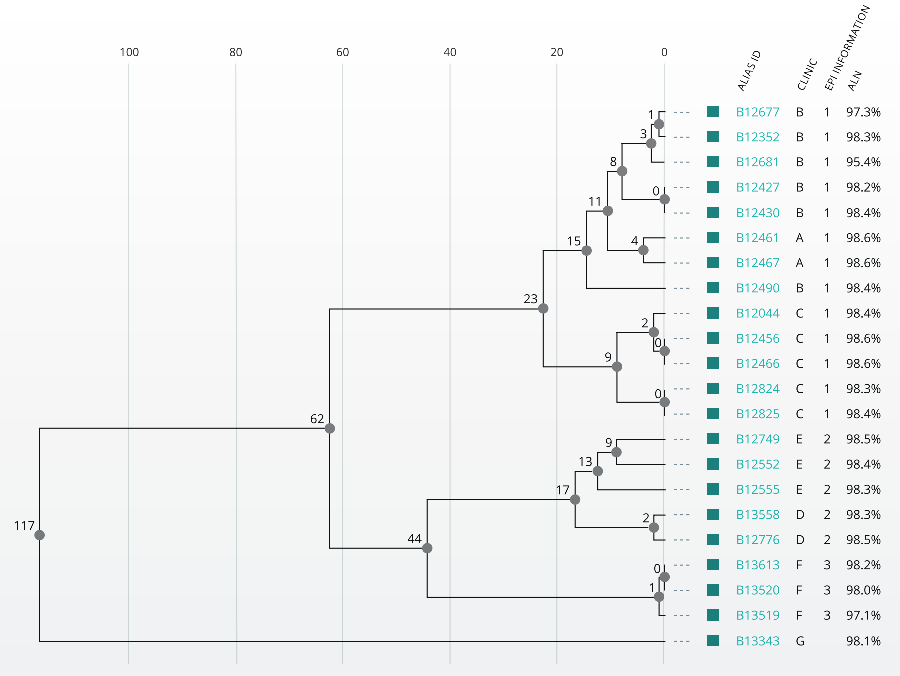


Figure 1. UPGMA tree – Distance shown at each node denotes average pairwise SNP difference between all samples in the subtree.

A neighbour-joining tree (Figure 2) was manually generated for easy comparison against the benchmark article results [1]. Within the three outbreak groups, the mean SNP distance was 15, ranging from 0 to 30.

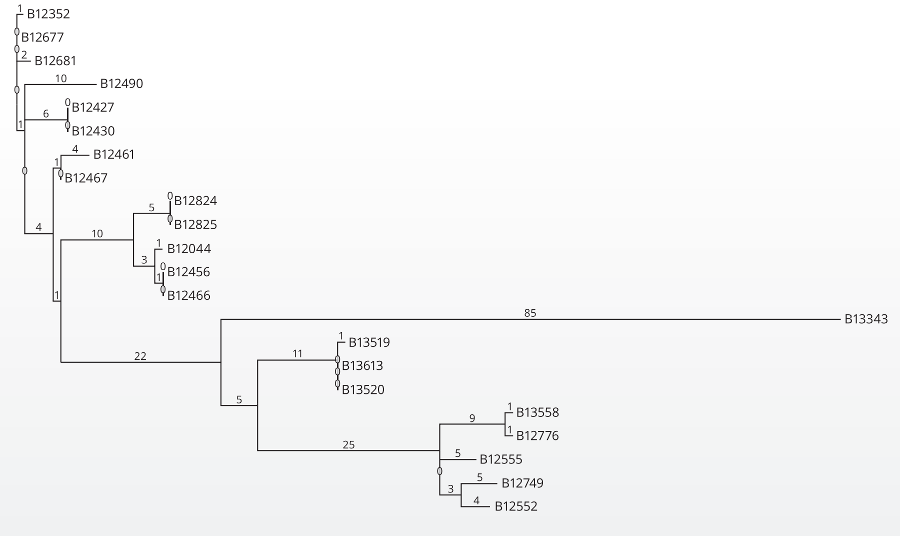
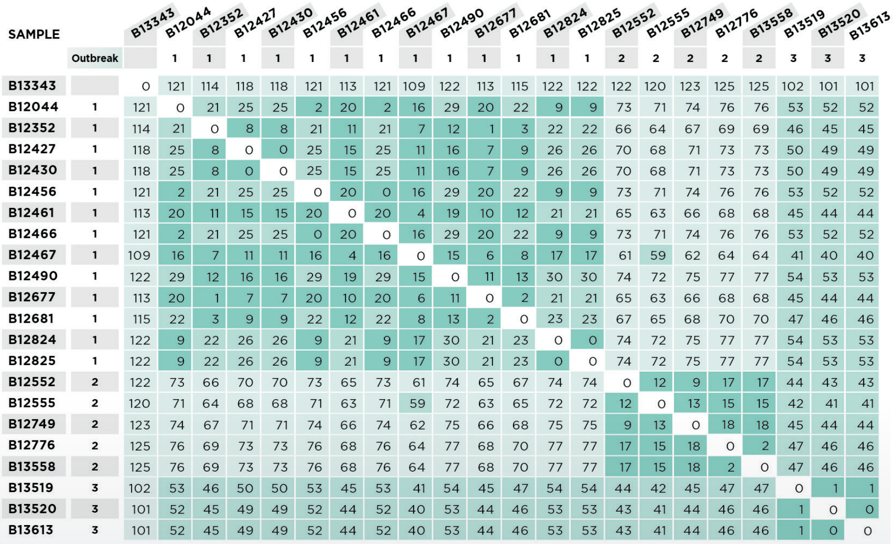


Figure 2. Neighbour-joining tree. Distances on branches shown as number of SNP differences.

continuation: **Results**

The complete distance matrix is shown in Figure 3.



SAMPLE		B13343	B12044	B12352	B12427	B12430	B12456	B12461	B12466	B12467	B12490	B12677	B12681	B12824	B12825	B12552	B12555	B12749	B12776	B13558	B13519	B13520	B13613
	Outbreak		1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	3	3	3
B13343		0	121	114	118	118	121	113	121	109	122	113	115	122	122	122	120	123	125	125	102	101	101
B12044	1	121	0	21	25	25	2	20	2	16	29	20	22	9	9	73	71	74	76	76	53	52	52
B12352	1	114	21	0	8	8	21	11	21	7	12	1	3	22	22	66	64	67	69	69	46	45	45
B12427	1	118	25	8	0	0	25	15	25	11	16	7	9	26	26	70	68	71	73	73	50	49	49
B12430	1	118	25	8	0	0	25	15	25	11	16	7	9	26	26	70	68	71	73	73	50	49	49
B12456	1	121	2	21	25	25	0	20	0	16	29	20	22	9	9	73	71	74	76	76	53	52	52
B12461	1	113	20	11	15	15	20	0	20	4	19	10	12	21	21	65	63	66	68	68	45	44	44
B12466	1	121	2	21	25	25	0	20	0	16	29	20	22	9	9	73	71	74	76	76	53	52	52
B12467	1	109	16	7	11	11	16	4	16	0	15	6	8	17	17	61	59	62	64	64	41	40	40
B12490	1	122	29	12	16	16	29	19	29	15	0	11	13	30	30	74	72	75	77	77	54	53	53
B12677	1	113	20	1	7	7	20	10	20	6	11	0	2	21	21	65	63	66	68	68	45	44	44
B12681	1	115	22	3	9	9	22	12	22	8	13	2	0	23	23	67	65	68	70	70	47	46	46
B12824	1	122	9	22	26	26	9	21	9	17	30	21	23	0	0	74	72	75	77	77	54	53	53
B12825	1	122	9	22	26	26	9	21	9	17	30	21	23	0	0	74	72	75	77	77	54	53	53
B12552	2	122	73	66	70	70	73	65	73	61	74	65	67	74	74	0	12	9	17	17	44	43	43
B12555	2	120	71	64	68	68	71	63	71	59	72	63	65	72	72	12	0	13	15	15	42	41	41
B12749	2	123	74	67	71	71	74	66	74	62	75	66	68	75	75	9	13	0	18	18	45	44	44
B12776	2	125	76	69	73	73	76	68	76	64	77	68	70	77	77	17	15	18	0	2	47	46	46
B13558	2	125	76	69	73	73	76	68	76	64	77	68	70	77	77	17	15	18	2	0	47	46	46
B13519	3	102	53	46	50	50	53	45	53	41	54	45	47	54	54	44	42	45	47	47	0	1	1
B13520	3	101	52	45	49	49	52	44	52	40	53	44	46	53	53	43	41	44	46	46	1	0	0
B13613	3	101	52	45	49	49	52	44	52	40	53	44	46	53	53	43	41	44	46	46	1	0	0

Figure 3. Distance matrix. Pairwise SNP differences between all samples.

Conclusions

Phylogenetic analysis of *Candida auris* using the 1928 platform clearly distinguishes all three expected outbreak clusters, and furthermore distinctly separates the different facilities within each outbreak group.

References

1. Welsh RM et al. *Candida auris* Whole-Genome Sequence Benchmark Dataset for Phylogenomic Pipelines. Journal of Fungi. MDPI AG; 2021. p. 214.
2. Werner A et al. Whole genome sequencing of Clostridioides difficile PCR ribotype 046 suggests transmission between pigs and humans. Chang Y-F, editor. PLOS ONE. Public Library of Science (PLoS); 2020. p. e0244227.

Conflict of interest statement

F.D. and S.S. are employees of 1928 Diagnostics AB.

Contact

fredrik.dyrkell@1928diagnostics.com