CanCOGeN RCanGeCO

CanCOGeN Interim Recommendations for Naming, Identifying, and Reporting SARS-CoV-2 Variants of Concern

Version: 1.0

Last Revision Date: 2021-01-15

Authors: Emma Griffiths, Jennifer Tanner, Natalie Knox, Will Hsiao, and Gary Van Domselaar on behalf of the CPHLN and CanCOGeN

Introduction

Since the publication of Public Health England's initial report describing a recently emerged variant of concern designated VOC202012/1 in December of 2020, the identification, tracking, and characterization of SARS-CoV-2 variants has become a global priority. Variants of Concern (VOCs) are subgroups of SARS-CoV-2 viruses containing combinations of mutations that have been associated with a clinically or epidemiologically significant phenotype. Naming VOCs is difficult because viral taxonomy is complex, and viral taxonomic naming schemes can evolve as the virus evolves. Terms such as strain, clade, and variant are often used interchangeably by both the public and scientific community. This taxonomic fluidity, in combination with the different nomenclature schemes implemented by different analytical platforms and a mixture of common names used in science and media communication, has created a patchwork of names and aliases for variants. Such variability poses significant challenges for public health reporting of VOCs in Canada, which in turn complicates data integration and analysis, and ultimately communication with the public. Notably, VOCs are often referred to by the location where they were first reported, such as the "UK variant" or the "South Africa variant". However, the association of pathogens and disease with geographic locations or populations has been shown to cause stigma and create xenophobia. There currently exists no international guidelines for naming VOCs, although The World Health Organization recognizes the need for one and has convened an expert working group to provide these guidelines. Canada's ability to effectively track and respond to VOCs requires a standard process for naming, identifying, and reporting VOCs. To address this pressing need, we provide here an interim variant-naming scheme along with conventions for identifying and reporting VOCs for use by the Canadian Public Health Laboratory Network and CanCOGeN. These guidelines may be revised after international guidelines become available.

Proposed Canadian SARS-CoV-2 Variant Nomenclature System

Variant Naming

To reduce stigma, the Canadian nomenclature system excludes any reference to place names and instead relies on scientifically derived designations. Two widely used genomic analysis platforms—Pangolin and Nextstrain—have had enormous influence on how variants are named and defined. Pangolin is a command-line tool and web application that assigns a lineage name to a genome using the "Pango" nomenclature scheme[1]. Nextstrain is a phylodynamic analysis platform that enables users to analyze and visualize SARS-CoV-2 genomes in a global context. Because of its utility and common usage in the scientific community and the public sphere, our proposed nomenclature system adopts the Pango nomenclature for naming VOCs. The lineage names of the three currently known VOCs are presented in Table 1, along with common aliases (synonyms), for reference and clarification.

Table 1: Pango lineage designations and common aliases for VOCs

Pango Lineage	**Alias**
B.1.1.7	clade 20I/501Y.V1, VUI 202012/01, VOC 202012/01, B1.17, B117 UK variant, Kent Variant
B.1.351	clade 501Y.V2, SA S.501Y.V2, 20H/501Y.V2, B1351, South African variant
P.1	clade 20J/501Y.V3, B.1.1.28.1, P1, Manaus variant, B.1.1.28(K417N/E484K/N501Y), B.1.1.248 (Brazil/Japan), Brazilian variant

Lineages, Variants, and Variants of Concern

A viral *lineage* is a group of viruses defined by a founding variant and its descendents. Names are assigned to SARS-CoV-2 lineages using manual and automated methods. Lineage designations are based on phylogenetic grouping followed by the identification of shared, common mutations, which are referred to as *lineage-defining mutations*. The significance of some mutations have been characterized; however, the vast majority of known mutations are uncharacterized.

A *Variant* is a distinct virus defined by the unique constellation of mutations contained in its genome. Most mutations are unremarkable; however, some variants contain mutations that may alter viral transmissibility, disease severity, or propensity for immune system and/or vaccine escape. Variants that contain these *Mutations of Concern* are candidates for national surveillance, but they have not yet been classified as a Varian of Concern, and thus are not considered reportable.

A *Variant of Concern (VOC)* is a variant associated with an experimentally verified functional change in the virus affecting transmissibility, disease severity, immune escape, vaccine escape, or any other important clinical or epidemiological trait. Because of their increased risk to public health, VOCs have been identified as a priority for surveillance and response. The process for raising the surveillance priority of any variant to that of a VOC has yet to be developed by the broader scientific community, although in general, the process proceeds through a formal investigation, during which the variant is considered a Variant Under Investigation (VUI). A VUI can be designated a VOC depending on the outcome of the investigation[2]. Efforts are underway by international health authorities to establish these processes, and Canada is defining similar standards and processes for defining VOCs, which will be addressed in a subsequent guidance document.

The three currently known VOCs are B.1.1.7, B.1.351, and P.1. The canonical lineage-defining mutations of these VOCs are presented in Table 2.

Table 2: Canonical lineage-defining mutations for known variants of concern

Protein Name	**B.1.1.7**	**B.1.351**	**P.1**
ORF1ab	T1001I, A1708D, I2230T, 3675-3677SGFdel,	K1655N	S1188L, K1795Q, 3675-3677SGFdel, E5662D (synT733C, synC2749T, syn C12778T, synC13860T)
Spike (S)	69-70HVdel, Y144del, **N501Y**, A570D, P681H, T716I, S982A, D1118H	K417N, **E484K**, **N501Y**, **D614G**, A701V	**L18F**, T20N, P26S, D138Y, R190S, K417T, **E484K**, **N501Y**, H655Y, T1027I
ORF8	Q27stop, R52I, Y73C		E92K (ins28269-28273)
Nucleocapsid (N)	D3L, S235F	T205I	P80R
Envelope (E)		P71L	

syn = synonymous genetic mutation
ins = genetic sequence insert
del = amino acid deletion
stop = mutation resulting in a stop codon
bold = mutations common to more than one VOC

Data Standard for Naming Variants

To capture and structure information pertaining to the results and methods used to determine the presence of VOCs in patient samples, a data standard prescribing a set of standardized fields and terms is recommended for recording and communicating results. Encoding this information in a standardized way facilitates and streamlines database queries, data analysis, and reporting. The fields and suggested values, along with their definitions and guidance for usage, are provided in Table 3.

Table 3: Standardized fields and terms for naming variants

Field Name	Definition	Values	Guidance
Lineage name	The lineage name of the virus	B.1.1.7 B.1.351 P.1 Undetermined	Determine the lineage using Pangolin, by assessing lineage-defining mutations, or by VOC-specific RT-qPCR assay. If the technique used cannot distinguish the lineage, use “Undetermined” or leave blank.
Variant designation	The designation used to classify lineages as Variants or Variants of Concern	Variant of Concern Variant	Track whether the lineage is a variant or variant of concern. If the lineage assigned is neither, leave this blank.
Variant evidence	The evidence used to determine the lineage of a Variant of Concern	Free text	List the assay/technique used for testing, and any mutations of concern/interest used as criteria for the lineage/variant designation. If the lineage is not a Variant/VOC, leave blank.

Laboratory Confirmation of Variants

Different techniques offer different thresholds of evidence for characterizing a variant of concern, with WGS offering the highest level of evidence. VOCs were first identified and characterized by WGS, and there is growing evidence that lineages continue to adapt under evolutionary pressure. As such, it is therefore recommended that a VOC be sequenced, when possible, in order to determine its spectrum of mutations. However, laboratories may confirm the presence of a viral variant from complete genome sequence, partial genomic sequence, and from RT-qPCR assays that target one or more variant-defining mutations, each with diminishing

confidence. Genome sequencing and analysis can take days to weeks to complete, while RT-qPCR-based assays can provide results in a much shorter time frame—a saving which can be critical for effective public health response. The definitions, testing criteria, and naming conventions of confirmed VOCs are provided in Table 4.

Confirming Variants

A flowchart depicting the process for confirming variants is provided in Figures 1 and 2.

Confirming Variants From Whole Genome Sequence Data

Whole-genome sequencing provides the strongest evidence for confirming a VOC. The most straightforward method to confirm a variant from WGS data is from its Pango lineage assignment; therefore, **we recommend that VOCs be confirmed from whole genome sequence data from its Pango lineage assignment.**

Confirming Variants From Partial Genome Sequence Data

Pango lineage assignment from a high quality, complete genome is the preferred method for confirming a variant; however, some specimens (e.g., specimens with a high Ct value) may yield an incomplete genome of insufficient quality for Pangolin to assign a lineage. Other partial genome sequencing approaches, such as metagenomic sequencing of wastewater samples, can also be unsuitable for lineage assignment by Pangolin. In these cases, it may be necessary to manually identify a VOC from the mutation data. The minimum set of mutations required to confirm a VOC depends on the power of the mutations in a given VOC to discriminate it from other circulating variants. For example, Pangolin assigns lineages from 5 of the 17 defining B.1.1.7 mutations[3], 5 of the 9 defining B.1.351 mutations[4], and 10 of the 17 defining P.1 mutations[5]. **We recommend that the same number of minimal lineage-defining mutations as those used by Pangolin be identified for each specific variant in order to confirm a VOC.** We also urge caution when using this approach, since variants that are not considered VOCs may share common mutations with VOCs. A good example is the recently emerged B.1.525 variant (see for example EPI_ISL_961609|2021-01-13), which shares the S:69/70del, S:Y144del, and Orf1ab:3675-3677SGFdel in common with the B.1.1.7 variant.

Confirming Variants From From RT-qPCR Assay

Some RT-qPCR Assays are currently being used as a proxy to detect VOCs, the most familiar example being the Thermo Fisher TaqPath 3-gene assay, which exhibits a reliable S-gene target failure due to the presence of the 69-70del mutation in the B.1.1.7 VOC. Other RT-qPCR assays are in development, such as the single-target N501Y SNP assay and multiplex RT-qPCR assays that can detect and discriminate the B.1.1.7, B.1.351, and P.1 VOCs[6].

There is some debate as to whether RT-qPCR assays are confirmatory for a VOC. Some feel that RT-qPCR assays based on the detection of a single target is insufficient to confirm a VOC since it is not definitive for a given variant[7]. Multiplex RT-qPCR assays that detect two or more

mutations in a given VOC can substantially increase its confirmatory power; however, they are not robust to viral evolution. For example, the E484K mutation, which is present in B1.351 and P.1 and is thought to be responsible for immune escape, has recently been detected in a subset of B.1.1.7 VOCs[8] and as of February 10, 2020, is considered a VOC by NERVTAG. Existing RT-qPCR assays for the B.1.1.7 VOC do not discriminate between B.1.1.7 and this evolved variant.

One option for addressing the problems associated with molecular diagnostic detection of VOCs is to label them as “presumptive” if they are detected by RT-qPCR assay, and “confirmed” by follow up genome sequencing. This approach is problematic, however, since the large number of VOCs detected by screening, and the delay in confirming by whole genome sequencing, would result in the initial reporting of many VOCs detected by RT-qPCR assay as “presumptive,” which can create confusion with reporting to public health authorities and the public. To avoid these problems, we recommend not to use the terms “presumptive" and “confirmed” when reporting VOCs. Instead, the evidence used to report the detection of a VOC should be provided in the Variant Evidence field. The evidence can be used to judge the confirmatory power of the method used to to detect the VOC. For this reason, we do not include a field to capture the classification status of VOCs in our interim naming standard.

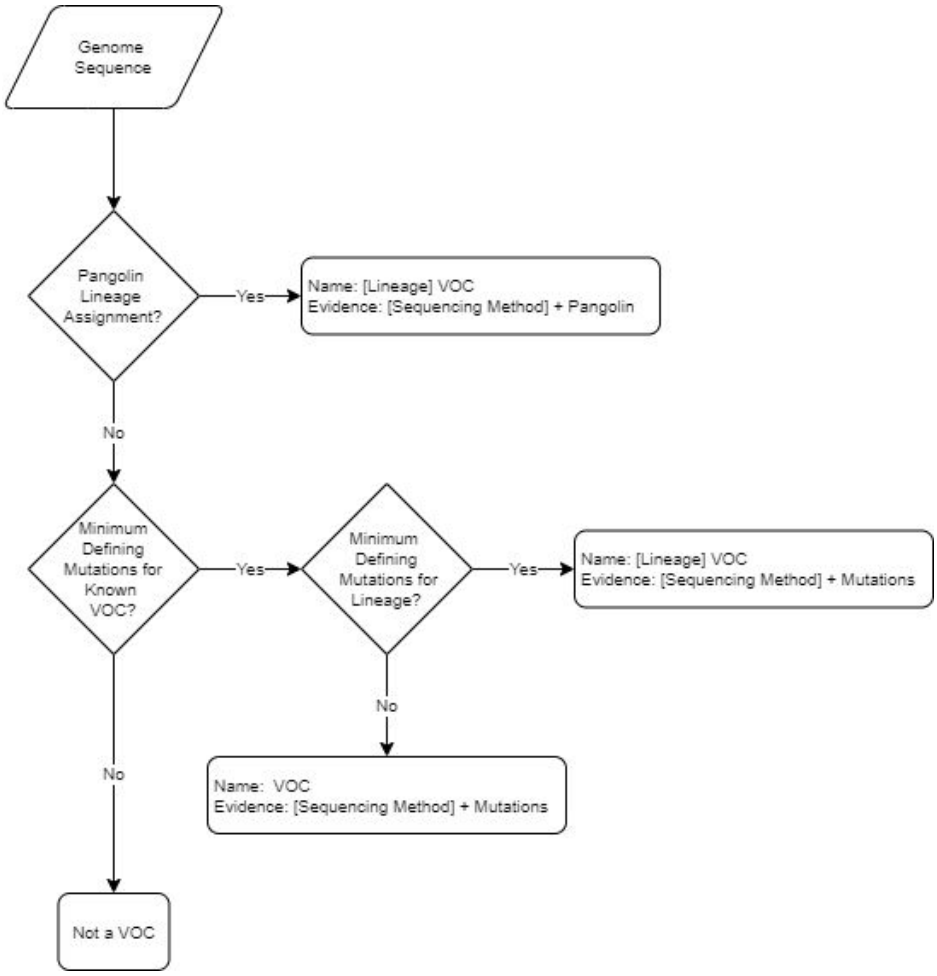

Figure 1. Flowchart for naming VOCs from sequence data.

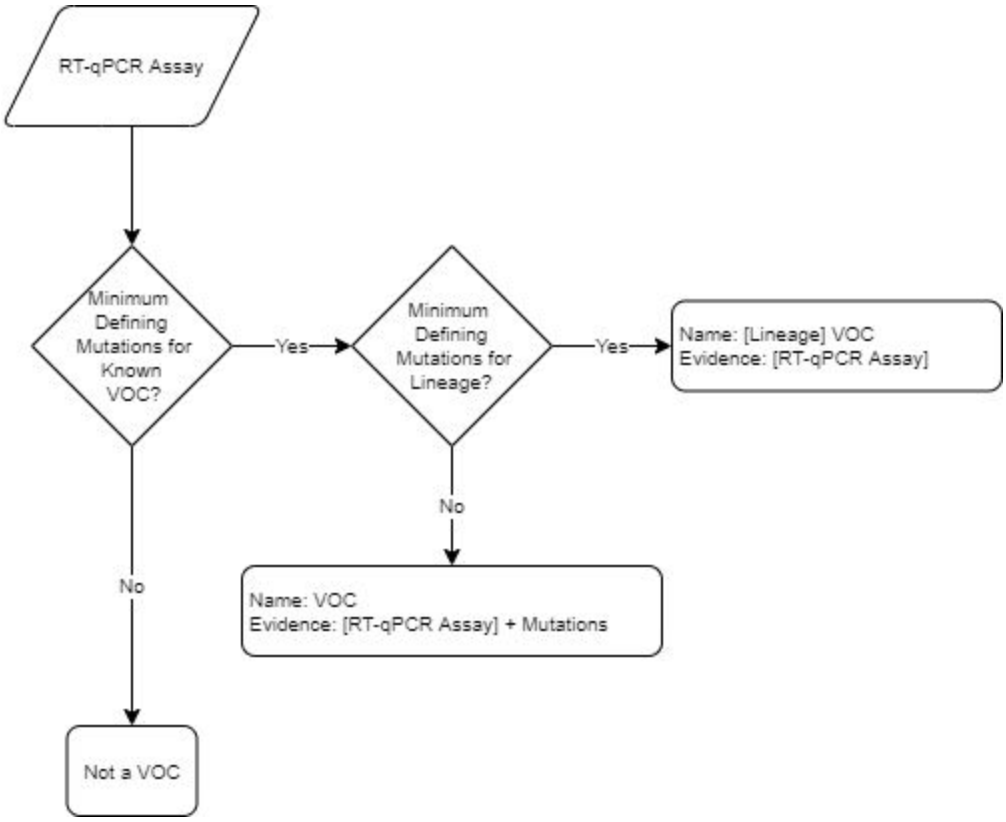

Figure 2. Flowchart for naming VOCs from RT-qPCR assay

Table 4: Definitions, testing criteria and naming conventions of SARS-CoV-2 variants

Nomenclature	Definition	Testing Criteria
B.1.1.7 VOC	A B.1.1.7 VOC confirmed by RT-qPCR assay or sequencing	RT-qPCR screening assay targeting B.1.1.7-specific mutations (e.g., TaqPath RT-qPCR assay) AND/OR Whole or partial genome sequencing; Pango lineage assignment or minimal set of VOC-defining mutations
B.1.351 VOC	A B.1.351 VOC confirmed by RT-qPCR assay or sequencing	RT-qPCR screening assay targeting B.1.351-specific mutations (e.g., A701V) AND/OR Whole or partial genome sequencing;

		Pango lineage assignment or minimal set of VOC-defining mutations
P.1 VOC	A P.1 VOC confirmed by RT-qPCR assay or sequencing	RT-qPCR screening assay targeting a P.1-specific mutation (e.g., P26S) AND/OR Whole or partial genome sequencing; Pango lineage assignment or minimal set of VOC-defining mutations
VOC	A test result containing a single VOC-defining mutation	RT-qPCR screening assay targeting non-discriminatory mutations (e.g., N501Y RT-qPCR assay)
[Lineage] Variant	Any variant of interest containing variant-defining mutations	Whole or partial genome sequencing Pango lineage assignment or minimal set of variant-defining mutations

Worked Examples for Data Standard

We provide here a number of worked examples to illustrate how the data standard should be implemented. A number of hypothetical scenarios are presented below, and the corresponding values for the fields defined by the data standard are shown in Table 5.

Sample Descriptions

Sample 1: The genome was fully sequenced and a lineage of B.1.1.7 assigned using Pangolin. Upon analysis, the genome was found to contain all canonical mutations.
Sample 2: The sample had a Ct value >34 and did not meet the quality threshold for genome sequencing. Targeted sequencing was used to analyze regions of the genome. Five B.1.351 lineage-defining mutations were found to be present.
Sample 3: The genome was fully sequenced and a lineage of B.1.1.7 assigned using Pangolin. The minimal set of variant-defining mutations were identified; however, a few non-lineage defining mutations were also present.
Sample 4: A TaqPath qPCR screen was carried out and all 3 targets indicate the sample is a B.1.1.7 VOC.
Sample 5: A N501Y qPCR screen was carried out and the sample gave a positive result. The sample is presumed to be a VOC but the lineage could not be determined (N501Y is present in all three lineages).
Sample 6: The genome was fully sequenced and a lineage of B.1.429 was assigned. The lineage contains a L452R (Spike) mutation of concern and is thought to be “the California variant,” which is not currently considered a VOC.

Sample 7: The genome was fully sequenced and a lineage of B.1.1.1 was assigned by Pangolin.

Table 5: Worked examples for naming variants

Sample	Lineage name	Variant designation	Variant evidence
Sample 1	B.1.1.7	Variant of Concern	Genome sequencing; Pango lineage assignment
Sample 2	B.1.351	Variant of Concern	Partial genome sequencing; minimal lineage-defining mutations (m_1,m_2...m_n)
Sample 3	B.1.1.7	Variant of Concern	Genome sequencing; Pangol lineage assignment
Sample 4	B.1.1.7	Variant of Concern	TaqPath; SGTF, N, ORF1ab mutations
Sample 5	Undetermined	Variant of Concern	N501Y RT-qPCR screen
Sample 6	B.1.429	Variant	Genome sequencing; Pango lineage assignment; L452R
Sample 7	B.1.1.1	Variant	Genome sequencing; Pango lineage assignment

Reporting Variants

Characterizing the genomic content of a lineage can provide information that can help tease out factors influencing increased viral spread such as types of human behaviour (phylogenetic analysis resolving transmission due close social interactions, travel, transport between hospitals and care facilities, etc.) and viral evolution (mutations resulting in increased transmissibility etc). However, even in the absence of genomic characterization, knowing where and when VOCs are identified is actionable information. As such, federal health authorities recommend that all VOCs be reported. Reporting should be carried out weekly through the submission of the COVID Variants of Concern Report form. Since not all variants of interest are variants of concern, **we recommend that only VOCs be officially reported to jurisdictional and federal health authorities.**

At present, there are no changes to treatment regimens based on VOC determinations. While there is thought to be an elevated risk for transmission and immune escape, there is little evidence to date that suggests an increase in clinical severity for those infected with VOCs. These findings should be kept in mind when communicating VOCs with the public.

Glossary of Terms

Mutation: a change of a nucleotide in the viral RNA genome, or an insertion or deletion event. Some mutations result in an amino acid substitution. Substitutions are denoted by the wildtype amino acid followed by the site in the amino acid sequence and the replacement amino acid (e.g., N501Y denotes and asparagine-to-tyrosine substitution at amino acid site 501). The mutation is sometimes presented with the gene name prepended (e.g., S:N501Y).

Mutation of Interest: a mutation *that may cause a functional change in the virus* affecting transmissibility, disease severity, immune escape, vaccine escape, or any other biological, clinical, or epidemiological trait.

Mutation of Concern: a mutation *associated with a known functional change in the virus* affecting transmissibility, disease severity, immune escape, vaccine escape, or any other biological, clinical, or epidemiological trait.

Variant: a distinct virus defined by the collection of mutations it harbours (the “variant-defining mutations”). Variants adopt the name of the lineage in which they reside (e.g., “a B.1.1.7 variant”).

Variant-Defining Mutation: a non-synonymous substitution or indel found in a variant in addition to the lineage-defining mutations that characterize a given variant.

Variant of Interest: a variant that warrants ongoing surveillance but is otherwise not a variant of concern (e.g., the P.2 variant).

Variant of Concern: a variant *associated with a known functional change in the virus* affecting transmissibility, disease severity, immune escape, vaccine escape, or any other important biological, clinical, or epidemiological trait. Variants of concern are classified as a national and global surveillance priority.

Lineage: a founding variant and its descendants. Lineages are assigned using genomic data and epidemiological data (e.g., “the B.1.1.7 lineage”). We have adopted the Pango lineage nomenclature for naming lineages and variants.

Lineage-Defining Mutation: a shared, common mutation found in at least five genomes of epidemiological significance (“genotypes” with epidemiological significance (geographical prevalence, implicated in outbreaks, etc.). Canonical lineage-defining mutations are assigned using manual and machine-learning approaches.

References

1. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* 1–5 (2020).
2. Coronavirus: The Global COVID-19 Observatory - Science Museum Group. https://www.sciencemuseumgroup.org.uk/blog/coronavirus-the-global-covid-19-observatory

3. B.1.1.7 report. https://cov-lineages.org/global_report_B.1.1.7.html.
4. B.1.351 report. https://cov-lineages.org/global_report_B.1.351.html.
5. P.1 report. https://cov-lineages.org/global_report_P.1.html.
6. Vogels, C. Multiplexed RT-qPCR to screen for SARS-COV-2 B.1.1.7 variants. (2021) doi:10.17504/protocols.io.brcrm2v6.
7. Vogels, C. B. F. *et al.* PCR assay to enhance global surveillance for SARS-CoV-2 variants of concern. *bioRxiv* (2021) doi:10.1101/2021.01.28.21250486.
8. Public Health England. Investigation of novel SARS-CoV-2 variant - Variant of Concern 202012/01. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957504/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5_England.pdf.