

Guidelines

Metadata category	Metadata	Requirement	Description	Syntax	Example value	Specificity to UViGs
General genome metadata	source of UViGs	Mandatory	Type of dataset from which the UViG was obtained	[metagenome (not viral targeted) viral fraction metagenome (virome) sequence-targeted metagenome metatranscriptome (not viral targeted) viral fraction RNA metagenome (RNA virome) sequence-targeted RNA metagenome microbial single amplified genome (SAG) viral single amplified genome (vSAG) isolate microbial genome other]	viral fraction metagenome (virome)	New and specific to UViGs
	assembly software	Mandatory	Tool(s) used for assembly and/or binning, including version number and parameters	{software};{version};{parameters}	metaSPAdes; 3.11.0; kmer set 21,33,55,77,99,121, default parameters otherwise	Identical MIMAG / MISAG
	viral identification software	Mandatory	Tool(s) used for the identification of UViG as a viral genome, software or protocol name including version number, parameters, and cutoffs used (see Table S2)	{software};{version};{parameters}	VirSorter; 1.0.4; Virome database, category 2	New and specific to UViGs
	predicted genome type	Mandatory	Type of genome predicted for the UViG	[DNA dsDNA ssDNA RNA dsRNA ssRNA ssRNA (+) ssRNA (-) mixed uncharacterized]	dsDNA	New and specific to UViGs
	predicted genome structure	Mandatory	Expected structure of the viral genome	[segmented non-segmented undetermined]	non-segmented	New and specific to UViGs
	detection type	Mandatory	Type of UViG detection	[independent sequence (UViG) provirus (UpViG)]	independent sequence (UViG)	New and specific to UViGs
Genome quality	assembly quality	Mandatory	The assembly quality categories, specific for virus genomes, are based on sets of criteria as follows. Finished: Single, validated, contiguous sequence per replicon without gaps or ambiguities, with extensive manual review and editing to annotate putative gene functions and transcriptional units. High-quality draft genome: One or multiple fragments, totalling ≥ 90% of the expected genome or replicon sequence or predicted complete. Genome fragment(s): One or multiple fragments, totalling < 90% of the expected genome or replicon sequence, or for which no genome size could be estimated.	[Finished genome High-quality draft genome Genome fragment(s)]	High-quality draft genome	Comparable to and adapted from MIMAG / MISAG
	number of contigs	Mandatory	Total number of contigs composing the UViG	{number}	1	Identical MIMAG / MISAG
	completeness score	Conditional (required for finished genomes and high-quality draft genomes, optional for other categories)	Estimated completeness of the UViG	{quality};{percentage}	high;92%	Comparable to and adapted from MIMAG / MISAG
	completeness approach	Conditional (required if a completeness estimation is provided)	Approach used to estimate the UViG completeness, including reference genome or group used, and contig feature suggesting a complete genome	{text}	UViG length compared to the average length of reference genomes from the P22virus genus (NCBI RefSeq v83)	Comparable to and adapted from MIMAG / MISAG
Genome annotation	feature prediction	Conditional (required if genome annotation is provided)	Method used to predict UViGs features such as ORFs, integration site, etc.	{software};{version};{parameters}	Prodigal; 2.6.3; default parameters	Comparable to and adapted from MIMAG / MISAG
	reference database(s)	Conditional (required if a viral-specific ORF annotation is provided)	List of database(s) used for ORF annotation, along with version number and reference to website or publication	{database};{version};{reference}	pVOGs; 5; http://dmk-brain.ecn.uiova.edu/pVOGs/ Graziotin et al. 2017 doi:10.1093/nar/gkw975	Comparable to and adapted from MIMAG / MISAG
	similarity search method	Conditional (required if a viral reference database is provided)	Tool used to compare ORFs with database, along with version and cutoffs used	{software};{version};{parameters}	HMMER3; 3.1b2; hmsearch, cutoff of 50 on score	Comparable to and adapted from MIMAG / MISAG
	taxonomic classification	Conditional (required if a taxonomic classification is provided)	Method used for taxonomic classification, along with reference database used, classification rank, and thresholds used to classify new genomes	{text}	e.g. vConTACT vContact2 (references from NCBI RefSeq v83, genus rank classification, default parameters)	Comparable to and adapted from MIMAG / MISAG
	vOTU classification approach	Conditional (required if a vOTU classification is provided)	Cutoffs and approach used when clustering new UViGs in "species-level" vOTUs. Note that results from standard 95% ANI / 85% AF clustering should be provided alongside vOTUs defined from another set of thresholds, even if the latter are the ones primarily used during the analysis.	{ANI cutoff};{AF cutoff};{clustering method}	95% ANI;85% AF; greedy incremental clustering	New and specific to UViGs
	vOTU sequence comparison approach	Conditional (required if a vOTU classification is provided)	Tool and thresholds used to compare sequences when computing "species-level" vOTUs.	{software};{version};{parameters}	blastn; 2.6.0+; e-value cutoff: 0.001	New and specific to UViGs
	vOTU database	Conditional (required if a vOTU classification is provided)	Reference database (i.e. sequences not generated as part of the current study) used to cluster new genomes in "species-level" vOTUs, if any	{database};{version}	NCBI Viral RefSeq; 83	New and specific to UViGs
	host prediction approach	Conditional (required if a predicted host is provided)	Tool or approach used for host prediction	[provirus host sequence similarity CRISPR spacer match kmer similarity co-occurrence combination other]	CRISPR spacer match	New and specific to UViGs
	host prediction estimated accuracy	Conditional (required if a host prediction is provided, except for proviruses)	For each tool or approach used for host prediction, estimated false discovery rates should be included, either computed de novo or from the literature (see Table S4)	{text}	CRISPR spacer match: 0 or 1 mismatches, estimated 8% FDR at the host genus rank (Edwards et al. 2016 doi:10.1093/femsre/fuv048)	New and specific to UViGs
viral SAG metadata	sorting technology	Conditional (required for UViG obtained from vSAGs)	Method used to sort/isolate cells or particles of interest	[flow cytometric cell sorting microfluidics laser-tweezing optical manipulation micromanipulation other]	flow cytometry cell sorting	Comparable to and adapted from MIMAG / MISAG
	single cell or viral particle lysis approach	Conditional (required for UViG obtained from vSAGs)	Method used to free DNA from interior of the cell(s) or particle(s)	[chemical enzymatic physical combination]	chemical	Comparable to and adapted from MIMAG / MISAG
	single cell or viral particle lysis kit protocol	Optional	Name of the kit or standard protocol used for cell(s) or particle(s) lysis	{text}	MagMAX™ Viral RNA Isolation Kit	Comparable to and adapted from MIMAG / MISAG
	WGA amplification approach	Conditional (required for UViG obtained from vSAGs)	Method used to amplify genomic DNA in preparation for sequencing	[pcr based mda based none]	mda based	Identical MIMAG / MISAG
	WGA amplification kit	Optional	Kit used to amplify genomic DNA in preparation for sequencing	{text}	REPL-g Mini Kit	Identical MIMAG / MISAG
viral MAG metadata	size fraction selected	Conditional (required for UViG assembled from metagenomes)	Filtering pore size used in sample preparation	{float}-{float} {unit}	0-0.22 µm	New and specific to UViGs
	virus enrichment approach	Conditional (required for UViG assembled from metagenomes)	List of approaches used to enrich the sample for viruses, if any	[filtration ultrafiltration centrifugation ultracentrifugation PEG Precipitation FeCl Precipitation CsCl density gradient DNase RNase targeted sequence capture other none]	filtration + FeCl Precipitation + ultracentrifugation + DNase	New and specific to UViGs
	nucleic acid extraction	Conditional (required for UViG assembled from metagenomes)	A link to a literature reference, electronic resource or a standard operating procedure (SOP), that describes the material separation to recover the nucleic acid fraction from a sample	{PMID} {DOI} {URL}	10.1111/j.1462-2920.2012.02836.x	Identical MIMAG / MISAG
	WGA amplification approach	Conditional (required for UViG assembled from metagenomes)	Description of the approach used for whole genome amplification, if any	[pcr based mda based none]	none	Identical MIMAG / MISAG
	binning parameters	Conditional (required if genome bin(s) were defined)	The parameters that have been applied during the extraction of genomes from metagenomic datasets	[homology search kmer coverage codon usage combination]	kmer and coverage	Identical MIMAG / MISAG
	binning software	Conditional (required if genome bin(s) were defined)	Tool(s) used for the extraction of genomes from metagenomic datasets	[metabat maxbin concoct groupm esom metawatt combination other]	metabat	Identical MIMAG / MISAG
	reassembly post binning	Conditional (required if genome bin(s) were defined)	Has an assembly been performed on a genome bin extracted from a metagenomic assembly?	[yes no]	yes	Identical MIMAG / MISAG
	MAG coverage software	Optional	Tool(s) used to determine the genome coverage if coverage is used as a binning parameter in the extraction of genomes from metagenomic datasets	[bwa bbmap bowtie other]	bowtie	Identical MIMAG / MISAG

Metadata category	Metadata	Requirement	Description	Syntax	Example value	Specificity to UViGs	UViGs in this study
General genome metadata	source of UViGs	Mandatory	Type of dataset from which the UViG was obtained	[metagenome (not viral targeted) viral fraction metagenome (virome) sequence-targeted metagenome metatranscriptome (not viral targeted) viral fraction RNA metagenome (RNA virome) sequence-targeted RNA metagenome microbial single amplified genome (SAG) viral single amplified genome (vSAG) isolate microbial genome other]	viral fraction metagenome (virome)	New and specific to UViGs	viral fraction metagenome (virome)
	assembly software	Mandatory	Tool(s) used for assembly and/or binning, including version number and parameters	{software};{version};{parameters}	metaSPAdes; 3.11.0; kmer set 21,33,55,77,99,121, default parameters otherwise	Identical MIMAG / MISAG	Flye; v.2.9, --meta
	viral identification software	Mandatory	Tool(s) used for the identification of UViG as a viral genome, software or protocol name including version number, parameters, and cutoffs used (see Table S2)	{software};{version};{parameters}	VirSorter; 1.0.4; Virome database, category 2	New and specific to UViGs	VirSorter2; 2.2.3; --min-score 0.5, --min-length 4000; CheckV; 0.7.0, viral genes > 0 or viral genes = 0 & host genes = 0 or VirSorter2 score = 0.95 or hallmark genes > 2.
	predicted genome type	Mandatory	Type of genome predicted for the UViG	[DNA dsDNA ssDNA RNA dsRNA ssRNA ssRNA (+) ssRNA (-) mixed uncharacterized]	dsDNA	New and specific to UViGs	dsDNA
	predicted genome structure detection type	Mandatory	Expected structure of the viral genome	[segmented non-segmented undetermined]	non-segmented	New and specific to UViGs	undetermined
		Mandatory	Type of UViG detection	[independent sequence (UViG) provirus (UpViG)]	independent sequence (UViG)	New and specific to UViGs	Independent sequence (UViG)
Genome quality	assembly quality	Mandatory	The assembly quality categories, specific for virus genomes, are based on sets of criteria as follows. Finished: Single, validated, contiguous sequence per replicon without gaps or ambiguities, with extensive manual review and editing to annotate putative gene functions and transcriptional units. High-quality draft genome: One or multiple fragments, totaling ≥ 90% of the expected genome or replicon sequence or predicted complete. Genome fragment(s): One or multiple fragments, totalling < 90% of the expected genome or replicon sequence, or for which no genome size could be estimated.	[Finished genome High-quality draft genome Genome fragment(s)]	High-quality draft genome	Comparable to and adapted from MIMAG / MISAG	reported in next sheet
	number of contigs	Mandatory	Total number of contigs composing the UViG	{number}	1	Identical MIMAG / MISAG	1
	completeness score	Conditional (required for finished genomes and high-quality draft genomes, optional for other categories)	Estimated completeness of the UViG	{quality};{percentage}	high;92%	Comparable to and adapted from MIMAG / MISAG	reported in next sheet
	completeness approach	Conditional (required if a completeness estimation is provided)	Approach used to estimate the UViG completeness, including reference genome or group used, and contig feature suggesting a complete genome	{text}	UViG length compared to the average length of reference genomes from the P22virus genus (NCBI RefSeq v83)	Comparable to and adapted from MIMAG / MISAG	reported in next sheet
Genome annotation	feature prediction	Conditional (required if genome annotation is provided)	Method used to predict UViGs features such as ORFs, integration site, etc.	{software};{version};{parameters}	Prodigal; 2.6.3; default parameters	Comparable to and adapted from MIMAG / MISAG	Prodigal; 2.6.3; default parameters
	reference database(s)	Conditional (required if a viral-specific ORF annotation is provided)	List of database(s) used for ORF annotation, along with version number and reference to website or publication	{database};{version};{reference}	pVOGs; 5; http://dmk-brain.ecn.uiova.edu/pVOGs/ / Graziotin et al. 2017 doi:10.1093/nar/gkw975	Comparable to and adapted from MIMAG / MISAG	
	similarity search method	Conditional (required if a viral reference database is provided)	Tool used to compare ORFs with database, along with version and cutoffs used	{software};{version};{parameters}	HMMER3; 3.1b2; hmmsearch, cutoff of 50 on score	Comparable to and adapted from MIMAG / MISAG	
	taxonomic classification	Conditional (required if a taxonomic classification is provided)	Method used for taxonomic classification, along with reference database used, classification rank, and thresholds used to classify new genomes	{text}	e.g. vContact vContact2 (references from NCBI RefSeq v83, genus rank classification, default parameters)	Comparable to and adapted from MIMAG / MISAG	vContact2 (References from NCBI RefSeq v201 and list of Crass-like phages https://doi.org/10.1038/s41467-021-21350-w , family rank classification, default parameters), HMM profile similarities (ViPhOGs database (29/06/2021), family rank classification, significant hits = sequence value <= 0.001 and domain value <= 0.1, >=20% of ORFs with significant hits and >=60% of those hits assigned to the same taxon)
	vOTU classification approach	Conditional (required if a vOTU classification is provided)	Cutoffs and approach used when clustering new UViGs in "species-level" vOTUs. Note that results from standard 95% ANI / 85% AF clustering should be provided alongside vOTUS defined from another set of thresholds, even if the latter are the ones primarily used during the analysis.	{ANI cutoff};{AF cutoff};{clustering method}	95% ANI;85% AF; greedy incremental clustering	New and specific to UViGs	
	vOTU sequence comparison approach	Conditional (required if a vOTU classification is provided)	Tool and thresholds used to compare sequences when computing "species-level" vOTUs.	{software};{version};{parameters}	blastn; 2.6.0+; e-value cutoff: 0.001	New and specific to UViGs	vContact2; 0.9.19; default parameters. Hmmer; 3.3.2; sequence value <= 0.001 & domain value <= 0.1
	vOTU database	Conditional (required if a vOTU classification is provided)	Reference database (i.e. sequences not generated as part of the current study) used to cluster new genomes in "species-level" vOTUs, if any	{database};{version}	NCBI Viral RefSeq; 83	New and specific to UViGs	NCBI Prokaryotic RefSeq; 201 ViPhOGs; (29/06/2021)
	host prediction approach	Conditional (required if a predicted host is provided)	Tool or approach used for host prediction	[provirus host sequence similarity CRISPR spacer match kmer similarity co-occurrence combination other]	CRISPR spacer match	New and specific to UViGs	VirHostMatcher-Net; 1.0
	host prediction estimated accuracy	Conditional (required if a host prediction is provided, except for proviruses)	For each tool or approach used for host prediction, estimated false discovery rates should be included, either computed de novo or from the literature (see Table S4)	{text}	CRISPR spacer match: 0 or 1 mismatches, estimated 8% FDR at the host genus rank (Edwards et al. 2016 doi:10.1093/femsrev/fv048)	New and specific to UViGs	VirHostMatcher-Net; Estimated 10% FDR at the host phylum level; https://doi.org/10.1093/nargab/lqaa044
viral SAG metadata	sorting technology	Conditional (required for UViG obtained from vSAGs)	Method used to sort/isolate cells or particles of interest	[flow cytometric cell sorting microfluidics laser-tweezing optical manipulation micromanipulation other]	flow cytometry cell sorting	Comparable to and adapted from MIMAG / MISAG	
	single cell or viral particle lysis approach	Conditional (required for UViG obtained from vSAGs)	Method used to free DNA from interior of the cell(s) or particle(s)	[chemical enzymatic physical combination]	chemical	Comparable to and adapted from MIMAG / MISAG	
	single cell or viral particle lysis kit protocol	Optional	Name of the kit or standard protocol used for cell(s) or particle(s) lysis	{text}	MagMAX™ Viral RNA Isolation Kit	Comparable to and adapted from MIMAG / MISAG	
	WGA amplification approach	Conditional (required for UViG obtained from vSAGs)	Method used to amplify genomic DNA in preparation for sequencing	[pcr based mda based none]	mda based	Identical MIMAG / MISAG	
WGA amplification kit	Optional	Kit used to amplify genomic DNA in preparation for sequencing	{text}	REPLI-g Mini Kit	Identical MIMAG / MISAG		
viral MAG metadata	size fraction selected	Conditional (required for UViG assembled from metagenomes)	Filtering pore size used in sample preparation	{float}-{float} {unit}	0-0.22 µm	New and specific to UViGs	None
	virus enrichment approach	Conditional (required for UViG assembled from metagenomes)	List of approaches used to enrich the sample for viruses, if any	[filtration ultrafiltration centrifugation ultracentrifugation PEG Precipitation FeCl Precipitation CsCl density gradient DNase RNase targeted sequence capture other none]	filtration + FeCl Precipitation + ultracentrifugation + DNase	New and specific to UViGs	centrifugation + PEG precipitation + chloroform + DNase
	nucleic acid extraction	Conditional (required for UViG assembled from metagenomes)	A link to a literature reference, electronic resource or a standard operating procedure (SOP), that describes the material separation to recover the nucleic acid fraction from a sample	{PMID} {DOI} {URL}	10.1111/j.1462-2920.2012.02836.x	Identical MIMAG / MISAG	
	WGA amplification approach	Conditional (required for UViG assembled from metagenomes)	Description of the approach used for whole genome amplification, if any	[pcr based mda based none]	none	Identical MIMAG / MISAG	
	binning parameters	Conditional (required if genome bin(s) were defined)	The parameters that have been applied during the extraction of genomes from metagenomic datasets	[homology search kmer coverage codon usage combination]	kmer and coverage	Identical MIMAG / MISAG	
	binning software	Conditional (required if genome bin(s) were defined)	Tool(s) used for the extraction of genomes from metagenomic datasets	[metabat maxbin concoct groupm esom metawatt combination other]	metabat	Identical MIMAG / MISAG	
	reassembly post binning	Conditional (required if genome bin(s) were defined)	Has an assembly been performed on a genome bin extracted from a metagenomic assembly?	[yes no]	yes	Identical MIMAG / MISAG	
	MAG coverage software	Optional	Tool(s) used to determine the genome coverage if coverage is used as a binning parameter in the extraction of genomes from metagenomic datasets	[bwa bbmap bowtie other]	bowtie	Identical MIMAG / MISAG	

UViG	VirSorter2 score	Predicted genome type	Provirus	Completeness quality	Completeness (%)	Assembly quality	Completeness approach	Family	Host phylum	Host species
UViG 2	0.973	dsDNAPhage	No	Low-quality	20.32	Genome fragment	AAI-based (high-confidence)			
UViG 3	1	dsDNAPhage	No	Low-quality	46.06	Genome fragment	AAI-based (high-confidence)			
UViG 6	1	dsDNAPhage	No	Low-quality	15.14	Genome fragment	AAI-based (high-confidence)		<i>Actinomycetota</i>	<i>Gardenerella vaginalis</i>
UViG 7	1	dsDNAPhage	No	Medium-quality	79.02	Genome fragment	AAI-based (high-confidence)		<i>Bacteroidota</i>	<i>Bacteroides xylanisolvens</i>
UViG 8	0.973	dsDNAPhage	No	Medium-quality	86.76	Genome fragment	AAI-based (high-confidence)	<i>Siphoviridae</i>		
UViG 10	0.993	dsDNAPhage	No	High-quality	95.86	High-quality draft genome	AAI-based (high-confidence)		<i>Bacillota</i>	<i>Staphylococcus aureus</i>
UViG 12	1	dsDNAPhage	No	High-quality	100	High-quality draft genome	AAI-based (high-confidence)	<i>Microviridae</i>	<i>Bacillota</i>	<i>Staphylococcus aureus</i>