New proposed PlasmoGenEpi targeted amplicon results datafields for PMO format

# Goal

Creating fields with efforts to be consistent with [MIxS standards](https://genomicsstandardsconsortium.github.io/mixs/)

Important aspects to keep in mind are creating an efficient, low-weight, and minimum amount of information about a run without losing any important data. We can build tools around this table to generate certain fields that are important but not necessary always to keep constantly stored in this base class (e.g. SNP/INDEL calls). Also since we are proposing to keep this data in a singular file in [JSON format](https://en.wikipedia.org/wiki/JSON) we are not limited to keeping data in a tabular format for organization; output generated from this file can certainly be a table but we don’t have to store things as a table, e.g. certain fields might be a single ID while other fields might be vectors/lists of data. What’s most important is that we agree on what fields are important and what they should store (and what values should be allowable or formatting, e.g. double/floats vs strings vs characters vs POSIX date vs URL etc)

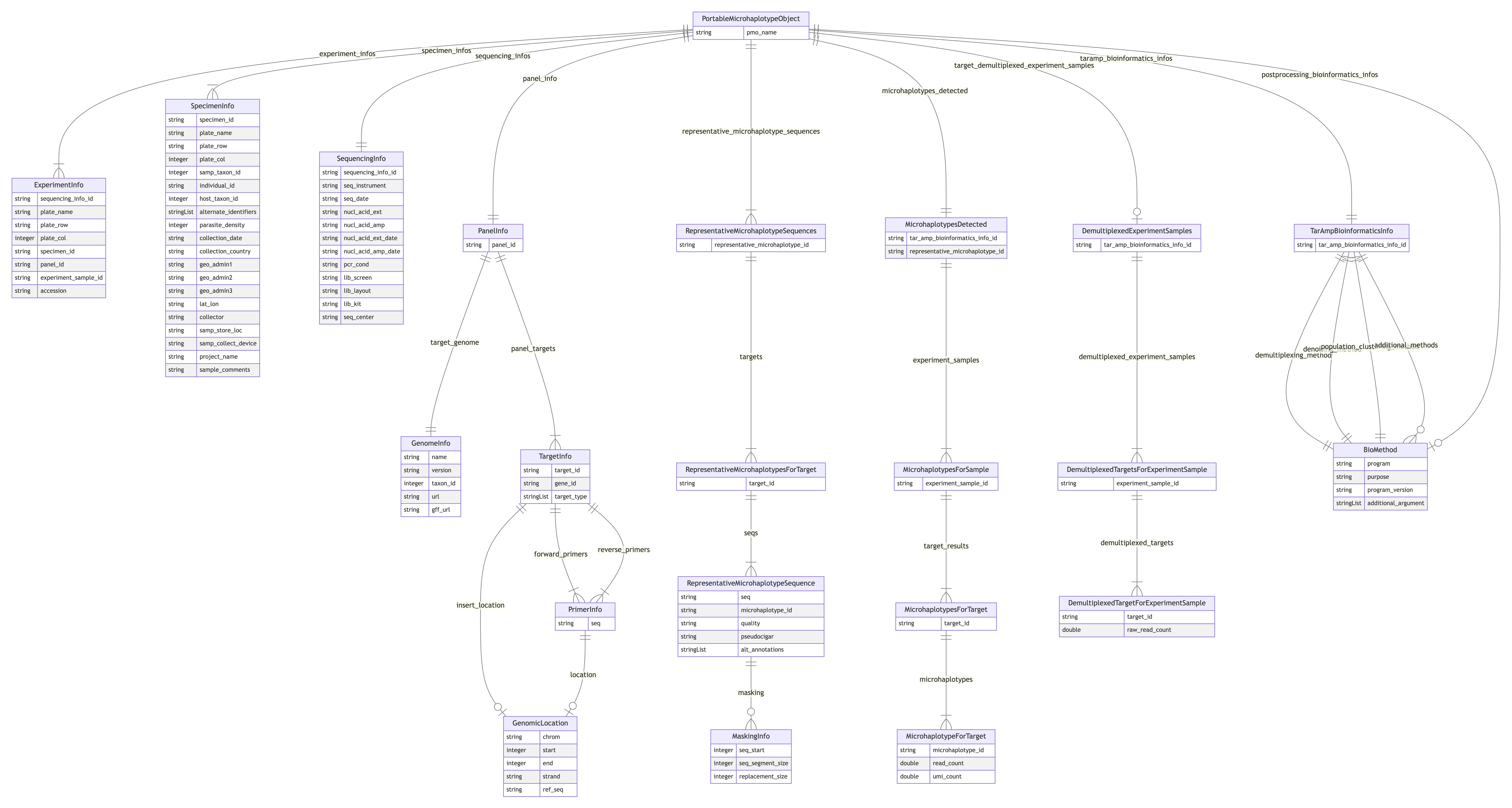
Format is defined by utilizing [LinkML](https://linkml.io/linkml/) to generate a general data scheme which creates convenient outputs like [JSON Schema](https://json-schema.org/) for validation tools.

Other notable users of LinkML/MIxS [National Microbiome Data Collaborative Schema](https://github.com/microbiomedata/nmdc-schema)

# Overview

Below is an overview of the entire format currently in alpha development and is still under heavy development and optimization so structure and names are subject to undergo significant changes as development continues. Please send questions to [info@plasmogenepi.org](mailto:info@plasmogenepi.org)

<https://github.com/PlasmoGenEpi/portable-microhaplotype-object>



## PortableMicrohaplotypeObject

<https://plasmogenepi.github.io/portable-microhaplotype-object/PortableMicrohaplotypeObject/>

### Required

* pmo\_name (***type=string***)
  + a name for this PMO, can be a concatenation of names if combined more than one PMO
* experiment\_infos (***type=***[***ExperimentInfo***](#experimentinfo))
  + a list of experiments of all the seq/amp of samples within this project
* specimen\_infos (***type=***[***SpecimenInfo***](#specimeninfo))
  + a list of SpecimenInfo of all the samples within this project
* sequencing\_infos (***type=***[***SequencingInfo***](#sequencinginfo))
  + the sequencing info for this project
* panel\_info (***type=***[***PanelInfo***](#panelinfo))
  + the info on the targeted sequencing panel used for this project
* representative\_microhaplotype\_sequences (***type=***[***RepresentativeMicrohaplotypeSequences***](#representativemicrohaplotypesequences))
  + a list of the representative sequences for the results for this project
* microhaplotypes\_detected (***type=***[***MicrohaplotypesDetected***](#microhaplotypesdetected))
  + the microhaplotypes detected in this projects
* taramp\_bioinformatics\_infos (***type=***[***TarAmpBioinformaticsInfo***](#tarampbioinformaticsinfo))
  + the bioinformatics pipeline/methods used to generated the amplicon analysis for this project

### Optional

* postprocessing\_bioinformatics\_infos (***type=***[***BioMethod***](#biomethod))
  + any additional methods that were applied to the processing of this file/analysis, this can be filtering, adding info etc
* target\_demultiplexed\_experiment\_samples (***type=***[***DemultiplexedExperimentSamples***](#demultiplexedexperimentsamples))
  + the raw demultiplex target counts for each sample

## TarAmpBioinformaticsInfo

<https://plasmogenepi.github.io/portable-microhaplotype-object/TarAmpBioinformaticsInfo/>

### Required

* tar\_amp\_bioinformatics\_info\_id (***type=string***)
  + a unique identifier for this targeted amplicon bioinformatics pipeline run
* demultiplexing\_method (***type=***[***BioMethod***](#biomethod))
  + the demultiplexing method used to separate raw reads from barcodes and primer targets
* denoising\_method (***type=***[***BioMethod***](#biomethod))
  + the method used to de-noise and/or cluster the raw reads
* population\_clustering\_method (***type=***[***BioMethod***](#biomethod))
  + the method used to compare clustered/de-noised reads across samples for a target

### Optional

* additional\_methods (***type=***[***BioMethod***](#biomethod))
  + any additional methods used to analyze the data

### Example

{  
 "demultiplexing\_method" :   
 {  
 "program" : "SeekDeep extractorPairedEnd",  
 "purpose" : "Takes raw paired-end reads and demultiplexes on primers and does QC filtering",  
 "version" : "v2.6.5"  
 },  
 "denoising\_method" :   
 {  
 "additional\_argument" : "--illumina --qualThres 25,20 --trimFront 1 --trimBack 1",  
 "program" : "SeekDeep qluster",  
 "purpose" : "Takes sequences per sample per target and clusters them",  
 "version" : "v2.6.5"  
 },  
 "population\_clustering\_method" :   
 {  
 "additional\_argument" : "--strictErrors --illumina --removeOneSampOnlyOneOffHaps --excludeCommonlyLowFreqHaplotypes --excludeLowFreqOneOffs --rescueExcludedOneOffLowFreqHaplotypes",  
 "program" : "SeekDeep processClusters",  
 "purpose" : "Compare across samples for each target to create population level identifiers and do post artifact cleanup",  
 "version" : "v2.6.5"  
 },  
 "tar\_amp\_bioinformatics\_info\_id" : "Mozambique2018-SeekDeep"  
 }

## BioMethod

<https://plasmogenepi.github.io/portable-microhaplotype-object/BioMethod/>

### Required

* program (***type=string***)
  + name of the program used for this portion of the pipeline
* purpose (***type=string***)
  + the purpose for this method
* program\_version (***type=string***)
  + versioning info for the program

### Optional

* additional\_argument (***type=array***)
  + any additional arguments that differ from the default

### Example

{  
 "additional\_argument" : "--strictErrors --illumina --removeOneSampOnlyOneOffHaps --excludeCommonlyLowFreqHaplotypes --excludeLowFreqOneOffs --rescueExcludedOneOffLowFreqHaplotypes",  
 "program" : "SeekDeep processClusters",  
 "purpose" : "Compare across samples for each target to create population level identifiers and do post artifact cleanup",  
 "version" : "v2.6.5"  
 },

## DemultiplexedExperimentSamples

<https://plasmogenepi.github.io/portable-microhaplotype-object/DemultiplexedExperimentSamples/>

### Required

* tar\_amp\_bioinformatics\_info\_id (***type=string***)
  + a unique identifier for this targeted amplicon bioinformatics pipeline run
* demultiplexed\_experiment\_samples (***type=***[***DemultiplexedTargetsForExperimentSample***](#demultiplexedtargetsforexperimentsample))
  + a list of the samples with the number of raw reads extracted

### Example

"tar\_amp\_bioinformatics\_info\_id" : "Mozambique2018-SeekDeep",   
 "demultiplexed\_experiment\_samples" :   
 {  
 "1112282540" :   
 {  
 "demultiplexed\_targets" :   
 {  
 "t1" :   
 {  
 "raw\_read\_count" : 34.0,  
 "target\_id" : "t1"  
 },  
 "t10" :   
 {  
 "raw\_read\_count" : 205.0,  
 "target\_id" : "t10"  
 },  
 "t100" :   
 {  
 "raw\_read\_count" : 159.0,  
 "target\_id" : "t100"  
 },

## DemultiplexedTargetForExperimentSample

<https://plasmogenepi.github.io/portable-microhaplotype-object/DemultiplexedTargetForExperimentSample/>

### Required

* target\_id (***type=string***)
  + name of the target
* raw\_read\_count (***type=number***)
  + the raw read counts extracted for a target for a experiment sample

### Example

"t100" :   
 {  
 "raw\_read\_count" : 159.0,  
 "target\_id" : "t100"  
 },

## DemultiplexedTargetsForExperimentSample

<https://plasmogenepi.github.io/portable-microhaplotype-object/DemultiplexedTargetsForExperimentSample/>

### Required

* experiment\_sample\_id (***type=string***)
  + a unique identifier for this sequence/amplification run on a specimen
* demultiplexed\_targets (***type=***[***DemultiplexedTargetForExperimentSample***](#demultiplexedtargetforexperimentsample))
  + a list of the targets extracted for a sample

### Example

"demultiplexed\_targets" :   
 {  
 "t1" :   
 {  
 "raw\_read\_count" : 34.0,  
 "target\_id" : "t1"  
 },  
 "t10" :   
 {  
 "raw\_read\_count" : 205.0,  
 "target\_id" : "t10"  
 },  
 "t100" :   
 {  
 "raw\_read\_count" : 159.0,  
 "target\_id" : "t100"  
 },  
 "t11" :   
 {  
 "raw\_read\_count" : 198.0,  
 "target\_id" : "t11"  
 },  
 "t12" :   
 {  
 "raw\_read\_count" : 19.0,  
 "target\_id" : "t12"  
 },  
 "t13" :   
 {  
 "raw\_read\_count" : 66.0,  
 "target\_id" : "t13"  
 },  
 "experiment\_sample\_id" : "1112282540"

## MicrohaplotypeForTarget

<https://plasmogenepi.github.io/portable-microhaplotype-object/MicrohaplotypeForTarget/>

### Required

* microhaplotype\_id (***type=string***)
  + name of the microhaplotype, should be unique to this microhaplotype
* read\_count (***type=number***)
  + the read count associated with this microhaplotype

### Optional

* umi\_count (***type=number***)
  + the unique molecular identifier (umi) count associated with this microhaplotype

### Example

{  
 "microhaplotype\_id" : "t1.0",  
 "read\_count" : 11600.0  
 }

## MicrohaplotypesDetected

<https://plasmogenepi.github.io/portable-microhaplotype-object/MicrohaplotypesDetected/>

### Required

* tar\_amp\_bioinformatics\_info\_id (***type=string***)
  + a unique identifier for this targeted amplicon bioinformatics pipeline run
* representative\_microhaplotype\_id (***type=string***)
  + an identifier for the representative microhaplotype object collection
* experiment\_samples (***type=***[***MicrohaplotypesForSample***](#microhaplotypesforsample))
  + a list of the microhaplotypes detected for a sample for various targets

### Example

"representative\_microhaplotype\_id" : "Mozambique2018-SeekDeep",  
 "tar\_amp\_bioinformatics\_info\_id" : "Mozambique2018-SeekDeep",   
 "experiment\_samples" :   
 {  
 "8025874217" :   
 {  
 "experiment\_sample\_id" : "8025874217",  
 "target\_results" :   
 {  
 "t1" :   
 {  
 "microhaplotypes" :   
 [  
 {  
 "microhaplotype\_id" : "t1.2",  
 "read\_count" : 34463.0  
 },  
 {  
 "microhaplotype\_id" : "t1.0",  
 "read\_count" : 11600.0  
 }  
 ],  
 "target\_id" : "t1"  
 },  
 "t10" :   
 {  
 "microhaplotypes" :   
 [  
 {  
 "microhaplotype\_id" : "t10.0",  
 "read\_count" : 49728.0  
 }  
 ],  
 "target\_id" : "t10"  
 },  
 "t100" :   
 {  
 "microhaplotypes" :   
 [  
 {  
 "microhaplotype\_id" : "t100.05",  
 "read\_count" : 49740.0  
 }  
 ],  
 "target\_id" : "t100"  
 },

## MicrohaplotypesForSample

<https://plasmogenepi.github.io/portable-microhaplotype-object/MicrohaplotypesForSample/>

### Required

* experiment\_sample\_id (***type=string***)
  + a unique identifier for this sequence/amplification run on a specimen
* target\_results (***type=***[***MicrohaplotypesForTarget***](#microhaplotypesfortarget))
  + a list of the microhaplotypes detected for a list of targets

### Example

{  
 "experiment\_sample\_id" : "8025874217",  
 "target\_results" :   
 {  
 "t1" :   
 {  
 "microhaplotypes" :   
 [  
 {  
 "microhaplotype\_id" : "t1.2",  
 "read\_count" : 34463.0  
 },  
 {  
 "microhaplotype\_id" : "t1.0",  
 "read\_count" : 11600.0  
 }  
 ],  
 "target\_id" : "t1"  
 },  
 "t10" :   
 {  
 "microhaplotypes" :   
 [  
 {  
 "microhaplotype\_id" : "t10.0",  
 "read\_count" : 49728.0  
 }  
 ],  
 "target\_id" : "t10"  
 },  
 "t100" :   
 {  
 "microhaplotypes" :   
 [  
 {  
 "microhaplotype\_id" : "t100.05",  
 "read\_count" : 49740.0  
 }  
 ],  
 "target\_id" : "t100"  
 },

## MicrohaplotypesForTarget

<https://plasmogenepi.github.io/portable-microhaplotype-object/MicrohaplotypesForTarget/>

### Required

* target\_id (***type=string***)
  + name of the target
* microhaplotypes (***type=***[***MicrohaplotypeForTarget***](#microhaplotypefortarget))
  + a list of the microhaplotypes detected for this target

### Example

{  
 "microhaplotypes" :   
 [  
 {  
 "microhaplotype\_id" : "t1.2",  
 "read\_count" : 34463.0  
 },  
 {  
 "microhaplotype\_id" : "t1.0",  
 "read\_count" : 11600.0  
 }  
 ],  
 "target\_id" : "t1"  
 },

## PanelInfo

<https://plasmogenepi.github.io/portable-microhaplotype-object/PanelInfo/>

### Required

* panel\_id (***type=string***)
  + name of the panel
* target\_genome (***type=***[***GenomeInfo***](#genomeinfo))
  + the info on the target reference genome for this panel
* panel\_targets (***type=***[***TargetInfo***](#targetinfo))
  + a list of the target infos for the targets in this panel

### Example

"panel\_info" :   
 {  
 "panel\_id" : "heomev1",  
 "target\_genome" :   
 {  
 "gff\_url" : "https://plasmodb.org/common/downloads/release-65/Pfalciparum3D7/gff/data/PlasmoDB-65\_Pfalciparum3D7.gff",  
 "name" : "3D7",  
 "taxon\_id" : 5833,  
 "url" : "https://plasmodb.org/common/downloads/release-65/Pfalciparum3D7/fasta/data/PlasmoDB-65\_Pfalciparum3D7\_Genome.fasta",  
 "version" : "2020-09-01"  
 },  
 "targets" :   
 {  
 "t1" :   
 {  
 "forward\_primers" :   
 [  
 {  
 "location" :   
 {  
 "chrom" : "Pf3D7\_01\_v3",  
 "end" : 145449,  
 "start" : 145416,  
 "strand" : "+"  
 },  
 "seq" : "TGTTCGATATGTTTAAATATATGATTCTCGAAA"  
 }  
 ],  
 "gene\_id" : "PF3D7\_0103300",  
 "insert\_location" :   
 {  
 "chrom" : "Pf3D7\_01\_v3",  
 "end" : 145622,  
 "start" : 145449,  
 "strand" : "+"  
 },  
 "reverse\_primers" :   
 [  
 {  
 "location" :   
 {  
 "chrom" : "Pf3D7\_01\_v3",  
 "end" : 145449,  
 "start" : 145416,  
 "strand" : "+"  
 },  
 "seq" : "CCAATATGTCAAGGTATATTAAAGTATGGTATC"  
 }  
 ],  
 "target\_id" : "t1"  
 },  
 "t10" :   
 {  
 "forward\_primers" :   
 [  
 {  
 "location" :   
 {  
 "chrom" : "Pf3D7\_02\_v3",  
 "end" : 109807,  
 "start" : 109776,  
 "strand" : "+"  
 },  
 "seq" : "CCACCATTTCTTCATTTTAATTTTGAATGGT"  
 }  
 ],  
 "gene\_id" : "PF3D7\_0202100",  
 "insert\_location" :   
 {  
 "chrom" : "Pf3D7\_02\_v3",  
 "end" : 109982,  
 "start" : 109807,  
 "strand" : "+"  
 },  
 "reverse\_primers" :   
 [  
 {  
 "location" :   
 {  
 "chrom" : "Pf3D7\_02\_v3",  
 "end" : 109807,  
 "start" : 109776,  
 "strand" : "+"  
 },  
 "seq" : "ACCATTTGGATTAAAACCTTCAGATTTAAATA"  
 }  
 ],  
 "target\_id" : "t10"  
 },

## TargetInfo

<https://plasmogenepi.github.io/portable-microhaplotype-object/TargetInfo/>

### Required

* target\_id (***type=string***)
  + name of the target
* forward\_primers (***type=***[***PrimerInfo***](#primerinfo))
  + A list of forward primers associated with this target
* reverse\_primers (***type=***[***PrimerInfo***](#primerinfo))
  + A list of reverse primers associated with this target

### Optional

* gene\_id (***type=string***)
  + an identifier of the gene, if any, is being covered with this targeted
* insert\_location (***type=***[***GenomicLocation***](#genomiclocation))
  + the intended genomic location of the insert of the amplicon (the location between the end of the forward primer and the beginning of the reverse primer)
* target\_type (***type=array***)
  + a list of classification type for the primer target

### Example

{  
 "forward\_primers" :   
 [  
 {  
 "location" :   
 {  
 "chrom" : "Pf3D7\_01\_v3",  
 "end" : 145449,  
 "start" : 145416,  
 "strand" : "+"  
 },  
 "seq" : "TGTTCGATATGTTTAAATATATGATTCTCGAAA"  
 }  
 ],  
 "gene\_id" : "PF3D7\_0103300",  
 "insert\_location" :   
 {  
 "chrom" : "Pf3D7\_01\_v3",  
 "end" : 145622,  
 "start" : 145449,  
 "strand" : "+"  
 },  
 "reverse\_primers" :   
 [  
 {  
 "location" :   
 {  
 "chrom" : "Pf3D7\_01\_v3",  
 "end" : 145449,  
 "start" : 145416,  
 "strand" : "+"  
 },  
 "seq" : "CCAATATGTCAAGGTATATTAAAGTATGGTATC"  
 }  
 ],  
 "target\_id" : "t1"  
 },

## PrimerInfo

<https://plasmogenepi.github.io/portable-microhaplotype-object/PrimerInfo/>

### Required

* seq (***type=string***)
  + the DNA sequence

### Optional

* location (***type=***[***GenomicLocation***](#genomiclocation))
  + what the intended genomic location of the primer is

### Example

{  
 "location" :   
 {  
 "chrom" : "Pf3D7\_01\_v3",  
 "end" : 145449,  
 "start" : 145416,  
 "strand" : "+"  
 },  
 "seq" : "TGTTCGATATGTTTAAATATATGATTCTCGAAA"  
 }

## MaskingInfo

<https://plasmogenepi.github.io/portable-microhaplotype-object/MaskingInfo/>

### Required

* seq\_start (***type=integer***)
  + the start of the masking
* seq\_segment\_size (***type=integer***)
  + the size of the masking
* replacement\_size (***type=integer***)
  + the size of replacement mask

### Example

[  
 {  
 seq\_start : 10,   
 seq\_segment\_size : 5,   
 replacement\_size : 3   
 },   
 {  
 seq\_start : 45,   
 seq\_segment\_size : 7,   
 replacement\_size : 7  
 },   
]

## GenomeInfo

<https://plasmogenepi.github.io/portable-microhaplotype-object/GenomeInfo/>

### Required

* name (***type=string***)
  + name of the genome
* version (***type=string***)
  + the genome version
* taxon\_id (***type=integer***)
  + the NCBI taxonomy number
* url (***type=string***)
  + a link to the where this genome file could be downloaded

### Optional

* gff\_url (***type=string***)
  + a link to the where this genome’s annotation file could be downloaded

### Example

{  
 "gff\_url" : "https://plasmodb.org/common/downloads/release-65/Pfalciparum3D7/gff/data/PlasmoDB-65\_Pfalciparum3D7.gff",  
 "name" : "3D7",  
 "taxon\_id" : 5833,  
 "url" : "https://plasmodb.org/common/downloads/release-65/Pfalciparum3D7/fasta/data/PlasmoDB-65\_Pfalciparum3D7\_Genome.fasta",  
 "version" : "2020-09-01"  
 },

## GenomicLocation

<https://plasmogenepi.github.io/portable-microhaplotype-object/GenomicLocation/>

### Required

* chrom (***type=string***)
  + the chromosome name
* start (***type=integer***)
  + the start of the location, 0-based positioning
* end (***type=integer***)
  + the end of the location, 0-based positioning

### Optional

* ref\_seq (***type=string***)
  + the reference sequence of this genomic location
* strand (***type=string***)
  + which strand the location is, either + for plus strand or - for negative strand

### Example

{  
 "chrom" : "Pf3D7\_01\_v3",  
 "end" : 145449,  
 "start" : 145416,  
 "strand" : "+"  
 },

## RepresentativeMicrohaplotypeSequences

<https://plasmogenepi.github.io/portable-microhaplotype-object/RepresentativeMicrohaplotypeSequences/>

### Required

* representative\_microhaplotype\_id (***type=string***)
  + an identifier for the representative microhaplotype object collection
* targets (***type=***[***RepresentativeMicrohaplotypesForTarget***](#representativemicrohaplotypesfortarget))
  + a list of the targets detected for this analysis

### Example

"representative\_microhaplotype\_id" : "Mozambique2018-SeekDeep",  
 "targets" :   
 {  
 "t1" :   
 {  
 "seqs" :   
 {  
 "t1.0" :   
 {  
 "microhaplotype\_id" : "t1.0",  
 "seq" : "AACTTTTTTTATTTTTTTTGTCAATAGATAAATGATCAATATTTTCTATATTTAATCTATCAAGTATTTTTATATATCTATTATTTCTTTCTTCGATGGATAAATTATAAGAATCAATATCCTTTCTTTCATCAACAAACTTTTTTATTGTTAACTCCATTTTTTTATTTA"  
 },  
 "t1.1" :   
 {  
 "microhaplotype\_id" : "t1.1",  
 "seq" : "AACTTTTTTTATTTTTTTTGTCAATAGATAAATGATCAATATTTTCTATATTTAATCTATCAAGAATTTTTATATATCTATTATTTCTTTCTTCGATGGATAAATTATATGAATCAATATCCTTTCTTTCATCAACAAACTTTTTTATTGTTAACTCCATTTTTTTATTTA"  
 },

## RepresentativeMicrohaplotypesForTarget

<https://plasmogenepi.github.io/portable-microhaplotype-object/RepresentativeMicrohaplotypesForTarget/>

### Required

* target\_id (***type=string***)
  + name of the target
* seqs (***type=***[***RepresentativeMicrohaplotypeSequence***](#representativemicrohaplotypesequence))
  + a list of the microhaplotypes detected for a target

### Example

{  
 "seqs" :   
 {  
 "t1.0" :   
 {  
 "microhaplotype\_id" : "t1.0",  
 "seq" : "AACTTTTTTTATTTTTTTTGTCAATAGATAAATGATCAATATTTTCTATATTTAATCTATCAAGTATTTTTATATATCTATTATTTCTTTCTTCGATGGATAAATTATAAGAATCAATATCCTTTCTTTCATCAACAAACTTTTTTATTGTTAACTCCATTTTTTTATTTA"  
 },  
 "t1.1" :   
 {  
 "microhaplotype\_id" : "t1.1",  
 "seq" : "AACTTTTTTTATTTTTTTTGTCAATAGATAAATGATCAATATTTTCTATATTTAATCTATCAAGAATTTTTATATATCTATTATTTCTTTCTTCGATGGATAAATTATATGAATCAATATCCTTTCTTTCATCAACAAACTTTTTTATTGTTAACTCCATTTTTTTATTTA"  
 },  
 "t1.2" :   
 {  
 "microhaplotype\_id" : "t1.2",  
 "seq" : "AACTTTTTTTATTTTTTTTGTCAATAGATAAATGATCAATATTTTCTATATTTAATCTATCAAGTATTTTTATATATCTATTATTTCTTTCTTCGATGGATAAATTATATGAATCAATATCCTTTCTTTCATCAACAAACTTTTTTATTGTTAACTCCATTTTTTTATTTA"  
 },  
 "t1.3" :   
 {  
 "microhaplotype\_id" : "t1.3",  
 "seq" : "AACTTTTTTTATTTTTTTTGTCAATAGATAAATGATCAATATTTTCTATATTTAATCTATCAAGAATTTTTATATATCTATTATTTCTTTCTTCGATGGATAAATTATAAGAATCAATATCCTTTCTTTCATCAACAAACTTTTTTATTGTTAACTCCATTTTTTTATTTA"  
 },  
 "t1.4" :   
 {  
 "microhaplotype\_id" : "t1.4",  
 "seq" : "AACTTTTTTTATTTTTTTTGTCAATAGATAAATGATCAATATTTTCTATATTTAATCTATCAAGTATTTTTATATATCTATTATTTCTTTCTTCGATGGATAAATTATAAGAATCAATATCCTTTCTTTCATCAACAAACTTTTTTATGGTTAACTCCATTTTTTTATTTA"  
 },  
 "t1.5" :   
 {  
 "microhaplotype\_id" : "t1.5",  
 "seq" : "AACTTTTTTTATTTTTTTTGTCAATAGATAAATGATCAATATTTTCTATATTTAATCTATCAAGAATTTTTATATATCTATTATTTCTTTCTTCGAAGGATAAATTATAAGAATCAATATCCTTTCTTTCATCAACAAACTTTTTTATTGTTAACTCCATTTTTTTATTTA"  
 }  
 },  
 "target\_id" : "t1"  
 },

## RepresentativeMicrohaplotypeSequence

<https://plasmogenepi.github.io/portable-microhaplotype-object/RepresentativeMicrohaplotypeSequence/>

### Required

* seq (***type=string***)
  + the DNA sequence
* microhaplotype\_id (***type=string***)
  + name of the microhaplotype, should be unique to this microhaplotype

### Optional

* alt\_annotations (***type=array***)
  + a list of additional annotations associated with this microhaplotype, e.g. wildtype, amino acid changes etc
* masking (***type=***[***MaskingInfo***](#maskinginfo))
  + masking info for the sequence
* pseudocigar (***type=string***)
  + the pseudocigar of the haplotype
* quality (***type=string***)
  + the ansi fastq per base quality score for this sequence, this is optional

### Example

{  
 "microhaplotype\_id" : "t1.0",  
 "seq" : "AACTTTTTTTATTTTTTTTGTCAATAGATAAATGATCAATATTTTCTATATTTAATCTATCAAGTATTTTTATATATCTATTATTTCTTTCTTCGATGGATAAATTATAAGAATCAATATCCTTTCTTTCATCAACAAACTTTTTTATTGTTAACTCCATTTTTTTATTTA"  
 },

## SequencingInfo

<https://plasmogenepi.github.io/portable-microhaplotype-object/SequencingInfo/>

### Required

* sequencing\_info\_id (***type=string***)
  + a unique identifier for this sequencing info
* seq\_instrument (***type=string***)
  + the sequencing instrument used to sequence the run, e.g. ILLUMINA, Illumina MiSeq
* seq\_date (***type=string***)
  + the date of sequencing, should be YYYY-MM or YYYY-MM-DD
* nucl\_acid\_ext (***type=string***)
  + Link to a reference or kit that describes the recovery of nucleic acids from the sample
* nucl\_acid\_amp (***type=string***)
  + Link to a reference or kit that describes the enzymatic amplification of nucleic acids,
* nucl\_acid\_ext\_date (***type=string***)
  + the date of the nucleoacide extraction
* nucl\_acid\_amp\_date (***type=string***)
  + the date of the nucleoacide amplification
* pcr\_cond (***type=string***)
  + the method/conditions for PCR, List PCR cycles used to amplify the target
* lib\_screen (***type=string***)
  + Describe enrichment, screening, or normalization methods applied during amplification or library preparation, e.g. size selection 390bp, diluted to 1 ng DNA/sample
* lib\_layout (***type=string***)
  + Specify the configuration of reads, e.g. paired-end
* lib\_kit (***type=string***)
  + Name, version, and applicable cell or cycle numbers for the kit used to prepare libraries and load cells or chips for sequencing. If possible, include a part number, e.g. MiSeq Reagent Kit v3 (150-cycle), MS-102-3001
* seq\_center (***type=string***)
  + Name of facility where sequencing was performed (lab, core facility, or company)

### Example

{  
 "lib\_kit" : "TruSeq i5/i7 barcode primers",  
 "lib\_layout" : "paired-end",  
 "lib\_screen" : "40 µL reaction containing 10 µL of bead purified digested product, 18μL of nuclease-free water, 8μL of 5X secondary PCR master mix, and 5 µL of 10 µM TruSeq i5/i7 barcode primers",  
 "nucl\_acid\_amp" : "https://www.paragongenomics.com/targeted-sequencing/amplicon-sequencing/cleanplex-ngs-amplicon-sequencing/",  
 "nucl\_acid\_date" : "2019-07-15",  
 "nucl\_acid\_ext" : "https://www.paragongenomics.com/targeted-sequencing/amplicon-sequencing/cleanplex-ngs-amplicon-sequencing/",  
 "pcr\_cond" : "10 min at 95°C, 13 cycles for high density samples (or 15 cycles for low density samples) of 15 sec at 98°C and 75 sec at 60°C",  
 "seq\_center" : "UCSF",  
 "seq\_date" : "2019-07-15",  
 "seq\_instrument" : "NextSeq 550 instrument",  
 "sequencing\_info\_id" : "Mozambique2018"  
 }

## ExperimentInfo

<https://plasmogenepi.github.io/portable-microhaplotype-object/ExperimentInfo/>

### Required

* sequencing\_info\_id (***type=string***)
  + a unique identifier for this sequencing info
* specimen\_id (***type=string***)
  + the name of the specimen of a individual
* panel\_id (***type=string***)
  + name of the panel
* experiment\_sample\_id (***type=string***)
  + a unique identifier for this sequence/amplification run on a specimen

### Optional

* accession (***type=string***)
  + ERA/SRA accession number for the sample if it was submitted
* plate\_col (***type=integer***)
  + the column the specimen was in
* plate\_name (***type=string***)
  + a name of plate the specimen was in
* plate\_row (***type=string***)
  + the row the specimen was in

### Example

{  
 "experiment\_sample\_id" : "8025874217",  
 "panel\_id" : "heomev1",  
 "plate\_col" : 12,  
 "plate\_name" : "8",  
 "plate\_row" : "C",  
 "sequencing\_info\_id" : "Mozambique2018",  
 "specimen\_id" : "8025874217"  
 },

## SpecimenInfo

<https://plasmogenepi.github.io/portable-microhaplotype-object/SpecimenInfo/>

### Required

* specimen\_id (***type=string***)
  + the name of the specimen of a individual
* samp\_taxon\_id (***type=integer***)
  + the NCBI taxonomy number of the organism of interest
* collection\_date (***type=string***)
  + the date of the sample collection
* collection\_country (***type=string***)
  + the name of country collected in, would be the same as admin level 0
* collector (***type=string***)
  + the name of the primary person managing the specimen
* samp\_store\_loc (***type=string***)
  + the sample store site, address or facility name
* samp\_collect\_device (***type=string***)
  + the way the sample was collected, e.g. whole blood, dried blood spot, etc
* project\_name (***type=string***)
  + a name of the project under which the sample is organized

### Optional

* alternate\_identifiers (***type=array***)
  + a list of optional alternative names for the samples
* geo\_admin1 (***type=string***)
  + geographical admin level 1, the secondary large demarcation of a nation (nation = admin level 0)
* geo\_admin2 (***type=string***)
  + geographical admin level 2, the third large demarcation of a nation (nation = admin level 0)
* geo\_admin3 (***type=string***)
  + geographical admin level 3, the third large demarcation of a nation (nation = admin level 0)
* host\_taxon\_id (***type=integer***)
  + optional the NCBI taxonomy number of the host of the organism
* individual\_id (***type=string***)
  + an identifier for the individual a specimen was collected from
* lat\_lon (***type=string***)
  + the latitude and longitude of the collection site of the specimen
* parasite\_density (***type=integer***)
  + the parasite density in microliters
* plate\_col (***type=integer***)
  + the column the specimen was in
* plate\_name (***type=string***)
  + a name of plate the specimen was in
* plate\_row (***type=string***)
  + the row the specimen was in
* sample\_comments (***type=string***)
  + any additional comments about the sample

### Example

{  
 "collection\_country" : "Mozambique",  
 "collection\_date" : "2018-06-07",  
 "collector" : "Greenhouse, Bryan",  
 "geo\_admin3" : "Inhassoro",  
 "host\_taxon\_id" : 1758,  
 "lat\_lon" : "-21.5535,35.1819",  
 "parasite\_density" : 477719.34375,  
 "plate\_col" : 12,  
 "plate\_name" : "8",  
 "plate\_row" : "C",  
 "project\_name" : "MOZ2018",  
 "samp\_collect\_device" : "dried blood spot",  
 "samp\_store\_loc" : "UCSF Greenhouse Lab",  
 "samp\_taxon\_id" : 5833,  
 "specimen\_id" : "8025874217"  
 },