PRO-GRACE

PROMOTING A PLANT GENETIC
RESOURCE COMMUNITY FOR EUROPE

DELIVERABLE 1.2

Standards for collecting and displaying genetic data

Call identifier:
PRO-GRACE
Grant agreement no: 101094738

Promoting a plant genetic resource community for Europe

Deliverable No. D1.2

Standards for collecting and displaying genetic data

Contractual delivery date:
M11

Actual delivery date:
M11

Responsible partner:
CSIC
Contributing partners:
(CSIC, CREA, UNITO, ENEA, INIAV, MAICH, NASC, CNR, NORDGEN, UEB)

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101094738.

Grant agreement no.	Horizon Europe – 101094738
Project full title	PRO-GRACE – Promoting a plant genetic resource community for Europe

Deliverable number	**D1.2**
Deliverable title	**Standards for collecting and displaying genetic data**
Type	R – Document, report
Dissemination level	Public
Work package number	WP1
Author(s)	CSIC (Clara Pons, Antonio Granell, Bruno Contreras Moreira, Ernesto Igartua Arregui, Luis Guash), IPK (Catherine Hazel Aguilar, Stephan Weise) with the contribution of CREA, UNITO, ENEA, INIAV, MAICH, NASC, CNR, NORDGEN, UEB
Keywords	

The research leading to these results has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101094738**.**

SUMMARY

1. Introduction

Plant genetic resources (PGR) are crucial for sustainable crop production and for ensuring food and nutrition security both, today and in the future. To support efficient conservation and accessibility, the integration of genetic data has become essential through practices known as 'next-generation genebanking' (van Treuren and van Hintum, 2014; Wambugu *et al.*, 2018; Nybom and Lācis, 2021; Aubry, 2023).

The European Search Catalogue for Plant Genetic Resources (*EURISCO;* http://eurisco.ecpgr.org/) serves as a pivotal component of the European PGR information system. Developed within the European Cooperative Programme for Plant Genetic Resources (ECPGR; https://www.ecpgr.cgiar.org/), EURISCO acts as a centralized repository for passport and phenotypic data related to plant genetic resources conserved in European collections. However, it currently lacks representation of genetic data, which has seen significant advancements over the last two decades through EU-funded projects. Such valuable genetic information remains scattered across literature, specific websites, and databases from different projects, highlighting the pressing need for its integration into EURISCO.

Genetic data and metadata encompasses a wide array of types generated using diverse technologies, including whole genome sequencing, reduced genome sequencing (exome sequencing, genotyping, target sequencing, transcriptome sequencing), single nucleotide polymorphisms (SNP), simple sequence repeats (SSR), structural variations (SV), copy number variation (CNV), epigenomics, alternative splicing, gene expression, QTLs and GWAs trait-associated variants are some examples of genetic data produced using different technologies. Harmonizing, integrating, and making this genetic data FAIR (findable, accessible, interoperable, and re-usable) (Wilkinson et al., 2016) has been a shared goal within the scientific community since 2000. To achieve this, employing Permanent Unique Identifiers, standardized metadata with controlled vocabularies, and open-source communication protocols and databases are essential.

By linking genetic data with EURISCO and adhering to FAIR principles (FAIRification,(Welter *et al.*, 2023)), researchers and stakeholders can leverage high-quality, curated data for better decision-making and sustainable use of plant genetic resources, contributing to global food security. Two models for data FAIRification exist: create new standards or exploiting (rather than replacing) existing resources and promote their use. In this deliverable we comprehensively reviewed existing standards and public repositories for nucleotide sequence, genetic polymorphisms, makers and, variants to identify currently implemented standards, requirements, and features related to the management and sharing of genetic data and its metadata. The **overall objective of D1.2 is to find and evaluate solutions to the challenges described above, in line with the ECPGR's Plant Genetic Resources Strategy for Europe, and to develop a viable concept for genetic (meta)data implementation and long-term operation to further establish EURISCO as a trusted open-access repository**.

2. Activities

We organized the D1.2 activities in two phases:

1) In the **first phase** we conducted a survey of the FAIRsharing repository (https://fairsharing.org), a curated catalogue on data and metadata standards, inter-related to databases and data policies (Sansone *et al.*, 2019). The FAIRsharing repository was filtered by the terms “genetic”, “nucleotide”, “sequence”, “DNA”, “genomics” to identify the existing

standards and databases for genetic data. Standards were then classified in four types according to Sansone (Sansone *et al.*, 2019):

- **Minimum reporting guidelines:** guiding principles or checklists that outline the necessary and sufficient information vital for contextualizing and understanding a digital object
- **Semantic standards:** ranging from dictionaries to ontologies, provide definitions and unambiguous identification for concepts and objects.
- **Model/formats or syntax:** define the structure and relationship of information for a conceptual model and include transmission formats to facilitate the exchange of data between different systems.
- **Identifier schemes:** formal systems for resources and other digital objects that allow their unique and unambiguous identification.

Then standards and databases for genetic data were annotated according to: (i) FAIRsharing status indicators (maintained, ready for use and recommended) (Sansone *et al.*, 2019), (ii) if they are organism-specific or not and, (iii) their source or provider (this only for non-organism specific). In addition, we included ELIXIR's (an European life sciences infrastructure to enable researchers to access and analyse life science data, elixir-europe.org) recommended resources for submission of experimental data. In addition, we extended the search for stable identifier schemes to identifiers.org (http://identifiers.org/) a resolving system for persistent identifiers in the form of Compact Identifiers and URLs that enable the referencing of scientific data for the scientific community.

2) In a **second phase**, we reviewed those databases and standards and evaluated which are the more suitable to link genetic data with EURISCO.

3. Results

3.1. FAIRsharing survey to find available databases and standards to store EURISCO genetic data.

In July 2023, FAIRsharing had a comprehensive collection of more than 3,868 entries for databases and standards. Within this collection, related to genetic data, there were 873 databases (knowledgebases and public repositories with standard-compliant data) and 297 standards (minimum reporting guidelines, models/formats or syntax and identifier schemes) (**Table 1** and **Annex 1**). A significant portion of these entries, constituting 32.5%, can handle data and metadata across all species for a particular domain or area of study (genome, molecule- or process-specific databases, variations and others). The remaining 67.5% of available standards and databases were species-specific, taxa-specific or family-specific, predominantly focussed on biomedical research and model species. For plants only, there are 129 entries (115 databases and public repositories and 14 standards) (**Table 1**), most of them capable to manage meta(data) across a wide spectrum of plant species (**Annex 1**).

Table 1. *Summary of databases, public repositories and standards in FAIRsharing.org related to genetic information*

Type	Total entries FAIRsharing	Maintained	Ready-to-use	Recommended	All spcies data	Plant data	ELIXIR recommended
Knowledgebases & public respositories	**873**	**440**	**668**	**117**	**200**	**115**	**13**

Standards	**297**	**139**	**243**	**25**	**181**	**14**	**5**
Identifier schemes	1		1				
Minimum reporting guidelines	50	23	39	10	24	4	
Model/formats or syntax:Format	74	29	66	2	61	1	1
Model/formats or syntax:models/syntax	45	30	36	5	31	2	4
Semantic standards	127	57	101	8	65	7	
Total	**1170**	**579**	**911**	**142**	**381**	**129**	**18**

Since the ultimate goal of D1.2 is to find a global solution to link genetic data with EURISCO, which comprises 6,737 genera and 45,175 species (Kotni *et al.*, 2023), we excluded species-specific, taxon-specific or family-specific entries related to genetic data to identify standards. In total, 440 databases, public repositories and standards related to genetic data and accepting an unlimited number of species remain. We used this list to review and evaluate the existing standards that can be used to link genetic data with EURISCO. Around 239 databases and 181 standards for genetic data and accepting all species are collected in the FAIRsharing catalogue (**Annex 1**, **Table 1** and **Table 2**). When we analyzed the sources and providers for databases and standards (**Figure 1**) related to genetic data accepting diverse range of species, including plants, we found that most are provided by the International Nucleotide Sequence Database Collaboration (INSDC, https://www.insdc.org/) (Nakamura *et al.*, 2013; Cochrane *et al.*, 2016; Karsch-Mizrachi *et al.*, 2018) and its standard-developer collaborators. INSDC partners are (in alphabetical order): the DNA Data Bank of Japan (DDBJ; http://www.ddbj.nig.ac.jp/) at the National Institute for Genetics in Mishima, Japan; the European Nucleotide Archive (ENA, www.ebi.ac.uk/ena) at the EMBL European Bioinformatics Institute (EMBL-EBI) in Cambridge, UK; and GenBank (http://www.ncbi.nlm.nih.gov/genbank) at the National Center for Biotechnology Information (NCBI) in Bethesda, MD, (Nakamura *et al.*, 2013; Cochrane *et al.*, 2016; Karsch-Mizrachi *et al.*, 2018).

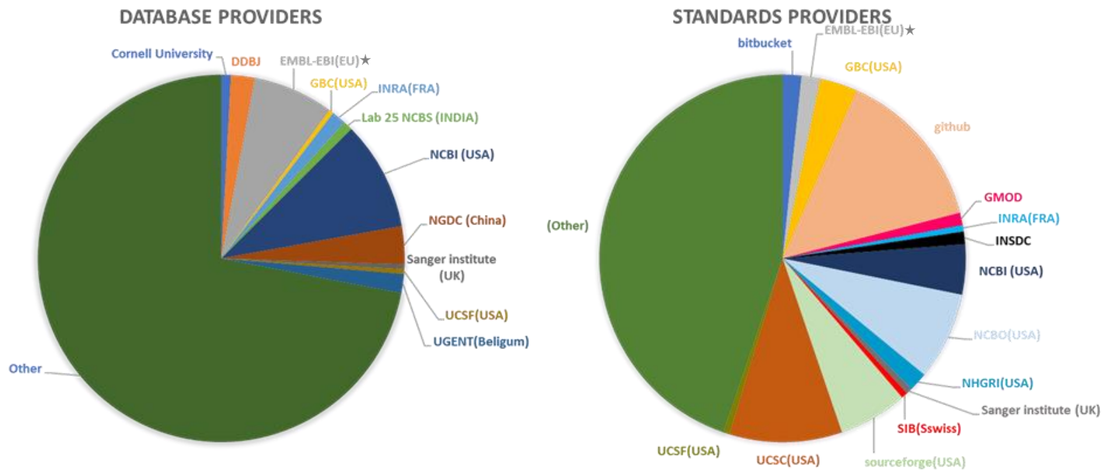


Figure 1. *Main sources of databases and standards related to genetic (meta)data that accept a diverse range of species, including plants. *EMBL include countries outside EU (see https://www.embl.org/about/member-states/)*

The INSDC repositories operate independently but synchronize periodically their data and apply similar quality standards (Nakamura *et al.*, 2013; Cochrane *et al.*, 2016; Karsch-Mizrachi *et al.*, 2018). These repositories are built to accommodate everything from raw data (i.e. next-generation sequencing reads) to assembly data, experimental design details, taxonomic information, functional annotation and information about the projects and biological samples associated with sequencing efforts (Nakamura *et al.*, 2013; Cochrane *et al.*, 2016; Karsch-Mizrachi *et al.*, 2018). Further, the INSDC repositories work in concert with appropriate standards communities to ensure rich metadata capture for understanding the origin of the sequences (Nakamura *et al.*, 2013; Cochrane *et al.*, 2016; Karsch-Mizrachi *et al.*, 2018). From this primary repositories, source of sequence data, many other secondary and tertiary databases (or knowledge bases) providing much more specific tools and information are constructed (**Table 2**).

Table 2. *Main public data repositories and knowledge bases for genetic data accepting all species or all plant species.*

Abbreviation	Name	Type	Scope	Species	Provider/maintainer	URL
GenBank	NIH genetic sequence database	Repository	raw and annotated sequebce data	all	NCBI	https://www.ncbi.nlm.nih.gov/genbank/
ENA	European Nucleotide Archive	Repository	nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation	all	EMBL-EBI	https://www.ebi.ac.uk/ena/browser/home
DDBJ	DNA Data Bank of Japan	Repository	nucleotide sequences, including raw data, annotated sequences, and metagenomic data	all	DDBJ	https://www.ddbj.nig.ac.jp/index-e.html
SRA	Sequence Read Archive	Repository	high-throughput sequencing data, including raw sequence reads	all	NCBI	https://www.ncbi.nlm.nih.gov/sra
EVA	European Variation Archive	Repository	all types of genetic variation data	all	EMBL-EBI	https://www.ebi.ac.uk/eva/
UCSC Genome Browser	UCSC Genome Browser	Knowledgebase	visualizing and analyzing genomic data	all	NCBI	https://genome.ucsc.edu/
Ensembl	Ensembl	Knowledgebase	visualizing and analyzing genomic data	all	EMBL-EBI	https://www.ensembl.org, https://plants.ensembl.org
ArrayExpress	ArrayExpress	Repository	all functional genomics data generated from microarray or next-generation sequencing (NGS)	all	EMBL-EBI	https://www.ebi.ac.uk/biostudies/arrayexpress
GEO	Gene Expression Omnibus	Repository	all functional genomics data generated from microarray or next-generation sequencing (NGS)	all	NCBI	https://www.ncbi.nlm.nih.gov/geo/
Track Hub Registry	**Track Hub Registry**	Repository	Traks for all kind of genetic data	all	EMBL-EBI	https://trackhubregistry.org/
PLAZA	PLAZA	Knowledgebase	comparative genomics, gene annotations, and sequence data	plants	GHENT University	https://bioinformatics.psb.ugent.be/plaza/
GERMINATE	GERMINATE	Repository	Phenotypic and genetic data	plants	The James	https://ics.hutton.ac.uk/germinate-demo/#/home

					Hutton Institute	
PlantGDB	Plant Genomics Database	Knowledgebase	comparative genomics, gene annotations, and sequence data	plants	NSF	https://www.plantgdb.org/
Gramene	**Gramene**	Knowledgebase	comparative genomics, gene annotations, and sequence data	plants	NSF	https://www.gramene.org/

Of course, there are many other databases (see **Table 2**, **Annex 1**) and standards related to genetic data provided by sources other than INSDC. databases or standards are specialized and crucial to the topic. These are typically maintained by researchers who specialize in the subject and are frequently generated as part of genome assembly/annotation projects. The data from these projects is ultimately submitted to INSDC.

3.3 Review of existing standards for models, formats and syntax:

Various scheme and syntax standards are used to structure and define the format of the data. These standards help ensure interoperability, consistent representation, and effective data sharing. They preferably should be widespread and open, readable and usable by a wide number of software and invisible to users (Sansone et al., 2019). A total of 91 standards for models, formats and syntax for genetic data and accepting data for all species are listed in FAIRsharing.

3.3.1 Standard formats

Several models and formats are commonly used for representing and storing genetic data. These models/syntaxes ensure that genetic information can be accurately captured, shared, and analysed across different tools and platforms. The choice of format often depends on the specific type of genetic data and the intended use case. Standard formats to store, interchange and display genetic data already exist for many types of data (see https://genome.ucsc.edu/FAQ/FAQformat.html). Some of these standards such as FASTA, FASTQ, SAM/BAM, GFF3 or VCF are stable, widely adopted and implemented in a diverse range of databases and analysis tools. The most common standards for file formats encoding genetic data are listed in **Table 3**.

Table 3. *The most common file formats to store genetic data*

	ACRONYM	NAME	Type of data	Codification	Extension
Sequence formats	**FASTA**	FASTA Sequence Format	nucleotide or peptide sequences	Plain text	.fa (others: .fasta,.fsa)
	FASTQ	FASTQ Sequence And Sequence Quality Format	Nucleotide sequence and PHRED quality scores	Plain text	.fq
	Genbank format	EMBL flat file format	Storing sequences nucleotide or peptide sequences and their associated meta-information, feature coordinates, and annotations	Plain text	.gb or .genebank
	EMBL format	EMBL flat file format	Storing sequences nucleotide or peptide sequences and their associated meta-information, feature coordinates, and annotations	Plain text	.embl
	DDBJ format	DDBJ flat file format	Storing sequences nucleotide or peptide sequences and their	Plain text	. ddbj

			associated meta-information, feature coordinates, and annotations		
Alignment formats	**SAM**	Sequence Alignment/Map Format	Alignments of sequencing reads to a reference genome	Plain text	.sam
	BAM	Binary Alignment Map Format	Alignments of sequencing reads to a reference genome	Binary	.bam
	CRAM	Compressed Reference-oriented Alignment Map	Alignments of sequencing reads to a reference genome	Byte	.cram
	VCF	Variant Call Format	DNA variation	Plain text	.vcf
	BCF2	Binary Variant Call Format version 2	DNA variation	Binary	.bcf
Discrete genomic feature data and tracks	**BED format**	Browser Extensible Data Format	Genomic intervals with additional information	Plain text	.bed
	Bedgraph	Browser Extensible Graphical Data Format	Genomic intervals with continuous-valued data	Plain text	.bb
	GTrack format	Genomic track	Genomic intervals with additional information	Plain text	.gtrak
	BTrack format	Genomic track	Genomic intervals with additional information	Binary	.btrak
	GFF3	Gene Feature File version 3	Features, positions and meta-feature information		.gff
	GTF	Gene transfer format	Genomic annotation with information and scores		.gtf

Sequence data files:

- **FASTA format**: created in 1988 (Pearson and Lipman, 1988), FASTA **is the standard** most implemented by databases and repositories (Sansone et al., 2019). It is a very basic format with two minimum lines **to represent either nucleotide or peptide sequences** (see **Annex 2** for an example of FASTA file). First line referred as comment line starts with '>' and gives basic information about sequence. There is no set format for comment line. Any other line that starts with ';' will be ignored. Lines with ';' are not a common feature of FASTA files. After comment line, sequence of nucleic acid or protein is included in standard one letter code. Any tabulators, spaces, asterisks etc in sequence will be ignored. A sequence file in FASTA format can contain several sequences. Each sequence will be separated by their "header" line, starting by ">".
- **FASTQ format**: common file format for sharing sequencing read data combining both the sequence and an associated per base quality score (Cock *et al.*, 2010). It is the **standard for storing the output of high-throughput sequencing instruments including third-generation sequencing instruments**. Sequences are represented by four lines (see **Annex 2** for an example of FASTQ file). Line 1 begins with "@" followed by the sequence identifier and optional descriptions. Line 2 is raw sequence letters while Line 3 begins with the "+" character optionally followed by the same sequence identifier and descriptions. Line 4 are PHRED quality scores for nucleotide sequence or the estimated probability for a base to be erroneous (Ewing and Green, 1998; Ewing *et al.*, 1998). PHRED scores use a logarithmic scale, and are represented by ASCII characters, mapping to a quality usually going from 0 to 93. Single ASCII character codes from 33 to 126 are used in place of numeric values to denote the PHRED scores in Line 4. See https://bioinformatics.uconn.edu/resources-and-events/tutorials-2/file-formats-tutorial/ for detailed information of PHRED scores and ASCII characters.
- **Genbank/EMBL/DDBJ formats**: Each of the **INSDC members** has a **sequence standard file format** that was designed to be rich and human readable. Despite some differences in the formatting of information, all of them contain almost the same fields and the same

information, making them interchangeable (see **Annex 2** for an example of Genbank/EMBL/DDBJ files). Each sequence entry in the INSDC sequence files is composed of lines (each with their own format) starting with a sequence unique identifier, followed by several annotation lines, the sequence and ended by two slashes ("//"). Annotation lines are used to record the various data including a concise description of the sequence, the scientific name and taxonomy of the source organism, and a table of features with controlled vocabulary (see https://www.insdc.org/submitting-standards/feature-table/ for format details) that identifies coding regions and other sites of biological significance, such as transcription units, sites of mutations or modifications, and repeats, protein translations and bibliographic references. For more details see https://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html,https://www.ddbj.nig.ac.jp/ddbj/flat-file-e.html and https://ena-docs.readthedocs.io/en/latest/submit/fileprep/flat-file-example.html.

Alignment data files

- **SAM format:** SAM is the **standard for reads alignments against reference sequences** (Li *et al.*, 2009). It is a tab-delimited text-based format for storing single- and paired-end reads alignments against reference sequences. **It can support short and third generation sequencing long reads produced by different sequencing platforms**. The format has been extended to include unmapped sequences, and it may contain other data such as base-call & alignment qualities. The SAM format (see **Annex 2** for an example) consists of a header section (~metadata) and an alignment section (raw mapping information). Header lines contain metadata about the reference sequences, read and sample information, and (optionally) processing steps and comments. Each header line begins with @, followed by a two-letter code that distinguishes the different types of metadata records in the header: @HD file level metadata, @SQ reference sequence dictionary, @RG read group, @PG program. Each alignment record in a SAM file contains 11 mandatory fields in tab-separated columns describing the read, quality of the read, and nature alignment of the read to a region of the genome. SAM standard is maintained by the Large Scale Genomics work stream of the Global Alliance for Genomics & Health (GA4GH; https://www.ga4gh.org/). For detailed info see https://samtools.github.io/hts-specs/SAMv1.pdf and https://pacbiofileformats.readthedocs.io/en/12.0/BAM.html.
- **BAM format:** It is the lossless compressed binary version of the Sequence Alignment/Map (SAM) format. The data between SAM and BAM are exactly same, but BAM is compacted and indexed. BAM file requires an associated BAM.bai index file. Binary BAM files are small and ideal to store alignment files. **BAM is the most commonly used format for submission of reads to INSDC sequence repositories and analysis**. BAM is maintained by the Large-Scale Genomics work stream of the Global Alliance for Genomics & Health (GA4GH; https://www.ga4gh.org/). For a detailed info see https://samtools.github.io/hts-specs/SAMv1.pdf.
- **CRAM format:** CRAM is primarily a reference-based compressed format of SAM file, meaning that only differences between the stored sequences and the reference are stored. Additionally, each column in the SAM format is separated into blocks, improving compression ratio. It holds the same information as their SAM equivalent, offering significantly better lossless compression whilst maintaining backwards compatibility with BAM. Since CRAM files are **typically 30 to 60% smaller than their BAM equivalents**, some labs and repositories are implementing them. Usually, **CRAM files are used to store alignments and converted to the BAM format for processing**. CRAM is maintained by the Large-Scale Genomics work stream of

the Global Alliance for Genomics & Health (GA4GH; https://www.ga4gh.org/). For a detailed info see https://samtools.github.io/hts-specs/CRAMv3.pdf.

- **VCF format:** is the **standard file format for storing DNA variation data** between a reference genome and sequences aligned to it, based on SAM/BAM alignments. VCF is a preferred format for databases and analysis tools because it is **unambiguous, scalable and flexible**, allowing extra information to be added to the info field and storing many millions of variants. The VCF format is a tab-delimited format that stores individual genotypes and variant calls including SNPs, insertions, deletions and structural variants, together with rich annotations (Danecek *et al.*, 2011). The current version VCF 4.3 consists of two parts: metadata lines and data section (See **Annex 2**). Metadata lines start with a ##, each line begins with a key=value pairs, and have a description of the settings, samples and general experimental design of the genotyping. The data section begins with a single #; this first line is the header line the eight fixed, mandatory fields, used to record the genotyping calls: chromosome (CHROM), a 1-base position of the start of the variant (POS), unique identifiers of the variant (ID), the reference allele (REF), a comma separated list of alternate non-reference alleles (ALT), a PHRED-scaled quality score (QUAL), site filtering information (FILTER) and a semicolon separated list of additional, user extensible annotation (INFO) and information contained within each subsequent genotype column (FORMAT). After the FORMAT field, the sample IDs define the samples included in the VCF file. The following lines of the data section contain the variant site records. Variant site records do not start with any symbol. VCF data and metadata information have a controlled vocabulary for describing variants, however, the VCF format defines a syntactic format but no vocabulary, unique identifier or recommended content for describing samples and reference genome. Recently, in the frame of ELIXIR, a minimal list of metadata recommended fields (**Table 4**) to describe plant samples in the VCF file has been proposed (Beier *et al.*, 2022). The goal of these recommendations is to link, through precise sample identifiers provided by BioSample (see below), variant data for plants to other information data sets (e.g. phenotypic, transcriptomic, metabolomic) submitted to EMBL-EBI repositories. The Global Alliance for Genomics & Health (GA4GH) maintains the specification of the VCF. See http://samtools.github.io/hts-specs/ for more details.

Table 4. *Summary of recommendations for metadata formatting (from Beier et al., 2022)*

Metadata field	Definition	Format	Example	Cardinality
##fileDate	Creation date of the VCF file	Date (ISO 8601, YYYYMMDD)	##fileDate=20120921	1
##bioinformatics_source	Chains of bioinformatics tools for creating the VCF file	URL, DOI	##bioinformatics_source="doi.org/10.1038/s41588-018-0266-x"	1
##reference_ac	Accession number of reference genome assembly used in the VCF file	/[(GCA/GCF)_(d){9}\.(0-9)*]/	##reference_ac=GCA_902498975.1	1
##reference_url	URL of the reference genome assembly used in the VCF file	URL, DOI	##reference_url="ftp.ncbi.nlm.nih.gov/genomes/all/GCA/902/498/975/GCA_902498975.1_Morex_v2.0/GCA_902498975.1_Morex_v2.0_genomic.fna.gz"	1
##contig	Metadata about a single sequence in the reference genome assembly	Composite (see below)	##contig=<ID=chr1H,length=522466905,assembly=GCA_902498975.1,md5=8d21a35cc68340ecf40e2a8dec9428fa,species=NCBITaxon:4513>	1:N
	The primary identifier of the sequence	String	ID=chr1H	1
	The length in base pairs (bp) of the sequence	Integer	length=522466905	1
	The assembly accession number this sequence belongs to	/[(GCA/GCF)_(d){9}\.(0-9)*]/	assembly=GCA_902498975.1	1

Metadata field	Definition	Format	Example	Cardinality
	The md5 checksum of the sequence	MD5	md5=8d21a35cc68340ecf40e2a8dec9428fa	1
	The species of the sequence (NCBI Taxon ID)	/[(NCBITaxon):(\d+)]/	species=NCBITaxon:4513	1
##SAMPLE	Metadata about a single sample genotype that is part of the genotyping experiment in the VCF file	Composite (see below)	##SAMPLE=<ID=SAMEA104646767,DOI="doi.org/10.25642/IPK/GBIS/7811152">	1:N
	The primary identifier (BioSamples Database identifier) of the genotyping sample	/[(SAM)(E|N|D)(A|G)(\d+)]/	ID=SAMEA104646767	1
	The DOI of the genotyping sample (if available)	URL, DOI	DOI="doi.org/10.25642/IPK/GBIS/7811152"	0-1
	The external identifiers under which this genotyping sample is registered in other databases (either 'FAO-WIEWS_instcode:genus:accession_number' or 'DNS:database_identifier:identifier_scheme:identifier')	See Definition	ext_ID="DEU146:Hordeum:HOR 1361 BRG" or ext_ID="ipk-gatersleben.de:GBIS:akzessionId:7811152"	0:N

- **BCF2 format:** is a binary, compressed equivalent of VCF that can be indexed with tabix and can be efficiently decoded from disk or stream. The relationship between BCF and VCF is similar to that between BAM and SAM. The Global Alliance for Genomics & Health (GA4GH) maintains the specification of the VCF. See http://samtools.github.io/hts-specs/ for more details.

Discrete genomic feature data and track files:

- **BED format:** is a simple, tab-delimited text format **for representing track genomic features and their associated annotations linked to the genomic coordinates of a reference genome**. Examples of such data are epigenetic DNA methylation data, ChIP-seq peaks, germline or somatic DNA variants, as well as RNA-seq expression levels. BED files are predominantly used for graphical display, but can also be queried by statistical analysis tools. BED files consist of one line per feature, with each line containing a minimum of three mandatory columns (chromosome, start, and end), and nine additional optional tab-separated columns for feature annotations (name, score, strand, etc). A BED file can optionally contain a header. However, there is no official description of the format of the header. It may contain one or more lines placed at the beginning of the list of features affected and be signified by different words or symbols, depending on its functional role or simply descriptive: genome browser options (browser); track name and configuration description (track) and other comments, such as the name of each column (#). **BED files do not contain any metadata associated to them excepting the track name**. See https://github.com/samtools/hts-specs and http://genome.ucsc.edu/FAQ/FAQformat.html#format1 for more details.
- **BedGraph format**: is a variation of the BED format **used for representing continuous-valued data in track format**, such as read coverage or signal intensities, in a genome browser. BedGraph files are tab-delimited text files, but they contain four columns (chromosome, start, end, and value) representing the genomic region and the associated data value. The bedGraph format is line-oriented. BedGraph data are preceded by a track definition line, which adds a number of options for controlling the default display of this track. Strand information is provided in the value column, negative values represent reverse strand, whereas positive

values represent forward strand. For more detailed information see http://genome.ucsc.edu/goldenPath/help/bedgraph.html

- **GTrack format**: a tabular format developed to provide a **uniform representation for tracks of most types of genomic datasets.** This track format data file format is capable of representing all 15 track types (Gundersen *et al.*, 2011) since they have a flexible column specification and declaring syntactic properties at the beginning of the file. The GTrack format consists of 5 different line types, distinguished by the leading characters and numbered here by order of appearance in the file. Comments preceded by the character # are ignored by parsers and may be present anywhere in the fil. Header lines, preceded by ##, provide detailed information for the type of track. Column specification line, preceded by ###, provide detailed information about the origin of the track (reference, genome, sample, genomic coordinates, values, graphical representation...). Bounding region specification line, preceded by ####, specifies the region where we have information. Data lines, a tab-separated list of values without any preceding character, are defined by the column definition line. **GTrack is able to replace most common track file formats and had the advantage that all GTrack types are supported by the FAIRtracks metadata standard** (see below). For more detail see https://github.com/gtrack/gtrack/tree/master/gtrack/spec and https://fairtracks.net/standards/#standards-01-fairtracks.
- **BTrack format**: supports the same variety of informational content as GTrack, but in binary form. See https://fairtracks.net/standards/#standards-01-fairtracks.
- **GTF/GFF3 formats**: are tab-delimited text file formats used for describe functional regions of genomes. **GFF3 is the most widely used format to annotate genomes**. GFF/GTF lines have nine required fields that *must* be tab-separated. GFF/GTF share the same structure for the first 7 fields, while differing in the format of the ninth field (See **Annex 2**). The first seven lines define the sequence name, the source of the sequence, the feature type (Transcripts, exon, intron, promoter, 3’ UTR, repetitive elements etc), the start and the end of the feature, the confidence score, the strand and the Frame (GTF) or Phase (GFF3). Column 9 contains attributes pertaining to the described feature. In GFF3 format, it contains a list of feature attributes (name, Id, alias, ontology, feature hierarchy, etc) in the format tag=value. Multiple tag=value pairs are separated by semicolons. In GTF, column 9 contains globally unique identifiers for the genomic locus of the transcript and the predicted transcript, respectively, separated by semicolon. For more details see https://github.com/The-Sequence-Ontology/Specifications/blob/master/gff3.md and http://genome.ucsc.edu/FAQ/FAQformat.html#format1.

3.3.2. Models/syntax

In FairSharing.org there are 45 models and syntax standards for genetic data storage and exchange across various fields and domains (**Annex 1**). These standards ensure consistency, interoperability, and efficient sharing of data among different systems and users. The most 5common models and syntax standards for genetic data storage and exchange are listed in **Table 5**.

Table 5. *Models and syntax standards most common in genetic databases.*

ACRONYM	NAME	Implementing databases or standards	Used for
XML	Extensible Markup Language,	NCBI Entrez Gene, ENA, DDBJ, UniProtKB, INSDC, Gene Expression Omnibus (GEO), InterPro, Phytozome, Gramene, PlantGBD, Planteome, BAR	data export, web services, or specific metadata representation

		(Bio-Analytic Resource) Plant Genome Database, MaizeGDB	
RDF	Resource Description Framework	Planteome, Agronomic Linked Data (AgroLD	representing ontologies, linking data, and providing structured metadata
Json	JavaScript Object Notation	Gramene, PlantGDB, Phytozome	data retrieval and representation, especially when interacting with web-based services
CHADO, CHADO-XML	Modular scheme for biological database	Gramene, Solgenomics, TAIR	store diverse types of biological data, including sequences, annotations, ontologies, and genetic markers, among others.
ISA, ISA-TAB	Investigation Study Assay Tabular	EVA, ArrayExpress, MIAME, MINSEQE, MIGS/MIMS	Capturing, describing and exchanging of experimental data and metadata
OWL	Web Ontology language Format	Planteome, Agronomy Ontology Service (AO), Gene Ontology (GO), Sequence ontology (SO)	define the concepts, relationships, and semantics within these ontologies,
BRAPI	Breeding API specification for the communication of plant breeding data	The Triticeae Toolbox (T3), CassavaBase, Integrated breeding platform	exchange of plant breeding and agricultural data across different software systems and platforms

- **XML**: is a versatile markup language used for structuring and encoding data in a machine-readable format. Several databases that store genetic data use **XML to represent, storage and exchange various types of genetic (meta)data, from sequences to annotations** (**Table 5**). It is often chosen for its flexibility in structuring complex data and its compatibility with web-based data exchange. The specific usage of XML can vary between databases. For example, Gramene uses XML to provide structured data about genes, genomes, pathways, and annotations while Planteome uses XML to define and exchange ontology terms, annotations, and relationships. See https://www.w3.org/TR/rdf-syntax-grammar/ for more detailed info
- **RDF:** is a standard for describing resources on the web, using a graph-based model to represent data and their relationships. It is commonly used in the context of semantic web technologies and linked data, where it enables the creation of interconnected and machine-readable data networks. Some resources related to genomics and plant genetics might use RDF for specific purposes, such as **representing ontologies, linking data, and providing structured metadata** (**Table 5**). Planteome uses RDF to represent and provide structured data about plant traits, anatomy, and growth stages. RDF is used to enable the integration of plant phenotype data across different species. For more details see https://www.w3.org/RDF/
- **JSON:** is a lightweight data interchange format that is easy for both humans and machines to read and write. In the field of genomics and plant genetics, JSON is often **used for data exchange through web services and APIs**. It is commonly used to provide structured data to users and applications over the internet. While JSON may not be the primary format for data storage in these databases, it could be used for certain types of data retrieval and representation, especially when interacting with web-based services. For more details see https://json-spec.readthedocs.io/index.html
- **CHADO/CHADOXML scheme:** is a relational database **modular scheme designed for storing, managing and integrating biological data. It is capable of representing many of the general classes of data frequently encountered in modern biology such as sequence, sequence comparisons, phenotypes, genotypes, ontologies, publications, and phylogeny** (Mungall *et al.*, 2007), particularly genomic and genetic data. It's an open-source scheme that provides a structured framework for representing diverse biological information in a consistent and standardized manner. Chado was developed to address the complex and interconnected nature of biological data and to support integration and querying across various data types and sources**.** CHADO is part of the Generic Model Organism Database (GMOD) suite

(www.gmod.org), a collection of open source software tools for managing, visualising, storing, and disseminating genetic and genomic data that include tools such as Jbrowse, Galaxy, Biomart or Tripal. For more detail see http://gmod.org/wiki/Chado and http://gmod.org/wiki/Chado_XML

- **ISA/ISA-tab:** The ISA (Investigation/Study/Assay) model is a general-purpose framework for **capturing and describing the metadata associated with experiments** in various scientific domains, including the life sciences (Sansone *et al.*, 2012). ISA-tab is the original format developed by the ISA team. It's a simple, tabular format for representing investigation, study, and assay metadata. It was developed to provide rich descriptions of experimental metadata (i.e. sample characteristics, technology and measurement types, sample-to-data relationships) so that the resulting data and discoveries are reproducible and reusable. The ISA model is not tied to a specific database or biological standard but rather serves as a flexible structure that can be implemented and adapted by various projects and databases Different databases or standards dealing with genetic data such as EVA, ArrayExpress, MIAME, MINSEQE, MIGS/MIMS have implemented the ISA model with slight variations or customizations to suit their specific needs, but the core principles of capturing experimental metadata in an organized and standardized way remain consistent. See https://isa-specs.readthedocs.io/en/latest/index.html for more details.
- **OWL language**: is a powerful and expressive language used **for creating ontologies**—formal models that represent knowledge about a domain and the relationships between different concepts within that domain. OWL provides a rich set of constructs for describing classes, properties, individuals, and relationships between them, supporting the creation of hierarchies of classes and sub-classes, reflecting the inherent structure of the domain being modelled. Further, OWL ontologies can be serialized using RDF (Resource Description Framework), making them compatible with other Semantic Web technologies and allowing for easy integration with existing data. Some genetic resources or projects related to ontologies, functional annotations, and semantic integration currently (?) use OWL, for instance Gene Ontology (GO) and Sequence Ontology (SO). OWL is part of the Semantic Web stack, a set of technologies aimed at enhancing the meaning and interpretation of data on the World Wide Web. For more details see https://www.w3.org/TR/owl-ref/
- **BrAPI:** is an open standard and API (Application Programming Interface) designed to facilitate a **standardized exchange of plant breeding and agricultural (meta)data, such as germplasm information, phenotypic and genotypic data, trial and experiment details**, across different software systems and platforms. (Selby *et al.*, 2019). BrAPI is not a database itself but rather a set of standardized APIs that various databases and systems can implement to enable data exchange and integration. BrAPI re-uses existing data standards as much as possible and is compatible with both the FAO Multi-Crop Passport Descriptors and the minimum information standards for plant phenotyping. BrAPI aims to improve data interoperability and enable data sharing and integration among researchers, breeders, and organizations involved in plant breeding and crop improvement. BrAPI has gained popularity in the plant breeding and agricultural research communities because it addresses the challenges of data fragmentation, lack of interoperability and allow to develop software applications, databases, and tools. For more detail see www.brapi.org.

3.4. Identifier (ID) schemes standards

ID scheme standards serve for providing unambiguous and stable identification for individual objects of interest. Fairsharing has over 33 entries registered as ID scheme standards, however none has a direct relation exclusively with genetic data. In the case of Identifiers.org, **there are more than 350 ID scheme related to genetic data or their annotation** (see **Annex 3**). With the exception of centralized repositories and ontologies (Nucleotide Sequence Database, SRA, EVA, RefSeq, Ensembl, UniProt, InterPro, KEGG, Gene Ontology, Sequence ontology, etc), most identifiers in Identifiers.org are species-specific, taxa-specific or family-specific.

ID standards encompass various formats such as text, numbers, URLs/DOIs, and primary keys for genetic related objects that are **assigned when the objects (raw sequence, genome, alignment, gene, marker, variant locus, ontology term) are created in the repositories, knowledgebases or in ontologies for a particular domain or area of study**.

In the case of INSDC, sequence accession numbers have specific formats that might vary slightly between databases, but they generally follow a similar pattern. Here's the general format for sequence accession numbers in INSDC repositories:

- Prefix: The prefix indicates the type of submission and provides context for the sequence. Different prefixes are assigned to different types of data, such as genomic DNA, mRNA, coding sequences, non-coding RNAs, and more.
- Numbers (Digits): The numeric portion of the accession number follows the prefix and is unique to each sequence entry. It's usually a sequence of digits.

Each type of data is assigned a unique accession number following a specific format. For example:

- **Genomes:** all INSDC members use a standardized format for annotating genome assemblies. This format involves using the "GCA_XXXX" identifier followed by a unique numerical accession number to represent the entire genome assembly sequence and version of an organism. This standardized format ensures consistency and ease of reference when dealing with genome assemblies across different organisms and databases within the INSDC collaboration.
- **Genes and Transcripts:** The specific format can vary between databases, but it often includes a prefix that denotes the type of sequence. Accession numbers from GenBank often start with a two-letter prefix followed by digits. For example, "NM_001123456" for a RefSeq mRNA entry or "CP012345" for a genomic DNA entry. Accession numbers from ENA start with a letter prefix followed by digits. For example, "X12345" for a nucleotide sequence or "Y56789" for a transcript. Accession numbers from DDBJ also start with a letter prefix followed by digits. For example, "D12345" for a DNA sequence entry or "AB67890" for a transcript.
- **Splicing Forms:** Splicing forms, which represent alternative splicing isoforms of a gene, are often assigned to unique identifiers that are distinct from the primary gene or transcript accession numbers. These identifiers might include version numbers or other annotations to differentiate between different splicing forms.
- **Variants:** variants submitted in the context of a genomic study are accessioned with submitted SNP ID number (“ss#”, that identifies the taxon, the reference genome, the study, the position and the ref/alt alleles) and a reference SNP ID number (“rs#”; "refSNP cluster"), that identifies a variant class in a genomic location of a taxon genome and which can be reported by one or more studies. Variants, such as single nucleotide polymorphisms (SNPs), insertions, deletions, and structural variants, are typically accessioned using unique identifiers.

Ideally, each genetic object should be assigned to a single ID to avoid confusion. However, most IDs found in Identifiers.org have synonyms in other databases, resulting in the assignation of different IDs to the same object. Detailed info of the IDs, such the pattern of the ID, resource URL and synonyms, used for different ontologies and databases are in **Annex 3**. To deal with multiple IDs, the concept of Persistent Unique Identifiers (PUID) has emerged as a potential solution. **PUIDs are designed to be**

permanent and steadfast identifiers, maintaining their validity on a global scale. It is crucial that PUIDs are never completely deleted but instead deprecated if necessary. Additionally, these identifiers should be effortlessly convertible between local IDs and URIs, allowing for seamless interoperability. A well-known example of a PID is a Digital Object Identifier (DOI) which is used to locate specific digital objects. Recently, the term Digital Genetic Object (DGO) has also been proposed (Manzella *et al.*, 2023), the idea is to create DGOs for each type of genetic information in INSDC repositories (e.g. nucleic acid sequence reads, data about gene expression and function, unsequenced markers and chromosomal segments) and link them to the DOI of the germplasm source material to facilitate discovery interoperability among genomic data and germplasm repositories. This proposal, however, is still in infancy (Manzella *et al.*, 2023).

Special case: Plant gene ID

The standardization description of gene nomenclature is of critical importance. It's important to note that the patterns for gene IDs can vary widely between organisms and databases (see **Annex 3**). The adoption of standardized nomenclature and IDs can be influenced by the research community, the organism's complexity, and the level of annotation available. Each one of the INSDC repositories uses their own centralized and stable ID pattern for genes. The gene ID patterns used in the repositories of the INSDC can vary depending on the organism and the type of gene annotation. Further, organism-specific communities have developed their own gene nomenclature standard system. **Table 6** shows plant gene nomenclature resources. Despite this, common formats and conventions are often followed:

- **Locus Tag or Systematic ID:** Many organisms use a systematic naming convention that often includes a short prefix followed by numbers. Organisms with well-defined genomes often use locus tags to identify genes in a sequential or systematic manner.
- **Gene Name:** Some genes are identified by their common or official names. They are often chosen based on their functions, characteristics, or roles in specific biological pathways. For example, the plant gene for Activating Transcription Factor 1 is often referred to as " ATF1."
- **Accession Number-Based:** Sometimes, genes might be linked to an accession number of a specific gene entry in the database. For instance, in GenBank, genes might be referenced using their associated accession number, such as "NM_001123456" for a RefSeq mRNA entry.
- **Synonyms or Aliases**: Genes might have synonyms or aliases that are used as identifiers. These can include both official and common names, as well as other identifiers used in the literature.
- **Standardized Gene Symbols:** For well-studied organisms, standardized gene symbols might be used. These symbols are usually short and informative names that represent the gene's function or role.
- **Unique Identifiers:** Some databases assign unique identifiers to genes that are independent of their names or other identifiers. These can be alphanumeric strings or combinations of letters and numbers.

Table 6. *Plant Gene Nomenclature Resources (Adapted from (Reiser et al., 2018))*

Database name	URL	Species covered
Alfalfa Breeders Toolbox	alfalfatoolbox.org	Alfalfa
CassavaBase	cassavabase.org	Cassava
Citrus Genome Database	citrusgenomedb.org	Lemons, oranges, and more
Cool Season Food Legume Database	coolseasonfoodlegume.org	Lentil, pea, fava bean, chickpea
CottonGen	cottongen.org	Cotton, many species
Genome Database for Rosaceae	rosaceae.org	Apple, strawberry, rose, plums, pears, and more
Genome Database for Vaccinium	vaccinium.org	Blueberry, cranberry, bilberry, and more
GrainGenes	wheat.pw.usda.gov/GG3	Wheat, barley, oats

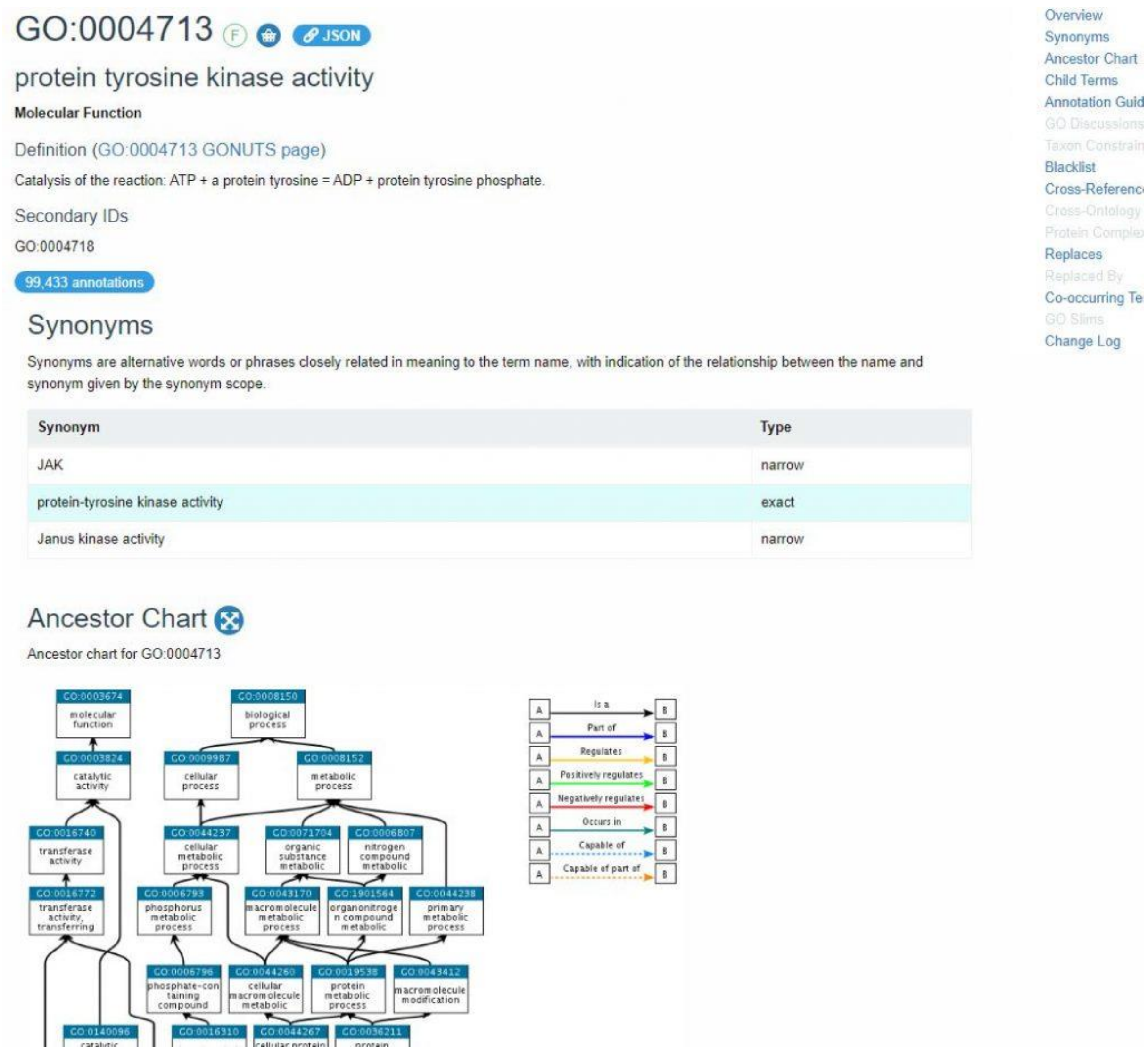
Gramene	https://www.gramene.org	Maize, rice
International Grape Genome Program	http://www.vitaceae.org	Gravepine
Hardwood Genomics	hardwoodgenomics.org	Oaks, poplars, maples, chestnuts, and more
KnowPulse	knowpulse.usask.ca	Chickpea, common bean, lentil
Legume Information System	legumeinfo.org	Soybean, Medicago, cowpea, chickpea, and more
MaizeGDB	maizegdb.org	Maize
Medicago truncatula Genome Database	medicagogenome.org	Clover
MusaBase	musabase.org	Banana
Oryzabase, Rap-Db	shigen.nig.ac.jp/rice/oryzabase/	Rice
PeanutBase	peanutbase.org	Peanuts
Solanaceae Genomics Network	solgenomics.net	Tomato, potato, eggplant, petunia, and more
Soybase	soybase.org	Soybean
Sweet Potato Database	sweetpotatobase.org	Sweet potato
T3	triticeaetoolbox.org/wheat	Wheat, barley, oats
TAIR	arabidopsis.org	*Arabidopsis thaliana*
TreeGenes	treegenesdb.org	1792 species of trees
YamBase	yambase.org	Yams

One of the main problems with gene ID is that researchers often encounter a mix of INSDC IDs and organism-specific gene IDs, names and gene products, which leads to confusion, missing information, and makes difficult to search or integrate data about genes (Reiser *et al.*, 2018). However, some resources provide cross-reference tools to help you map between different gene ID systems, enabling you to work with the IDs that best suit your research context. **Whether to use organism-specific gene IDs or INSDC gene IDs depends on the context of your research and the goals of data integration and analysis.** Both types of IDs have their advantages and considerations. Organism-specific gene IDs provide direct information about genes (gene annotations, including functional annotations, pathways, and other relevant data for that organism) in the context of a specific organism. Further, they are more intuitive and familiar to researchers working with that organism. INSDC gene IDs provide standardized access to genes across different organisms. This can be beneficial when working with multiple species or conducting comparative genomics or integrating genetic data from different organisms.

A solution to reduce the number of IDs assigned to the same gene and facilitate the integration gene data could be those used in Ensembl Plants (EMBL-EBI genome-centric portal for *plant* species of scientific interest, see below). Ensembl Plants, with genome for more than 150 species (see **Annex 4**), has adopted a combination of organism-specific gene IDs and INSDC IDs, depending on the gene and the organism being studied**. For well-studied model organisms and species with established annotations, Ensembl use organism-specific gene IDs that are consistent with the existing nomenclature for those species.** This helps Ensembl users from different organism communities to easily identify and locate genes they are familiar with, within the context of the species they are interested in.

3.5 Review of existing Semantic standards

Semantic standards play a crucial role in organizing, annotating, and exchanging metadata in a structured and machine-readable manner. They range from dictionaries to ontologies and provide definitions and unambiguous identification for concepts and objects (Sansone *et al.*, 2019). In addition, ontologies can provide a hierarchical structure of the terms. Hierarchical ontologies are the most common used, they organize concepts into a hierarchy of parent-child semantic relationships which provide a richer context for understanding the relationships between concepts (**Figure 2**). For identifying and accessing concepts, terms, and relationships ontologies use URLs and IDs. Both URLs and IDs are important components of ontology design and usage. URLs provide access to additional information and context, while IDs provide efficient and unique references for ontology terms.



GO:0004713

protein tyrosine kinase activity

Molecular Function

Definition (GO:0004713 GONUTS page)

Catalysis of the reaction: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.

Secondary IDs

GO:0004718

99,433 annotations

Synonyms

Synonyms are alternative words or phrases closely related in meaning to the term name, with indication of the relationship between the name and synonym given by the synonym scope.

Synonym	Type
JAK	narrow
protein-tyrosine kinase activity	exact
Janus kinase activity	narrow

Ancestor Chart

Ancestor chart for GO:0004713

Figure 2. *Snapshot of the protein tyrosine kinase entry in GO, showing the richness of an ontology and highlighting the terms and hierarchies that are used to populate controlled vocabularies.*

In Fairsharing.org there are 63 entries for semantic standards dealing with genetic data for all species (**Annex 1**). These ontologies provide a common language to enhance data integration, interoperability, and automated reasoning in genomics research.

In the context of genetics and genomics, especially for plant genetic metadata, several semantic standards are commonly used (**Table 7**).

Table 7. *Common semantic standards used for annotating and standardizing genetic metadata*

ACRONYM	NAME	Used for
GO	Gene Ontology	describing gene functions, biological processes, and cellular components in a hierarchical model
SO	Sequence Ontology	sequence features
OBI	Ontology for Biomedical Investigations	biological investigations, experiments, and assays (including genetics)
BAO	BioAssay Ontology	describing various types of assays
EDAM	Sequence Types and Features Ontology	analysis, modelling, optimisation, and data life-cycle
EFO	Experimental Factor Ontology	annotation, analysis and visualization of samples, phenotypes and studies
	DDBJ/ENA/GenBank Feature Table Definition	features and characteristics of genetic sequences
PATO	Phenotype And Trait Ontology	phenotypes

PO	Plant Ontology	plant anatomy, morphology and growth and development
TO	Trait Ontology	phenotypic traits in plants
AgrO	Agronomy Ontology	agronomic practices, techniques, variables
CO	Crop Ontology	crop-specific agronomic, morphological, physiological, quality, and stress traits
PECO	Plant Experimental Condition Ontology	Abiotic treatments, growing conditions and study types
NCBI Taxonomy	NCBI Taxonomic Database	Species taxonomic classification and nomenclature
ChEBI	Chemical Entities of Biological Interest	chemical compounds

Some of them such as GO and SO are strictly used for genetic annotations, while others like OBI, BAO, EDAM, EFO are used to describe metadata associated to the samples and protocols used to generate the genetic data:

- **GO:** it provides a standardized vocabulary for describing more than 84.220 terms related to gene functions, biological processes, and cellular components associated with genes and gene products (The Gene Ontology Consortium, 2017). GO is one of the **most widely used ontologies in genomics and implemented** in major public databases**.** See more details in http://www.geneontology.org
- **SO:** The Sequence Ontology (Eilbeck *et al.*, 2005) defines 1,454 terms to describe sequence features, such as exons, introns, and regulatory elements. It's used to annotate genomic sequences and variants. While not exclusively for experimental features, SO includes terms related to experimental features, such as "assembly", "SNP," "SSR," and "AFLP," allowing genetic markers to be annotated within a genomic context. See http://www.sequenceontology.org/ for more details.
- **OBI**: While broader in scope, OBI is an ontology for investigations, materials, protocols and instrumentation used, the data generated and the types of analysis performed on it (Bandrowski *et al.*, 2016). It contains 4,887 terms, among them terms related to genomics research as well as terms for various biological and molecular entities as methods, which can be used to annotate genetic markers and their properties. This ontology is the foundry of new ontologies such as Next generation biobanking ontology (NGBO) (Alghamdi *et al.*, 2023). See https://obi-ontology.org/ for more detailed description and https://github.com/Dalalghamdi/NGBO.
- **BAO:** is a a semantic description of bioassays and high-throughput screening results (Visser *et al.*, 2011). It includes more than 8700 terms for describing various types of assays, which can include genetic assays and techniques used to analyze genetic material. See http://bioassayontology.org/
- **EDAM:** is domain ontology of data analysis and data management in bio- and other sciences, and science-based applications (Ison *et al.*, 2013). It comprises more around 3500 concepts related to analysis, modelling, optimisation, and data life-cycle including those used in genomics. EDAM is used in numerous resources, for example Bio.tools, Galaxy, Debian, or the ELIXIR Europe training portal TeSS. See http://edamontology.org/page# and https://edamontology.github.io/edam-browser/#topic_3175 for more details.
- **EFO:** provides a systematic description of more than 50.600 experimental variables available in EBI databases (Malone *et al.*, 2010), and for projects such as the EMBL-EBI GWAS catalogue. It combines parts of several biological ontologies such as OBI, GO, Cell ontology, Environment ontology (see https://agroportal.lirmm.fr/ontologies/EFO/?p=mappings) and aims to support the annotation, analysis and visualization of samples, phenotypes and studies. See https://www.ebi.ac.uk/efo/ for detailed info.
- **DDBJ/ENA/GenBank Feature Table Definition:** The Feature Table Definition provides a **structured format for annotating and describing the features and characteristics of genetic sequences** in a standardized manner in the frame of INSDC (Cochrane *et al.*, 2016). This format

allows to provide detailed information about various elements within a genetic sequence, such as genes, coding regions, promoters, introns, exons, and more. While not include the rich semantics and relationships seen in more complex semantic ontologies, it does establish a standardized way of representing and communicating important information about genetic features within a sequence. For detailed info see https://www.insdc.org/submitting-standards/feature-table/.

- **Phenotypic Quality Ontology (PATO):** PATO provides a standardized vocabulary for describing phenotypic qualities, such as color, size, shape, and behaviour (Gkoutos *et al.*, 2004). It's used to describe phenotypes associated to genetic data in a consistent and structured manner. See http://obofoundry.org/ontology/pato.html for more detail

In addition, there are several ontologies specifically designed for representing plant genetics and genomics concepts and relationships related to phenotype. While most common are generic, some of them are designed specifically for an organism (for specific ontologies see databases in **Table 6**). The most common are:

- **PO:** it is a structured vocabulary and database resource that links plant anatomy, morphology and growth and development to plant genomics data (Jaiswal *et al.*, 2005). It contains more that 200 terms used to annotate metadata for genetic studies, gene expression data, phenotypic traits, and other plant-related information. For detailed info see https://github.com/Planteome/plant-ontology
- **TO:** it provides more than 5000 standardized terms for describing phenotypic traits in plants, including morphological, physiological, and molecular traits (Jaiswal *et al.*, 2002; Arnaud *et al.*, 2012). It supports data integration and analysis across different plant species. See https://github.com/Planteome/plant-trait-ontology for more information.
- **AgrO:** is an ontology for representing agronomic practices, techniques, variables and related entities. AgrO provides more than 4000 terms provides terms from the agronomy domain such data on field management, soil, weather and crop phenotypes. For detailed info see https://bigdata.cgiar.org/resources/agronomy-ontology/ for more information.
- **CO:** provides crop-specific ontology descriptions of agronomic, morphological, physiological, quality, and stress traits along with a standard nomenclature for different plant species (Shrestha *et al.*, 2012). In August 2023 CO comprises 4,235 traits and 6,151 variables for 31 plant species. See https://cropontology.org/ for more detailed info.
- **PECO:** it describes the treatments, growing conditions, and/or study types used in plant biology experiments (Jaiswal *et al.*, 2002). Currently has more than 3000 entries. See https://github.com/Planteome/plant-experimental-conditions-ontology for more information.
- **NCBI Taxonomy:** is a comprehensive hierarchical classification system that organizes and categorizes all known living organisms. It provides a curated and standardized classification and nomenclature for all of the organisms in the public sequence databases. NCBI Taxonomy provides a unique taxonomic identifier (TaxID) for each species, allowing researchers to accurately identify and reference organisms in genetic studies and to integrate genetic data from different sources by providing a common framework for species classification. Currently represents about 10% of the described species of life on the planet. For more details see https://www.ncbi.nlm.nih.gov/taxonomy.
- **ChEBI: is a** dictionary of molecular entities focused on ‘small’ chemical compounds (Hastings *et al.*, 2016). It incorporates an ontological classification of 183,420 compounds, whereby the

relationships between molecular entities or classes of entities and their parents and/or children are specified. See https://www.ebi.ac.uk/chebi/init.do.

3.6 Review of existing Minimum reporting guidelines

Minimum reporting guidelines outline the necessary and sufficient information vital for contextualizing and understanding a digital object. They are conformed by rules, guidelines, or recommendations established by consensus and approved by a recognized body. They accurately describe the plurality of relevant attributes of experiments, protocols, and their results by harmonizing both the structure of the (meta)data and their descriptions using a controlled vocabulary. The ISA (Investigation/Study/Assay) model is often used as a framework for capturing and organizing experimental metadata to ensure compliance with Minimum Information Reporting Standards. Our survey has rendered 26 Minimum reporting guidelines for genetic data that can be used for all species (**Annex 1**).

3.6.1. Minimum reporting guidelines for sequences

The most accepted at this moment for genetic data are Minimum Information about any (x) Sequence (MIxS) developed (Field *et al.*, 2008; Yilmaz *et al.*, 2011) by the Genomic Standards Consortium (GSC) (http://www.gensc.org/pages/standards/checklists.html). They have been adopted by several public repositories for genetic data, including those appertaining to the INSDC and The Genomes Online Database (GOLD; https://gold.jgi.doe.gov/). MixS initiative provides a set of standards and guidelines for reporting biological sequence data in a consistent and informative manner. These standards are designed to ensure that essential contextual information accompanies sequence data submissions to public databases, improving the quality and usability of the data for researchers and data users. The following checklists are under the MIxS umbrella:

- MIGS: Minimum information about a genome sequence
- MIMS: Minimum information about a metagenome sequence
- MIMARKS: Minimum information about a marker gene sequence
- MISAG: Minimum information about a single amplified genome sequence
- MIMAG: Minimum information about a metagenome-assembled genome sequence
-

MIGS and MIMARKS are further divided into additional subchecklists, based on the genome sequence in question, or the sequencing type.

- MIGS-EU: MIGS for eukaryotic genome sequences
- MIGS-BA: MIGS for bacterial and archaeal genome sequences
- MIGS-PL: MIGS for plasmid sequences
- MIGS-VI: MIGS for viral genome sequences
- MIGS-ORG: MIGS for organelle sequences
- MIMARKS-SP: MIMARKS-survey for marker gene sequences obtained directly from the environment
- MIMARKS-SU: MIMARKS-specimen for marker gene sequences from cultured or voucher-identifiable specimens

In general, all MIxS checklist covered the following concepts for describe and structure metadata:

Sample, Environmental Context and Investigation:
- **Project name**: Name of the project within which the sequencing was organized.
- **Biological Source:** Information about the biological origin of the sample, including taxonomic classification and ecological context.
- **Geographical Location:** Geographic coordinates and environmental context of the sampling site.
- **Environmental Parameters:** Relevant environmental conditions of the sampling site.

Sequencing and Experimental Details:
- **Sequencing Platform:** Information about the sequencing instrument or platform used.
- **Sequencing Method:** Details about the sequencing method employed, such as shotgun or targeted sequencing.
- **Library Preparation:** Information about how the sequencing library was prepared.

Bioinformatics and Data Processing:
- **Data Processing Steps:** Description of the steps taken during data processing and analysis.
- **Assembly and Annotation:** Details about genome assembly and annotation methods used.
- **Quality and Reliability:**
- **Quality Control:** Information about quality control measures applied to the data.
- **Coverage and Read Statistics:** Data on read depth, coverage, and other relevant statistics.

Accession and Submission:
- **Accession Numbers:** Any accession numbers associated with the data submission.
- **Database and Repository:** Information about the public database or repository where the data is being submitted.

The whole list of minimum information required for the current release (v6.0) for each MIxS checklist is in **Annex 5**. This file contains the complete list of core descriptors, which are shared by all MIxS checklists and, the checklist sequence type-specific descriptors. Users of MIxS should first determine the type of genome or marker gene sequence that they have, and then proceed to complement the core descriptors with the checklist specific mandatory descriptors to achieve MIxS compliant metadata. See http://www.gensc.org/pages/standards/checklists.html and https://github.com/GenomicsStandardsConsortium/mixs/tree/main for more detailed information.
In addition, MIxS team have developed a series of environmental packages which expand on the basic set of mandatory terms to help flesh out metadata expectations for more specific environmental samples (**Annex 5**). These packages and the combination of MixS with enviromental packages are described in more detail on https://genomicsstandardsconsortium.github.io/mixs/

One of the weaknesses of MIxS in relation to linkage with EURISCO is that MIxS Sample, Environmental Context and Investigation information is not fully compliant with other standards widely recognized for plants such as The FAO/Biodiversity Multicrop Passport Descriptors (MCPD; https://alliancebioversityciat.org/publications-data/faoipgri-multi-crop-passport-descriptors-mcpd) and Minimal Information About Plant Phenotyping Experiments (MIAPPE, http://www.miappe.org/). Despite this, MIxS enable to link plant genetic resources in EURISCO to sequences in INSDC repositories (such as BioSamples) through source materials identifiers (MIXS:0000026: source_mat_id), that can be used for PGR DOIs and (MIXS:0001107: samp_name) for BiosampleID.

3.6.1. Review of existing Minimum reporting guidelines for tracks metadata

A genetic data track refers to a visual representation of genetic or sequence-related information plotted along a linear or two-dimensional axis. These tracks are commonly used in genomics and bioinformatics to display various types of genetic and genomic data, such as DNA sequences, gene annotations, sequence variants, epigenetic modifications but they can also be queried by statistical analysis tools (Raney *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013; Martin *et al.*, 2023). The visualization of data in tracks allows researchers to explore and analyse complex genetic information in a comprehensible manner. Genetic data tracks are often displayed in genome browsers or visualization tools, where they are stacked vertically or horizontally to provide an overview of the genome or a specific region of interest. Each track may represent different types of data, with its own unique visual representation and annotations. File formats used to represent tracks were not designed with FAIR data principles https://f1000research.com/articles/10-268/v1(Wilkinson *et al.*, 2016), in particular with respect to metadata (Gundersen *et al.*, 2011). Thus, their potential impact through re-use is greatly limited. Recently a draft for genomic track metadata (FAIRtracks) has been proposed (Gundersen *et al.*, 2021). Metadata are implemented as a set of JSON Schemas organized around four (meta)data models, namely studies, samples, experiments and track (see https://github.com/fairtracks/fairtracks_standard#overview-of-structure-of-the-fairtracks-standard and https://fairtracks.net/fair/#fair-02-metadata-models). The minimal set of attributes of each metadata model are:

- FAIRtracks document: Version id, version date, ontology versions, URL to original source
- Track Collection: Name, description, URL to original source, contact info
- Study: Name, publications, contact info
- Sample: Species, biospecimen class, sample type (e.g., specific cell type, cell line, tissue), phenotype
- Experiment: (Sample OR upstream experiment), technique, biological target (e.g., gene, motif, phenotype), lab/compute protocols
- Track: Assembly details, file URL, label, description, ID of source collection, IDs of raw files, file format, type of condensed data, genometric track type, checksum

Each metadata FAIRtrack model can be mapped to objects in other revelat metadata models (**Table 8**).

Table 8. *Mapping of FAIRtracks objects to objects in other metadata standards (Source (Gundersen et al., 2021)).*

FAIRtrack	INSDC	ISA	Other	Comments
Track collection	SRA: Submission EGA: Dataset	Investigation	Track Hub Registry: Track Hub GSuite: Track collection	Can represent both original repository submissions, as well as other sets of track files, *e.g.*, as analysed in a research paper.
Study	Study	Study		
Sample	Sample	Sample		
Experiment	Experiment & Analysis	Assay & Process		"aggregated_from" attribute allows provenance through all experimental steps
Track	Analysis	Data	Track Hub Registry/ GSuite: Track	"raw_file_ids" can link to original data files in case a full experiment trace is not available "source_coll_ref" links to the source track collection if current collection is an *ad hoc* mix

Metadata recomendations for FAIRtracks applied to Ensembl TrackHub tracks (https://www.ensembl.org/info/website/trackhubs/index.html) recommend to store track file in binary formats such as BigBed, BigWig or BCF, link experiments and studies to the European Genome-phenome Archive (EGA)https://f1000research.com/articles/10-268/v1 and Samples to as described in BioSamples. For more detail of FairTracks and tools see https://fairtracks.net/.

3.6.2 Review of existing standards for visual representation of genetic data

Visual representation of genetic data plays a crucial role in understanding the complexity and relationships within genomes, including the presentation of variants. There are some standards developed.

One of the most used standards for visual representation of genetic data is the Synthetic Biology Open Language Visual (SBOL Visual) (Baig *et al.*, 2021). This is a standard notation and visual representation system designed specifically to represent synthetic biology designs and genetic constructs in a standardized and intuitive way. It provides a graphical language for depicting biological parts, devices, and genetic circuits, making it easier for researchers to communicate complex synthetic biology concepts and designs. See more details in https://sbolstandard.org/.

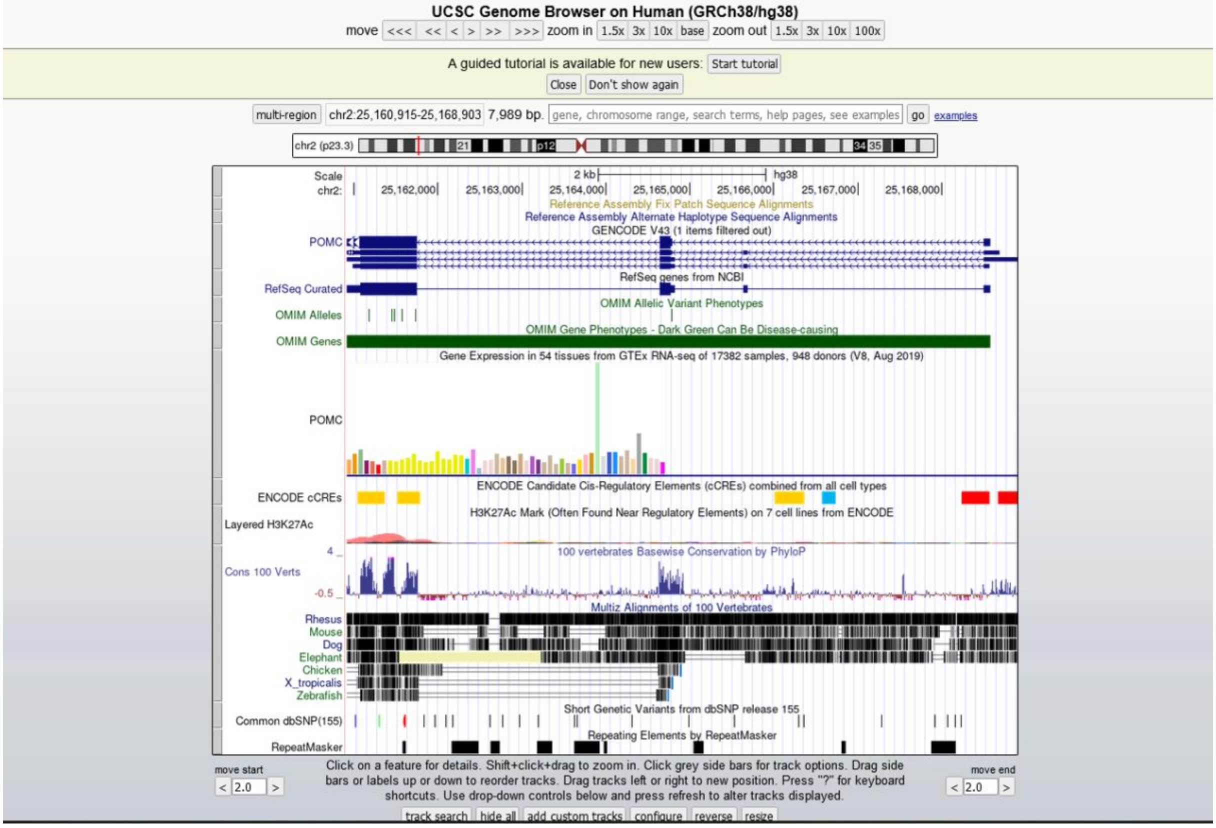
Key features of SBOL Visual include:

- **Standardized Symbols**: SBOL Visual provides a set of standardized symbols and glyphs that represent various biological components, such as promoters, coding sequences, terminators, and more. These symbols are designed to be easily recognizable and consistent across different design representations.
- **Hierarchy and Composition:** SBOL Visual allows for the hierarchical representation of genetic constructs. Complex designs, such as genetic circuits composed of multiple components, can be depicted in a way that shows the arrangement and connectivity of the different parts.
- **Visual Annotation:** Annotations can be added to the symbols to provide additional information, such as the name of the component, its sequence, and other relevant metadata.

Since this SBOL Visual aims to enhance the communication and sharing of synthetic biology designs is not directly applicable for the integration of genetic data into EURISCO, however it would be useful for illustrating future genetic engineered plants.

In addition, genome browsers offer powerful **visualization** tools for genetic data, an interactive interface that allows users to navigate, analyse, and interpret genomic information, including genes, regulatory elements, variants, and more **in the context of the genome** and gain insights into **their functional effects and relationships**, which is crucial for interpreting complex genomic information and making informed research decisions. Some of the most widely used genome browsers include:

- **UCSC Genome Browser** (https://genome.ucsc.edu/: is one of the most widely used genome browsers. It provides a user-friendly interface to explore genomes of various species and offers a wide range of tracks for genes, transcripts, regulatory elements, genetic variants, epigenetic marks, and more, allowing multiple annotation tracks and data visualization options (**Figure 3**). UCSC Genome Browser allows users to upload and visualize their own data alongside the reference genome.



UCSC Genome Browser on Human (GRCh38/hg38)

A guided tutorial is available for new users:

Click on a feature for details. Shift+click+drag to zoom in. Click grey side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts. Use drop-down controls below and press refresh to alter tracks displayed.

Figure 3. *Snapshot of UCSC Genome Browser multiple track view*

- **Ensembl Genome Browser** (http://www.ensembl.org, https://plants.ensembl.org) (**Figure 4**): it offers a comprehensive genome annotation for a wide range of species. Ensembl provides gene annotations, transcripts, regulatory elements, variation data, and more (Yates *et al.*, 2022). It allows users to visualize their data in the context of the Ensembl genome, supports advanced searching and filtering options and offers a REST API for programmatic access to data.

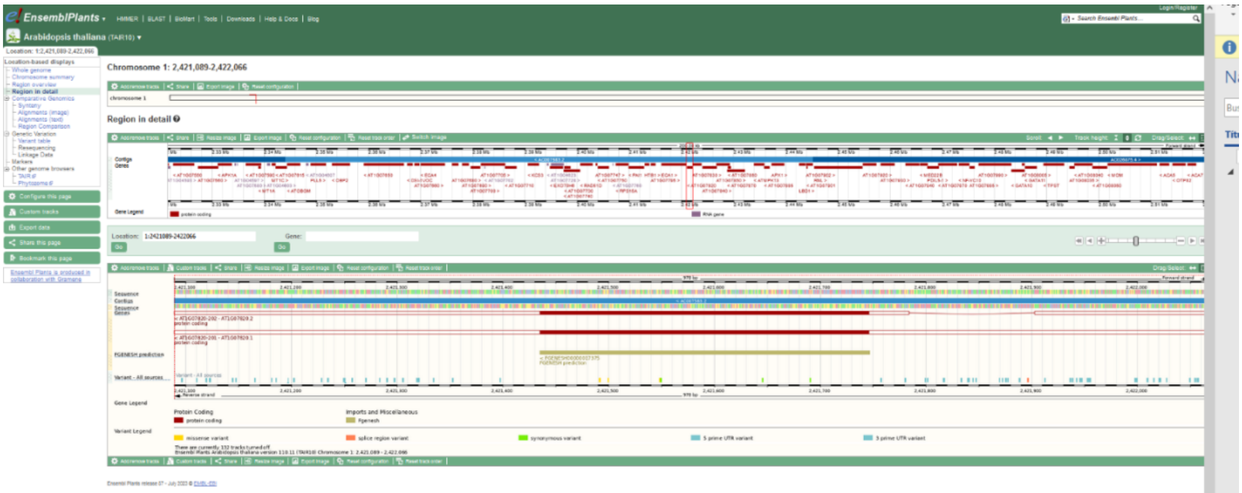


Figure 4. *Snapshot of Ensembl Genome Browser*

- **NCBI Genome Data Viewer** (https://www.ncbi.nlm.nih.gov/genome/gdv/) (**Figure 5**): developed by the National Center for Biotechnology Information (NCBI), it integrates data from GenBank, RefSeq, and other NCBI databases and provides interactive visualization of genomes with features such as genes, transcripts, and genetic variations. Also allows users to upload custom tracks.

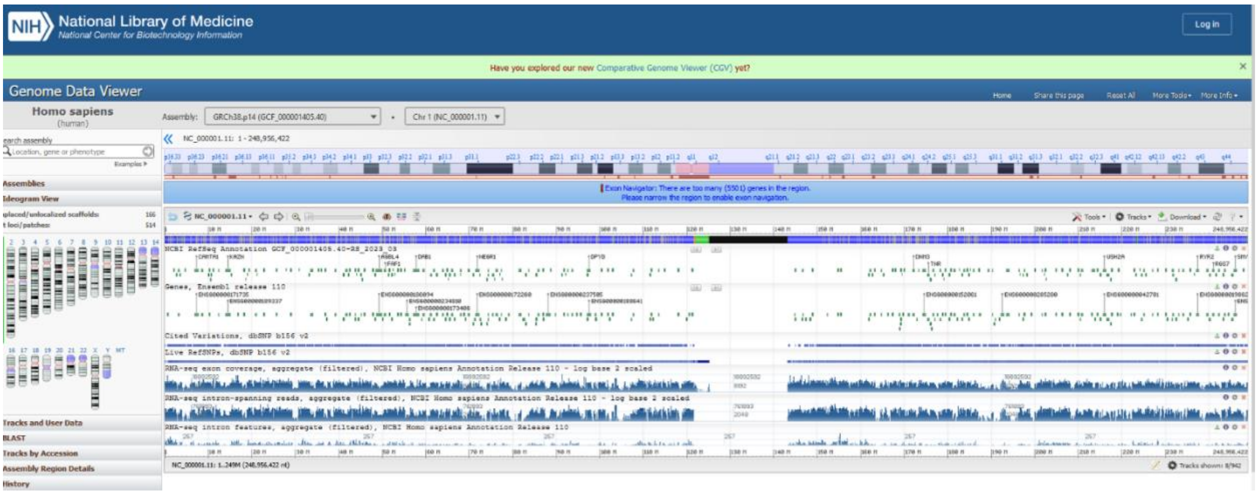
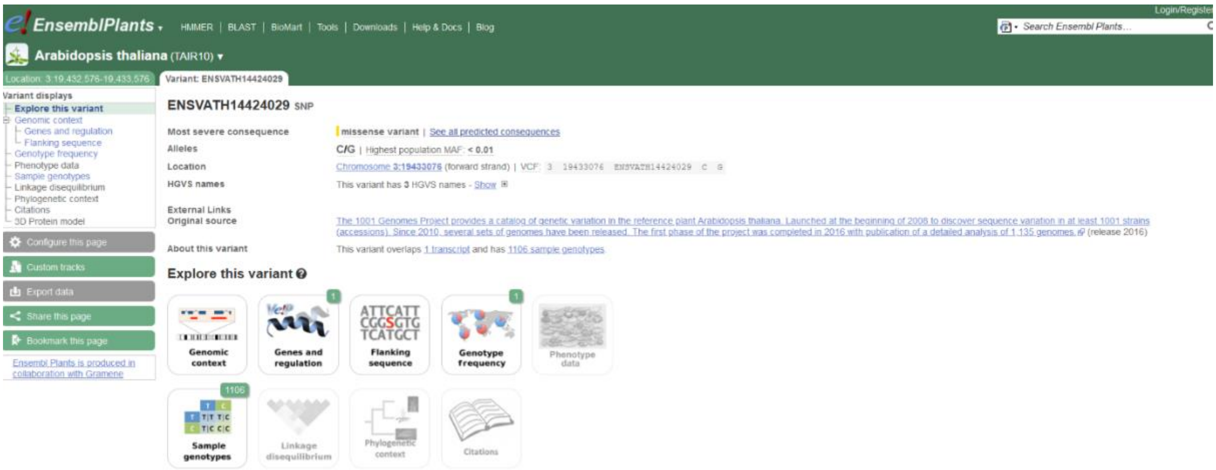


Figure 5. Snapshot of NCBI Genome Data Viewer

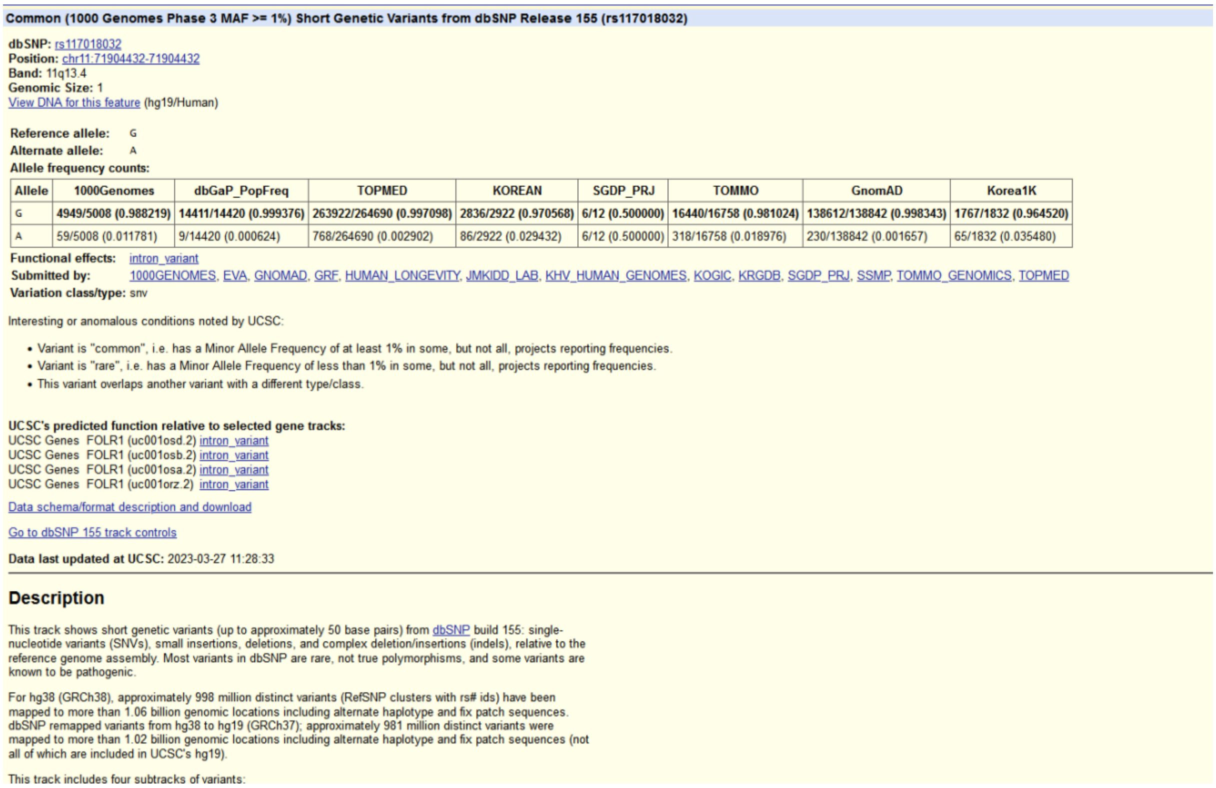
In addition, genetic variants, such as single nucleotide polymorphisms (SNPs) and insertions/deletions (indels), can be visually represented in genome browsers. Besides visual representation, these browsers also often include the position of the variant on the genome, allele frequencies, functional annotations, linkage disequilibrium (LD) pattern and phenotypic associations (**Figures 6-7**)



ENSVATH14424029 SNP

Most severe consequence: missense variant | See all predicted consequences
Alleles: C/G | Highest population MAF: < 0.01
Location: Chromosome 3:19433076 (forward strand) | VCF: 3 19433076 ENSVATH14424029 C G
HGVS names: This variant has 3 HGVS names - Show
External Links
Original source: The 1001 Genomes Project provides a catalog of genetic variation in the reference plant Arabidopsis thaliana. Launched at the beginning of 2008 to discover sequence variation in at least 1001 strains (accessions). Since 2010, several sets of genomes have been released. The first phase of the project was completed in 2016 with publication of a detailed analysis of 1,135 genomes. (release 2016)
About this variant: This variant overlaps 1 transcript and has 1106 sample genotypes.

Figure 6. Snapshot variant description in EnsemblPlants.



Common (1000 Genomes Phase 3 MAF >= 1%) Short Genetic Variants from dbSNP Release 155 (rs117018032)

dbSNP: rs117018032
Position: chr11:71904432-71904432
Band: 11q13.4
Genomic Size: 1
View DNA for this feature (hg19/Human)

Reference allele: G
Alternate allele: A
Allele frequency counts:

Allele	1000Genomes	dbGaP_PopFreq	TOPMED	KOREAN	SGDP_PRJ	TOMMO	GnomAD	Korea1K
G	**4949/5008 (0.988219)**	**14411/14420 (0.999376)**	**263922/264690 (0.997098)**	**2836/2922 (0.970568)**	**6/12 (0.500000)**	**16440/16758 (0.981024)**	**138612/138842 (0.998343)**	**1767/1832 (0.964520)**
A	59/5008 (0.011781)	9/14420 (0.000624)	768/264690 (0.002902)	86/2922 (0.029432)	6/12 (0.500000)	318/16758 (0.018976)	230/138842 (0.001657)	65/1832 (0.035480)

Functional effects: intron_variant
Submitted by: 1000GENOMES, EVA, GNOMAD, GRF, HUMAN_LONGEVITY, JMKIDD_LAB, KHV_HUMAN_GENOMES, KOGIC, KRGDB, SGDP_PRJ, SSMP, TOMMO_GENOMICS, TOPMED
Variation class/type: snv

Interesting or anomalous conditions noted by UCSC:

- Variant is "common", i.e. has a Minor Allele Frequency of at least 1% in some, but not all, projects reporting frequencies.
- Variant is "rare", i.e. has a Minor Allele Frequency of less than 1% in some, but not all, projects reporting frequencies.
- This variant overlaps another variant with a different type/class.

UCSC's predicted function relative to selected gene tracks:
UCSC Genes FOLR1 (uc001osd.2) intron_variant
UCSC Genes FOLR1 (uc001osb.2) intron_variant
UCSC Genes FOLR1 (uc001osa.2) intron_variant
UCSC Genes FOLR1 (uc001orz.2) intron_variant

Data schema/format description and download

Go to dbSNP 155 track controls

Data last updated at UCSC: 2023-03-27 11:28:33

Description

This track shows short genetic variants (up to approximately 50 base pairs) from dbSNP build 155: single-nucleotide variants (SNVs), small insertions, deletions, and complex deletion/insertions (indels), relative to the reference genome assembly. Most variants in dbSNP are rare, not true polymorphisms, and some variants are known to be pathogenic.

For hg38 (GRCh38), approximately 998 million distinct variants (RefSNP clusters with rs# ids) have been mapped to more than 1.06 billion genomic locations including alternate haplotype and fix patch sequences. dbSNP remapped variants from hg38 to hg19 (GRCh37); approximately 981 million distinct variants were mapped to more than 1.02 billion genomic locations including alternate haplotype and fix patch sequences (not all of which are included in UCSC's hg19).

This track includes four subtracks of variants:

Figure 7. *Snapshot variant description in UCSC Genome Browser*

3.6.3 Review of existing standards and public repositories for linked genotype and phenotype data.

Characterization/sequencing of PGRFA is regularly performed on cultivars, landraces, farmer's breeds, or even crop wild relatives. This aims at linking phenotype and genotype (QTL and GWAS) and generally produces large amounts of (digital) data, but underlying genotype data and the associated sample and population metadata have not been routinely submitted to appropriate archives. Actually there are some standards for (meta)data of GWAs such as seqGWAS (McMahon *et al.*, 2021) and those used in GWAS Catalog (https://www.ebi.ac.uk/gwas/home) or in the database of Genotypes and Phenotypes (dbGaP) dbGAP (https://www.ncbi.nlm.nih.gov/gap/) that can handle with array-based genotyping and sequencing-based association studies, however they are only dedicated to human biomedical research.

Several other international initiatives are on-going to agree on data models that specify standards and for different types of data in relation to linking plant phenotype to genotype facilitating the exploration of genetic variations, QTLs, genetic markers, and phenotypic traits, enabling researchers to better understand the genetic basis of plant traits and develop improved crop varieties. Most of these initiatives to standardize QTL and GWAs studies and their results are private or organism-specific initiatives , such as Gramene (https://archive.gramene.org/qtl/; (Tello-Ruiz *et al.*, 2021)), SoyBase (https://www.soybase.org/search/qtllist_by_symbol.php, https://www.soybase.org/GWAS/list.php) , MaizeGDB (https://curation.maizegdb.org/qtl.php), WheatIS (https://urgi.versailles.inrae.fr/wheatis/), CottonGen (https://www.cottongen.org/tripal_megasearch?datatype=tripal_megasearch_qtl) or Solgenomics (https://solgenomics.net/search/phenotypes/qtl)

Only a few initiatives aim to involve many species. The earliest one was **Minimum Information for QTLs and Association Studies (MIQAS). MIQAS** is a standard composed by a series of file formats designed to describe in a standardized way the results from a QTL or association study (**Figure 8**). They include

- **Experiment description file**: Contains meta-information and provides links to the rest of the data.
- ***Population description file***: Describes the population used for the analysis.
- **Pedigree file:** Contains sex and pedigree for the individuals.
- **Trait description file:** Describes the trait(s) under investigation. More than one trait can be covered in a single file.
- **Marker description file:** Provides accession numbers and possible alleles for all markers.
- **Map file:** Provides the genetic map for the markers as generated on the study sample.
- **Genotype data file:** Contains the genotypes for all markers vs individuals.
- **Phenotype data file:** Contains the phenotypes for all traits vs individuals.
- **Results file:** Contains specifics for the analysis and location of qtls and associated markers.

Despite the huge effort done for built MIQAS, it is not currently maintained and no widely implemented.

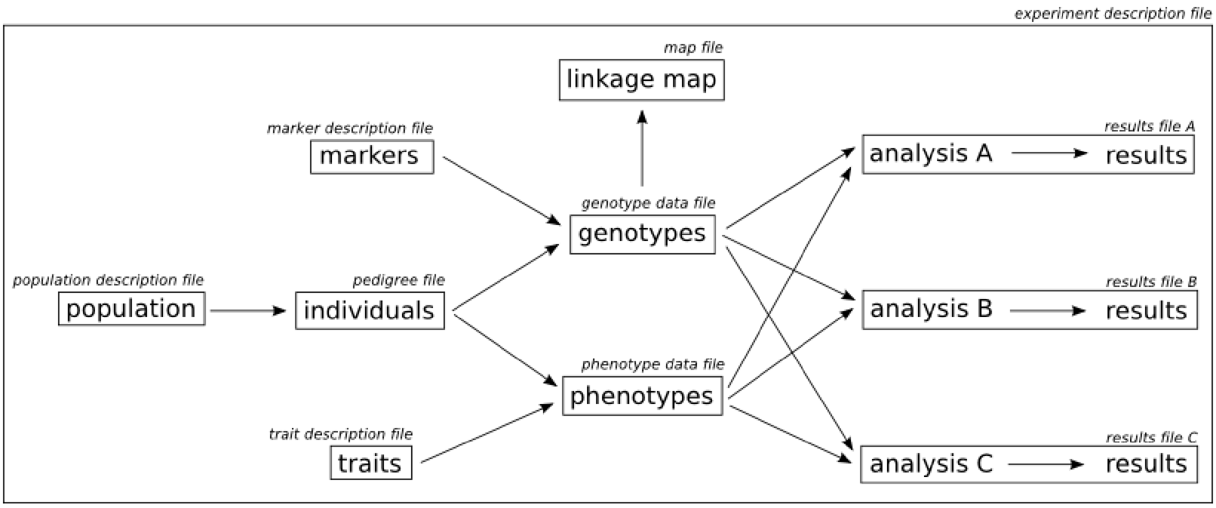


Figure 8. *Overview of experiment design that is covered by the MIQAS_TAB specification (source https://miqas.sourceforge.net/specification/MIQAS_TAB/MIQAS_TAB_specification.html)*

In addition, the French National Institute for Agricultural, Food and Environment (INRAE) initative GnpIS (Steinbach *et al.*, 2013)(http://urgi.versailles.inra.fr/gnpis) links phenomic, genetic, and genomic data and handles different types of data in the scope of genetics and genomics for plants including forest trees and fungi: genetic resources, polymorphisms and genotyping data, phenotyping data, association data, genetic maps and QTLs data.

GnpIS provides a series of complete and agreed FAIR compilant standards and ontologies for QTLs and GWAs (meta)data (**GnpisQTL, GnpisMetaQTL and GnpisAssso**) that are relevant for all species. See https://urgi.versailles.inrae.fr/Data for more details. See **Table 9** for a summary of main (meta)data to be submitted in these standards.

Table 9. *Main information required for submission in GnpisQTL, GnpisMetaQTL and GnpisAssso standards*

Sheet	Description	QTL	MetaQTL	Association
Contact	Professional information about persons involved in the project	**Mandatory**	**Mandatory**	**Mandatory**
Project	Information about projects	**Mandatory**	**Mandatory**	**Mandatory**

Institution	Information about institutions	-	-	**Mandatory**
Map	Information about maps	**Mandatory**	**Mandatory**	
Plant material	Information about plant accessions studied	-	-	**Mandatory**
Population	Information about populations studied	**Mandatory**	**Mandatory**	
Experiment	Information on experimentations	**Mandatory**	**Mandatory**	**Mandatory**
Variable	List of variables observed in this trial. Ideally they are part of a well identified ontology.	-	-	**Mandatory**
Measure	Measures for each experimentation	**Mandatory**	**Mandatory**	-
Marker	Description of the marker genomic location associated with the data.	**Mandatory**	**Mandatory**	**Mandatory**
Mapping	Localization of markers on a map	**Mandatory**	**Optional**	-
QTLDetection	Information on QTL detections	**Mandatory**	**Mandatory**	-
QTL	Localisation of QTLs on a map	**Mandatory**	**Mandatory**	-
QTL XREF	Description of the references for the QTL	**Optional**	-	-
MetaQTL	Information about metaQTLs	-	**Mandatory**	-
MetaAnalysis	Information about meta-analysis	-	**Mandatory**	-
MappingExperiment	Parameters needed for MetaQTL algorithm	-	**Mandatory**	-
Software	Software used	-	**Mandatory**	
Association analysis	associations analyses performed for this experiment	-	-	**Mandatory**
Linkage disequilibrium	Description of the Linkage Desequilibrium associated with the data.	-	-	**Optional**
Kinship	Description of the kinship associated with the data.	-	-	**Optional**
Structure	Description of the structure associated with the data.	-	-	**Mandatory**
Association results	results of association associated with the data	-	-	**Mandatory**
Software	Software used	-	**Mandatory**	-
Publication	References of each publication	**Optional**	**Optional**	**Optional**

The last initiative for linking plant genetic data to phenotypic data comes from the ELIXIR Plant Community (https://elixir-europe.org/communities/plant-sciences). ELIXIR Plant Community has created **The plant genomics tool assembly** (https://rdmkit.elixir-europe.org/plant_genomics_assembly), a Research Data Management Kit (RDMkit) for FAIR-ification of Plant Genotyping Data and its linking to Phenotyping (**Figure 9**). This online guide contains a set of instructions for managing plant genomics and genotyping data throughout their life cycle, with a particular focus on ensuring traceability of the biological material to enable interoperability with plant phenotyping data.

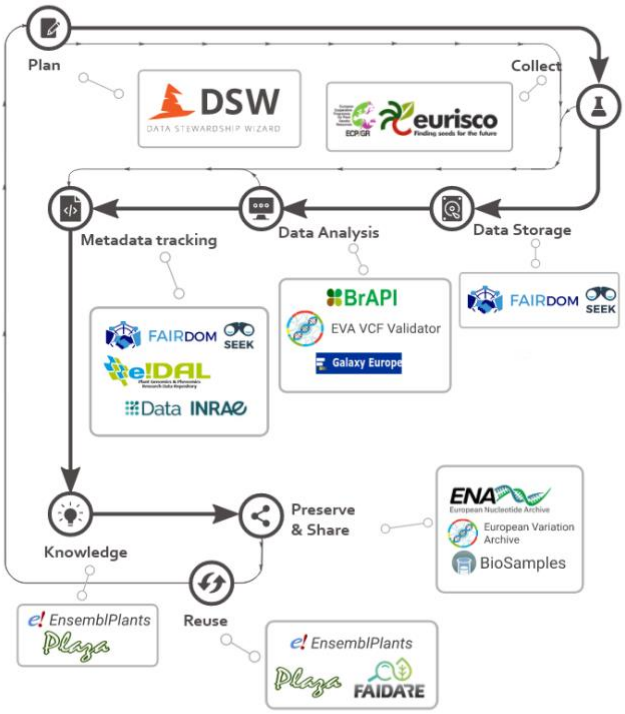


Figure 9. *The plant genomics tool assembly (source https://rdmkit.elixir-europe.org/plant_genomics_assembly).*

The roadmap of the Plant Genomics Tool Assembly recommends to submit samples to Biosamples, before submitting genomic data, with MIAPPE compliance validated using BioSamples' plant-miappe.json template. To ensure interoperability with plant phenotyping and genotyping data, the same PUIDs must be used in both the genotyping and phenotyping experiments as well in germplasm repository, preferably EURISCO. It is recommended that the biological plant material is accurately described using MCPD. Submission of sample descriptions to BioSamples can be done as early as the data collection stage, but at the latest, must accompany submission of the genomic data to the European Nucleotide Archive (ENA) or of genotyping data to the European Variation Archive (EVA). Reference genomes for genome assembly and annotation should be obtained from Ensembl Plants or PLAZA (https://bioinformatics.psb.ugent.be/plaza/), if available. Genetic variant data must be produced in the VCF format enriched with VCF metadata specifications (Beier *et al.*, 2022). All sequencing data collected in plant genotyping experiments should be submitted to ENA together with metadata compliant to the GSC MIxS plant associated checklist. The final results of such studies in the form of VCF files should be submitted to EVA. Variant data and its associated phenotype can be visualized and reused from Ensembl Plants, PLAZA or FAIR Data-finder for Agronomic REsearch FAIDARE (https://urgi.versailles.inrae.fr/faidare/).

3.7 Quality control standards: how to QC data

Quality control (QC) thresholds are often used to assess the quality and reliability of genetic data, including data related to plant genetics. One point does not record in standards and file reviewed is quality control (QC) thresholds for submitting genetic data to public repositories. While there is no specific universal QC threshold standards that apply exclusively to plant genetics, many general QC practices and guidelines are commonly followed to ensure the quality of genetic data across various organisms, including plants. In fact, the responsibility to submit high quality sequence data to public repositories lies on the authors. Some common QC steps and considerations for plant genetics data might include:

- Sequence Quality:
    - Evaluating sequence quality scores and base-call accuracy.
    - Trimming low-quality bases and adapters from sequencing reads.
- Read Mapping:
    - Assessing the alignment quality of sequencing reads to the reference genome.
    - Filtering out poorly aligned or ambiguous reads.
- Variant Calling:
    - Applying filters to variant calls to exclude low-quality or spurious variants.
    - Considering variant depth and allele frequency thresholds.
- Duplicate Removal:
    - Identifying and removing PCR or sequencing duplicates that can affect downstream analyses.
- Coverage Analysis:
    - Evaluating the coverage depth across the genome to ensure even and sufficient coverage.
- Contamination Detection:
    - Checking for potential sample cross-contamination or contamination with other species.
- Batch Effects:
    - Identifying and addressing potential batch effects or technical variations in data.

- Replication and Validation:
  - Ensuring that results are reproducible and validated using independent methods or replicates.
- Genotype Concordance:
  - Comparing genotypes with known standards or replicate samples to assess concordance.
- Variant Annotation:
  - Annotating variants with functional information and considering the impact on genes and regulatory regions.

3.8 Recommendations for linkage of genetic data with EURISCO

Following the principle of requiring minimal changes to existing systems while maximizing advantages for users (standardization, long term storage, interoperability and funding), **one practical pathway of linking genetic data with germplasm sources, such as EURISCO, is to use repositories, knowledgebases and standards implemented by the International INSDC repositories** (Manzella *et al.*, 2023), e**specially those hosted in the EMBL-EBI.**

For linking genetic metadata between INSDC and EURISCO, we must consider the following steps:

- **Data Identification and Linking:** Use shared identifiers or accession numbers to link genetic data in INSDC with corresponding metadata in EURISCO.
- **Use Common Standards:** both INSDC and EURISCO support common metadata standards, we recommend using those standards to ensure data consistency and exchangeability.
- **Custom Integration:** Develop custom integration methods or middleware that can bridge the gap between the two systems, translating and transferring data as needed.
- **Data Transformation:** Convert the data from the format used by one system to a format understood by the other, ensuring data integrity and consistency.
- **Web Services and APIs:** Explore the use of web services or APIs that are specific to the types of data you're exchanging. INSDC and EURISCO might offer APIs or services that facilitate data exchange.
- **Consult Experts:** It's advisable to involve experts in data integration, biodiversity informatics, and genomics to guide the integration process effectively.

Considering these steps, we recommend:

- Follow The plant genomics tool assembly (https://rdmkit.elixir-europe.org/plant_genomics_assembly) for genetic data submission to EMBL-EBI
- Use the standard file formats (or their compressed versions) described in **Table 3** and if possible to use for metadata on these file header the same identifier scheme and vocabulary proposed in minimal list of metadata fields for VCF (**Table 4**) (Beier *et al.*, 2022), specially for the equivalents of reference, contig and sample (BioSample, germplasm DOI). This will allow to link, through precise sample identifiers provided by BioSamples, raw and aligned sequences and variant data to other information recorded in the same sample (e.g. phenotypic, transcriptomic, metabolomic) but also to passport data in germplasm resources such as EURISCO.

- For Gene ID use the established nomenclature standards for the genome version of the organism you are studying (**Table 6**). In addition, it should ensure that gene IDs correspond exactly to the gene/gene products for a given assembly.
- Genetic marker data use as unique ID the EVA #SS and #RS codes.
- For tracks, we recommend following FAIRtracks metadata standards, use GTtrack files with rich metadata and submit them to the ENA or SRA and register in Track hub.
- Use preferentially ontologies reported in **Table 7**
- Redefine plants MixS standards to be MCPD and MIAPPE compliant.
- For GWAs and QTLs there is a need to agree standards for the plant community. A good solution would be the GnpisQTL, GnpisMetaQTL and GnpisAssso standards, however they need be MCPD and MIAPPE compliant.
- Create a centralized repository similar to those GWAS Catalog (https://www.ebi.ac.uk/gwas/home) or database of Genotypes and Phenotypes (dbGaP) dbGAP (https://www.ncbi.nlm.nih.gov/gap/) for submitting standardized QTL and GWAS data, including those that are published in the literature
- Use Ensembl plants as standard for genome, genes and variant visualization tool.
- Define QC thresholds standards for each type of sequence submitted to the EMBL-EBI repositories.
- Promote the use of BrAPI by developing APIs to exchange associated metadata or phenotypic data linked to genetic data between INSDC and EURISCO and between INSDC and EURISCO and other repositories
- Join PRO-GRACE to ELIXIR plant community

3.9. Discussion

To better describe and understand the genome and PGR genetic diversity stored in EURISCO, whether through genotyping/re-sequencing studies or de novo sequencing of new genotypes, it is essential to integrate passport, phenotypic, and genotypic (meta)data. Both data and metadata should also be formatted in standardized representations to enable automated processing and prevent errors arising from manual manipulations, particularly in the case of extensive datasets. This necessitates a community-wide consensus on annotation guidelines and tools for data preparation.

In this deliverable, we assess the strengths and weaknesses of the most common standards and databases to evaluate their suitability for linking genetic data with EURISCO. Additionally, we propose a set of recommendations that the PRO-GRACE community can follow in pursuing this objective.

The success of these standards resides on community adoption, influenced by factors such as publisher enforcement and the availability and user-friendliness of associated tools (Adam-Blondon et al., 2016; Pommier et al., 2019; Papoutsoglou et al., 2020). To foster universal adoption of the recommended standards and their future implementation, we propose the following steps:

- Develop detailed guides and tutorials
- Store them at the EURISCO central hub for direct download, along with comprehensive guides and tutorials
- Host them on GitHub for collaborative editing and versioning
- Include them in RDMKit and Fairsharing collections to achieve wider dissemination
- Conduct training courses in collaboration with WP1 partners to educate and encourage the PGR research community to embrace reporting standards from the outset of their research

- Advocate for journal adherence to PGR reporting standards to ensure consistency in published datasets

Deviations

No deviations

References

Alghamdi D, Dooley DM, Samman M, Hsiao WWL. 2023. NGBO: Introducing -omics metadata to biobanking ontology. bioRxiv, 2023.05.09.539725.

Arnaud E, Cooper L, Shrestha R, Menda N, Nelson RT, Matteis L, Skofic M, Bastow R, Jaiswal P, Mueller L. 2012. Proceedings of the International Conference on Knowledge Engineering and Ontology Development.

Aubry S. 2023. Genebanking plant genetic resources in the postgenomic era. Agriculture and Human Values.

Baig H, Fontanarossa P, McLaughlin J, Scott-Brown J, Vaidyanathan P, Gorochowski T, Misirli G, Beal J, Myers C. 2021. Synthetic biology open language visual (SBOL visual) version 3.0. **18**.

Bandrowski A, Brinkman R, Brochhausen M, *et al.* 2016. The Ontology for Biomedical Investigations. PLOS ONE **11**, e0154556.

Beier S, Fiebig A, Pommier C, *et al.* 2022. Recommendations for the formatting of Variant Call Format (VCF) files to make plant genotyping data FAIR [version 2; peer review: 2 approved]. F1000Research **11(ELIXIR)**, 231.

Cochrane G, Karsch-Mizrachi I, Takagi T, Sequence Database Collaboration IN. 2016. The International Nucleotide Sequence Database Collaboration. Nucleic Acids Research **44**, D48–D50.

Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research **38**, 1767–1771.

Danecek P, Auton A, Abecasis G, *et al.* 2011. The variant call format and VCFtools. Bioinformatics **27**, 2156–2158.

Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, Ashburner M. 2005. The Sequence Ontology: a tool for the unification of genome annotations. Genome Biology **6**, R44.

Ewing B, Green P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. Genome Research **8**, 186–194.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-Calling of Automated Sequencer Traces UsingPhred. I. Accuracy Assessment. Genome Research **8**, 175–185.

Field D, Garrity G, Gray T, *et al.* 2008. The minimum information about a genome sequence (MIGS) specification. Nature Biotechnology **26**, 541–547.

Gkoutos G V, Green ECJ, Mallon A-M, Hancock JM, Davidson D. 2004. Using ontologies to describe mouse phenotypes. Genome Biology **6**, R8.

Gundersen S, Boddu S, Capella-Gutierrez S, Drabløs F, Fernández JM, Kompova R, Taylor K, Titov D, Zerbino D, Hovig E. 2021. Recommendations for the FAIRification of genomic track metadata [version 1; peer review: 2 approved]. F1000Research **10(ELIXIR)**, 268.

Gundersen S, Kalaš M, Abul O, Frigessi A, Hovig E, Sandve GK. 2011. Identifying elemental genomic track types and representing them uniformly. BMC Bioinformatics **12**, 494.

Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P,

Steinbeck C. 2016. ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic acids research **44**, D1214-9.

Ison J, Kalaš M, Jonassen I, Bolser D, Uludag M, McWilliam H, Malone J, Lopez R, Pettifer S, Rice P. 2013. EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. Bioinformatics **29**, 1325–1332.

Jaiswal P, Avraham S, Ilic K, *et al.* 2005. Plant Ontology (PO): a Controlled Vocabulary of Plant Structures and Growth Stages. Comparative and Functional Genomics **6**, 373740.

Jaiswal P, Ware D, Ni J, *et al.* 2002. Gramene: Development and Integration of Trait and Gene Ontologies for Rice. Comparative and Functional Genomics **3**, 432674.

Karsch-Mizrachi I, Takagi T, Cochrane G, Collaboration on behalf of the INSD. 2018. The international nucleotide sequence database collaboration. Nucleic Acids Research **46**, D48–D51.

Kotni P, van Hintum T, Maggioni L, Oppermann M, Weise S. 2023. EURISCO update 2023: the European Search Catalogue for Plant Genetic Resources, a pillar for documentation of genebank material. Nucleic Acids Research **51**, D1465–D1469.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**, 2078–2079.

Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, Zhukova A, Brazma A, Parkinson H. 2010. Modeling sample variables with an Experimental Factor Ontology. Bioinformatics **26**, 1112–1118.

Manzella D, Marsella M, Jaiswal P, Arnaud E, King B. 2023. Digital Sequence Information and Plant Genetic Resources: Global Policy Meets Interoperability BT - Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development. In: Williamson HF,, In: Leonelli S, eds. Cham: Springer International Publishing, 183–200.

Martin FJ, Amode MR, Aneja A, *et al.* 2023. Ensembl 2023. Nucleic Acids Research **51**, D933–D941.

McMahon A, Lewis E, Buniello A, Cerezo M, Hall P, Sollis E, Parkinson H, Hindorff LA, Harris LW, MacArthur JAL. 2021. Sequencing-based genome-wide association studies reporting standards. Cell Genomics **1**.

Mungall CJ, Emmert DB, Consortium TF. 2007. A Chado case study: an ontology-based modular scheme for representing genome-associated biological information. Bioinformatics **23**, i337–i346.

Nakamura Y, Cochrane G, Karsch-Mizrachi I, Collaboration on behalf of the INSD. 2013. The International Nucleotide Sequence Database Collaboration. Nucleic Acids Research **41**, D21–D24.

Nybom H, Lācis G. 2021. Recent Large-Scale Genotyping and Phenotyping of Plant Genetic Resources of Vegetatively Propagated Crops. Plants **10**.

Pearson WR, Lipman DJ. 1988. PNAS. Proceedings of the National Academy of Sciences **85**, 2444–2448.

Raney BJ, Cline MS, Rosenbloom KR, *et al.* 2011. ENCODE whole-genome data in the UCSC genome browser (2011 update). Nucleic Acids Research **39**, D871–D875.

Reiser L, Harper L, Freeling M, Han B, Luan S. 2018. FAIR: A Call to Make Published Data More Findable, Accessible, Interoperable, and Reusable. Molecular Plant **11**, 1105–1108.

Sansone S-A, McQuilton P, Rocca-Serra P, Gonzalez-Beltran A, Izzo M, Lister AL, Thurston M, Community the Fair. 2019. FAIRsharing as a community approach to standards, repositories and policies. Nature Biotechnology **37**, 358–367.

Sansone S-A, Rocca-Serra P, Field D, *et al.* 2012. Toward interoperable bioscience data. Nature Genetics **44**, 121–126.

Selby P, Abbeloos R, Backlund JE, *et al.* 2019. BrAPI—an application programming interface for plant breeding applications. Bioinformatics **35**, 4147–4155.

Shrestha R, Matteis L, Skofic M, Portugal A, McLaren G, Hyman G, Arnaud E. 2012. Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice . Frontiers in Physiology **3**.

Steinbach D, Alaux M, Amselem J, *et al.* 2013. GnpIS: an information system to integrate genetic and genomic data from plants and fungi. Database **2013**, bat058.

Tello-Ruiz MK, Naithani S, Gupta P, *et al.* 2021. Gramene 2021: harnessing the power of comparative genomics and pathways for plant research. Nucleic Acids Research **49**, D1452–D1463.

The Gene Ontology Consortium. 2017. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Research **45**, D331–D338.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics **14**, 178–192.

van Treuren R, van Hintum TJL. 2014. Next-generation genebanking: plant genetic resources management and utilization in the sequencing era. Plant Genetic Resources **12**, 298–307.

Visser U, Abeyruwan S, Vempati U, Smith RP, Lemmon V, Schürer SC. 2011. BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results. BMC Bioinformatics **12**, 257.

Wambugu PW, Ndjiondjop M-N, Henry RJ. 2018. Role of genomics in promoting the utilization of plant genetic resources in genebanks. Briefings in Functional Genomics **17**, 198–206.

Welter D, Juty N, Rocca-Serra P, *et al.* 2023. FAIR in action - a flexible framework to guide FAIRification. Scientific Data **10**, 291.

Wilkinson MD, Dumontier M, Aalbersberg IjJ, *et al.* 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data **3**, 160018.

Yates AD, Allen J, Amode RM, *et al.* 2022. Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. Nucleic Acids Research **50**, D996–D1003.

Yilmaz P, Kottmann R, Field D, *et al.* 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nature Biotechnology **29**, 415–420.

Annex 1. Databases, public repositories and standards in FAIRsharing.org related to genetic information

FAIRsharing URL	FAIRsharing DOI	Record Name	Record homepage URL	Resource Type	Manintenace	Ready-to-use	FAIRorg recommended	plant species accepted	Applied to species	Type	ELIXIR recommended	provider
		BrAPI The Breeding API	https://brapi.org/	standard	maintained	ready-to-use	recommended	All plant species		Model/formats or syntax:models/syntax	recommended	
https://fairsharing.org/93	**10.25504/FAIRsharing.hfgman**	**.ACE format**	http://en.wikipedia.org/wiki/ACE_%28genomic_file_format%29	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/2115	**10.25504/FAIRsharing.anpa6**	**A Systematic Annotation Package**	https://asap.ahabs.wisc.edu/asap/home.php	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1494	**10.25504/FAIRsharing.7tx4ac**	**Access to Biological Collection Data DNA extension**	http://www.tdwg.org/standards/640/	standard		ready-to-use			All species	Minimum reporting guidelines		
https://fairsharing.org/1738	**10.25504/FAIRsharing.8hcczk**	**Addgene**	http://www.addgene.org/	database	maintained	ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/2584	**10.25504/FAIRsharing.83wDfe**	**Ag Data Commons**	https://data.nal.usda.gov/	database	maintained	ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/1742	**10.25504/FAIRsharing.kn4ycg**	**AgBase**	http://www.agbase.msstate.edu/	database	maintained			All plant species		Knowledgebase		
https://fairsharing.org/2847	**10.25504/FAIRsharing.ZPRtfG**	**Agronomic Linked Data**	http://www.agrold.org/	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/2673	**10.25504/FAIRsharing.aIe9SA**	**Allosteric Mutation Analysis and Polymorphism of Signaling database**	http://allomaps.bii.a-star.edu.sg	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/4894	**N/A**	**AlphaFill**	https://www.alphafill.eu	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1050	**10.25504/FAIRsharing.9zpcgj**	**Amino Acid Ontology**	https://robertdavidstevens.wordpress.com/2010/08/09/an-ontology-of-amino-acids/	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/1552	**10.25504/FAIRsharing.s8vrb1**	**Apo and Holo structures DataBase**	http://ahdb.ee.ncku.edu.tw/	database					All species	Knowledgebase		
https://fairsharing.org/2894	**10.25504/FAIRsharing.a308a0**	**DDBJ BioProject**	https://www.ddbj.nig.ac.jp/bioproject/index-e.html	database		ready-to-use	recommended		All species	Data repository		DDBJ
https://fairsharing.org/1572	**10.25504/FAIRsharing.k337f0**	**DNA Data Bank of Japan**	https://www.ddbj.nig.ac.jp/ddbj	database	maintained	ready-to-use	recommended		All species	Data repository		DDBJ
https://fairsharing.org/3013	**10.25504/FAIRsharing.23823e**	**DDBJ Sequence Read Archive**	https://www.ddbj.nig.ac.jp/dra/index-e.html	database		ready-to-use	recommended		All species	Data repository		DDBJ
https://fairsharing.org/3188	**10.25504/FAIRsharing.b67925**	**DDBJ Trace Archive**	https://www.ddbj.nig.ac.jp/dta/index-e.html	database		ready-to-use			All species	Data repository		DDBJ
https://fairsharing.org/2743	**10.25504/FAIRsharing.hESBcy**	**Genomic Expression Archive**	https://www.ddbj.nig.ac.jp/gea	database	maintained	ready-to-use	recommended		All species	Data repository		DDBJ
https://fairsharing.org/1577	**10.25504/FAIRsharing.j1wj7d**	**Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups**	http://eggnog.embl.de	database		ready-to-use			All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/861	**10.25504/FAIRsharing.cp0fs9**	**ARLEQUIN Project Format**	http://cmpg.unibe.ch/software/arlequin35/	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/1645	**10.25504/FAIRsharing.hsbpq3**	**Simple Modular Architecture Research Tool**	http://smart.embl.de	database		ready-to-use			All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/1895	**10.25504/FAIRsharing.sed5tq**	**BAliBASE**	http://www.lbgi.fr/balibase/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2380	**10.25504/FAIRsharing.k4yzh**	**BCCM/GeneCorner Plasmid Collection**	http://www.genecorner.ugent.be/	database	maintained	ready-to-use			All species	Knowledgebase		UGENT(Beligum)
https://fairsharing.org/1859	**10.25504/FAIRsharing.hjybww**	**Ligand-Gated Ion Channel database**	http://www.ebi.ac.uk/compneur-srv/LGICdb/	database					All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/2156	**10.25504/FAIRsharing.txkh36**	**Database of Genomic Variants Archive**	http://www.ebi.ac.uk/dgva/	database	maintained		recommended		All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/2401	**10.25504/FAIRsharing.4njvzy**	**BCCM/MUCL Agro-food & Environmental Fungal Collection**	http://bccm.belspo.be/about-us/bccm-mucl	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/188	**10.25504/FAIRsharing.kfqVQ7**	**BcForms Grammar**	https://www.bcforms.org/#grammar	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/2630	**10.25504/FAIRsharing.1ef98c**	**BCL-2 Database**	https://bcl2db.lyon.inserm.fr/BCL2DB/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2240	**10.25504/FAIRsharing.k5a3yt**	**Big Data Nucleic Acid Simulations Database**	http://mmb.irbbarcelona.org/BIGNASim	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/380	**10.25504/FAIRsharing.b52795**	**BIM**	https://www.cog-genomics.org/plink2/formats#bim	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/4822	**N/A**	**Binary Call Format**	http://samtools.github.io/hts-specs/BCFv2_qref.pdf	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/4475	**10.25504/FAIRsharing.b2f583**	**BioData.pt Data Management Portal**	https://dmportal.biodata.pt/	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/2592	**10.25504/FAIRsharing.qqFfXF**	**BioInformatics Platform for Agroecosystem Arthropods**	https://bipaa.genouest.org	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1246	**10.25504/FAIRsharing.3fyavr**	**Bioinformatics Web Service Ontology**	https://github.com/obi-webservice/OBIws	standard					All species	Semantic standards		github
https://fairsharing.org/2512	**10.25504/FAIRsharing.pjj4gd**	**Biological and Chemical Oceanography Data Management Office**	https://www.bco-dmo.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1080	**10.25504/FAIRsharing.8ktkqy**	**Biological Collections Ontology**	https://github.com/BiodiversityOntologies/bco	standard	maintained	ready-to-use			All species	Semantic standards		github
https://fairsharing.org/2114	**10.25504/FAIRsharing.p06nme**	**Biological Magnetic Resonance Data Bank**	https://bmrb.io/	database	maintained	ready-to-use	recommended		All species	Data repository		
https://fairsharing.org/2212	**10.25504/FAIRsharing.73reht**	**Bio-Mirror**	http://www.bio-mirror.net/	database	maintained				All species	Knowledgebase		
https://fairsharing.org/467	**N/A**	**Bioversity Molecular Markers Ontology**	https://cropontology.org/term/CO_500:ROOT	standard				All plant species		Semantic standards		
https://fairsharing.org/3604	**10.25504/FAIRsharing.0ef9f5**	**Bologna Annotation Resource**	https://bar.biocomp.unibo.it/bar3/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2856	**10.25504/FAIRsharing.2cw3HU**	**BonaRes Repository**	https://datenzentrum.bonares.de/research-data.php	database	maintained	ready-to-use			All species	Data repository		
https://fairsharing.org/1850	**10.25504/FAIRsharing.dj8nt8**	**European Nucleotide Archive**	http://www.ebi.ac.uk/ena	database	maintained	ready-to-use	recommended		All species	Data repository	recommended	EMBL-EBI(EU)
https://fairsharing.org/1098	**10.25504/FAIRsharing.Hbvvle**	**BpForms Grammar**	https://www.bpforms.org/#grammar	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/3186	**10.25504/FAIRsharing.ac95d5**	**Bridges**	https://bridges.monash.edu/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/3566	**10.25504/FAIRsharing.313630**	**CABRI Guidelines for Catalogue Production**	http://cabri.org/guidelines/catalogue/CPcover.html	standard	maintained	ready-to-use		All plant species		Model/formats or syntax:models/syntax		
https://fairsharing.org/1333	**10.25504/FAIRsharing.2b3at8**	**Marine Microbial Biodiversity, Bioinformatics, Biotechnology Checklist**	http://www.ebi.ac.uk/ena/submit/microb3-checklist	standard		ready-to-use			All species	Minimum reporting guidelines		EMBL-EBI(EU)
https://fairsharing.org/2155	**10.25504/FAIRsharing.6824pv**	**European Variation Archive**	http://www.ebi.ac.uk/eva/	database	maintained	ready-to-use	recommended		All species	Data repository	recommended	EMBL-EBI(EU)
https://fairsharing.org/1560	**10.25504/FAIRsharing.r3pbp3**	**CAPS-DB : a structural classification of helix-capping motifs**	http://www.bioinsilico.org/CAPSDB	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1730	**10.25504/FAIRsharing.dkbt9j**	**Carbohydrate Structure Database**	http://csdb.glycoscience.ru/database/	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1852	**10.25504/FAIRsharing.7zffgc**	**Gene Ontology Annotation Database**	http://www.ebi.ac.uk/GOA	database	maintained	ready-to-use	recommended		All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/1588	**10.25504/FAIRsharing.f5zx00**	**Expression Atlas**	http://www.ebi.ac.uk/gxa	database	maintained	ready-to-use	recommended		All species	Data repository		EMBL-EBI(EU)
https://fairsharing.org/2047	**10.25504/FAIRsharing.2s4n8r**	**MEROPS**	http://www.ebi.ac.uk/merops/	database		ready-to-use	recommended		All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/1847	**10.25504/FAIRsharing.e58gcm**	**Chemical Component Dictionary**	http://www.ebi.ac.uk/pdbe-srv/pdbechem/	database		ready-to-use			All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/227	**10.25504/FAIRsharing.qz0rvm**	**Variant Effect Predictor data format**	http://www.ensembl.org/info/docs/tools/vep/vep_formats.html	standard		ready-to-use			All species	Model/formats or syntax:Format		EMBL-EBI(EU)
https://fairsharing.org/2386	**10.25504/FAIRsharing.923a0p**	**Ensembl Genomes**	https://ensemblgenomes.org/	database	maintained	ready-to-use		All plant species		Knowledgebase	recommended	EMBL-EBI(EU)

FAIRsharing URL	FAIRsharing DOI	Record Name	Record homepage URL	Resource Type	Manintenace	Ready-to-use	FAIRorg recommended	plant species accepted	Applied to species	Type	ELIXIR recommended	provider
https://fairsharing.org/2404	**10.25504/FAIRsharing.j8g2cv**	**Ensembl Plants**	https://plants.ensembl.org/index.html	database	maintained	ready-to-use		All plant species		Knowledgebase	recommended	EMBL-EBI(EU)
https://fairsharing.org/1845	**10.25504/FAIRsharing.6k0kwd**	**ArrayExpress**	https://www.ebi.ac.uk/arrayexpress/	database	maintained	ready-to-use	recommended		All species	Data repository	recommended	EMBL-EBI(EU)
https://fairsharing.org/1554	**10.25504/FAIRsharing.ewjdq6**	**BioSamples at the European Bioinformatics Institute**	https://www.ebi.ac.uk/biosamples/	database	maintained	ready-to-use	recommended		All species	Data repository	recommended	EMBL-EBI(EU)
https://fairsharing.org/1280	**10.25504/FAIRsharing.q72e3w**	**Short Read Archive eXtensible Markup Language**	https://www.ebi.ac.uk/ena/submit/read-xml-format-1-5	standard		ready-to-use	recommended		All species	Model/formats or syntax:models/syntax		EMBL-EBI(EU)
https://fairsharing.org/2691	**10.25504/FAIRsharing.A0ozDj**	**Genome Properties**	https://www.ebi.ac.uk/interpro/genomeproperties	database	maintained	ready-to-use			All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/2129	**10.25504/FAIRsharing.dxj07r**	**MGnify**	https://www.ebi.ac.uk/metagenomics/	database	maintained	ready-to-use	recommended		All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/1851	**10.25504/FAIRsharing.p3n3fn**	**Patent Data Resources**	https://www.ebi.ac.uk/patentdata	database		ready-to-use			All species	Knowledgebase		EMBL-EBI(EU)
https://fairsharing.org/378	**10.25504/FAIRsharing.h4rs6h**	**GO-PLUS**	http://geneontology.org/docs/download-ontology/#go_plus_owl	standard		ready-to-use			All species	Semantic standards		GBC(USA)
https://fairsharing.org/1466	**10.25504/FAIRsharing.rtm51**	**Gene Product Annotation Data**	http://geneontology.org/docs/gene-product-association-data-gpad-format/	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		GBC(USA)
https://fairsharing.org/1468	**10.25504/FAIRsharing.820ebm**	**Gene Product Information Format**	http://geneontology.org/docs/gene-product-information-gpi-format/	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		GBC(USA)
https://fairsharing.org/1095	**10.25504/FAIRsharing.IIMCe0**	**Gene Ontology Gene Association File Format 2.0**	http://geneontology.org/docs/go-annotation-file-gaf-format-2.0/	standard	maintained				All species	Model/formats or syntax:Format		GBC(USA)
https://fairsharing.org/766	**10.25504/FAIRsharing.FGJ2T8**	**Gene Ontology (GO) Gene Association File Format 2.1**	http://geneontology.org/docs/go-annotation-file-gaf-format-2.1/	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		GBC(USA)
https://fairsharing.org/1418	**10.25504/FAIRsharing.77fbbf**	**Gene Ontology (GO) Gene Association File Format 2.2**	http://geneontology.org/docs/go-annotation-file-gaf-format-2.2/	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		GBC(USA)
https://fairsharing.org/2077	**10.25504/FAIRsharing.s1ne3g**	**UniProt Knowledgebase**	https://www.uniprot.org	database	maintained	ready-to-use	recommended		All species	Knowledgebase		GBC(USA)
https://fairsharing.org/1237	**10.25504/FAIRsharing.xhwrnr**	**Cell Cycle Ontology**	https://www.biogateway.eu/tools/#CCO	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/2069	**10.25504/FAIRsharing.ntv8pe**	**Chemical Effects in Biological Systems**	https://doi.org/10.22427/NTP-DATA-1	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2748	**10.25504/FAIRsharing.9btRvC**	**China National GeneBank DataBase**	https://db.cngb.org/	database	maintained	ready-to-use	recommended		All species	Data repository		
https://fairsharing.org/3966	**10.25504/FAIRsharing.b52cce**	**CIRAD Dataverse**	https://doi.org/10.18167/infrastructure/00003	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/122	**10.25504/FAIRsharing.jx6ea0**	**CLUSTAL-W Alignment Format**	http://www.clustal.org/clustal2/	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/1429	**10.25504/FAIRsharing.8y5ayx**	**Common Workflow Language**	http://www.commonwl.org	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/1155	**10.25504/FAIRsharing.kay31r**	**Comparative Data Analysis Ontology**	https://github.com/evoinfo/cdao	standard	maintained	ready-to-use			All species	Semantic standards		github
https://fairsharing.org/1902	**10.25504/FAIRsharing.wyz5he**	**Conformation Angles Database**	http://cluster.physics.iisc.ernet.in/cadb/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/292	**10.25504/FAIRsharing.695abb**	**Connectivity Table file format**	http://rna.urmc.rochester.edu/Text/File_Formats.html#CT	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/889	**10.25504/FAIRsharing.VJDEtl**	**Cooler file format**	https://cooler.readthedocs.io/en/latest/datamodel.html	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/2688	**10.25504/FAIRsharing.02xjaW**	**CyVerse Data Common Repository**	http://datacommons.cyverse.org/	database	maintained	ready-to-use			All species	Data repository		
https://fairsharing.org/841	**10.25504/FAIRsharing.hgsFLe**	**Darwin Core Germplasm**	https://github.com/dagendresen/darwincore-germplasm	standard	maintained	ready-to-use			All species	Semantic standards		github
https://fairsharing.org/1566	**10.25504/FAIRsharing.evfe2s**	**Database of Aligned Ribosomal Complexes**	http://darcsite.genzentrum.lmu.de/darc/	database					All species	Knowledgebase		
https://fairsharing.org/2385	**10.25504/FAIRsharing.m58329**	**Database Of Local Biomolecular Conformers**	http://dolbico.org/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2350	**10.25504/FAIRsharing.jhjnp0**	**Database of local DNA conformers**	http://ich.vscht.cz/projects/dolce/viewHome	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1743	**10.25504/FAIRsharing.1bnhyh**	**Databases of Orthologous Promoters**	http://doop.abc.hu/	database				All plant species		Knowledgebase		
https://fairsharing.org/3010	**10.25504/FAIRsharing.NmIgg9**	**Datanator**	https://www.datanator.info	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2414	**10.25504/FAIRsharing.yyf78h**	**DataONE**	https://www.dataone.org	database	maintained	ready-to-use	recommended	All plant species		Knowledgebase		
https://fairsharing.org/2040	**10.25504/FAIRsharing.52qw6p**	**Description of Plant Viruses**	http://www.dpvweb.net	database		ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/3314	**10.25504/FAIRsharing.8d2e5c**	**DIGITAL.CSIC**	https://digital.csic.es/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2765	**10.25504/FAIRsharing.23bdba**	**DISNOR**	https://disnor.uniroma2.it/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2526	**10.25504/FAIRsharing.rt4gyp**	**Disordered Binding Sites Database**	http://dibs.enzim.ttk.mta.hu	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/143	**10.25504/FAIRsharing.mcgj5b**	**Distributed Sequence Annotation System**	http://www.biodas.org/documents/spec-1.6.html	standard		ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/2860	**10.25504/FAIRsharing.qgMKai**	**DNA Modification Database**	https://dnamod.hoffmanlab.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2865	**10.25504/FAIRsharing.VXoFLf**	**Domain-centric GO**	http://supfam.org/SUPERFAMILY/dcGO/index.html	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/144	**10.25504/FAIRsharing.4xrzw1**	**Dot Bracket Notation (DBN) - Vienna Format**	https://www.tbi.univie.ac.at/RNA/ViennaRNA/doc/html/rna_structure_notations.html#dot-bracket-notation	standard					All species	Model/formats or syntax:Format		
https://fairsharing.org/3015	**10.25504/FAIRsharing.04fcf5**	**EMDataResource**	http://www.emdataresource.org/	database		ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/1478	**10.25504/FAIRsharing.q9nh66**	**ENA Sequence Flat File Format**	https://ena-docs.readthedocs.io/en/latest/submit/fileprep/flat-file-example.html	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format	recommended	
https://fairsharing.org/1061	**10.25504/FAIRsharing.9ed958**	**ENA Sequence XML Schema**	https://ena-docs.readthedocs.io/en/latest/submit/general-guide/programmatic.html	standard		ready-to-use			All species	Model/formats or syntax:models/syntax	recommended	
https://fairsharing.org/2683	**10.25504/FAIRsharing.0VV4KB**	**Environmental Data Portal**	https://www.envidat.ch/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/511	**10.25504/FAIRsharing.xmmsmr**	**Epigenome Ontology**	https://github.com/EGO-ontology	standard		ready-to-use			All species	Semantic standards		github
https://fairsharing.org/4329	**10.25504/FAIRsharing.1a0bbf**	**ERIC/open**	https://opendata.eawag.ch/	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1484	**10.25504/FAIRsharing.9rhr9j**	**Eukaryotic Pathogen, Host & Vector Genomics Resource Ontology**	https://github.com/VEuPathDB-ontology/VEuPathDB-ontology	standard	maintained	ready-to-use			All species	Semantic standards		github
https://fairsharing.org/4715	**10.25504/FAIRsharing.c3521d**	**EvolclustDB**	http://evolclustdb.org/	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1757	**10.25504/FAIRsharing.4eanvm**	**Evolutionary Trace**	http://mammoth.bcm.tmc.edu/ETserver.html	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1581	**10.25504/FAIRsharing.2z0e8b**	**ExoCarta**	http://www.exocarta.org	database		ready-to-use			All species	Knowledgebase		

FAIRsharing URL	FAIRsharing DOI	Record Name	Record homepage URL	Resource Type	Manintenace	Ready-to-use	FAIRorg recommended	plant species accepted	Applied to species	Type	ELIXIR recommended	provider
https://fairsharing.org/2262	**10.25504/FAIRsharing.nnvcr9**	**FAIRDOMHub**	https://fairdomhub.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/3640	**10.25504/FAIRsharing.807b9d**	**FAIRtracks**	https://fairtracks.net	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax	recommended	
https://fairsharing.org/985	**10.25504/FAIRsharing.zc2tfa**	**Fanconi Anemia Ontology**	http://lab.rockefeller.edu/smogorzewska/ifar/	standard		ready-to-use			All species	Semantic standards		
https://fairsharing.org/114	**10.25504/FAIRsharing.r2ts5t**	**FASTQ Sequence and Sequence Quality Format**	http://news.open-bio.org/news/2009/12/nar-fastq-format/	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/478	**10.25504/FAIRsharing.haxp7g**	**Feature Annotation Location Description Ontology**	http://biohackathon.org/resource/faldo	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/2516	**10.25504/FAIRsharing.7zfq1a**	**FlavorDB**	http://cosylab.iiitd.edu.in/flavordb	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/3102	**10.25504/FAIRsharing.SfXDoZ**	**French Biodiversity (meta)data repository**	https://data.pndb.fr/	database	maintained	ready-to-use			All species	Data repository		
https://fairsharing.org/868	**10.25504/FAIRsharing.8xecy7**	**FuGEFlow**	http://flowcyt.sourceforge.net/fugeflow/	standard					All species	Minimum reporting guidelines		sourceforge(USA)
https://fairsharing.org/1270	**10.25504/FAIRsharing.namxfm**	**Functional Genomics Experiment Markup Language**	http://fuge.sourceforge.net/dev/index.php	standard					All species	Model/formats or syntax:models/syntax		sourceforge(USA)
https://fairsharing.org/1586	**10.25504/FAIRsharing.556qpw**	**FunTree: A Resource For Exploring The Functional Evolution Of Structurally Defined Enzyme Superfamilies**	http://www.funtree.info/FunTree/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2163	**10.25504/FAIRsharing.zv11j3**	**GBIF.org**	https://www.gbif.org	database	maintained	ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/1660	**10.25504/FAIRsharing.fyqk5z**	**Gene3D**	http://gene3d.biochem.ucl.ac.uk	database					All species	Knowledgebase		
https://fairsharing.org/414	**10.25504/FAIRsharing.z057ar**	**GenePattern GeneSet Table Format**	http://software.broadinstitute.org/cancer/software/genepattern/file-formats-guide	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/2136	**10.25504/FAIRsharing.qmygaa**	**GeneProf**	http://www.geneprof.org	database	maintained				All species	Knowledgebase		
https://fairsharing.org/1348	**10.25504/FAIRsharing.dnk0f6**	**Generic Feature Format Version 3**	https://github.com/The-Sequence-Ontology/Specifications/blob/master/gff3.md	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		github
https://fairsharing.org/1590	**10.25504/FAIRsharing.1ety1h**	**GeneSigDB: a manually curated database and resource for analysis of gene expression signatures**	http://www.genesigdb.org	database					All species	Knowledgebase		
https://fairsharing.org/1062	**10.25504/FAIRsharing.cjzx96**	**Genetic Testing Ontology**	http://qianzhu-lab.umbc.edu/	standard					All species	Semantic standards		
https://fairsharing.org/57	**10.25504/FAIRsharing.sbbbf6**	**Genome Biology Ontology Language**	http://gbol.life	standard	maintained				All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/379	**10.25504/FAIRsharing.pkw5bj**	**Genome Variation Format**	https://github.com/The-Sequence-Ontology/Specifications/blob/master/gvf.md	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		github
https://fairsharing.org/69	**10.25504/FAIRsharing.2hh7g7**	**Genomic Contextual Data Markup Language**	https://www.liebertpub.com/doi/10.1089/omi.2008.0A10	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/279	**10.25504/FAIRsharing.y1mmbv**	**Genomic Epidemiology Ontology**	https://genepio.org/	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/1328	**10.25504/FAIRsharing.vbejb6**	**Genomic Feature and Variation Ontology**	https://github.com/BioInterchange/Ontologies	standard		ready-to-use			All species	Semantic standards		github
https://fairsharing.org/3019	**10.25504/FAIRsharing.wOcTaM**	**Genomic Observatories Meta-Database**	https://geome-db.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/3626	**10.25504/FAIRsharing.3da56b**	**Genomics Cohorts Knowledge Ontology**	https://github.com/IHCC-cohorts/GECKO	standard	maintained	ready-to-use			All species	Semantic standards		github
https://fairsharing.org/688	**10.25504/FAIRsharing.kpbna7**	**Genotype Ontology**	https://github.com/monarch-initiative/GENO-ontology/	standard		ready-to-use			All species	Semantic standards		github
https://fairsharing.org/2125	**10.25504/FAIRsharing.rcbwsf**	**Giga Science Database**	http://gigadb.org/	database	maintained	ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/4445	**10.25504/FAIRsharing.84b518**	**Gigwa**	https://gigwa.southgreen.fr/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/787	**10.25504/FAIRsharing.eff3b2**	**Global Genome Biodiversity Network Data Standard**	https://wiki.ggbn.org/ggbn/GGBN_Data_Standard	standard		ready-to-use			All species	Minimum reporting guidelines		
https://fairsharing.org/632	**10.25504/FAIRsharing.9ry4cz**	**GoMapMan**	https://gomapman.nib.si/	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/2086	**10.25504/FAIRsharing.wfrsvq**	**gpDB**	http://bioinformatics.biol.uoa.gr/gpDB	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2235	**10.25504/FAIRsharing.fe9kyy**	**Green Non-Coding Database**	http://greenc.sequentiabiotech.com/wiki2/Main_Page	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1947	**10.25504/FAIRsharing.bpxgb6**	**Greengenes**	http://greengenes.lbl.gov	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1827	**10.25504/FAIRsharing.j98570**	**GreenPhylDB: A phylogenomic database for plant comparative genomics**	http://www.greenphyl.org	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1084	**10.25504/FAIRsharing.y49yj6**	**Herbarium information standards and protocols for interchange of data**	https://github.com/tdwg/prior-standards/tree/master/hispid3	standard		ready-to-use		All plant species		Minimum reporting guidelines		github
https://fairsharing.org/698	**10.25504/FAIRsharing.gaqfk4**	**HGeneCodonOntology**	http://blog.51.ca/u-345129/	standard					All species	Semantic standards		
https://fairsharing.org/472	**10.25504/FAIRsharing.n3rt95**	**HMMER Profile File Format**	http://www.hmmer.org	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/4454	**10.25504/FAIRsharing.a945b0**	**HostDB**	https://hostdb.org/hostdb/app	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1595	**10.25504/FAIRsharing.m4a6d3**	**HotRegion**	http://prism.ccbb.ku.edu.tr/hotregion	database					All species	Knowledgebase		
https://fairsharing.org/342	**10.25504/FAIRsharing.8drfwh**	**ICM binary file Format**	http://www.molsoft.com/icm/files.html	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/4293	**10.25504/FAIRsharing.bbd7df**	**IEEE 2791-2020**	https://standards.ieee.org/ieee/2791/7337/	standard	maintained	ready-to-use			All species	Minimum reporting guidelines		
https://fairsharing.org/2777	**10.25504/FAIRsharing.PeLjos**	**iGEM Registry of Standard Biological Parts**	http://parts.igem.org	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2303	**10.25504/FAIRsharing.6wf1zw**	**Image Data Resource**	https://idr.openmicroscopy.org	database	maintained	ready-to-use	recommended	All plant species		Knowledgebase		
https://fairsharing.org/1664	**10.25504/FAIRsharing.a4ybgg**	**Indel Flanking Region Database**	http://indel.bioinfo.sdu.edu.cn	database					All species	Knowledgebase		
https://fairsharing.org/1600	**10.25504/FAIRsharing.smqf7y**	**InterEvol database**	http://biodev.cea.fr/interevol/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2908	**N/A**	**Internal Control Genes**	http://icg.big.ac.cn/index.php/Main_Page	database					All species	Knowledgebase		
https://fairsharing.org/1501	**10.25504/FAIRsharing.yhLgTV**	**Investigation Study Assay JSON**	https://isa-specs.readthedocs.io/en/latest/isajson.html	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/1497	**10.25504/FAIRsharing.53gp75**	**Investigation Study Assay Tabular**	https://isa-specs.readthedocs.io/en/latest/isatab.html	standard	maintained	ready-to-use	recommended		All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/2750	**N/A**	**ionomicHUB**	https://www.ionomicshub.org/home/PiiMS	database				All plant species		Knowledgebase		
https://fairsharing.org/2492	**10.25504/FAIRsharing.ekdqe5**	**iReceptor Public Archive**	https://www.ireceptor.org/	database	maintained	ready-to-use			All species	Knowledgebase		

FAIRsharing URL	FAIRsharing DOI	Record Name	Record homepage URL	Resource Type	Manintenace	Ready-to-use	FAIRorg recommended	plant species accepted	Applied to species	Type	ELIXIR recommended	provider
https://fairsharing.org/950	**10.25504/FAIRsharing.dee5fb**	**IUPAC-IUB Commission on Biochemical Nomenclature - Abbreviations and Symbols for Nucleic Acids, Polynucleotides and their Constituents**	https://www.qmul.ac.uk/sbcs/iupac/misc/naabb.html	standard		ready-to-use			All species	Semantic standards		
https://fairsharing.org/959	**10.25504/FAIRsharing.d5a210**	**IUPAC-IUB Joint Commission on Biochemical Nomenclature - Nomenclature and Symbolism for Amino Acids and Peptides**	https://www.qmul.ac.uk/sbcs/iupac/misc/naabb.html	standard		ready-to-use			All species	Semantic standards		
https://fairsharing.org/1936	**10.25504/FAIRsharing.vd25jp**	**J-GLOBAL**	https://jglobal.jst.go.jp/en	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1661	**10.25504/FAIRsharing.5q1p14**	**Joint Genome Institute, Genomes OnLine Database**	https://gold.jgi.doe.gov/	database	maintained	ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/3618	**10.25504/FAIRsharing.826b4a**	**KnetMiner**	https://knetminer.com	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/4783	**10.25504/FAIRsharing.d30a2a**	**KNOTTIN database**	https://www.dsimb.inserm.fr/KNOTTIN/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2189	**10.25504/FAIRsharing.327nbg**	**Kyoto Encyclopedia of Genes and Genomes**	http://www.genome.jp/kegg/	database		ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/2325	**10.25504/FAIRsharing.dq46p7**	**Life Science Database Archive**	http://dbarchive.biosciencedbc.jp/index-e.html	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2043	**10.25504/FAIRsharing.2ma4gq**	**Ligand Expo**	http://ligand-depot.rutgers.edu/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2623	**N/A**	**Longhorn Array Database**	http://darwin.biochem.okstate.edu/lad/ilat/	database					All species	Knowledgebase		
https://fairsharing.org/3674	**N/A**	**LUCApedia**	http://eebgroups.princeton.edu/lucapedia/index.html	database					All species	Knowledgebase		
https://fairsharing.org/2469	**10.25504/FAIRsharing.gdj215**	**Madagascar Biodiversity Information Facility IPT - GBIF France**	http://ipt.madbif.mg/	database	maintained				All species	Knowledgebase		
https://fairsharing.org/430	**10.25504/FAIRsharing.53edcc**	**MAP**	https://www.cog-genomics.org/plink2/formats#map	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/2419	**10.25504/FAIRsharing.j1eyq2**	**mentha**	http://mentha.uniroma2.it/	database	maintained				All species	Knowledgebase		
https://fairsharing.org/115	**10.25504/FAIRsharing.wjwb5c**	**CHADO XML**	http://gmod.org/wiki/Chado_XML	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		GMOD
https://fairsharing.org/1347	**10.25504/FAIRsharing.bfqdc8**	**Generic Feature Format Version 2**	http://gmod.org/wiki/GFF2	standard					All species	Model/formats or syntax:Format		GMOD
https://fairsharing.org/2051	**10.25504/FAIRsharing.z1czxj**	**PeroxiBase**	http://peroxibase.toulouse.inra.fr/	database		ready-to-use			All species	Knowledgebase		INRA(FRA)
https://fairsharing.org/2608	**10.25504/FAIRsharing.178BmT**	**Portail Data INRAE**	https://data.inrae.fr/	database	maintained			All plant species		Data repository		INRA(FRA)
https://fairsharing.org/1499	**10.25504/FAIRsharing.ld9wne**	**Open Data for Access and Mining Structural Metadata**	https://inrae.github.io/ODAM/data-preparation/	standard	maintained	ready-to-use			All species	Minimum reporting guidelines		INRA(FRA)
https://fairsharing.org/2263	**10.25504/FAIRsharing.dw22y3**	**Genetic and Genomic Information System**	https://urgi.versailles.inra.fr/gnpis	database	maintained	ready-to-use		All plant species		Data repository		INRA(FRA)
https://fairsharing.org/1427	**10.25504/FAIRsharing.4c7c1c**	**DDBJ/ENA/GenBank Feature Table**	http://www.insdc.org/documents/feature-table	standard		ready-to-use			All species	Model/formats or syntax:models/syntax	recommended	INSDC
https://fairsharing.org/438	**10.25504/FAIRsharing.tn873z**	**INSD sequence record XML**	http://www.insdc.org/documents/xml-status	standard		ready-to-use			All species	Model/formats or syntax:models/syntax		INSDC
https://fairsharing.org/4211	**10.25504/FAIRsharing.209c8e**	**Metadata Schema for the Persistent Identification of Instruments (1.0)**	https://doi.org/10.15497/RDA00070	standard		ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/768	**10.25504/FAIRsharing.xfje21**	**Metagenome Sample Vocabulary**	http://mdb.bio.titech.ac.jp/msv	standard					All species	Semantic standards		
https://fairsharing.org/1682	**10.25504/FAIRsharing.htzwqf**	**PASS2**	http://caps.ncbs.res.in/pass2/	database		ready-to-use			All species	Knowledgebase		Lab 25 NCBS (INDIA)
https://fairsharing.org/354	**10.25504/FAIRsharing.rvvbCB**	**Meta-omics Data and Collection Objects**	https://www.mod-co.net/wiki/Schema_Representations	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/3569	**10.25504/FAIRsharing.fa85ba**	**MetaPhOrs**	http://orthology.phylomedb.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/849	**10.25504/FAIRsharing.k893xa**	**MHC Restriction Ontology**	https://github.com/IEDB/MRO	standard	maintained	ready-to-use			All species	Semantic standards		github
https://fairsharing.org/1193	**10.25504/FAIRsharing.qs4x5m**	**Microarray and Gene Expression Data Ontology**	http://mged.sourceforge.net/ontologies/MGEDontology.php	standard	maintained		recommended		All species	Semantic standards		sourceforge(USA)
https://fairsharing.org/1962	**10.25504/FAIRsharing.cwn40y**	**Sequence-Structural Templates of Single-member Superfamilies**	http://caps.ncbs.res.in/SSTOSS/index.htm	database		ready-to-use			All species	Knowledgebase		Lab 25 NCBS (INDIA)
https://fairsharing.org/620	**10.25504/FAIRsharing.rz4vfg**	**FASTA Sequence Format**	http://blast.ncbi.nlm.nih.gov/blastcgihelp.shtml	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		NCBI (USA)
https://fairsharing.org/1973	**10.25504/FAIRsharing.wk5azf**	**Database of Sequence Tagged Sites**	http://www.ncbi.nlm.nih.gov/dbSTS/index.html	database					All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/1222	**10.25504/FAIRsharing.x964fb**	**MicroArray Gene Expression Markup Language**	http://www.mged.org/Workgroups/MAGE/mage-ml.html	standard					All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/1254	**10.25504/FAIRsharing.ak8p5g**	**MicroArray Gene Expression Tabular Format**	https://www.fged.org/projects/mage-tab	standard		ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/4456	**10.25504/FAIRsharing.a73460**	**MicrobiomeDB**	https://microbiomedb.org/mbio/app/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1606	**10.25504/FAIRsharing.bv0zjz**	**Mimotope Database**	http://immunet.cn/mimodb	database					All species	Knowledgebase		
https://fairsharing.org/1607	**10.25504/FAIRsharing.wqtfkv**	**MINAS - A Database of Metal Ions in Nucleic AcidS**	http://www.minas.uzh.ch	database					All species	Knowledgebase		
https://fairsharing.org/1986	**10.25504/FAIRsharing.b5ann2**	**Plant Genome Central**	http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html	database				All plant species		Data repository		NCBI (USA)
https://fairsharing.org/217	**10.25504/FAIRsharing.gaegy8**	**MIAME Notation in Markup Language**	http://www.ncbi.nlm.nih.gov/geo/info/MINiML.html	standard	maintained	ready-to-use	recommended		All species	Model/formats or syntax:models/syntax		NCBI (USA)
https://fairsharing.org/1206	**10.25504/FAIRsharing.a55z32**	**Minimal Information about a high throughput SEQuencing Experiment**	https://doi.org/10.5281/zenodo.5706412	standard	maintained	ready-to-use	recommended		All species	Minimum reporting guidelines		
https://fairsharing.org/94	**10.25504/FAIRsharing.ca48xs**	**Minimal Information About a Phylogenetic Analysis**	http://www.evoio.org/wiki/MIAPA	standard		ready-to-use			All species	Minimum reporting guidelines		
https://fairsharing.org/599	**10.25504/FAIRsharing.vy4h4j**	**Minimal Information for QTLs and Association Studies**	http://miqas.sourceforge.net/	standard		ready-to-use			All species	Minimum reporting guidelines		sourceforge(USA)
https://fairsharing.org/375	**10.25504/FAIRsharing.wgvr02**	**Minimal Information for QTLs and Association Studies Tabular**	http://miqas.sourceforge.net/specification/MIQAS_TAB/MIQAS_TAB_specification.html	standard		ready-to-use			All species	Minimum reporting guidelines		sourceforge(USA)

FAIRsharing URL	FAIRsharing DOI	Record Name	Record homepage URL	Resource Type	Manintenace	Ready-to-use	FAIRorg recommended	plant species accepted	Applied to species	Type	ELIXIR recommended	provider
https://fairsharing.org/881	**10.25504/FAIRsharing.ckg5a2**	**Minimal Metagenome Sequence Analysis Standard**	http://mibbi.sf.net/projects/MINIMESS.shtml	standard					All species	Minimum reporting guidelines		
https://fairsharing.org/250	**10.25504/FAIRsharing.9mfexc**	**Minimum Information About a Genotyping Experiment**	http://migen.sourceforge.net/	standard		ready-to-use			All species	Minimum reporting guidelines		sourceforge(USA)
https://fairsharing.org/96	**10.25504/FAIRsharing.32b10v**	**Minimum Information About a Microarray Experiment**	http://www.fged.org/projects/miame/	standard		ready-to-use	recommended		All species	Minimum reporting guidelines		
https://fairsharing.org/36	**10.25504/FAIRsharing.3gxr9b**	**Simple Omnibus Format in Text**	http://www.ncbi.nlm.nih.gov/geo/info/soft2.html	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		NCBI (USA)
https://fairsharing.org/1980	**10.25504/FAIRsharing.sgqf4n**	**MapViewer**	http://www.ncbi.nlm.nih.gov/mapview/	database					All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/1971	**10.25504/FAIRsharing.64bk6p**	**NCBI Probe Database**	http://www.ncbi.nlm.nih.gov/probe	database			recommended		All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/1991	**10.25504/FAIRsharing.4jg0qw**	**Reference Sequence Database**	http://www.ncbi.nlm.nih.gov/RefSeq/	database		ready-to-use			All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/1978	**10.25504/FAIRsharing.g7t2hv**	**Sequence Read Archive**	http://www.ncbi.nlm.nih.gov/sra	database		ready-to-use	recommended		All species	Data repository		NCBI (USA)
https://fairsharing.org/1974	**10.25504/FAIRsharing.6brrxe**	**Genetic Codes**	http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi	database		ready-to-use			All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/2164	**10.25504/FAIRsharing.abwvhp**	**NCBI Trace Archive**	http://www.ncbi.nlm.nih.gov/Traces/home/	database		ready-to-use	recommended		All species	Data repository		NCBI (USA)
https://fairsharing.org/877	**10.25504/FAIRsharing.nn7bf2**	**Standard Flowgram Format**	http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?cmd=show&f=formats&m=doc&s=format#sff	standard		ready-to-use			All species	Model/formats or syntax:Format		NCBI (USA)
https://fairsharing.org/1993	**10.25504/FAIRsharing.ge1c3p**	**UniGene gene-oriented nucleotide sequence clusters**	http://www.ncbi.nlm.nih.gov/unigene	database					All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/1994	**10.25504/FAIRsharing.81dw2c**	**NCBI UniSTS**	http://www.ncbi.nlm.nih.gov/unists	database					All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/1995	**10.25504/FAIRsharing.p4dtt2**	**UniVec**	http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html	database		ready-to-use			All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/664	**10.25504/FAIRsharing.mg1mdc**	**A Gold Path format**	https://www.ncbi.nlm.nih.gov/assembly/agp/AGP_Specification/	standard		ready-to-use			All species	Model/formats or syntax:Format		NCBI (USA)
https://fairsharing.org/2210	**10.25504/FAIRsharing.aqhv1y**	**NCBI BioProject**	https://www.ncbi.nlm.nih.gov/bioproject/	database		ready-to-use	recommended		All species	Data repository		NCBI (USA)
https://fairsharing.org/2211	**10.25504/FAIRsharing.qr6pqk**	**NCBI BioSample**	https://www.ncbi.nlm.nih.gov/biosample	database	maintained	ready-to-use	recommended		All species	Data repository		NCBI (USA)
https://fairsharing.org/95	**10.25504/FAIRsharing.pvdrmz**	**Minimum Information About a Microarray Experiment involving Plants**	http://miame-plant.sourceforge.net/	standard		ready-to-use		All plant species		Minimum reporting guidelines		sourceforge(USA)
https://fairsharing.org/1215	**10.25504/FAIRsharing.mad142**	**Minimum Information about a Nutrigenomics experiment**	http://mibbi.sourceforge.net/projects/MIAME-Nutr.shtml	standard					All species	Minimum reporting guidelines		sourceforge(USA)
https://fairsharing.org/180	**10.25504/FAIRsharing.my19zk**	**Minimum Information About a RNAi Experiment**	http://miare.sourceforge.net	standard		ready-to-use			All species	Minimum reporting guidelines		sourceforge(USA)
https://fairsharing.org/1258	**10.25504/FAIRsharing.zrmjr7**	**Minimum Information about an array-based toxicogenomics experiment**	http://mibbi.sourceforge.net/projects/MIAME-Tox.shtml	standard					All species	Minimum reporting guidelines		sourceforge(USA)
https://fairsharing.org/560	**10.25504/FAIRsharing.x6dcc7**	**Minimum Information about an ENVironmental transcriptomic experiment**	http://nebc.nox.ac.uk/miame/miame_env.html	standard					All species	Minimum reporting guidelines		
https://fairsharing.org/106	**10.25504/FAIRsharing.t8g7yc**	**Minimum Information About an RNAi Experiment Tabular**	http://miare.sourceforge.net/HomePage	standard		ready-to-use			All species	Minimum reporting guidelines		sourceforge(USA)
https://fairsharing.org/262	**10.25504/FAIRsharing.9aa0zp**	**Minimum Information about any (x) Sequence**	https://www.gensc.org/pages/standards-intro.html	standard	maintained	ready-to-use	recommended		All species	Minimum reporting guidelines		
https://fairsharing.org/3922	**10.25504/FAIRsharing.c2e4ab**	**Minimum Information about Tissue Imaging**	https://www.miti-consortium.org/	standard		ready-to-use			All species	Minimum reporting guidelines		
https://fairsharing.org/865	**10.25504/FAIRsharing.mxz4jy**	**Minimum Information for Publication of Quantitative Real-Time PCR Experiments**	https://rdml.org/miqe.html	standard		ready-to-use	recommended		All species	Minimum reporting guidelines		
https://fairsharing.org/45	**10.25504/FAIRsharing.jw7rq3**	**Minimum Information for Reporting Next Generation Sequencing Genotyping**	http://miring.immunogenomics.org	standard	maintained	ready-to-use			All species	Minimum reporting guidelines		
https://fairsharing.org/4740	**10.25504/FAIRsharing.9a2f48**	**MIRAGE reporting guidelines for lectin microarray analysis**	https://www.beilstein-institut.de/en/projects/mirage/guidelines/#lectin_microarrays	standard	maintained	ready-to-use			All species	Minimum reporting guidelines		
https://fairsharing.org/2100	**10.25504/FAIRsharing.hmgte8**	**miRBase**	http://www.mirbase.org/	database		ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/1345	**N/A**	**mirGFF3**	https://github.com/miRTop/mirGFF3	standard	maintained				All species	Model/formats or syntax:Format		github
https://fairsharing.org/1667	**10.25504/FAIRsharing.5pfx4r**	**miRNEST**	http://mirnest.amu.edu.pl	database					All species	Knowledgebase		
https://fairsharing.org/3195	**10.25504/FAIRsharing.Ql6K87**	**ModelSEED Biochemistry Database**	https://modelseed.org/biochem	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1964	**10.25504/FAIRsharing.w2eeqr**	**NCBI BioSystems Database**	https://www.ncbi.nlm.nih.gov/biosystems/	database					All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/2205	**10.25504/FAIRsharing.djyk2c**	**MoonProt**	http://www.moonlightingproteins.org	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/299	**10.25504/FAIRsharing.1dd23d**	**Multiple Alignment using Fast Fourier Transform**	https://mafft.cbrc.jp/alignment/software/	standard		ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/2520	**10.25504/FAIRsharing.bwswdf**	**Mutual Folding Induced by Binding Database**	http://mfib.enzim.ttk.mta.hu	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/548	**10.25504/FAIRsharing.95n1ns**	**National Agricultural Library Thesaurus**	https://agclass.nal.usda.gov/	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/2658	**10.25504/FAIRsharing.EivZJw**	**National Omics Data Encyclopedia**	https://www.biosino.org/node	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/490	**10.25504/FAIRsharing.tej5wx**	**New Hampshire eXtended Format**	http://www.phylosoft.org/NHX/nhx.pdf	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/4	**10.25504/FAIRsharing.ezznsh**	**Newick tree Format**	http://evolution.genetics.washington.edu/phylip/newicktree.html	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/297	**10.25504/FAIRsharing.knrb06**	**Nexus XML**	https://github.com/nexml	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		github
https://fairsharing.org/436	**10.25504/FAIRsharing.rk682s**	**BioProject XML Schema**	https://www.ncbi.nlm.nih.gov/data_specs/schema/other/bioproject/	standard		ready-to-use			All species	Model/formats or syntax:models/syntax		NCBI (USA)
https://fairsharing.org/1547	**10.25504/FAIRsharing.9kahy4**	**GenBank Nucleotide Sequence Database**	https://www.ncbi.nlm.nih.gov/genbank/	database		ready-to-use	recommended		All species	Data repository		NCBI (USA)
https://fairsharing.org/1969	**10.25504/FAIRsharing.v9fya8**	**Expressed Sequence Tags database**	https://www.ncbi.nlm.nih.gov/genbank/dbest/	database		ready-to-use	recommended		All species	Data repository		NCBI (USA)
https://fairsharing.org/1983	**10.25504/FAIRsharing.5h3maw**	**NCBI Gene**	https://www.ncbi.nlm.nih.gov/gene	database		ready-to-use	recommended		All species	Knowledgebase		NCBI (USA)

FAIRsharing URL	FAIRsharing DOI	Record Name	Record homepage URL	Resource Type	Manintenace	Ready-to-use	FAIRorg recommended	plant species accepted	Applied to species	Type	ELIXIR recommended	provider
https://fairsharing.org/3653	**10.25504/FAIRsharing.7ad252**	**NCBI Genome**	https://www.ncbi.nlm.nih.gov/genome	database		ready-to-use			All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/3337	**10.25504/FAIRsharing.b05f99**	**NCBI Genome Data Viewer**	https://www.ncbi.nlm.nih.gov/genome/gdv/	database		ready-to-use			All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/1975	**10.25504/FAIRsharing.5hc8vt**	**Gene Expression Omnibus**	https://www.ncbi.nlm.nih.gov/geo/	database	maintained	ready-to-use	recommended		All species	Data repository		NCBI (USA)
https://fairsharing.org/1977	**10.25504/FAIRsharing.mzc066**	**NCBI HomoloGene Database**	https://www.ncbi.nlm.nih.gov/homologene	database		ready-to-use	recommended		All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/997	**10.25504/FAIRsharing.kx4bgf**	**NGS ontology**	https://github.com/mickaelsilva/NGSOnto	standard					All species	Semantic standards		github
https://fairsharing.org/833	**10.25504/FAIRsharing.rg2vmt**	**GenBank Sequence Format**	https://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html	standard		ready-to-use	recommended		All species	Model/formats or syntax:Format		NCBI (USA)
https://fairsharing.org/1981	**10.25504/FAIRsharing.zqzvyc**	**NCBI Structure**	https://www.ncbi.nlm.nih.gov/structure	database	maintained	ready-to-use			All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/1967	**10.25504/FAIRsharing.b9st5p**	**Conserved Domain Database**	https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml	database	maintained	ready-to-use			All species	Knowledgebase		NCBI (USA)
https://fairsharing.org/731	**10.25504/FAIRsharing.xk98ez**	**NimbleGen Gene Description**	https://dmac.cbs.dtu.dk/service/data-formats	standard					All species	Model/formats or syntax:Format		
https://fairsharing.org/1620	**10.25504/FAIRsharing.hmb1f4**	**NONCODE**	http://www.noncode.org	database					All species	Knowledgebase		
https://fairsharing.org/1670	**10.25504/FAIRsharing.hgrpah**	**NRG-CING**	http://nmr.cmbi.ru.nl/NRG-CING	database					All species	Knowledgebase		
https://fairsharing.org/1621	**10.25504/FAIRsharing.8ecbxx**	**NucleaRDB**	http://www.receptors.org/nucleardb	database					All species	Knowledgebase		
https://fairsharing.org/2046	**10.25504/FAIRsharing.bh0k78**	**Nucleic Acid Database**	http://ndbserver.rutgers.edu/	database			recommended		All species	Knowledgebase		
https://fairsharing.org/1220	**10.25504/FAIRsharing.fj07xj**	**NCBI Taxonomy**	https://www.ncbi.nlm.nih.gov/taxonomy	standard	maintained	ready-to-use	recommended		All species	Semantic standards		NCBI (USA)
https://fairsharing.org/395	**10.25504/FAIRsharing.cfhbvn**	**Observ-Tab**	http://wiki.gcc.rug.nl/wiki/ObservStart	standard	maintained				All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/2519	**10.25504/FAIRsharing.re1278**	**Omics Discovery Index**	http://www.omicsdi.org/home	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2704	**10.25504/FAIRsharing.58lHsV**	**OmicsDB**	http://www.omicsdb.org/	database	maintained				All species	Knowledgebase		
https://fairsharing.org/282	**10.25504/FAIRsharing.2hqa97**	**OmicsDI XML format**	https://github.com/OmicsDI/specifications/blob/master/docs/schema/OmicsDISchema.xsd	standard		ready-to-use			All species	Model/formats or syntax:Format		github
https://fairsharing.org/1622	**10.25504/FAIRsharing.hsy066**	**Online GEne Essentiality database**	https://v3.ogee.info/#/home	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2568	**10.25504/FAIRsharing.9btkvp**	**Online Resource for Community Annotation of Eukaryotes**	http://bioinformatics.psb.ugent.be/orcae/	database	maintained	ready-to-use		All plant species		Knowledgebase		UGENT(Beligum)
https://fairsharing.org/1488	**10.25504/FAIRsharing.284e1z**	**Ontology for Biomedical Investigations**	http://obi-ontology.org/	standard	maintained	ready-to-use	recommended		All species	Semantic standards		
https://fairsharing.org/1250	**N/A**	**Ontology for Experimental Phenotypic Objects**	https://github.com/OpenSILEX/ontology-phis-oepo-field	standard					All species	Semantic standards		github
https://fairsharing.org/363	**10.25504/FAIRsharing.zmx7nn**	**Ontology of Genes and Genomes**	https://bitbucket.org/hegroup/ogg	standard	maintained	ready-to-use			All species	Semantic standards		bitbucket
https://fairsharing.org/3937	**N/A**	**Open Digital Specimen**	https://github.com/DiSSCo/openDS	standard	maintained				All species	Model/formats or syntax:models/syntax		github
https://fairsharing.org/1790	**10.25504/FAIRsharing.8dak0r**	**Open Regulatory Annotation**	http://www.oreganno.org/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2393	**10.25504/FAIRsharing.g3j5qj**	**Orthologous MAtrix**	http://omabrowser.org/oma/about/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/77	**10.25504/FAIRsharing.4877h0**	**Orthology Ontology**	https://github.com/qfo/OrthologyOntology	standard		ready-to-use			All species	Semantic standards		github
https://fairsharing.org/3692	**10.25504/FAIRsharing.45bf5b**	**Pacific Northwest National Laboratory DataHub: Scientific Data Repository**	https://data.pnnl.gov/	database	maintained	ready-to-use	recommended		All species	Data repository		
https://fairsharing.org/3306	**10.25504/FAIRsharing.9289d4**	**Paired Omics Data Platform**	https://pairedomicsdata.bioinformatics.nl/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/813	**10.25504/FAIRsharing.RvBRMp**	**Pairs file format**	https://github.com/4dn-dcic/pairix/blob/master/pairs_format_specification.md	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		github
https://fairsharing.org/2987	**10.25504/FAIRsharing.oaLCAI**	**PathosSystems Resource Integration Center Repository**	https://www.patricbrc.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2032	**10.25504/FAIRsharing.33yggg**	**PAZAR**	http://www.pazar.info/	database					All species	Knowledgebase		
https://fairsharing.org/1014	**10.25504/FAIRsharing.zsnv69**	**Multiple alignment**	http://bioportal.bioontology.org/ontologies/1026	standard					All species	Semantic standards		NCBO(USA)
https://fairsharing.org/3301	**10.25504/FAIRsharing.d02054**	**PDB-REDO**	https://pdb-redo.eu	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1839	**10.25504/FAIRsharing.k4sp7m**	**PDB-REPRDB**	http://mbs.cbrc.jp/pdbreprdb-cgi/reprdb_menu.pl	database					All species	Knowledgebase		
https://fairsharing.org/1925	**10.25504/FAIRsharing.rxe7z2**	**PDZBase**	http://abc.med.cornell.edu/pdzbase	database					All species	Data repository		Cornell University
https://fairsharing.org/2648	**N/A**	**PDZscape**	http://www.actrec.gov.in:8080/pdzscape/	database					All species	Knowledgebase		
https://fairsharing.org/2613	**10.25504/FAIRsharing.654822**	**PGDBj Ortholog Database**	http://pgdbj.jp/pages/index.html?dir=&page=od&ln=en	database		ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/398	**10.25504/FAIRsharing.kfs4pj**	**Pseudogene**	http://bioportal.bioontology.org/ontologies/1135	standard		ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/930	**10.25504/FAIRsharing.23th83**	**GeoSpecies Ontology**	http://bioportal.bioontology.org/ontologies/1247	standard		ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/635	**10.25504/FAIRsharing.ngn3a6**	**Genome Component Ontology**	http://bioportal.bioontology.org/ontologies/GCO	standard		ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/48	**10.25504/FAIRsharing.efv7gw**	**HOMology**	http://bioportal.bioontology.org/ontologies/HOM/?p=summary	standard		ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/899	**10.25504/FAIRsharing.b7evv9**	**MeGO**	http://bioportal.bioontology.org/ontologies/MEGO/?p=summary	standard		ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/535	**10.25504/FAIRsharing.vppyga**	**Non-coding RNA Ontology**	http://bioportal.bioontology.org/ontologies/NCRO	standard		ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/550	**10.25504/FAIRsharing.vdenk6**	**Phylogenetic Ontology**	http://bioportal.bioontology.org/ontologies/PHYLONT	standard		ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/703	**10.25504/FAIRsharing.fbn4sh**	**Reference Sequence Annotation**	http://bioportal.bioontology.org/ontologies/RSA	standard		ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/280	**10.25504/FAIRsharing.zbff4z**	**Gene Regulation Ontology**	http://purl.bioontology.org/ontology/GRO	standard					All species	Semantic standards		NCBO(USA)
https://fairsharing.org/903	**10.25504/FAIRsharing.vh9jbb**	**Ontology for Genetic Interval**	http://purl.bioontology.org/ontology/OGI	standard	maintained	ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/88	**10.25504/FAIRsharing.ngv2xx**	**Apollo-SV**	http://purl.obolibrary.org/obo/apollo_sv.owl	standard	maintained	ready-to-use			All species	Semantic standards		NCBO(USA)

FAIRsharing URL	FAIRsharing DOI	Record Name	Record homepage URL	Resource Type	Manintenace	Ready-to-use	FAIRorg recommended	plant species accepted	Applied to species	Type	ELIXIR recommended	provider
https://fairsharing.org/1303	**10.25504/FAIRsharing.bnx9cf**	**Molecular Connections Cell Line Ontology**	https://bioportal.bioontology.org/ontologies/MCCL	standard					All species	Semantic standards		NCBO(USA)
https://fairsharing.org/263	**10.25504/FAIRsharing.4qyf0f**	**Regulation of Transcription Ontolology**	https://bioportal.bioontology.org/ontologies/RETO	standard	maintained	ready-to-use			All species	Semantic standards		NCBO(USA)
https://fairsharing.org/2415	**10.25504/FAIRsharing.tdhkc6**	**Genome Sequence Archive**	https://bigd.big.ac.cn/gsa/	database	maintained	ready-to-use	recommended		All species	Data repository		NGDC (China)
https://fairsharing.org/2903	**10.25504/FAIRsharing.91c279**	**Genome Variation Map**	https://bigd.big.ac.cn/gvm/	database		ready-to-use		All plant species		Data repository		NGDC (China)
https://fairsharing.org/2905	**10.25504/FAIRsharing.7qBaJ0**	**Genome Warehouse**	https://bigd.big.ac.cn/gwh/	database	maintained	ready-to-use			All species	Data repository		NGDC (China)
https://fairsharing.org/2910	**10.25504/FAIRsharing.86ec7a**	**Leaf Senescence Database**	https://bigd.big.ac.cn/lsd/	database		ready-to-use		All plant species		Knowledgebase		NGDC (China)
https://fairsharing.org/2913	**10.25504/FAIRsharing.4ef690**	**Plant Editosome Database**	https://bigd.big.ac.cn/ped	database		ready-to-use		All plant species		Knowledgebase		NGDC (China)
https://fairsharing.org/2763	**10.25504/FAIRsharing.52da68**	**National Genomics Data Center Repository**	https://ngdc.cncb.ac.cn/	database		ready-to-use			All species	Data repository		NGDC (China)
https://fairsharing.org/2911	**10.25504/FAIRsharing.d1a9bd**	**MethBank**	https://ngdc.cncb.ac.cn/methbank/	database		ready-to-use			All species	Data repository		NGDC (China)
https://fairsharing.org/4333	**10.25504/FAIRsharing.45990b**	**Open Archive for Miscellaneous Data**	https://ngdc.cncb.ac.cn/omix/	database	maintained	ready-to-use			All species	Knowledgebase		NGDC (China)
https://fairsharing.org/1417	**10.25504/FAIRsharing.6xq0ee**	**Gene Ontology**	http://www.geneontology.org	standard	maintained	ready-to-use	recommended		All species	Semantic standards		NHGRI(USA)
https://fairsharing.org/554	**10.25504/FAIRsharing.7y955w**	**Gene Ontology Extension**	http://www.geneontology.org	standard		ready-to-use			All species	Semantic standards		NHGRI(USA)
https://fairsharing.org/1072	**10.25504/FAIRsharing.9gx9at**	**Gene Ontology Annotation File Format 1.0**	https://github.com/geneontology/geneontology.github.io/blob/e6e554c42cf75312784ecdbca44f858633990e99/_docs/go-annotation-file-gaf-format-10.md	standard	maintained				All species	Model/formats or syntax:Format		NHGRI(USA)
https://fairsharing.org/2048	**10.25504/FAIRsharing.xshwbf**	**Sanger Pfam Mirror**	http://pfam.sanger.ac.uk/	database	maintained				All species	Knowledgebase		Sanger institute (UK)
https://fairsharing.org/953	**10.25504/FAIRsharing.qaaj0q**	**Gene Feature File version 1**	http://www.sanger.ac.uk/resources/software/gff/spec.html	standard					All species	Model/formats or syntax:Format		Sanger institute (UK)
https://fairsharing.org/427	**10.25504/FAIRsharing.820ef3**	**GenePix Results Format**	https://support.moleculardevices.com/s/article/GenePix-File-Formats#gpr	standard		ready-to-use			All species	Model/formats or syntax:Format		SIB(Sswiss)
https://fairsharing.org/311	**10.25504/FAIRsharing.rbpqg**	**PGxO**	https://pgxo.loria.fr	standard		ready-to-use			All species	Semantic standards		
https://fairsharing.org/1273	**10.25504/FAIRsharing.9xsgcr**	**Phenotype eXchange Format**	https://github.com/phenopackets/phenopacket-format	standard		ready-to-use	recommended		All species	Model/formats or syntax:Format		github
https://fairsharing.org/1760	**10.25504/FAIRsharing.73cqdk**	**PHI-base**	http://www.phi-base.org/	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/3297	**10.25504/FAIRsharing.625fde**	**Phylogenes**	http://www.phylogenes.org/	database		ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1838	**10.25504/FAIRsharing.7hxxc4**	**PhylomeDB**	http://phylomedb.org	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2727	**10.25504/FAIRsharing.w7Yuyx**	**PhytoMine**	https://phytozome.jgi.doe.gov/phytomine/	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/2734	**10.25504/FAIRsharing.83d06b**	**Phytozome**	https://phytozome.jgi.doe.gov/pz/portal.html	database		ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1288	**10.25504/FAIRsharing.tghhc4**	**Pipeline Patterns Ontology**	https://bitbucket.org/PlantExpAssay/ontology	standard		ready-to-use			All species	Semantic standards		bitbucket
https://fairsharing.org/2035	**10.25504/FAIRsharing.vssch2**	**PIR SuperFamily**	https://proteininformationresource.org/pirwww/dbinfo/pirsf.shtml	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1025	**10.25504/FAIRsharing.nft558**	**Plant Diversity Ontology**	http://gooa.las.ac.cn/external/index.jsp	standard				All plant species		Semantic standards		
https://fairsharing.org/2141	**10.25504/FAIRsharing.7qexb2**	**Plant DNA C-values database**	http://data.kew.org/cvalues/	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1314	**10.25504/FAIRsharing.fgmzk8**	**Plant Experimental Assay Ontology**	https://bitbucket.org/PlantExpAssay/ontology	standard		ready-to-use		All plant species		Semantic standards		bitbucket
https://fairsharing.org/1184	**10.25504/FAIRsharing.6yNXYK**	**Plant Experimental Condition Ontology**	https://github.com/Planteome/plant-experimental-conditions-ontology	standard	maintained	ready-to-use		All plant species		Semantic standards		github
https://fairsharing.org/2867	**10.25504/FAIRsharing.BnLUyq**	**Plant Genome Integrative Explorer**	https://plantgenie.org	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1840	**10.25504/FAIRsharing.391trx**	**Plant Genome Network**	http://pgn.cornell.edu/	database				All plant species		Data repository		Cornell University
https://fairsharing.org/2195	**10.25504/FAIRsharing.rf3m4g**	**Plant Genomics and Phenomics Research Data Repository**	http://edal-pgp.ipk-gatersleben.de	database	maintained	ready-to-use		All plant species		Data repository		
https://fairsharing.org/1629	**10.25504/FAIRsharing.rn1pxg**	**Plant Natural Antisense Transcripts Database**	http://bis.zju.edu.cn/pnatdb/	database				All plant species		Knowledgebase		
https://fairsharing.org/237	**10.25504/FAIRsharing.3ngg40**	**Plant Ontology**	http://plantontology.org/	standard	maintained	ready-to-use		All plant species		Semantic standards		
https://fairsharing.org/2133	**10.25504/FAIRsharing.7k8zh0**	**Plant Promoter Database**	http://ppdb.agr.gifu-u.ac.jp	database		ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1721	**10.25504/FAIRsharing.ysp7ke**	**Plant Resistance Gene Database**	http://prgdb.org/prgdb4/	database		ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1807	**10.25504/FAIRsharing.ex3fqk**	**Plant Transcription Factor Database**	http://planttfdb.cbi.pku.edu.cn	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/2528	**10.25504/FAIRsharing.k0kcjw**	**Planteome**	http://www.planteome.org	database	maintained	ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/1713	**10.25504/FAIRsharing.fy2ebk**	**PlantRNA**	http://plantrna.ibmp.cnrs.fr/	database		ready-to-use		All plant species		Knowledgebase		
https://fairsharing.org/2766	**10.25504/FAIRsharing.wBOua0**	**PLAZA**	https://bioinformatics.psb.ugent.be/plaza/	database	maintained	ready-to-use		All plant species		Knowledgebase	recommended	UGENT(Beligum)
https://fairsharing.org/3814	**10.25504/FAIRsharing.23ac74**	**PlutoF**	https://plutof.ut.ee/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1671	**10.25504/FAIRsharing.etqbej**	**Pocket Similarity Search using Multiple-Sketches**	http://possum.cbrc.jp/PoSSuM/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1631	**10.25504/FAIRsharing.tc6df8**	**Pocketome: an encyclopedia of small-molecule binding sites in 4D**	http://pocketome.org	database					All species	Knowledgebase		
https://fairsharing.org/1632	**10.25504/FAIRsharing.s9ztmd**	**Polbase**	https://polbase.neb.com/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1053	**10.25504/FAIRsharing.vq28qp**	**Population and Community Ontology**	https://github.com/PopulationAndCommunityOntology/pco	standard	maintained	ready-to-use			All species	Semantic standards		github
https://fairsharing.org/1684	**10.25504/FAIRsharing.znvzhj**	**Predictive Networks**	http://predictivenetworks.org/	database					All species	Knowledgebase		
https://fairsharing.org/2790	**10.25504/FAIRsharing.22e3be**	**Princeton University MicroArray database**	http://puma.princeton.edu	database		ready-to-use			All species	Data repository		
https://fairsharing.org/2099	**10.25504/FAIRsharing.h8r843**	**PRINTS**	http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/PRINTS.html	database		ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/2232	**10.25504/FAIRsharing.t6wjn7**	**probeBase**	http://www.probebase.net	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1143	**10.25504/FAIRsharing.v033mj**	**proBed**	http://www.psidev.info/probed	standard	maintained				All species	Model/formats or syntax:Format		
https://fairsharing.org/1532	**10.25504/FAIRsharing.zve9cc**	**Proteasix Ontology**	https://dx.doi.org/10.1186%2Fs13326-016-0078-9	standard					All species	Semantic standards		
https://fairsharing.org/1637	**10.25504/FAIRsharing.wv5q9d**	**ProtoNet**	http://www.protonet.cs.huji.ac.il/	database					All species	Knowledgebase		
https://fairsharing.org/1244	**10.25504/FAIRsharing.dmhjcg**	**Quality Control Markup Language**	http://www.psidev.info/groups/quality-control	standard	maintained		recommended		All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/2327	**10.25504/FAIRsharing.txcn6k**	**RDF Portal**	https://rdfportal.org	database	maintained	ready-to-use			All species	Knowledgebase		

FAIRsharing URL	FAIRsharing DOI	Record Name	Record homepage URL	Resource Type	Manintenace	Ready-to-use	FAIRorg recommended	plant species accepted	Applied to species	Type	ELIXIR recommended	provider
https://fairsharing.org/1269	**10.25504/FAIRsharing.a4wgws**	**Real-time PCR Data Markup Language**	http://rdml.org/	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/2177	**10.25504/FAIRsharing.pxnqyt**	**RepeatsDB**	http://repeatsdb.bio.unipd.it/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/4548	**10.25504/FAIRsharing.fa2cf3**	**Resource Identification Portal**	https://scicrunch.org/resources	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2009	**10.25504/FAIRsharing.9sb9qh**	**Restriction enzymes and methylases database**	http://rebase.neb.com/rebase/rebase.html	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2396	**10.25504/FAIRsharing.fex4c8**	**Rfam**	http://rfam.xfam.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2039	**10.25504/FAIRsharing.tn11nn**	**RIKEN Bioresource Research Center**	https://web.brc.riken.jp/en/	database		ready-to-use	recommended	Arabidopsis	All species	Knowledgebase		
https://fairsharing.org/3200	**10.25504/FAIRsharing.H3pwqc**	**RNA Atlas of Structure Probing**	http://rasp.zhanglab.net	database	maintained	ready-to-use		Arabidopsis/rice	All species	Knowledgebase		
https://fairsharing.org/1640	**10.25504/FAIRsharing.vebj4e**	**RNA Characterization of Secondary Structure Motifs**	https://www.rnacossmos.com/index.php	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/308	**10.25504/FAIRsharing.e729c4**	**RNA Markup Language**	http://www-lbit.iro.umontreal.ca/rnaml/	standard		ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/65	**10.25504/FAIRsharing.kqt2h2**	**RNA Ontology**	https://github.com/bgsu-rna/rnao	standard	maintained	ready-to-use			All species	Semantic standards		github
https://fairsharing.org/2891	**10.25504/FAIRsharing.KcCjL7**	**RNAcentral**	https://rnacentral.org	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1739	**10.25504/FAIRsharing.zzgvrv**	**RNAJunction**	https://rnajunction.ncifcrf.gov/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2853	**10.25504/FAIRsharing.6mMhZp**	**rPredictor**	http://rpredictordb.elixir-czech.cz/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1810	**10.25504/FAIRsharing.x16th8**	**RTPrimerDB**	http://medgen.ugent.be/rtprimerdb/	database					All species	Knowledgebase		UGENT(Beligum)
https://fairsharing.org/2231	**10.25504/FAIRsharing.bb5rna**	**Sea scientific open data publication**	https://www.seanoe.org/	database	maintained	ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/138	**10.25504/FAIRsharing.k97xzh**	**Sequence Alignment Map**	https://github.com/samtools/samtools	standard		ready-to-use			All species	Model/formats or syntax:Format		github
https://fairsharing.org/238	**10.25504/FAIRsharing.6bc7h9**	**Sequence Ontology**	http://www.sequenceontology.org/	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/22	**10.25504/FAIRsharing.cnbjkn**	**Single Nucleotide Polymorphism Ontology**	https://members.loria.fr/ACoulet/files/snpontology1.6_description.html	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/1673	**10.25504/FAIRsharing.kq47fy**	**SpliceDisease**	http://cmbi.bjmu.edu.cn/Sdisease	database					All species	Knowledgebase		
https://fairsharing.org/2283	**10.25504/FAIRsharing.ekj9zx**	**Standards for Reporting Enzymology Data Database**	https://www.beilstein-strenda-db.org/strenda/	standard	maintained	ready-to-use	recommended		All species	Minimum reporting guidelines		
https://fairsharing.org/2057	**10.25504/FAIRsharing.x93ckv**	**Stanford Microarray Database**	http://smd.princeton.edu/	database					All species	Data repository		
https://fairsharing.org/1877	**10.25504/FAIRsharing.q7bkqr**	**STING**	http://sms.cbi.cnptia.embrapa.br/SMS/index_s.html	database					All species	Knowledgebase		
https://fairsharing.org/139	**10.25504/FAIRsharing.dpgp6t**	**Stockholm Multiple Alignment Format**	http://sonnhammer.sbc.su.se/Stockholm.html	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/4524	**N/A**	**Stress Knowledge Map**	https://skm.nib.si/	database	maintained			All plant species		Knowledgebase		
https://fairsharing.org/1875	**10.25504/FAIRsharing.9b7wvk**	**STRING**	https://string-db.org/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2239	**10.25504/FAIRsharing.bya6z**	**Super-Enhancer Archive**	http://sea.edbc.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/116	**10.25504/FAIRsharing.mwmbpq**	**Browser Extensible Data Format**	http://genome.ucsc.edu/FAQ/FAQformat.html#format1	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/1952	**10.25504/FAIRsharing.frryvx**	**SUPERFAMILY**	http://supfam.org	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/132	**10.25504/FAIRsharing.gnga3t**	**Personal Genome SNP Format**	http://genome.ucsc.edu/FAQ/FAQformat.html#format10	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/32	**10.25504/FAIRsharing.20df5w**	**ENCODE peak information Format**	http://genome.ucsc.edu/FAQ/FAQformat.html#format13	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/901	**10.25504/FAIRsharing.9sdcx8**	**Multiple Alignment Format**	http://genome.ucsc.edu/FAQ/FAQformat.html#format5	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/98	**10.25504/FAIRsharing.ckjmt3**	**Binary sequence information Format**	http://genome.ucsc.edu/FAQ/FAQformat.html#format7	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/131	**10.25504/FAIRsharing.wkh9j2**	**nucleotide inFormation binary Format**	http://genome.ucsc.edu/FAQ/FAQformat.html#format8	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/148	**10.25504/FAIRsharing.fe4pja**	**Gene Prediction File Format**	http://genome.ucsc.edu/FAQ/FAQformat.html#format9	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/1691	**10.25504/FAIRsharing.se4zhk**	**SuperTarget**	http://bioinformatics.charite.de/supertarget	database					All species	Knowledgebase		
https://fairsharing.org/1087	**10.25504/FAIRsharing.cwf4py**	**Synthetic Biology Open Language**	https://sbolstandard.org/	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/3	**10.25504/FAIRsharing.dg76vs**	**Synthetic Biology Open Language Visual**	https://sbolstandard.org/visual-about/	standard	maintained	ready-to-use			All species	Model/formats or syntax:models/syntax		
https://fairsharing.org/1786	**10.25504/FAIRsharing.9fz3g3**	**Tandem Repeats Database**	http://tandem.bu.edu/cgi-bin/trdb/trdb.exe	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1773	**10.25504/FAIRsharing.en9npn**	**The Barcode of Life Data Systems**	https://www.boldsystems.org/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2381	**10.25504/FAIRsharing.jsxqrk**	**The Department of Energy Systems Biology Knowledgebase**	https://www.kbase.us	database	maintained	ready-to-use	recommended	All plant species		Knowledgebase		
https://fairsharing.org/1889	**10.25504/FAIRsharing.aftff2**	**The Gene Index Project**	http://compbio.dfci.harvard.edu/tgi/	database					All species	Knowledgebase		
https://fairsharing.org/2533	**10.25504/FAIRsharing.8nq9t6**	**The Network Data Exchange**	https://www.ndexbio.org	database	maintained	ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/4907	**N/A**	**The Nucleic Acid Knowledgebase**	http://nakb.org/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1643	**10.25504/FAIRsharing.6648ht**	**The SEQanswers wiki**	https://www.seqanswers.com/	database	maintained				All species	Knowledgebase		
https://fairsharing.org/2736	**10.25504/FAIRsharing.a1de61**	**The Track Hub Registry**	https://www.trackhubregistry.org/	database		ready-to-use			All species	Knowledgebase		
https://fairsharing.org/482	**10.25504/FAIRsharing.3pxg2f**	**Axt Alignment Format**	http://genome.ucsc.edu/goldenPath/help/axt.html	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/78	**10.25504/FAIRsharing.hza1ec**	**Binary Alignment Map Format**	http://genome.ucsc.edu/goldenPath/help/bam.html	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/525	**10.25504/FAIRsharing.vttygv**	**BEDgraph**	http://genome.ucsc.edu/goldenPath/help/bedgraph.html	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/666	**10.25504/FAIRsharing.wzp79x**	**Big Chain**	http://genome.ucsc.edu/goldenPath/help/bigChain.html	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)

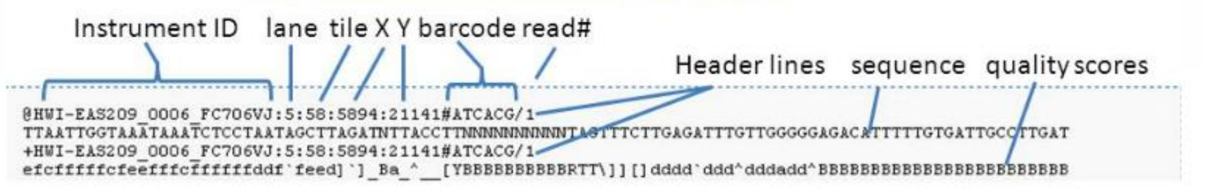
FAIRsharing URL	FAIRsharing DOI	Record Name	Record homepage URL	Resource Type	Manintenace	Ready-to-use	FAIRorg recommended	plant species accepted	Applied to species	Type	ELIXIR recommended	provider
https://fairsharing.org/660	**10.25504/FAIRsharing.6eg9a3**	**Big Gene Prediction**	http://genome.ucsc.edu/goldenPath/help/bigGenePred.html	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/615	**10.25504/FAIRsharing.78make**	**Big Multiple Alignment Format**	http://genome.ucsc.edu/goldenPath/help/bigMaf.html	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/517	**10.25504/FAIRsharing.c9fakh**	**Big Pattern Space Layout**	http://genome.ucsc.edu/goldenPath/help/bigPsl.html	standard	maintained	ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/518	**10.25504/FAIRsharing.x9k6a1**	**bigWig Track Format**	http://genome.ucsc.edu/goldenPath/help/bigWig.html	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/110	**10.25504/FAIRsharing.72e4we**	**Chain Format for pairwise alignment**	http://genome.ucsc.edu/goldenPath/help/chain.html	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/128	**10.25504/FAIRsharing.evxcfb**	**net alignment annotation Format**	http://genome.ucsc.edu/goldenPath/help/net.html	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/864	**10.25504/FAIRsharing.2nrf9f**	**Wiggle Track Format**	http://genome.ucsc.edu/goldenPath/help/wiggle.html	standard		ready-to-use			All species	Model/formats or syntax:Format		UCSC(USA)
https://fairsharing.org/405	**10.25504/FAIRsharing.9y3gv0**	**Enzyme Structure Function Ontology**	http://sfld.rbvi.ucsf.edu/	standard		ready-to-use			All species	Semantic standards		UCSF(USA)
https://fairsharing.org/2641	**N/A**	**Structure Function Linkage Database Archive**	http://sfld.rbvi.ucsf.edu/	database					All species	Knowledgebase		UCSF(USA)
https://fairsharing.org/2873	**10.25504/FAIRsharing.83jovl**	**T-psi-C**	http://tpsic.igcz.poznan.pl/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1360	**10.25504/FAIRsharing.ylYltV**	**Transcription Factor Class Schema**	http://www.edgar-wingender.de/TFClass_schema.html	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/3655	**10.25504/FAIRsharing.de533c**	**Translational Data Catalog**	https://datacatalog.elixir-luxembourg.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/1903	**10.25504/FAIRsharing.1nwy41**	**Transmembrane Helices in Genome Sequences**	http://pranag.physics.iisc.ernet.in/thgs/	database					All species	Knowledgebase		
https://fairsharing.org/2022	**10.25504/FAIRsharing.p3bzqb**	**Transporter Classification Database**	https://tcdb.org/	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/2072	**10.25504/FAIRsharing.zcn4w4**	**TreeBase**	https://www.treebase.org	database		ready-to-use	recommended		All species	Knowledgebase		
https://fairsharing.org/2338	**10.25504/FAIRsharing.x6nr7d**	**Tropical Data Hub**	https://tropicaldatahub.org/	database					All species	Knowledgebase		
https://fairsharing.org/935	**10.25504/FAIRsharing.wf28wm**	**UniProt Taxonomy**	http://www.uniprot.org/taxonomy/	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/2264	**10.25504/FAIRsharing.j97pjn**	**Validated Antibody Database**	http://www.labome.com	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/125	**10.25504/FAIRsharing.cfzz0h**	**Variant Call Format**	https://www.ga4gh.org/product/genetic-variation-formats-vcf/	standard		ready-to-use			All species	Model/formats or syntax:Format		
https://fairsharing.org/546	**10.25504/FAIRsharing.65xkbs**	**Variation Ontology**	http://variationontology.org	standard	maintained	ready-to-use			All species	Semantic standards		
https://fairsharing.org/2819	**10.25504/FAIRsharing.EGn1ut**	**WALTZ-DB 2.0**	http://waltzdb.switchlab.org	database	maintained	ready-to-use			All species	Knowledgebase		
https://fairsharing.org/710	**10.25504/FAIRsharing.j4encb**	**XEML Environment Ontology**	http://xeo.codeplex.com/	standard					All species	Semantic standards		
https://fairsharing.org/17	**10.25504/FAIRsharing.3mprr8**	**XML for evolutionary biology and comparative genomics.**	http://www.phyloxml.org/	standard		ready-to-use			All species	Model/formats or syntax:models/syntax		

Annex 2. Examples of file format standards

FASTA

```
>Solyc05g012020.2.1 Ripening Inhibitor (RIN) MADS-box transcription factor MADS-MC (AHRD V1 *-** Q8S4L5_SOLLC); contains Interpro domain(s) IPR002487  Transcription factor, K-box
TCATGATATATAGTAAAACAAAAAAGTCTTTATTCTCTTTTCTTCTTGACTAGGGAACCATTAGATTTTAAAGACATTAA
ATCTATTACCCTTACCCTAAGAATAAGAAGATGTAAAGTAGAAGAGAAAACAACCAAAACCATATATATACATATATATA
ATTACATTATATTGTCTTATAACATATAGTCTTTTAAGGAAAAACAAATTTAGAAAAAAATAATATTATTTTACATTTTT
TTTCTTCATACAATATGGGTAGAGGGAAAGTAGAATTGAAGAGAATTGAGAACAAAATAAATAGACAAGTTACCTTTGCA
```

FASTQ



```
@HWI-EAS209_0006_FC706VJ:5:58:5894:21141#ATCACG/1
TTAATTGGTAAATAAATCTCCTAATAGCTTAGATNTTACCTTNNNNNNNNNNTAGTTTCTTGAGATTTGTTGGGGAGACATTTTTGTGATTGCCTTGAT
+HWI-EAS209_0006_FC706VJ:5:58:5894:21141#ATCACG/1
efcfffffcfeefffcffffffddf`feed]`]_Ba_^__[YBBBBBBBBBBRTT\]][]dddd`ddd^dddadd^BBBBBBBBBBBBBBBBBBBBBBBB
```

EMBL Format

```
ID   X64011; SV 1; linear; genomic DNA; STD; PRO; 756 BP.
XX
AC   X64011; S78972;
XX
SV   X64011.1
XX
DT   28-APR-1992 (Rel. 31, Created)
DT   30-JUN-1993 (Rel. 36, Last updated, Version 6)
XX
DE   Listeria ivanovii sod gene for superoxide dismutase
XX
KW   sod gene; superoxide dismutase.
XX
OS   Listeria ivanovii
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Listeria.
XX
RN   [1]
RX   MEDLINE; 92140371.
RA   Haas A., Goebel W.;
RT   "Cloning of a superoxide dismutase gene from Listeria ivanovii by
RT   functional complementation in Escherichia coli and characterization of the
RT   gene product.";
RL   Mol. Gen. Genet. 231:313-322(1992).
XX
RN   [2]
RP   1-756
RA   Kreft J.;
RT   ;
RL   Submitted (21-APR-1992) to the EMBL/GenBank/DDBJ databases.
RL   J. Kreft, Institut f. Mikrobiologie, Universitaet Wuerzburg, Biozentrum Am
RL   Hubland, 8700 Wuerzburg, FRG
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..756
FT                   /db_xref="taxon:1638"
FT                   /organism="Listeria ivanovii"
FT                   /strain="ATCC 19119"
```

```
FT              /mol_type="genomic DNA"
FT   regulatory      95..100
FT              /gene="sod"
FT              /regulatory_class="ribosome_binding_site"
FT   regulatory      723..746
FT              /gene="sod"
FT              /regulatory_class="terminator"
FT   CDS           109..717
FT              /transl_table=11
FT              /gene="sod"
FT              /EC_number="1.15.1.1"
FT              /db_xref="GOA:P28763"
FT              /db_xref="HSSP:P00448"
FT              /db_xref="InterPro:IPR001189"
FT              /db_xref="UniProtKB/Swiss-Prot:P28763"
FT              /product="superoxide dismutase"
FT              /protein_id="CAA45406.1"
FT              /translation="MTYELPKLPYTYDALEPNFDKETMEIHYTKHHNIYVTKLNEAVSG
FT              HAELASKPGEELVANLDSVPEEIRGAVRNHGGGHANHTLFWSSLSPNGGGAPTGNLKAA
FT              IESEFGTFDEFKEKFNAAAAARFGSGWAWLVVNNGKLEIVSTANQDSPLSEGKTPVLGL
FT              DVWEHAYYLKFQNRRPEYIDTFWNVINWDERNKRFDAAK"
XX
SQ   Sequence 756 BP; 247 A; 136 C; 151 G; 222 T; 0 other;
     cgttatttaa ggtgttacat agttctatgg aaatagggtc tatacctttc gccttacaat   60
     gtaatttctt ..........                                               120
//
```

GenBank Format

```
LOCUS       LISOD                   756 bp    DNA     linear   BCT 30-JUN-1993
DEFINITION  Listeria ivanovii sod gene for superoxide dismutase.
ACCESSION   X64011 S78972
VERSION     X64011.1  GI:44010
KEYWORDS    sod gene; superoxide dismutase.
SOURCE      Listeria ivanovii
  ORGANISM  Listeria ivanovii
            Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE   1  (bases 1 to 756)
  AUTHORS   Haas,A. and Goebel,W.
  TITLE     Cloning of a superoxide dismutase gene from Listeria ivanovii by
            functional complementation in Escherichia coli and characterization
            of the gene product
  JOURNAL   Mol. Gen. Genet. 231 (2), 313-322 (1992)
  MEDLINE   92140371
REFERENCE   2  (bases 1 to 756)
  AUTHORS   Kreft,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1992) J. Kreft, Institut f. Mikrobiologie,
            Universitaet Wuerzburg, Biozentrum Am Hubland, 8700 Wuerzburg, FRG
FEATURES             Location/Qualifiers
     source          1..756
                     /organism="Listeria ivanovii"
                     /strain="ATCC 19119"
                     /db_xref="taxon:1638"
                     /mol_type="genomic DNA"
     regulatory      95..100
                     /gene="sod"
                     /regulatory_class="ribosome_binding_site"
```

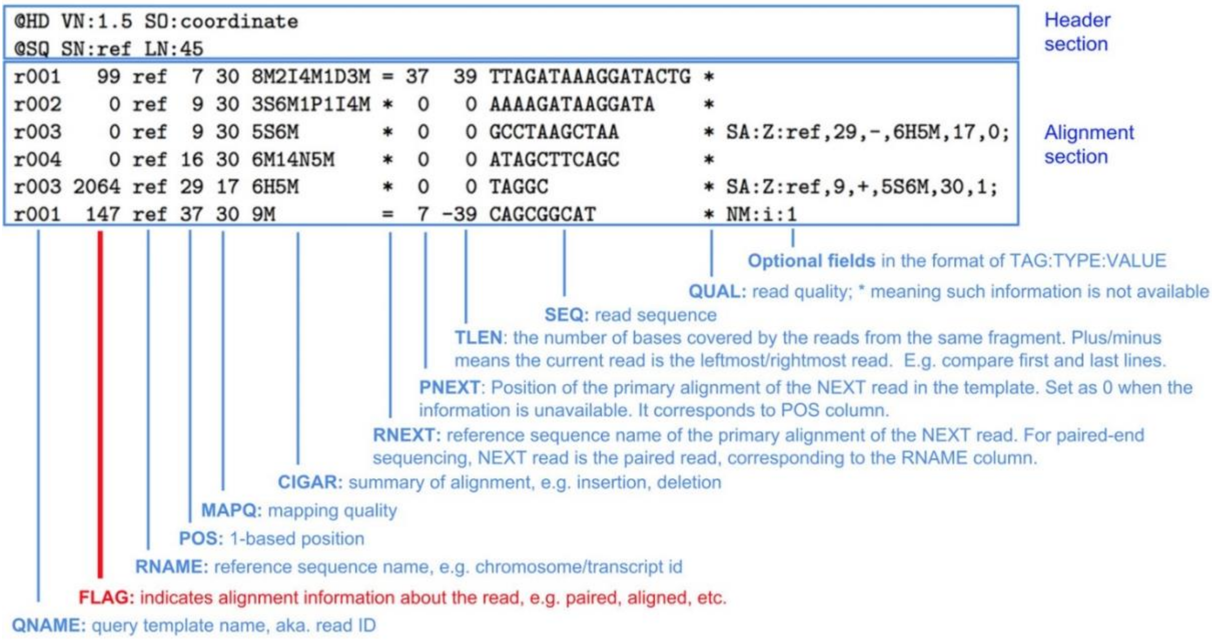
```
   gene          95..746
                 /gene="sod"
   CDS          109..717
                 /gene="sod"
                 /EC_number="1.15.1.1"
                 /codon_start=1
                 /transl_table=11
                 /product="superoxide dismutase"
                 /db_xref="GI:44011"
                 /db_xref="GOA:P28763"
                 /db_xref="InterPro:IPR001189"
                 /db_xref="UniProtKB/Swiss-Prot:P28763"
                 /protein_id="CAA45406.1"
                 /translation="MTYELPKLPYTYDALEPNFDKETMEIHYTKHHNIYVTKLNEAVS
                 GHAELASKPGEELVANLDSVPEEIRGAVRNHGGGHANHTLFWSSLSPNGGGAPTGNLK
                 AAIESEFGTFDEFKEKFNAAAAARFGSGWAWLVVNNGKLEIVSTANQDSPLSEGKTPV
                 LGLDVWEHAYYLKFQNRRPEYIDTFWNVINWDERNKRFDAAK"
   regulatory    723..746
                 /gene="sod"
                 /regulatory_class="terminator"
ORIGIN
      1 cgttatttaa ggtgttacat agttctatgg aaatagggtc tataccttc gccttacaat
     61 gtaatttctt ..........
//
```

DDBJ flat file format

```
LOCUS       LISOD                    756 bp    DNA     linear   BCT 30-JUN-1993
DEFINITION  Listeria ivanovii sod gene for superoxide dismutase.
ACCESSION   X64011 S78972
VERSION     X64011.1  GI:44010
KEYWORDS    sod gene; superoxide dismutase.
SOURCE      Listeria ivanovii
  ORGANISM  Listeria ivanovii
            Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE   1  (bases 1 to 756)
  AUTHORS   Haas,A. and Goebel,W.
  TITLE     Cloning of a superoxide dismutase gene from Listeria ivanovii by
            functional complementation in Escherichia coli and characterization
            of the gene product
  JOURNAL   Mol. Gen. Genet. 231 (2), 313-322 (1992)
  MEDLINE   92140371
REFERENCE   2  (bases 1 to 756)
  AUTHORS   Kreft,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1992) J. Kreft, Institut f. Mikrobiologie,
            Universitaet Wuerzburg, Biozentrum Am Hubland, 8700 Wuerzburg, FRG
FEATURES             Location/Qualifiers
   source        1..756
                 /organism="Listeria ivanovii"
                 /strain="ATCC 19119"
                 /db_xref="taxon:1638"
                 /mol_type="genomic DNA"
   regulatory    95..100
                 /gene="sod"
                 /regulatory_class="ribosome_binding_site"
   gene          95..746
                 /gene="sod"
   CDS          109..717
```

```
                     /gene="sod"
                     /EC_number="1.15.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="superoxide dismutase"
                     /db_xref="GOA:P28763"
                     /db_xref="HSSP:P00448"
                     /db_xref="InterPro:IPR001189"
                     /db_xref="UniProtKB/Swiss-Prot:P28763"
                     /protein_id="CAA45406.1"
                     /translation="MTYELPKLPYTYDALEPNFDKETMEIHYTKHHNIYVTKLNEAVS
                     GHAELASKPGEELVANLDSVPEEIRGAVRNHGGGHANHTLFWSSLSPNGGGAPTGNLK
                     AAIESEFGTFDEFKEKFNAAAAARFGSGWAWLVVNNGKLEIVSTANQDSPLSEGKTPV
                     LGLDVWEHAYYLKFQNRRPEYIDTFWNVINWDERNKRFDAAK"
     regulatory      723..746
                     /gene="sod"
                     /regulatory_class="terminator"
BASE COUNT        247 a        136 c        151 g        222 t
ORIGIN
        1 cgttatttaa ggtgttacat agttctatgg aaatagggtc tatacctttc gccttacaat
       61 gtaatttctt ..........
//
```

SAM (from https://www.samformat.info/sam-format-flag)



```
@HD VN:1.5 SO:coordinate
@SQ SN:ref LN:45
r001   99 ref  7 30 8M2I4M1D3M = 37  39 TTAGATAAAGGATACTG *
r002    0 ref  9 30 3S6M1P1I4M *  0   0 AAAAGATAAGGATA    *
r003    0 ref  9 30 5S6M       *  0   0 GCCTAAGCTAA       * SA:Z:ref,29,-,6H5M,17,0;
r004    0 ref 16 30 6M14N5M    *  0   0 ATAGCTTCAGC       *
r003 2064 ref 29 17 6H5M       *  0   0 TAGGC             * SA:Z:ref,9,+,5S6M,30,1;
r001  147 ref 37 30 9M         =  7 -39 CAGCGGCAT         * NM:i:1
```

Optional fields in the format of TAG:TYPE:VALUE

QUAL: read quality; * meaning such information is not available

SEQ: read sequence

TLEN: the number of bases covered by the reads from the same fragment. Plus/minus means the current read is the leftmost/rightmost read. E.g. compare first and last lines.

PNEXT: Position of the primary alignment of the NEXT read in the template. Set as 0 when the information is unavailable. It corresponds to POS column.

RNEXT: reference sequence name of the primary alignment of the NEXT read. For paired-end sequencing, NEXT read is the paired read, corresponding to the RNAME column.

CIGAR: summary of alignment, e.g. insertion, deletion

MAPQ: mapping quality

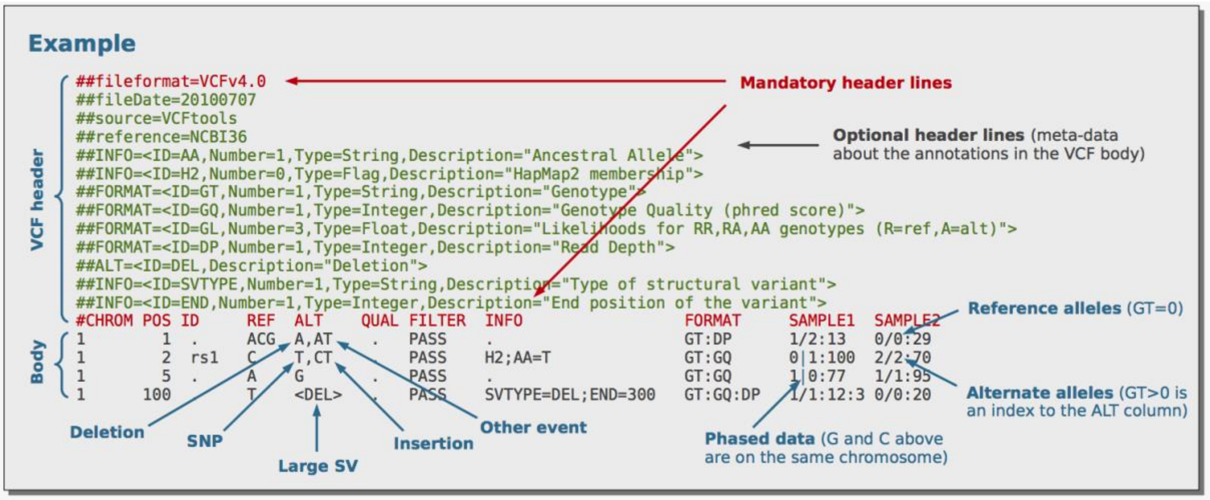
POS: 1-based position

RNAME: reference sequence name, e.g. chromosome/transcript id

FLAG: indicates alignment information about the read, e.g. paired, aligned, etc.

QNAME: query template name, aka. read ID

VCF format (from https://davetang.github.io/learning_vcf_file/)



```
##fileformat=VCFv4.0
##fileDate=20100707
##source=VCFtools
##reference=NCBI36
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality (phred score)">
##FORMAT=<ID=GL,Number=3,Type=Float,Description="Likelihoods for RR,RA,AA genotypes (R=ref,A=alt)">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##ALT=<ID=DEL,Description="Deletion">
##INFO=<ID=SVTYPE,Number=1,Type=String,Description="Type of structural variant">
##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant">
#CHROM POS ID    REF  ALT    QUAL FILTER  INFO                FORMAT    SAMPLE1   SAMPLE2
1        1 .     ACG  A,AT    .   PASS    .                   GT:DP     1/2:13    0/0:29
1        2 rs1   C    T,CT    .   PASS    H2;AA=T             GT:GQ     0|1:100   2/2:70
1        5 .     A    G       .   PASS    .                   GT:GQ     1|0:77    1/1:95
1      100 .     T    <DEL>   .   PASS    SVTYPE=DEL;END=300  GT:GQ:DP  1/1:12:3  0/0:20
```

BED format

Track line

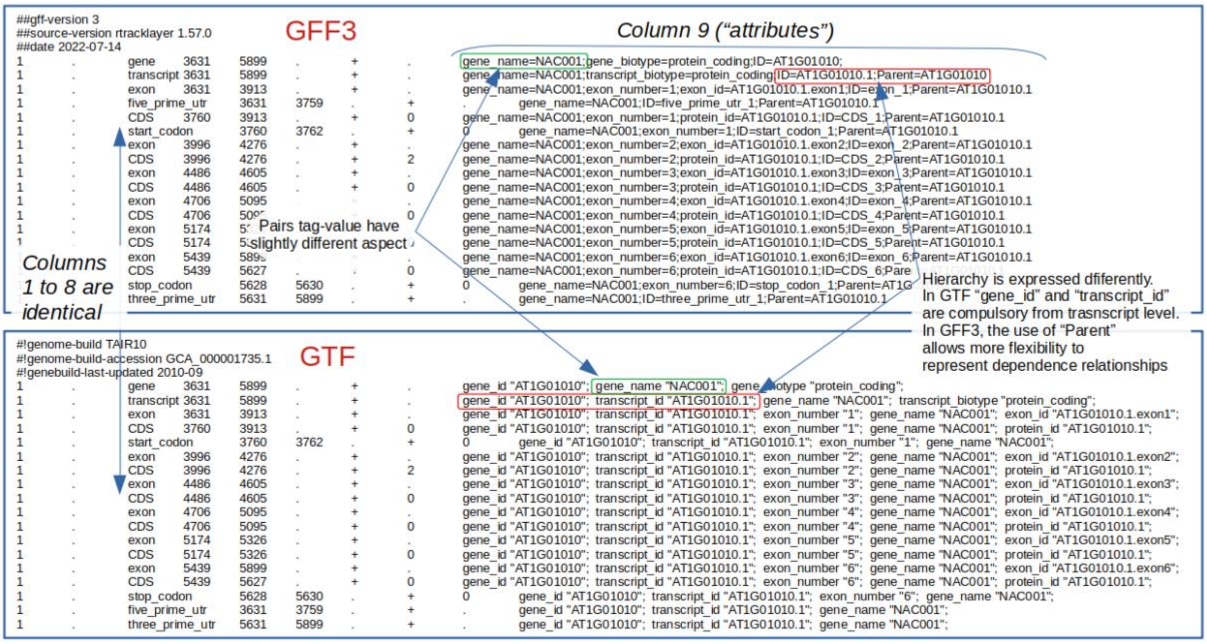
```
track name="ItemRGBDemo" description="Item RGB demonstration" itemRgb="On"
chr7  127471196  127472363  Pos1  0  +  127471196  127472363  255,0,0
chr7  127472363  127473530  Pos2  0  +  127472363  127473530  255,0,0
chr7  127473530  127474697  Pos3  0  +  127473530  127474697  255,0,0
chr7  127474697  127475864  Pos4  0  +  127474697  127475864  255,0,0
chr7  127475864  127477031  Neg1  0  -  127475864  127477031  0,0,255
chr7  127477031  127478198  Neg2  0  -  127477031  127478198  0,0,255
chr7  127478198  127479365  Neg3  0  -  127478198  127479365  0,0,255
chr7  127479365  127480532  Pos5  0  +  127479365  127480532  255,0,0
chr7  127480532  127481699  Neg4  0  -  127480532  127481699  0,0,255
```

Mandatory columns   Optional columns

BedGraph format

```
browser position chr19:49302001-49304701
browser hide all
browser pack refGene encodeRegions
browser full altGraph
#       300 base wide bar graph, autoScale is on by default == graphing
#       limits will dynamically change to always show full range of data
#       in viewing window, priority = 20 positions this as the second graph
#       Note, zero-relative, half-open coordinate system in use for bedGraph format
track type=bedGraph name="BedGraph Format" description="BedGraph format" visibility=full color=200,100,0 altColor=0,100,200 priority=20
chr19 49302000 49302300 -1.0
chr19 49302300 49302600 -0.75
chr19 49302600 49302900 -0.50
chr19 49302900 49303200 -0.25
chr19 49303200 49303500 0.0
chr19 49303500 49303800 0.25
chr19 49303800 49304100 0.50
chr19 49304100 49304400 0.75
chr19 49304400 49304700 1.00
```

GFF3/GTF format (from https://bioinfogp.cnb.csic.es/tools/seqnjoy/help/)



GFF3

Column 9 (“attributes”)

```
##gff-version 3
##source-version rtracklayer 1.57.0
##date 2022-07-14
1	.	gene	3631	5899	.	+	.	gene_name=NAC001;gene_biotype=protein_coding;ID=AT1G01010;
1	.	transcript	3631	5899	.	+	.	gene_name=NAC001;transcript_biotype=protein_coding;ID=AT1G01010.1;Parent=AT1G01010
1	.	exon	3631	3913	.	+	.	gene_name=NAC001;exon_number=1;exon_id=AT1G01010.1.exon1;ID=exon_1;Parent=AT1G01010.1
1	.	five_prime_utr	3631	3759	.	+	.	gene_name=NAC001;ID=five_prime_utr_1;Parent=AT1G01010.1
1	.	CDS	3760	3913	.	+	0	gene_name=NAC001;exon_number=1;protein_id=AT1G01010.1;ID=CDS_1;Parent=AT1G01010.1
1	.	start_codon	3760	3762	.	+	0	gene_name=NAC001;exon_number=1;ID=start_codon_1;Parent=AT1G01010.1
1	.	exon	3996	4276	.	+	.	gene_name=NAC001;exon_number=2;exon_id=AT1G01010.1.exon2;ID=exon_2;Parent=AT1G01010.1
1	.	CDS	3996	4276	.	+	2	gene_name=NAC001;exon_number=2;protein_id=AT1G01010.1;ID=CDS_2;Parent=AT1G01010.1
1	.	exon	4486	4605	.	+	.	gene_name=NAC001;exon_number=3;exon_id=AT1G01010.1.exon3;ID=exon_3;Parent=AT1G01010.1
1	.	CDS	4486	4605	.	+	0	gene_name=NAC001;exon_number=3;protein_id=AT1G01010.1;ID=CDS_3;Parent=AT1G01010.1
1	.	exon	4706	5095		+	.	gene_name=NAC001;exon_number=4;exon_id=AT1G01010.1.exon4;ID=exon_4;Parent=AT1G01010.1
1	.	CDS	4706	509			0	gene_name=NAC001;exon_number=4;protein_id=AT1G01010.1;ID=CDS_4;Parent=AT1G01010.1
1	.	exon	5174	5				gene_name=NAC001;exon_number=5;exon_id=AT1G01010.1.exon5;ID=exon_5;Parent=AT1G01010.1
1	.	CDS	5174	5				gene_name=NAC001;exon_number=5;protein_id=AT1G01010.1;ID=CDS_5;Parent=AT1G01010.1
1	.	exon	5439	589			.	gene_name=NAC001;exon_number=6;exon_id=AT1G01010.1.exon6;ID=exon_6;Parent=AT1G01010.1
1	.	CDS	5439	5627	.	+	0	gene_name=NAC001;exon_number=6;protein_id=AT1G01010.1;ID=CDS_6;Pare
1	.	stop_codon	5628	5630	.	+	0	gene_name=NAC001;exon_number=6;ID=stop_codon_1;Parent=AT1G
1	.	three_prime_utr	5631	5899	.	+	.	gene_name=NAC001;ID=three_prime_utr_1;Parent=AT1G01010.1
```

Pairs tag-value have slightly different aspect

Columns 1 to 8 are identical

Hierarchy is expressed dfiferently.
In GTF “gene_id” and “transcript_id” are compulsory from trasnscript level.
In GFF3, the use of “Parent” allows more flexibility to represent dependence relationships

GTF

```
#!genome-build TAIR10
#!genome-build-accession GCA_000001735.1
#!genebuild-last-updated 2010-09
1	.	gene	3631	5899	.	+	.	gene_id "AT1G01010";  gene_name "NAC001";  gene_biotype "protein_coding";
1	.	transcript	3631	5899	.	+	.	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  gene_name "NAC001";  transcript_biotype "protein_coding";
1	.	exon	3631	3913	.	+	.	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "1";  gene_name "NAC001";  exon_id "AT1G01010.1.exon1";
1	.	CDS	3760	3913	.	+	0	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "1";  gene_name "NAC001";  protein_id "AT1G01010.1";
1	.	start_codon	3760	3762	.	+	0	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "1";  gene_name "NAC001";
1	.	exon	3996	4276	.	+	.	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "2";  gene_name "NAC001";  exon_id "AT1G01010.1.exon2";
1	.	CDS	3996	4276	.	+	2	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "2";  gene_name "NAC001";  protein_id "AT1G01010.1";
1	.	exon	4486	4605	.	+	.	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "3";  gene_name "NAC001";  exon_id "AT1G01010.1.exon3";
1	.	CDS	4486	4605	.	+	0	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "3";  gene_name "NAC001";  protein_id "AT1G01010.1";
1	.	exon	4706	5095	.	+	.	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "4";  gene_name "NAC001";  exon_id "AT1G01010.1.exon4";
1	.	CDS	4706	5095	.	+	0	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "4";  gene_name "NAC001";  protein_id "AT1G01010.1";
1	.	exon	5174	5326	.	+	.	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "5";  gene_name "NAC001";  exon_id "AT1G01010.1.exon5";
1	.	CDS	5174	5326	.	+	0	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "5";  gene_name "NAC001";  protein_id "AT1G01010.1";
1	.	exon	5439	5899	.	+	.	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "6";  gene_name "NAC001";  exon_id "AT1G01010.1.exon6";
1	.	CDS	5439	5627	.	+	0	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "6";  gene_name "NAC001";  protein_id "AT1G01010.1";
1	.	stop_codon	5628	5630	.	+	0	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  exon_number "6";  gene_name "NAC001";
1	.	five_prime_utr	3631	3759	.	+	.	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  gene_name "NAC001";
1	.	three_prime_utr	5631	5899	.	+	.	gene_id "AT1G01010";  transcript_id "AT1G01010.1";  gene_name "NAC001";
```

Annex3. Identinfier.org ID schemes related to genetic data or their annotation

id	name	Domain	pattern	definition	prefix	url	prefixed	resources										synonyms
MIR:00000003	Ensembl	All species	^((ENS[FPTG]\\d{11}(\\.\\d+)?)|(FB\\w{2}\\d{7})|(Y[A-Z]{2}\\d{3}[a-zA-Z](\\-[A-Z])?)|([A-Z_a-z0-9]+(\\.)?(t)?(\\d+)?([a-z])?))$	Ensembl is a joint project between EMBL - EBI and the Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes. This collections also references outgroup organisms.	ensembl	http://identifiers.org/ensembl	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100011	https://www.ensembl.org/id/{$id}	Ensembl at Sanger/EMBL-EBI	Sanger Institute and European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	ENSG00000139618	Homo sapiens	https://www.ensembl.org/	
								MIR:00100561	http://uswest.ensembl.org/id/{$id}	Ensembl US West mirror		USA	false		ENSG00000139618	Homo sapiens	http://uswest.ensembl.org/	
								MIR:00100562	http://useast.ensembl.org/id/{$id}	Ensembl US East mirror		USA	false		ENSG00000139618	Homo sapiens	http://useast.ensembl.org/	
								MIR:00100563	http://asia.ensembl.org/id/{$id}	Ensembl Asia mirror		Singapore	false		ENSG00000139618	Homo sapiens	http://asia.ensembl.org/	
MIR:00000005	UniProt Knowledgebase	All species	^([A-N,R-Z][0-9]([A-Z][A-Z, 0-9][A-Z, 0-9][0-9]){1,2})|([O,P,Q][0-9][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9])(\.\d+)?$	The UniProt Knowledgebase (UniProtKB) is a comprehensive resource for protein sequence and functional information with extensive cross-references to more than 120 external databases. Besides amino acid sequence and a description, it also provides taxonomic data and citation information.	uniprot	http://identifiers.org/uniprot	0	id	accessURL	info	institution	official	localId	testString	resourceURL	location	resourcePrefix	UniProtKB
								MIR:00100164	http://purl.uniprot.org/uniprot/{$id}	Universal Protein Resource using Persistent URL system	UniProt Consortium	true	P0DP23	CALM1_HUMAN	http://www.uniprot.org/			UniProt
								MIR:00100330	https://www.ncbi.nlm.nih.gov/protein/{$id}	UniProt through NCBI	National Center for Biotechnology Information (NCBI)	false	P0DP23	Calmodulin	https://www.ncbi.nlm.nih.gov/protein/	USA	ncbi	Protein Knowledgebase
																		UniProt-TrEMBL
																		UniProt/TrEMBL
																		UniProtKB/Swiss-Prot
MIR:00000010	IntAct	All species	^EBI\-[0-9]+$	IntAct provides a freely available, open source database system and analysis tools for protein interaction data.	intact	http://identifiers.org/intact	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100017	https://www.ebi.ac.uk/intact/interaction/{$id}	IntAct at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	EBI-2307691	ATP hydrolysis by RAD51 - inhibition by BRCA2	https://www.ebi.ac.uk/intact/	
MIR:00000011	InterPro	All species	^IPR\d{6}$	InterPro is a database of protein families, domains and functional sites in which identifiable features found in known proteins can be applied to unknown protein sequences.	interpro	http://identifiers.org/interpro	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100018	https://www.ebi.ac.uk/interpro/entry/{$id}	InterPro	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	IPR000100	Ribonuclease P	https://www.ebi.ac.uk/interpro/	
								MIR:00100697	http://interpro.bio2rdf.org/describe/?url=http://bio2rdf.org/interpro:{$id}	Bio2RDF	Bio2RDF.org		false		IPR000100	Ribonuclease P	http://interpro.bio2rdf.org/fct/	
MIR:00000012	KEGG Pathway	All species	^\w{2,4}\d{5}$	KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks.	kegg.pathway	http://identifiers.org/kegg.pathway	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		KEGG
								MIR:00100020	https://www.kegg.jp/entry/{$id}	KEGG PATHWAY Database	Kyoto University Bioinformatics Center	Japan	true	hsa00620	Pyruvate metabolism	https://www.genome.jp/kegg/pathway.html		
MIR:00000014	KEGG Reaction	All species	^R\d+$	KEGG reaction contains our knowledge on the universe of reactions that are relevant to life.	kegg.reaction	http://identifiers.org/kegg.reaction	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		KEGG
								MIR:00100022	https://www.kegg.jp/entry/{$id}	KEGG Reaction Database	Kyoto University Bioinformatics Center	Japan	true	R00100	NADH:ferricytochrome-b5 oxidoreductase	https://www.genome.jp/kegg/reaction/		
MIR:00000016	OMIM	Human	^[*#+%^]?\d{6}$	Online Mendelian Inheritance in Man is a catalog of human genes and genetic disorders.	mim	http://identifiers.org/mim	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Online Mendelian Inheritance in Man
								MIR:00100024	https://omim.org/entry/{$id}	OMIM at John Hopkins	Johns Hopkins University, Baltimore, Maryland	USA	true	603903	SICKLE CELL ANEMIA	https://omim.org/		Mendelian Inheritance in Man
								MIR:00100694	http://omim.bio2rdf.org/describe/?url=http://bio2rdf.org/omim:{$id}	Bio2RDF	Bio2RDF.org		false	603903	Scriver and Waugh	http://omim.bio2rdf.org/fct		MIM
								MIR:00100777	http://mirror.omim.org/entry/{$id}	OMIM mirror at John Hopkins	Johns Hopkins University, Baltimore, Maryland	USA	false	603903	SICKLE CELL ANEMIA	http://mirror.omim.org/		
MIR:00000017	PIRSF	All species	^PIRSF\d{6}$	The PIR SuperFamily concept is being used as a guiding principle to provide comprehensive and non-overlapping clustering of UniProtKB sequences into a hierarchical order to reflect their evolutionary relationships.	pirsf	http://identifiers.org/pirsf	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Protein Information Resource
								MIR:00100025	https://pir.georgetown.edu/cgi-bin/ipcSF?id={$id}	PIRSF	Georgetown University Medical Center	USA	false	PIRSF000100	shikimate 5-dehydrogenase	https://pir.georgetown.edu/		
MIR:00000018	Reactome	Human	(^R-[A-Z]{3}-\d+(-\d+)?(\.\d+)?$)\|(^REACT_\d+(\.\d+)?$)	The Reactome project is a collaboration to develop a curated resource of core pathways and reactions in human biology.	reactome	http://identifiers.org/reactome	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Reactome Stable ID
								MIR:00100026	https://reactome.org/content/detail/{$id}	Reactome, a curated knowledgebase of biological pathways	Ontario Institute for Cancer Research, NYU Medical School, Cold Spring Harbor Laboratory and European Bioinformatics Institute	Canada / USA / UK	false	R-HSA-201451	TGF-beta/BMP	https://www.reactome.org/		
MIR:00000019	DOI	all species	^(doi\:)?\d{2}\.\d{4}.*$	The Digital Object Identifier System is for identifying content objects in the digital environment.	doi	http://identifiers.org/doi	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Digital Object Identifier
								MIR:00100010	https://doi.org/{$id}	Digital Object Identifier	International DOI Foundation	United Kingdom	true	10.1038/nbt1156	Minimum information requested in the annotation of biochemical models	https://www.doi.org/		
MIR:00000020	Protein Data Bank	All species	^[0-9][A-Za-z0-9]{3}$	The Protein Data Bank is the single worldwide archive of structural data of biological macromolecules.	pdb	http://identifiers.org/pdb	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	PDB
								MIR:00100029	https://www.rcsb.org/pdb/explore/explore.do?structureId={$id}	RCSB PDB	Rutgers, The State University of New Jersey	USA	false	rcsb	2gc4	methylamine dehydrogenase	https://www.pdb.org/	
								MIR:00100037	http://www.pdbe.org/{$id}	Protein Databank in Europe (PDBe)	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	pdbe	2gc4	Methylamine dehydrogenase heavy chain	http://www.pdbe.org/	
								MIR:00100096	http://proteopedia.org/wiki/index.php/{$id}	Proteopedia	Weizmann Institute of Science	Israel	false		2gc4	Paracoccus denitrificans	http://www.proteopedia.org/	
								MIR:00100165	http://pdbj.org/mine/summary/{$id}	Protein Data Bank Japan (PDBj)	Institute for Protein Research, Osaka University	Japan	false	pdbj	2gc4	methylamine dehydrogenase	http://www.pdbj.org/	
								MIR:00100166	https://www.ebi.ac.uk/pdbsum/{$id}	Protein Databank through PDBsum	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	2gc4	Oxidoreductase, electron transport	https://www.ebi.ac.uk/pdbsum/	
MIR:00000022	Gene Ontology	All species	^GO:\d{7}$	The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism.	go	http://identifiers.org/go	1	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	GO
								MIR:00100012	https://www.ebi.ac.uk/QuickGO/GTerm?id=GO:{$id}	QuickGO (Gene Ontology browser)	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	quickgo	6915	apoptosis	https://www.ebi.ac.uk/QuickGO/	
								MIR:00100013	http://amigo.geneontology.org/amigo/term/GO:{$id}	AmiGO 2	The Gene Ontology Consortium	USA	true	amigo	6915	apoptotic process	http://amigo.geneontology.org/	
								MIR:00100015	http://www.informatics.jax.org/searches/GO.cgi?id=GO:{$id}	GO Browser	The Jackson Laboratory	USA	false		6915	apoptosis	http://www.informatics.jax.org/searches/GO_form.shtml	
								MIR:00100237	http://purl.bioontology.org/ontology/GO/GO:{$id}	GO through BioPortal	National Center for Biomedical Ontology, Stanford	USA	false	bptl	6915	apoptosis	http://bioportal.bioontology.org/ontologies/GO	
								MIR:00100585	http://www.pantherdb.org/panther/category.do?categoryAcc=GO:{$id}	GO through PANTHER	Keck School of Medicine, University of Southern California	USA	false		3	reproduction	http://www.pantherdb.org/	
								MIR:00100675	https://www.ebi.ac.uk/ols/ontologies/go/terms?obo_id=GO:{$id}	GO through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ols	6915	apoptosis	https://www.ebi.ac.uk/ols/ontologies/go	
MIR:00000023	SGD	Yeast	^((S\d+$)\|(Y[A-Z]{2}\d{3}[a-zA-Z](\-[A-Z])?))$	The Saccharomyces Genome Database (SGD) project collects information and maintains a database of the molecular biology of the yeast Saccharomyces cerevisiae.	sgd	http://identifiers.org/sgd	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	Saccharomyces Genome Database
								MIR:00100033	https://www.yeastgenome.org/locus/{$id}	SGD	Stanford University	USA	true	S000003909	BAT2	https://www.yeastgenome.org/		
								MIR:00100703	http://sgd.bio2rdf.org/describe/?url=http://bio2rdf.org/sgd:{$id}	Bio2RDF	Bio2RDF.org		false	S000003909	DNA-binding transcription factor	http://sgd.bio2rdf.org/fct		
								MIR:00100892	https://bioentity.link/#/lexicon/public/{$id}	SGD through BioEntity Link	InSilico Inc, in Eugene, Oregon	USA	false	S000003909		https://bioentity.link/	bioentitylink	
								MIR:00100906	https://www.alliancegenome.org/gene/SGD:{$id}	SGD through the Alliance of Genome Resources	Alliance of Genome Resources	USA	false	S000003909		https://www.alliancegenome.org	agr	
MIR:00000024	Systems Biology Ontology	All species	^SBO:\d{7}$	The goal of the Systems Biology Ontology is to develop controlled vocabularies and ontologies tailored specifically for the kinds of problems being faced in Systems Biology, especially in the context of computational modeling. SBO is a project of the BioModels.net effort	sbo	http://identifiers.org/sbo	1	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	SBO
								MIR:00100034	https://www.ebi.ac.uk/sbo/main/SBO:{$id}	SBO	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	262	enzymatic rate law for simple uncompetitive inhibition of irreversible unireactant	https://www.ebi.ac.uk/sbo/	
								MIR:00100176	https://www.ebi.ac.uk/ols/ontologies/sbo/terms?obo_id=SBO:{$id}	SBO through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ols	262	enzymatic rate law for simple uncompetitive inhibition of irreversible unireactant	https://www.ebi.ac.uk/ols/ontologies/sbo	

				of the BioModels.net effort.				MIR:00100242	http://purl.bioontology.org/ontology/SBO/SBO:{$id}	SBO through BioPortal	National Center for Biomedical Ontology, Stanford	USA	false	bptl	262	enzymatic rate law for simple uncompetitive inhibition of irreversible unireactant	http://bioportal.bioontology.org/ontologies/SBO	
MIR:00000027	WormBase	worm	^WB[A-Z][a-z]+\d+$	WormBase is an online bioinformatics database of the biology and genome of the model organism Caenorhabditis elegans and other nematodes. It is used by the C. elegans research community both as an information resource and as a mode to publish and distribute their results. This collection references WormBase-accessioned entities.	wb	http://identifiers.org/wb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	
								MIR:00100038	https://www.wormbase.org/get?name={$id}	WormBase	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	WBGene00000001	aap-1	https://www.wormbase.org/		
								MIR:00100704	http://wormbase.bio2rdf.org/describe/?url=http://bio2rdf.org/wormbase:{$id}	Bio2RDF	Bio2RDF.org		false	WBGene00000001	aap-1 encodes the C. elegans ortholog of the phosphoinositide 3-kinase	http://wormbase.bio2rdf.org/fct		
								MIR:00100890	https://bioentity.link/#/lexicon/public/{$id}	WormBase through BioEntity Link	InSilico Inc, in Eugene, Oregon	USA	false	WBGene00021291		https://bioentity.link/	bioentitylink	
								MIR:00100901	https://www.alliancegenome.org/gene/WB:{$id}	WormBase through the Alliance of Genome Resources	Alliance of Genome Resources	USA	false	WBGene00000001		https://www.alliancegenome.org	agr	
MIR:00000028	Pfam	All species	^PF\d{5}$	The Pfam database contains information about protein domains and families. For each entry a protein sequence alignment and a Hidden Markov Model is stored.	pfam	http://identifiers.org/pfam	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Protein Family Database
								MIR:00100685	https://pfam.xfam.org/family/{$id}	Pfam at EMBL-EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	PF01234	Methyltransf	https://pfam.xfam.org/	
MIR:00000029	Nucleotide Sequence Database	All species	^([A-Z]\d{5}|[A-Z]{2}\d{6}|[A-Z]{4}\d{8}|[A-J][A-Z]{2}\d{5})(\.\d+)?$	The International Nucleotide Sequence Database Collaboration (INSDC) consists of a joint effort to collect and disseminate databases containing DNA and RNA sequences.	insdc	http://identifiers.org/insdc	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	International Nucleotide Sequence Database Collaboration
								MIR:00100049	https://www.ncbi.nlm.nih.gov/nuccore/{$id}	INSDC through GenBank	National Center for Biotechnology Information (NCBI)	USA	false	X58356	N.sylvestris gene for 33kd chloroplast ribonucleoprotein	https://www.ncbi.nlm.nih.gov/Genbank/		INSDC
								MIR:00100066	http://getentry.ddbj.nig.ac.jp/getentry?database=ddbj&accession_number={$id}	INSDC through DDBJ	DNA Data Bank of Japan, Mishima, Shizuoka	Japan	false	X58356	N.sylvestris gene for 33kd chloroplast ribonucleoprotein	http://www.ddbj.nig.ac.jp/		NCBI nucleotide
								MIR:00100487	https://www.ebi.ac.uk/ena/data/view/{$id}	INSDC through European Nucleotide Archive (ENA)	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	X58356	N.sylvestris gene for 33kd chloroplast ribonucleoprotein	https://www.ebi.ac.uk/ena/	ebi	GenBank
								MIR:00100490	https://www.ncbi.nlm.nih.gov/nuccore/{$id}	INSDC through Nucleotide database at NCBI	National Center for Biotechnology Information (NCBI)	USA	true	X58356	N.sylvestris gene for 33kd chloroplast ribonucleoprotein	https://www.ncbi.nlm.nih.gov/nuccore/	ncbi	
MIR:00000030	FlyBase	Drosophila	^FB\w{2}\d{7}$	FlyBase is the database of the Drosophila Genome Projects and of associated literature.	fb	http://identifiers.org/fb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	
								MIR:00100050	http://flybase.org/reports/{$id}.html	The FlyBase Database	The FlyBase Consortium	USA / United Kingdom	false	FBgn0011293	antennal protein 10	http://flybase.org/		
								MIR:00100893	https://bioentity.link/#/lexicon/public/{$id}	FlyBase through BioEntity Link	InSilico Inc, in Eugene, Oregon	USA	false	FBgn0011293		https://bioentity.link/	bioentitylink	
								MIR:00100904	https://www.alliancegenome.org/gene/FB:{$id}	FlyBase through the Alliance of Genome Resources	Alliance of Genome Resources	USA	false	FBgn0011293		https://www.alliancegenome.org	agr	
MIR:00000031	Wormpep	worm	^CE\d{5}$	Wormpep contains the predicted proteins from the Caenorhabditis elegans genome sequencing project.	wormpep	http://identifiers.org/wormpep	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100051	https://www.wormbase.org/db/seq/protein?name=WP:{$id}	Wormpep (Master)	Cold Spring Harbor Laboratory	USA	true	CE28239	UNC-26, isoform a	https://www.wormbase.org/db/seq/protein		
MIR:00000032	PROSITE	All species	^PS\d{5}$	PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them.	prosite	http://identifiers.org/prosite	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	resourceURL		
								MIR:00100056	https://prosite.expasy.org/{$id}	ExPASy PROSITE	Swiss Institute of Bioinformatics (SIB)	Switzerland	false	expasy	PS00001	https://www.expasy.org/prosite/		
MIR:00000036	ArrayExpress	All species	^[AEP]-\w{4}-\d+$	ArrayExpress is a public repository for microarray data, which is aimed at storing MIAME-compliant data in accordance with Microarray Gene Expression Data (MGED) recommendations.	arrayexpress	http://identifiers.org/arrayexpress	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100061	https://www.ebi.ac.uk/arrayexpress/experiments/{$id}	ArrayExpress	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	E-MEXP-1712	Transcription profiling of Streptococcus gordonii wildtype and Brf- deletion and	https://www.ebi.ac.uk/arrayexpress/	
								MIR:00100854	https://www.omicsdi.org/dataset/arrayexpress-repository/{$id}	ArrayExpress through OmicsDI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	omicsdi	E-MEXP-1712		https://www.omicsdi.org/	
MIR:00000037	Mouse Genome Database	Mouse	^MGI:\d+$	The Mouse Genome Database (MGD) project includes data on gene characterization, nomenclature, mapping, gene homologies among mammals, sequence links, phenotypes, allelic variants and mutants, and strain data.	mgi	http://identifiers.org/mgi	1	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	MGD
								MIR:00100062	http://www.informatics.jax.org/accession/MGI:{$id}	Mouse Genome Informatics Database	The Jackson Laboratory	USA	true	2442292	Dickie's small eye	http://www.informatics.jax.org/		MGI
								MIR:00100894	https://bioentity.link/#/lexicon/public/MGI:{$id}	MGD through BioEntity Link	InSilico Inc, in Eugene, Oregon	USA	false	2442292		https://bioentity.link/	bioentitylink	Mouse Genome Informatics
								MIR:00100903	https://www.alliancegenome.org/gene/MGI:{$id}	MGI through the Alliance of Genome Resources	Alliance of Genome Resources	USA	false	2442292		https://www.alliancegenome.org	agr	
MIR:00000039	RefSeq	All species	^((AC|AP|NC|NG|NM|NP|NR|NT|NW|XM|XP|XR|YP|ZP)_\d+|(NZ_[A-Z]{4}\d+))(\.\d+)?$	The Reference Sequence (RefSeq) collection aims to provide a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products.	refseq	http://identifiers.org/refseq	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100067	https://www.ncbi.nlm.nih.gov/protein/{$id}	The NCBI Reference Sequence database	NCBI, NIH, Bethesda, Maryland	USA	false	ncbi	NP_012345	Saccharomyces cerevisiae S288c	https://www.ncbi.nlm.nih.gov/projects/RefSeq/	
MIR:00000040	Transport Classification Database	All species	^\d+(\.[A-Z])?(\.\d+)?(\.\d+)?(\.\d+)?$	The database details a comprehensive IUBMB approved classification system for membrane transport proteins known as the Transporter Classification (TC) system. The TC system is analogous to the Enzyme Commission (EC) system for classification	tcdb	http://identifiers.org/tcdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		TCDB
								MIR:00100068	http://www.tcdb.org/search/result.php?tc={$id}	TCDB	Division of Biological Sciences, UCSD	USA	false	5.A.1.1.1	Disulfide bond oxidoreductase-D	http://www.tcdb.org/		
MIR:00000041	UniParc	All species	^UPI[A-F0-9]{10}$	The UniProt Archive (UniParc) is a database containing non-redundant protein sequence information from many sources. Each unique sequence is given a stable and unique identifier (UPI) making it possible to identify the same protein from different source databases.	uniparc	http://identifiers.org/uniparc	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100069	https://www.ebi.ac.uk/cgi-bin/dbfetch?db=uniparc&id={$id}	UniProt Archive (UniParc)	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	UPI000000000A	MAETKEFKTLYNLFIDSYLQKLAQHSIPTNVTCAIHIGEVIGQFKNCALRITNKCMSNSR	https://www.ebi.ac.uk/uniparc/	
								MIR:00100476	http://www.uniprot.org/uniparc/{$id}	UniParc through UniProt	UniProt Consortium	USA, UK and Switzerland	false		UPI000000000A	MAETKEFKTLYNLFIDSYLQKLAQHSIPTNVTCAIHIGEVIGQFKNCALRI	http://www.uniprot.org/uniparc/	
MIR:00000042	MINT	All species	^MINT\-\d{1,7}$	The Molecular INTeraction database (MINT) stores, in a structured format, information about molecular interactions by extracting experimental details from work published in peer-reviewed journals.	mint	http://identifiers.org/mint	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	
								MIR:00100070	http://mint.bio.uniroma2.it/mint/search/inFrameInteraction.do?interactionAc={$id}	The Molecular INTeraction database (MINT)	University of Rome Tor Vergata, Rome	Italy	false	MINT-10000	Protein transport protein SEC13	http://mint.bio.uniroma2.it/mint/		
								MIR:00100654	https://www.ebi.ac.uk/intact/query/interaction_id:{$id}	MINT subset through IntAct	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	MINT-7905142	http://mint.bio.uniroma2.it/mint/search/interaction.do?ac=MINT-7905142	https://www.ebi.ac.uk/intact/	ebi	
MIR:00000045	Signaling Gateway	All species	A\d{6}$	The Signaling Gateway provides information on mammalian proteins involved in cellular signaling.	signaling-gateway	http://identifiers.org/signaling-gateway	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		SGMP
								MIR:00100073	http://www.signaling-gateway.org/molecule/query?afcsid={$id}	Signaling Gateway Molecule Pages	UC San Diego	USA	false	A001094	Adrenergic receptor kinase	http://www.signaling-gateway.org/molecule		Signaling Gateway Molecule Pages
																		AfCS
																		UCSD Signaling Gateway Molecule Pages
																		UCSD - Molecule Pages
MIR:00000047	Rat Genome Database	rat	^\d{4,}$	Rat Genome Database seeks to collect, consolidate, and integrate rat genomic and genetic data with curated functional and physiological data and make these data widely available to the scientific community. This collection references genes.	rgd	http://identifiers.org/rgd	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	RGD
								MIR:00100075	http://rgd.mcw.edu/rgdweb/report/gene/main.html?id={$id}	Rat Genome Database	Medical College of Wisconsin	USA	true	2018	ATP citrate lyase	http://rgd.mcw.edu/		
								MIR:00100905	https://www.alliancegenome.org/gene/RGD:{$id}	RGD through the Alliance of Genome Resources	Alliance of Genome Resources	USA	false	2018		https://www.alliancegenome.org	agr	
MIR:00000048	TAIR Protein	Arabidopsis	^AASequence:\d{10}$	The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant Arabidopsis thaliana. This provides protein information for a given gene model and provides links to other	tair.protein	http://identifiers.org/tair.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100076	http://arabidopsis.org/servlets/TairObject?accession={$id}	The Arabidopsis Information Resource (TAIR) Protein	Carnegie Institution of Washington Department of Plant Biology and National Center for Genome Resources (NCGR)	USA	false	AASequence:1009107926	AASequence:1009107926	http://arabidopsis.org/index.jsp		
MIR:00000049	TAIR Gene	Arabidopsis	^Gene:\d{7}$	The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant Arabidopsis thaliana. This is the reference gene model for a given locus.	tair.gene	http://identifiers.org/tair.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100077	http://arabidopsis.org/servlets/TairObject?accession={$id}	The Arabidopsis Information Resource (TAIR) Gene	Carnegie Institution of Washington Department of Plant Biology and National Center for Genome Resources (NCGR)	USA	false	Gene:2200934	NAC domain containing protein 1	http://arabidopsis.org/index.jsp		
MIR:00000050	TAIR Locus	Arabidopsis	^\d{7}$	The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant Arabidopsis thaliana. The name of a Locus is unique and used by TAIR, TIGR, and MIPS.	tair.locus	http://identifiers.org/tair.locus	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100078	http://www.arabidopsis.org/servlets/TairObject?accession=Locus:{$id}	The Arabidopsis Information Resource (TAIR) Locus	Carnegie Institution of Washington Department of Plant Biology and National Center for Genome Resources (NCGR)	USA	false	2200950	NGATHA3	http://arabidopsis.org/index.jsp		

MIR:00000054	GEO	All species	^G(PL|SM|SE|DS)\d+$	The Gene Expression Omnibus (GEO) is a gene expression repository providing a curated, online resource for gene expression data browsing, query and retrieval.	geo	http://identifiers.org/geo	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Gene Expression Omnibus
								MIR:00100082	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc={$id}	Gene Expression Omnibus at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	GDS1234	Fibroblast growth factor receptor 3 K650E mutation for type II thanatophoric dysplasia	https://www.ncbi.nlm.nih.gov/geo/	
MIR:00000057	Saccharomyces genome database pathways	Yeast	^PWY\w{2}\-\d{3}$	Curated biochemical pathways for Saccharomyces cerevisiae at Saccharomyces genome database (SGD).	sgd.pathways	http://identifiers.org/sgd.pathways	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100085	http://pathway.yeastgenome.org/YEAST/new-image?type=PATHWAY&object={$id}	Saccharomyces cerevisiae pathways at SGD.	SGD, Stanford University	USA	false	PWY3O-214	tryptophan degradation	http://pathway.yeastgenome.org/		
MIR:00000058	BioGRID	Yeast	^\d+$	BioGRID is a database of physical and genetic interactions in Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster, Homo sapiens, and Schizosaccharomyces pombe.	biogrid	http://identifiers.org/biogrid	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100086	http://thebiogrid.org/{$id}	BioGRID database of physical and genetic interactions	Ontario Cancer Institute	Canada	false	31623	YLR362W	http://thebiogrid.org/		
MIR:00000060	PANTHER Family	All species	^PTHR\d{5}(\:SF\d{1,3})?$	The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System is a resource that classifies genes by their functions, using published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct	panther.family	http://identifiers.org/panther.family	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100088	http://www.pantherdb.org/panther/family.do?clsAccession={$id}	PANTHER Family at USC (Los Angeles)	Keck School of Medicine, University of Southern California	USA	false	PTHR12345	membrane trafficking regulatory protein	http://www.pantherdb.org/		
MIR:00000061	PRINTS	All species	^PR\d{5}$	PRINTS is a compendium of protein fingerprints. A fingerprint is a group of conserved motifs used to characterise a protein family; its diagnostic power is refined by iterative scanning of a SWISS-PROT/TrEMBL composite. Usually the motifs do not overlap, but are separated	prints	http://identifiers.org/prints	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		SPRINT
								MIR:00100089	http://www.bioinf.manchester.ac.uk/cgi-bin/dbbrowser/sprint/searchprintss.cgi?prints_accn={$id}&display_opts=Prints&category=None&queryform=false&regexpr=off	PRINTS through SPRINT	University of Manchester	United Kingdom	false	PR00001	Coagulation factor GLA domain signature	http://www.bioinf.manchester.ac.uk/dbbrowser/sprint/		
MIR:00000069	NCBI Gene	All species	^\d+$	Entrez Gene is the NCBI's database for gene-specific information, focusing on completely sequenced genomes, those with an active research community to contribute gene-specific information, or those that are scheduled for intense sequence analysis.	ncbigene	http://identifiers.org/ncbigene	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Entrez Gene
								MIR:00100099	https://www.ncbi.nlm.nih.gov/gene/{$id}	Entrez Gene (NCBI)	National Center for Biotechnology Information (NCBI)	USA	true	ncbi	100010	expressed sequence AA589556	https://www.ncbi.nlm.nih.gov/gene	
								MIR:00100693	http://ncbigene.bio2rdf.org/describe/?url=http://bio2rdf.org/ncbigene:{$id}	Bio2RDF	Bio2RDF.org		false		100010	expressed sequence AA589556	http://ncbigene.bio2rdf.org/fct	
MIR:00000070	KEGG Genes	All species	^\w+:[\w\d\.-]*$	KEGG GENES is a collection of gene catalogs for all complete genomes and some partial genomes, generated from publicly available resources.	kegg.genes	http://identifiers.org/kegg.genes	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		KEGG
								MIR:00100100	http://www.kegg.jp/entry/{$id}	KEGG GENES Database	Kyoto University Bioinformatics Center	Japan	true	syn:ssr3451	cytochrome b559 subunit alpha	http://www.genome.jp/kegg/genes.html		
MIR:00000071	BRENDA	All species	^((\d+\.-\.-\.-)|(\d+\.\d+\.-\.-)|(\d+\.\d+\.\d+\.-)|(\d+\.\d+\.\d+\.\d+))$	BRENDA is a collection of enzyme functional data available to the scientific community. Data on enzyme function are extracted directly from the primary literature The database covers information on classification and nomenclature, reaction and specificity	brenda	http://identifiers.org/brenda	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100101	http://www.brenda-enzymes.org/php/result_flat.php4?ecno={$id}	Brenda enzyme database	Technical University Braunschweig, Institute for Bioinformatics and Biochemistry	Germany	false	1.1.1.1	alcohol dehydrogenase	http://www.brenda-enzymes.org/		
MIR:00000073	Pathway Commons	human	^\d+$	Pathway Commons is a convenient point of access to biological pathway information collected from public pathway databases, which you can browse or search. It is a collection of publicly available pathways from multiple organisms that provides researchers with	pathwaycommons	http://identifiers.org/pathwaycommons	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100103	http://www.pathwaycommons.org/pc/record2.do?id={$id}	Pathway Commons	Memorial Sloan-Kettering Cancer Center	USA	false	485991	Meiotic Recombination	http://www.pathwaycommons.org/pc/		
MIR:00000074	HOVERGEN	vertebrate	^HBG\d+$	HOVERGEN is a database of homologous vertebrate genes that allows one to select sets of homologous genes among vertebrate species, and to visualize multiple alignments and phylogenetic trees.	hovergen	http://identifiers.org/hovergen	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100104	http://pbil.univ-lyon1.fr/cgi-bin/view-tree.pl?query={$id}&db=HOVERGEN	Hovergen	Laboratoire de Biométrie, Génétique et Biologie des Populations	France	false	HBG004341	HBG004341 gene family	http://pbil.univ-lyon1.fr/databases/hovergen.php		
MIR:00000075	Melanoma Molecular Map Project Biomaps	Human	^\d+$	A collection of molecular interaction maps and pathways involved in cancer development and progression with a focus on melanoma.	mmmp:biomaps	http://identifiers.org/mmmp:biomaps	0	id	accessURL	info	institution	official	localId	testString	resourceURL			MMMP Biomaps
								MIR:00100105	http://www.mmmp.org/MMMP/public/biomap/viewBiomap.mmmp?id={$id}	main webpage	Melanoma Molecular Map Project	false	37	Upstream of Akt is PI3K (phosphatidyl inositol 3-kinase)	http://www.mmmp.org/MMMP/public/biomap/listBiomap.mmmp			
MIR:00000078	miRBase Sequence	All species	MI\d{7}	The miRBase Sequence Database is a searchable database of published miRNA sequences and annotation. The data were previously provided by the miRNA Registry. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA transcript	mirbase	http://identifiers.org/mirbase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100135	http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc={$id}	miRBase Sequence Database	Faculty of Life Sciences, University of Manchester	UK	false	MI0000001	Caenorhabditis elegans let-7 stem-loop	http://www.mirbase.org/		
MIR:00000079	ZFIN Bioentity	zebrafish	^ZDB\-\w+\-\d+\-\d+$	ZFIN serves as the zebrafish model organism database. This collection references all zebrafish biological entities in ZFIN.	zfin	http://identifiers.org/zfin	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	
								MIR:00100110	http://zfin.org/{$id}	ZFIN at University of Oregon	Zebrafish Information Network, University of Oregon	USA	true	ZDB-GENE-041118-11	pcdh2ab12	http://zfin.org		
								MIR:00100891	https://bioentity.link/#/lexicon/public/{$id}	ZFIN through BioEntity Link	InSilico Inc, in Eugene, Oregon	USA	false	ZDB-GENE-041118-11		https://bioentity.link/	bioentitylink	
								MIR:00100902	https://test.alliancegenome.org/gene/ZFIN:{$id}	ZFIN through the Alliance of Genome Resources	Alliance of Genome Resources	USA	false	ZDB-GENE-041118-11		https://www.alliancegenome.org	agr	
MIR:00000080	HGNC	Human	^((HGNC|hgnc):)?\d{1,5}$	The HGNC (HUGO Gene Nomenclature Committee) provides an approved gene name and symbol (short-form abbreviation) for each known human gene. All approved symbols are stored in the HGNC database, and each symbol is unique. HGNC identifiers refer to records in the HGNC symbol database.	hgnc	http://identifiers.org/hgnc	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	HUGO Gene Nomenclature Committee
								MIR:00100111	https://www.genenames.org/cgi-bin/gene_symbol_report?hgnc_id={$id}	HUGO Genome Nomenclature Committee	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	2674	CCDS43842	https://www.genenames.org	
								MIR:00100698	http://hgnc.bio2rdf.org/describe/?url=http://bio2rdf.org/{$id}	Bio2RDF	Bio2RDF.org		false		hgnc:2674	Gene Symbol for DAPK1	http://hgnc.bio2rdf.org/fct	
MIR:00000081	Sequence Ontology	All species	^SO:\d{7}$	The Sequence Ontology (SO) is a structured controlled vocabulary for the parts of a genomic annotation. It provides a common set of terms and definitions to facilitate the exchange, analysis and management of genomic data.	so	http://identifiers.org/so	1	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	SO
								MIR:00100112	http://www.sequenceontology.org/miso/current_release/term/SO:{$id}	Sequence Ontology	Department of Molecular and Cellular Biology, University of California, Berkeley	USA	false	704	necessary to encode a functional transcript.	http://www.sequenceontology.org/		Sequence Types and Features
								MIR:00100241	http://purl.bioontology.org/ontology/SO/SO:{$id}	Sequence Ontology through BioPortal	National Center for Biomedical Ontology, Stanford	USA	false	704	necessary to encode a functional transcript.	http://bioportal.bioontology.org/ontologies/SO	bptl	
								MIR:00100628	https://www.ebi.ac.uk/ols/ontologies/so/terms?obo_id=SO:{$id}	Sequence Ontology through Ontology Lookup Service (OLS)	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	704	necessary to encode a functional transcript.	https://www.ebi.ac.uk/ols/ontologies/so	ols	
MIR:00000083	Unipathway	All species	^UPA\d{5}$	UniPathway is a manually curated resource of metabolic pathways for the UniProtKB/Swiss-Prot knowledgebase. It provides a structured controlled vocabulary to describe the role of a protein in a metabolic pathway.	unipathway	http://identifiers.org/unipathway	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100114	http://www.grenoble.prabi.fr/obiwarehouse/unipathway/upa?upid={$id}	UniPathway	Swiss Institute of Bioinformatics (SIB) and French National Institute for Research in Computer Science and Control	Switzerland	false	UPA00206	ethanol fermentation	http://www.grenoble.prabi.fr/obiwarehouse/unipathway		
MIR:00000089	PharmGKB Pathways	Human	^PA\d+$	The PharmGKB database is a central repository for genetic, genomic, molecular and cellular phenotype data and clinical information about people who have participated in pharmacogenomics research studies. The data includes, but is not limited to, clinical and basic	pharmgkb.pathways	http://identifiers.org/pharmgkb.pathways	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100120	http://www.pharmgkb.org/pathway/{$id}	PharmGKB at Stanford	Department of Genetics, School of Medicine, Stanford University, Stanford, California	USA	false	PA146123006	Codeine and Morphine	http://www.pharmgkb.org/		
MIR:00000097	BioSystems	All species	^\d+$	The NCBI BioSystems database centralizes and cross-links existing biological systems databases, increasing their utility and target audience by integrating their pathways and systems into NCBI resources.	biosystems	http://identifiers.org/biosystems	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100128	https://www.ncbi.nlm.nih.gov/biosystems/{$id}	BioSystems database at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	1	Aspartate undergoes a number of reactions that are not in any linear order.	https://www.ncbi.nlm.nih.gov/biosystems/	
MIR:00000100	CTD Gene	Human	^\d+$	The Comparative Toxicogenomics Database (CTD) presents scientifically reviewed and curated information on chemicals, relevant genes and proteins, and their interactions in vertebrates and invertebrates. It integrates sequence, reference, species, microarray, and	ctd.gene	http://identifiers.org/ctd.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100131	http://ctdbase.org/detail.go?type=gene&acc={$id}	Comparative Toxicogenomics Database (Gene)	The Mount Desert Island Biological Laboratory	USA	false	101	ADAM metallopeptidase domain	http://ctdbase.org/		
MIR:00000106	GeneDB	human patogen	^[\w\d\.-]*$	GeneDB is a genome database for prokaryotic and eukaryotic organisms and provides a portal through which data generated by the "Pathogen Genomics" group at the Wellcome Trust Sanger Institute and other collaborating sequencing centres can be accessed.	genedb	http://identifiers.org/genedb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100139	http://www.genedb.org/gene/{$id}	GeneDB at Sanger Institute	Pathogen Genomics, Sanger Institute and European Bioinformatics Institute	UK	false	Cj1536c	Cj1536c - GeneDB	http://www.genedb.org/		
MIR:00000109	Molecular Interactions Ontology	All species	^MI:\d{4}$	The Molecular Interactions (MI) ontology forms a structured controlled vocabulary for the annotation of experiments concerned with protein-protein interactions. MI is developed by the HUPO Proteomics Standards Initiative.	psimi	http://identifiers.org/psimi	1	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	MI
								MIR:00100142	https://www.ebi.ac.uk/ols/ontologies/mi/terms?obo_id=MI:{$id}	Molecular Interactions Ontology through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ols	308	permeabilising cell	https://www.ebi.ac.uk/ols/ontologies/mi	
MIR:00000116	KEGG Orthology	All species	^K\d+$	KEGG Orthology (KO) consists of manually defined, generalised ortholog groups that correspond to KEGG pathway nodes and BRITE hierarchy nodes in all organisms.	kegg.orthology	http://identifiers.org/kegg.orthology	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		KEGG
								MIR:00100149	http://www.kegg.jp/entry/{$id}	KEGG Orthology Database	Department of Computational Biology, University of Tokyo, Tokyo	Japan	true	K00001	alcohol dehydrogenase	http://www.genome.jp/kegg/ko.html		
MIR:00000118	SMART	All species	^SM\d{5}$	The Simple Modular Architecture Research Tool (SMART) is an online tool for the identification and annotation of protein domains, and the analysis of domain architectures.	smart	http://identifiers.org/smart	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100151	http://smart.embl-heidelberg.de/smart/do_annotation.pl?DOMAIN={$id}	SMART (Simple Modular Architecture Research Tool)	EMBL, Heidelberg	Germany	false	SM00015	Calmodulin-binding motif.	http://smart.embl-heidelberg.de/		

MIR:00000119	Conserved Domain Database	All species	^(cd)?\d{5}$	The Conserved Domain Database (CDD) is a collection of multiple sequence alignments and derived database search models, which represent protein domains conserved in molecular evolution.	cdd	http://identifiers.org/cdd	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	CDD
								MIR:00100152	https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid={$id}	Conserved Domain Database at NCBI	National Library of Medicine, National Institutes of Health, Maryland	USA	false	ncbi	cd00400	The ClC chloride channels	https://www.ncbi.nlm.nih.gov/sites/entrez?db=cdd	
MIR:00000120	Relation Ontology	Human	^RO_\d+{7}$	The OBO Relation Ontology provides consistent and unambiguous formal definitions of the relational expressions used in biomedical ontologies.	ro	http://identifiers.org/ro	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	
								MIR:00100764	http://purl.obolibrary.org/obo/{$id}	OBO Relation Ontology through OntoBee	University of Michigan Medical School (MI), Lawrence Berkeley National Laboratory (CA) and Science Commons (MA)	USA	false	RO_0002533	collection of like units arranged in a linear order	http://obofoundry.org/ontology/ro.html		
								MIR:00100767	https://www.ebi.ac.uk/ols/ontologies/ro/terms?short_form={$id}	OBO Relation Ontology through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	RO_0002533	collection of like units arranged in a linear order	https://www.ebi.ac.uk/ols/ontologies/ro	ols	
MIR:00000121	Molecular Modeling Database	All species	^\d{1,5}$	The Molecular Modeling Database (MMDB) is a database of experimentally determined structures obtained from the Protein Data Bank (PDB). Since structures are known for a large fraction of all protein families, structure homologs may facilitate inference of biological function.	mmdb	http://identifiers.org/mmdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		MMDB
								MIR:00100154	http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?uid={$id}	MMDB at NCBI	National Library of Medicine, National Institutes of Health, Maryland	USA	false	50885	Structural Basis of Aspirin	http://www.ncbi.nlm.nih.gov/sites/entrez?db=structure		Entrez Structure
MIR:00000124	Microbial Protein Interaction Database	Microbial organisms	^\d+$	The microbial protein interaction database (MPIDB) provides physical microbial interaction data. The interactions are manually curated from the literature or imported from other databases, and are linked to supporting experimental evidence, as well as evidences based on interaction conservation, protein complex membership, and 3D domain contacts.	mpid	http://identifiers.org/mpid	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	MPID
								MIR:00100157	http://www.jcvi.org/mpidb/experiment.php?interaction_id={$id}	Microbial Protein Interaction Database	J. Craig Venter Institute, Maryland	USA	false	172	Thermotoga maritima	http://www.jcvi.org/mpidb/about.php		
								MIR:00100655	https://www.ebi.ac.uk/intact/query/interaction_id:MPIDB-INT-{$id}	Microbial Protein Interaction Database subset through IntAct	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	1776	spoke expanded co-complexes	https://www.ebi.ac.uk/intact/	ebi	
MIR:00000127	Ontology for Biomedical Investigations	Human	(^OBI:\d{7}$)\|(^OBI_\d{7}$)	The Ontology for Biomedical Investigations (OBI) project is developing an integrated ontology for the description of biological and clinical investigations. The ontology will represent the design of an investigation, the protocols and instrumentation used, the material used, the data generated and the type analysis performed on it. Currently OBI is being built under the Basic Formal Ontology (BFO).	obi	http://identifiers.org/obi	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	OBI
								MIR:00100162	http://purl.obolibrary.org/obo/{$id}	OBI through Ontobee	University of Michigan Medical School (MI), awrence Berkeley National Laboratory (CA) and Science Commons (MA)	USA	true	OBI_0000070	A planned process with the objective to produce	http://www.ontobee.org/		
								MIR:00100177	http://purl.bioontology.org/ontology/OBI/{$id}	Ontology for Biomedical Investigations through Bioportal	National Center for Biomedical Ontology, Stanford	USA	false	OBI:0000070	scientific observation	http://bioportal.bioontology.org/ontologies/OBI	bptl	
								MIR:00100760	https://www.ebi.ac.uk/ols/ontologies/obi/terms?obo_id={$id}	OBI through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	OBI:0000070	A planned process with the objective to produce	https://www.ebi.ac.uk/ols/ontologies/obi	ols	
MIR:00000132	SubtiWiki	Bacillus subtilis	^BSU\d{5}$	SubtiWiki is a scientific wiki for the model bacterium Bacillus subtilis. It provides comprehensive information on all genes and their proteins and RNA products, as well as information related to the current investigation of the gene/protein. Note: Currently, direct access to RNA products is	subtiwiki	http://identifiers.org/subtiwiki	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100172	http://www.subtiwiki.uni-goettingen.de/wiki/index.php/{$id}	SubtiWiki website	University of Göttingen, Department for General Microbiology	Germany	false	BSU29180	pyruvate kinase, glycolytic enzyme	http://www.subtiwiki.uni-goettingen.de/wiki/index.php/Main_Page		
MIR:00000133	NCI Pathway Interaction Database: Pathway	Human	^\w+$	The Pathway Interaction Database is a highly-structured, curated collection of information about known human biomolecular interactions and key cellular processes assembled into signaling pathways. This datatype provides access to pathway information.	pid.pathway	http://identifiers.org/pid.pathway	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		NCI_Nature
								MIR:00100173	http://pid.nci.nih.gov/search/pathway_landing.shtml?what=graphic&jpg=on&pathway_id={$id}	NCI Pathway Interaction Database (Pathway)	National Cancer Institute, Rockville, Maryland	USA	false	pi3kcipathway	Class I PI3K signaling events	http://pid.nci.nih.gov/		PID
																		Pathway Interaction Database NCI-Nature Curated Data
																		NCI-Nature PID
																		NCI-Nature Pathway Interaction Database
																		NCI_Nature curated
MIR:00000141	Protein Ontology	All species	^PR\:P?\d+$	The PRotein Ontology (PRO) has been designed to describe the relationships of proteins and protein evolutionary classes, to delineate the multiple protein forms of a gene locus (ontology for protein forms), and to interconnect existing ontologies.	pr	http://identifiers.org/pr	1	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	
								MIR:00100184	http://pir.georgetown.edu/cgi-bin/pro/entry_pro?id=PR:{$id}	Protein Ontology at Georgetown	Georgetown University Medical Center, Washington	USA	false	24	rho-associated protein kinase	http://pir.georgetown.edu/pro/pro.shtml		
								MIR:00100240	http://purl.bioontology.org/ontology/PR/PR:{$id}	Protein Ontology through BioPortal	National Center for Biomedical Ontology, Stanford	USA	false	24	rho-associated protein kinase	http://bioportal.bioontology.org/ontologies/PR	bptl	
								MIR:00100755	https://www.ebi.ac.uk/ols/ontologies/pr/terms?obo_id=PR:{$id}	Protein Ontology through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	24	rho-associated protein kinase	https://www.ebi.ac.uk/ols/ontologies/pr	ols	
MIR:00000142	OMIA	Animal	^\d+$	Online Mendelian Inheritance in Animals is a a database of genes, inherited disorders and traits in animal species (other than human and mouse).	omia	http://identifiers.org/omia	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Online Mendelian Inheritance in Animals
								MIR:00100185	http://omia.angis.org.au/{$id}/	OMIA at University of Sydney	Reprogen, Faculty of Veterinary Science, University of Sydney	Australia	false	1000	Coat colour, albinism	http://omia.angis.org.au/		
MIR:00000145	Candida Genome Database	candida	^CAL\d{7}$	The Candida Genome Database (CGD) provides access to genomic sequence data and manually curated functional information about genes and proteins of the human pathogen Candida albicans. It collects gene names and aliases, and assigns gene ontology terms to describe	cgd	http://identifiers.org/cgd	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		CGD
								MIR:00100189	http://www.candidagenome.org/cgi-bin/locus.pl?dbid={$id}	Candida Genome Database at Stanford University	Stanford University,	USA	false	CAL0003079	Heat Shock Protein	http://www.candidagenome.org/		
MIR:00000148	AmoebaDB	Eukaryotic pathogens	^EDI_\d+$	AmoebaDB is one of the databases that can be accessed through the EuPathDB (http://EuPathDB.org; formerly ApiDB) portal, covering eukaryotic pathogens of the genera Cryptosporidium, Giardia, Leishmania, Neospora, Plasmodium, Toxoplasma, Trichomonas and	amoebadb	http://identifiers.org/amoebadb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100193	http://amoebadb.org/amoeba/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id={$id}	AmoebaDB at EuPathDB	Center for Tropical & Emerging Global Diseases, University of Georgia	USA	false	EDI_244000	phosphatidylcholine-sterol acyltransferase precursor, putative	http://amoebadb.org/amoeba/		
MIR:00000149	CryptoDB	Eukaryotic pathogens	^\w+	CryptoDB is one of the databases that can be accessed through the EuPathDB (http://EuPathDB.org; formerly ApiDB) portal, covering eukaryotic pathogens of the genera Cryptosporidium, Giardia, Leishmania, Neospora, Plasmodium, Toxoplasma, Trichomonas and	cryptodb	http://identifiers.org/cryptodb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100194	http://cryptodb.org/cryptodb/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id={$id}	CryptoDB at EuPathDB	Center for Tropical & Emerging Global Diseases, University of Georgia	USA	false	cgd7_230	Cryptosporidium parvum Iowa II	http://cryptodb.org/cryptodb/		
MIR:00000150	PlasmoDB	Eukaryotic pathogens	^\w+$	AmoebaDB is one of the databases that can be accessed through the EuPathDB (http://EuPathDB.org; formerly ApiDB) portal, covering eukaryotic pathogens of the genera Cryptosporidium, Giardia, Leishmania, Neospora, Plasmodium, Toxoplasma, Trichomonas and	plasmodb	http://identifiers.org/plasmodb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100195	http://plasmodb.org/plasmo/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id={$id}	PlasmoDB at EuPathDB	Center for Tropical & Emerging Global Diseases, University of Georgia	USA	false	PF11_0344	apical membrane antigen 1	http://plasmodb.org/plasmo/		
MIR:00000151	GiardiaDB	Eukaryotic pathogens	^\w+$	GiardiaDB is one of the databases that can be accessed through the EuPathDB (http://EuPathDB.org; formerly ApiDB) portal, covering eukaryotic pathogens of the genera Cryptosporidium, Giardia, Leishmania, Neospora, Plasmodium, Toxoplasma, Trichomonas and	giardiadb	http://identifiers.org/giardiadb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100196	http://giardiadb.org/giardiadb/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id={$id}	GiardiaDB at EuPathDB	Center for Tropical & Emerging Global Diseases, University of Georgia	USA	false	GL50803_102438	Plasma membrane calcium-transporting ATPase 2	http://giardiadb.org/giardiadb/		
MIR:00000152	MicrosporidiaDB	Eukaryotic pathogens	^\w+$	MicrosporidiaDB is one of the databases that can be accessed through the EuPathDB (http://EuPathDB.org; formerly ApiDB) portal, covering eukaryotic pathogens of the genera Cryptosporidium, Giardia, Leishmania, Neospora, Plasmodium, Toxoplasma, Trichomonas	microsporidia	http://identifiers.org/microsporidia	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100197	http://microsporidiadb.org/micro/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id={$id}	MicrosporidiaDB at EuPathDB	Center for Tropical & Emerging Global Diseases, University of Georgia	USA	false	ECU03_0820i	TUBULIN BETA CHAIN	http://microsporidiadb.org/micro/		
MIR:00000153	ToxoDB	Eukaryotic pathogens	^\w+$	ToxoDB is one of the databases that can be accessed through the EuPathDB (http://EuPathDB.org; formerly ApiDB) portal, covering eukaryotic pathogens of the genera Cryptosporidium, Giardia, Leishmania, Neospora, Plasmodium, Toxoplasma, Trichomonas and	toxoplasma	http://identifiers.org/toxoplasma	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100198	http://toxodb.org/toxo/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id={$id}	ToxoDB at EuPathDB	Center for Tropical & Emerging Global Diseases, University of Georgia	USA	false	TGME49_053730	importin-alpha re-exporter, putative	http://toxodb.org/toxo/		
MIR:00000154	TrichDB	Eukaryotic pathogens	^\w+$	TrichDB is one of the databases that can be accessed through the EuPathDB (http://EuPathDB.org; formerly ApiDB) portal, covering eukaryotic pathogens of the genera Cryptosporidium, Giardia, Leishmania, Neospora, Plasmodium, Toxoplasma, Trichomonas and	trichdb	http://identifiers.org/trichdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100199	http://trichdb.org/trichdb/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id={$id}	TrichDB at EuPathDB	Center for Tropical & Emerging Global Diseases, University of Georgia	USA	false	TVAG_386080	Clan MG, familly M24, aminopeptidase P-like metallopeptidase	http://trichdb.org/trichdb/		
MIR:00000155	TriTrypDB	Eukaryotic pathogens	^\w+(\.)?\w+(\.)?\w+	TriTrypDB is one of the databases that can be accessed through the EuPathDB (http://EuPathDB.org; formerly ApiDB) portal, covering eukaryotic pathogens of the genera Cryptosporidium, Giardia, Leishmania, Neospora, Plasmodium, Toxoplasma, Trichomonas and	tritrypdb	http://identifiers.org/tritrypdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Kinetoplastid Genomics Resource
								MIR:00100200	http://tritrypdb.org/tritrypdb/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id={$id}	TriTrypDB at EuPathDB	Wellcome Trust Sanger Institute, Hinxton	UK	false	Tb927.8.620	Trypanosoma brucei TREU927	http://tritrypdb.org/tritrypdb/		
MIR:00000156	BDGP insertion DB	Drosophila	^\w+$	BDGP gene disruption collection provides a public resource of gene disruptions of Drosophila genes using a single transposable element.	bdgp.insertion	http://identifiers.org/bdgp.insertion	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		BDGP gene disruption
								MIR:00100201	http://flypush.imgen.bcm.tmc.edu/pscreen/details.php?line={$id}	BDGP Gene Disruption Project	Baylor College of Medicine, Texas	USA	false	KG09531	5634250	http://flypush.imgen.bcm.tmc.edu/pscreen/		
MIR:00000157	BeetleBase	Tribolium	^TC\d+$	BeetleBase is a comprehensive sequence database and community resource for Tribolium genetics, genomics and developmental biology. It incorporates information about genes, mutants, genetic markers, expressed sequence tags and publications.	beetlebase	http://identifiers.org/beetlebase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100202	http://beetlebase.org/cgi-bin/gbrowse/BeetleBase3.gff3/?name={$id}	BeetleBase at Kansas State University	Bioinformatics Center, Kansas State University	USA	false	TC010103	positions 17,144,898	http://beetlebase.org/		
MIR:00000158	BOLD Taxonomy	All species	^\d+$	The Barcode of Life Data System (BOLD) is an informatics workbench aiding the acquisition, storage, analysis and publication of DNA barcode records. The associated taxonomy browser shows the progress of DNA barcoding and provides sample collection site distribution, and	bold.taxonomy	http://identifiers.org/bold.taxonomy	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100203	http://www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid={$id}	BOLD taxonomy browser	Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario	Canada	false	27267	Doliolida	http://www.boldsystems.org/		

MIR:00000159	dbEST	All species	^([A-Z]+)?\d+(\.\d+)?$	The dbEST contains sequence data and other information on "single-pass" cDNA sequences, or "Expressed Sequence Tags", from a number of organisms.	dbest	http://identifiers.org/dbest	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Database of Expressed Sequence Tags
								MIR:00100204	https://www.ncbi.nlm.nih.gov/nucest/{$id}	dbEST at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	BP100000	BP100000 Chaetomium cDNA	https://www.ncbi.nlm.nih.gov/nucest	NCBI EST
								MIR:00100331	https://www.ebi.ac.uk/ena/data/view/{$id}	dbEST through European Nucleotide Archive (ENA)	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	BP100000.1	Chaetomium globosum	https://www.ebi.ac.uk/ena	
								MIR:00100651	http://getentry.ddbj.nig.ac.jp/getentry/na/{$id}	dbEST through DNA Data Bank of Japan (DDBJ)	DNA Data Bank of Japan, Mishima, Shizuoka	Japan	false		BP100000	Chaetomium globosum mRNA	http://www.ddbj.nig.ac.jp/	
MIR:00000160	dbProbe	Human	^\d+$	The NCBI Probe Database is a public registry of nucleic acid reagents designed for use in a wide variety of biomedical research applications, together with information on reagent distributors, probe effectiveness, and computed sequence similarities.	dbprobe	http://identifiers.org/dbprobe	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100205	https://www.ncbi.nlm.nih.gov/probe/?term={$id}	dbProbe at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	1000000	Homo sapiens bead probe	https://www.ncbi.nlm.nih.gov/sites/entrez?db=probe	
MIR:00000161	dbSNP	Human	^rs\d+$	The dbSNP database is a repository for both single base nucleotide subsitutions and short deletion and insertion polymorphisms.	dbsnp	http://identifiers.org/dbsnp	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100206	https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs={$id}	dbSNP at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	121909098	Cluster Report: rs121909098	https://www.ncbi.nlm.nih.gov/sites/entrez?db=snp	
MIR:00000162	UniSTS	human and mouse	^\d+$	UniSTS is a comprehensive database of sequence tagged sites (STSs) derived from STS-based maps and other experiments. STSs are defined by PCR primer pairs and are associated with additional information such as genomic position, genes and sequences.	unists	http://identifiers.org/unists	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	dbSTS
								MIR:00100207	https://www.ncbi.nlm.nih.gov/genome/sts/sts.cgi?uid={$id}	UniSTS at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	456789	voltage-gated channel	https://www.ncbi.nlm.nih.gov/sites/entrez?db=unists	
MIR:00000163	EcoGene	E.coli	^EG\d+$	The EcoGene database contains updated information about the E. coli K-12 genome and proteome sequences, including extensive gene bibliographies. A major EcoGene focus has been the re-evaluation of translation start sites.	ecogene	http://identifiers.org/ecogene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100208	http://www.ecogene.org/gene/{$id}	EcoGene at University of Miami	Miller School of Medicine, University of Miami	USA	true	EG10173	Cytochrome d	http://ecogene.org/		
MIR:00000164	GABI	Plant species	^\w+$	GabiPD (Genome Analysis of Plant Biological Systems Primary Database) constitutes a repository for a wide array of heterogeneous data from high-throughput experiments in several plant species. These data (i.e. genomics, transcriptomics, proteomics and	gabi	http://identifiers.org/gabi	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		GABI PD
								MIR:00100210	http://www.gabipd.org/database/cgi-bin/GreenCards.pl.cgi?BioObjectId={$id}&Mode=ShowBioObject	GABI at Max Planck	Max Planck Institute of Molecular Plant Physiology, Berlin	Germany	true	2679240	AT1G06390	http://www.gabipd.org/		
MIR:00000165	GreenGenes	All species	^\d+$	A 16S rRNA gene database which provides chimera screening, standard alignment, and taxonomic classification using multiple published taxonomies.	greengenes	http://identifiers.org/greengenes	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100211	http://greengenes.lbl.gov/cgi-bin/show_one_record_v2.pl?prokMSA_id={$id}	GreenGenes at Lawrence Berkeley	Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, California	USA	false	100000	bellerophon	http://greengenes.lbl.gov/		
MIR:00000167	H-InvDb Locus	Human	^HIX\d{7}(\.\d+)?$	H-Invitational Database (H-InvDB) is an integrated database of human genes and transcripts. It provides curated annotations of human genes and transcripts including gene structures, alternative splicing isoforms, non-coding functional RNAs, protein functions	hinv.locus	http://identifiers.org/hinv.locus	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100213	http://h-invitational.jp/hinv/spsoup/locus_view?hix_id={$id}	H-Invitational Locus View at BIRC	BIRC, Tokyo	Japan	false	HIX0004394	Alcohol dehydrogenase class-3	http://h-invitational.jp/hinv/ahg-db/index.jsp		
MIR:00000168	H-InvDb Transcript	Human	^HIT\d{9}(\.\d+)?$	H-Invitational Database (H-InvDB) is an integrated database of human genes and transcripts. It provides curated annotations of human genes and transcripts including gene structures, alternative splicing isoforms, non-coding functional RNAs, protein functions	hinv.transcript	http://identifiers.org/hinv.transcript	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100214	http://h-invitational.jp/hinv/spsoup/transcript_view?hit_id={$id}	H-Invitational Transcript View at BIRC	BIRC, Tokyo	Japan	false	HIT000195363	Alcohol dehydrogenase class-3	http://h-invitational.jp/hinv/ahg-db/index.jsp		
MIR:00000169	H-InvDb Protein	Human	^HIP\d{9}(\.\d+)?$	H-Invitational Database (H-InvDB) is an integrated database of human genes and transcripts. It provides curated annotations of human genes and transcripts including gene structures, alternative splicing isoforms, non-coding functional RNAs, protein functions	hinv.protein	http://identifiers.org/hinv.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100215	http://h-invitational.jp/hinv/protein/protein_view.cgi?hip_id={$id}	H-Invitational Protein View at BIRC	BIRC, Tokyo	Japan	false	HIP000030660	Alcohol dehydrogenase class-3	http://h-invitational.jp/hinv/ahg-db/index.jsp		
MIR:00000170	HOMD Sequence Metainformation	Human	^SEQF\d+$	The Human Oral Microbiome Database (HOMD) provides a site-specific comprehensive database for the more than 600 prokaryote species that are present in the human oral cavity. It contains genomic information based on a curated 16S rRNA gene-based provisional	homd.seq	http://identifiers.org/homd.seq	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Human Oral Microbiome Database
								MIR:00100216	http://www.homd.org/modules.php?op=modload&name=GenomeList&file=index&link=detailinfo&seqid={$id}	HOMD sequence metainformation at The Forsyth Institute	The Forsyth Institute, Boston	USA	false	SEQF1003	ATCC 49626	http://www.homd.org/index.php		
MIR:00000172	IRD Segment Sequence	Influenza	^\w+(_)?\d+(\.\d+)?$	Influenza Research Database (IRD) contains information related to influenza virus, including genomic sequence, strain, protein, epitope and bibliographic information. The Segment Details page contains descriptive information and annotation data about a particular	ird.segment	http://identifiers.org/ird.segment	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Influenza Research Database Segment
								MIR:00100218	http://www.fludb.org/brc/fluSegmentDetails.do?ncbiGenomicAccession={$id}	IRD at BioHealthBase	Southwestern Medical Center, University of Texas, Dallas	USA	false	CY077097	Influenza Segment/Protein Details	http://www.fludb.org/		
MIR:00000173	ISFinder	bacteria	^IS\w+(\-\d)?$	ISfinder is a database of bacterial insertion sequences (IS). It assigns IS nomenclature and acts as a repository for ISs. Each IS is annotated with information such as the open reading frame DNA sequence, the sequence of the ends of the element and target sites, its origin and	isfinder	http://identifiers.org/isfinder	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Insertion Sequence Finder
								MIR:00100219	https://www-is.biotoul.fr/scripts/ficheIS.php?name={$id}	IS Finder at CNRS	Laboratoire de Microbiologie et Génétique Moléculaires, CNRS, Toulouse	France	false	ISA1083-2	Archaeoglobus fulgidus	http://www-is.biotoul.fr/i		
MIR:00000175	Integrated Microbial Genomes Taxon	Microbial organisms	^\d+$	The Integrated Microbial Genomes (IMG) system is a data management, analysis and annotation platform for all publicly available genomes. IMG contains both draft and complete JGI (DoE Joint Genome Institute) microbial genomes integrated with all other publicly available genomes	img.taxon	http://identifiers.org/img.taxon	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		IMG Taxon
								MIR:00100221	http://img.jgi.doe.gov/cgi-bin/w/main.cgi?section=TaxonDetail&taxon_oid={$id}	Integrated Microbial Genomes Taxon at Lawrence Berkeley National Laboratory	Lawrence Berkeley National Laboratory, Berkeley	USA	false	648028003	Acidilobus saccharovorans 345-15	http://img.jgi.doe.gov/		
MIR:00000176	Integrated Microbial Genomes Gene	Microbial organisms	^\d+$	The Integrated Microbial Genomes (IMG) system is a data management, analysis and annotation platform for all publicly available genomes. IMG contains both draft and complete JGI (DoE Joint Genome Institute) microbial genomes integrated with all other publicly available genomes	img.gene	http://identifiers.org/img.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100222	http://img.jgi.doe.gov/cgi-bin/w/main.cgi?section=GeneDetail&gene_oid={$id}	Integrated Microbial Genomes Taxon at Lawrence Berkeley National Laboratory	Lawrence Berkeley National Laboratory, Berkeley	USA	false	638309541	polynucleotide phosphatase/kinase	http://img.jgi.doe.gov/		
MIR:00000177	MaizeGDB Locus	Maize	^\d+$	MaizeGDB is the maize research community's central repository for genetics and genomics information.	maizegdb.locus	http://identifiers.org/maizegdb.locus	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100223	http://www.maizegdb.org/cgi-bin/displaylocusrecord.cgi?id={$id}	MaizeGDB at Missouri	Division of Plant Sciences, Missouri	USA	false	25011	aldolase1	http://www.maizegdb.org/		
MIR:00000179	NITE Biological Research Center Catalogue	Microbial organisms	^\d+$	NITE Biological Research Center (NBRC) provides a collection of microbial resources, performing taxonomic characterization of individual microorganisms such as bacteria including actinomycetes and archaea, yeasts, fungi, algae, bacteriophages and DNA resources	nbrc	http://identifiers.org/nbrc	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		NBRC Catalogue
								MIR:00100225	http://www.nbrc.nite.go.jp/NBRC2/NBRCCatalogueDetailServlet?ID=NBRC&CAT={$id}	NBRC Catalogue at National Institute of Technology Evaluation	National Institute of Technology Evaluation, Tokyo	Japan	false	1234	NBRC 1234	http://www.nbrc.nite.go.jp/e/index.html		
MIR:00000180	Pseudomonas Genome Database	Pseudomonas	^P\w+$	The Pseudomonas Genome Database is a resource for peer-reviewed, continually updated annotation for all Pseudomonas species. It includes gene and protein sequence information, as well as regulation and predicted function and annotation.	pseudomonas	http://identifiers.org/pseudomonas	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		PGD
								MIR:00100226	http://v2.pseudomonas.com/getAnnotation.do?locusID={$id}	Pseudomonas Genome Database at Simon Fraser University	Simon Fraser University, British Columbia	Canada	false	PSEEN0001	dnaA	http://www.pseudomonas.com/		Pseudocap
MIR:00000181	Gramene protein	Gramineae	^\d+$	Gramene is a comparative genome mapping database for grasses and crop plants. It combines a semi-automatically generated database of cereal genomic and expressed sequence tag sequences, genetic maps, map relations, quantitative trait loci (QTL), and publications, with a	gramene.protein	http://identifiers.org/gramene.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100227	http://www.gramene.org/db/protein/protein_search?protein_id={$id}	Gramene Protein at Cold Spring Harbor	Cold Spring Harbor Laboratory, New York	USA	false	78073	WAXY	http://www.gramene.org/		
MIR:00000182	Gramene genes	Gramineae	^GR\:\d+$	Gramene is a comparative genome mapping database for grasses and crop plants. It combines a semi-automatically generated database of cereal genomic and expressed sequence tag sequences, genetic maps, map relations, quantitative trait loci (QTL), and publications, with a	gramene.gene	http://identifiers.org/gramene.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100228	http://www.gramene.org/db/genes/search_gene?acc={$id}	Gramene Genes at Cold Spring Harbor	Cold Spring Harbor Laboratory, New York	USA	false	GR:0080039	bZIP transcription factor	http://www.gramene.org/		
MIR:00000184	Gramene QTL	Gramineae	^\w+$	Gramene is a comparative genome mapping database for grasses and crop plants. It combines a semi-automatically generated database of cereal genomic and expressed sequence tag sequences, genetic maps, map relations, quantitative trait loci (QTL), and publications, with a	gramene.qtl	http://identifiers.org/gramene.qtl	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100230	http://www.gramene.org/db/qtl/qtl_display?qtl_accession_id={$id}	Gramene Quantitative Trait Loci at Cold Spring Harbor	Cold Spring Harbor Laboratory, New York	USA	false	CQG5	abaxial stomatal frequency	http://www.gramene.org/		
MIR:00000185	Sol Genomics Network	Solanaceae	^\d+$	The Sol Genomics Network (SGN) is a database and website dedicated to the genomic information of the nightshade family, which includes species such as tomato, potato, pepper, petunia and eggplant.	sgn	http://identifiers.org/sgn	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		SGN
								MIR:00100231	http://solgenomics.net/phenome/locus_display.pl?locus_id={$id}	Sol Genomics Network at Boyce Thompson Institute for Plant Research	Boyce Thompson Institute for Plant Research, New York	USA	false	1	Tomato locus	http://solgenomics.net/		
MIR:00000186	Xenbase	Xenopus	^(XB-GENE-)?\d+$	Xenbase is the model organism database for Xenopus laevis and X. (Silurana) tropicalis. It contains genomic, development data and community information for Xenopus research. it includes gene expression patterns that incorporates image data from the	xenbase	http://identifiers.org/xenbase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100232	http://www.xenbase.org/gene/showgene.do?method=displayGeneSummary&geneId={$id}	Xenbase at University of Calgary	Department of Biological Sciences, University of Calgary, Alberta	Canada	false	922462	protogenin	http://www.xenbase.org/		
MIR:00000187	BioPortal	All species	^\d+$	BioPortal is an open repository of biomedical ontologies that provides access via Web services and Web browsers to ontologies developed in OWL, RDF, OBO format and Protégé frames. BioPortal functionality includes the ability to browse, search and visualize	bioportal	http://identifiers.org/bioportal	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100244	http://bioportal.bioontology.org/ontologies/{$id}	BioPortal at National Center for Biomedical Ontology	National Center for Biomedical Ontology, Stanford University	USA	false	bptl	1046	Systems Biology Ontology	http://bioportal.bioontology.org/	
MIR:00000188	MIRIAM Registry resource	All species	^MIR:001\d{5}$	MIRIAM Registry is an online resource created to catalogue data types (Gene Ontology, Taxonomy or PubMed are some examples), their URIs and the corresponding resources (or physical locations), whether these are controlled vocabularies or databases.	miriam.resource	http://identifiers.org/miriam.resource	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100245	https://www.ebi.ac.uk/miriam/main/resources/{$id}	MIRIAM Registry (resources)	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	MIR:00100005	MIRIAM Resources (data collection)	https://www.ebi.ac.uk/miriam/	

MIR:00000189	EDAM Ontology	All species	^(data|topic|operation|format)_\d{4}$	EDAM is an ontology of general bioinformatics concepts, including topics, data types, formats, identifiers and operations. EDAM provides a controlled vocabulary for the description, in semantic terms, of things such as: web services (e.g. WSDL files), applications, tool collections and packages, workbenches and workflow software, databases and ontologies, XSD data	edam	http://identifiers.org/edam	0		id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	EMBRACE Data and Methods Ontology
									MIR:00100536	http://purl.bioontology.org/ontology/EDAM/{$id}	EDAM through BioPortal	National Center for Biomedical Ontology, Stanford	USA	false	bptl	data_1664	Minimal Information Requested in the Annotation of Biochemical Models	http://bioportal.bioontology.org/ontologies/EDAM	
									MIR:00100762	https://www.ebi.ac.uk/ols/ontologies/edam/terms?short_form={$id}	EDAM through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ols	data_1664	Minimal Information Requested in the Annotation of Biochemical Models	https://www.ebi.ac.uk/ols/ontologies/edam	
MIR:00000193	ArachnoServer	Spider	^AS\d{6}$	ArachnoServer (www.arachnoserver.org) is a manually curated database providing information on the sequence, structure and biological activity of protein toxins from spider venoms. It include a molecular target ontology designed specifically for venom toxins, as well as	arachnoserver	http://identifiers.org/arachnoserver	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100250	http://www.arachnoserver.org/toxincard.html?id={$id}	ArachnoServer at Brisbane	Institute for Molecular Bioscience, The University of Queensland, Brisbane	Australia	false	AS000060	Latrotoxin-Lt1a	http://www.arachnoserver.org/		
MIR:00000194	BioCyc	All species	^[A-Z-0-9]+(\:)?[A-Za-z0-9+_.%-:]+$	BioCyc is a collection of Pathway/Genome Databases (PGDBs) which provides an electronic reference source on the genomes and metabolic pathways of sequenced organisms.	biocyc	http://identifiers.org/biocyc	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100251	http://biocyc.org/getid?id={$id}	BioCyc at SRI International	Bioinformatics Research Group, SRI International, California	USA	false	ECOLI:CYT-D-UBIOX-CPLX	3 cytochrome oxidase enzymes	http://biocyc.org		
MIR:00000195	CAZy	All species	^(GT|GH|PL|CE|CBM)\d+(_\d+)?$	The Carbohydrate-Active Enzyme (CAZy) database is a resource specialized in enzymes that build and breakdown complex carbohydrates and glycoconjugates. These enzymes are classified into families based on structural features	cazy	http://identifiers.org/cazy	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		Carbohydrate-Active enzYmes
									MIR:00100252	http://www.cazy.org/{$id}.html	CAZy at CNRS	Architecture et Fonction des Macromolécules Biologiques, CNRS, Marseille	France	false	GT10	fucosyltransferase	http://www.cazy.org/		
MIR:00000196	GOA	All species	^(([A-N,R-Z][0-9][A-Z][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9])|([O,P,Q][0-9][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9]))|(URS[0-9A-F]{10}(_[0-9]+){0,1})|(EBI-[0-9]+)$	The GOA (Gene Ontology Annotation) project provides high-quality Gene Ontology (GO) annotations to proteins in the UniProt Knowledgebase (UniProtKB) and International Protein Index (IPI). This involves electronic annotation and the integration of high-quality manual GO	goa	http://identifiers.org/goa	0		id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Gene Ontology Annotation
									MIR:00100253	https://www.ebi.ac.uk/QuickGO/GProtein?ac={$id}	GOA through QuickGO	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	quickgo	P12345	Oryctolagus cuniculus	https://www.ebi.ac.uk/GOA/	UniProtKB-GOA
MIR:00000197	PaleoDB	Animal	^\d+$	The Paleobiology Database seeks to provide researchers and the public with information about the entire fossil record. It stores global, collection-based occurrence and taxonomic data for marine and terrestrial animals and plants of any geological age, as well as web	paleodb	http://identifiers.org/paleodb	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		Paleobiology database
									MIR:00100254	http://fossilworks.org/?a=taxonInfo&taxon_no={$id}	PaleoDB at Macquarie University	Macquarie University, Sydney	Australia	false	83088	ground dwelling omnivore	http://paleodb.org/		
MIR:00000200	EchoBASE	E.coli	^EB\d+$	EchoBASE is a database designed to contain and manipulate information from post-genomic experiments using the model bacterium Escherichia coli K-12. The database is built on an enhanced annotation of the updated genome sequence of strain MG1655 and the	echobase	http://identifiers.org/echobase	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100259	http://www.york.ac.uk/res/thomas/Gene.cfm?recordID={$id}	EchoBASE at the University of York	Department of Biology (Area 10), University of York, York	UK	false	EB0170	Electron acceptor	http://www.york.ac.uk/		
MIR:00000201	eggNOG	All species	^\w+$	eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) is a database of orthologous groups of genes. The orthologous groups are annotated with functional description lines (derived by identifying a common denominator for the genes based on their various	eggnog	http://identifiers.org/eggnog	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100260	http://eggnog.embl.de/version_3.0/cgi/search.py?search_term_0={$id}	eggNOG Version 3 at European Molecular Biology Laboratory	European Molecular Biology Laboratory, Heidelberg	Germany	false	veNOG12876	MAGI family member	http://eggnog.embl.de/version_3.0/		
MIR:00000202	Ensembl Bacteria	bacteria	^((EB\w+)|([A-Z0-9]+_[A-Z0-9]+))$	Ensembl Genomes consists of five sub-portals (for bacteria, protists, fungi, plants and invertebrate metazoa) designed to complement the availability of vertebrate genomes in Ensembl. This collection is concerned with bacterial genomes.	ensembl.bacteria	http://identifiers.org/ensembl.bacteria	0		id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Ensembl Genomes Bacteria
									MIR:00100261	https://bacteria.ensembl.org/id/{$id}	Ensembl Bacteria at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	MU9_3181	Cysteinyl-tRNA synthetase	https://bacteria.ensembl.org/	
MIR:00000203	Ensembl Protists	protists	^\w+$	Ensembl Genomes consists of five sub-portals (for bacteria, protists, fungi, plants and invertebrate metazoa) designed to complement the availability of vertebrate genomes in Ensembl. This collection is concerned with protist genomes.	ensembl.protist	http://identifiers.org/ensembl.protist	0		id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Ensembl Genomes Protists
									MIR:00100262	https://protists.ensembl.org/id/{$id}	Ensembl Protists at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	PFC0120w	Cytoadherence linked asexual	https://protists.ensembl.org	
MIR:00000204	Ensembl Metazoa	Metazoa	^\w+(\.)?\d+$	Ensembl Genomes consists of five sub-portals (for bacteria, protists, fungi, plants and invertebrate metazoa) designed to complement the availability of vertebrate genomes in Ensembl. This collection is concerned with metazoa genomes.	ensembl.metazoa	http://identifiers.org/ensembl.metazoa	0		id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Ensembl Genomes Metazoa
									MIR:00100263	https://metazoa.ensembl.org/id/{$id}	Enzembl Metazoa at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	FBtr0084214	Pyruvate kinase	https://metazoa.ensembl.org/	
MIR:00000205	Ensembl Plants	Plant species	^\w+(\.\d+)?(\.\d+)?$	Ensembl Genomes consists of five sub-portals (for bacteria, protists, fungi, plants and invertebrate metazoa) designed to complement the availability of vertebrate genomes in Ensembl. This collection is concerned with plant genomes.	ensembl.plant	http://identifiers.org/ensembl.plant	0		id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Ensembl Genomes Plants
									MIR:00100264	https://plants.ensembl.org/id/{$id}	Enzembl Plants at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	AT1G73965	CLAVATA3/ESR (CLE)-related protein	https://plants.ensembl.org/	
MIR:00000206	Ensembl Fungi	Fungi	^[A-Z-a-z0-9]+$	Ensembl Genomes consists of five sub-portals (for bacteria, protists, fungi, plants and invertebrate metazoa) designed to complement the availability of vertebrate genomes in Ensembl. This collection is concerned with fungal genomes.	ensembl.fungi	http://identifiers.org/ensembl.fungi	0		id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Ensembl Genomes Fungi
									MIR:00100265	https://fungi.ensembl.org/id/{$id}	Enzembl Fungi at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	CADAFLAT00006211	RING finger protein	https://fungi.ensembl.org/	
MIR:00000207	HCVDB	Human	^M\d{5}$	the European Hepatitis C Virus Database (euHCVdb, http://euhcvdb.ibcp.fr), a collection of computer-annotated sequences based on reference genomes.mainly dedicated to HCV protein sequences, 3D structures and functional analyses	hcvdb	http://identifiers.org/hcvdb	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		euHCVDB
									MIR:00100266	https://euhcvdb.ibcp.fr/euHCVdb/do/displayHCVEntry?primaryAC={$id}	HCVDB at CNRS	Institut de Biologie et Chimie des Protéines, CNRS, University of Lyon , Lyon-Gerland	France	false	M58335	Hepatitis C virus	http://euhcvdb.ibcp.fr/euHCVdb/		
MIR:00000208	Genatlas	Human	^\w+$	GenAtlas is a database containing information on human genes, markers and phenotypes.	genatlas	http://identifiers.org/genatlas	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100267	http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol={$id}	Genatlas at Paris Descartes University	Paris Descartes University, Paris	France	false	HBB	Heinz body anemia	http://genatlas.medecine.univ-paris5.fr/		
MIR:00000209	CATH superfamily	All species	^\d+(\.\d+(\.\d+(\.\d+)?)?)?$	The CATH database is a hierarchical domain classification of protein structures in the Protein Data Bank. Protein structures are classified using a combination of automated and manual procedures. There are four major levels in this hierarchy: Class (secondary structure	cath.superfamily	http://identifiers.org/cath.superfamily	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100268	http://www.cathdb.info/cathnode/{$id}	CATH superfamily at UCL	Institute of Structural and Molecular Biology, University College London	United Kingdom	false	1.10.10.200	1mw7A01	http://www.cathdb.info/		
MIR:00000210	CATH domain	All species	^\w+$	The CATH database is a hierarchical domain classification of protein structures in the Protein Data Bank. Protein structures are classified using a combination of automated and manual procedures. There are four major levels in this hierarchy: Class (secondary structure	cath.domain	http://identifiers.org/cath.domain	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100269	http://www.cathdb.info/domain/{$id}	CATH domain at UCL	Institute of Structural and Molecular Biology, University College London	UK	false	1cukA01	Nucleic acid-binding proteins	http://www.cathdb.info/		
MIR:00000211	GeneFarm	Arabidopsis	^\d+$	GeneFarm is a database whose purpose is to store traceable annotations for Arabidopsis nuclear genes and gene products.	genefarm	http://identifiers.org/genefarm	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100270	https://urgi.versailles.inra.fr/Genefarm/Gene/display_gene.htpl?GENE_ID={$id}	GeneFarm at Unité de Recherche en Génomique Végétale	Unité de Recherche en Génomique Végétale, Evry	France	false	4892	Gene 'ALDO1'	http://urgi.versailles.inra.fr/Genefarm/		
MIR:00000213	HOGENOM	All species	^\w+$	HOGENOM is a database of homologous genes from fully sequenced organisms (bacteria, archeae and eukarya). This collection references phylogenetic trees which can be retrieved using either UniProt accession numbers, or HOGENOM tree family identifier	hogenom	http://identifiers.org/hogenom	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100272	http://pbil.univ-lyon1.fr/cgi-bin/view-tree.pl?db=HOGENOM5&query={$id}	HOGENOM at CNRS	Laboratoire de Biométrie et Biologie Evolutive, CNRS, Lyon	France	false	HBG284870	Chlamydiales	http://pbil.univ-lyon1.fr/databases/hogenom/		
MIR:00000214	GeneTree	All species	^ENSGT\d+$	Genetree displays the maximum likelihood phylogenetic (protein) trees representing the evolutionary history of the genes. These are constructed using the canonical protein for every gene in Ensembl.	genetree	http://identifiers.org/genetree	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		Ensembl GeneTree
									MIR:00100273	http://www.ensembl.org/Multi/GeneTree/Image?db=core;gt={$id}	GeneTree at Ensembl	Sanger Institute and European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ENSGT00550000074763	101	http://www.ensembl.org/		
MIR:00000216	MycoBrowser tuberculosis	Mycobacterium	^Rv\d{4}(A|B|c)?$	Mycobrowser is a resource that provides both in silico generated and manually reviewed information within databases dedicated to the complete genomes of Mycobacterium tuberculosis, Mycobacterium leprae, Mycobacterium marinum and Mycobacterium smegmatis.	myco.tuber	http://identifiers.org/myco.tuber	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		TubercuList
									MIR:00100275	http://tuberculist.epfl.ch/quicksearch.php?gene+name={$id}	MycoBrowser tuberculosis at Global Health Institute	Global Health Institute, Lausanne	Switzerland	false	Rv1908c	both a catalase	http://tuberculist.epfl.ch/		
MIR:00000217	MycoBrowser leprae	Mycobacterium	^ML\w+$	Mycobrowser is a resource that provides both in silico generated and manually reviewed information within databases dedicated to the complete genomes of Mycobacterium tuberculosis, Mycobacterium leprae, Mycobacterium marinum and Mycobacterium smegmatis.	myco.lepra	http://identifiers.org/myco.lepra	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		Leproma
									MIR:00100276	http://mycobrowser.epfl.ch/leprosysearch.php?gene+name={$id}	MycoBrowser leprae at Global Health Institute	Global Health Institute, Lausanne	Switzerland	false	ML0224	dihydrofolate biosynthesis	http://mycobrowser.epfl.ch/leprosy.html		
MIR:00000218	MycoBrowser marinum	Mycobacterium	^MMAR_\d+$	Mycobrowser is a resource that provides both in silico generated and manually reviewed information within databases dedicated to the complete genomes of Mycobacterium tuberculosis, Mycobacterium leprae, Mycobacterium marinum and Mycobacterium smegmatis.	myco.marinum	http://identifiers.org/myco.marinum	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		marinolist
									MIR:00100277	http://mycobrowser.epfl.ch/marinosearch.php?gene+name={$id}	MycoBrowser marinum at Global Health Institute	Global Health Institute, Lausanne	Switzerland	false	MMAR_2462	arginine biosynthesis	http://mycobrowser.epfl.ch/marinolist.html		
MIR:00000219	MycoBrowser smegmatis	Mycobacterium	^MSMEG\w+$	Mycobrowser is a resource that provides both in silico generated and manually reviewed information within databases dedicated to the complete genomes of Mycobacterium tuberculosis, Mycobacterium leprae, Mycobacterium marinum and Mycobacterium smegmatis.	myco.smeg	http://identifiers.org/myco.smeg	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		smegmalist
									MIR:00100278	http://mycobrowser.epfl.ch/smegmasearch.php?gene+name={$id}	MycoBrowser smegmatis at Global Health Institute	Global Health Institute, Lausanne	Switzerland	false	MSMEG_3769	HMM TIGR00838	http://mycobrowser.epfl.ch/smegmalist.html		
MIR:00000221	OrthoDB	vertebrates, arthropods, and fungi	^\w+$	OrthoDB presents a catalog of eukaryotic orthologous protein-coding genes across vertebrates, arthropods, and fungi. Orthology refers to the last common ancestor of the species under consideration, and thus OrthoDB explicitly delineates orthologs at each radiation	orthodb	http://identifiers.org/orthodb	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		
									MIR:00100280	http://cegg.unige.ch/orthodb/results?searchtext={$id}	OrthoDB at Swiss Institute of Bioinformatics	Department of Genetic Medicine and Development, Swiss Institute of Bioinformatics, Geneva	Switzerland	false	Q9P0K8	Transcription factor, fork head	http://cegg.unige.ch/orthodb4		
MIR:00000222	Peroxibase	All species	^\d+$	Peroxibase provides access to peroxidase sequences from all kingdoms of life, and provides a series of bioinformatics tools	peroxibase	http://identifiers.org/peroxibase	0		id	accessURL	info	institution	location	official	localId	testString	resourceURL		

MIR:00000222	Peroxibase	All species	^\d+$	provides a series of bioinformatics tools and facilities suitable for analysing these sequences.	peroxibase	http://identifiers.org/peroxibase	0	MIR:00100281	http://peroxibase.toulouse.inra.fr/browse/process/view_perox.php?id={$id}	Peroxibase at University of Geneva	Department of Plant Biology, University of Geneva, Geneva	Switzerland	false	5282	CatA_Human	http://peroxibase.toulouse.inra.fr/		
MIR:00000223	PhylomeDB	All species	^\w+$	PhylomeDB is a database of complete phylomes derived for different genomes within a specific taxonomic range. It provides alignments, phylogentic trees and tree-based orthology predictions for all encoded proteins.	phylomedb	http://identifiers.org/phylomedb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100282	http://phylomedb.org/?seqid={$id}	PhylomeDB at Centro de Investigación Príncipe Felipe	Bioinformatics Department,Centro de Investigación Príncipe Felipe, Valencia	Spain	false	Phy000CLXM_RAT	Phy00085K5_HUMAN	http://phylomedb.org/		
MIR:00000226	ProtClustDB	All species	^\w+$	ProtClustDB is a collection of related protein sequences (clusters) consisting of Reference Sequence proteins encoded by complete genomes. This database contains both curated and non-curated clusters.	protclustdb	http://identifiers.org/protclustdb	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Protein Clusters Database
								MIR:00100285	https://www.ncbi.nlm.nih.gov/sites/entrez?Db=proteinclusters&Cmd=DetailsSearch&Term={$id}	ProtClustDB at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	O80725	P-glycoprotein 21	https://www.ncbi.nlm.nih.gov/proteinclusters?db=proteinclusters	Entrez ProtClustDB
MIR:00000227	PMP	All species	^([A-N,R-Z][0-9][A-Z][A-Z, 0-9][A-Z, 0-9][0-9])|([O,P,Q][0-9][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9])$	The number of known protein sequences exceeds those of experimentally solved protein structures. Homology (or comparative) modeling methods make use of experimental protein structures to build models for evolutionary related proteins. The Protein Model Portal (PMP)	pmp	http://identifiers.org/pmp	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Protein Model Portal
								MIR:00100286	http://www.proteinmodelportal.org/query/uniprot/{$id}	PMP at University of Basel	Biozentrum, University of Basel, Basel	Switzerland	false	Q0VCA6	158	http://www.proteinmodelportal.org/		
MIR:00000228	ProtoNet ProteinCard	All species	^\d+$	ProtoNet provides automatic hierarchical classification of protein sequences in the UniProt database, partitioning the protein space into clusters of similar proteins. This collection references protein information.	protonet.proteincard	http://identifiers.org/protonet.proteincard	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100287	http://www.protonet.cs.huji.ac.il/requested/protein_card.php?protein_id={$id}	ProtoNet at Hebrew University	Hebrew University, Jerusalem	Israel	false	16941567	Colicin	http://www.protonet.cs.huji.ac.il/		
MIR:00000229	ProtoNet Cluster	All species	^\d+$	ProtoNet provides automatic hierarchical classification of protein sequences in the UniProt database, partitioning the protein space into clusters of similar proteins. This collection references cluster information.	protonet.cluster	http://identifiers.org/protonet.cluster	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100288	http://www.protonet.cs.huji.ac.il/requested/cluster_card.php?cluster={$id}	ProtoNet at Hebrew University	Hebrew University, Jerusalem	Israel	false	4349895	Cyclotide	http://www.protonet.cs.huji.ac.il/		
MIR:00000230	REBASE	All species	^\d+$	REBASE is a comprehensive database of information about restriction enzymes, DNA methyltransferases and related proteins involved in the biological process of restriction-modification (R-M). It contains fully referenced information about recognition and cleavage sites.	rebase	http://identifiers.org/rebase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Restriction Enzyme Database
								MIR:00100289	http://rebase.neb.com/rebase/enz/{$id}.html	REBASE at New England Biolabs	New England Biolabs, Ipswich, Massachusetts	USA	false	101	Asp14H	http://rebase.neb.com/rebase/		
MIR:00000232	VectorBase	invertebrate	^\D{4}\d{6}(\-\D{2})?$	VectorBase is an NIAID-funded Bioinformatic Resource Center focused on invertebrate vectors of human pathogens. VectorBase annotates and curates vector genomes providing a web accessible integrated resource for the research community. Currently VectorBase	vectorbase	http://identifiers.org/vectorbase	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100291	https://www.vectorbase.org/search/site/{$id}?&site="Genome"	Vectorbase at EMBL-EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	ISCW007415	hypothetical protein	https://www.vectorbase.org/	
MIR:00000233	Human Disease Ontology	Human	^DOID\:\d+$	The Disease Ontology has been developed as a standardized ontology for human disease with the purpose of providing the biomedical community with consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts.	doid	http://identifiers.org/doid	1	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	DO
								MIR:00100292	http://purl.bioontology.org/ontology/DOID/DOID:{$id}	Human Disease Ontology through BioPortal	National Center for Biomedical Ontology, Stanford University	USA	false	bptl	11337	commensal bacterial	http://bioportal.bioontology.org/ontologies/DOID	
								MIR:00100293	https://www.ebi.ac.uk/ols/ontologies/doid/terms?obo_id=DOID:{$id}	Human Disease Ontology through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ols	11337	commensal bacterial	https://www.ebi.ac.uk/ols/ontologies/doid	
								MIR:00100297	http://disease-ontology.org/term/DOID:{$id}	Human Disease Ontology at Northwestern University	University of Maryland (Maryland) and Northwestern University (Illinois)	USA	false		11337	commensal bacterial	http://disease-ontology.org/	
MIR:00000234	Cell Cycle Ontology	All species	^CCO\:\w+$	The Cell Cycle Ontology is an application ontology that captures and integrates detailed knowledge on the cell cycle process.	cco	http://identifiers.org/cco	1	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100294	https://www.ebi.ac.uk/ontology-lookup/?termId=CCO:{$id}	Cell Cycle Ontology through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ols	P0000023	last of the stages of mitosis	https://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=CCO	
MIR:00000235	miRBase mature sequence	All species	MIMAT\d{7}	The miRBase Sequence Database is a searchable database of published miRNA sequences and annotation. This collection refers specifically to the mature miRNA sequence.	mirbase.mature	http://identifiers.org/mirbase.mature	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100296	http://www.mirbase.org/cgi-bin/mature.pl?mature_acc={$id}	miRBase	Faculty of Life Sciences, University of Manchester	UK	false	MIMAT0000001	cel-let-7-5p	http://www.mirbase.org/		
MIR:00000236	nextProt	Human	^NX_\w+	neXtProt is a resource on human proteins, and includes information such as proteins' function, subcellular location, expression, interactions and role in diseases.	nextprot	http://identifiers.org/nextprot	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100300	https://www.nextprot.org/db/entry/{$id}	nextProt at Swiss Institute of Bioinformatics	Swiss Institute of Bioinformatics, Geneva	Switzerland	false	NX_O00165	Promotes cell survival	https://www.nextprot.org/		
MIR:00000238	KEGG Genome	All species	^(T0\d+|\w{3,5})$	KEGG Genome is a collection of organisms whose genomes have been completely sequenced.	kegg.genome	http://identifiers.org/kegg.genome	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		KEGG Organism
								MIR:00100303	http://www.kegg.jp/entry/{$id}	KEGG Genome Database	Kyoto University Bioinformatics Center, Kyoto	Japan	true	eco	MG1655	http://www.genome.jp/kegg/catalog/org_list.html		KEGG
MIR:00000239	KEGG Metagenome	All species	^T3\d+$	The KEGG Metagenome Database collection information on environmental samples (ecosystems) of genome sequences for multiple species.	kegg.metagenome	http://identifiers.org/kegg.metagenome	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		KEGG
								MIR:00100304	http://www.kegg.jp/entry/{$id}	KEGG Metagenome Database	Kyoto University Bioinformatics Center, Kyoto	Japan	true	T30002	microbial communities	http://www.genome.jp/kegg/catalog/org_list3.html		
MIR:00000242	Pathway Ontology	Human	^PW:\d{7}$	The Pathway Ontology captures information on biological networks, the relationships between netweorks and the alterations or malfunctioning of such networks within a hierarchical structure. The five main branches of the ontology are: classic metabolic pathways, regulatory, signaling, drug, and disease pathwaysfor complex human conditions.	pw	http://identifiers.org/pw	1	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	
								MIR:00100309	http://rgd.mcw.edu/rgdweb/ontology/annot.html?acc_id=PW:{$id}	Pathway Ontology at Rat Genome Database	Medical College of Wisconsin, Wisconsin	USA	false	208	Diabetes mellitus	http://rgd.mcw.edu/rgdweb/ontology/search.html		
								MIR:00100310	https://www.ebi.ac.uk/ols/ontologies/pw/terms?obo_id=PW:{$id}	Pathway Ontology through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	208	Diabetes mellitus	https://www.ebi.ac.uk/ols/ontologies/pw	ols	
								MIR:00100311	http://purl.bioontology.org/ontology/PW/PW:{$id}	Pathway Ontology through BioPortal	National Center for Biomedical Ontology, Stanford, California	USA	false	208	Diabetes mellitus	http://bioportal.bioontology.org/ontologies/PW	bptl	
MIR:00000243	Sequence Read Archive	All species	^[SED]R[APRSXZ]\d+$	The Sequence Read Archive (SRA) stores raw sequencing data from the next generation of sequencing platforms Data submitted to SRA. It is organized using a metadata model consisting of six objects: study, sample, experiment, run, analysis and submission. The SRA study contains high-level information including goals of the study and literature references, and may be linked to the INSDC BioProject database.	insdc.sra	http://identifiers.org/insdc.sra	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100312	https://www.ncbi.nlm.nih.gov/sra/{$id}?&report=full	Sequence Read Archive at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	SRX000007	SID2748	https://www.ncbi.nlm.nih.gov/sra	
								MIR:00100313	https://www.ebi.ac.uk/ena/data/view/{$id}	European Nucleotide Archive (ENA)	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	SRX000007	SID2748	https://www.ebi.ac.uk/ena	
								MIR:00100314	http://trace.ddbj.nig.ac.jp/DRASearch/experiment?acc={$id}	DDBJ Sequence Read Archive (DRA)	DNA Data Bank of Japan, Mishima, Shizuoka	Japan	false		SRX000007	SID2748	http://trace.ddbj.nig.ac.jp/dra/	
MIR:00000244	ScerTF	yeast	^\w+$	ScerTF is a database of position weight matrices (PWMs) for transcription factors in Saccharomyces species. It identifies a single matrix for each TF that best predicts in vivo data, providing metrics related to the performance of that matrix in accurately representing the DNA	scretf	http://identifiers.org/scretf	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100315	http://stormo.wustl.edu/ScerTF/details/{$id}/	ScerTF at Washington University	Department of Genetics, Washington University Medical School, St Louis, Missouri	USA	false	RSC3	RSC chromatin remodeling	http://stormo.wustl.edu/ScerTF/		
MIR:00000245	PharmGKB Gene	Human	^PA\w+$	The PharmGKB database is a central repository for genetic, genomic, molecular and cellular phenotype data and clinical information about people who have participated in pharmacogenomics research studies. The data includes, but is not limited to, clinical and basic	pharmgkb.gene	http://identifiers.org/pharmgkb.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100316	http://www.pharmgkb.org/gene/{$id}	PharmGKB Gene at Stanford	Department of Genetics, School of Medicine, Stanford University, Stanford, California	USA	false	PA131	CYP3A5	http://www.pharmgkb.org/		
MIR:00000246	miRNEST	All species	^MNEST\d+$	miRNEST is a database of animal, plant and virus microRNAs, containing miRNA predictions conducted on Expressed Sequence Tags of animal and plant species.	mirnest	http://identifiers.org/mirnest	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100317	http://rhesus.amu.edu.pl/mirnest/copy/details.php?id={$id}	miRNEST at Adam Mickiewicz University	Institute of Molecular Biology, Adam Mickiewicz University, Poznan	Poland	false	MNEST029358	hsa-mir-3190	http://rhesus.amu.edu.pl/mirnest/copy/		
MIR:00000247	NAPP	bacteria	^\d+$	NAPP (Nucleic Acids Phylogenetic Profiling is a clustering method based on conserved noncoding RNA (ncRNA) elements in a bacterial genomes. Short intergenic regions from a reference genome are compared with other genomes to identify RNA rich clusters.	napp	http://identifiers.org/napp	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Nucleic Acid Phylogenetic Profiling
								MIR:00100318	http://rna.igmors.u-psud.fr/NAPP/Niveau2.php?specie={$id}	NAPP at CNRS	Institut de Génétique et Microbiologie, UMR 8621, CNRS, Université Paris Sud, Paris	France	false	351	CP000247	http://napp.u-psud.fr/		
MIR:00000248	NONCODE v3		^\d+$	NONCODE is a database of expression and functional lncRNA (long noncoding RNA) data obtained from microarray studies. LncRNAs have been shown to play key roles in various biological processes such as imprinting control, circuitry controlling pluripotency and differentiation, immune	noncodev3	http://identifiers.org/noncodev3	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100319	http://www.noncode.org/NONCODERv3/ncrna.php?ncid={$id}	NONCODE at Chinese Academy of Sciences	Bioinformatics Research Group, Chinese Academy of Sciences, Beijing	China	false	377550	Human lincRNA	http://www.noncode.org/		
MIR:00000249	VIRsiRNA	virus	^virsi\d+$	The VIRsiRNA database contains details of siRNA/shRNA which target viral genome regions. It provides efficacy information where available, as well as the siRNA sequence, viral target and subtype, as well as the target genomic region.	virsirna	http://identifiers.org/virsirna	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Virus siRNA database
								MIR:00100320	http://crdd.osdd.net/servers/virsirnadb/record.php?details={$id}	VIRsiRNA at Institute of Microbial Technology	Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Chandigarh, India	India	false	virsi1909	gaauguaacucgaucugcu	http://crdd.osdd.net/servers/virsirnadb		VIRsiRNAdb
MIR:00000250	ELM	All species	^[A-Za-z_0-9]+$	Linear motifs are short, evolutionarily plastic components of regulatory proteins. Mainly focused on the eukaryotic sequences,the Eukaryotic Linear Motif resource (ELM) is a database of curated motif classes and instances.	elm	http://identifiers.org/elm	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Eukaryotic Linear Motifs
								MIR:00100321	http://elm.eu.org/elms/elmPages/{$id}.html	ELM at EMBL (Germany)	Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg	Germany	false	CLV_MEL_PAP_1	Melanisation activation	http://elm.eu.org/		
MIR:00000253	BYKdb	bacteria	^[A-Z0-9]+$	The bacterial tyrosine kinase database (BYKdb) that collects sequences of putative and authentic bacterial tyrosine	bykdb	http://identifiers.org/bykdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Bacterial protein tYrosine Kinase database

MIR:00000253	BYKdb	bacteria	^[A-Z0-9]+$	putative and authentic bacterial tyrosine kinases, providing structural and functional information.	bykdb	http://identifiers.org/bykdb	0	MIR:00100324	https://bykdb.ibcp.fr/data/html/annotated/{$id}.html	BYKdb at CNRS	Unité Bases Moléculaires et Structurales des Systèmes Infectieux; CNRS - Université Claude Bernard, Lyon	France	false	A0A010GJI9	is a bacterial tyrosine kinase	http://bykdb.ibcp.fr/BYKdb/		
MIR:00000255	TopFind	human, mouse, yeast and E.coli	^([A-N,R-Z][0-9][A-Z][A-Z, 0-9][A-Z, 0-9][0-9])|([O,P,Q][0-9][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9])$	TopFIND is a database of protein termini, terminus modifications and their proteolytic processing in the species: Homo sapiens, Mus musculus, Arabidopsis thaliana, Saccharomyces cerevisiae and Escherichia coli.	topfind	http://identifiers.org/topfind	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100326	http://clipserve.clip.ubc.ca/topfind/proteins/{$id}	TopFind at University of British Columbia	Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver	Canada	false	Q9UKQ2	Disintegrin and metalloproteinase domain-containing protein 28	http://clipserve.clip.ubc.ca/topfind		
MIR:00000256	MIPModDB	All species	^\w+$	MIPModDB is a database of comparative protein structure models of MIP (Major Intrinsic Protein) family of proteins, identified from complete genome sequence. It provides key information of MIPs based on their sequence and structures	mipmod	http://identifiers.org/mipmod	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100327	http://bioinfo.iitk.ac.in/MIPModDB/result.php?code={$id}	MIPModDB at Indian Institute of Technology Kanpur	Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur	India	false	HOSAPI0399	major intrinsic protein of lens fiber	http://bioinfo.iitk.ac.in/MIPModDB		
MIR:00000261	CABRI	All species	^([A-Za-z]+)?(_)?([A-Za-z-]+)\:([A-Za-z0-9]+)$	CABRI (Common Access to Biotechnological Resources and Information) is an online service where users can search a number of European Biological Resource Centre catalogues. It lists the availability of a particular organism or genetic resource and defines the set of technical specifications and procedures which should be used to handle it	cabri	http://identifiers.org/cabri	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100334	http://www.cabri.org/CABRI/srs-bin/wgetz?-e+-page+EntryPage+[{$id}]	CABRI Cell Lines catalogue in Genova (SRS)	National Cancer Research Institute of Genova, Genova	Italy	true	dsmz_mutz-id:ACC 291	human glioblastoma	http://www.cabri.org/		
								MIR:00100335	http://www.be.cabri.org/CABRI/srs-bin/wgetz?-e+-page+EntryPage+[{$id}]	CABRI Cell Lines catalogue in Brussels (SRS)	BCCM, Brussels	Belgium	false	dsmz_mutz-id:ACC 291	human glioblastoma	http://www.be.cabri.org/		
MIR:00000262	CYGD	Yeast	^\w{2,3}\d{2,4}(\w)?$	The MIPS Comprehensive Yeast Genome Database (CYGD) provides information on the molecular structure and functional network of the entirely sequenced the budding yeast, Saccharomyces cerevisiae, as well as on related yeasts which are used for comparative analysis	cygd	http://identifiers.org/cygd	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100684	http://pedant.helmholtz-muenchen.de/pedant3htmlview/pedant3view?Db=p3_p13838_Sac_cerev&Method=ReportGene&GeneticelemCode={$id}	CYGD PEDANT Interface at Biomax Informatics	Biomax Informatics AG, Planegg	Germany	false	YFL039c	Actin	http://pedant.helmholtz-muenchen.de/pedant3htmlview/pedant3view?Method=start_method&Db=p3_p13838_Sac		
MIR:00000263	HUGE	human	^KIAA\d{4}$	The Human Unidentified Gene-Encoded (HUGE) protein database contains results from sequence analysis of human novel large (>4 kb) cDNAs identified in the Kazusa cDNA sequencing project.	huge	http://identifiers.org/huge	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100339	http://www.kazusa.or.jp/huge/gfpage/{$id}/	HUGE at Kazusa DNA Research Institute	Kazusa DNA Research Institute, Chiba	Japan	false	KIAA0001	P2Y purinoceptor 14	http://www.kazusa.or.jp/huge/		
MIR:00000265	STRING	All species	^([A-N,R-Z][0-9][A-Z][A-Z, 0-9][A-Z, 0-9][0-9])|([O,P,Q][0-9][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9])|([0-9][A-Za-z0-9]{3})$	STRING (Search Tool for Retrieval of Interacting Genes/Proteins) is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:Genomic Context, High-throughput Experiments,(Conserved) Coexpression, Previous Knowledge. STRING quantitatively integrates	string	http://identifiers.org/string	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Search Tool for Retrieval of Interacting Genes/Proteins
								MIR:00100341	http://string.embl.de/interactions/{$id}	STRING at Heidelberg	European Molecular Biology Laboratory, Heidelberg	Germany	false	P53350	polo-like kinase 1	http://string.embl.de/		
								MIR:00100342	http://string-db.org/interactions/{$id}	STRING Mirror at Heidelberg	European Molecular Biology Laboratory, Heidelberg	Germany	false	P53350	polo-like kinase 1	http://string-db.org/		
MIR:00000269	SNOMED CT	human	^(\w+)?\d+$	SNOMED CT (Systematized Nomenclature of Medicine -- Clinical Terms), is a systematically organized computer processable collection of medical terminology covering most areas of clinical information such as diseases, findings, procedures, microorganisms	snomedct	http://identifiers.org/snomedct	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100405	http://www.snomedbrowser.com/Codes/Details/{$id}	SNOMED-CT at The National Pathology Exchange	The Health Informatics Service, NHS, West Yorkshire	UK	false	284196006	Burn of skin (disorder)	http://www.snomedbrowser.com/		
MIR:00000270	MeSH 2012	human	^[A-Za-z0-9]+$	MeSH (Medical Subject Headings) is the National Library of Medicine's controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. This thesaurus is used by NLM for indexing articles from	mesh.2012	http://identifiers.org/mesh.2012	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Medical Subject Headings 2012
								MIR:00100349	http://www.nlm.nih.gov/cgi/mesh/2012/MB_cgi?mode=&index={$id}&view=expanded	MeSH at National Library of Medicine	U.S. National Library of Medicine, National Institute of Health, Maryland	USA	false	17186	Codon, Terminator	http://www.nlm.nih.gov/mesh/		
MIR:00000277	ABS	animal	^A\d+$	The database of Annotated regulatory Binding Sites (from orthologous promoters), ABS, is a public database of known binding sites identified in promoters of orthologous vertebrate genes that have been manually curated from bibliography	abs	http://identifiers.org/abs	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Annotated regulatory Binding Sites database
								MIR:00100358	http://genome.crg.es/datasets/abs2005/entries/{$id}.html	ABS at IMIM	Institut Municipal d'Investigació Mèdica, Barcelona	Spain	false	A0014	Collagen	http://genome.crg.es/datasets/abs2005/		
MIR:00000280	DPV	plant viruses	^\d+$	Description of Plant Viruses (DPV) provides information about viruses, viroids and satellites of plants, fungi and protozoa. It provides taxonomic information, including brief descriptions of each family and genus, and classified lists of virus sequences. The database also	dpv	http://identifiers.org/dpv	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Description of Plant Viruses
								MIR:00100361	http://www.dpvweb.net/dpv/showdpv.php?dpvno={$id}	DPV at Rothamsted	Plant-Pathogen Interactions Division, Wheat Pathogenesis Programme, Rothamsted Research Harpenden, Hertfordshire	UK	false	100	Rice transitory yellowing	http://www.dpvweb.net/		DPVweb
MIR:00000281	IUPHAR receptor	mamalian	^\d+$	The IUPHAR Compendium details the molecular, biophysical and pharmacological properties of identified mammalian sodium, calcium and potassium channels, as well as the related cyclic nucleotide-modulated ion channels and transient receptor potential channels	iuphar.receptor	http://identifiers.org/iuphar.receptor	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		International Union of Basic and Clinical Pharmacology
								MIR:00100362	http://www.guidetopharmacology.org/GRAC/ObjectDisplayForward?objectId={$id}	IUPHAR Receptor at University of Edinburgh	Centre for Cardiovascular Science, University of Edinburgh, Edinburgh	UK	true	101	GPR34	http://www.guidetopharmacology.org/targets.jsp		IUPHAR-DB
MIR:00000282	Aceview Worm	worm	^[a-z0-9-]+$	AceView provides a curated sequence representation of all public mRNA sequences (mRNAs from GenBank or RefSeq, and single pass cDNA sequences from dbEST and Trace). These are aligned on the genome and clustered into a minimal number of alternative transcript	aceview.worm	http://identifiers.org/aceview.worm	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	AceView WormGenes
								MIR:00100363	https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=worm&c=Gene&l={$id}	AceView Worm at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	aap-1	phosphoinositide 3-kinase	https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html?worm	
MIR:00000283	ASAP	bacteria	^[A-Za-z0-9-]+$	ASAP (a systematic annotation package for community analysis of genomes) stores bacterial genome sequence and functional characterization data. It includes multiple genome sequences at various stages of analysis, corresponding experimental data and access to	asap	http://identifiers.org/asap	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100364	http://asap.ahabs.wisc.edu/asap/feature_info.php?LocationID=WIS&FeatureID={$id}	ASAP at University of Wisconsin	Animal Health and Biomedical Sciences, University of Wisconsin, Madison	USA	false	ABE-0009634	biosynthetic arginine decarboxylase, PLP-binding	http://asap.ahabs.wisc.edu/asap/home.php		
MIR:00000285	BDGP EST	Drosophila	^\w+(\.)?(\d+)?$	The BDGP EST database collects the expressed sequence tags (ESTs) derived from a variety of tissues and developmental stages for Drosophila melanogaster. All BDGP ESTs are available at dbEST (NCBI).	bdgp.est	http://identifiers.org/bdgp.est	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Berkeley Drosophila Genome Project EST
								MIR:00100366	https://www.ncbi.nlm.nih.gov/nucest/{$id}	BDGP EST at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	EY223054.1	Moniliophthora perniciosa cDNA	https://www.ncbi.nlm.nih.gov/dbEST/index.html	
MIR:00000286	Dictybase Gene	Dictyostelium	^DDB_G\d+$	The dictyBase database provides data on the model organism Dictyostelium discoideum and related species. It contains the complete genome sequence, ESTs, gene models and functional annotations. This collection references gene information	dictybase.gene	http://identifiers.org/dictybase.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100367	http://dictybase.org/gene/{$id}	Dictybase Gene at Northwestern University	Northwestern University Biomedical Informatics Center and Center for Genetic Medicine, Chicago, Illinois	USA	false	DDB_G0267522	Gene Information for DDB_G0267522	http://dictybase.org/		
MIR:00000287	IMGT LIGM	vertebrates	^M\d+$	IMGT, the international ImMunoGeneTics project, is a collection of high-quality integrated databases specialising in Immunoglobulins, T cell receptors and the Major Histocompatibility Complex (MHC) of all vertebrate species. IMGT/LIGM is a comprehensive database of fully annotated sequences of Immunoglobulins and T cell receptors from human and other vertebrates	imgt.ligm	http://identifiers.org/imgt.ligm	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		IMGT/LIGM-DB
								MIR:00100369	http://genius.embnet.dkfz-heidelberg.de/menu/cgi-bin/srs7.1.3.1/wgetz?-id+4ktAl1fNMTA+-e+[IMGT:'{$id}']	IMGT LIGM through DKFZ (SRS)	DKFZ TP3, Heidelberg	Germany	false	M94112	Homo sapiens Ig germline	http://genius.embnet.dkfz-heidelberg.de/		
								MIR:00100428	http://www.imgt.org/ligmdb/view?id={$id}	IMGT LIGM at CNRS	CNRS, Montpellier University, Montpellier	France	true	M94112	Homo sapiens Ig germline	http://www.imgt.org/		
MIR:00000288	Worfdb	worm	^\w+(\.\d+)?	WorfDB (Worm ORFeome DataBase) contains data from the cloning of complete set of predicted protein-encoding Open Reading Frames (ORFs) of Caenorhabditis elegans. This collection describes experimentally defined transcript structures of unverified genes	worfdb	http://identifiers.org/worfdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Worm Orf DB
								MIR:00100370	http://worfdb.dfci.harvard.edu/index.php?search_type=name&page=showresultrc&race_query={$id}	Worm Orf DB at Harvard	Center for Cancer Systems Biology, Harvard Medical School, Massachusetts	USA	false	T01B6.1	CTGGAGTCACAGAGGACATAA	http://worfdb.dfci.harvard.edu/		
MIR:00000289	NEXTDB	Caenorhabditis	^[A-Za-z0-9]+$	NextDB is a database that provides information on the expression pattern map of the 100Mb genome of the nematode Caenorhabditis elegans. This was done through EST analysis and systematic whole mount in situ hybridization. Information available	nextdb	http://identifiers.org/nextdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Nematode Expression Pattern Database
								MIR:00100371	http://nematode.lab.nig.ac.jp/db2/ShowCloneInfo.php?clone={$id}	NEXTDB at Shizuoka	National Institute of Genetics, Shizuoka	Japan	false	6b1	transketolase	http://nematode.lab.nig.ac.jp/		
MIR:00000291	SoyBase	Soybean	^\w+(\-)?\w+(\-)?\w+$	SoyBase is a repository for curated genetics, genomics and related data resources for soybean.	soybase	http://identifiers.org/soybase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100373	http://www.soybase.org/sbt/search/search_results.php?category=SNP&search_term={$id}	SoyBase at Iowa State University	USDA-ARS-CICGRU, Iowa State University, Iowa	USA	false	BARC-013845-01256	Glycine max	http://soybase.org/		
MIR:00000292	HAMAP	Microbial organisms	^MF_\d+$	HAMAP is a system that identifies and semi-automatically annotates proteins that are part of well-conserved and orthologous microbial families or subfamilies. These are used to build rules which are used to propagate annotations to member bacterial, archaeal and plastid-	hamap	http://identifiers.org/hamap	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		High-quality Automated and Manual Annotation of microbial Proteomes
								MIR:00100374	https://hamap.expasy.org/unirule/{$id}	HAPMAP at Swiss Institute of Bioinformatics	Swiss Institute of Bioinformatics, Geneva	Switzerland	false	MF_01400	Peptide-L-methionine	https://hamap.expasy.org/		
MIR:00000293	Rouge	human and mouse	^m\w+$	The Rouge protein database contains results from sequence analysis of novel large (>4 kb) cDNAs identified in the Kazusa cDNA sequencing project.	rouge	http://identifiers.org/rouge	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Rodent Unidentified Gene-Encoded Large Proteins
								MIR:00100375	http://www.kazusa.or.jp/rouge/gfpage/{$id}/	Rouge at Kazusa DNA Research Institute	Kazusa DNA Research Institute, Chiba	Japan	false	mKIAA4200	DEP domain containing	http://www.kazusa.or.jp/rouge/		
MIR:00000294	ArrayExpress Platform	All species	^[AEP]-\w{4}-\d+$	ArrayExpress is a public repository for microarray data, which is aimed at storing MIAME-compliant data in accordance with Microarray Gene Expression Data (MGED) recommendations.This collection references the specific platforms used in the generation of experimental results	arrayexpress.platform	http://identifiers.org/arrayexpress.platform	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100376	https://www.ebi.ac.uk/arrayexpress/arrays/{$id}	ArrayExpress Platform at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	A-GEOD-50	spotted DNA/cDNA	https://www.ebi.ac.uk/arrayexpress/	
MIR:00000295	CGSC Strain	E.coli	^\d+$	The CGSC Database of E. coli genetic information includes genotypes and reference information for the strains in the CGSC collection, the names, synonyms, properties, and map position for genes, gene product information, and information on specific mutations and	cgsc	http://identifiers.org/cgsc	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Coli Genetic Stock Center
								MIR:00100377	http://cgsc.biology.yale.edu/Strain.php?ID={$id}	CGSC at Yale university	Dept. of Molecular, Cellular, and Developmental Biology, Yale University, Conneticut	USA	false	11042	araC14	http://cgsc.biology.yale.edu/index.php		
MIR:00000297	DragonDB DNA	Antirrhinum	^\d\w+$	DragonDB is a genetic and genomic database for Antirrhinum majus (Snapdragon). This collection refers to DNA sequence information.	dragondb.dna	http://identifiers.org/dragondb.dna	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100379	http://antirrhinum.net/cgi-bin/ace/generic/tree/DragonDB?name={$id};class=DNA	DragonDB DNA at University of British Columbia	University of British Columbia, Vancouver	Canada	false	3hB06	gagagcatt	http://www.antirrhinum.net/		

MIR:00000298	DragonDB Protein	Antirrhinum	^\w+$	DragonDB is a genetic and genomic database for Antirrhinum majus (Snapdragon). This collection refers to protein sequence information.	dragondb.protein	http://identifiers.org/dragondb.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100380	http://antirrhinum.net/cgi-bin/ace/generic/tree/DragonDB?name={$id};class=Peptide	DragonDB Protein at University of British Columbia	University of British Columbia, Vancouver	Canada	false	AMDEFA	ARGKIQIKRIENQ	http://www.antirrhinum.net/		
MIR:00000299	DragonDB Locus	Antirrhinum	^\w+$	DragonDB is a genetic and genomic database for Antirrhinum majus (Snapdragon). This collection refers to Locus information.	dragondb.locus	http://identifiers.org/dragondb.locus	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100381	http://antirrhinum.net/cgi-bin/ace/generic/tree/DragonDB?name={$id}&class=Locus	DragonDB Locus at University of British Columbia	University of British Columbia, Vancouver	Canada	false	DEF	marker (RFLP)	http://www.antirrhinum.net/		
MIR:00000300	DragonDB Allele	Antirrhinum	^\w+$	DragonDB is a genetic and genomic database for Antirrhinum majus (Snapdragon). This collection refers to allele information.	dragondb.allele	http://identifiers.org/dragondb.allele	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100382	http://antirrhinum.net/cgi-bin/ace/generic/tree/DragonDB?name={$id}&class=Allele	DragonDB Allele at University of British Columbia	University of British Columbia, Vancouver	Canada	false	cho	FIM extends to first whorl	http://www.antirrhinum.net/		
MIR:00000303	MGED Ontology	All species	^\w+$	The MGED Ontology (MO) provides terms for annotating all aspects of a microarray experiment from the design of the experiment and array layout, through to the preparation of the biological sample and the protocols used to hybridize the RNA and analyze the data.	mo	http://identifiers.org/mo	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	MO
								MIR:00100385	http://purl.bioontology.org/ontology/MO/{$id}	MGED Ontology at BioPortal	National Center for Biomedical Ontology, Stanford University	USA	false	bptl	ArrayGroup	physical platform	http://bioportal.bioontology.org/	
								MIR:00100568	https://mged.sourceforge.net/ontologies/MGEDontology.php#{$id}	MGED Ontology at SourceForge	(Ontology Working Group), European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	ArrayGroup	physical platform	https://mged.sourceforge.net/ontologies/MGEDontology.php	
MIR:00000304	NASC code	Arabidopsis	^(\w+)?\d+$	The Nottingham Arabidopsis Stock Centre (NASC) provides seed and information resources to the International Arabidopsis Genome Programme and the wider research community.	nasc	http://identifiers.org/nasc	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		The Nottingham Arabidopsis Stock Centre
								MIR:00100386	http://arabidopsis.info/StockInfo?NASC_id={$id}	NASC at University of Nottingham	The Nottingham Arabidopsis Stock Centre, University of Nottingham	UK	false	N1899	Recombinant inbred line	http://arabidopsis.info/		
MIR:00000305	NIAEST	mouse	^\w\d{4}\w\d{2}(\-[35])?$	A catalog of mouse genes expressed in early embryos, embryonic and adult stem cells, including 250000 ESTs, was assembled by the NIA (National Institute on Aging) assembled.This collection represents the name and sequence from individual cDNA clones.	niaest	http://identifiers.org/niaest	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100387	http://lgsun.grc.nia.nih.gov/cgi-bin/pro3?sname1={$id}	NIAEST at National Institute on Aging	National Institute on Aging, NIH	USA	false	J0705A10	GTACCTTTAAAAGAA	http://lgsun.grc.nia.nih.gov/cDNA/		
MIR:00000307	Plant Ontology	Plant species	^PO:\d+$	The Plant Ontology is a structured vocabulary and database resource that links plant anatomy, morphology and growth and development to plant genomics data.	po	http://identifiers.org/po	1	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	PO
								MIR:00100389	http://www.plantontology.org/amigo/go.cgi?view=details&query=PO:{$id}	Plant Ontology through Amigo	Cold Spring Harbor Laboratory, Cold Spring Harbor, New York	USA	false	amigo	9089	endosperm	http://www.plantontology.org/	
								MIR:00100390	http://purl.bioontology.org/ontology/PO/PO:{$id}	Plant Ontology through BioPortal	National Center for Biomedical Ontology, Stanford	USA	false	bptl	9089	endosperm	http://bioportal.bioontology.org/ontologies/PO	
								MIR:00100676	https://www.ebi.ac.uk/ols/ontologies/po/terms?obo_id=PO:{$id}	Plant Ontology through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ols	9089	endosperm	https://www.ebi.ac.uk/ols/ontologies/po	
MIR:00000308	RNA Modification Database	All species	^\d{3}$	The RNA modification database provides a comprehensive listing of post-transcriptionally modified nucleosides from RNA. The database consists of all RNA-derived ribonucleosides of known structure, including those from established sequence positions, as well as	rnamods	http://identifiers.org/rnamods	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		RNAmods
								MIR:00100391	http://mods.rna.albany.edu/mods/modifications/view/{$id}	RNA Modification Database at University of Utah	Department of Medicinal Chemistry and Department of Biochemistry, University of Utah, Salt Lake City, Utah	USA	false	101	isowyosine	http://rna-mdb.cas.albany.edu/RNAmods/rnaover.htm		
MIR:00000312	TreeBASE	All species	^TB[1,2]?:[A-Z][a-z]?\d+$	TreeBASE is a relational database designed to manage and explore information on phylogenetic relationships. It includes phylogenetic trees and data matrices, together with information about the relevant publication, taxa, morphological and	treebase	http://identifiers.org/treebase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100395	http://purl.org/phylo/treebase/phylows/study/{$id}?format=html	TreeBASE (hosted at National Evolutionary Synthesis Center)	Phyloinformatics Research Foundation, Durham, North Carolina	USA	false	TB2:S1000	Out of Anatolia: Longitudina	http://treebase.org/		
MIR:00000313	Tetrahymena Genome Database	Tetrahymena	^TTHERM_\d+$	The Tetrahymena Genome Database (TGD) Wiki is a database of information about the Tetrahymena thermophila genome sequence. It provides information curated from the literature about each published gene, including a standardized gene name, a link to the genomic locus	tgd	http://identifiers.org/tgd	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		TGD
								MIR:00100396	http://ciliate.org/index.php/feature/details/{$id}	TGD at Stanford University	Department of Genetics, School of Medicine, Stanford University, Stanford, California	USA	false	TTHERM_00648910	Intraflagellar transport 52	http://ciliate.org/index.php/		
MIR:00000315	TIGRFAMS	All species	^TIGR\d+$	TIGRFAMs is a resource consisting of curated multiple sequence alignments, Hidden Markov Models (HMMs) for protein sequence classification, and associated information designed to support automated annotation of (mostly prokaryotic) proteins.	tigrfam	http://identifiers.org/tigrfam	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100398	http://www.jcvi.org/cgi-bin/tigrfams/HmmReportPage.cgi?acc={$id}	TIGRFAM at JCVI	The Institute for Genomic Research, Rockville, Maryland	USA	false	TIGR00010	starting with a urease alpha subunit	http://www.jcvi.org/cgi-bin/tigrfams/Listing.cgi		
MIR:00000316	Animal TFDB Family	animal	^\w+$	The Animal Transcription Factor DataBase (AnimalTFDB) classifies TFs in sequenced animal genomes, as well as collecting the transcription co-factors and chromatin remodeling factors of those genomes. This collections refers to transcription factor families, and the species in which they are	atfdb.family	http://identifiers.org/atfdb.family	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Animal Transcription Factor Database
								MIR:00100399	http://www.bioguo.org/AnimalTFDB/family.php?fam={$id}	Animal TFDB at Hubei Bioinformatics & Molecular Imaging Key Laboratory	Hubei Bioinformatics and Molecular Imaging Key Laboratory, Wuhan	China	false	CUT	Ailuropoda melanoleuca	http://www.bioguo.org/AnimalTFDB/family_index.php		
MIR:00000318	DBG2 Introns	bacteria	^\w{1,2}\.(\w{1,2}\.)?[A-Za-z0-9]+$	The Database for Bacterial Group II Introns provides a catalogue of full-length, non-redundant group II introns present in bacterial DNA sequences in GenBank.	dbg2introns	http://identifiers.org/dbg2introns	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Database for Bacterial Group II Introns
								MIR:00100401	http://webapps2.ucalgary.ca/~groupii/cgi-bin/intron.cgi?name={$id}	Bacterial Group II Introns at University of Calgary	Department of Biological Sciences, University of Calgary, Calgary, Alberta	Canada	false	Cu.me.I1	cgagcagctt	http://webapps2.ucalgary.ca/~groupii/		
MIR:00000320	Vbase2	human and mouse	^\w+$	The database VBASE2 provides germ-line sequences of human and mouse immunoglobulin variable (V) genes.	vbase2	http://identifiers.org/vbase2	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100403	http://www.vbase2.org/vgene.php?id={$id}	Vbase2 at German Research Centre for Biotechnology	Department of Experimental Immunology, German Research Centre for Biotechnology, Braunschweig	Germany	false	humIGHV025	GAGGTGCAGCTGG	http://www.vbase2.org/vbase2.php		
MIR:00000321	SPIKE Map	human	^spike\d{5}$	SPIKE (Signaling Pathways Integrated Knowledge Engine) is a repository that can store, organise and allow retrieval of pathway information in a way that will be useful for the research community. The database currently focuses primarily on pathways describing DNA damage	spike.map	http://identifiers.org/spike.map	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Signaling Pathways Integrated Knowledge Engine
								MIR:00100404	http://www.cs.tau.ac.il/~spike/maps/{$id}.html	SPIKE database at Tel Aviv University	School of Computer Sciences and School of Medicine, Tel Aviv University	Israel	false	spike00001	Cdk4-Cyclin D complex	http://www.cs.tau.ac.il/~spike/		
MIR:00000323	GeneCards	human	^[A-Za-z-0-9_]+(\@)?$	The GeneCards human gene database stores gene related transcriptomic, genetic, proteomic, functional and disease information. It uses standard nomenclature and approved gene symbols. GeneCards presents a complete summary for each human gene	genecards	http://identifiers.org/genecards	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100407	http://www.genecards.org/cgi-bin/carddisp.pl?gene={$id}	GeneCards at Weizmann Institute	Department of Molecular Genetics, Weizmann Institute of Science, Rehovot	Israel	true	ABL1	protooncogene	http://www.genecards.org/		
MIR:00000329	mirEX	Arabidopsis	^\d+(\w+)?$	mirEX is a comprehensive platform for comparative analysis of primary microRNA expression data, storing RT–qPCR-based gene expression profile over seven development stages of Arabidopsis. It also provides RNA structural models, publicly available deep sequencing results and	mirex	http://identifiers.org/mirex	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100423	http://comgen.pl/mirex1/?page=results/record&name={$id}&exref=pp2a&limit=yes	mirEX at Adam Mickiewicz University	Institute of Molecular Biology, Adam Mickiewicz University, Poznan	Poland	false	165a	CUGCUAAGAUC	http://comgen.pl/mirex/?page=home		
MIR:00000330	Dictybase EST	Dictyostelium	^DDB\d+$	The dictyBase database provides data on the model organism Dictyostelium discoideum and related species. It contains the complete genome sequence, ESTs, gene models and functional annotations. This collection references expressed sequence tag (EST) information.	dictybase.est	http://identifiers.org/dictybase.est	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100424	http://dictybase.org/db/cgi-bin/feature_page.pl?primary_id={$id}	Dictybase Gene at Northwestern University	Northwestern University Biomedical Informatics Center and Center for Genetic Medicine, Chicago, Illinois	USA	false	DDB0016567	CATTAATCAA	http://dictybase.org/		
MIR:00000331	IMGT HLA	Dictyostelium	^[A-Z0-9*:]+$	IMGT, the international ImMunoGeneTics project, is a collection of high-quality integrated databases specialising in Immunoglobulins, T cell receptors and the Major Histocompatibility Complex (MHC) of all vertebrate species. IMGT/HLA is a database for sequences of the human	imgt.hla	http://identifiers.org/imgt.hla	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	IPD IMGT/HLA
								MIR:00100425	https://www.ebi.ac.uk/cgi-bin/imgt/hla/get_allele.cgi?{$id}	IMGT HLA at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	A*01:01:01:01	Caucasoid, Oriental	https://www.ebi.ac.uk/imgt/hla/allele.html	
MIR:00000335	PomBase	Yeast	^S\w+(\.)?\w+(\.)?$	PomBase is a model organism database established to provide access to molecular data and biological information for the fission yeast Schizosaccharomyces pombe. It encompasses annotation of genomic sequence and features, comprehensive manual literature curation and genome	pombase	http://identifiers.org/pombase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100430	http://www.pombase.org/spombe/result/{$id}	PomBase at University of Cambridge	Cambridge Systems Biology Centre, Department of Biochemistry, University of Cambridge	UK	false	SPCC13B11.01	alcohol dehydrogenase	http://www.pombase.org/		
MIR:00000336	HPA	human	^ENSG\d{11}$	The Human Protein Atlas (HPA) is a publicly available database with high-resolution images showing the spatial distribution of proteins in different normal and cancer human cell lines. Primary access to this collection is through Ensembl Gene identifiers.	hpa	http://identifiers.org/hpa	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Human Protein Atlas
								MIR:00100431	http://www.proteinatlas.org/{$id}	Human Protein Atlas at AlbaNova University	Department of Proteomics, School of Biotechnology, AlbaNova University Center, Stockholm	Sweden	false	ENSG00000026508	Blood group antigen proteins	http://www.proteinatlas.org/		
MIR:00000337	JAX Mice	mouse	^\d+$	JAX Mice is a catalogue of mouse strains supplied by the Jackson Laboratory.	jaxmice	http://identifiers.org/jaxmice	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100432	http://jaxmice.jax.org/strain/{$id}.html	JAX Mice at Jackson Laboratory	The Jackson Laboratory, Sacremento, California	USA	false	5012	Chemically Induced Mutation	http://jaxmice.jax.org/		
MIR:00000339	YeTFasCo	Yeast	^\w+_\d+(\.\d+)?$	The Yeast Transcription Factor Specificity Compendium (YeTFasCO) is a database of transcription factor specificities for the yeast Saccharomyces cerevisiae in Position Frequency Matrix (PFM) or Position Weight Matrix (PWM) formats.	yetfasco	http://identifiers.org/yetfasco	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Yeast Transcription Factor Specificity Compendium
								MIR:00100434	http://yetfasco.ccbr.utoronto.ca/showPFM.php?mot={$id}	YeTFasCo at University of Toronto	Department of Molecular Genetics, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto	Canada	false	YOR172W_571.0	Motif 571 for YOR172W	http://yetfasco.ccbr.utoronto.ca/		
MIR:00000340	TarBase	All species	^[a-z]{3}\-(mir|let|lin)\-\w+(\-\w+\-\w+)?	TarBase stores microRNA (miRNA) information for miRNA–gene interactions, as well as miRNA- and gene-related facts to information specific to the interaction and the experimental validation methodologies used.	tarbase	http://identifiers.org/tarbase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100713	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index&mirnas={$id}	TarBase v7 at University of Thessaly	DIANA-Lab, Department of Electrical & Computer Engineering, University of Thessaly	Greece	false	hsa-let-7a-2-3p	IGFBP3	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index		
MIR:00000342	OMA Protein	All species	^[A-Z0-9]{5}\d+$	OMA (Orthologous MAtrix) is a database that identifies orthologs among publicly available, complete genome sequences. It identifies orthologous relationships which	oma.protein	http://identifiers.org/oma.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Orthologous MAtrix

MIR:00000342	OMA Protein	All species	^[A-Z0-9]{5}\d+$	identifies orthologous relationships which can be accessed either group-wise, where all group members are orthologous to all other group members, or on a sequence	oma.protein	http://identifiers.org/oma.protein	0	MIR:00100437	http://omabrowser.org/cgi-bin/gateway.pl?f=DisplayEntry&p1={$id}	OMA Protein through OMA browser at ETH Zurich	ETH Zurich, Computer Science, Zurich	Switzerland	false	HUMAN16963	coiled-coil domain	http://omabrowser.org/cgi-bin/gateway.pl		
MIR:00000343	OMA Group	All species	^[A-Z]+$	OMA (Orthologous MAtrix) is a database that identifies orthologs among publicly available, complete genome sequences. It identifies orthologous relationships which can be accessed either group-wise, where all group members are orthologous to all other group members, or on a sequence-	oma.grp	http://identifiers.org/oma.grp	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100438	http://omabrowser.org/cgi-bin/gateway.pl?f=DisplayGroup&p1={$id}	OMA Group through OMA browser at ETH Zurich	ETH Zurich, Computer Science, Zurich	Switzerland	false	LCSCCPN	SARHA02316	http://omabrowser.org/cgi-bin/gateway.pl		
MIR:00000344	NCBI Protein	All species	^(\w+\d+(\.\d+)?)|(NP_\d+)$	The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB.	ncbiprotein	http://identifiers.org/ncbiprotein	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100439	https://www.ncbi.nlm.nih.gov/protein/{$id}	NCBI Protein at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	CAA71118.1	Klebsiella pneumoniae	https://www.ncbi.nlm.nih.gov/protein	
MIR:00000345	GenPept	All species	^\w{3}\d{5}(\.\d+)?$	The GenPept database is a collection of sequences based on translations from annotated coding regions in GenBank.	genpept	http://identifiers.org/genpept	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	protein genbank identifier
								MIR:00100440	https://www.ncbi.nlm.nih.gov/protein/{$id}?report=genpept	GenPept at NCBI	National Center for Biotechnology Information (NCBI)	USA	false	ncbi	CAA71118.1	Klebsiella pneumoniae	https://www.ncbi.nlm.nih.gov/protein	
MIR:00000346	UniGene	All species	^\d+$	A UniGene entry is a set of transcript sequences that appear to come from the same transcription locus (gene or expressed pseudogene), together with information on protein similarities, gene expression, cDNA clone reagents, and genomic location	unigene	http://identifiers.org/unigene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100441	http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?UGID={$id}	UniGene at NCBI	National Center for Biotechnology , Maryland	USA	false	4900	Arabidopsis thaliana	http://www.ncbi.nlm.nih.gov/unigene		
MIR:00000351	PiroplasmaDB	Eukaryotic pathogens	^TA\d+$	PiroplasmaDB is one of the databases that can be accessed through the EuPathDB (http://EuPathDB.org; formerly ApiDB) portal, covering eukaryotic pathogens of the genera Cryptosporidium, Giardia, Leishmania, Neospora, Plasmodium, Toxoplasma, Trichomonas and	piroplasma	http://identifiers.org/piroplasma	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100448	http://piroplasmadb.org/piro/showRecord.do?name=GeneRecordClasses.GeneRecordClass&source_id={$id}	PiroplasmaBD at EuPathDB	Center for Tropical & Emerging Global Diseases, University of Georgia, Georgia	USA	false	TA14985	topoisomerase IV subunit A	http://piroplasmadb.org/		
MIR:00000352	Unite	fungus	^UDB\d{6}$	UNITE is a fungal rDNA internal transcribed spacer (ITS) sequence database. It focuses on high-quality ITS sequences generated from fruiting bodies collected and identified by experts and deposited in public herbaria. Entries may be supplemented with metadata on	unite	http://identifiers.org/unite	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100449	http://unite.ut.ee/bl_forw.php?nimi={$id}	Unite at University of Tartu	Institute of Botany, University of Tartu, Tartu	Estonia	false	UDB000691	albatrellus	http://unite.ut.ee/		
MIR:00000354	ProGlycProt	bacteria and archea	^[A-Z]C\d{1,3}$	ProGlycProt (Prokaryotic GlycoProtein) is a repository of bacterial and archaeal glycoproteins with at least one experimentally validated glycosite (glycosylated residue). Each entry in the database is fully cross-referenced and enriched with available published	proglyc	http://identifiers.org/proglyc	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100451	http://www.proglycprot.org/detail.aspx?ProId={$id}	ProGlycProt at IMT	Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Chandigarh	India	false	AC119	Staphylothermus marinus F1	http://www.proglycprot.org/		
MIR:00000357	SUPFAM	All species	^\w+$	SUPERFAMILY provides structural, functional and evolutionary information for proteins from all completely sequenced genomes, and large sequence collections such as UniProt.	supfam	http://identifiers.org/supfam	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		SUPERFAMILY
								MIR:00100454	http://supfam.org/SUPERFAMILY/cgi-bin/scop.cgi?ipid={$id}	SUPFAM at MRC and University of Bristol	MRC Laboratory of Molecular Biology, Cambridge	UK	false	SSF57615	cellulose binding domain	http://supfam.org/SUPERFAMILY/		
MIR:00000358	Rice Genome Annotation Project	rice	^LOC_Os\d{1,2}g\d{5}$	The objective of this project is to provide high quality annotation for the rice genome Oryza sativa spp japonica cv Nipponbare. All genes are annotated with functional annotation including expression data, gene ontologies, and tagged lines.	ricegap	http://identifiers.org/ricegap	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100455	http://rice.plantbiology.msu.edu/cgi-bin/ORF_infopage.cgi?&orf={$id}	Rice Genome Annotation Project at TIGR	The Institute for Genomic Research, Rockville, Maryland	USA	false	LOC_Os02g13300	mitochondrial carrier protein	http://rice.plantbiology.msu.edu/annotation_pseudo_current.shtml		
MIR:00000359	PINA	human	^([A-N,R-Z][0-9][A-Z][A-Z, 0-9][A-Z, 0-9][0-9])|([O,P,Q][0-9][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9])$	Protein Interaction Network Analysis (PINA) platform is an integrated platform for protein interaction network construction, filtering, analysis, visualization and management. It integrates protein-protein interaction data from six public curated databases	pina	http://identifiers.org/pina	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Protein Interaction Network Analysis
								MIR:00100456	http://cbg.garvan.unsw.edu.au/pina/interactome.oneP.do?ac={$id}&showExtend=null	Protein Interaction Network Analysis (PINA) at Garvan Institute	Garvan Institute of Medical Research, Sydney	Australia	false	Q13485	SMAD4	http://cbg.garvan.unsw.edu.au/pina/		
MIR:00000361	BacMap Biography	bacteria	^\d+$	BacMap is an electronic, interactive atlas of fully sequenced bacterial genomes. It contains labeled, zoomable and searchable chromosome maps for sequenced prokaryotic (archaebacterial and eubacterial) species. Each map can be zoomed to the level of individual genes	bacmap.biog	http://identifiers.org/bacmap.biog	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100459	http://bacmap.wishartlab.com/organisms/{$id}	BacMap Biography at University of Alberta	Department of Computing Science, Food and Nutritional Science, University of Alberta, Edmonton	Canada	false	1050	thermoacidophilic, anaerobic	http://bacmap.wishartlab.com/		
MIR:00000362	HGNC Symbol	human	^[A-Za-z-0-9_]+(\@)?$	The HGNC (HUGO Gene Nomenclature Committee) provides an approved gene name and symbol (short-form abbreviation) for each known human gene. All approved symbols are stored in the HGNC database, and each symbol is unique. This collection refers to records	hgnc.symbol	http://identifiers.org/hgnc.symbol	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	HUGO Gene Nomenclature Committee Symbol
								MIR:00100460	https://www.genenames.org/cgi-bin/gene_symbol_report?match={$id}	HGNC Symbol at HUGO Genome Nomenclature Committee	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	DAPK1	CCDS43842	https://www.genenames.org/	
MIR:00000363	PANTHER Pathway	All species	^P\d{5}$	The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System is a resource that classifies genes by their functions, using published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct	panther.pathway	http://identifiers.org/panther.pathway	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100461	http://www.pantherdb.org/pathway/pathwayDiagram.jsp?catAccession={$id}	PANTHER Pathway at USC (Los Angeles)	Keck School of Medicine, University of Southern California	USA	false	P00024	Glycolysis	http://www.pantherdb.org/		
MIR:00000365	FungiDB	fungi	^[A-Za-z_0-9]+$	FungiDB is a genomic resource for fungal genomes. It contains contains genome sequence and annotation from several fungal classes, including the Ascomycota classes, Eurotiomycetes, Sordariomycetes, Saccharomycetes and the Basidiomycota orders, Pucciniomycetes and	fungidb	http://identifiers.org/fungidb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100464	http://fungidb.org/gene/{$id}	FungiDB at University of California	Department of Plant Pathology & Microbiology, University of California, Riverside, California	USA	false	CNBG_0001	2,143 to 5,911	http://FungiDB.org		
MIR:00000367	DRSC	Drosophila	^DRSC\d+$	The DRSC (Drosophila RNAi Screening Cente) tracks both production of reagents for RNA interference (RNAi) screening in Drosophila cells and RNAi screen results. It maintains a list of Drosophila gene names, identifiers, symbols and synonyms and provides information for cell-based or	drsc	http://identifiers.org/drsc	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		FlyRNAi
								MIR:00100466	http://www.flyrnai.org/cgi-bin/RNAi_gene_lookup_public.pl?gname={$id}	DRSC at Harvard Medical School	Department of Genetics, Harvard Medical School, Boston, Massachusetts	USA	false	DRSC05221	Longitudinals lacking	http://flyrnai.org/		Drosophila RNAi Screening Cente
MIR:00000368	OriDB Schizosaccharomyces	Yeast	^\d+$	OriDB is a database of collated genome-wide mapping studies of confirmed and predicted replication origin sites in Saccharomyces cerevisiae and the fission yeast Schizosaccharomyces pombe. This collection references Schizosaccharomyces pombe.	oridb.schizo	http://identifiers.org/oridb.schizo	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100467	http://pombe.oridb.org/details.php?id={$id}	OriDB Schizosaccharomyces at University of Nottingham	Centre for Genetics and Genomics, The University of Nottingham, Queen's Medical Centre, Nottingham	UK	false	1	tandem intergenic	http://pombe.oridb.org/index.php		
MIR:00000369	OriDB Saccharomyces	Yeast	^\d+$	OriDB is a database of collated genome-wide mapping studies of confirmed and predicted replication origin sites in Saccharomyces cerevisiae and the fission yeast Schizosaccharomyces pombe. This collection references Saccharomyces cerevisiae.	oridb.sacch	http://identifiers.org/oridb.sacch	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100468	http://cerevisiae.oridb.org/details.php?id={$id}	OriDB Saccharomyces at University of Nottingham	Centre for Genetics and Genomics, The University of Nottingham, Queen's Medical Centre, Nottingham	UK	false	1	not chromosomally active	http://cerevisiae.oridb.org/index.php		
MIR:00000371	SCOP	All species	^\d+$	The SCOP (Structural Classification of Protein) database is a comprehensive ordering of all proteins of known structure according to their evolutionary, functional and structural relationships. The basic classification unit is the protein domain. Domains are hierarchically classified into species, proteins, families, superfamilies, folds, and classes.	scop	http://identifiers.org/scop	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Structural Classification of Proteins
								MIR:00100470	http://scop.mrc-lmb.cam.ac.uk/scop/search.cgi?sunid={$id}	SCOP at MRC	MRC Laboratory of Molecular Biology, Centre for Protein Engineering, Hills Road, Cambridge	UK	false	47419	HTH repressors	http://scop.mrc-lmb.cam.ac.uk/scop/		
								MIR:00100471	http://scop.berkeley.edu/sunid={$id}	SCOP at Berkeley	University of California, Berkeley	USA	false	47419	HTH repressors	http://scop.berkeley.edu/		
MIR:00000372	ENA	All species	^[A-Z]+[0-9]+(\.\d+)?$	The European Nucleotide Archive (ENA) captures and presents information relating to experimental workflows that are based around nucleotide sequencing. ENA is made up of a number of distinct databases that includes EMBL-Bank, the Sequence Read Archive (SRA) and the Trace Archive each with their own data formats and standards. This collection references Embl-Bank identifiers.	ena.embl	http://identifiers.org/ena.embl	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	European Nucleotide Archive
								MIR:00100473	https://www.ebi.ac.uk/ena/data/view/{$id}	ENA at European Bioinformatics Institute	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	BN000065	Homo sapiens SMP1 gene	https://www.ebi.ac.uk/ena/	Embl-Bank
								MIR:00100889	https://www.ncbi.nlm.nih.gov/nuccore/{$id}	ENA through GenBank	National Center for Biotechnology Information (NCBI)	USA	false		BN000065		https://www.ncbi.nlm.nih.gov/Genbank/	
MIR:00000374	PANTHER Node	All species	^PTN\d{9}$	The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System is a resource that classifies genes by their functions, using published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct	panther.node	http://identifiers.org/panther.node	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100475	http://www.pantree.org/node/annotationNode.jsp?id={$id}	PANTHER Node at USC (Los Angeles)	Keck School of Medicine, University of Southern California	USA	false	PTN000000026	PHOSPHOSERINE PHOSPHATASE	http://pantree.org/		
MIR:00000375	Consensus CDS	human and mouse	^CCDS\d+\.\d+$	The Consensus CDS (CCDS) project is a collaborative effort to identify a core set of human and mouse protein coding regions that are consistently annotated and of high quality. The CCDS set is calculated following coordinated whole genome annotation updates carried out	ccds	http://identifiers.org/ccds	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	CCDS
								MIR:00100477	http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi?REQUEST=CCDS&DATA={$id}	Consensus CDS at NCBI	National Center for Biotechnology Information (NCBI)	USA	true	ncbi	CCDS13573.1	ENST00000352957	http://www.ncbi.nlm.nih.gov/CCDS/	
MIR:00000376	Locus Reference Genomic	human	^LRG_\d+$	Locus Reference Genomic (LRG) provides identifiers to stable genomic DNA sequences for regions of the human genome, providing a recognized reference-sequence standard for reporting sequence variants. LRG is maintained by the NCBI and the European Bioinformatics Institute (EBI).	lrg	http://identifiers.org/lrg	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		LRG
								MIR:00100478	ftp://ftp.ebi.ac.uk/pub/databases/lrgex/{$id}.xml	Locus Reference Genomic at University of Leicester	Department of Genetics, University of Leicester, Leicester	UK	false	LRG_1	Osteogenesis Imperfecta	http://www.lrg-sequence.org/		
								MIR:00100627	http://www.ensembl.org/Homo_sapiens/LRG/Summary?lrg={$id}	Locus Reference Genomic through Ensembl	EnSembl, Sanger Institute and European Bioinformatics Institute, Hinxton, Cambridge	UK	false	LRG_1	Chromosome 17: 50,182,096-50,206,639	http://www.ensembl.org/		
								MIR:00100629	http://asia.ensembl.org/Homo_sapiens/LRG/Summary?lrg={$id}	Locus Reference Genomic through Ensembl mirror (asia)		Singapore	false	LRG_1	Chromosome 17: 50,182,096-50,206,639	http://asia.ensembl.org/		
								MIR:00100630	http://uswest.ensembl.org/Homo_sapiens/LRG/Summary?lrg={$id}	Locus Reference Genomic through Ensembl mirror (US west)		USA	false	LRG_1	Chromosome 17: 50,182,096-50,206,639	http://uswest.ensembl.org/		
								MIR:00100631	http://useast.ensembl.org/Homo_sapiens/LRG/Summary?lrg={$id}	Locus Reference Genomic through Ensembl mirror (US east)		USA	false	LRG_1	Chromosome 17: 50,182,096-50,206,639	http://useast.ensembl.org/		
MIR:00000377	HPRD	human	^\d+$	The Human Protein Reference Database (HPRD) represents a centralized platform to visually depict and integrate information pertaining to domain	hprd	http://identifiers.org/hprd	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Human Protein Reference Database

MIR:00000377	HPRD	human	^\d+$	information pertaining to domain architecture, post-translational modifications, interaction networks and disease association for each protein in the	hprd	http://identifiers.org/hprd	0	MIR:00100479	http://www.hprd.org/protein/{$id}	Human Protein Reference Database	Institute of Bioinformatics, International Tech Park, Bangalore	India	false	1	Aldehyde dehydrogenase 1	http://www.hprd.org/		
MIR:00000378	GXA Gene	All species	^\w+$	The Gene Expression Atlas (GXA) is a semantically enriched database of meta-analysis based summary statistics over a curated subset of ArrayExpress Archive, servicing queries for condition-specific gene expression patterns as well as broader exploratory searches for	gxa.gene	http://identifiers.org/gxa.gene	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Gene Expression Atlas Gene
								MIR:00100482	https://www.ebi.ac.uk/gxa/genes/{$id}	GXA Gene at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ebi	AT4G01080	Protein trichome birefringence-like 26	https://www.ebi.ac.uk/gxa/	
MIR:00000379	GXA Expt	All species	^[AEP]-\w{4}-\d+$	The Gene Expression Atlas (GXA) is a semantically enriched database of meta-analysis based summary statistics over a curated subset of ArrayExpress Archive, servicing queries for condition-specific gene expression patterns as well as broader exploratory searches for biologically interesting genes/samples. This collection references experiments.	gxa.expt	http://identifiers.org/gxa.expt	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Gene Expression Atlas Experiment
								MIR:00100483	https://www.ebi.ac.uk/gxa/experiments/{$id}	GXA Expt at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	E-MTAB-2037	Oryza sativa	https://www.ebi.ac.uk/gxa/	
								MIR:00100857	https://www.omicsdi.org/dataset/atlas-experiments/{$id}	GXA Expt through OmicsDI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	omicsdi	E-MTAB-2037		https://www.omicsdi.org/	
MIR:00000385	PhosphoPoint Kinase	human	^\w+$	PhosphoPOINT is a database of the human kinase and phospho-protein interactome. It describes the interactions among kinases, their potential substrates and their interacting (phospho)-proteins. It also incorporates gene expression and uses gene ontology (GO) terms to	phosphopoint.kinase	http://identifiers.org/phosphopoint.kinase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100499	http://kinase.bioinformatics.tw/showall.jsp?type=Kinase&info=Gene&name={$id}&drawing=0&sorting=0&kinome=1	PhosphoPoint Kinase at National Taiwan University	Department of Computer Science and Information Engineering, National Taiwan University	Republic of China	false	AURKA	STK6_HUMAN	http://kinase.bioinformatics.tw/		
MIR:00000386	PhosphoPoint Phosphoprotein	human	^\w+$	PhosphoPOINT is a database of the human kinase and phospho-protein interactome. It describes the interactions among kinases, their potential substrates and their interacting (phospho)-proteins. It also incorporates gene expression and uses gene ontology (GO) terms to	phosphopoint.protein	http://identifiers.org/phosphopoint.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100500	http://kinase.bioinformatics.tw/showall.jsp?type=PhosphoProtein&info=Gene&name={$id}&drawing=0&sorting=0&kinome=0	PhosphoPoint Phosphoprotein at National Taiwan University	Department of Computer Science and Information Engineering, National Taiwan University	Republic of China	false	AURKA	20q13.2-q13.3	http://kinase.bioinformatics.tw/		
MIR:00000388	UniProt Isoform	All species	^([A-N,R-Z][0-9][A-Z][A-Z, 0-9][A-Z, 0-9][0-9])|([O,P,Q][0-9][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9])(\-\d+)$	The UniProt Knowledgebase (UniProtKB) is a comprehensive resource for protein sequence and functional information with extensive cross-references to more than 120 external databases. This collection is a subset of UniProtKB, and provides a means to reference isoform information.	uniprot.isoform	http://identifiers.org/uniprot.isoform	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		UniProtKB Isoform
								MIR:00100502	http://www.uniprot.org/uniprot/{$id}	UniProt Isoform through Universal Protein Resource	UniProt Consortium	USA, UK and Switzerland	false	Q5BJF6-3	Cenexin 1	http://www.uniprot.org/		
								MIR:00100503	http://purl.uniprot.org/uniprot/{$id}	UniProt Isoform through Universal Protein Resource using Persistent URL system	UniProt Consortium	USA, UK and Switzerland	false	Q5BJF6-3	Cenexin 1	http://purl.uniprot.org/		
								MIR:00100504	http://www.uniprot.org/uniparc/?query={$id}	UniProt Isoform through UniParc	UniProt Consortium	USA, UK and Switzerland	false	Q5BJF6-3	synthetic construct	http://www.uniprot.org/uniparc/		
MIR:00000392	HGMD	human	^[A-Z_0-9]+$	The Human Gene Mutation Database (HGMD) collates data on germ-line mutations in nuclear genes associated with human inherited disease. It includes information on single base-pair substitutions in coding, regulatory and splicing-relevant regions; micro-deletions	hgmd	http://identifiers.org/hgmd	0	id	accessURL	info	institution	location	official	localId	resourceURL			Human Gene Mutation Database
								MIR:00100512	http://www.hgmd.cf.ac.uk/ac/gene.php?gene={$id}	HGMD at Cardiff University	Cardiff University, Cardiff	Wales	false	CALM1	http://www.hgmd.cf.ac.uk/ac/index.php			
MIR:00000393	AphidBase Transcript	aphids	^ACYPI\d{6}(-RA)?$	AphidBase is a centralized bioinformatic resource that was developed to facilitate community annotation of the pea aphid genome by the International Aphid Genomics Consortium (IAGC). The AphidBase Information System was designed to organize and distribute	aphidbase.transcript	http://identifiers.org/aphidbase.transcript	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100513	http://bipaa.genouest.org/apps/grs-2.3/grs?reportID=aphidbase_transcript_report&objectID={$id}	AphidBase at INRA	INRA UMR, Le Rheu	France	false	ACYPI000159	GL349778	http://www.aphidbase.com/aphidbase		
MIR:00000395	TreeFam	animal	^\w{1,2}\d+$	TreeFam is a database of phylogenetic trees of gene families found in animals. Automatically generated trees are curated, to create a curated resource that presents the accurate evolutionary history of all animal gene families, as well as reliable ortholog and paralog	treefam	http://identifiers.org/treefam	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100515	http://www.treefam.org/family/{$id}	TreeFam database	Beijing Genomics Institute, Beijing	China	false	TF101014	Cyclin T	http://www.treefam.org/		
MIR:00000397	Cube db	vertebrates	^[A-Za-z_0-9]+$	Cube-DB is a database of pre-evaluated results for detection of functional divergence in human/vertebrate protein families. It analyzes comparable taxonomical samples for all paralogues under consideration, storing functional specialization at the level of residues. The	cubedb	http://identifiers.org/cubedb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100517	http://epsf.bmad.bii.a-star.edu.sg/cube/db/data/{$id}/	Cube db at Bioinformatics Institute (Singapore)	Bioinformatics Institute, Singapore	China	false	AKR	AKR1B1	http://epsf.bmad.bii.a-star.edu.sg/cube/db/html/home.html		
MIR:00000398	IDEAL		^IID\d+$	IDEAL provides a collection of knowledge on experimentally verified intrinsically disordered proteins. It contains manual annotations by curators on intrinsically disordered regions, interaction regions to other molecules, post-translational modification sites, references and	ideal	http://identifiers.org/ideal	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Intrinsically Disordered proteins with Extensive Annotations and Literature
								MIR:00100518	http://idp1.force.cs.is.nagoya-u.ac.jp/IDEAL/idealItem.php?id={$id}	IDEAL at Maebashi Institute of Technology	Faculty of Engineering, Maebashi Institute of Technology, Maebashi	Japan	false	IID00001	Nuclear receptor	http://www.ideal.force.cs.is.nagoya-u.ac.jp/IDEAL/		
MIR:00000401	GOLD genome	All species	^[Gi|Gc]\d+$	The GOLD (Genomes OnLine Database)is a resource for centralized monitoring of genome and metagenome projects worldwide. It stores information on complete and ongoing projects, along with their associated metadata. This collection references the sequencing	gold.genome	http://identifiers.org/gold.genome	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100521	http://www.genomesonline.org/cgi-bin/GOLD/GOLDCards.cgi?goldstamp={$id}	GOLD genome at Department of Energy Joint Genome Institute	Department of Energy Joint Genome Institute, Microbial Genomics and Metagenomics Program, California	USA	false	Gi07796	halophilic archaeon	http://www.genomesonline.org/cgi-bin/GOLD/index.cgi		
MIR:00000406	VariO	All species	^VariO:\d+$	The Variation Ontology (VariO) is an ontology for the standardized, systematic description of effects, consequences and mechanisms of variations. It describes the effects of variations at the DNA, RNA and/or protein level.	vario	http://identifiers.org/vario	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	Variation Ontology
								MIR:00100526	http://www.variationontology.org/cgi-bin/amivario/term-details.cgi?term={$id}	VariO at Lund University	Department of Experimental Medical Science, Lund University	Sweden	false	VariO:0294	disease causing	http://www.variationontology.org/		
								MIR:00100566	http://purl.bioontology.org/ontology/VARIO/{$id}	VariO through BioPortal	National Center for Biomedical Ontology, Stanford	USA	false	VariO:0294	disease causing	http://bioportal.bioontology.org/ontologies/VARIO	bptl	
								MIR:00100677	https://www.ebi.ac.uk/ols/ontologies/vario/terms?obo_id={$id}	VariO through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	VariO:0294	disease causing	https://www.ebi.ac.uk/ols/ontologies/vario	ols	
MIR:00000408	EPD		^[A-Z-_0-9]+$	The Eukaryotic Promoter Database (EPD) is an annotated non-redundant collection of eukaryotic POL II promoters, for which the transcription start site has been determined experimentally. Access to promoter sequences is provided by pointers to positions in nucleotide	epd	http://identifiers.org/epd	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Eukaryotic Promoter Database
								MIR:00100528	http://epd.vital-it.ch/cgi-bin/query_result.pl?out_format=NICE&Entry_0={$id}	EPD at Swiss Institute of Bioinformatics	Swiss Institute of Bioinformatics (SIB), Geneva	Switzerland	false	TA_H3	Chromosomal protein	http://epd.vital-it.ch/		
MIR:00000409	RFAM	All species	^RF\d{5}$	The Rfam database is a collection of RNA families, each represented by multiple sequence alignments, consensus secondary structures and covariance models (CMs). The families in Rfam break down into three broad functional classes: non-coding RNA genes, structured cis-	rfam	http://identifiers.org/rfam	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	RNA Family Database
								MIR:00100686	https://rfam.xfam.org/family/{$id}	Rfam at EMBL-EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	RF00230	element that controls gene expression	https://rfam.xfam.org/	
MIR:00000410	Fungal Barcode	fungi	^\d+$	DNA barcoding is the use of short standardised segments of the genome for identification of species in all the Kingdoms of Life. The goal of the Fungal Barcoding site is to promote the DNA barcoding of fungi and other fungus-like organisms.	fbol	http://identifiers.org/fbol	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		International Fungal Working Group
								MIR:00100532	http://www.fungalbarcoding.org/BioloMICS.aspx?Table=Fungal barcodes&Rec={$id}&Fields=All&ExactMatch=T	Fungal Barcode at Fungal Biodiversity Centre	Fungal Biodiversity Centre, Utrecht	Netherlands	false	2224	Fusarium poae	http://www.fungalbarcoding.org/		FBOL
																		Fungal Barcode of Life
MIR:00000412	AspGD Locus	Aspergilus	^[A-Za-z_0-9]+$	The Aspergillus Genome Database (AspGD) is a repository for information relating to fungi of the genus Aspergillus, which includes organisms of clinical, agricultural and industrial importance. AspGD facilitates comparative genomics by providing a full-featured genomics	aspgd.locus	http://identifiers.org/aspgd.locus	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Aspergillus Genome Database Locus
								MIR:00100534	http://www.aspergillusgenome.org/cgi-bin/locus.pl?dbid={$id}	AspGD at Stanford Medical School	Department of Genetics, Stanford University Medical School, Stanford, California	USA	false	ASPL0000349247	histone methyltransferase activity	http://www.aspgd.org/		
MIR:00000413	AspGD Protein	Aspergilus	^[A-Za-z_0-9]+$	The Aspergillus Genome Database (AspGD) is a repository for information relating to fungi of the genus Aspergillus, which includes organisms of clinical, agricultural and industrial importance. AspGD facilitates comparative genomics by providing a full-featured genomics	aspgd.protein	http://identifiers.org/aspgd.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100535	http://www.aspergillusgenome.org/cgi-bin/protein/proteinPage.pl?dbid={$id}	AspGD Protein at Stanford Medical School	Department of Genetics, Stanford University Medical School, Stanford, California	USA	false	ASPL0000349247	histone methyltransferase activity	http://www.aspgd.org/		
MIR:00000415	AutDB	human	^[A-Z]+[A-Z-0-9]{2,}$	AutDB is a curated database for autism research. It is built on information extracted from the studies on molecular genetics and biology of Autism Spectrum Disorders (ASD). The four modules of AutDB include information on Human Genes, Animal models, Protein	autdb	http://identifiers.org/autdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Autism Database
								MIR:00100538	http://autism.mindspec.org/GeneDetail/{$id}	AutDB at MindSpec	MindSpec Inc., Fairfax, Virginia	USA	false	ADA	adenosine deaminase	http://autism.mindspec.org/autdb/		
MIR:00000416	BacMap Map	bacteria	^\w+(_)?\d+(\.\d+)?$	BacMap is an electronic, interactive atlas of fully sequenced bacterial genomes. It contains labeled, zoomable and searchable chromosome maps for sequenced prokaryotic (archaebacterial and eubacterial) species. Each map can be zoomed to the level of individual genes	bacmap.map	http://identifiers.org/bacmap.map	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100539	http://bacmap.wishartlab.com/maps/{$id}/index.html	BacMap Genome Map at University of Alberta	Department of Computing Science, Food and Nutritional Science, University of Alberta, Edmonton	Canada	false	AP011135	urease accessory protein UreF	http://bacmap.wishartlab.com/		
MIR:00000417	Bgee family	animal	^(ENSFM|ENSGTV:)\d+$	Bgee is a database of gene expression patterns within particular anatomical structures within a species, and between different animal species. This collection refers to expression across species.	bgee.family	http://identifiers.org/bgee.family	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100540	http://bgee.unil.ch/bgee/bgee?page=gene_family&action=family_details&gene_family_id={$id}	Bgee at Lausanne	Swiss Institute of Bioinformatics, Lausanne	Switzerland	false	ENSFM00500000270089	HOMEOBOX HOX	http://bgee.unil.ch/bgee/bgee		
MIR:00000418	Bgee gene	animal	^[A-Za-z]+\d+$	Bgee is a database to retrieve and compare gene expression patterns in multiple species, produced from multiple data types (RNA-Seq, Affymetrix, in situ hybridization, and EST data). This collection references genes in Bgee.	bgee.gene	http://identifiers.org/bgee.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100541	https://bgee.org/?page=gene&gene_id={$id}	Bgee gene	SIB Swiss Institute of Bioinformatics and University of Lausanne	Switzerland	false	FBgn0000015	expression in 21 developmental stages	https://bgee.org/		
MIR:00000422	PANTHER Pathway Component	All species	^G|P|U|C|S\d{5}$	The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System is a resource that classifies genes by their functions, using published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct	panther.pthcmp	http://identifiers.org/panther.pthcmp	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100545	http://www.pantherdb.org/pathway/pathCatDetail.do?clsAccession={$id}	PANTHER Pathway Component at USC (Los Angeles)	Keck School of Medicine, University of Southern California	USA	false	P00266	MAP kinase superfamily	http://www.pantherdb.org/		
MIR:00000428	DEPOD		^[A-Z0-9]+$	The human DEPhOsphorylation Database (DEPOD) contains information on known human active phosphatases and their experimentally verified protein and	depod	http://identifiers.org/depod	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		human DEPhOsphorylation Database

MIR:00000428	DEPOD		^[A-Z0-9]+$	experimentally verified protein and nonprotein substrates. Reliability scores are provided for dephosphorylation	depod	http://identifiers.org/depod	0	MIR:00100551	http://www.koehnlab.de/depod/showp.php?gene={$id}	DEPOD at EMBL	European Molecular Biology Laboratory EMBL, Heidelberg	Germany	false	PTPN1	Tyrosine-protein phosphatase	http://www.koehnlab.de/depod/		
MIR:00000432	Phytozome Locus	Plant species	^[A-Za-z0-9]+$	Phytozome is a project to facilitate comparative genomic studies amongst green plants. Famlies of orthologous and paralogous genes that represent the modern descendents of ancestral gene sets are constructed at key phylogenetic	phytozome.locus	http://identifiers.org/phytozome.locus	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100556	http://www.phytozome.net/genePage.php?crown&method=0&search=1&detail=1&searchText=locusname:{$id}	Phytozome Locus at Joint Genome Institute	Joint Genome Institute, California, and the Center for Integrative Genomics, Lausanne	USA and Switzerland	false	Glyma0021s00410	Histones H3 and H4	http://www.phytozome.net/		
MIR:00000433	SubtiList	bacillus subtilis	^BG\d+$	SubtiList serves to collate and integrate various aspects of the genomic information from B. subtilis, the paradigm of sporulating Gram-positive bacteria. SubtiList provides a complete dataset of DNA and protein sequences derived from	subtilist	http://identifiers.org/subtilist	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100557	http://genolist.pasteur.fr/SubtiList/genome.cgi?external_query+{$id}	SubtiList at Pasteur Institute	Pasteur Institute, Paris	France	false	BG11523	naringenin-chalcone	http://genolist.pasteur.fr/SubtiList/		
MIR:00000437	WikiGenes	All species	^\d+$	WikiGenes is a collaborative knowledge resource for the life sciences, which is based on the general wiki idea but employs specifically developed technology to serve as a rigorous scientific tool. The rationale behind WikiGenes is to provide	wikigenes	http://identifiers.org/wikigenes	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100567	http://www.wikigenes.org/e/gene/e/{$id}.html	WikiGenes at Massachusetts Institute of Technology	Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts	USA	false	3771877	Alcohol dehydrogenase	http://www.wikigenes.org/		
MIR:00000438	Broad Fungal Genome Initiative	fungi	^S\d+$	Magnaporthe grisea, the causal agent of rice blast disease, is one of the most devasting threats to food security worldwide and is a model organism for studying fungal phytopathogenicity and host-parasite interactions. The	broad	http://identifiers.org/broad	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Magnaporthe comparative genomics database
								MIR:00100569	http://www.broadinstitute.org/annotation/genome/magnaporthe_grisea/GeneDetails.html?sp={$id}	Broad Fungal Genome Initiative at Broad Institute	Broad Institute of MIT and Harvard, Cambridge, Massachuchetts	USA	false	S7000002168151102	dicarboxylic amino acid permease	http://www.broadinstitute.org/annotation/genome/magnaporthe_grisea/		
MIR:00000442	EcoliWiki	E.coli	^[A-Za-z0-9-]+$	EcoliWiki is a wiki-based resource to store information related to non-pathogenic E. coli, its phages, plasmids, and mobile genetic elements. This collection references genes.	ecoliwiki	http://identifiers.org/ecoliwiki	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100573	http://ecoliwiki.net/colipedia/index.php/{$id}:Gene	EcoliWiki at Texas Agrilife Research	Texas Agrilife Research, Texas A&M University College Station	USA	false	aaeA	Aromatic (carboxylic) acid efflux	http://ecoliwiki.net/colipedia/		
MIR:00000448	VBRC	virus	^\d+$	The VBRC provides bioinformatics resources to support scientific research directed at viruses belonging to the Arenaviridae, Bunyaviridae, Filoviridae, Flaviviridae, Paramyxoviridae, Poxviridae, and Togaviridae families. The Center	vbrc	http://identifiers.org/vbrc	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Viral Bioinformatics Resource Center
								MIR:00100582	http://vbrc.org/gene_detail.asp?gene_id={$id}	VBRC at University of Alabama	University of Alabama, Birmingham and the University of Victoria, British Columbia	USA and Canada	false	35742	mp_NS3	http://vbrc.org/		
MIR:00000449	ViralZone	virus	^\d+$	ViralZone is a resource bridging textbook knowledge with genomic and proteomic sequences. It provides fact sheets on all known virus families/genera with easy access to sequence data. A selection of reference strains (RefStrain) provides	viralzone	http://identifiers.org/viralzone	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100583	http://viralzone.expasy.org/all_by_protein/{$id}.html	ViralZone at SIB	Swiss Institute of Bioinformatics, Centre Médical Universitaire, Geneva	Switzerland	false	992	virus membrane	http://www.expasy.org/viralzone/		
MIR:00000450	Gene Ontology Reference	All species	^GO_REF:\d{7}$	The GO reference collection is a set of abstracts that can be cited in the GO ontologies (e.g. as dbxrefs for term definitions) and annotation files (in the Reference column). It provides two types of reference; It can be used to provide	go.ref	http://identifiers.org/go.ref	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		GO reference collection
								MIR:00100584	http://www.geneontology.org/cgi-bin/references.cgi#{$id}	Gene Ontology Reference at The Gene Ontology Consortium	The Gene Ontology Consortium	USA	false	GO_REF:0000041	Gene Ontology annotation based on UniPathway vocabulary mapping	http://www.geneontology.org/cgi-bin/references.cgi		
MIR:00000451	Rat Genome Database qTL	rat	^\d+$	Rat Genome Database seeks to collect, consolidate, and integrate rat genomic and genetic data with curated functional and physiological data and make these data widely available to the scientific community. This collection references	rgd.qtl	http://identifiers.org/rgd.qtl	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100586	http://rgd.mcw.edu/rgdweb/report/qtl/main.html?id={$id}	Rat Genome Database qTL at Medical College of Wisconsin	Medical College of Wisconsin, Milwaukee, Wisconsin	USA	false	1354581	Blood pressure QTL 247	http://rgd.mcw.edu/		
MIR:00000453	DOOR	procariotes	^\d+$	DOOR (Database for prOkaryotic OpeRons) contains computationally predicted operons of all the sequenced prokaryotic genomes. It includes operons for RNA genes.	door	http://identifiers.org/door	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Database for prOkaryotic OpeRons
								MIR:00100588	http://csbl1.bmb.uga.edu/OperonDB/genedetail.php?id={$id}	DOOR v1 at University of Georgia	Computational Systems Biology Laboratory, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia	USA	false	60101	MTARETLMPVLLEKVYQLIQDKLELAQQPLVTQLGQHLFSNISQDDLVERNESDLYGAVLSLWHHINEKK	http://csbl1.bmb.uga.edu/OperonDB/DOOR.php		
MIR:00000454	Degradome Database	mamalian	^[AMCST][0-9x][0-9]$	The Degradome Database contains information on the complete set of predicted proteases present in a a variety of mammalian species that have been subjected to whole genome sequencing. Each protease sequence is curated and,	degradome	http://identifiers.org/degradome	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100589	http://degradome.uniovi.es/cgi-bin/protease/{$id}	Degradome Database at	Departamento de Bioqu&amp;iacute;mica y Biolog&amp;iacute;a Molecular, Facultad de Medicina, Universidad de Oviedo	Spain	false	Ax1	seminal vesicle antigen-like 1	http://degradome.uniovi.es/		
MIR:00000455	DBD	All species	^\d+$	The DBD (transcription factor database) provides genome-wide transcription factor predictions for organisms across the tree of life. The prediction method identifies sequence-specific DNA-binding transcription factors through homology	dbd	http://identifiers.org/dbd	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Transcription Factor Database
								MIR:00100590	http://www.transcriptionfactor.org/index.cgi?Search/Domain+domain:{$id}+cat:DBD	DBD at MRC Laboratory of Molecular Biology	MRC Laboratory of Molecular Biology, Cambridge	UK	false	45310	CC1G_00911.1	http://www.transcriptionfactor.org/		
MIR:00000456	DATF	Arabidopsis	^AT[1-5]G\d{5}(\.\d+)?$	DATF contains known and predicted Arabidopsis transcription factors (1827 genes in 56 families) with the unique information of 1177 cloned sequences and many other features including 3D structure templates, EST expression	datf	http://identifiers.org/datf	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Database of Arabidopsis transcription factors
								MIR:00100591	http://planttfdb.cbi.pku.edu.cn/tf.php?sp=Ath&did={$id}	DATF through PlantTFDB	Center for Bioinformatics, Peking	Peoples Republic of China.	false	AT1G01030.1	NGA3	http://datf.cbi.pku.edu.cn/		
MIR:00000459	YRC PDR	Yeast	^\d+$	The Yeast Resource Center Public Data Repository (YRC PDR) serves as a single point of access for the experimental data produced from many collaborations typically studying Saccharomyces cerevisiae (baker's yeast). The	yrcpdr	http://identifiers.org/yrcpdr	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Yeast Resource Center Public Data Repository
								MIR:00100594	http://yeastrc.org/pdr/viewProtein.do?id={$id}	YRC PDR at University of Washington	Department of Biochemistry, University of Washington, Seattle	USA	false	2673500	PKHD-type hydroxylase	http://www.yeastrc.org/pdr/		
MIR:00000460	Yeast Intron Database v3	Yeast	^[A-Z0-9]+$	The Yeast Intron Database (version 3) contains information on the spliceosomal introns of the yeast Saccharomyces cerevisiae. It includes expression data that relates to the efficiency of splicing relative to other processes in strains of yeast	yid	http://identifiers.org/yid	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100595	http://metarray.ucsc.edu/cgi-bin/intron/yirIntrondb?orfName={$id}	Yeast Intron Database version 3 at Baskin School of Engineering	Baskin School of Engineering, Center for Molecular Biology of RNA, University of California, Santa Cruz, California	USA	false	SNR17A	U3A snoRNA	http://compbio.soe.ucsc.edu/yeast_introns.html		
MIR:00000461	FuncBase Fly	Drosophila	^\d+$	Computational gene function prediction can serve to focus experimental resources on high-priority experimental tasks. FuncBase is a web resource for viewing quantitative machine learning-based gene function annotations. Quantitative	funcbase.fly	http://identifiers.org/funcbase.fly	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100596	http://func.mshri.on.ca/fly/genes/list_functional_scores/{$id}	FuncBase Fly at Harvard Medical School	Harvard Medical School, Boston, Massachusetts	USA	false	10194	Wnt5	http://func.mshri.on.ca/fly		
MIR:00000462	FuncBase Human	human	^\d+$	Computational gene function prediction can serve to focus experimental resources on high-priority experimental tasks. FuncBase is a web resource for viewing quantitative machine learning-based gene function annotations. Quantitative	funcbase.human	http://identifiers.org/funcbase.human	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100597	http://func.mshri.on.ca/human/genes/list_functional_scores/{$id}	FuncBase Human at Harvard Medical School	Harvard Medical School, Boston, Massachusetts	USA	false	119514	GALNT12	http://func.mshri.on.ca/human/		
MIR:00000463	FuncBase Mouse	mouse	^\d+$	Computational gene function prediction can serve to focus experimental resources on high-priority experimental tasks. FuncBase is a web resource for viewing quantitative machine learning-based gene function annotations. Quantitative	funcbase.mouse	http://identifiers.org/funcbase.mouse	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100598	http://func.mshri.on.ca/mouse/genes/list_functional_scores/{$id}	FuncBase Mouse at Harvard Medical School	Harvard Medical School, Boston, Massachusetts	USA	false	1351341	Grm4	http://func.mshri.on.ca/mouse/		
MIR:00000464	FuncBase Yeast	Yeast	^\d+$	Computational gene function prediction can serve to focus experimental resources on high-priority experimental tasks. FuncBase is a web resource for viewing quantitative machine learning-based gene function annotations. Quantitative	funcbase.yeast	http://identifiers.org/funcbase.yeast	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100599	http://func.mshri.on.ca/yeast/genes/list_functional_scores/{$id}	FuncBase Yeast at Harvard Medical School	Harvard Medical School, Boston, Massachusetts	USA	false	2701	SSD1	http://func.mshri.on.ca/yeast		
MIR:00000465	YDPM	Yeast	^Y[A-Z]{2}\d+[CW]$	The YDPM database serves to support the Yeast Deletion and the Mitochondrial Proteomics Project. The project aims to increase the understanding of mitochondrial function and biogenesis in the context of the cell. In the Deletion	ydpm	http://identifiers.org/ydpm	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Yeast Deletion Project and Proteomics of Mitochondria Database
								MIR:00100600	http://www-deletion.stanford.edu/cgi-bin/YDPM/YDPM_search.cgi?thelist={$id}	YDPM at Stanford University School of Medicine	Stanford University School of Medicine, Stanford, California	USA	false	YAL001C	TFIIIC (transcription initiation factor)	http://www-deletion.stanford.edu/YDPM/		
MIR:00000466	WormBase RNAi	worm	^WBRNAi\d{8}$	WormBase is an online bioinformatics database of the biology and genome of the model organism Caenorhabditis elegans and related nematodes. It is used by the C. elegans research community both as an information resource and as a	wb.rnai	http://identifiers.org/wb.rnai	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100601	https://www.wormbase.org/species/c_elegans/rnai/{$id}	WormBase RNAi	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	WBRNAi00086878	C. elegans	https://www.wormbase.org/		
MIR:00000467	FunCat	All species	^\d{2}|\d{2}\.\d{2}|\d{2}\.\d{2}\.\d{2}|\d{2}\.\d{2}\.\d{2}\d{2}$	The Functional Catalogue (FunCat) is a hierarchically structured, organism-independent, flexible and scalable controlled classification system enabling the functional description of proteins from any organism. It has been applied	funcat	http://identifiers.org/funcat	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		MIPS Functional Catalogue Database
								MIR:00100602	http://mips.helmholtz-muenchen.de/funcatDB/cgi-bin/search_advanced.pl?action=2&wert={$id}	FunCat at MIPS (Neuherberg)	Institute for Bioinformatics (MIPS), GSF National Research Center for Environment and Health, Neuherberg	Germany	false	32.01.01	oxidative stress response	http://mips.helmholtz-muenchen.de/funcatDB/		
MIR:00000469	ICEberg element	bacteria	^\d+$	ICEberg (integrative and conjugative elements) is a database of integrative and conjugative elements (ICEs) found in bacteria. ICEs are conjugative self-transmissible elements that can integrate into and excise from a host chromosome,	iceberg.element	http://identifiers.org/iceberg.element	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100604	http://db-mml.sjtu.edu.cn/ICEberg/feature_page.php?ice_id={$id}	ICEberg element at Shanghai Jiaotong University	State Key Laboratory of Microbial Metabolism and School of Life Sciences &amp;amp;amp; Biotechnology, Shanghai Jiaotong University, Shanghai	China	false	100	Tn5383	http://db-mml.sjtu.edu.cn/ICEberg/		
MIR:00000470	ICEberg family	bacteria	^\d+$	ICEberg (integrative and conjugative elements) is a database of integrative and conjugative elements (ICEs) found in bacteria. ICEs are conjugative self-transmissible elements that can integrate into and excise from a host chromosome,	iceberg.family	http://identifiers.org/iceberg.family	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100605	http://db-mml.sjtu.edu.cn/ICEberg/browse_result.php?type=fam&fam_id={$id}	ICEberg family at Shanghai Jiaotong University	State Key Laboratory of Microbial Metabolism and School of Life Sciences &amp; Biotechnology, Shanghai Jiaotong University, Shanghai	China	false	1	SXT/R391	http://db-mml.sjtu.edu.cn/ICEberg/		
MIR:00000472	VFDB Gene	bacteria	^\w+$	VFDB is a repository of virulence factors (VFs) of pathogenic bacteria.This collection references VF genes.	vfdb.gene	http://identifiers.org/vfdb.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100607	http://www.mgc.ac.cn/cgi-bin/VFs/gene.cgi?GeneID={$id}	VFDB Gene at Institute of Pathogen Biology	State Key Laboratory for Molecular Virology and Genetic Engineering, Institute of Pathogen Biology, Chinese Academy Medical Sciences and Peking Union Medical	China	false	VFG2154	Chlamydia trachomatis	http://www.mgc.ac.cn/VFs/		
MIR:00000473	MeSH 2013	All species	^[A-Za-z0-9]+$	MeSH (Medical Subject Headings) is the National Library of Medicine's controlled vocabulary thesaurus. It consists of sets of	mesh.2013	http://identifiers.org/mesh.2013	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		

MIR:00000473	MeSH 2013	All species	^[A-Za-z0-9]+$	terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. This thesaurus is used by NLM for indexing articles from	mesh.2013	http://identifiers.org/mesh.2013	0	MIR:00100608	http://www.nlm.nih.gov/cgi/mesh/2013/MB_cgi?mode=&index={$id}&view=expanded	MeSH 2013 at National Library of Medicine	U.S. National Library of Medicine, National Institute of Health, Maryland	USA	false	17165	Codon, Terminator	http://www.nlm.nih.gov/mesh/		
MIR:00000474	KEGG Module	All species	^([a-z]{3,5}_)?M\d{5}$	KEGG Modules are manually defined functional units used in the annotation and biological interpretation of sequenced genomes. Each module corresponds to a set of 'KEGG Orthology' (MIR:00000116) entries. KEGG Modules can represent pathway, structural	kegg.module	http://identifiers.org/kegg.module	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		KEGG
								MIR:00100609	http://www.kegg.jp/entry/{$id}	KEGG Module at Kyoto University Bioinformatics Center	Department of Computational Biology, University of Tokyo, Tokyo	Japan	true	M00002	Glycolysis	http://www.kegg.jp/kegg/module.html		
MIR:00000479	JCGGDB	animal	^JCGG-STR\d{6}$	JCGGDB (Japan Consortium for Glycobiology and Glycotechnology DataBase) is a database that aims to integrate all glycan-related data held in various repositories in Japan. This includes databases for large-quantity synthesis of glycogenes and glycans	jcggdb	http://identifiers.org/jcggdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Japan Consortium for Glycobiology and Glycotechnology DataBase
								MIR:00100614	http://jcggdb.jp/idb/jcggdb/{$id}	JCGGDB at Advanced Science Institute (Japan)	Advanced Science Institute(ASI), Chemical Biology Department Systems Glycobiology research Group, Saitama	Japan	false	JCGG-STR008690	GlyTouCan	http://jcggdb.jp/index_en.html		
MIR:00000480	NONCODE v4 Gene	All species	^NONHSAG\d{5}$	NONCODE is a database of expression and functional lncRNA (long noncoding RNA) data obtained from microarray studies. LncRNAs have been shown to play key roles in various biological processes such as imprinting control, circuitry controlling pluripotency and differentiation, immune	noncodev4.gene	http://identifiers.org/noncodev4.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100616	http://www.bioinfo.org/NONCODEv4/show_gene.php?id={$id}	NONCODE v4 Gene at Chinese Academy of Sciences	Bioinformatics Research Group, Chinese Academy of Sciences, Beijing	China	false	NONHSAG00001	11869	http://www.bioinfo.org/NONCODEv4/		
MIR:00000481	NONCODE v4 Transcript	All species	^NONHSAT\d{6}$	NONCODE is a database of expression and functional lncRNA (long noncoding RNA) data obtained from microarray studies. LncRNAs have been shown to play key roles in various biological processes such as imprinting control, circuitry controlling pluripotency and differentiation, immune	noncodev4.rna	http://identifiers.org/noncodev4.rna	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100617	http://www.bioinfo.org/NONCODEv4/show_rna.php?id={$id}	NONCODE v4 Transcript at Chinese Academy of Sciences	Bioinformatics Research Group, Chinese Academy of Sciences, Beijing	China	false	NONHSAT000001	11868	http://www.bioinfo.org/NONCODEv4/		
MIR:00000482	Oryzabase Gene	rice	^\d+$	Oryzabase provides a view of rice (Oryza sativa) as a model monocot plant by integrating biological data with molecular genomic information. It contains information about rice development and anatomy, rice mutants, and genetic resources, especially for wild varieties of	oryzabase.gene	http://identifiers.org/oryzabase.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100618	http://www.shigen.nig.ac.jp/rice/oryzabaseV4/gene/detail/{$id}	Oryzabase v4 Gene at National Institute of Genetics (Japan)	Genetic Strains Research Center, National Institute of Genetics, Mishima, Shizuoka	Japan	false	117	frizzy panicle, BRANCHED SILKLESS	http://www.shigen.nig.ac.jp/rice/oryzabaseV4/		
MIR:00000483	Oryzabase Mutant	rice	^\d+$	Oryzabase provides a view of rice (Oryza sativa) as a model monocot plant by integrating biological data with molecular genomic information. It contains information about rice development and anatomy, rice mutants, and genetic resources, especially for wild varieties of	oryzabase.mutant	http://identifiers.org/oryzabase.mutant	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100619	http://www.shigen.nig.ac.jp/rice/oryzabaseV4/strain/inducedMutationLine/detail/{$id}	Oryzabase v4 Mutant at National Institute of Genetics (Japan)	Genetic Strains Research Center, National Institute of Genetics, Mishima, Shizuoka	Japan	false	21393	CM1	http://www.shigen.nig.ac.jp/rice/oryzabaseV4/		
MIR:00000487	Gene Wiki	All species	^\d+$	The Gene Wiki is project which seeks to provide detailed information on human genes. Initial 'stub' articles are created in an automated manner, with further information added by the community. Gene Wiki can be accessed in wikipedia using Gene identifiers from NCBI.	genewiki	http://identifiers.org/genewiki	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100623	http://plugins.biogps.org/cgi-bin/wp.cgi?id={$id}	Gene Wiki through BioGPS	Wikimedia Foundation, San Francisco, California	USA	false	1017	protein in G1-S transition	http://en.wikipedia.org/wiki/Gene_Wiki		
MIR:00000489	PaxDb Protein	All species	^\d+$	PaxDb is a resource dedicated to integrating information on absolute protein abundance levels across different organisms. Publicly available experimental data are mapped onto a common namespace and, in the case of tandem mass spectrometry data, re-processed	paxdb.protein	http://identifiers.org/paxdb.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100625	http://pax-db.org/#!protein/{$id}	PaxDb v3 at University of Zurich	Institute of Molecular Life Sciences, University of Zurich, Zurich	Switzerland	false	977869	apolipoprotein C-I	http://pax-db.org/		
MIR:00000497	Homeodomain Research	human	^\d+$	The Homeodomain Resource is a curated collection of sequence, structure, interaction, genomic and functional information on the homeodomain family. It contains sets of curated homeodomain sequences from fully sequenced genomes, including experimentally derived	hdr	http://identifiers.org/hdr	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100638	http://research.nhgri.nih.gov/apps/homeodomain/web/index.cgi?mode=view&view=proteins&id={$id}	Homeodomain Research at National Human Genome Research Institute (NIH)	Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland	USA	false	63	Protein coding	http://research.nhgri.nih.gov/apps/homeodomain/web/		
MIR:00000498	NORINE	bacteria and fungi	^NOR\d+$	Norine is a database dedicated to nonribosomal peptides (NRPs). In bacteria and fungi, in addition to the traditional ribosomal proteic biosynthesis, an alternative ribosome-independent pathway called NRP synthesis allows peptide production. The molecules	norine	http://identifiers.org/norine	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100639	http://bioinfo.lifl.fr/norine/result.jsp?ID={$id}	NORINE at Computer Science Laboratory of Lille	Computer Science Laboratory of Lille, INRIA and ProBioGEM, University of Sciences and Technologies of Lille, Lille	France	false	NOR00681	Vancomycin is a glycopeptide antibiotic	http://bioinfo.lifl.fr/norine/		
MIR:00000499	Olfactory Receptor Database	animal , fungi and bacteria	^\d+$	The Olfactory Receptor Database (ORDB) is a repository of genomics and proteomics information of olfactory receptors (ORs). It includes a broad range of chemosensory genes and proteins, that includes in addition to ORs the taste papilla receptors (TPRs), vomeronasal	ordb	http://identifiers.org/ordb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		OrDB
								MIR:00100640	http://senselab.med.yale.edu/ORDB/Data/{$id}	Olfactory Receptor Database at Yale University School of Medicine	Center for Medical Informatics, Yale University School of Medicine, New Haven, Connecticut	USA	false	8497	ATGAGGGGGGAG	http://senselab.med.yale.edu/OrDB/		
MIR:00000501	P3DB Protein	Plant species	^\d+$	Plant Protein Phosphorylation DataBase (P3DB) is a database that provides information on experimentally determined phosphorylation sites in the proteins of various plant species. This collection references plant proteins that contain phosphorylation sites.	p3db.protein	http://identifiers.org/p3db.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100642	http://www.p3db.org/protein.php?id={$id}&ref=0	P3DB Protein at University of Missouri	Department of Computer Science, University of Missouri, Columbia, Missouri	USA	false	70	receptor-like kinase 2	http://www.p3db.org/		
MIR:00000502	P3DB Site	Plant species	^\d+$	Plant Protein Phosphorylation DataBase (P3DB) is a database that provides information on experimentally determined phosphorylation sites in the proteins of various plant species. This collection references phosphorylation sites in proteins.	p3db.site	http://identifiers.org/p3db.site	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100643	http://www.p3db.org/phosphosite.php?id={$id}&ref=0	P3DB Site at University of Missouri	Department of Computer Science, University of Missouri, Columbia, Missouri	USA	false	65	somatic embryogenesis	http://www.p3db.org/		
MIR:00000503	TOPDB	All species	^[A-Z0-9]+$	The Topology Data Bank of Transmembrane Proteins (TOPDB) is a collection of transmembrane protein datasets containing experimentally derived topology information. It contains information gathered from the literature and from public databases available	topdb	http://identifiers.org/topdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100644	http://topdb.enzim.hu/?m=show&id={$id}	TOPDB at Hungarian Academy of Sciences	Institute of Enzymology, Hungarian Academy of Sciences, Budapest	Hungary	false	AP00378	MVGLTTLFWL	http://topdb.enzim.hu/		
MIR:00000504	Animal Genome Cattle QTL	cattle	^\d+$	The Animal Quantitative Trait Loci (QTL) database (Animal QTLdb) is designed to house publicly all available QTL and single-nucleotide polymorphism/gene association data on livestock animal species. This collection references cattle QTLs.	cattleqtldb	http://identifiers.org/cattleqtldb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100645	http://www.animalgenome.org/cgi-bin/QTLdb/BT/qdetails?QTL_ID={$id}	Animal QTL Cattle at Iowa State University	Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Iowa	USA	false	4685	Holstein-Friesian cattle	http://www.animalgenome.org/cgi-bin/QTLdb/BT/index		
MIR:00000505	Animal Genome Chicken QTL	chicken	^\d+$	The Animal Quantitative Trait Loci (QTL) database (Animal QTLdb) is designed to house publicly all available QTL and single-nucleotide polymorphism/gene association data on livestock animal species. This collection references chicken QTLs.	chickenqtldb	http://identifiers.org/chickenqtldb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100646	http://www.animalgenome.org/cgi-bin/QTLdb/GG/qdetails?QTL_ID={$id}	Animal QTL Chicken at Iowa State University	Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Iowa	USA	false	14362	cross between Xinghua and White Recessive Rock	http://www.animalgenome.org/cgi-bin/QTLdb/GG/index		
MIR:00000506	Animal Genome Pig QTL	Pig	^\d+$	The Animal Quantitative Trait Loci (QTL) database (Animal QTLdb) is designed to house publicly all available QTL and single-nucleotide polymorphism/gene association data on livestock animal species. This collection references pig QTLs.	pigqtldb	http://identifiers.org/pigqtldb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100647	http://www.animalgenome.org/cgi-bin/QTLdb/SS/qdetails?QTL_ID={$id}	Animal QTL Pig at Iowa State University	Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Iowa	USA	false	14	back-cross of Meishan x White composite	http://www.animalgenome.org/cgi-bin/QTLdb/SS/index		
MIR:00000507	Animal Genome Sheep QTL	Sheep	^\d+$	The Animal Quantitative Trait Loci (QTL) database (Animal QTLdb) is designed to house publicly all available QTL and single-nucleotide polymorphism/gene association data on livestock animal species. This collection references sheep QTLs.	sheepqtldb	http://identifiers.org/sheepqtldb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100648	http://www.animalgenome.org/cgi-bin/QTLdb/OA/qdetails?QTL_ID={$id}	Animal QTL Sheep at Iowa State University	Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Iowa	USA	false	19803	Red Maasai and Dorper	http://www.animalgenome.org/cgi-bin/QTLdb/OA/index		
MIR:00000512	European Genome-phenome Archive Dataset	human	^EGAD\d{11}$	The EGA is a service for permanent archiving and sharing of all types of personally identifiable genetic and phenotypic data resulting from biomedical research projects. The EGA contains exclusive data collected from individuals whose consent agreements authorize data release only for specific research use or to bona fide researchers. Strict protocols govern how information is	ega.dataset	http://identifiers.org/ega.dataset	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	EGA Dataset
								MIR:00100658	https://www.ebi.ac.uk/ega/datasets/{$id}	EGA Dataset at European Bioinformatics Institute	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	EGAD00000000001	WTCCC case-control study for Bipolar Disorder	https://www.ebi.ac.uk/ega/dataset	
								MIR:00100852	https://www.omicsdi.org/dataset/ega/{$id}	EGA Dataset through OmicsDI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	omicsdi	EGAD00000000001		https://www.omicsdi.org/	
MIR:00000515	PRIDE Project		^P(X|R)D\d+{6}$	The PRIDE PRoteomics IDEntifications database is a centralized, standards compliant, public data repository that provides protein and peptide identifications together with supporting evidence. This collection references projects.	pride.project	http://identifiers.org/pride.project	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100662	https://www.ebi.ac.uk/pride/archive/projects/{$id}	PRIDE Project at EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ebi	PXD000440	Qualitatively Different T Cell Phenotypic Responses	https://www.ebi.ac.uk/pride/	
								MIR:00100858	https://www.omicsdi.org/dataset/pride/{$id}	PRIDE Project through OmicsDI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	omicsdi	PXD000440		https://www.omicsdi.org/	
MIR:00000520	HGNC Family	human	^[A-Z0-9-]+(#[A-Z0-9-]+)?$	The HGNC (HUGO Gene Nomenclature Committee) provides an approved gene name and symbol (short-form abbreviation) for each known human gene. All approved symbols are stored in the HGNC database, and each symbol is unique. In addition, HGNC also provides	hgnc.family	http://identifiers.org/hgnc.family	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100671	https://www.genenames.org/genefamilies/{$id}	HGNC Family at HUGO Genome Nomenclature Committee	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	PADI	Peptidyl arginine deiminases	https://www.genenames.org/		
MIR:00000521	Yeast Intron Database v4.3	Yeast	^[A-Z0-9]+$	The Yeast Intron Database (version 4.3) contains information on the spliceosomal introns of the yeast Saccharomyces cerevisiae. It includes expression data that relates to the efficiency of splicing relative to other processes in strains of yeast lacking nonessential splicing factors. The	yeastintron	http://identifiers.org/yeastintron	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100673	http://intron.ucsc.edu/cgi-bin/yeast4.3/intronreports.pl?outputformat=full&orfName={$id}	Yeast Intron Database version 4.3 at Baskin School of Engineering	Baskin School of Engineering, Center for Molecular Biology of RNA, University of California, Santa Cruz, California	USA	false	SNR17A	Oct 12, 1998	http://intron.ucsc.edu/yeast4.3/		
MIR:00000522	Antibiotic Resistance Genes Database	human	^[A-Z_]{3}[0-9]{4,}$	The Antibiotic Resistance Genes Database (ARDB) is a manually curated database which characterises genes involved in antibiotic resistance. Each gene and resistance type is annotated with information, including resistance profile, mechanism of action, ontology, COG and	ardb	http://identifiers.org/ardb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		ARDB
								MIR:00100678	http://ardb.cbcb.umd.edu/cgi/search.cgi?db=L&field=ni&term={$id}	ARDB at University of Maryland	Center for Bioinformatics and Computational Biology, University of Maryland, Maryland	USA	false	CAE46076	Ribosomal protection	http://ardb.cbcb.umd.edu/		
MIR:00000524	ProteomicsDB Protein	human	^\d+$	ProteomicsDB is an effort dedicated to expedite the identification of the human proteome and its use across the scientific community. This human proteome data is assembled primarily using information from liquid chromatography tandem-mass-spectrometry (LC-MS/MS) experiments.	proteomicsdb.protein	http://identifiers.org/proteomicsdb.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100680	https://www.proteomicsdb.org/#human/proteinDetails/{$id}/summary	Proteomics DB Protein at Center for Integrated Protein Science	Center for Integrated Protein Science, Munich,	Germany	false	53504	Calmodulin-binding	https://www.proteomicsdb.org/#human		
MIR:00000525	ProteomicsDB Peptide	human	^\d+$	ProteomicsDB is an effort dedicated to expedite the identification of the human proteome and its use across the scientific community. This human proteome data is	proteomicsdb.peptide	http://identifiers.org/proteomicsdb.peptide	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		

MIR:00000525	ProteomicsDB Peptide	human	^\d+$	community. This human proteome data is assembled primarily using information from liquid chromatography tandem-mass-spectrometry (LC-MS/MS) experiments	proteomicsdb.peptide	http://identifiers.org/proteomicsdb.peptide	0	MIR:00100681	https://www.proteomicsdb.org/#human/proteinDetails/{$id}/peptides/	Proteomics DB Peptide at Center for Integrated Protein Science	Center for Integrated Protein Science, Munich	Germany	false	53504	Neuromodulin	https://www.proteomicsdb.org/#peptideSearch		
MIR:00000526	Human Proteome Map Protein	human	^\d+$	The Human Proteome Map (HPM) portal integrates the peptide sequencing result from the draft map of the human proteome project. The project was based on LC-MS/MS by utilizing of high resolution and high accuracy Fourier transform mass spectrometry. The HPM	hpm.protein	http://identifiers.org/hpm.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		HPM
								MIR:00100682	http://www.humanproteomemap.org/protein.php?hpm_id={$id}	Human Proteome Map Protein at Institute of Bioinformatics (Bangalore)	Institute of Bioinformatics, International Tech Park, Bangalore	India	false	1968	Eukaryotic translation initiation	http://www.humanproteomemap.org/index.php		
MIR:00000527	Human Proteome Map Peptide	human	^\d+$	The Human Proteome Map (HPM) portal integrates the peptide sequencing result from the draft map of the human proteome project. The project was based on LC-MS/MS by utilizing of high resolution and high accuracy Fourier transform mass spectrometry. The HPM	hpm.peptide	http://identifiers.org/hpm.peptide	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100683	http://www.humanproteomemap.org/spectrum.php?pep_id={$id}	Human Proteome Map Peptide at Institute of Bioinformatics (Bangalore)	Institute of Bioinformatics, International Tech Park, Bangalore	India	false	9606117	AGGE	http://www.humanproteomemap.org/index.php		
MIR:00000528	DrugBank Target v4	human	^BE\d{7}$	The DrugBank database is a bioinformatics and chemoinformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. This collection	drugbankv4.target	http://identifiers.org/drugbankv4.target	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100687	http://www.drugbank.ca/biodb/bio_entities/{$id}	DrugBank Target information version 4	Departments of Computing Science, Biological Sciences, University of Alberta	Canada	false	BE0000048	Prothrombin	http://www.drugbank.ca/targets		
MIR:00000532	Orphanet Rare Disease Ontology	All species	^Orphanet(_|:)C?\d+$	The Orphanet Rare Disease Ontology (ORDO) is a structured vocabulary for rare diseases, capturing relationships between diseases, genes and other relevant features which will form a useful resource for the computational analysis of rare diseases. It integrates a nosology	orphanet.ordo	http://identifiers.org/orphanet.ordo	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	ORDO
								MIR:00100705	https://www.ebi.ac.uk/ols/ontologies/ordo/terms?short_form={$id}	ORDO via OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	true	ols	Orphanet_C023	age of onset	https://www.ebi.ac.uk/ols/ontologies/ordo	
MIR:00000534	ClinVar Record	human	^RCV\d+(\.\d+)?$	ClinVar archives reports of relationships among medically important variants and phenotypes. It records human variation, interpretations of the relationship specific variations to human health, and supporting evidence for each interpretation. Each ClinVar record (RCV	clinvar.record	http://identifiers.org/clinvar.record	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100710	http://www.ncbi.nlm.nih.gov/clinvar/{$id}/	ClinVar Record at NCBI	National Center for Biotechnology Information (NCBI), NIH, Maryland	USA	false	ncbi	RCV000033555.3	protein tyrosine phosphatase	http://www.ncbi.nlm.nih.gov/clinvar/	
MIR:00000541	ExAC Variant	human	^\d{1,2}\-\d+\-[GATC]\-[GATC]$	The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a variety of large-scale sequencing projects, and to make summary data available for the wider scientific community. The data	exac.variant	http://identifiers.org/exac.variant	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100718	http://exac.broadinstitute.org/variant/{$id}	ExAC Variant at Exome Aggregation Consortium	Exome Aggregation Consortium (ExAC), Cambridge, Massachusetts	USA	false	22-46615880-T-C	1209806.67	http://exac.broadinstitute.org/		
MIR:00000542	GWAS Central Marker	human	^HGVM\d+$	GWAS Central (previously the Human Genome Variation database of Genotype-to-Phenotype information) is a database of summary level findings from genetic association studies, both large and small. It gathers datasets from public domain projects, and accepts direct data	gwascentral.marker	http://identifiers.org/gwascentral.marker	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100719	http://www.gwascentral.org/marker/{$id}	GWAS Central Marker at University of Leicester	Department of Genetics, University of Leicester, Leicester	UK	false	HGVM15354	rs8580	http://www.gwascentral.org/markers/		
MIR:00000545	LINCS Protein	animal	^\d+$	The HMS LINCS Database currently contains information on experimental reagents (small molecule perturbagens, cells, and proteins). It aims to collect and disseminate information relating to the fundamental principles of cellular response in humans to perturbation. This	lincs.protein	http://identifiers.org/lincs.protein	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100722	http://lincs.hms.harvard.edu/db/proteins/{$id}/	LINCS Protein at Harvard Medical School	Harvard Medical School, Boston, Massachusetts	USA	false	200282	Myosin IIIA	http://lincs.hms.harvard.edu/db/proteins/		
MIR:00000546	LINCS Small Molecule	animal	^LSM-\d+{4}$	The Library of Network-Based Cellular Signatures (LINCS) Program aims to create a network-based understanding of biology by cataloging changes in gene expression and other cellular processes that occur when cells are exposed to a variety of perturbing agents. The LINCS small	lincs.smallmolecule	http://identifiers.org/lincs.smallmolecule	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100723	http://lincsportal.ccs.miami.edu/SmallMolecules/#/view/{$id}	LINCS Portal	University of Miami, BD2K-LINCS DCIC, USA	USA	false	LSM-6306	Dinaciclib	http://lincsportal.ccs.miami.edu/SmallMolecules/		
MIR:00000547	ExAC Transcript	human	^ENST\d{11}$	The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a variety of large-scale sequencing projects, and to make summary data available for the wider scientific community. The data	exac.transcript	http://identifiers.org/exac.transcript	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100724	http://exac.broadinstitute.org/transcript/{$id}	ExAC Transcript at Exome Aggregation Consortium	Exome Aggregation Consortium (ExAC), Cambridge, Massachusetts	USA	false	ENST00000407236	PPARA	http://exac.broadinstitute.org/		
MIR:00000548	ExAC Gene	human	^ENSG\d{11}$	The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a variety of large-scale sequencing projects, and to make summary data available for the wider scientific community. The data	exac.gene	http://identifiers.org/exac.gene	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100725	http://exac.broadinstitute.org/gene/{$id}	ExAC Gene at Exome Aggregation Consortium	Exome Aggregation Consortium (ExAC), Cambridge, Massachusetts	USA	false	ENSG00000169174	proprotein convertase	http://exac.broadinstitute.org/		
MIR:00000552	SEED Subsystem	Microbial organisms	^\w+$	This cooperative effort, which includes Fellowship for Interpretation of Genomes (FIG), Argonne National Laboratory, and the University of Chicago, focuses on the development of the comparative genomics environment called the SEED. It is a framework to support comparative	seed	http://identifiers.org/seed	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100729	http://seed-viewer.theseed.org/seedviewer.cgi?page=Subsystems&subsystem={$id}	SEED Subsystem at Argonne National Laboratory	Argonne National Laboratory, University of Chicago, Argonne, Illinois	USA	false	Biotin_biosynthesis	salvage	http://seed-viewer.theseed.org/seedviewer.cgi		
MIR:00000553	SEED Compound	Microbial organisms	^cpd\d+$	This cooperative effort, which includes Fellowship for Interpretation of Genomes (FIG), Argonne National Laboratory, and the University of Chicago, focuses on the development of the comparative genomics environment called the SEED. It is a framework to support comparative	seed.compound	http://identifiers.org/seed.compound	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100730	http://modelseed.org/biochem/compounds/{$id}	SEED Compound at Argonne National Laboratory	Argonne National Laboratory, University of Chicago, Argonne, Illinois	USA	false	cpd15380	5'-deoxyribose	http://modelseed.org/		
MIR:00000554	BiGG Model	All species	^[a-z_A-Z0-9]+$	BiGG is a knowledgebase of Biochemically, Genetically and Genomically structured genome-scale metabolic network reconstructions. It more published genome-scale metabolic networks into a single database with a set of stardized identifiers called BiGG IDs. Genes in the	bigg.model	http://identifiers.org/bigg.model	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100731	http://bigg.ucsd.edu/models/{$id}	BiGG Model at University of California	Systems Biology Research Group at the University of California, San Diego	USA	false	iECABU_c1320	Escherichia coli	http://bigg.ucsd.edu/models		
MIR:00000559	UMLS	human	^C\d+$	The Unified Medical Language System is a repository of biomedical vocabularies. Vocabularies integrated in the UMLS Metathesaurus include the NCBI taxonomy, Gene Ontology, the Medical Subject Headings (MeSH), OMIM and the Digital Anatomist Symbolic Knowledge	umls	http://identifiers.org/umls	0	id	accessURL	info	institution	location	official	localId	resourceURL			Unified Medical Language System
								MIR:00100736	http://linkedlifedata.com/resource/umls/id/{$id}	Unified Medical Language System at LinkedLifeData	Linkedlifedata, Ontotext, Sofia	Bulgaria	true	C0021390	http://linkedlifedata.com/			
MIR:00000560	MeSH	human	^(C|D)\d{6}$	MeSH (Medical Subject Headings) is the National Library of Medicine's controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. This thesaurus is used by NLM for indexing articles from	mesh	http://identifiers.org/mesh	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Medical Subject Headings
								MIR:00100737	http://id.nlm.nih.gov/mesh/{$id}	MeSH Linked Data at National Library of Medicine	U.S. National Library of Medicine, National Institute of Health, Maryland	USA	false	C000100	aquayamycin	http://id.nlm.nih.gov/mesh/		
MIR:00000562	miRTarBase	All species	^MIRT\d{6}$	miRTarBase is a database of miRNA-target interactions (MTIs), collected manually from relevant literature, following Natural Language Processing of the text to identify research articles related to functional studies of miRNAs. Generally, the collected MTIs are validated	mirtarbase	http://identifiers.org/mirtarbase	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100739	http://mirtarbase.mbc.nctu.edu.tw/php/detail.php?mirtid={$id}	miRTarBase at National Chiao Tung University	Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu	Taiwan	false	MIRT000002	Homo sapiens miR-20a stem-loop	http://mirtarbase.mbc.nctu.edu.tw/		
MIR:00000564	DASHR	human	^(hsa-(let|mir)-\w+(-\w+)?)|(piR-\d+)|(chr\w+.tRNA\d+-\w+)|(chr\w+.tRNA\d+-\w+-tRF\d)|((SNORD|SNORA|ACA|HBII|HBI|U)(-)?\w+)|(HY\d\+(-L\d+)?)|((LSU|SSU|5S)(-rRNA_Hsa)?(-L\d+)?)$	DASHR reports the annotation, expression and evidence for specific RNA processing (cleavage specificity scores/entropy) of human sncRNA genes, precursor and mature sncRNA products across different human tissues and cell types. DASHR integrates information from multiple	dashr	http://identifiers.org/dashr	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Database of small human noncoding RNAs
								MIR:00100741	http://lisanwanglab.org/DASHR/entry/{$id}	DASHR at University of Pennsylvania	Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia	USA	false	hsa-mir-200a	testiculargerm1	http://lisanwanglab.org/DASHR/		
MIR:00000565	DASHR expression	human	^(hsa-(let|mir)-\w+(-\w+)?)|(piR-\d+)|(chr\w+.tRNA\d+-\w+)|(chr\w+.tRNA\d+-\w+-tRF\d)|((SNORD|SNORA|ACA|HBII|HBI|U)(-)?\w+)|(HY\d\+(-L\d+)?)|((LSU|SSU|5S)(-rRNA_Hsa)?(-L\d+)?)$	DASHR reports the annotation, expression and evidence for specific RNA processing (cleavage specificity scores/entropy) of human sncRNA genes, precursor and mature sncRNA products across different human tissues and cell types. DASHR integrates information from multiple	dashr.expression	http://identifiers.org/dashr.expression	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100742	http://lisanwanglab.org/DASHR/entry/{$id}#exprPerTissueTable	DASHR expression at University of Pennsylvania	Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia	USA	false	hsa-mir-200a	brainog1	http://lisanwanglab.org/DASHR/show-expression-table.php?start=0		
MIR:00000570	UniPathway Reaction	All species	^UCR\d{5}$	UniPathway is a manually curated resource of enzyme-catalyzed and spontaneous chemical reactions. It provides a hierarchical representation of metabolic pathways and a controlled vocabulary for pathway annotation in UniProtKB. UniPathway data are cross-	unipathway.reaction	http://identifiers.org/unipathway.reaction	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100752	http://www.grenoble.prabi.fr/obiwarehouse/unipathway/ucr?upid={$id}	UniPathway Reaction at Swiss Institute of Bioinformatics (SIB)	Swiss Institute of Bioinformatics (SIB) and French National Institute for Research in Computer Science and Control	Switzerland	false	UCR00226	2-acetolactate	http://www.grenoble.prabi.fr/obiwarehouse/unipathway		
MIR:00000573	HGNC gene family	human	^\d+$	The HGNC (HUGO Gene Nomenclature Committee) provides an approved gene name and symbol (short-form abbreviation) for each known human gene. All approved symbols are stored in the HGNC database, and each symbol is unique. In addition, HGNC also provides a	hgnc.genefamily	http://identifiers.org/hgnc.genefamily	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100765	https://www.genenames.org/cgi-bin/genefamilies/set/{$id}	HGNC gene family at HUGO Genome Nomenclature Committee	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	141	Olfactory receptors	https://www.genenames.org/		
MIR:00000576	APID Interactomes	All species	^([A-N,R-Z][0-9]([A-Z][A-Z, 0-9][A-Z, 0-9][0-9]){1,2})|([O,P,Q][0-9][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9])(\.\d+)?$	APID (Agile Protein Interactomes DataServer) provides information on the protein interactomes of numerous organisms, based on the integration of known experimentally validated protein-protein physical interactions (PPIs). Interactome data includes a report on	apid.interactions	http://identifiers.org/apid.interactions	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Agile Protein Interactomes DataServer
								MIR:00100769	http://cicblade.dep.usal.es:8080/APID/Interactions.action?protein={$id}	APID at Salamanca	Cancer Research Center (CiC-IBMCC), Consejo Superior de Investigaciones Científicas (CSIC) and Universidad de Salamanca (USAL), Salamanca	Spain	false	P01116	IntAct	http://cicblade.dep.usal.es:8080/APID/		
MIR:00000578	Kyoto Encyclopedia of Genes and Genomes	All species	^([CHDEGTMKR]\d+)|(\w+:[\w\d\.-]*)|([a-z]{3,5})|(\w{2,4}\d{5})$	Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale	kegg	http://identifiers.org/kegg	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		KEGG
								MIR:00100773	http://www.kegg.jp/entry/{$id}	KEGG at Kyoto University Bioinformatics Center	Kyoto University Bioinformatics Center	Japan	false	hsa00190	phosphorylation	http://www.kegg.jp/		
MIR:00000579	Plant Transcription Factor Database	Plant species	^[A-Z][a-z]{2}_([A-Za-z]{3}[0-9]{6})|([A-Za-z0-9\._\-#]*)$	The Plant TF database (PlantTFDB) systematically identifies transcription factors for plant species. It includes annotation for identified TFs, including information on expression, regulation, interaction, conserved elements, phenotype information. It also provides	planttfdb	http://identifiers.org/planttfdb	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		PlantTFDB
								MIR:00100772	http://planttfdb.cbi.pku.edu.cn/tf.php?uid={$id}	PlantTFDB at Peking University	Center for Bioinformatics, Peking University, Beijing	Peoples Republic of China	false	Ath_AT1G01030.1	Arabidopsis	http://planttfdb.cbi.pku.edu.cn		
MIR:00000594	MedGen	human	^[CN]*\d{4,7}$	MedGen is a portal for information about conditions and phenotypes related to Medical Genetics. Terms from multiple sources are aggregated into concepts, each of which is assigned a unique identifier and a preferred name and symbol. The core content of the record	medgen	http://identifiers.org/medgen	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	
								MIR:00100794	https://www.ncbi.nlm.nih.gov/medgen/{$id}	MedGen at NCBI	National Center for Biotechnology Information (NCBI), NIH, Maryland	USA	false	ncbi	760050	Caffeine	https://www.ncbi.nlm.nih.gov/medgen/	
MIR:00000596	ClinVar Variant	human	^\d+$	ClinVar archives reports of relationships among medically important variants and phenotypes. It records human variation, interpretations of the relationship specific	clinvar	http://identifiers.org/clinvar	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	

MIR:00000596	ClinVar Variant	human	^\d+$	interpretations of the relationship specific variations to human health, and supporting evidence for each interpretation. Each ClinVar record (RCV	clinvar	http://identifiers.org/clinvar	0	MIR:00100796	https://www.ncbi.nlm.nih.gov/clinvar/variation/{$id}	ClinVar Variant at NCBI	National Center for Biotechnology Information (NCBI), NIH, Maryland	USA	false	ncbi	12345	Pathogenic	http://www.ncbi.nlm.nih.gov/clinvar/	
MIR:00000600	Progenetix	human	^[\w\-:,]{3,64}$	The Progenetix database provides an overview of copy number abnormalities in human cancer from currently 32317 array and chromosomal Comparative Genomic Hybridization (CGH) experiments, as well as Whole Genome or Whole Exome Sequencing (WGS, WES) studies. The	pgx	http://identifiers.org/pgx	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Cancer Genome Knowledge Resource
								MIR:00100801	http://progenetix.org/pgx:{$id}	Progenetix	University of Zurich	Switzerland	false	icdom:8500_3	8500/3	http://progenetix.org/		
MIR:00000601	COSMIC Gene	human	^[A-Z0-9]+$	COSMIC is a comprehensive global resource for information on somatic mutations in human cancer, combining curation of the scientific literature with tumor resequencing data from the Cancer Genome Project at the Sanger Institute, U.K. This collection references genes	cosmic	http://identifiers.org/cosmic	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Catalogue of Somatic Mutations in Cancer Gene
								MIR:00100802	http://cancer.sanger.ac.uk/cosmic/gene/overview?ln={$id}	COSMIC Gene at Sanger	Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge	UK	false	BRAF	BRAF	http://cancer.sanger.ac.uk/cosmic/		
MIR:00000602	dbGaP	human	^phs[0-9]{6}(.v\d+.p\d+)?$	The database of Genotypes and Phenotypes (dbGaP) archives and distributes the results of studies that have investigated the interaction of genotype and phenotype.	dbgap	http://identifiers.org/dbgap	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Database of Genotypes and Phenotypes
								MIR:00100803	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id={$id}	dbGaP through NCBI	National Center for Biotechnology Information (NCBI)	USA	true	ncbi	phs000768.v2.p1	Genomic	https://www.ncbi.nlm.nih.gov/projects/gap	
MIR:00000603	Information Artifact Ontology	human	^\d{7}$	An ontology of information entities, originally driven by work by the Ontology of Biomedical Investigation (OBI) digital entity and realizable information entity branch.	iao	http://identifiers.org/iao	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100804	http://purl.obolibrary.org/obo/IAO_{$id}	IAO through Ontobee	University of Pennsylvania	USA	false	30	continuant	http://www.ontobee.org/ontology/IAO		
MIR:00000604	Genomic Data Commons Data Portal	human	^[a-z0-9]{8}-[a-z0-9]{4}-[a-z0-9]{4}-[a-z0-9]{4}-[a-z0-9]{12}$	The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis.	gdc	http://identifiers.org/gdc	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100805	https://portal.gdc.cancer.gov/cases/{$id}	Genomic Data Commons	National Cancer Institute Genomic Data Commons, NIH	USA	false	ae8c77fe-e6c8-44d5-8265-4a38c637bbef	https://gdc.cancer.gov			
MIR:00000605	OMIT	human	^\d{7}$	The purpose of the OMIT ontology is to establish data exchange standards and common data elements in the microRNA (miR) domain. Biologists (cell biologists in particular) and bioinformaticians can make use of OMIT to leverage emerging semantic technologies in knowledge	omit	http://identifiers.org/omit	0	id	accessURL	info	institution	location	official	resourcePrefix	localId	testString	resourceURL	Ontology for miRNA Target
								MIR:00100806	https://www.ebi.ac.uk/ols/ontologies/omit/terms?short_form=OMIT_{$id}	Ontology for miRNA Target (OMIT) through OLS	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	ols	5506	DNA Damage	https://www.ebi.ac.uk/ols/ontologies/omit/	
MIR:00000606	INSDC CDS	All species	^([A-Z]\d{5}|[A-Z]{2}\d{6}|[A-Z]{4}\d{8}|[A-J][A-Z]{2}\d{5})(\.\d+)?$	The coding sequence or protein identifiers as maintained in INSDC.	insdc.cds	http://identifiers.org/insdc.cds	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100807	https://www.ebi.ac.uk/ena/data/view/{$id}	INSDC CDS through ENA	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	AAA35559	Homo sapiens	https://www.ebi.ac.uk/ena		
								MIR:00100808	https://www.ncbi.nlm.nih.gov/protein/{$id}	INSDC CDS through NCBI	National Center for Biotechnology Information (NCBI)	USA	false	AAA35559	beta-arrestin	https://www.ncbi.nlm.nih.gov/protein/		
								MIR:00100809	http://getentry.ddbj.nig.ac.jp/getentry/dad/{$id}	INSDC CDS through DDBJ	DNA Data Bank of Japan, Mishima, Shizuoka	Japan	false	AAA35559	Homo sapiens	http://getentry.ddbj.nig.ac.jp		
MIR:00000607	Genome assembly database	All species	^GCA_[0-9]{9}\.[0-9]$	The genome assembly database contains detailed information about genome assemblies for eukaryota, bacteria and archaea. The scope of the genome collections database does not extend to viruses, viroids and bacteriophage.	insdc.gca	http://identifiers.org/insdc.gca	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL	resourcePrefix	
								MIR:00100810	https://www.ebi.ac.uk/ena/data/view/{$id}	Genome assembly database through ENA	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	GCA_000155495.1	ASM15549v1	https://www.ebi.ac.uk/ena/browse/genome-assembly-database		
								MIR:00100811	https://www.ncbi.nlm.nih.gov/assembly/{$id}	Genome assembly database NCBI	National Center for Biotechnology Information (NCBI)	USA	false	GCA_000155495.1	ASM15549v1	https://www.ncbi.nlm.nih.gov/assembly/	ncbi	
MIR:00000610	MetaCyc Reaction	All species	^RXN-\d{5}$	MetaCyc is a curated database of experimentally elucidated metabolic pathways from all domains of life. MetaCyc contains 2526 pathways from 2844 different organisms. MetaCyc contains pathways involved in both primary and secondary metabolism	metacyc.reaction	http://identifiers.org/metacyc.reaction	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100814	https://metacyc.org/META/NEW-IMAGE?type=REACTION&object={$id}	MetaCyc Metabolic Pathway Database - Reaction	Bioinformatics Research Group at SRI International, CA	USA	false	RXN-14904	ribose phosphorylation	https://metacyc.org		
MIR:00000612	CATH Protein Structural Domain Superfamily		^[1-4]\.[0-9]+\.[0-9]+\.[0-9]+$	CATH is a classification of protein structural domains. We group protein domains into superfamilies when there is sufficient evidence they have diverged from a common ancestor. CATH can be used to predict structural and functional information directly from protein	cath	http://identifiers.org/cath	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100816	http://www.cathdb.info/cathnode/{$id}	CATH Node through UCL	University College London	UK	false	1.10.8.10	DNA helicase RuvA subunit	http://www.cathdb.info		
MIR:00000613	MarRef	Microbial organisms	^MMP\d+.\d+$	MarRef is a manually curated marine microbial reference genome database that contains completely sequenced genomes.	mmp.ref	http://identifiers.org/mmp.ref	0	id	accessURL	info	institution	location	official	localId	resourceURL			Marine Metagenomics Portal
								MIR:00100817	https://mmp.sfb.uit.no/databases/marref/#/records/{$id}	MarRef through SfB	Center for Bioinformatics (SfB), UiT The Arctic university of Norway	Norway	false	MMP3312982.2	https://mmp.sfb.uit.no/databases/marref/			
MIR:00000614	MarDB	Microbial organisms	^MMP\d+.\d+$	MarDB includes all sequenced marine microbial genomes regardless of level of completeness.	mmp.db	http://identifiers.org/mmp.db	0	id	accessURL	info	institution	location	official	localId	resourceURL			Marine Metagenomics Portal
								MIR:00100818	https://mmp.sfb.uit.no/databases/mardb/#/records/{$id}	MarDB through SfB	Center for Bioinformatics (SfB), UiT The Arctic university of Norway	Norway	false	MMP02954345.1	https://mmp.sfb.uit.no/databases/mardb/			
MIR:00000615	MarCat	Microbial organisms	^MMP\d+.\d+$	MarCat is a gene (protein) catalogue of uncultivable and cultivable marine genes and proteins derived from metagenomics samples.	mmp.cat	http://identifiers.org/mmp.cat	0	id	accessURL	info	institution	location	official	localId	resourceURL			Marine Metagenomics Portal
								MIR:00100819	https://mmp.sfb.uit.no/databases/marcat/#/records/{$id}	MarCat through SfB	Center for Bioinformatics (SfB), UiT The Arctic university of Norway	Norway	false	MMP743597.11	https://mmp.sfb.uit.no/databases/marcat/			
MIR:00000618	NASA GeneLab	All species	^GLDS-\d+$	NASA's GeneLab gathers spaceflight genomic data, RNA and protein expression, and metabolic profiles, interfaces with existing databases for expanded research, will offer tools to conduct data analysis, and is in the process of creating a robust online	ngl	http://identifiers.org/ngl	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100822	https://genelab-data.ndc.nasa.gov/genelab/accession/{$id}	NASA GeneLab	NASA's Ames Research Center	USA	false	GLDS-141	https://genelab-data.ndc.nasa.gov/genelab/			
MIR:00000619	Human Endogenous Retrovirus Database	virus	^[A-Za-z0-9\-_]+$	Endogenous retroviruses (ERVs) are common in vertebrate genomes; a typical mammalian genome contains tens to hundreds of thousands of ERV elements. Most ERVs are evolutionarily old and have accumulated multiple mutations, playing important roles in physiology and disease	erv	http://identifiers.org/erv	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		
								MIR:00100823	https://herv.img.cas.cz/s/{$id}	Human Endogenous Retrovirus Database	Institute of Molecular Genetics, Academy of Sciences of the Czech Republic	Czech Republic	false	THE1B	THE1B	https://herv.img.cas.cz/		
MIR:00000621	ArrayMap	human	^[\w\-:,]{3,64}$	arrayMap is a collection of pre-processed oncogenomic array data sets and CNA (somatic copy number aberrations) profiles. CNA are a type of mutation commonly found in cancer genomes. arrayMap data is assembled from public repositories and supplemented with	arraymap	http://identifiers.org/arraymap	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100825	https://www.arraymap.org/pgx:{$id}	ArrayMap	Institute of Molecular Life Sciences, University of Zurich, Zurich	Switzerland	false	icdom:8500_3	https://www.arraymap.org			
MIR:00000624	SISu	human	^[0-9]+:[0-9]+$	The Sequencing Initiative Suomi (SISu) project is an international collaboration to harmonize and aggregate whole genome and exome sequence data from Finnish samples, providing data for researchers and clinicians. The SISu project allows for the search of variants to	sisu	http://identifiers.org/sisu	0	id	accessURL	info	institution	location	official	localId	resourceURL			Sequence Initiative Suomi
								MIR:00100828	http://search.sisuproject.fi/#/variant/{$id}	SISu	Sequencing Initiative Suomi project (SISu), Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki	Finland	false	17:41223048	http://www.sisuproject.fi/			
MIR:00000626	CRISPRdb	human and mouse	^[0-9]+$	Repeated CRISPR ("clustered regularly interspaced short palindromic repeats") elements found in archaebacteria and eubacteria are believed to defend against viral infection, potentially targeting invading DNA for degradation. CRISPRdb is a database that stores information on	crisprdb	http://identifiers.org/crisprdb	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100830	http://crispr.i2bc.paris-saclay.fr/cgi-bin/crispr/SpecieProperties_db.cgi?Taxon_id[]={$id}	CRISPRdb	Université Paris-Sud, Institut de Génétique et Microbiologie, Orsay,	France	false	551115	http://crispr.i2bc.paris-saclay.fr/			
MIR:00000627	GnpIS	Plant species	^[A-Za-z0-9]+$	GnpIS is an integrative information system focused on plants and fungal pests. It provides both genetic (e.g. genetic maps, quantitative trait loci, markers, single nucleotide polymorphisms, germplasms and genotypes) and genomic data (e.g. genomic sequences, physical maps	gnpis	http://identifiers.org/gnpis	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100831	https://urgi.versailles.inra.fr/gnpis/#result/term={$id}	GnpIS	URGI (Research Unit in Genomics-Info), INRA de Versailles, Route de Saint-Cyr, Versailles	France	false	AY109603	https://urgi.versailles.inra.fr/gnpis/			
MIR:00000628	LiceBase	Sea lice	^[A-Za-z0-9\-\/]+$	Sea lice (Lepeophtheirus salmonis and Caligus species) are the major pathogens of salmon, significantly impacting upon the global salmon farming industry. Lice control is primarily accomplished through chemotherapeutants, though emerging resistance necessitates the development	licebase	http://identifiers.org/licebase	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100832	https://licebase.org/?q={$id}	LiceBase	Sea Lice Research Centre (SLRC), University of Bergen	Norway	false	EMLSAT00000003403	https://licebase.org			
MIR:00000630	RNAcentral	All species	^URS[0-9A-F]{10}(_\d+)?$	RNAcentral is a public resource that offers integrated access to a comprehensive and up-to-date set of non-coding RNA sequences provided by a collaborating group of Expert Databases.	rnacentral	http://identifiers.org/rnacentral	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100834	https://rnacentral.org/rna/{$id}	RNAcentral through EMBL-EBI	European Bioinformatics Institute, Hinxton, Cambridge	UK	false	URS0000759CF4	https://rnacentral.org/			
MIR:00000631	MicroScope	Microbial organisms	^\d+$	MicroScope is an integrative resource that supports systematic and efficient revision of microbial genome annotation, data management and comparative analysis.	microscope	http://identifiers.org/microscope	0	id	accessURL	info	institution	location	official	localId	testString	resourceURL		Microbial Genome Annotation & Analysis Platform
								MIR:00100836	http://www.genoscope.cns.fr/agc/microscope/mage/info.php?id={$id}	MicroScope	Université Evry-Val-d'Essonne, CEA, Institut de Génomique - Genoscope, Laboratoire d'Analyses Bioinformatiques pour la Génomique et le Métabolisme, Evry	France	false	5601141	Quinoprotein glucose dehydrogenase B	http://www.genoscope.cns.fr/agc/microscope		
MIR:00000632	SwissRegulon	animal, yeast and prokaryotes	^[A-Za-z0-9]+$	A database of genome-wide annotations of regulatory sites. It contains annotations for 17 prokaryotes and 3 eukaryotes. The database frontend offers an intuitive interface showing genomic information in a graphical form.	swissregulon	http://identifiers.org/swissregulon	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100837	http://swissregulon.unibas.ch/query/{$id}	SwissRegulon	Swiss Institute of Bioinformatics	Switzerland	false	AHR	http://swissregulon.unibas.ch			
MIR:00000662	ENCODE: Encyclopedia of DNA Elements	human	^ENC[A-Za-z]{2}[0-9]{3}[A-Za-z]{3}$	The ENCODE Consortium is integrating multiple technologies and approaches in a collective effort to discover and define the functional elements encoded in the human genome, including genes, transcripts, and transcriptional regulatory regions, together with their attendant	encode	http://identifiers.org/encode	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100880	https://www.encodeproject.org/{$id}	ENCODE: Encyclopedia of DNA Elements	ENCODE Consortium	USA	false	ENCSR163RYW	www.encodeproject.org			

MIR:00000663	GTEx	human	^\w.+$	The Genotype-Tissue Expression (GTEx) project aims to provide to the scientific community a resource with which to study human gene expression and regulation and its relationship to genetic variation.	gtex	http://identifiers.org/gtex	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100881	https://www.gtexportal.org/home/gene/{$id}	The GTEx Project	The Broad Institute of MIT and Harvard	USA	false	BRIP1	https://www.gtexportal.org			
MIR:00000664	RiceNetDB Gene	rice	^LOC_Os\d{1,2}g\d{5}\.\d$	RiceNetDB is currently the most comprehensive regulatory database on Oryza Sativa based on genome annotation. It was displayed in three levels: GEM, PPIs and GRNs to facilitate biomolecular regulatory analysis and gene-metabolite mapping.	ricenetdb.gene	http://identifiers.org/ricenetdb.gene	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100882	http://bis.zju.edu.cn/ricenetdb/genedetails.php?ID={$id}	RiceNetDB	College of Life Sciences, Zhejiang University	China	false	LOC_Os01g49190.1	http://bis.zju.edu.cn/ricenetdb/			
MIR:00000665	RiceNetDB Protein	rice	^LOC_Os\d{1,2}g\d{5}$	RiceNetDB is currently the most comprehensive regulatory database on Oryza Sativa based on genome annotation. It was displayed in three levels: GEM, PPIs and GRNs to facilitate biomolecular regulatory analysis and gene-metabolite mapping.	ricenetdb.protein	http://identifiers.org/ricenetdb.protein	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100883	http://bis.zju.edu.cn/ricenetdb/proteindetails.php?ID={$id}	RiceNetDB	College of Life Sciences, Zhejiang University	China	false	LOC_Os01g49190	http://bis.zju.edu.cn/ricenetdb/			
MIR:00000668	RiceNetDB miRNA	rice	^osa-miR\d{3,5}[a-z]{0,1}$	RiceNetDB is currently the most comprehensive regulatory database on Oryza Sativa based on genome annotation. It was displayed in three levels: GEM, PPIs and GRNs to facilitate biomolecular regulatory analysis and gene-metabolite mapping.	ricenetdb.miRNA	http://identifiers.org/ricenetdb.miRNA	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100886	http://bis.zju.edu.cn/ricenetdb/miRNAdetails.php?ID={$id}	RiceNetDB miRNA	College of Life Sciences, Zhejiang University	China	false	osa-miR446	http://bis.zju.edu.cn/ricenetdb			
MIR:00000675	Addgene Plasmid Repository	plasmid	^[0-9]{5}(-[a-zA-Z0-9-]{0,7})?$\|^[0-9]{10}$	Addgene is a non-profit plasmid repository. Addgene facilitates the exchange of genetic material between laboratories by offering plasmids and their associated cloning data to not-for-profit laboratories around the world.	addgene	http://identifiers.org/addgene	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100899	http://addgene.org/{$id}	Addgene Plasmid Repository	Addgene	USA	false	50943	http://addgene.org/			
MIR:00000676	Bacterial Diversity Metadatabase	bacteria	^[0-9]+$	BacDive—the Bacterial Diversity Metadatabase merges detailed strain-linked information on the different aspects of bacterial and archaeal biodiversity.	bacdive	http://identifiers.org/bacdive	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100900	https://bacdive.dsmz.de/strain/{$id}	Bacterial Diversity Metadatabase	Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen	Germany	false	131392	https://bacdive.dsmz.de/			
MIR:00000679	The cBioPortal for Cancer Genomics	human	^[a-z0-9_]+$	The cBioPortal for Cancer Genomics provides visualization, analysis and download of large-scale cancer genomics data sets.	cbioportal	http://identifiers.org/cbioportal	0	id	accessURL	info	institution	location	official	localId	resourceURL			
								MIR:00100909	http://www.cbioportal.org/study?id={$id}#summary	The cBioPortal for Cancer Genomics	Memorial Sloan Kettering Cancer Center	USA	false	laml_tcga_pub	http://www.cbioportal.org			

Annex4 List of plant genomes stored in Ensembl Plants

Name	Classification	Taxon ID	Assembly	Accession	Variation database	Regulation database	Whole genome alignments	Other alignments	In peptide compara	In pan-taxonomic compara
Actinidia chinensis	eudicotyledons	1590841	Red5_PS1_1.69.0	GCA_003024255.1	-	-		-		-
Aegilops tauschii	Liliopsida	200361	Aet_v4.0	GCA_002575655.1	-					-
Amborella trichopoda	Amborellales	13333	AMTR1.0	GCA_000471905.1	-	-				
Ananas comosus	Liliopsida	4615	F153	GCA_902162155.1	-	-		-		-
Arabidopsis halleri	eudicotyledons	63677	Ahal2.2	GCA_900078215.1	-					-
Arabidopsis lyrata	eudicotyledons	81972	v.1.0	GCA_000004255.1	-	-				-
Arabidopsis thaliana	eudicotyledons	3702	TAIR10	GCA_000001735.1						
Arabis alpina	eudicotyledons	50452	A_alpina_V4	GCA_000733195.1	-	-		-		-
Asparagus officinalis	Liliopsida	4686	Aspof.V1	GCA_001876935.1	-	-		-		-
Avena sativa OT3098	Liliopsida	4498	Oat_OT3098_v2	GCA_022788535.1	-	-		-		-
Avena sativa Sang	Liliopsida	4498	Asativa_sang.v1.1	GCA_910574605.1	-	-		-		-
Beta vulgaris	eudicotyledons	3555	RefBeet-1.2.2	GCA_000511025.2	-	-				-
Brachypodium distachyon	Liliopsida	15368	Brachypodium_distachyon_v3.0	GCA_000005505.4		-				
Brassica juncea	eudicotyledons	3707	ASM1870372v1	GCA_018703725.1	-	-		-		-
Brassica napus	eudicotyledons	3708	AST_PRJEB5043_v1	GCA_000751015.1	-					-
Brassica oleracea	eudicotyledons	109376	BOL	GCA_000695525.1	-					-
Brassica rapa	eudicotyledons	3711	Brapa_1.0	GCA_000309985.1	-		-		-	-
Brassica rapa R-o-18	eudicotyledons	1813537	SCU_BraROA_2.3	GCA_017639395.1	-	-		-		-
Cajanus cajan (pigeon pea) - GCA_000340665.1	eudicotyledons	3821	C.cajan_V1.0	GCA_000340665.1	-	-		-		-
Camelina sativa	eudicotyledons	90675	Cs	GCA_000633955.1	-	-		-		-
Cannabis sativa female	eudicotyledons	3483	cs10	GCA_900626175.1	-	-		-		-
Capsicum annuum	eudicotyledons	4072	ASM51225v2	GCA_000512255.2	-	-		-		-
Chara braunii	Streptophyta	69332	Cbr_1.0	GCA_003427395.1	-	-	-	-		-
Chenopodium quinoa	eudicotyledons	63459	ASM168347v1	GCA_001683475.1	-	-		-		-
Chlamydomonas reinhardtii	Chlorophyta	3055	Chlamydomonas_reinhardtii_v5.5	GCA_000002595.3	-	-				
Chondrus crispus	Rhodophyta	2769	ASM35022v2	GCA_000350225.2	-	-		-		-
Citrullus lanatus	eudicotyledons	3654	Cla97_v1	GCA_000238415.2	-	-		-		-
Citrus clementina	eudicotyledons	85681	Citrus_clementina_v1.0	GCA_000493195.1	-	-		-		-
Coffea canephora	eudicotyledons	49390	AUK_PRJEB4211_v1	GCA_900059795.1	-	-		-		-
Corchorus capsularis	eudicotyledons	210143	CCACVL1_1.0	GCA_001974805.1	-	-				-
Corylus avellana	eudicotyledons	13451	CavTom2PMs-1.0	GCA_901000735.2	-	-		-		-
Corymbia citriodora	eudicotyledons	360336	Ccitriodora_v2_1	GCA_014858505.1	-	-		-		-
Cucumis melo	eudicotyledons	3656	Melonv4	GCA_902497455.1	-	-		-		-
Cucumis sativus	eudicotyledons	3659	ASM407v2	GCA_000004075.2		-				-
Cyanidioschyzon merolae	Rhodophyta	280699	ASM9120v1	GCA_000091205.1	-	-		-		
Cynara cardunculus	eudicotyledons	59895	CcrdV1	GCA_001531365.1	-	-		-		-
Daucus carota	eudicotyledons	79200	ASM162521v1	GCA_001625215.1	-	-				-
Digitaria exilis	Liliopsida	1010633	Fonio_CM05836	GCA_902859565.1	-	-		-		-
Dioscorea rotundata	Liliopsida	55577	TDr96_F1_v2_PseudoChromosome	GCA_009730915.1	-	-		-		-
Echinochloa crus-galli	Liliopsida	90397	ec_v3	GCA_020466025.1	-	-		-		-
Eragrostis curvula	Liliopsida	38414	CERZOS_E.curvula1.0	GCA_007726485.1	-	-		-		-
Eragrostis tef	Liliopsida	110835	Salk_teff_dabbi_3.0	GCA_024500355.1	-	-		-		-
Eucalyptus grandis	eudicotyledons	71139	Egrandis1_0	GCA_000612305.1	-	-		-		-
Eutrema salsugineum	eudicotyledons	72664	Eutsalg1_0	GCA_000478725.1	-	-				-
Ficus carica	eudicotyledons	3494	UNIPI_FiCari_1.0	GCA_009761775.1	-	-		-		-
Fraxinus excelsior	eudicotyledons	38873	BATG-0.5	GCA_900149125.1	-	-		-		-
Galdieria sulphuraria	Rhodophyta	130081	ASM34128v1	GCA_000341285.1	-	-		-		-
Glycine max	eudicotyledons	3847	Glycine_max_v2.1	GCA_000004515.4	-					-
Gossypium raimondii	eudicotyledons	29730	Graimondii2_0_v6	GCA_000327365.1	-	-				-
Helianthus annuus	eudicotyledons	4232	HanXRQr2.0-SUNRISE	GCA_002127325.2		-		-		-
Hordeum vulgare	Liliopsida	112509	MorexV3_pseudomolecules_assembly	GCA_904849725.1						-
Hordeum vulgare GoldenPromise	Liliopsida	112509	GPv1	GCA_902500625.1	-	-	-	-	-	-
Hordeum vulgare TRITEX	Liliopsida	112509	Morex_V2_scaf	GCA_903813605.1	-	-	-	-	-	-
Ipomoea triloba	eudicotyledons	35885	ASM357664v1	GCA_003576645.1	-	-		-		-
Juglans regia	eudicotyledons	51240	Walnut_2.0	GCA_001411555.2	-	-		-		-

Name	Classification	Taxon ID	Assembly	Accession	Variation database	Regulation database	Whole genome alignments	Other alignments	In peptide compara	In pan-taxonomic compara
Kalanchoe fedtschenkoi	eudicotyledons	63787	K_fedtschenkoi_M2_v1	GCA_002312845.1	-	-		-		-
Lactuca sativa	eudicotyledons	4236	Lsat_Salinas_v7	GCA_002870075.2		-		-		-
Leersia perrieri	Liliopsida	77586	Lperr_V1.4	GCA_000325765.3	-	-				-
Lolium perenne	Liliopsida	4522	MPB_Lper_Kyuss_1697	GCA_019359855.1	-	-		-		-
Lupinus angustifolius	eudicotyledons	3871	LupAngTanjil_v1.0	GCA_001865875.1	-	-				-
Malus domestica Golden	eudicotyledons	3750	ASM211411v1	GCA_002114115.1		-		-		-
Manihot esculenta	eudicotyledons	3983	Manihot_esculenta_v6	GCA_001659605.1	-	-				-
Marchantia polymorpha	Embryophyta	3197	Marchanta_polymorpha_v1	GCA_003032435.1	-	-		-		
Medicago truncatula	eudicotyledons	3880	MedtrA17_4.0	GCA_000219495.2	-					-
Musa acuminata	Liliopsida	214687	Musa_acuminata_v2	GCA_904845865.1	-	-				-
Nicotiana attenuata	eudicotyledons	49451	NIATTr2	GCA_001879085.1	-					-
Nymphaea colorata	Embryophyta	210225	ASM883128v1	GCA_008831285.1	-	-		-		-
Olea europaea	eudicotyledons	158383	OLEA9	GCA_902713445.1	-	-		-		-
Olea europaea var. sylvestris	eudicotyledons	158386	O_europaea_v1	GCA_002742605.1	-	-	-	-	-	-
Oryza barthii	Liliopsida	65489	O.barthii_v1	GCA_000182155.2	-					-
Oryza brachyantha	Liliopsida	4533	Oryza_brachyantha.v1.4b	GCA_000231095.2	-	-				-
Oryza glaberrima	Liliopsida	4538	Oryza_glaberrima_V1	GCA_000147395.1						-
Oryza glumipatula	Liliopsida	40148	Oryza_glumaepatula_v1.5	GCA_000576495.1						-
Oryza longistaminata	Liliopsida	4528	O_longistaminata_v1.0	GCA_000789195.1	-					-
Oryza meridionalis	Liliopsida	40149	Oryza_meridionalis_v1.3	GCA_000338895.2	-					-
Oryza nivara	Liliopsida	4536	Oryza_nivara_v1.0	GCA_000576065.1	-					-
Oryza punctata	Liliopsida	4537	Oryza_punctata_v1.2	GCA_000573905.1	-					-
Oryza rufipogon	Liliopsida	4529	OR_W1943	GCA_000817225.1	-					-
Oryza sativa (Geng/Japonica-sbtrp var. Chao Meo)	Liliopsida	1736656	Os132278RS1	GCA_009831315.1	-	-	-	-	-	-
Oryza sativa (Geng/Japonica-trop1 var. Azucena)	Liliopsida	1736656	AzucenaRS1	GCA_009830595.1	-	-	-	-	-	-
Oryza sativa (Geng/Japonica-trop2 var. Ketan Nangka)	Liliopsida	1736656	Os128077RS1	GCA_009831275.1	-	-	-	-	-	-
Oryza sativa (Xian/Indica-1A var. Zhenshan 97)	Liliopsida	39946	ZS97RS3	GCA_001618795.1	-	-	-	-	-	-
Oryza sativa (Xian/Indica-1B1 var. IR64)	Liliopsida	39946	OsIR64RS1	GCA_009914875.1	-	-	-	-	-	-
Oryza sativa (Xian/Indica-1B2 var. PR106)	Liliopsida	39946	Os127742RS1	GCA_009831045.1	-	-	-	-	-	-
Oryza sativa (Xian/Indica-2A var. Gobol Sail)	Liliopsida	39946	Os132424RS1	GCA_009831025.1	-	-	-	-	-	-
Oryza sativa (Xian/Indica-2B var. Larha Mugad)	Liliopsida	39946	Os125619RS1	GCA_009831355.1	-	-	-	-	-	-
Oryza sativa (Xian/Indica-3A var. Lima)	Liliopsida	39946	Os127564RS1	GCA_009829395.1	-	-	-	-	-	-
Oryza sativa (Xian/Indica-3B1 var. Khao Yai Guang)	Liliopsida	39946	Os127518RS1	GCA_009831295.1	-	-	-	-	-	-
Oryza sativa (Xian/Indica-3B2 var. Liu Xu)	Liliopsida	39946	Os125827RS1	GCA_009829375.1	-	-	-	-	-	-
Oryza sativa (Xian/Indica-adm var. Minghui 63)	Liliopsida	39946	MH63RS2	GCA_001618785.1	-	-	-	-	-	-
Oryza sativa (circum-Aus1 var. N22)	Liliopsida	1736659	OsN22RS2	GCA_001952365.2	-	-	-	-	-	-
Oryza sativa (circum-Aus2 var. Natel Boro)	Liliopsida	1736659	Os127652RS1	GCA_009831335.1	-	-	-	-	-	-
Oryza sativa (circum-Basmati var. ARC 10497)	Liliopsida	1736658	Os117425RS1	GCA_009831255.1	-	-	-	-	-	-
Oryza sativa Indica Group	Liliopsida	39946	ASM465v1	GCA_000004655.2						-
Oryza sativa Japonica Group	Liliopsida	39947	IRGSP-1.0	GCA_001433935.1						
Ostreococcus lucimarinus	Chlorophyta	436017	ASM9206v1	GCA_000092065.1	-	-	-			-
Panicum hallii FIL2	Liliopsida	206008	PHallii_v3.1	GCA_002211085.2	-	-	-	-	-	-
Panicum hallii HAL2	Liliopsida	1504633	PhalliiHAL_v2.1	GCA_003061485.1	-	-		-		-
Papaver somniferum	Embryophyta	3469	ASM357369v1	GCA_003573695.1	-	-		-		-
Phaseolus vulgaris	eudicotyledons	3885	PhaVulg1_0	GCA_000499845.1						-
Physcomitrium patens	Embryophyta	3218	Phypa_V3	GCA_000002425.2	-	-				
Pistacia vera	eudicotyledons	55513	PisVer_v2	GCA_008641045.1	-	-		-		-
Pisum sativum	eudicotyledons	3888	Pisum_sativum_v1a	GCA_900700895.2	-	-		-		-
Populus trichocarpa	eudicotyledons	3694	Pop_tri_v4	GCA_000002775.4	-	-		-		-
Prunus avium	eudicotyledons	42229	PAV_r1.0	GCA_002207925.1	-	-		-		-
Prunus dulcis	eudicotyledons	3755	ALMONDv2	GCA_902201215.1	-	-		-		-
Prunus persica	eudicotyledons	3760	Prunus_persica_NCBIv2	GCA_000346465.2	-	-				-
Quercus lobata	eudicotyledons	97700	ValleyOak3.0	GCA_001633185.2	-	-		-		-
Quercus suber	eudicotyledons	58331	CorkOak1.0	GCA_002906115.4	-	-		-		-
Rosa chinensis	eudicotyledons	74649	RchiOBHm-V2	GCA_002994745.2	-	-		-		-

Name	Classification	Taxon ID	Assembly	Accession	Variation database	Regulation database	Whole genome alignments	Other alignments	In peptide compara	In pan-taxonomic compara
Saccharum spontaneum	Liliopsida	62335	Sspon.HiC_chr_asm	GCA_003544955.1	-	-		-		-
Secale cereale	Liliopsida	4550	Rye_Lo7_2018_v1p1p1	GCA_902687465.1	-	-		-		-
Selaginella moellendorffii	Embryophyta	88036	v1.0	GCA_000143415.1	-	-				
Sesamum indicum	eudicotyledons	4182	S_indicum_v1.0	GCA_000512975.1	-	-		-		-
Setaria italica	Liliopsida	4555	Setaria_italica_v2.0	GCA_000263155.2	-	-				-
Setaria viridis	Liliopsida	4556	Setaria_viridis_v2.0	GCA_005286985.1	-	-		-		-
Solanum lycopersicum	eudicotyledons	4081	SL3.0	GCA_000188115.3						
Solanum tuberosum	eudicotyledons	4113	SolTub_3.0	GCA_000226075.1	-	-				-
Solanum tuberosum RH89-039-16	eudicotyledons	4113	ASM1418947v1	GCA_014189475.1	-	-	-	-	-	-
Sorghum bicolor	Liliopsida	4558	Sorghum_bicolor_NCBIv3	GCA_000003195.3		-				-
Theobroma cacao Belizian Criollo B97-61/B2	eudicotyledons	3641	Criollo_cocoa_genome_V2	GCA_000208745.2	-	-	-	-	-	-
Theobroma cacao Matina 1-6	eudicotyledons	3641	Theobroma_cacao_20110822	GCA_000403535.1		-				-
Trifolium pratense	eudicotyledons	57577	Trpr	GCA_900079335.1	-	-				-
Triticum aestivum	Liliopsida	4565	IWGSC	GCA_900519105.1						-
Triticum aestivum Arinalrfor	Liliopsida	4565	PGSBv2.1	GCA_903993985.1	-	-		-		-
Triticum aestivum Cadenza	Liliopsida	4565	EIv1.1	GCA_902810645.1	-	-	-	-	-	-
Triticum aestivum Claire	Liliopsida	4565	EIv1.1	GCA_902810655.1	-	-	-	-	-	-
Triticum aestivum Jagger	Liliopsida	4565	PGSBv2.1	GCA_903993795.1	-	-		-		-
Triticum aestivum Julius	Liliopsida	4565	PGSBv2.1	GCA_903994195.1	-	-		-		-
Triticum aestivum Kariega	Liliopsida	4565	Tae_Kariega_v1	GCA_910594105.1	-	-	-	-	-	-
Triticum aestivum Lancer	Liliopsida	4565	PGSBv2.1	GCA_903993975.1	-	-		-		-
Triticum aestivum Landmark	Liliopsida	4565	PGSBv2.1	GCA_903995565.1	-	-		-		-
Triticum aestivum Mace	Liliopsida	4565	PGSBv2.1	GCA_903994175.1	-	-		-		-
Triticum aestivum Norin61	Liliopsida	4565	PGSBv2.1	GCA_904066035.1	-	-		-		-
Triticum aestivum Paragon	Liliopsida	4565	EIv1.1	GCA_902810665.1	-	-	-	-	-	-
Triticum aestivum Refseqv2	Liliopsida	4565	IWGSC_RefSeq_v2.1	GCA_018294505.1	-	-	-	-	-	-
Triticum aestivum Renan	Liliopsida	4565	Triticum_aestivum_Renan_v2.1	GCA_937894285.1	-	-	-	-	-	-
Triticum aestivum Robigus	Liliopsida	4565	EIv1.1	GCA_902810685.1	-	-	-	-	-	-
Triticum aestivum Stanley	Liliopsida	4565	PGSBv2.2	GCA_903994155.1	-	-		-		-
Triticum aestivum Sy Mattis	Liliopsida	4565	PGSBv2.1	GCA_903994185.1	-	-	-	-	-	-
Triticum aestivum Weebill	Liliopsida	4565	WeebilV1	GCA_902810675.1	-	-	-	-	-	-
Triticum dicoccoides	Liliopsida	85692	WEWSeq_v.1.0	GCA_002162155.1	-					-
Triticum spelta	Liliopsida	58933	PGSBv2.0	GCA_903994165.1	-	-	-	-	-	-
Triticum turgidum	Liliopsida	4567	Svevo.v1	GCA_900231445.1		-		-		-
Triticum urartu	Liliopsida	4572	IGDB	GCA_003073215.1	-	-				-
Vigna angularis	eudicotyledons	3914	Vigan1.1	GCA_001190045.1	-					-
Vigna radiata	eudicotyledons	3916	Vradiata_ver6	GCA_000741045.2	-					-
Vigna unguiculata	eudicotyledons	3917	ASM411807v1	GCA_004118075.1	-	-		-		-
Vitis vinifera	eudicotyledons	29760	PN40024.v4	GCA_910591555.1				-		
Zea mays	Liliopsida	4577	Zm-B73-REFERENCE-NAM-5.0	GCA_902167145.1				-		-

Annex5 Minimum information required for each each MIxS v6.0 checklist

TYPE OF INFORMATION			DESCRIPTOR DEFINITION AND SYNTAX									CHECKLIST REQUIREMENT (M:mandatory; C: conditional mandatory;X:optional; E:Environment dependent; (-): not apllicable)										
Type	**Section**	sub-section	**MIXS ID**	**Structured comment name**	**Item (rdfs:label)**	**Definition**	**Expected value**	**Value syntax**	**Example**	**Preferred unit**	**Occurence**	**migs_eu**	**migs_ba**	**migs_pl**	**migs_vi**	**migs_org**	**mims**	**mimarks_s**	**mimarks_c**	**misag**	**mimag**	**miuvig**
Sample, Environmental Context and investigation	investigation	**Project name**	MIXS:0000092	project_name	project name	Name of the project within which the sequencing was organized		{text}	Forest soil metagenome		1	M	M	M	M	M	M	M	M	M	M	M
Sample, Environmental Context and investigation	investigation	**Biological Source**	MIXS:0001107	samp_name	sample name	A local identifier or name that for the material sample used for extracting nucleic acids, and subsequent sequencing. It can refer either to the original material collected or to any derived sub-samples. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. INSDC requires every sample name from a single Submitter to be unique. Use of a globally unique identifier for the field source_mat_id is recommended in addition to sample_name.	text	{text}	ISDsoil1		1	M	M	M	M	M	M	M	M	M	M	M
Sample, Environmental Context and investigation	investigation	**Biological Source**	MIXS:0001320	samp_taxon_id	taxonomy ID of DNA sample	NCBI taxon id of the sample. Maybe be a single taxon or mixed taxa sample. Use 'synthetic metagenome' for mock community/positive controls, or 'blank sample' for negative controls.	Taxonomy ID	{text} [NCBI:txid]	Gut Metagenome [NCBI:txid749906]		1	M	M	M	M	M	M	M	M	M	M	M
Sample, Environmental Context and investigation	investigation	**Environmental Parameters**	MIXS:0000008	experimental_factor	experimental factor	Experimental factors are essentially the variable aspects of an experiment design which can be used to describe an experiment, or set of experiments, in an increasingly detailed manner. This field accepts ontology terms from Experimental Factor Ontology (EFO) and/or Ontology for Biomedical Investigations (OBI). For a browser of EFO (v 2.95) terms, please see http://purl.bioontology.org/ontology/EFO; for a browser of OBI (v 2018-02-12) terms please see http://purl.bioontology.org/ontology/OBI	text or EFO and/or OBI	{termLabel} {[termID]}|{text}	time series design [EFO:EFO_0001779]		1	X	X	X	X	X	C	C	X	C	C	C
Sample, Environmental Context and investigation	environment	**Geographical Location**	MIXS:0000009	lat_lon	geographic location (latitude and longitude)	The geographical origin of the sample as defined by latitude and longitude. The values should be reported in decimal degrees and in WGS84 system	decimal degrees, limit to 8 decimal points	{float} {float}	50.586825 6.408977		1	M	M	M	M	M	M	M	M	M	M	M
Sample, Environmental Context and investigation	environment	**Geographical Location**	MIXS:0000018	depth	depth	The vertical distance below local surface, e.g. for sediment or soil samples depth is measured from sediment or soil surface, respectively. Depth can be reported as an interval for subsurface samples.	measurement value	{float} {unit}	10 meter		1	E	E	E	E	E	E	E	E	E	E	E
Sample, Environmental Context and investigation	environment	**Geographical Location**	MIXS:0000094	alt	altitude	Altitude is a term used to identify heights of objects such as airplanes, space shuttles, rockets, atmospheric balloons and heights of places such as atmospheric layers and clouds. It is used to measure the height of an object which is above the earth's surface. In this context, the altitude measurement is the vertical distance between the earth's surface above sea level and the sampled position in the air	measurement value	{float} {unit}	100 meter		1	E	E	E	E	E	E	E	E	E	E	E
Sample, Environmental Context and investigation	environment	**Geographical Location**	MIXS:0000093	elev	elevation	Elevation of the sampling site is its height above a fixed reference point, most commonly the mean sea level. Elevation is mainly used when referring to points on the earth's surface, while altitude is used for points above the surface, such as an aircraft in flight or a spacecraft in orbit.	measurement value	{float} {unit}	100 meter		1	E	E	E	E	E	E	E	E	E	E	E
Sample, Environmental Context and investigation	environment	**Environmental Parameters**	MIXS:0000113	temp	temperature	Temperature of the sample at the time of sampling.	measurement value	{float} {unit}	25 degree Celsius	degree Celsius	1	E	E	E	E	E	E	E	E	E	E	E
Sample, Environmental Context and investigation	environment	**Geographical Location**	MIXS:0000010	geo_loc_name	geographic location (country and/or sea,region)	The geographical origin of the sample as defined by the country or sea name followed by specific region name. Country or sea names should be chosen from the INSDC country list (http://insdc.org/country.html), or the GAZ ontology (http://purl.bioontology.org/ontology/GAZ)	country or sea name (INSDC or GAZ): region(GAZ), specific location name	{term}: {term}, {text}	USA: Maryland, Bethesda		1	M	M	M	M	M	M	M	M	M	M	M
Sample, Environmental Context and investigation	environment	**Biological Source**	MIXS:0000011	collection_date	collection date	The time of sampling, either as an instance (single point in time) or interval. In case no exact time is available, the date/time can be right truncated i.e. all of these are valid times: 2008-01-23T19:23:10+00:00; 2008-01-23T19:23:10; 2008-01-23; 2008-01; 2008; Except: 2008-01; 2008 all are ISO8601 compliant	date and time	{timestamp}	2018-05-11T10:00:00+01:00; 2018-05-11		1	M	M	M	M	M	M	M	M	M	M	M
Sample, Environmental Context and investigation	investigation	**Biological Source**	MIXS:0001321	neg_cont_type	negative control type	The substance or equipment used as a negative control in an investigation	enumeration or text	[distilled water|phosphate buffer|empty collection device|empty collection tube|DNA-free PCR mix|sterile swab |sterile syringe]			1	C	C	C	C	C	C	C	C	C	C	C
Sample, Environmental Context and investigation	investigation	**Biological Source**	MIXS:0001322	pos_cont_type	positive control type	The substance, mixture, product, or apparatus used to verify that a process which is part of an investigation delivers a true positive.		{term} or {text}			1	C	C	C	C	C	C	C	C	C	C	C
Sample, Environmental Context and investigation	environment	**Environmental Parameters**	MIXS:0000012	env_broad_scale	broad-scale environmental context	Report the major environmental system the sample or specimen came from. The system(s) identified should have a coarse spatial grain, to provide the general environmental context of where the sampling was done (e.g. in the desert or a rainforest). We recommend using subclasses of EnvO's biome class: http://purl.obolibrary.org/obo/ENVO_00000428. EnvO documentation about how to use the field: https://github.com/EnvironmentOntology/envo/wiki/Using-ENVO-with-MIxS	The major environment type(s) where the sample was collected. Recommend subclasses of biome [ENVO:00000428]. Multiple terms can be separated by one or more pipes.	{termLabel} {[termID]}	oceanic epipelagic zone biome [ENVO:01000033] for annotating a water sample from the photic zone in middle of the Atlantic Ocean		1	M	M	M	M	M	M	M	M	M	M	M
Sample, Environmental Context and investigation	environment	**Environmental Parameters**	MIXS:0000013	env_local_scale	local environmental context	Report the entity or entities which are in the sample or specimen's local vicinity and which you believe have significant causal influences on your sample or specimen. We recommend using EnvO terms which are of smaller spatial grain than your entry for env_broad_scale. Terms, such as anatomical sites, from other OBO Library ontologies which interoperate with EnvO (e.g. UBERON) are accepted in this field. EnvO documentation about how to use the field: https://github.com/EnvironmentOntology/envo/wiki/Using-ENVO-with-MIxS.	Environmental entities having causal influences upon the entity at time of sampling.	{termLabel} {[termID]}	litter layer [ENVO:01000338]; Annotating a pooled sample taken from various vegetation layers in a forest consider: canopy [ENVO:00000047]|herb and fern layer [ENVO:01000337]|litter layer [ENVO:01000338]|understory [01000335]|shrub layer [ENVO:01000336].		1	M	M	M	M	M	M	M	M	M	M	M

TYPE OF INFORMATION			DESCRIPTOR DEFINITION AND SYNTAX									CHECKLIST REQUIREMENT (M:mandatory; C: conditional mandatory;X:optional; E:Environment dependent; (-): not apllicable)										
Type	**Section**	sub-section	**MIXS ID**	**Structured comment name**	**Item (rdfs:label)**	**Definition**	**Expected value**	**Value syntax**	**Example**	**Preferred unit**	**Occurence**	**migs_eu**	**migs_ba**	**migs_pl**	**migs_vi**	**migs_org**	**mims**	**mimarks_s**	**mimarks_c**	**misag**	**mimag**	**miuvig**
Sample, Environmental Context and investigation	environment	**Environmental Parameters**	MIXS:0000014	env_medium	environmental medium	Report the environmental material(s) immediately surrounding the sample or specimen at the time of sampling. We recommend using subclasses of 'environmental material' (http://purl.obolibrary.org/obo/ENVO_00010483). EnvO documentation about how to use the field: https://github.com/EnvironmentOntology/envo/wiki/Using-ENVO-with-MIxS . Terms from other OBO ontologies are permissible as long as they reference mass/volume nouns (e.g. air, water, blood) and not discrete, countable entities (e.g. a tree, a leaf, a table top).	The material displaced by the entity at time of sampling. Recommend subclasses of environmental material [ENVO:00010483].	{termLabel} {[termID]}	soil [ENVO:00001998]; Annotating a fish swimming in the upper 100 m of the Atlantic Ocean, consider: ocean water [ENVO:00002151]. Example: Annotating a duck on a pond consider: pond water [ENVO:00002228]|air [ENVO_00002005]		1	M	M	M	M	M	M	M	M	M	M	M
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000020	subspecf_gen_lin	subspecific genetic lineage	Information about the genetic distinctness of the sequenced organism below the subspecies level, e.g., serovar, serotype, biotype, ecotype, or any relevant genetic typing schemes like Group I plasmid. Subspecies should not be recorded in this term, but in the NCBI taxonomy. Supply both the lineage name and the lineage rank separated by a colon, e.g., biovar:abc123.	Genetic lineage below lowest rank of NCBI taxonomy, which is subspecies, e.g. serovar, biotype, ecotype, variety, cultivar.	{rank name}:{text}	serovar:Newport		1	C	C	C	C	C	-	-	C	-	-	-
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000021	ploidy	ploidy	The ploidy level of the genome (e.g. allopolyploid, haploid, diploid, triploid, tetraploid). It has implications for the downstream study of duplicated gene and regions of the genomes (and perhaps for difficulties in assembly). For terms, please select terms listed under class ploidy (PATO:001374) of Phenotypic Quality Ontology (PATO), and for a browser of PATO (v 2018-03-27) please refer to http://purl.bioontology.org/ontology/PATO	PATO	{termLabel} {[termID]}	allopolyploidy [PATO:0001379]		1	X	-	-	-	-	-	-	-	-	-	-
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000022	num_replicons	number of replicons	Reports the number of replicons in a nuclear genome of eukaryotes, in the genome of a bacterium or archaea or the number of segments in a segmented virus. Always applied to the haploid chromosome count of a eukaryote	for eukaryotes and bacteria: chromosomes (haploid count); for viruses: segments	{integer}	2		1	X	M	-	C	-	-	-	-	-	-	-
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000023	extrachrom_elements	extrachromosomal elements	Do plasmids exist of significant phenotypic consequence (e.g. ones that determine virulence or antibiotic resistance). Megaplasmids? Other plasmids (borrelia has 15+ plasmids)	number of extrachromosomal elements	{integer}	5		1	X	C	-	-	C	-	-	X	-	-	-
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000024	estimated_size	estimated size	The estimated size of the genome prior to sequencing. Of particular importance in the sequencing of (eukaryotic) genome which could remain in draft form for a long or unspecified period.	number of base pairs	{integer} bp	300000 bp		1	X	X	X	X	X	-	-	-	-	-	X
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000025	ref_biomaterial	reference for biomaterial	Primary publication if isolated before genome publication; otherwise, primary genome report.	PMID, DOI or URL	{PMID}|{DOI}|{URL}	doi:10.1016/j.syapm.2018.01.009		1	X	M	X	X	X	X	-	-	X	X	X
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000026	source_mat_id	source material identifiers	A unique identifier assigned to a material sample (as defined by http://rs.tdwg.org/dwc/terms/materialSampleID, and as opposed to a particular digital record of a material sample) used for extracting nucleic acids, and subsequent sequencing. The identifier can refer either to the original material collected or to any derived sub-samples. The INSDC qualifiers /specimen_voucher, /bio_material, or /culture_collection may or may not share the same value as the source_mat_id field. For instance, the /specimen_voucher qualifier and source_mat_id may both contain 'UAM:Herps:14' , referring to both the specimen voucher and sampled tissue with the same identifier. However, the /culture_collection qualifier may refer to a value from an initial culture (e.g. ATCC:11775) while source_mat_id would refer to an identifier from some derived culture from which the nucleic acids were extracted (e.g. xatc123 or ark:/2154/R2).	for cultures of microorganisms: identifiers for two culture collections; for other material a unique arbitrary identifer	{text}	MPI012345		m	C	C	C	C	C	C	C	C	C	C	C
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000027	pathogenicity	known pathogenicity	To what is the entity pathogenic	names of organisms that the entity is pathogenic to	{text}	human, animal, plant, fungi, bacteria		1	C	C	-	C	-	-	-	-	-	-	X
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000028	biotic_relationship	observed biotic relationship	Description of relationship(s) between the subject organism and other organism(s) it is associated with. E.g., parasite on species X; mutualist with species Y. The target organism is the subject of the relationship, and the other organism(s) is the object	enumeration	[free living|parasitism|commensalism|symbiotic|mutualism]	free living		1	X	C	-	X	-	-	-	C	-	-	X
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000029	specific_host	host scientific name	Report the host's taxonomic name and/or NCBI taxonomy ID.	host scientific name, taxonomy ID	{text}|{NCBI taxid}	Homo sapiens and/or 9606		1	X	C	C	C	-	-	-	-	-	-	X
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000030	host_spec_range	host specificity or range	The range and diversity of host species that an organism is capable of infecting, defined by NCBI taxonomy identifier.	NCBI taxid	{integer}	9606		m	X	X	X	C	-	-	-	-	-	-	X
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000031	host_disease_stat	host disease status	List of diseases with which the host has been diagnosed; can include multiple diagnoses. The value of the field depends on host; for humans the terms should be chosen from the DO (Human Disease Ontology) at https://www.disease-ontology.org, non-human host diseases are free text	disease name or Disease Ontology term	{termLabel} {[termID]}|{text}	rabies [DOID:11260]		m	X	C	-	C	-	-	-	-	-	-	C

TYPE OF INFORMATION			DESCRIPTOR DEFINITION AND SYNTAX									CHECKLIST REQUIREMENT (M:mandatory; C: conditional mandatory;X:optional; E:Environment dependent; (-): not apllicable)										
Type	**Section**	sub-section	**MIXS ID**	**Structured comment name**	**Item (rdfs:label)**	**Definition**	**Expected value**	**Value syntax**	**Example**	**Preferred unit**	**Occurence**	**migs_eu**	**migs_ba**	**migs_pl**	**migs_vi**	**migs_org**	**mims**	**mimarks_s**	**mimarks_c**	**misag**	**mimag**	**miuvig**
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000032	trophic_level	trophic level	Trophic levels are the feeding position in a food chain. Microbes can be a range of producers (e.g. chemolithotroph)	enumeration	[autotroph|carboxydotroph|chemoautotroph|chemoheterotroph|chemolithoautotroph|chemolithotroph|chemoorganoheterotroph|chemoorganotroph|chemosynthetic|chemotroph|copiotroph|diazotroph|facultative|autotroph|heterotroph|lithoautotroph|lithoheterotroph|lithotroph|methanotroph|methylotroph|mixotroph|obligate|chemoautolithotroph|oligotroph|organoheterotroph|organotroph|photoautotroph|photoheterotroph|photolithoautotroph|photolithotroph|photosynthetic|phototrop	heterotroph		1	C	C	-	-	-	-	-	C	-	-	-
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000033	propagation	propagation	The type of reproduction from the parent stock. Values for this field is specific to different taxa. For phage or virus: lytic/lysogenic/temperate/obligately lytic. For plasmids: incompatibility group. For eukaryotes: sexual/asexual.	for virus: lytic, lysogenic, temperate, obligately lytic; for plasmid: incompatibility group; for eukaryote: asexual, sexual; other more specific values (e.g., incompatibility group) are allowed	{text}	lytic		1	C	-	M	M	-	-	-	-	-	-	-
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000034	encoded_traits	encoded traits	Should include key traits like antibiotic resistance or xenobiotic degradation phenotypes for plasmids, converting genes for phage	for plasmid: antibiotic resistance; for phage: converting genes	{text}	beta-lactamase class A		1	-	X	C	C	-	-	-	-	-	-	-
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000015	rel_to_oxygen	relationship to oxygen	Is this organism an aerobe, anaerobe? Please note that aerobic and anaerobic are valid descriptors for microbial environments	enumeration	[aerobe|anaerobe|facultative|microaerophilic|microanaerobe|obligate aerobe|obligate anaerobe]	aerobe		1	-	C	-	-	-	X	X	C	X	X	-
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000003	isol_growth_condt	isolation and growth condition	Publication reference in the form of pubmed ID (pmid), digital object identifier (doi) or url for isolation and growth condition specifications of the organism/material	PMID,DOI or URL	{PMID}|{DOI}|{URL}	doi: 10.1016/j.syapm.2018.01.009		1	M	M	M	M	M	-	-	M	-	-	-
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000002	samp_collec_device	sample collection device	The device used to collect an environmental sample. This field accepts terms listed under environmental sampling device (http://purl.obolibrary.org/obo/ENVO). This field also accepts terms listed under specimen collection device (http://purl.obolibrary.org/obo/GENEPIO_0002094).	device name	{termLabel} {[termID]}|{text}	swab, biopsy, niskin bottle, push core, drag swab [GENEPIO:0002713]		1	X	X	X	X	X	C	C	X	C	C	C
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0001225	samp_collec_method	sample collection method	The method employed for collecting the sample.	PMID,DOI,url , or text	{PMID}|{DOI}|{URL}|{text}	swabbing		1	X	X	X	X	X	C	C	X	C	C	C
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000016	samp_mat_process	sample material processing	A brief description of any processing applied to the sample during or after retrieving the sample from environment, or a link to the relevant protocol(s) performed.	text	{text}	filtering of seawater, storing samples in ethanol		1	X	X	X	X	X	C	C	C	C	C	C
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000017	size_frac	size fraction selected	Filtering pore size used in sample preparation	filter size value range	{float}-{float} {unit}	0-0.22 micrometer		1	-	-	-	-	-	X	X	-	X	X	C
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000001	samp_size	amount or size of sample collected	The total amount or size (volume (ml), mass (g) or area (m2)) of sample collected.	measurement value	{float} {unit}	5 liter	millliter, gram, milligram, liter	1	X	X	X	X	X	C	C	X	C	C	C
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000111	samp_vol_we_dna_ext	sample volume or weight for DNA extraction	Volume (ml) or mass (g) of total collected sample processed for DNA extraction. Note: total sample collected should be entered under the term Sample Size (MIXS:0000001).	measurement value	{float} {unit}	1500 milliliter	millliter, gram, milligram, square centimeter	1	X	X	X	X	X	X	X	X	X	X	X

TYPE OF INFORMATION			DESCRIPTOR DEFINITION AND SYNTAX									CHECKLIST REQUIREMENT (M:mandatory; C: conditional mandatory;X:optional; E:Environment dependent; (-): not apllicable)										
Type	**Section**	**sub-section**	**MIXS ID**	**Structured comment name**	**Item (rdfs:label)**	**Definition**	**Expected value**	**Value syntax**	**Example**	**Preferred unit**	**Occurence**	**migs_eu**	**migs_ba**	**migs_pl**	**migs_vi**	**migs_org**	**mims**	**mimarks_s**	**mimarks_c**	**misag**	**mimag**	**miuvig**
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000035	source_uvig	source of UViGs	Type of dataset from which the UViG was obtained	enumeration	[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)		1	-	-	-	-	-	-	-	-	-	-	M
Sample, Environmental Context and investigation	nucleic acid sequence source	**Biological Source**	MIXS:0000036	virus_enrich_appr	virus enrichment approach	List of approaches used to enrich the sample for viruses, if any	enumeration	[filtration|ultrafiltration|centrifugation|ultracentrifugation|PEG Precipitation|FeCl Precipitation|CsCl density gradient|DNAse|RNAse|targeted sequence capture|other|none]	filtration + FeCl Precipitation + ultracentrifugation + DNAse		1	-	-	-	C	-	-	-	-	-	-	M
Sequencing and Experimental Details	sequencing	**Library Preparation**	MIXS:0000037	nucl_acid_ext	nucleic acid extraction	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 or URL	{PMID}|{DOI}|{URL}	https://mobio.com/media/wysiwyg/pdfs/protocols/12888.pdf		1	C	C	C	C	C	C	C	C	C	C	C
Sequencing and Experimental Details	sequencing	**Library Preparation**	MIXS:0000038	nucl_acid_amp	nucleic acid amplification	A link to a literature reference, electronic resource or a standard operating procedure (SOP), that describes the enzymatic amplification (PCR, TMA, NASBA) of specific nucleic acids	PMID, DOI or URL	{PMID}|{DOI}|{URL}	https://phylogenomics.me/protocols/16s-pcr-protocol/		1	C	C	C	C	C	C	C	C	C	C	C
Sequencing and Experimental Details	sequencing	**Library Preparation**	MIXS:0000039	lib_size	library size	Total number of clones in the library prepared for the project	number of clones	{integer}	50		1	X	X	X	X	X	C	C	-	C	C	C
Sequencing and Experimental Details	sequencing	**Library Preparation**	MIXS:0000040	lib_reads_seqd	library reads sequenced	Total number of clones sequenced from the library	number of reads sequenced	{integer}	20		1	X	X	X	X	X	C	C	-	C	C	C
Sequencing and Experimental Details	sequencing	**Library Preparation**	MIXS:0000041	lib_layout	library layout	Specify whether to expect single, paired, or other configuration of reads	enumeration	[paired|single|vector|other]	paired		1	X	X	X	X	X	C	C	-	C	C	C
Sequencing and Experimental Details	sequencing	**Library Preparation**	MIXS:0000042	lib_vector	library vector	Cloning vector type(s) used in construction of libraries	vector	{text}	Bacteriophage P1		1	X	X	X	X	X	C	C	-	C	C	C
Sequencing and Experimental Details	sequencing	**Library Preparation**	MIXS:0000043	lib_screen	library screening strategy	Specific enrichment or screening methods applied before and/or after creating libraries	screening strategy name	{text}	enriched, screened, normalized		1	X	X	X	X	X	C	C	-	C	C	C
Sequencing and Experimental Details	sequencing	**sequencing**	MIXS:0000044	target_gene	target gene	Targeted gene or locus name for marker gene studies	gene name	{text}	16S rRNA, 18S rRNA, nif, amoA, rpo		1	-	-	-	-	-	-	M	M	-	-	-
Sequencing and Experimental Details	sequencing	**sequencing**	MIXS:0000045	target_subfragment	target subfragment	Name of subfragment of a gene or locus. Important to e.g. identify special regions on marker genes like V6 on 16S rRNA	gene fragment name	{text}	V6, V9, ITS		1	-	-	-	-	-	-	C	C	-	-	-
Sequencing and Experimental Details	sequencing	**sequencing**	MIXS:0000046	pcr_primers	pcr primers	PCR primers that were used to amplify the sequence of the targeted gene, locus or subfragment. This field should contain all the primers used for a single PCR reaction if multiple forward or reverse primers are present in a single PCR reaction. The primer sequence should be reported in uppercase letters	FWD: forward primer sequence;REV:reverse primer sequence	FWD:{dna};REV:{dna}	FWD:GTGCCAGCMGCCGCGGTAA;REV:GGACTACHVGGGTWTCTAAT		1	-	-	-	-	-	-	C	C	-	-	-
Sequencing and Experimental Details	sequencing	**sequencing**	MIXS:0000047	mid	multiplex identifiers	Molecular barcodes, called Multiplex Identifiers (MIDs), that are used to specifically tag unique samples in a sequencing run. Sequence should be reported in uppercase letters	multiplex identifier sequence	{dna}	GTGAATAT		1	-	-	-	-	-	C	C	-	C	C	C
Sequencing and Experimental Details	sequencing	**sequencing**	MIXS:0000048	adapters	adapters	Adapters provide priming sequences for both amplification and sequencing of the sample-library fragments. Both adapters should be reported; in uppercase letters	adapter A and B sequence	{dna};{dna}	AATGATACGGCGACCACCGAGATCTACACGCT;CAAGCAGAAGACGGCATACGAGAT		1	C	C	C	C	C	C	C	-	C	C	C
Sequencing and Experimental Details	sequencing	**sequencing**	MIXS:0000049	pcr_cond	pcr conditions	Description of reaction conditions and components of PCR in the form of 'initial denaturation:94degC_1.5min; annealing=...'	initial denaturation:degrees_minutes;annealing:degrees_minutes;elongation:degrees_minutes;final elongation:degrees_minutes;total cycles	initial denaturation:degrees_minutes;annealing:degrees_minutes;elongation:degrees_minutes;final elongation:degrees_minutes;total cycles	initial denaturation:94_3;annealing:50_1;elongation:72_1.5;final elongation:72_10;35		1	-	-	-	-	-	-	C	C	-	-	-
Sequencing and Experimental Details	sequencing	**sequencing**	MIXS:0000050	seq_meth	sequencing method	Sequencing machine used. Where possible the term should be taken from the OBI list of DNA sequencers (http://purl.obolibrary.org/obo/OBI_0400103).	Text or OBI	{termLabel} {[termID]}|{text}	454 Genome Sequencer FLX [OBI:0000702]		1	M	M	M	M	M	M	M	M	M	M	M
Bioinformatics and Data Processing	sequencing	**sequencing**	MIXS:0000051	seq_quality_check	sequence quality check	Indicate if the sequence has been called by automatic systems (none) or undergone a manual editing procedure (e.g. by inspecting the raw data or chromatograms). Applied only for sequences that are not submitted to SRA,ENA or DRA	none or manually edited	[none|manually edited]	none		1	-	-	-	-	-	-	C	C	-	-	-
Bioinformatics and Data Processing	sequencing	**sequencing**	MIXS:0000052	chimera_check	chimera check software	Tool(s) used for chimera checking, including version number and parameters, to discover and remove chimeric sequences. A chimeric sequence is comprised of two or more phylogenetically distinct parent sequences.	name and version of software, parameters used	{software};{version};{parameters}	uchime;v4.1;default parameters		1	-	-	-	-	-	-	C	C	-	-	-

TYPE OF INFORMATION			DESCRIPTOR DEFINITION AND SYNTAX									CHECKLIST REQUIREMENT (M:mandatory; C: conditional mandatory;X:optional; E:Environment dependent; (-): not apllicable)										
Type	Section	sub-section	MIXS ID	Structured comment name	Item (rdfs:label)	Definition	Expected value	Value syntax	Example	Preferred unit	Occurence	migs_eu	migs_ba	migs_pl	migs_vi	migs_org	mims	mimarks_s	mimarks_c	misag	mimag	miuvig
Bioinformatics and Data Processing	sequencing	**sequencing**	MIXS:0000053	tax_ident	taxonomic identity marker	The phylogenetic marker(s) used to assign an organism name to the SAG or MAG	enumeration	[16S rRNA gene|multi-marker approach|other]	other: rpoB gene		1	C	C	C	C	C	-	-	-	M	M	X
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000056	assembly_qual	assembly quality	The assembly quality category is based on sets of criteria outlined for each assembly quality category. For MISAG/MIMAG; Finished: Single, validated, contiguous sequence per replicon without gaps or ambiguities with a consensus error rate equivalent to Q50 or better. High Quality Draft:Multiple fragments where gaps span repetitive regions. Presence of the 23S, 16S and 5S rRNA genes and at least 18 tRNAs. Medium Quality Draft:Many fragments with little to no review of assembly other than reporting of standard assembly statistics. Low Quality Draft:Many fragments with little to no review of assembly other than reporting of standard assembly statistics. Assembly statistics include, but are not limited to total assembly size, number of contigs, contig N50/L50, and maximum contig length. For MIUVIG; 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	enumeration	[Finished genome|High-quality draft genome|Medium-quality draft genome|Low-quality draft genome|Genome fragment(s)]	High-quality draft genome		1	M	M	X	X	X	C	-	-	M	M	M
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000057	assembly_name	assembly name	Name/version of the assembly provided by the submitter that is used in the genome browsers and in the community	name and version of assembly	{text} {text}	HuRef, JCVI_ISG_i3_1.0		1	C	C	C	C	C	C	-	-	C	C	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000058	assembly_software	assembly software	Tool(s) used for assembly, including version number and parameters	name and version of software, parameters used	{software};{version};{parameters}	metaSPAdes;3.11.0;kmer set 21,33,55,77,99,121, default parameters otherwise		1	M	M	M	M	M	C	C	-	M	M	M
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000059	annot	annotation	Tool used for annotation, or for cases where annotation was provided by a community jamboree or model organism database rather than by a specific submitter	name of tool or pipeline used, or annotation source description	{text}	prokka		1	C	C	C	C	C	C	-	-	X	X	X
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000060	number_contig	number of contigs	Total number of contigs in the cleaned/submitted assembly that makes up a given genome, SAG, MAG, or UViG	value	{integer}	40		1	M	M	X	X	X	C	-	-	X	X	M
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000061	feat_pred	feature prediction	Method used to predict UViGs features such as ORFs, integration site, etc.	names and versions of software(s), parameters used	{software};{version};{parameters}	Prodigal;2.6.3; default parameters		1	X	X	X	X	X	X	-	-	X	X	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000062	ref_db	reference database(s)	List of database(s) used for ORF annotation, along with version number and reference to website or publication	names, versions, and references of databases	{database};{version};{reference}	pVOGs;5;http://dmk-brain.ecn.uiowa.edu/pVOGs/ Grazziotin et al. 2017 doi:10.1093/nar/gkw975		1	X	X	X	X	X	X	-	-	X	X	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000063	sim_search_meth	similarity search method	Tool used to compare ORFs with database, along with version and cutoffs used	names and versions of software(s), parameters used	{software};{version};{parameters}	HMMER3;3.1b2;hmmsearch, cutoff of 50 on score		1	X	X	X	X	X	X	-	-	X	X	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000064	tax_class	taxonomic classification	Method used for taxonomic classification, along with reference database used, classification rank, and thresholds used to classify new genomes	classification method, database name, and other parameters	{text}	vConTACT vContact2 (references from NCBI RefSeq v83, genus rank classification, default parameters)		1	X	X	X	X	X	X	-	-	X	X	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000065	16s_recover	16S recovered	Can a 16S gene be recovered from the submitted SAG or MAG?	boolean	{boolean}	yes		1	-	-	-	-	-	-	-	-	X	X	-
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000066	16s_recover_software	16S recovery software	Tools used for 16S rRNA gene extraction	names and versions of software(s), parameters used	{software};{version};{parameters}	rambl;v2;default parameters		1	-	-	-	-	-	-	-	-	X	X	-
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000067	trnas	number of standard tRNAs extracted	The total number of tRNAs identified from the SAG or MAG	value from 0-21	{integer}	18		1	-	-	-	-	-	-	-	-	X	X	X
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000068	trna_ext_software	tRNA extraction software	Tools used for tRNA identification	names and versions of software(s), parameters used	{software};{version};{parameters}	infernal;v2;default parameters		1	-	-	-	-	-	-	-	-	X	X	X
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000069	compl_score	completeness score	Completeness score is typically based on either the fraction of markers found as compared to a database or the percent of a genome found as compared to a closely related reference genome. High Quality Draft: >90%, Medium Quality Draft: >50%, and Low Quality Draft: < 50% should have the indicated completeness scores	quality;percent completeness	[high|med|low];{percentage}	med;60%		1	X	X	X	X	X	-	-	-	M	M	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000070	compl_software	completeness software	Tools used for completion estimate, i.e. checkm, anvi'o, busco	names and versions of software(s) used	{software};{version}	checkm		1	X	X	X	X	X	-	-	-	M	M	X
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000071	compl_appr	completeness approach	The approach used to determine the completeness of a given genomic assembly, which would typically make use of a set of conserved marker genes or a closely related reference genome. For UViG completeness, include reference genome or group used, and contig feature suggesting a complete genome	text	[marker gene|reference based|other]	other: UViG length compared to the average length of reference genomes from the P22virus genus (NCBI RefSeq v83)		1	-	-	-	-	-	-	-	-	X	X	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000072	contam_score	contamination score	The contamination score is based on the fraction of single-copy genes that are observed more than once in a query genome. The following scores are acceptable for; High Quality Draft: < 5%, Medium Quality Draft: < 10%, Low Quality Draft: < 10%. Contamination must be below 5% for a SAG or MAG to be deposited into any of the public databases	value	{float} percentage	1%		1	-	-	-	-	-	-	-	-	M	M	-
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000005	contam_screen_input	contamination screening input	The type of sequence data used as input	enumeration	[reads| contigs]	contigs		1	-	-	-	-	-	-	-	-	X	X	-
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000073	contam_screen_param	contamination screening parameters	Specific parameters used in the decontamination sofware, such as reference database, coverage, and kmers. Combinations of these parameters may also be used, i.e. kmer and coverage, or reference database and kmer	enumeration;value or name	[ref db|kmer|coverage|combination];{text|integer}	kmer		1	-	-	-	-	-	-	-	-	X	X	-
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000074	decontam_software	decontamination software	Tool(s) used in contamination screening	enumeration	[checkm/refinem|anvi'o|prodege|bbtools:decontaminate.sh|acdc|combination]	anvi'o		1	-	-	-	-	-	-	-	-	X	X	-

TYPE OF INFORMATION			DESCRIPTOR DEFINITION AND SYNTAX									CHECKLIST REQUIREMENT (M:mandatory; C: conditional mandatory;X:optional; E:Environment dependent; (-): not apllicable)										
Type	Section	sub-section	MIXS ID	Structured comment name	Item (rdfs:label)	Definition	Expected value	Value syntax	Example	Preferred unit	Occurence	migs_eu	migs_ba	migs_pl	migs_vi	migs_org	mims	mimarks_s	mimarks_c	misag	mimag	miuvig
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000075	sort_tech	sorting technology	Method used to sort/isolate cells or particles of interest	enumeration	[flow cytometric cell sorting|microfluidics|lazer-tweezing|optical manipulation|micromanipulation|other]	optical manipulation		1	-	-	-	-	-	-	-	-	M	-	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000076	single_cell_lysis_appr	single cell or viral particle lysis approach	Method used to free DNA from interior of the cell(s) or particle(s)	enumeration	[chemical|enzymatic|physical|combination]	enzymatic		1	-	-	-	-	-	-	-	-	M	-	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000054	single_cell_lysis_prot	single cell or viral particle lysis kit protocol	Name of the kit or standard protocol used for cell(s) or particle(s) lysis	kit, protocol name	{text}	ambion single cell lysis kit		1	-	-	-	-	-	-	-	-	X	-	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000055	wga_amp_appr	WGA amplification approach	Method used to amplify genomic DNA in preparation for sequencing	enumeration	[pcr based|mda based]	mda based		1	-	-	-	-	-	-	-	-	M	-	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000006	wga_amp_kit	WGA amplification kit	Kit used to amplify genomic DNA in preparation for sequencing	kit name	{text}	qiagen repli-g		1	-	-	-	-	-	-	-	-	X	-	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000077	bin_param	binning parameters	The parameters that have been applied during the extraction of genomes from metagenomic datasets	enumeration	[homology search|kmer|coverage|codon usage|combination]	coverage and kmer		1	-	-	-	-	-	-	-	-	-	M	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000078	bin_software	binning software	Tool(s) used for the extraction of genomes from metagenomic datasets, where possible include a product ID (PID) of the tool(s) used.	names and versions of software(s) used	{software};{version}{PID}	MetaCluster-TA (RRID:SCR_004599), MaxBin (biotools:maxbin)		1	-	-	-	-	-	-	-	-	-	M	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000079	reassembly_bin	reassembly post binning	Has an assembly been performed on a genome bin extracted from a metagenomic assembly?	boolean	{boolean}	no		1	-	-	-	-	-	-	-	-	-	X	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000080	mag_cov_software	MAG coverage software	Tool(s) used to determine the genome coverage if coverage is used as a binning parameter in the extraction of genomes from metagenomic datasets	enumeration	[bwa|bbmap|bowtie|other]	bbmap		1	-	-	-	-	-	-	-	-	-	X	X
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000081	vir_ident_software	viral identification software	Tool(s) used for the identification of UViG as a viral genome, software or protocol name including version number, parameters, and cutoffs used	software name, version and relevant parameters	{software};{version};{parameters}	VirSorter; 1.0.4; Virome database, category 2		1	-	-	-	-	-	-	-	-	-	-	M
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000082	pred_genome_type	predicted genome type	Type of genome predicted for the UViG	enumeration	[DNA|dsDNA|ssDNA|RNA|dsRNA|ssRNA|ssRNA (+)|ssRNA (-)|mixed|uncharacterized]	dsDNA		1	-	-	-	-	-	-	-	-	-	-	M
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000083	pred_genome_struc	predicted genome structure	Expected structure of the viral genome	enumeration	[segmented|non-segmented|undetermined]	non-segmented		1	-	-	-	-	-	-	-	-	-	-	M
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000084	detec_type	detection type	Type of UViG detection	enumeration	[independent sequence (UViG)|provirus (UpViG)]	independent sequence (UViG)		1	-	-	-	-	-	-	-	-	-	-	M
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000085	otu_class_appr	OTU classification approach	Cutoffs and approach used when clustering “species-level” OTUs. Note that results from standard 95% ANI / 85% AF clustering should be provided alongside OTUS defined from another set of thresholds, even if the latter are the ones primarily used during the analysis	cutoffs and method used	{ANI cutoff};{AF cutoff};{clustering method}	95% ANI;85% AF; greedy incremental clustering		1	-	-	-	-	-	-	-	-	-	-	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000086	otu_seq_comp_appr	OTU sequence comparison approach	Tool and thresholds used to compare sequences when computing "species-level" OTUs	software name, version and relevant parameters	{software};{version};{parameters}	blastn;2.6.0+;e-value cutoff: 0.001		1	-	-	-	-	-	-	-	-	-	-	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000087	otu_db	OTU database	Reference database (i.e. sequences not generated as part of the current study) used to cluster new genomes in "species-level" OTUs, if any	database and version	{database};{version}	NCBI Viral RefSeq;83		1	-	-	-	-	-	-	-	-	-	-	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000088	host_pred_appr	host prediction approach	Tool or approach used for host prediction	enumeration	[provirus|host sequence similarity|CRISPR spacer match|kmer similarity|co-occurrence|combination|other]	CRISPR spacer match		1	-	-	-	-	-	-	-	-	-	-	C
Bioinformatics and Data Processing	sequencing	**data processing**	MIXS:0000089	host_pred_est_acc	host prediction estimated accuracy	For each tool or approach used for host prediction, estimated false discovery rates should be included, either computed de novo or from the literature	false discovery rate	{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)		1	-	-	-	-	-	-	-	-	-	-	C
Accession and Submission	sequencing	**Accession and Submission**	MIXS:0000091	associated resource	relevant electronic resources	A related resource that is referenced, cited, or otherwise associated to the sequence.	reference to resource	{PMID} | {DOI} | {URL}	http://www.earthmicrobiome.org/		m	C	C	C	C	C	C	C	C	C	C	C
Accession and Submission	sequencing	**Accession and Submission**	MIXS:0000090	sop	relevant standard operating procedures	Standard operating procedures used in assembly and/or annotation of genomes, metagenomes or environmental sequences	reference to SOP	{PMID}|{DOI}|{URL}	http://press.igsb.anl.gov/earthmicrobiome/protocols-and-standards/its/		m	C	C	C	C	C	C	C	C	C	C	C