

Genomic Analysis of SARS-CoV-2 Variants of Concern Circulating in Hawai'i to Facilitate Public-Health Policies

David P Maison

University of Hawaii at Manoa <https://orcid.org/0000-0003-3428-1931>

Sean B. Cleveland

University of Hawaii - System <https://orcid.org/0000-0002-7130-3434>

Vivek R Nerurkar (✉ nerurkar@hawaii.edu)

University of Hawaii at Manoa <https://orcid.org/0000-0002-2044-2784>

Research Article

Keywords: SARS-CoV-2, COVID-19, quarantine, variant, tracking, origin, lineage, phylogenetic, Hawai'i

Posted Date: July 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-378702/v3>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Using genomics, bioinformatics and statistics, herein we demonstrate the effect of statewide and nationwide quarantine on the introduction of SARS-CoV-2 variants of concern (VOC) in Hawai'i. To define the origins of introduced VOC, we analyzed 260 VOC sequences from Hawai'i, and 301,646 VOC sequences worldwide, deposited in the GenBank and global initiative on sharing all influenza data (GISAID), and constructed phylogenetic trees. The trees define the most recent common ancestor as the origin. Further, the multiple sequence alignment used to generate the phylogenetic trees identified the consensus single nucleotide polymorphisms in the VOC genomes. These consensus sequences allow for VOC comparison and identification of mutations of interest in relation to viral immune evasion and host immune activation. Of note is the P71L substitution within the E protein, the protein sensed by TLR2 to produce cytokines, found in the B.1.351 VOC may diminish the efficacy of some vaccines. Based on the phylogenetic trees, the B.1.1.7, B.1.351, B.1.427, and B.1.429 VOC have been introduced in Hawai'i multiple times since December 2020 from several definable geographic regions. From the first worldwide report of VOC in GenBank and GISAID, to the first arrival of VOC in Hawai'i, averages 320 days with quarantine, and 132 days without quarantine. As such, the effect of quarantine is shown to significantly affect the time to arrival of VOC in Hawai'i, both during and following quarantine. Further, the collective 2020 quarantine of 43-states in the United States demonstrates a profound impact in delaying the arrival of VOC in states that did not practice quarantine, such as Utah. Our data demonstrates that at least 76% of all definable SARS-CoV-2 VOC have entered Hawai'i from California, with the B.1.351 variant in Hawai'i originating exclusively from the United Kingdom. These data provide a foundation for policy-makers and public-health officials to apply precision public health genomics to real-world policies such as mandatory screening and quarantine.

Introduction

Precision public health genomics is a public health policy tool to track the spread of viruses. In the age of fast-evolving digital information, precision public health genomics became prominent during the West Africa Ebola outbreak from 2014-2016.¹ This tool has not been efficiently and effectively used during the COVID-19 pandemic due to the overwhelming worldwide sequencing effort. A testament to this effort is the deposition of over 2 million severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) whole-genome sequences (WGS) in the GenBank and global initiative on sharing all influenza data (GISAID) since January 2020. For precision public health genomics to be effective during the COVID-19 pandemic, high-throughput sequencing and high-speed, low-cost sequence data analysis, and robust phylogenetics are necessary. Also, a fast, effective, consistent, and economical method is required to analyze the vast amount of SARS-CoV-2 sequences and determine the origin of SARS-CoV-2 variants of concern (VOC) and lineages in populations worldwide.

The Pangolin/Phylogenetic Assignment of Named Global Outbreak (PANGO) Lineage nomenclature system developed by Rambaut and colleagues² for SARS-CoV-2 lineages has allowed for manageable and efficient partitioning of sequences for the rapid determination of SARS-CoV-2 origin. The partitioning

converts the vast number of sequences into smaller collections of pre-defined similar sequences. These can further generate multiple sequence alignments (MSA) to produce phylogenetic trees efficiently and at low cost.

Using the CDC classified SARS-CoV-2 VOC (B.1.1.7, B.1.351, B.1.427, B.1.429, and P.1), identified in Hawai'i³⁻⁷ as an example, we demonstrate a method to define the origin of SARS-CoV-2 lineages and VOC. This method works using either open-source or licensed software with either a personal computer or a supercomputer. Using this method and associated analysis to define the origin of introduction of VOC, public-health officials can develop evidence-based policies to curtail the spread of VOC.

Methods

Defining the Origin of SARS-CoV-2 VOC

Here, we present a method to define SARS-CoV-2 VOC origin. We used 14 B.1.1.7, five B.1.351, 34 B.1.427, and 207 B.1.429 SARS-CoV-2 VOC sequences from Hawai'i deposited as of April 2, 2021, in the GenBank and GISAID as an example. In brief, the method includes: 1) The lineage-defining sequences of SARS-CoV-2 Lineage A and Lineage B act as the most ancestral roots.² Lineage A (EPI_ISL_406801) is from GISAID, and Lineage B (MN908947) is from GenBank. 2) To identify lineages of interest in an area, filter GISAID by location (e.g.: North America/USA/Hawai'i) and download all sequences. For VOC with >10,000 sequences, GISAID sequences were downloaded in batches due to GISAID maximum download size. Similarly, all geographically similar sequences reported in GenBank were downloaded using the search term SARS-CoV-2 and state abbreviation (e.g., "SARS-CoV-2 HI") and the sequence length filter (20,000 - 40,000). 3) Combine the GISAID and GenBank sequences into one .fasta file using AliView, Geneious Prime 2021.0.3 (<http://www.geneious.com>), or a text editor, and assign lineages using Pangolin Lineage Assigner (pangolin.cog-uk.io).^{2,8-10} 4) Download the results to Microsoft Excel, use advanced filter to copy unique records of lineages to a new column (ex: column M), then use COUNTIF (e.g., =COUNTIF(\$B\$2:\$B\$1432,M2)) to determine prevalence of each lineage. Alternatively, upload the results to Google Sheets and use the =UNIQUE command (e.g., =UNIQUE(B2:B1432)) followed by the above COUNTIF command. 5) Identify sequences that are the lineage of interest (e.g.: B.1.429), from using Pangolin Lineage Assigner, and download those sequences individually from GenBank. Filter GISAID by the lineage of interest (e.g., B.1.429) and download all sequences. 6) Combine lineage of interest (B.1.429) GenBank sequences, GISAID (B.1.429) sequences, and EPI_ISL_406801 into one fasta file. 7) Align sequences using multiple alignment using fast Fourier transform (MAFFT) program or server (https://mafft.cbrc.jp/alignment/server/add_fragments.html?frommanualnov6)¹¹⁻¹³ with MN908947 as a reference and do not remove any uninformative sequences and all parameters set as "same as input." 8) Remove the newly added MN908947 sequence that MAFFT places at the beginning of the alignment using AliView, Geneious Prime, or a text editor. If not, the sRNA toolbox will remove the MN908947 sequence during the duplicate removal step, and Lineage B will not serve as an ancestral root in the phylogenetic tree. 9) Import MSA file into Geneious Prime or AliView,¹⁰ search for the orf1a 5' start of the

entire alignment (5'-atggagagccttgctccctggttca-3') and remove the 5' untranslated region (UTR) by deleting the upstream region (~265 bp) from the MSA. Next, search for ORF10 3' end (5'-tgtagttaactttaatctcacatag-3') and remove the entire 3' UTR by deleting the downstream region (~229 bp) from the MSA. 10) Create a duplicate file for the MN908947 sequence and remove the 5' UTR and 3' UTR from MN908947 as described above. 11) Using MAFFT, align the trimmed MSA with the trimmed MN908947 as a reference and delete sequences with uncalled nucleotides 'n'. Set the "remove uninformative sequences" parameter in the MAFFT at >0%. 12) Using sRNAtoolbox program or server (<https://arn.ugr.es/srnatoolbox/helper/removedup/>),¹⁴ load the updated alignment to remove duplicate sequences and merge identifications (also referred to as sequence accession numbers) of duplicates. This merger will create "appendages" in the phylogenetic tree where the sRNA toolbox will line up identical sequences together with equal signs (=). 13) Import the final alignment into Geneious Prime and create an approximately maximum-likelihood phylogenetic tree using the FastTree program.¹⁵ Alternatively, FastTree can run as standalone software, and FastTreeMP is appropriate when multiple CPU cores/threads are available. 14) Root the tree with Lineage A (EPI_ISL_406801), which should then be the most recent common ancestor (MRCA) to Lineage B (MN908947) if performing phylogenetics on a Lineage B subgroup. Identify the MRCA of each sequence of interest.

For the B.1.1.7 VOC, which began with 272,732 sequences, we partitioned the sequences into seven sub-MSAs of ~50,000 sequences and performed the above method on each sub-MSA. After unambiguous sequences and duplicates were removed from each group, sub-MSA were recombined using AliView.¹⁰ Duplicates were removed after each recombination of two sub-MSA, except for the final MSA due to size restrictions. Four sub-MSA were combined to create the final MSA. All sequences in the final MSA are unambiguous. However, there are likely duplicate sequences present. FastTreeMP generated the phylogenetic tree for B.1.1.7 in the University of Hawai'i MANA High-Performance Computing Cluster (HPC). All other steps, including the phylogenetic trees for B.1.351, B.1.427, and B.1.429, were done using a 2014 Apple MacBook Pro (2.6 GHz Intel Core i5, 8 GB RAM) or a Dell OptiPlex 3070 (3.0 GHz Intel Core i5, 8 GB RAM).

Identifying the Consensus of Each VOC

To identify consensus SNPs of SARS-CoV-2 VOCs, assign Lineage B as the reference sequence, and use the Geneious Prime "Find Variations/SNPs" Annotate and Predict function to identify consensus SNPs. Input SNPs into the SnapGene (Insightful Science, snapgene.com) to identify the nucleotide and amino acid number and substitution as described previously.¹⁶

Evaluating the Effect of Quarantine

To test the hypothesis that the 67-day (2020-03-25 to 2020-05-31)¹⁷ quarantine in Hawai'i, and the collective 43-state quarantine in the United States that occurred from 2020-03-11 to 2020-06-16, significantly delayed the arrival of VOC, we partitioned the VOC into two categories. The first category of "quarantine" are those VOC (B.1.1.7, B.1.351, and B.1.429) that emerged worldwide before and during the

43-state collective mandatory quarantine was in effect. The second category of “post-quarantine” are those VOC (B.1.427 and P.1) that emerged worldwide after all quarantines were lifted. To determine the earliest collection date of each VOC worldwide, we analyzed all VOC (764,134) reported in the GISAID as of May 20, 2021. Since only 10 WGS of VOC B.1.351 have been reported from Hawai’i as of May 20, 2021, we analyzed the 10 first reported, and genetically distinct, of each VOC (B.1.429, B.1.427, B.1.1.7, B.1.351, and P.1) (total 47, since only 7 of the 10 B.1.351 were distinctive genetically) from Hawai’i as of May 20, 2021. For each VOC, we calculated the days between the first worldwide report and each of the first ten genetically distinct strains of VOC introduced in Hawai’i. We then compared the quarantine group to the post-quarantine group by days-to-arrival using an independent t-test in RStudio version 1.3.1093 (R version 4.0.3) and plotted with ggplot2 and ggstatsplot packages.¹⁸⁻²⁰ Further, we evaluated one of the seven states (AR, IA, ND, NE, SD, UT, WY)²¹ that did not participate in the collective 43-state quarantine to determine if the delay in the arrival of VOC between quarantine and non-quarantine time periods exists for a state that did not quarantine. The nearest geographic state to Hawai’i of the remaining seven states which did not quarantine is Utah. Therefore, we evaluated Utah’s first ten genetically distinct reports of each of the five VOC (Utah only had three strains of B.1.351 VOC, with two being unique (total 42)), as mentioned above, and compared them to the groups from Hawai’i.

Results

As of April 02, 2021, 43 unique lineages were identified with Pangolin Lineage Assigner from the 1,431 total sequences deposited in the GenBank and GISAID from Hawai’i (Table 1). Based on this analysis, the B.1.429, B.1.427, B.1.1.7, and B.1.351 were 14.47%, 2.38%, 0.98%, and 0.35%, respectively, prevalent in Hawai’i. Moreover, the SARS-CoV-2 VOC, as classified by the CDC, were overall 18.2% (260/1,431) prevalent among all SARS-CoV-2 sequences in Hawai’i (Table 1). As of May 20, 2021, the VOC prevalence had increased overall to 45.2% (1,069/2,367) (Table 1). This increase includes the emergence of the P.1 VOC, a 11.78% increase in B.1.1.7, and a 14.13% increase in B.1.429 (Table 1).

Table 1. SARS-CoV-2 Lineage Prevalence and Reporting Dates in Hawaii as of April 02, 2021, and May 20, 2021

Lineage	VOC* Identifier	Prevalence of Lineages			VOC Reporting Dates		
		April 02, 2021	May 20, 2021	Δ%	First Reported VOC	Arrival of VOC in Hawaii	Δ (Days)
		n (%)	n (%)				
A.1	-	6 (0.42)	6 (0.25)	-0.17	-	-	-
A.2.2	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
A.3	-	2 (0.14)	2 (0.08)	-0.06	-	-	-
B	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1	-	66 (4.61)	54 (2.28)	-2.33	-	-	-
B.1.1	-	11 (0.77)	11 (0.47)	-0.30	-	-	-
B.1.1.207	-	1 (0.07)	3 (0.13)	+0.06	-	-	-
B.1.1.222	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.1.304	-	-	2 (0.08)	+0.08	-	-	-
B.1.1.316	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.1.380	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.1.416	-	7 (0.07)	7 (0.3)	+0.23	-	-	-
B.1.1.519	-	18 (1.26)	43 (1.82)	+0.56	-	-	-
B.1.1.7	United Kingdom	14 (0.98)	302 (12.76)	+11.78	2020-02-07	2021-01-21	349
B.1.108	-	2 (0.14)	2 (0.08)	-0.06	-	-	-
B.1.139	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.160	-	-	9 (0.38)	+0.38	-	-	-
B.1.2	-	171 (11.95)	214 (9.04)	-2.91	-	-	-
B.1.234	-	7 (0.49)	7 (0.3)	-0.19	-	-	-
B.1.241	-	2 (0.14)	3 (0.13)	-0.01	-	-	-
B.1.243	-	745 (52.06)	751 (31.73)	-20.33	-	-	-
B.1.265	-	2 (0.14)	2 (0.08)	-0.06	-	-	-
B.1.298	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.340	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.351	South Africa	5 (0.35)	10 (0.42)	+0.07	2020-05-11	2021-02-16	281
B.1.357	-	61 (4.26)	61 (2.58)	-1.68	-	-	-
B.1.36.8	-	5 (0.35)	5 (0.21)	-0.14	-	-	-
B.1.369	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.37	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.400	-	9 (0.63)	9 (0.38)	-0.25	-	-	-

B.1.413	-	4 (0.28)	4 (0.17)	-0.11	-	-	-
B.1.427	California	34 (2.38)	46 (1.94)	-0.44	2020-09-17	2020-12-07	81
B.1.429	California	207 (14.47)	677 (28.6)	+14.13	2020-04-15	2020-12-31	260
B.1.517	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.526	-	3 (0.21)	17 (0.72)	+0.51	-	-	-
B.1.526.1	-	-	3 (0.13)	+0.13	-	-	-
B.1.526.2	-	-	18 (0.76)	+0.76	-	-	-
B.1.561	-	1 (0.07)	8 (0.34)	+0.27	-	-	-
B.1.568	-	-	3 (0.13)	+0.13	-	-	-
B.1.575	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.588	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.595	-	4 (0.28)	4 (0.17)	-0.11	-	-	-
B.1.596	-	18 (1.26)	24 (1.01)	-0.25	-	-	-
B.1.601	-	1 (0.07)	1 (0.04)	-0.03	-	-	-
B.1.609	-	3 (0.21)	3 (0.13)	-0.08	-	-	-
B.6	-	5 (0.35)	5 (0.21)	-0.14	-	-	-
P.1	Brazil	-	34 (1.44)	+1.44	2020-11-03	2021-03-21	138
P.2	-	2 (0.14)	2 (0.08)	-0.06	-	-	-
R.1	-	2 (0.14)	2 (0.08)	-0.06	-	-	-
Total	-	1,431	2,367				
*VOC = Variant of concern (depicted in bold)							

Variants of Concern Consensus

Figure 1, shows the phylogenetic analysis of all VOC prevalent in Hawai'i reported worldwide rooted with the Lineage A reference sequence (EPI_ISL_406801).² These trees were generated using FastTree in MANA HPC and Geneious Prime.¹⁵ Based on this analysis, 228 of the 260 (87.69%) VOC found in Hawai'i have identifiable origins. Figure 2, shows the states in the continental United States, as well as the countries worldwide, that were identified as being the source of the B.1.429, B.1.427, B.1.1.7, and B.1.351 SARS-CoV-2 VOC introductions into Hawai'i. The consensus (>90%) of the B.1.429 MSA, B.1.427 MSA, B.1.1.7 MSA, and B.1.351 MSA revealed 20, 16, 27, and 19, respectively, genomic mutations as compared to the MN908947 Lineage B reference sequence (Table 2).

Table 2. Single Nucleotide Polymorphisms, Amino Acid Substitutions, and Deletions Among SARS-CoV-2 Variants of Concern Circulating in Hawai'i

Gene or region	Nucleotides and Amino Acids				VOC Circulating in Hawai'i			
	Nucleotide Loci	Nucleotide Change	Amino Acid Position	Amino Acid Change	B.1.427	B.1.429	B.1.1.7	B.1.351
orf1ab	913	C → T	216	-				
	1,059	C → T	265	Thr → Ile				
	2,395	C → T	710	-				
	2,597	T → C	778	-				
	3,037	C → T	924	-				
	3,267	C → T	1,001	Thr → Ile				
	5,230	G → T	1,655	Lys → Asn				
	5,388	C → A	1,708	Ala → Asp				
	5,986	C → T	1,907	-				
	6,954	T → C	2,230	Ile → Thr				
	8,947	C → T	2,894	-				
	9,738	G → C	3,158	Ser → Thr				
	10,323	A → G	3,353	Lys → Arg				
	11,288 - 11,296	Δ	3,675 - 3,677	ΔSerΔGlyΔPhe				
	12,100	C → T	3,945	-				
	12,878	A → G	4,205	Ile → Val				
	13,713	G → A	4,483	-				
	14,408	C → T	4,715	Pro → Leu				
	14,676	C → T	4,804	-				
	15,279	C → T	5,005	-				
16,176	T → C	5,304	-					
16,394	C → T	5,377	Pro → Leu					
17,014	G → T	5,584	Asp → Tyr					
S	21,600	G → T	13	Ser → Ile				
	21,765 - 21,770	Δ	69 - 70	ΔHisΔVal				

	21,801	A → C	80	Asp → Ala				
	21,991 - 21,993	Δ	144	ΔTyr				
	22,018	G → T	152	Trp → Cys				
	22,206	A → G	215	Asp → Gly				
	22,281-22,289	Δ	242 - 244	ΔLeuΔAlaΔLeu				
	22,813	G → T	417	Lys → Asn				
	22,917	T → G	452	Leu → Arg				
	23,012	G → A	484	Glu → Lys				
	23,063	A → T	501	Asn → Tyr				
	23,271	C → A	570	Ala → Asp				
	23,403	A → G	614	Asp → Gly				
	23,604	C → A	681	Pro → His				
	23,664	C → T	701	Ala → Val				
	23,709	C → T	716	Thr → Ile				
	24,349	T → C	929	-				
	24,506	T → G	982	Ser → Ala				
	24,914	G → C	1,118	Asp → His				
ORF3a	25,563	G → T	57	Gln → His				
	25,904	C → T	171	Ser → Leu				
E	26,456	C → T	71	Pro → Leu				
M	26,681	C → T	53	-				
ORF7b/ORF8 intron	27,890	G → T	-	-				
ORF8	27,972	C → T	27	Gln → stop				
	28,048	G → T	52	Arg → Ile				
	28,111	A → G	73	Tyr → Cys				
	28,253	C → T	120	-				
ORF8/N intron	28,271	Δ	-	-				
	28,272	A → T	-	-				
N	28,280 - 28,282	GAT → CTA	3	Asp → Leu				

28,881 - 28,883	GGG → AAC	203 - 204	ArgGly → LysArg					
28,887	C → T	205	Thr → Ile					
28,977	C → T	235	Ser → Phe					
29,362	C → T	363	-					

Consensus determined for B.1.427 VOC, B.1.429 VOC, B.1.1.7 VOC, and B.1.351 VOC from multiple sequence alignment of 1,019, 2,809, 83,471, and 778 strains, respectively. Mutations present in the corresponding VOC are shaded with black color.

B.1.429 California VOC

In the GenBank, as of April 02, 2021, 21 of 97 Hawai'i sequences were of lineage B.1.429 as determined with the Pangolin Lineage Assigner. One-hundred eighty-six sequences in GISAID from Hawai'i were identified as the B.1.429 VOC. Total B.1.429 VOC reported worldwide in GISAID was 15,393 as determined by applying the GISAID lineage filter. Thus, the starting sequence count was 15,416 (21 GenBank + 15,393 GISAID + 2 lineage origin = 15,416). Of the 15,416 sequences, 11,648 sequences were removed for being uninformative and containing incomplete sequences. Further, the sRNAtoolbox server removed 944 sequences containing duplicate sequences and 15 sequences of duplicate ID. The final alignment of 2,809 (15,416 - 11,648 ambiguous - 959 duplicate sequence and ID = 2,809) strains and subsequent phylogenetic analysis defined the origin of the B.1.429 variant introduced into Hawai'i (Figure 1A). Using this method, we were able to identify the origin of 183 of 207 B.1.429 sequences introduced into Hawai'i (Figure 2).

B.1.427 California VOC

In the GenBank, as of April 02, 2021, 8 of 97 Hawai'i sequences were B.1.427 VOC as determined with the Pangolin Lineage Assigner. Twenty-six sequences in GISAID from Hawai'i were identified as the B.1.427 VOC. Total B.1.427 lineages reported worldwide in GISAID were 6,562 as determined by applying the GISAID lineage filter. Thus, the starting sequence count was 6,572 (8 GenBank + 6,562 GISAID + 2 lineage origin = 6,572). Of the 6,572 sequences, 5,273 sequences were removed for being ambiguous. Further, the sRNAtoolbox server removed 278 sequences containing duplicate sequences and 3 sequences of duplicate ID. One duplicate ID was created by the duplicate sequence merger. The final alignment of 1,019 strains (6,572 - 5,273 ambiguous - (281 duplicate sequence and ID - 1 duplicate ID created by duplicate sequence merger) = 1,019) and subsequent phylogenetic analysis defined the origin of the B.1.427 variant introduced into Hawai'i (Figure 1B). Using this method, we were able to identify the origin of 22 of 34 B.1.427 sequences introduced into Hawai'i (Figure 2).

B.1.1.7 United Kingdom VOC

In the GenBank, as of April 02, 2021, 0 of 97 Hawai'i sequences were of lineage B.1.1.7 as determined with the Pangolin Lineage Assigner. Fourteen strains in GISAID from Hawai'i were identified as the B.1.1.7 VOC. Total B.1.1.7 lineages reported worldwide in GISAID were 272,730 as determined by applying the GISAID lineage filter. Thus, the starting sequence count was 272,734 (272,730 GISAID sequences + 4 lineage origin). The aforementioned method was applied to each of the seven sub-MSA, thereby removing a total of 113,685 ambiguous sequences, and 75,590 duplicate sequences and ID. Twelve duplicate IDs were created by the duplicate sequence merger. The final alignment of 83,471 sequences (272,734 - 113,685 ambiguous - (75,590 duplicate sequence and ID - 12 duplicate ID created by duplicate sequence merger) = 83,471) and subsequent phylogenetic analysis defined the origin of the B.1.1.7 VOC introduced into Hawai'i using phylogenetic analysis (Figure 1C). Using this method, we were able to identify the origin of 10 of 14 B.1.1.7 sequences introduced into Hawai'i (Figure 2).

B.1.351 South Africa VOC

In the GenBank, as of April 02, 2021, 0 of 97 Hawai'i sequences were B.1.351 VOC as determined with the Pangolin Lineage Assigner. Five sequences in GISAID from Hawai'i were identified as the B.1.351 VOC. Total B.1.351 sequences reported worldwide in GISAID were 6,961 as determined by applying the GISAID lineage filter. Of the 6,963 sequences aligned, 5,758 sequences were removed for being uninformative and containing incomplete sequences. Further, the sRNAtoolbox server removed 373 sequences containing duplicate sequences and 55 sequences of duplicate ID. One duplicate ID was created by the duplicate sequence merger. The final alignment of 778 strains (6,961 GISAID + 2 lineage origin - 5,758 ambiguous - (428 duplicate sequence and ID - 1 duplicate ID created by duplicate sequence merger) = 778) and subsequent phylogenetic analysis defined the origin of the B.1.1.7 VOC introduced into Hawai'i using phylogenetic analysis (Figure 1D). Using this method, we were able to identify the origin of all five B.1.351 sequences introduced into Hawai'i (Figure 2).

High-Quality Sequencing

Several published sequences worldwide have missing nucleotides between the 5' UTR and 3' UTR, such as 41.7%, 82.7%, 80.2%, and 75.6% of B.1.1.7, B.1.351, B.1.427, and B.1.429 sequences, respectively. Therefore, these sequences are not useful for phylogenetic analysis.

Quarantine

As of May 20, 2021, the first reported collection dates worldwide for B.1.429, B.1.427, B.1.1.7, B.1.351, and P.1 VOC, respectively, are 2020-04-15,²² 2020-09-17,²³ 2020-02-07,²⁴ 2020-05-11,²⁵ and 2020-11-03.²⁶ In Hawai'i, the first reported collection dates for B.1.429, B.1.427, B.1.1.7, B.1.351, and P.1 VOC were on 2020-12-31, 2020-12-07, 2021-01-21, 2021-02-16, and 2021-03-21, respectively.²⁷ From first reported worldwide collection date to first reported collection date in Hawai'i, for the B.1.429, B.1.427, B.1.1.7, B.1.351, and P.1 VOC, are 260, 81, 349, 281, and 138 days, respectively (Figure 3).

Partitioning days to VOC arrival in Hawai'i into quarantine (2020-03-11 to 2020-06-16) ($M = 320$ days, $SD = 45$) and post-quarantine (2020-06-16 to 2021-05-20) ($M = 132$ days, $SD = 21$) time periods demonstrates that quarantine significantly delayed the arrival of VOC to Hawai'i, $t(45) = 17.38$, $p = 1.38e-21$ (Figure 4A). Utah, a non-quarantine state, also demonstrated difference in days to VOC arrival between the time period defined by the collective quarantine of the 43 states (2020-03-11 to 2020-06-16) ($M = 285$ days, $SD = 71$) when compared to the nationwide no-quarantine time period (2020-06-16 to 2021-05-20) ($M = 116$ days, $SD = 40$), $t(40) = 9.37$, $p = 1.2e-11$ (Figure 4B). Comparing the group defined by the collective 43-state quarantine demonstrates that VOC arrival in Hawai'i ($M = 320$ days, $SD = 45$) was delayed when compared to Utah ($M = 285$ days, $SD = 71$), $t(47) = 2.11$, $p = 0.04$ (Figure 4C). Comparing the group defined by the time period following the collective 43-state quarantine demonstrates no significant difference in days to VOC arrival between Hawai'i ($M = 132$ days, $SD = 21$) when compared to Utah ($M = 116$ days, $SD = 40$), $t(38) = 1.59$, $p = 0.12$ (Figure 4D).

Discussion

Precision public health genomics has been a tool in past outbreaks that has yet to be applied for the COVID-19 pandemic. These data and the method serve as a foundation for policy-makers to apply precision public health genomics tools by discerning trends related to the source of SARS-CoV-2 introductions. By identifying the origin of SARS-CoV-2, policies can be reasonably constructed with evidence-based decisions.

The Origins of Variants of Concern in Hawai'i

Fourteen geographical locations are definable as the origin of SARS-CoV-2 VOC in Hawai'i. From the most to the least VOC introductions in Hawai'i, these locations are California (174 introduced VOC), Washington (19), Texas (10), Louisiana (6), United Kingdom (5), Pennsylvania (4), Colorado (2), Florida (1), Arizona (1), Illinois (1), Indiana (1), Nevada (1), North Carolina (1), Ohio (1) and Oregon (1).

The MRCA branch to the sequences from Hawai'i indicates that the B.1.429, B.1.427, and B.1.1.7 VOCs were introduced into Hawai'i independently at different times from the aforementioned states in the continental United States. The B.1.351 VOC was introduced to Hawai'i from the United Kingdom. The overwhelming majority of the VOC entering Hawai'i are exclusively from California.

Defining VOC Using Genetic Characterization

As an effect of performing the origin defining method, described in the methods section, the MSA of all unambiguous VOC genomes can characterize the genome of VOC. Additionally, with this method, we can robustly compare the genomic similarities and differences between VOC. Within the S gene of the B.1.1.7 VOC, there are nine nucleotide and amino acid changes compared to the wild-type. These are $\Delta 69-70$, $\Delta Y144$, N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H. Within the S gene of the B.1.427 and B.1.429 VOC, there is a consensus of four non-synonymous amino acid substitutions: S13I, W152C, L452R, D614G. Also, the S gene of the B.1.351 VOC encodes eight substitutions or deletions: D80A,

D215G, Δ242-244, K417N, E484K, N501Y, D614G, and A701V. Further, the B.1.351 contains a substitution in the E gene (P71L) that results in the slightly stabilizing, based on the changes of Gibbs free energy, loss of a proline in the envelope protein.²⁸ The envelope protein was recently shown to interact with TLR2 and initiate inflammatory response.²⁹ Prolines are known to be involved in beta-turns, and the P71L substitution could significantly change the secondary and tertiary protein structures. This proline loss is striking for vaccines, since some vaccines effective against wild-type SARS-CoV-2, have diminished efficacy against the B.1.351 variant.^{30,31}

Efforts to understand variants still focus primarily on identifying the effect of individual mutations and substitutions. Many of the other mutations have yet to be evaluated experimentally, either individually or in concert with the D614G substitution. The ubiquitous D614G substitution increases the fitness of SARS-CoV-2, even at the cost of increased susceptibility to neutralization.³²⁻³⁴ As such, SARS-CoV-2 has been evolving, and substitutions in the B.1.351 VOC and P.2 variant of interest (VOI), such as the E484K, are shown to confer resistance to neutralization and exhibit 50% increased transmission.^{35,36} Recent publications focused on L452R mutation demonstrate significant decrease in the effectiveness of mAb treatments and neutralization by convalescent and vaccine sera.^{37,38} Other mutated genes within VOC are orf1ab, ORF3a, M, ORF8, and N, as well as two introns. Mutations and substitutions in these genes and proteins, annotated in Table 2, are enigmatic, and warrant further studies. Conclusively, what is certain is that tracking the spread of these VOC, and determining the effects of their substitutions, is paramount in the effort to control the pandemic.

Ambiguous Sequences

This method demonstrates the need for high-quality sequencing and the need for enrichment and deep coverage. For example, the first B.1.429 sequence deposited from Hawai'i was from a sample collected on December 31, 2020 (EPI_ISL_967766) is presently unusable. Without resequencing the whole genome or filling in with Sanger sequencing, this sequence is currently not of use in phylogenetics and origin determination due to ambiguous nucleotides in the S gene. While uninformative sequences may be useful for tracking the emergence of individual mutations, they are not useful in tracking VOC.

Public Policy Recommendations and Impacts

As the scientific community continues to understand the vaccine and healthcare consequences of VOC, of crucial importance is to control and limit the spread of these VOC. Policy-makers must first ascertain the source of the spread before they can control and limit the spread of VOC. After policy-makers contain the source, healthcare providers will treat patients with the most effective therapy, and vaccines will uphold their efficacy. However, without such precision public health genomics in practice, society risks losing progress in the fight against this pandemic. As only 5.4% of the global population is fully vaccinated against SARS-CoV-2 as of May 28, 2021,³⁹ the possibility of new infections, even in vaccinated populations, is ever present without appropriate measures to deter the spread and evolution of SARS-CoV-2. To exemplify this healthcare point, the B.1.427 and B.1.429 are resistant to specific

monoclonal antibody (mAb) therapies.^{38,40} Clinically, of treatment importance, the Food and Drug Administration and California Department of Health and Human Services have stopped using Bamlanivimab due to reduced clinical activity against the B.1.427 and B.1.429 variants.^{40,41} Furthermore, the B.1.351 VOC demonstrates some resistance to mAbs, convalescent sera, and vaccine sera neutralizing antibodies, and the B.1.1.7 is resistant to some mAbs.⁴²

Much of Hawai'i was under a mandatory stay-at-home order (2020-03-22 - 2020-05-31) from the Governor of Hawai'i's Third⁴³ (2020-03-22) through Eighth⁴⁴ (2020-05-18) COVID-19 Proclamations, with intermittent and inter-island quarantines continuing through the Thirteenth⁴⁵ (2020-09-23) Proclamation. Trans-Pacific travelers were no longer required to quarantine as of October 15, 2020.⁴⁶ Of interest is that even though VOC were circulating worldwide, none entered Hawai'i during the stay-at-home order or intermittent quarantines; all entered after the conclusion of the Thirteenth Proclamation and the reinstatement of Trans-Pacific travel. Furthermore, the average time it took for the VOC to arrive into Hawai'i decreased from an average of 320 days to 132 days when compared between the quarantine and post-quarantine groups, respectively. The data here argues in favor of the success of the quarantine here in Hawai'i in preventing the influx of the SARS-CoV-2 VOC. Without public-health measure and quarantine, the B.1.617+ India VOI and B.1.525 New York VOI have already arrived in the Hawaiian Islands, or will do so by August 2021. Further, from 2020-03-11 to 2020-06-16, 43 states participated in a collective nationwide quarantine averaging 46 days per state, although there was no point in time where all 43 states were under simultaneous quarantine. We observed that even the seven states not participating in quarantine were able to significantly delay arrival of VOC, analogous to the "concept of herd immunity." Furthermore, in the post-collective-quarantine period (2020/06/16 - 2021/05/20), wherein no states were practicing quarantine, there was no significant difference in days to VOC arrival. This finding supports the data that quarantine (2020/03/11 - 2020/06/16) indeed delayed the arrival of VOC. While individual states benefit from separate quarantine, the data here indicate the success and impact of collective quarantine of the 43-states within the United States.

As the purpose of this article is to facilitate evidence based public-health policy, the vast number of VOC (194/228) entering Hawai'i from the West Coast of the United States is alarming and warrants policies directed at controlling the spread of the VOC in the Hawaiian Islands.

From the analysis of the SARS-CoV-2 sequence data, a policy-maker could reasonably consider focusing on additional screening, contact tracing, and quarantine efforts among travelers arriving from the West Coast of the continental United States. In terms of public policies and precision public health genomics, this information can direct funding into scientific studies evaluating the effects of mutations prevalent in specific populations, particularly mutations within genes that affect SARS-CoV-2 immunogenic epitopes. Policies should encourage research focusing on developing pseudoviruses⁴⁷ and infectious clones⁴⁸ to evaluate kinetics, virulence, anatomical localization, transmission, and neutralization by mAbs, convalescent sera, and vaccine sera. Funding to conduct the aforementioned research should be directed at local and national levels.

Limitations

As SARS-CoV-2 sequences continue to be submitted retrospectively, these data will evolve. As a tool for precision public health genomics, the highest value is the trends that this method elucidates. The quarantine data will also change, this analysis is a snapshot in time.

Conclusions

These methods demonstrate the ability of precision public health genomics to identify the origin of SARS-CoV-2, the success of quarantine in Hawai'i, and the concern of emerging VOC. The conclusion from defining the origin of VOC in Hawai'i is that California is the primary source of VOC circulating in Hawai'i. Additional screening and quarantining of the travelers from California while vacationing in Hawai'i will protect the local population from evasive SARS-CoV-2 VOC. A tool was needed to evaluate and make use of the vast worldwide sequencing effort and the tool herein fills that need. Moreover, our methodology demonstrates the ability of sequencing and phylogenetic analysis to provide precision public health genomics in policy-making decisions. As SARS-CoV-2 VOC spreads asymptotically across the United States and worldwide, it is essential to use fast and accurate SARS-CoV-2 VOC, lineage, and origin assignment for making evidence-based public-policy decisions.

Declarations

Acknowledgements

This research was supported by a grant (P30GM114737-05) from the Pacific Center for Emerging Infectious Diseases Research, COBRE and a grant (P20GM103466-20S1) from the INBRE, National Institute of General Medical Sciences, NIH. Computation was supported by NSF grant #1920304 on the University of Hawai'i MANA High Performance Computing Cluster. We thank Dr. Vedbar Khadka for assistance with identifying appropriate statistical tests to be used in this study.. The viral genome sequences used in this publication are publicly available from GenBank (<https://www.ncbi.nlm.nih.gov/sars-cov-2/>) and GISAID (<https://gisaid.org>). Tables of acknowledgements for the genome sequences from GISAID are available at: <https://github.com/dpmaison/Genomic-Analysis-of-SARS-CoV-2-Variants-of-Concern-Circulating-in-Hawai-i-to-Facilitate-Public-Health>

References

1. The Dutch-Covid-19 response team, Oude Munnink BB, Nieuwenhuijse DF, et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med*. 2020;26(9):1405-1410. doi:10.1038/s41591-020-0997-y
2. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*. 2020;5(11):1403-1407. doi:10.1038/s41564-020-0770-5

3. Wu BN. New COVID variant found in Hawaii, while uptick on Maui attributed to predominate California variant. Honolulu Star-Advertiser. Published February 26, 2021. Accessed March 14, 2021. <https://www.staradvertiser.com/2021/02/26/breaking-news/new-covid-variant-found-in-hawaii/>
4. DOH confirms new COVID-19 variant first detected in PH, first case of Brazil variant | ABS-CBN News. Accessed March 14, 2021. <https://news.abs-cbn.com/news/03/13/21/doh-confirms-new-covid-19-variant-first-detected-in-ph-first-case-of-brazil-variant>
5. DOH NEWS RELEASE: STATE'S MICROBIOLOGISTS PROVIDE GENOMIC VIRUS DATA TO SHAPE FUTURE PANDEMIC RESPONSE. Accessed May 20, 2021. <https://governor.hawaii.gov/newsroom/doh-news-release-states-microbiologists-provide-genomic-virus-data-to-shape-future-pandemic-response/>
6. Wu BN, Jan. 25 2021, am U 9:34. Hawaii health officials report two cases of new variant of COVID-19 – one on Oahu and one on Maui. Honolulu Star-Advertiser. Published January 25, 2021. Accessed January 26, 2021. <https://www.staradvertiser.com/2021/01/25/breaking-news/health-officials-confirm-denmark-variant-of-covid-19-in-hawaii/>
7. Consillio BK, March 8 2021. Health Department says highly transmissible COVID-19 variant present in Hawaii. Honolulu Star-Advertiser. Published March 8, 2021. Accessed March 14, 2021. <https://www.staradvertiser.com/2021/03/08/breaking-news/health-department-says-highly-transmissible-south-african-covid-19-variant-present-in-hawaii/>
8. O'Toole Á, Scher E, Underwood A, et al. pangolin: lineage assignment in an emerging pandemic as an epidemiological tool. PANGO lineages. Accessed March 11, 2021. github.com/cov-lineages/pangolin
9. *Pangolin (Version v.2.3.2)*. Centre for Genomic Pathogen Surveillance pangolin.cog-uk.io
10. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014;30(22):3276-3278. doi:10.1093/bioinformatics/btu531
11. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. 2019;20(4):1160-1166. doi:10.1093/bib/bbx108
12. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research*. 2019;47(W1):W5-W10. doi:10.1093/nar/gkz342
13. MAFFT version 7 - Multiple alignment program for amino acid or nucleotide sequences. MAFFT. Accessed March 20, 2021. https://mafft.cbrc.jp/alignment/server/add_fragments.html?frommanualnov6
14. Aparicio-Puerta E, Lebrón R, Rueda A, et al. sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression. *Nucleic Acids Research*. 2019;47(W1):W530-W535. doi:10.1093/nar/gkz415

15. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*. 2009;26(7):1641-1650. doi:10.1093/molbev/msp077
16. Maison DP, Ching LL, Shikuma CM, Nerurkar VR. Genetic Characteristics and Phylogeny of 969-bp S Gene Sequence of SARS-CoV-2 from Hawaii Reveals the Worldwide Emerging P681H Mutation. *Hawai'i Journal of Health and Social Welfare*. 80(3):52-61. doi:https://doi.org/10.1101/2021.01.06.425497
17. Coronavirus closures: Map of where US states are tightening restrictions. Accessed June 2, 2021. <https://www.usatoday.com/storytelling/coronavirus-reopening-america-map/>
18. R Core Team. *R: A Language and Environment for Statistical ## Computing*. R Foundation for Statistical Computing; 2020. <https://www.R-project.org/>
19. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. 2016. Springer International Publishing: Imprint: Springer; 2016. doi:10.1007/978-3-319-24277-4
20. Patil I. *Visualizations with Statistical Details: The "ggstatsplot" Approach*. PsyArXiv; 2021. doi:10.31234/osf.io/p7mku
21. States that did not issue stay-at-home orders in response to the coronavirus (COVID-19) pandemic, 2020. Ballotpedia. Accessed June 3, 2021. [https://ballotpedia.org/States_that_did_not_issue_stay-at-home_orders_in_response_to_the_coronavirus_\(COVID-19\)_pandemic,_2020](https://ballotpedia.org/States_that_did_not_issue_stay-at-home_orders_in_response_to_the_coronavirus_(COVID-19)_pandemic,_2020)
22. Gorzalski A, Pandori M. *Virus Name: HCoV-19/USA/NV-NSPHL-357909/2020 / Accession ID: EPI_ISL_1675148*. Nevada State Public Health Laboratory; 2020. <https://www.epicov.org/epi3/frontend#2aa2ed>
23. Gangavarapu K. *Virus Name: HCoV-19/Mexico/BCN-ALSR-8438/2020 / Accession ID: EPI_ISL_1531902*. Centro de Diagnostico COVID-19 UABC Tijuana; 2020. <https://www.epicov.org/epi3/frontend#1da1c7>
24. Hernandez M. *Virus Name: HCoV-19/Spain/CL-COV00781/2020 / Accession ID: EPI_ISL_1603195*. SARS-CoV-2 Sequencing Castilla y Leon-Spain Consortium; 2020. <https://www.epicov.org/epi3/frontend#378251>
25. Engelbrecht S. *Virus Name: HCoV-19/South Africa/Tygerberg_650/2020 / Accession ID: EPI_ISL_1591449*. Division of Medical Virology, National Health Laboratory Service (NHLS), Tygerberg Hospital / Stellenbosch University; 2020. <https://www.epicov.org/epi3/frontend#88164>
26. Sacchi CT. *Virus Name: HCoV-19/Brazil/SP-2389/2020 / Accession ID: EPI_ISL_1821208*. Hospital de Base de Bauru Hospital de Base Setima Regiao; 2020. <https://www.epicov.org/epi3/frontend#14e2e4>

27. Wu N. Variants of concern now the majority of COVID-19 cases in state, with UK variant now dominant in Hawaii. Honolulu Star-Advertiser. Published May 26, 2021. Accessed June 8, 2021. <https://www.staradvertiser.com/2021/05/26/breaking-news/variants-of-concern-now-the-majority-of-covid-19-cases-in-state-with-uk-variant-now-dominant-in-hawaii/>
28. Rahman MS, Hoque MN, Islam MR, et al. Mutational insights into the envelope protein of SARS-CoV-2. *Gene Rep.* 2021;22:100997. doi:10.1016/j.genrep.2020.100997
29. Zheng M, Karki R, Williams EP, et al. TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. *Nature Immunology.* Published online May 7, 2021. doi:10.1038/s41590-021-00937-x
30. Scobie H. Update on Emerging SARS-CoV-2 Variants and Vaccine Considerations. Published online May 12, 2021:30.
31. Madhi SA, Baillie V, Cutland CL, et al. Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. *N Engl J Med.* 2021;384(20):1885-1898. doi:10.1056/NEJMoa2102214
32. Weissman D, Alameh M-G, de Silva T, et al. D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host Microbe.* 2021;29(1):23-31.e4. doi:10.1016/j.chom.2020.11.012
33. Plante JA, Liu Y, Liu J, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature.* Published online October 26, 2020. doi:10.1038/s41586-020-2895-3
34. Korber B, Fischer WM, Gnanakaran S, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell.* 2020;182(4):812-827.e19. doi:10.1016/j.cell.2020.06.043
35. Jangra S, Ye C, Rathnasinghe R, et al. *The E484K Mutation in the SARS-CoV-2 Spike Protein Reduces but Does Not Abolish Neutralizing Activity of Human Convalescent and Post-Vaccination Sera.* *Infectious Diseases (except HIV/AIDS);* 2021. doi:10.1101/2021.01.26.21250543
36. Xie X, Liu Y, Liu J, et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat Med.* Published online February 8, 2021. doi:10.1038/s41591-021-01270-4
37. Deng X, Garcia-Knight MA, Khalid MM, et al. Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation. *medRxiv.* Published online March 9, 2021:2021.03.07.21252647. doi:10.1101/2021.03.07.21252647
38. Li Q, Wu J, Nie J, Li X, Huang W, Wang Y. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell.* 2020;182:1284-1294. doi:10.1016/j.cell.2020.07.012

39. Coronavirus (COVID-19) Vaccinations - Statistics and Research. Our World in Data. Accessed May 29, 2021. <https://ourworldindata.org/covid-vaccinations>
40. Aragón TJ, Newsom G. California Department of Public Health - Health Alert: Concerns re: the Use of Bamlanivimab Monotherapy in the Setting of SARS-CoV2 Variants. :4.
41. Moruf A. Fact Sheet For Health Care Providers Emergency Use Authorization (Eua) Of Bamlanivimab. :26.
42. Wang P, Nair MS, Liu L, et al. Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. *Nature*. Published online March 8, 2021. doi:10.1038/s41586-021-03398-2
43. Ige D. *THIRD SUPPLEMENTARY PROCLAMATION RELATED TO THE COVID-19 EMERGENCY;* 2020:8. https://governor.hawaii.gov/wp-content/uploads/2020/03/2003162-ATG_Third-Supplementary-Proclamation-for-COVID-19-signed.pdf
44. Ige D. *EIGHTH SUPPLEMENTARY PROCLAMATION RELATED TO THE COVID-19 EMERGENCY;* 2020. https://governor.hawaii.gov/wp-content/uploads/2020/05/2005088-ATG_Eighth-Supplementary-Proclamation-for-COVID-19-distribution-signed.pdf
45. Ige D. *THIRTEENTH PROCLAMATION RELATED TO THE COVID-19 EMERGENCY;* 2020. https://governor.hawaii.gov/wp-content/uploads/2020/09/2009139-ATG_Thirteenth-Supplementary-Proclamation-for-COVID-19-distribution-signed.pdf
46. Lund C. TIMELINE: It's been one year since Hawaii issued its first stay-at-home order. <https://www.hawaiinewsnow.com>. Published March 25, 2021. Accessed May 26, 2021. <https://www.hawaiinewsnow.com/2021/03/26/timeline-its-been-one-year-since-hawaii-issued-its-first-stay-at-home-order/>
47. Tsai W-Y, Ching LL, Hsieh S-C, Melish ME, Nerurkar VR, Wang W-K. A real-time and high-throughput neutralization test based on SARS-CoV-2 pseudovirus containing monomeric infrared fluorescent protein as reporter. *Emerging Microbes & Infections*. Published online April 30, 2021:1-38. doi:10.1080/22221751.2021.1925163
48. Xie X, Muruato A, Lokugamage KG, et al. An Infectious cDNA Clone of SARS-CoV-2. *Cell Host & Microbe*. 2020;27(5):841-848.e3. doi:10.1016/j.chom.2020.04.004

Figures

A) B.1.429 California VOC

Appendages



B) B.1.427 California VOC

C) B.1.1.7 United Kingdom VOC

D) B.1.351 South Africa VOC

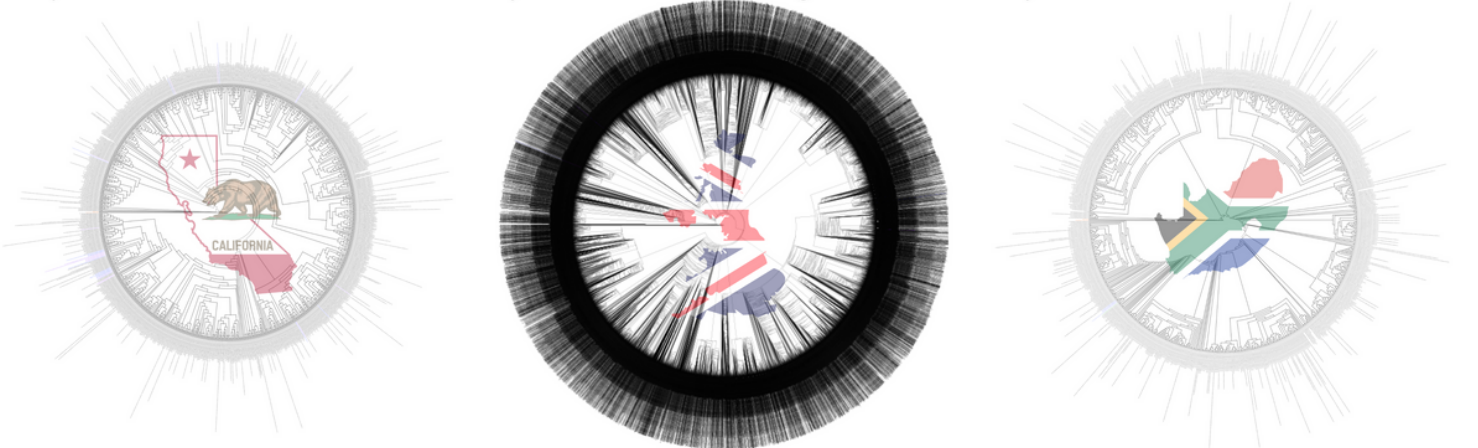


Figure 1

Phylogenetic Trees of Worldwide SARS-CoV-2 Variants of Concerns The phylogenetic trees show the approximately maximum-likelihood trees generated by FastTree in Geneious Prime 2021.0.3 (<http://www.geneious.com>) and FastTreeMP in the University of Hawai'i MANA High Performance Computing Cluster. The trees were rooted to the SARS-CoV-2 Lineage A reference sequence (EPI_ISL_406801). Clusters and strains from Hawai'i are identified with the colored text (blue), and were evaluated for the most recent common ancestor to define the origin of Hawai'i strains (pink). Text shown in black indicates global variant strains not necessarily directly affiliated with variants found in Hawai'i. Appendage sequences designated by an equal sign (=) indicate identical sequences as generated by the sRNAtoolbox. A) Phylogenetic tree of B.1.429 California variant of concern (VOC) generated using Geneious Prime from 2,809 unambiguous and unique sequences. B) Phylogenetic tree of B.1.427 California VOC generated using Geneious Prime from 1,019 unique and unambiguous sequences. C) Phylogenetic tree of B.1.1.7 United Kingdom VOC generated with FastTreeMP in the MANA HPC using 83,471 unambiguous sequences. D) Phylogenetic tree of B.1.351 South Africa VOC generated in Geneious Prime using 778

unambiguous and unique sequences. Stock images by Clker-Free-Vector-Images, OpenClipart-Vectors, and Gordon Johnson from Pixabay. Stock image editing was done in Adobe Photoshop 21.2.0, and the final figure was created with BioRender.com.

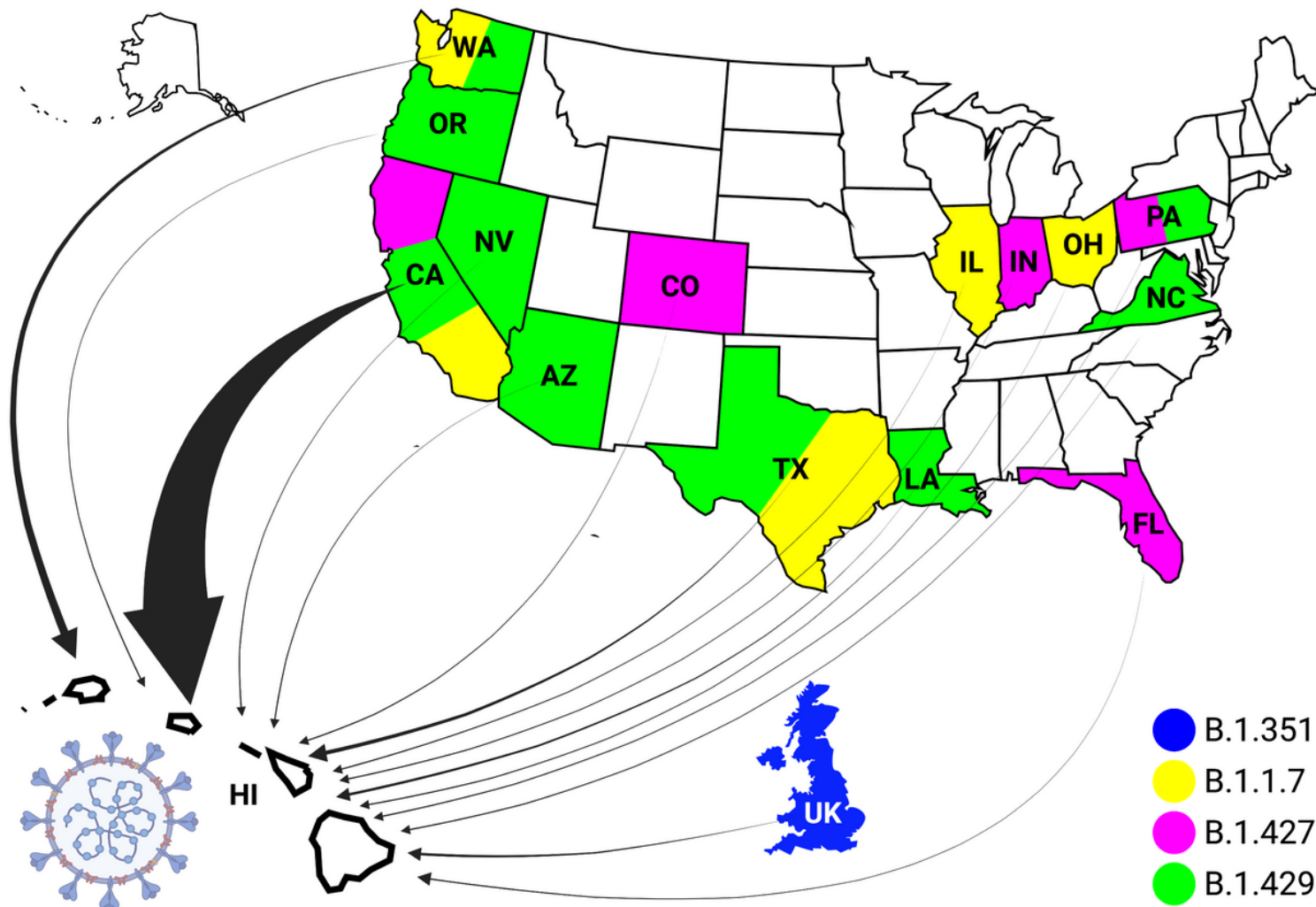


Figure 2

Introduction of the SAR-CoV-2 Variants of Concern in Hawai'i This figure shows the states within the United States, and countries worldwide, from which the SARS-CoV-2 variants of concern (VOC) have been introduced into Hawai'i. Represented geographical locations are: California (CA), Washington (WA), Texas (TX), Louisiana (LA), the United Kingdom (UK), Pennsylvania (PA), Colorado (CO), Florida (FL), Arizona (AZ), Illinois (IL), Indiana (IN), Nevada (NV), North Carolina (NC), Ohio (OH) and Oregon (OR). Blue represents the origin of B.1.351 VOC, yellow represents the origin of B.1.1.7 VOC, pink represents the origin of B.1.427 VOC, and the green represents the origin of B.1.429 VOC. The width of the arrows is proportional to the abundance of introductions of VOC from various geographic areas. Figure made with stock images by fajarbudi86 and Clker-Free-Vector-Images from Pixabay. Stock image editing was done in Adobe Photoshop 21.2.0 and the final figure was created with Biorender.com.

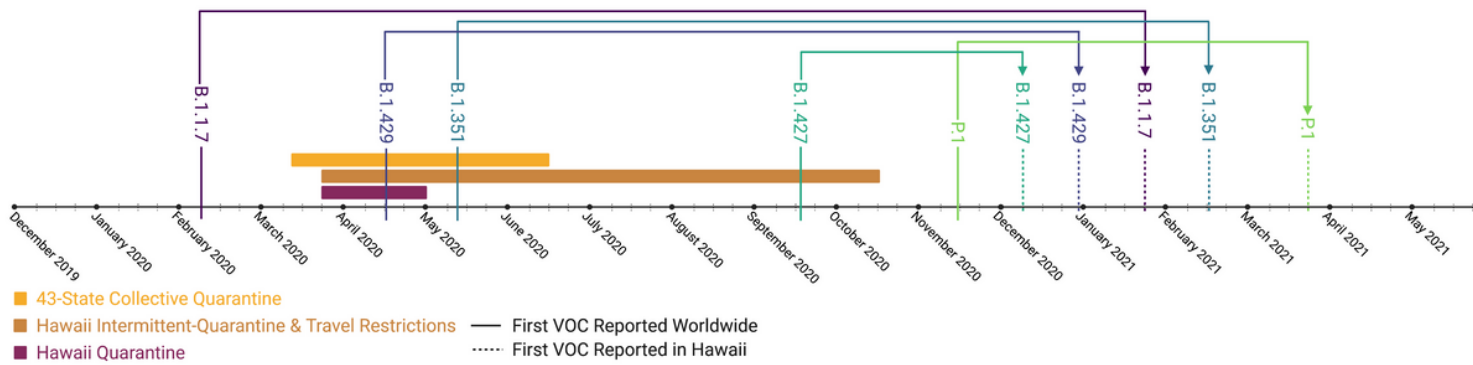


Figure 3

Emergence of SARS-CoV-2 VOC Timeline This figure shows the timeline of the emergence of SARS-CoV-2 variants of concern (VOC) with respect to Hawai'i. The x-axis is a calendar from December 2019 through May 20, 2021. Solid lines with the corresponding VOC name (B.1.1.7, B.1.429, B.1.351, B.1.427, and P.1) depict the worldwide emergence dates of the VOC. Dashed lines with the corresponding VOC name depict the first reported VOC date in Hawai'i. Dates of quarantine in Hawai'i (2020-03-22 - 2020-05-31), dates of intermittent quarantine in Hawai'i and trans-Pacific travel restrictions (2020-03-22 - 2020-10-15), and dates of the collective 43-state quarantine (2020-03-11 to 2020-06-16) are shown in colored boxes. As of May 20, 2021, the first reported collection dates worldwide for B.1.429, B.1.427, B.1.1.7, B.1.351, and P.1 VOC, respectively, are 2020-04-15, 2020-09-17, 2020-02-07, 2020-05-11, and 2020-11-03. In Hawai'i, the first reported collection dates for B.1.429, B.1.427, B.1.1.7, B.1.351, and P.1 VOC were on 2020-12-31, 2020-12-07, 2021-01-21, 2021-02-16, and 2021-03-21, respectively. The figure was created with Biorender.com.

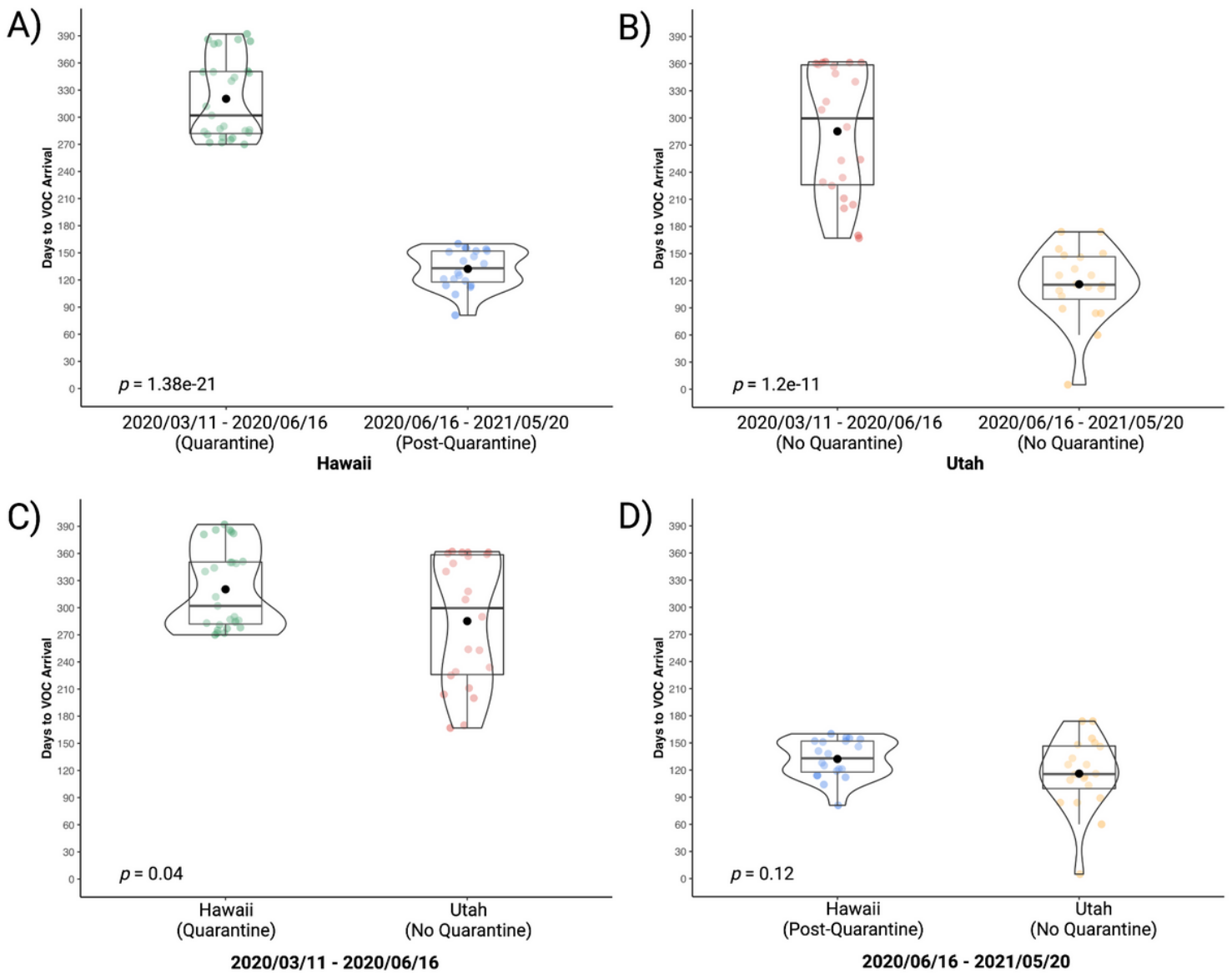


Figure 4

Post-Quarantine Days to VOC Arrival vs. Quarantine Days to VOC Arrival in Hawai'i and Utah as of May, 20, 2021 This figure shows the days-to-arrival of variants of concern (VOC) compared to VOC categorized as 'quarantine' and 'post-quarantine.' Depicted as a box-and-whisker/violin plot displaying median (line), mean (black circle), interquartile range, minimum, and maximum. Quarantine VOC are those VOC that emerged worldwide during the 43-state collective quarantine in the United States (2020-03-11 to 2020-06-16). Post-quarantine VOC are those VOC that emerged worldwide following the collective quarantine in the United States (2020-06-16 to 2021-05-20). The comparison was done using an independent t-test. A) Partitioning days to VOC arrival in Hawai'i into quarantine (M = 320 days, SD = 45) and post-quarantine (M = 132 days, SD = 21) time periods demonstrates that quarantine delayed the arrival of VOC to Hawai'i ($t(45) = 17.38$, $p = 1.38e-21$). B) For Utah, a state that did not quarantine during the 43-state collective quarantine, a difference is also demonstrated between the two time periods (quarantine M = 285 days, SD = 71; post-quarantine M = 116 days, SD = 40) regarding time to VOC arrival ($t(40) = 9.37$, $p = 1.2e-11$). C)

Comparing Hawai'i and Utah by the quarantine time period demonstrates that Hawai'i (M = 320 days, SD = 45), a quarantine state, saw a greater delay in days to VOC arrival compared to Utah (M = 285 days, SD = 71), a non-quarantine state ($t(47) = 2.11, p = 0.04$). D) Comparing Hawai'i and Utah by the post-quarantine time period demonstrates no difference between Hawai'i (M = 132 days, SD = 40) and Utah (M = 116 days, SD = 40) ($t(38) = 1.59, p = 0.12$) in days to VOC arrival. Created with RStudio version 1.3.1093 (R version 4.0.3) using ggplot2 and ggstatsplot packages. The final figure was created with BioRender.com.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.pdf](#)