**MIGS Version 1.1**

|  |
| --- |
| **INVESTIGATION**  **1 = strong support, 2 = discuss further, 3 =drop, 4= EMBL to take forward as future INSDC qualifier, CV = controlled vocabulary** |
| **Taxa** | | | | | | | | | | | |
| **EU** | | **BA** | | **PL** | | **VI** | | **OR** | | **ME** | |
| ***Study*** |
| **ORGANISM** |
| Complete genetic lineage (below lowest rank of NCBI taxonomy)**1, CV** | M | | M | | M | | M | | M | | - | |
| Ploidy level **1, CV** | M | | - | |  | |  | |  | |  | |
| Number of replicons (nuclear genome: chromosomes, Virus: number of segments) (refers to haploid chromosome count)**1, integer** | M | | M | | - | | M | | - | | - | |
| Extrachromosomal elements **1, integer** | X | | M | |  | |  | |  | |  | |
| Estimated size (prior to sequencing; to apply to all draft genomes) **1, integer** | M | | X | | X | | X | | X | | - | |
| Reference for biomaterial (primary publication if isolated prior to genome publication) **1, PMID or DOI** | X | | M | | X | | X | | X | | X | |
| Source material identifiers: (cultures of micro-organisms: identifiers **alphanumeric** for two culture collections **CV**, specimens (e.g. organelles and eukarya): voucher condition and location **CV**) **1,2** | M | | M | | M | | M | | M | | M | |
| Biotic Relationship (e.g. free-living, pathogen, commensal, symbiont etc) **1,2, CV** | Xhisve edhe option to input domv | | M | |  | | X | |  | |  | |
| Specific Host – (host taxid, unknown, or environmental; laboratory or natural host or both) **1, CV** | X | | M | | M | | M | |  | |  | |
| Host specificity/range **2 with a view to 3? taxid** | X | | X | | X | | M | |  | |  | |
| Health/disease status of specific host at time of collection **1,4, Phenotype ontology** |  | | M | |  | | M | |  | |  | |
| Whether normally pathogenic or not **2** | X | | X | |  | | M | |  | |  | |
| Trophic level **2 , CV** | M | | M | | - | | - | | - | | - | |
| Estimated community diversity and abundances of specific taxonomic groups **1, 3 (strong support, but perhaps too ambiguous to define well enough to make useful)** | - | | - | | - | | - | | - | | M | |
| **PHENOTYPE** |
| Propagation (Phage: lytic/lysogenic: Plasmids: incompatibility group) – **1,2, CV** | M | | M | | M | | M | | - | | - | |
| Encoded traits (e.g plasmid=antibiotic resistance, phage= converting genes) **1,2, CV, GO** |  | | X | | M | | M | |  | | X | |
| Relationship to oxygen (e.g. aerobic, anaerobic etc) **1,2, CV** |  | | M | | - | | - | | - | | - | |
| **ENVIRONMENT** |
| Geographic location (latitude and longitude **float**, depth / altitude of sample **?**) **1,** | M | | M | | M | | M | | M | | M | |
| Date and time of sample collection **1** | M | | M | | M | | M | | M | | M | |
| Habitat type **1, CV** | M | | M | | M | | M | | M | | M | |
| **SAMPLE PROCESSING** |
| Isolation and Growth conditions **1,2, PMID or DOI** |  | | M | |  | |  | |  | |  | |
| Volume of sample **1,2, integer** |  | |  | |  | |  | |  | | M | |
| Sampling strategy (enriched, screened, normalized) **1, CV** |  | |  | |  | |  | |  | | M | |
| Nucleic acid preparation (extraction method **CV** ; amplification *e.g.* MDA, emPCR, etc **CV**) **1** | M | | M | | M | | M | | M | | M | |
| Library Construction (library size **integer**, number of clones sequenced **integer**, vector **CV**) **1** |  | |  | |  | |  | |  | | M | |
| Sequencing Method (e.g. dideoxysequencing, pyrosequencing, polony) **1, CV** | M | | M | | M | | M | | M | | M | |
| *Assay* |
| **DATA PROCESSING** |
| Assembly (assembly method **CV**, estimated error rate and method of calculation **CV**) **1** | M | | M | | M | | M | | M | | M | |
| Finishing strategy (status *e.g*. complete or draft **CV**, coverage **integer**, contigs **integer**) **1** | M | | M | | X | | X | | X | | X | |
| Classification (binning) method for fragments - **1,2, PMID or DOI or URL for SOP** | - | | - | | - | | - | | - | | M | |

The proposed contents of the MIGS checklist 1.1. All proposed descriptors in MIGS and the taxonomic groups to which they apply are listed. Taxa abbreviations: EU=Eukarya, BA=Bacteria and Archaea, PL=Plasmid, VI=Virus, OR=Organelle, and ME=Metagenome. Descriptors in grey are common to all taxonomic groups and are considered the 'core' of MIGS. “Source Material Identifier” is an exception; GSC recommends this to be a core descriptor, but as of yet physical archives (deposits in at least two culture collections for viable samples is recommended 21 and vouchers for specimens) are not yet routinely created for all cases/types of biological material subjected to genome sequencing. This is due to both cultural and technical issues. The need for universal and unique identifiers for metagenomic samples is an idea recently discussed in an exploratory workshop organized by the MetaFunctions group (www.metafunctions.org). In fact, the application of MIGS to our complete genome collection will require the designation of permanent and unique identifiers for all genome projects, something the INSDC is working to implement 16. All descriptors deemed to be essential are marked “M” (Minimal) and others which could be optionally applied to other groups with high priority are marked “X” (eXtra). Taxonomic groups for which a descriptor is not meaningfully are marked with a dash. This list of minimal information is recognized by the GSC as just a starting point for the description of genomes.