

Structuring Strain Data for Storage and Retrieval of Information on Bacteria in MINE, the Microbial Information Network Europe*

J. A. STALPERS¹, M. KRACHT², D. JANSSENS³, J. DE LEY³, J. VAN DER TOORN⁴, J. SMITH⁵, D. CLAUS², and H. HIPPE²

¹ Centraalbureau voor Schimmelcultures, 3740 AG Baarn, The Netherlands

² DSM – Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, 3300 Braunschweig, Federal Republic of Germany

³ Laboratorium voor Microbiologie en Microbiële Genetica, 9000 Gent, Belgium

⁴ Laboratorium voor Microbiologie, Technische Universiteit, 2628 BC Delft, The Netherlands

⁵ University of Amsterdam, Department for Medical Informatics, Academic Medical Centre, 1105 AZ Amsterdam, The Netherlands

Received October 14, 1989

Summary

A number of major European culture collections, supported by the "Biotechnology Action Programme" (BAP) of the European Community (EC) are involved in the establishment of the multinational "Microbial Information Network Europe" (MINE). It has been realized that the adoption of a uniform format for computer storage and retrieval of strain data is essential, considering that it is the scope of the project to allow the electronic combination of data from different collections, and thus facilitate the publication of integrated catalogues or searching different databases using common commands.

The format for bacteria is different to that previously published for fungi and yeasts (Gams et al., 1988) in several important respects, largely relating to the differences in cytological properties and taxonomic treatment. In addition to the main data set, "species records", two other data sets, "strain records" and "synonym records" may be linked together. A minimum data set of 24 fields (similar to the fields used in catalogue production) is defined which facilitates the exchange of data between the national nodes and serves as a directory to strains available at other nodes. A total of 133 fields are proposed that can be grouped in the following ten blocks: internal administration – name – strain administration – environment and history – properties (cytological, biomolecular data) – mutants, plasmids, and phages – growth conditions – chemistry and enzymes – biological interactions – practical applications.

Several fields are to be divided into subfields at different ranks. Delimiters are used to either separate a range of entries that are to be indexed or to divide an entry from the reference to its source or remarks that are not to be indexed. In this paper the contents and structure of the fields proposed for bacteria are characterized and often illustrated by examples. Uniformity of input is essential for indexed fields and desirable for non-indexed fields. Five thesaurus files are envisaged to ensure consistency.

Key words: Bioinformatics – Bacteria – Format

Introduction

During the last few decades a wealth of information on a large number of microbial strains has been gathered, covering taxonomy, ecology, biochemistry, physiology, molecular biology and genetics. In some cases it has been possible to exploit this information for biotechnological purposes. However, the results of research work are

widely dispersed, being found in the literature, in culture collections, and laboratories in the private and public sector.

While some progress has been made in the computerizing of data, so that local access is faster, there are problems associated with the transfer of data between two sites, since no common database system is in use. Consequently it was decided that there was a need to fill this gap,

* MINE Technical Communication No. III

with emphasis on the needs of culture collections and the European biotechnological industry.

Studies funded by the Commission of the European Communities (CEC) on a European information system on microorganisms by Excerpta Medica (1983) and Environmental Resources Ltd. (1984) resulted in the establishment, in 1985 of the Microbial Information Network Europe (MINE). This project is funded by the CEC, within the Biotechnology Action Programme (BAP), until the end of 1989. Participants are culture collections from Belgium, France, Greece, Italy, The Netherlands, Portugal, Spain, The United Kingdom, and West-Germany. Cooperation with the international referral system MSDN (Microbial Strain Data Network) is envisaged.

The MINE participants agreed to the basic design of the databases including records for species (SP) and strain (STR). The whole data base comprises more than one hundred fields which can be filled with information on species only, strains only, or on species and strains. A special field is used for linking the strain record to the appropriate species record; so the taxonomic name has to be entered only once for strains of the same species. A synonym record (SYN) provides linkage to the corresponding SP record in the case of organisms having two or more synonymous names. The general characteristics of the MINE databases are published in the first Technical Communication of MINE (Gams et al., 1988), which deals with the formats for fungi and yeasts.

In the case of bacteria the data format was supplemented by some fields while other fields were redundant. Some fields relating to cytological properties were added to the full data format, like GRAM (Gram staining), MOT (Motility), PIGM (Pigments), and RESTST (Resting structures).

A total of 133 fields have been proposed; 126 of these form the strain record, 83 the species record; 50 fields are indexed. The Minimum Data Set is a subset of the full format containing the basic data of the collection; this means particularly the information included in the current catalogues of strains. At least this subset will be exchanged between all national MINE nodes, so the user will be able to ascertain in which collections a certain strain is held, or where further information can be obtained. The Minimum Data Set also had to be changed, to contain 24 fields (30 fields have been defined for fungi). The fields PVAR (Pathovar), SVAR (Serovar), GENOT (Genotype, MUTMET (Method of mutation, HYBR (Hybrids), PLS (Plasmids), PHAGE, RESTR (Restrictions and precautions), CONDL (Conditions for growth on liquid media) were added.

Subfields and subsubfields have been defined for the new fields analogous to the previously defined ones. Delimiters are used to separate a range of equivalent entries, e.g. ";" for separation of synonymous collection numbers. Delimiters make it also possible to index parts of a field only and/or to use them for catalogue printing. The "<" for example excludes remarks, literature references or references to other sources from indexing.

To facilitate entering of data (for instance translation of abbreviations to full names) and to check the entered data

for correctness (e.g. spelling) several thesauri are considered. Validation of data is very important for indexed fields, but also useful for non-indexed fields. Thesauri for geographical names, enzymes, and collection acronyms are ready (for bacteria some additional data have to be added) or planned (chemical substances, culture media).

The Proposed Fields

In the following list all fields are grouped in ten blocks. These blocks are so constructed that fields of SP records appear next to related fields of STR records. Each field is characterized by the field tag (label, prefix) and a short circumscription. They may be followed by a **O** indicating that the field must contain data for uptake in the database, by a **A** indicating automatic generation of information, or by a **I** indicating the field is indexed. In several fields references to the origin of data may be entered after a "<". There the symbol "I" means work done at the same institute, "D" work done by the depositor, otherwise literature references may be given.

Fields included in the minimum data set are marked by an asterisk (*).

A. Internal Administration

ACCN	Accession number of the record: O I A Occurs automatically in all records. It is a unique number.
APOS	Alternate posting: O I Occurs obligately in all STR and SYN records and contains the accession number (ACCN) of connected SP record.
RT	Record type: O I Occurs in all records. Contains an abbreviation indicating type of record. There are only 3 possibilities: SP = species; STR = strain; SYN = synonym
OT	Organism type: O A Must occur in species records. Contains abbreviation indicating type of databank. B = bacteria; F = fungi; P = plasmids; Y = yeasts
EDI	Date of input: I A Occurs automatically in all records. Contains input date. Format DDMMYY
EDM	Date of latest update: I A Occurs in all records after an update. Contains date of last update. Format DDMMYY
ID	Typist identification: I A Occurs automatically in all records. Name of user/authority who inserted the record or who was the last one to change it.
SEC	Security code May occur in any record. Different codes may prevent records from inclusion in the catalogue or from external access. No entry means open access, otherwise the following

OTHNAM	<p>Example: [in <i>Salmonella</i> sp. LMD 73.6] Serovar dublin (1,9,12:g,p:-) Other names: I *</p> <p>Only in STR. The field contains:</p> <p>a. names that are not synonymous with the adopted name, but have been used for it in the literature.</p> <p>b. names based on taxonomic rearrangements which cannot be dealt with by a SYN record.</p>	<p>STAT either refers to the name mentioned in the NAME fields, and then only one of the above abbreviations is given, or it refers to another taxon, then the name of that taxon with author(s) (no comma!) is given, preceded by "of". The whole field is indexed but the string "T of" is excluded from the index.</p>
<p>Since 1980 many names are no longer regarded as synonyms of valid names (objective or subjective synonyms). The use of such a name is contrary to the rules. However, information on strains may be searched under such names and many references in the literature are tied to such names. The format is in concordance with SP. Author(s), year and approbation should be added.</p> <p>For valid new names that are obviously misclassified the SYN record should be used (for example: a new name for <i>Marinomonas vaga</i> is <i>Oceanospirillum vagum</i>; however the strains do not belong to <i>Oceanospirillum</i> and the collection will therefore continue listing them as <i>Marinomonas</i>).</p>	<p>Examples: [in <i>Acetobacter pasteurianus</i> LMD 22.1] T [in <i>Acetobacter pasteurianus</i> LMD 51.1] T of <i>Acetobacter pasteurianus</i> ssp. <i>ascendens</i> (Henneberg 1898) De Ley and Frateur 1974 AL [in <i>Pseudomonas syringae</i> pv. <i>syringae</i> LMG 1247] PVRS; T of <i>Pseudomonas syringae</i>, van Hal 1902 [in <i>Acetobacter pasteurianus</i> LMD 31.3] T of <i>Acetobacter aceti</i> ssp. <i>orleanensis</i>, (Henneberg 1906) De Ley and Frateur 1974 AL</p>	<p>TAX Taxonomy Occurs only in SP. Contains names of suprageneric taxa separated by commas. May be used for the production of special catalogues. (To be superseded by Thesaurus)</p> <p>TAXREM Taxonomic remarks In SP, STR and SYN. May harbour all kind of taxonomic remarks, for example synonyms if the relation is equivocal. For abbreviations concerning the status see field STAT. Example: [in <i>Clostridium beijerinckii</i> LMD 25.101] OR of <i>Clostridium beijerinckii</i> Donker 1926 AL (This strain is the original monotype strain of Donker but not mentioned in the AL as the type strain.)</p>
STAT	<p>Status: I *</p> <p>Occurs in SP and STR. T = type strain (in STR) or type species (in SP); NT = neotype; PVRS = pathovar reference strain.</p> <p>Since 1980 only type strains and neotype strains have status in nomenclature. Use of the terms pathotype (PTT), lectotype (LT), paratype (PT) and authentic strain (AUT) is not recommended. OR (original strain) and REFS (reference strain for infrasubspecific taxa) also have no special status; if these terms are to be used, they should be entered in field TAXREM except for pathovar reference strain (PVRS = PTT) for practical reasons since pathovar is treated here at the same level as subspecies.</p>	<p>TAXLIT Taxonomic literature In SP, STR and SYN. Reference(s) to taxonomic literature. Subfields are separated by a semicolon. Example: Wiegand, J. D. et al. Int. J. System. Bact. (1978) 28: 573-581.</p> <p>ILL Illustration In STR and SP. Reference number of future graphic database in which figures are stored or to illustrations in literature from bibliographic data base.</p> <p>REM General remarks In SP, STR and SYN.</p> <p>LIT General references In SP and STR.</p>

C. Strain Administration

STN	Strain number in collection: O I * Must occur in all STR records. Number only. (In affiliated collections also acronym). Example: LMD 22.1		
OCC	Other collections: I * Occurs only in STR. Contains collection acronym + number applied to the same strain in other collections. (Abbreviation according to <i>McGowan</i> and <i>Skerman</i> , 1982; <i>Pridham</i> , 1981). Useful personal numbers or other strain designations can be entered here too. Strain designations are separated by a semicolon. After "<" strain designations will not be indexed and will be excluded from printed catalogues. Collection acronyms contain no full stops, initials of persons are separated by a full stop. Example: [in <i>Acetobacter aceti</i> LMD 23.1] LMG 1504; DSM 3508; CECT 298; PDDCC 8807; ATCC 15973; NCIB 8621; CCUG 18122; Visser't Hooft X<Delft L 40; Frateur 4 (I)	PRE	Preservation In STR. The date is not entered in this database. AG = agar slants (interval of transfers and storage temperature in parentheses); LI = liquid media (interval of transfers in parentheses); LY = lyophilized; LN = liquid nitrogen (the cryoprotectant in parentheses); FR = frozen, followed by figure to indicate temperature and the cryoprotectant between parentheses; LD = L-drying; MO = mineral oil; SG = silica gel; CY = cysts; ES = endospores Examples: AG (6 mo), LY, LN (glycerol 10%) ES, FR80 (DMSO 10%)
EDR	Date of receipt: I Occurs in STR. Contains date of receipt. Format is DD MMM YYYY. Example: 12 Feb 1982	PREREM	Remarks on preservation In STR and SP (when results of other workers with other strains are reported). Special care or results in preservation and remarks on the general ability to survive after preservation. Examples: LY- (does not survive freeze-drying) 4C- (does not endure storage at 4C), requires DMSO for LN.
EDA	Date of accession: I Occurs in STR. Contains date of accession. Format is DD MMM YYYY (or MMM YY when collections only register month and year). Example: Mar 1982	PRELIT	Literature on preservation In STR and SP
CORR	Correspondence Occurs in STR records. Refers to numbers of letters in archive(s). Example: letter 415/30-11-1985 letter 322/17 Mar 1987.	RESTR	Precautions and restrictions: * In SP and rarely in STR (genetically manipulated strains). 1. Precautions for internal policy and distribution of cultures. Example: transfer only in biohazard safety cabinet 2. Concerning mainly export of cultures. a. Pathogenicity class (WHO standard and/or national standard) or recombinant category and handling precautions b. distribution and transportation restrictions and precautions c. import regulations (restrictions imposed by laws of the country in which the collection is located) d. export regulations (restrictions imposed by laws of other countries, e.g. permits, packing conditions)
SPINF	Special information: A * In STR, indicating special information available. The acronym(s) of the further fields are introduced automatically during VALIDATE, if they contain information (in principle all fields, but not the LIT and REM fields, nor those of the MDS or those concerning internal administration).	SUPPLY	Form of supply Only in STR. Specifies how strains are sent to the customer. AC = active culture; DR = freeze-dried or L-dried; DI = dried ice
CHK	Collection check Only in STR. Initials, date; collections might also enter kind and result of check or other remarks. ID = check for identity (confirmation of the name under which the strain is actually listed); VIAB = check for viability (control for good recuperation, date of next check can be entered too); PUR = check for purity (only to be performed once for each new batch). Result: default is '+' (need not to be entered), in case of VIAB	ADMREM	Administrative remarks In STR

besides '+' and '-' also 'o' (= reduced viability) can be entered.

Examples: JT, 5 Jun 1972; recheck JT, 18 Dec 1986.

HH, 10 Nov 1985, VIAB o

- D. Environment and History*
- HABI** Specific habitat: I
Only in SP. Indicates restrictions to substrates, when consistent for a particular species.
Example: [in *Azotobacter paspali*] rhizosphere of *Paspalum notatum*
- ISOFR** Isolated from: I
Only in STR. Indicates substrate or source of isolation. Indications of biotope of a particular substrate may be entered here, but any local names in field LOC. Main information first, to be separated by semicolon “;” from less important information; the latter is not indexed.
Examples: [In *Beijerinckia indica* ssp. *indica* LMD 50.24] Rice field soil; surface layer
- LOC** Geographic location: I *
Only in STR. Geographic location of source. Sequence: country, state; local details. Data after a semicolon not to be indexed. Spelling of the country name is in English, other names according to the national spelling (thus “Hannover, München, Warszawa, Moskva”), cyrillic names are transliterated according to ISO Norm 833–1974. A thesaurus has been built for geographical names.
Examples: Canada, Quebec; Montreal, La Salle Woods, 400 m alt
- COLL** Collector of sample
Only in STR. Gives name of the collector, original number of specimen, date of collecting. Information separated with commas. Collector’s initials separated with full stops, no spaces.
Example: H. S. Sinclair, H.S.S. 1783, 12 Oct 1983.
- ISOL** Isolated by: *
Only in STR. Name of isolator, original designation (usually a number) of isolate, date. Microbiologist’s initials separated by full stops, no spaces. Designations are preceded by collectors initials or “No”.
Example: [In *Comamonas acidovorans* LMD 26.37] L.E. den Dooren de Jong, No. 7, Mar 1925.
- ISOM** Isolation method
Only in STR. Technique, medium (in parentheses). DP = dilution plate; ES = endospores; CC = continuous culture; CY = cysts; DI = direct plating; EM = enrichment method; PAST = pasteurisation
Examples: EM (67 + mannitol) PAST, DP, ES (112)
- DET** Identified by
Only in STR. Name(s) of identifier, date and name of identified organism (if deviating from current name) between brackets. After “<” a method may be specified.
Example: A. Fuchs (1957), redet. J.v.d. Toorn (1985) < API
- DEP** Depositor: I *
Only in STR. Individual name or institution (abbreviated, without punctuation marks between the letters). Two depositors may be cited, separated with “and”. A date is not given here but in field EDA and EDR. Microbiologist’s initials need not be separated from the name by a space, they are not indexed. After “<” the depositors strain number and/or institute may be entered.
Example: [in *Acetobacter aceti* LMG 1504] LMD < 23.1
- HIS** History: *
Only in STR. Fate of the isolate between isolation and deposit in the present collection. The backward sequence of deposits is used, separated by “<” meaning “received from”. Each entry thus may contain: name of collection, (month and) year of acquisition. Between parentheses can be entered: strain designation(s) or collection numbers (only when confusion is possible between two or more numbers from the same collection) and/or a name when a name change has occurred).
Example: [in *Bacillus sphaericus* DSM 488] NCTC, Nov 1973 (*Bacillus loehnisii*) < T. Gibson, 1935 < Kral Collection (*Bacillus probatus*)
- NAMCH** Name changes: *
In (SP or) STR. Name changes introduced in the present collection, date.
Example: [In *Klebsiella planticola* LMD 75.73] until 1984 preserved as *Klebsiella pneumoniae*
- ENHISLIT** Literature on environment and history
Mainly in STR but also in SP (environmental data)
- ENHISREM** Remarks on environment and history
Mainly in STR but also in SP (environmental data)
- E. Properties (morphological, cultural and chemotaxonomical data)*
- FSTR** Fine structure data
In SP and STR. Data on cell wall, flagella or organelles.
- WCONS** Cell wall constituents: I
In Sp. and STR. Name of component, percentage, (in parentheses technique used).
Example: peptidoglycan: A3a L-Lys-L-Ser-L-Ala₂; teichoic acid; glycerol

CINC	Cell inclusions In SP and STR. Name of inclusion (also gas vacuoles) Example: glycogen; PHB; PP (PHB = poly- β -hydroxybutyrate; PP = polyphosphate)	PIGM	Pigments In SP and STR. Colour of pigment, name (if available), (non)diffusing, fluorescence (wavelength), medium (number). Various pigments separated by semicolons. Examples: green, pyoverdin, dif, fl (254 nm) +, King B Yellow-green, bioverdin, dif, fl (254 nm) -, fl (365) nm +, King B Yellow, nondif, fl (254 nm) -, King A
CHTAX	Chemotaxonomical markers: I In STR. Name of marker (e.g. fatty acids, phospholipids, polyamines, quinones). Example: Q9 (co-enzyme Q with 9 isoprene units)	GRWTHL	Growth characteristics in liquid media In SP and STR. Typical growth characteristics in liquid media, e.g. adhesion, flocculation, cell lysis.
GRAM	Gram behaviour In SP and STR. Contains Gram type, Gram stain, result (+, -, v(ar)), method. Remarks can be added after "<".	GRWTHS	Growth characteristics on solid media In SP and STR. Typical growth characteristics on solid media, e.g. shape circular.
KOH	KOH test In SP and STR. Contains result (+, -, v(ar)) of reaction with KOH. Remarks can be added after "<".	QUADGER	Quantitative data on growth In STR. Quantitative growth characteristics like doubling time, growth rate and yield. Information on growth conditions (in general medium, temperature) should be indicated. GR = growth rate (μ), DR = division rate (v), GT = generation time, DT = doubling time, Y = growth yield (g cells/g substrate), Y/S = g cells/mol substrate, Y/ATP = g cells/mol ATP. Example: GR = 0.12, medium 1,30 C
AMPEP	Aminopeptidase test In SP and STR. Contains result of reaction with L-alanin-4-nitro-anilide hydrochloride. Remarks can be added after "<".	RESTST	Resting structures In SP and STR. Name of resting structure, medium, temperature, light and shaking). C = cyst, ENS = endospore, EXS = exospore, MS = myxospore, F = fruiting body. Remarks or references to be entered after "<". Example: ENS, 116, 24 C < oval, after 6 days
STAIN	Staining reactions In SP and STR. Contains method(s) and results. Various staining methods are separated by a semicolon. Remarks can be added after "<". Example: acid fast +;	GC	Guanine-cytosine content of DNA Mainly in STR. Percentage of GC in DNA (omit "%"). After "<" the method or authority can be entered. Example: [in <i>Clostridium butyricum</i> DSM 552] 28 < Cummins and Johnson 1971
CELL	Shape, arrangement and size of cell: I In STR. Three subfields for respectively shape, arrangement and size, separated by a semicolon. The first is indexed and contains data on shape. Abbreviations: C = coccus; CF = coryneform; R = rod; SP = spirillum; V = vibrio. The second subfield contains data on arrangement. Abbreviations: m = monocells; d = diplocells (be careful, as diplocells are always present in growing cultures); neis = diplocells with adjacent sides flattened (as in <i>Neisseria</i>); t = tetrads (2-dimensional); s = sarcinae (3-dimensional); str = streptocells (chains, > 4 cells); pal = palisades; myc = mycelium; fil = filamentous forms; vy = V and Y forms; irr = irregular; pl = pleomorphic. The third subfield gives data on size ranges. After "<" remarks on conditions (medium) can be added. Example: R; m; (2-2.4 \times 5-5.5) μ m < medium 122	DNAD	Hybridization DNA-DNA with other strains Only in STR. Strain number, percentage binding; different data separated by ";" (omit "%"). Example: [in <i>Pseudomonas cepacia</i> LMG 1222] LMG 1200, 15 < De Vos and De Ley 1983; LMG 1275, 63 < De Vos and De Ley 1983
MOT	Motility In SP and STR. Two subfields: motile (+ or -) and mechanism of motility (for example: flagella, gliding, twitching). In case of flagella, kind of flagellation and number of flagella can be added after a colon. Subfields are separated by a semicolon. Example: +; flagella: bipolar 3-5	DNAR	Hybridization DNA-r.RNA with other strains Only in STR. Strain number, type of RNA, temperature of denaturation [Tm(e) in $^{\circ}$ C]; different data in subfields separated by ";".

	Example: [in <i>Pseudomonas cepacia</i> LMG 1275] LMG 1794, 23sRNA, 64.5 < De Vos and De Ley 1983; LMG 1226, 23sRNA, 71.0 < De Vos and De Ley 1983	MUTREM	Remarks on genetics, mutations, DNA data Only in STR. Remarks on mutations and genetics.
		MUTLIT	Literature on genetics and mutants Only in STR. References to literature containing data on genetics and mutants.
RNAD	Hybridization r.RNA-DNA with other strains Only in STR. Type of RNA, strain number, temperature of denaturation [T _m (e) in °C]; different data in subfields separated by “;”. Note: in field DNAR the RNA comes from another strain; in field RNAD the RNA comes from the actual strain. Example: [in <i>Pseudomonas solanacearum</i> LMG 2299] 23sRNA, LMG 1242, 61.5 < De Vos and De Ley 1983; 23sRNA, LMG 2155, 75.5 < De Vos and De Ley 1983	GENOT	Genotype, chromosomal marker: I * Only in STR. Example: [In <i>E. coli</i> PC 0043] thyra mal (lam)phx
		SEXST	Sexual state: * Only in STR. Indicates the sexual condition of the strain. Examples: (in <i>E. coli</i> PC 1299) F- [in <i>Pseudomonas aeruginosa</i> PC 1412] FP+
		HYBR	Hybrid: * Only in STR. In “SP” records the hybrid nature appears in field SP (with an “x” between the strain numbers and/or names of the parents) according to the International Code of Bacteriological Nomenclature. After “<” author, date and methodology may be entered. Example: PC 0428 × PC 0856
DNASQ	DNA sequences Only in STR. Contains references to Heidelberg database.	PLS	Plasmid: I * Only in STR. Contains plasmid name, recombinant or natural plasmid, cryptic or non-cryptic and accession number in plasmid database. RP = recombinant plasmid, NP = natural (original) plasmid, C+ = cryptic, C- = non-cryptic. Information on various plasmids is separated by a semicolon. If plasmid is unnamed enter ‘unnamed’.
RNASQ	RNA sequences Only in STR. Contains references to Heidelberg database.		
PHYLO	Other phylogenetic data In STR. Contains for instance SAB or DAB values of strain A (the actual strain) with respect to strain B (for which the strain number has to be provided here).		
TKIT	Result of testkit, e.g. API test: I In STR. Strip indication / colon (:) / code. Several tests delimited with semicolon (;). Conditions for non-commercial kits have to be described (in general by a reference number). Example: API 20NE: 1.004.447 API 20E: 5.144.552-57; API 20EC: 5.444.402	PHAGE	Phage: I Only in STR. Contains phage name, recombinant or natural phage and accession number in phage database. RP = recombinant phage, NP = natural (original) phage. Information on various phages is separated by a semicolon.
PROPLIT	Literature on properties In SP and STR. Literature on morphological, cultural and chemotaxonomical data.	PLPREM	Remarks on plasmids and phages In SP and STR.
PROPREM	Remarks on properties In SP and STR. Remarks on morphological, cultural and chemotaxonomical data.	PLPLIT	Literature on plasmids and phages In SP and STR. In case of various plasmids and/or phages insert name of plasmid/phage between brackets.
<i>F. Mutants, Plasmids and Phages</i>			
MUT	Mutant: * Only in STR. The word “mutant” or equivalents (e.g. mutated characteristics) are to be entered. Indication of number of the parent strain may also be given. Example: PQQ-mutant of LMD 79.41		
MUTMET	Method of mutation: * Only in STR. After “<” the author of the mutation and the date may be given. Examples: UV 30 min 10% isothiocyanate 10 min.		
<i>G. Growth Conditions</i>			
		CONDS	Conditions for growth on solid media: * Mainly in SP but also in STR. Numbers of media (+ additives), temperature, light, gas atmosphere.
		CONDL	Conditions for growth on liquid media: * In SP (and STR). Numbers of media (+ additives), temperature, light, gas atmosphere, shaking.

- CONDGER** Conditions for germination
In SP (and STR). Numbers of media (+ additives), temperature, light, gas atmosphere, shaking.
- NUGR** Nutritional requirements and growth factors: I
In SP and STR. Amino acids, trace elements, growth factors (vitamins), blood, serum. Abbreviations for input to be organized through thesaurus for chemicals.
- TOL** Tolerances and sensitivities: I
In SP and STR. Agent: method, concentration (give unit!), result (S = sensitive, R = resistant, M = mesic). The colon (:) after the agent has the effect that only the first part of the field is indexed. Subfields containing agents, methods, concentrations and results are separated with ";". Use chemical symbols, not commercial names (cf. thesaurus!). In case of non-standard methods give reference. For disc method, give inhibitory zone (in mm) before indicating S, R or M.
Examples: penicillin: 5 units/ml S
streptomycin: 15 µg/ml R; crystalviolet: 0.0001% S (or 1 ppm or 1 µg/ml)
phage Tula; phage T1
- DEF** Deficiencies: I
In SP and STR. Name of compound (colon) concentration. To be used for naturally occurring or artificially induced mutational deficiencies, other than the regular growth factor requirements in field NUGR. Characteristics of artificially induced mutants to be entered in field MUT. Subfields containing compounds and concentrations are separated by ";". Abbreviations to be organized through a thesaurus.
- METY** Metabolic type: I
In SP and STR. Basic metabolic type. DT = diazotrophic, MT = methylotrophic, DG = diazogenic, MG = methanogenic, AC = acetogenic, SR = sulphate reducing, NR = nitrate reducing.
- CNUT** Carbon nutrition type
In SP and STR. A = autotrophic (CO₂, CO); H = heterotrophic; M = microtrophic.
- CSOR** Single carbon sources tested for growth: I
Only in STR. Name of carbon source and result of test. No abbreviations in the database. For input abbreviations (e.g. as adopted from API systems) may be used, when organized in a thesaurus. Results are indicated with +, -, or v. Subfields are separated by a semicolon. If the test has been performed under anaerobic conditions, "(AN)" can be added after the result.
Example: glucose +; galactose +; trehalose -
- NSOR** Single nitrogen sources tested for growth: I
Only in STR. Name of nitrogen source and result of test. Data include also hydrolysis of urea. No abbreviations in the database. Abbreviations for input to be organized through thesaurus. Subfields separated by a semicolon. If the test has been performed under anaerobic conditions, "(AN)" can be added after the result.
Examples: ammonium +; nitrate + nitrite +; ethylamine +; cadaverine -
- CNSOR** Single compounds tested as C and N source for growth: I
In SP and STR. Name of compound and result of test. No abbreviations in the database. Abbreviations for input to be organized through thesaurus. Subfields separated by a semicolon. If the test has been performed under anaerobic conditions, "(AN)" can be added after the result.
Examples: dimethylamine +; diethylamine -; methylurea +
- OXR** Oxygen relationships
In SP and STR. AE = aerobic (< = 100% air); MA = microaerophilic (5% air); FAN = facultative anaerobic; AT = aerotolerant (prefers anaerobic conditions for good growth, but tolerates aerobic conditions); MAT = microaerotolerant (anaerobic bacteria tolerating microaerophilic conditions); AN = anaerobic. (Definitions according to *H. G. Schlegel*, *Allgemeine Mikrobiologie*, 5th ed. 1981).
- GAS** Gas phase
In SP and STR. Percentage of gases. Air is default. N₂ = nitrogen; CO₂ = carbon dioxide; H₂ = hydrogen; O₂ = oxygen; ME = methane.
- TEMPR** Temperature relationships: I
Only in SP. PT = psychrotroph (T opt. < 20 C, slow growth at T > 20 C); PS = psychrophilic (T opt. < 20 C, no growth at T > 20 C); ME = mesophilic (T opt. between 20-40 C, no growth at 50 C); TT = thermotolerant (T opt. between 20-40 C, growth at 50 C); TH = thermophilic (T opt. between 40-65 C); ET = extremely thermophilic (T op. > 65 C). (Definitions according to *H.G.Schlegel*, *Allgemeine Mikrobiologie*, ed. 5, 1981).
- TEGR** Cardinal temperatures for growth
Only in STR, single value or range (no "degree C") in three subfields for minimum, optimum and maximum temperature (in that order, subfields separated by ";". Recommended temperature for preservation in field CONDS. If a subfield contains no data, care must be taken that the semicolon which terminates this empty subfield is entered. If

	growth has been tested at a particular temperature without establishing the range, the temperature is indicated with the growth response (+, -, v). Examples: 12.5; 20-24; 30 ;47		
HEATR	Heat resistance In SP and STR. Temperature and time (separated by "/" to survive with "+", to kill with "-"; after "<" the survival structure and cultivation condition may be entered. A "+" in the SP record indicates heat resistance in general. Example: 70/15 min + < endospores	SUGC	Sugar concentration In STR. Names of sugars and optimum and maximum concentration (in that order) in % (without "%", but with indication w/w, v/v, w/v). The subfields containing the name of a sugar and the values of the concentrations are separated by ";". Subsubfields containing the values are separated by ",". If a subsubfield contains no data, care must be taken that the comma which terminates this empty subfield is entered. If growth has been tested for a particular sugar at a particular concentration without establishing the range, name and concentration are indicated with the growth response (+, -, v). Example: NaCl 2,5-8.5, 14 w/v;
PHR	pH relationships: I In SP and STR. Initial pH for a good growth. In batch cultures the pH may change (eg. by acid production). EA = extremely acidophilic, pH < 4; AC = acidophilic, pH 4-6; NE = neutrophilic, pH 6-8; AL = alkaliphilic, pH 8-11; EL = extremely alkaliphilic, pH > 11.		
PHC	pH conditions Only in STR. Value or range in three subfields for minimum, optimum and maximum pH (in that order, subfields separated by ";". If a subfield contains no data, care must be taken that the semicolon which terminates this empty subfield is entered. If growth has been tested at a particular pH without establishing the range, the pH value is indicated with the growth response (+, -, v). Examples: 4.5; 5.5-6.5; 9.3 ;4.5-5.0	WATC	Conditions of water activity In SP and STR. Minimal water activity (a_w) for growth and, after "<", osmoticum and concentration used.
LC	Light conditions In SP and STR. Value of brightness (in L (LUX) or E (Einstein)) and wavelength in three subfields for minimum, optimum and maximum (in that order, subfields separated by a semicolon. If a subfield contains no data, care must be taken that the semicolon which terminates this empty subfield is entered. Illumination favouring growth in general is (also) entered in CONDS.	ELDO	Electron donor In SP and STR. Two subfields for organic and inorganic compounds. Specification after a colon with abbreviations. Subfields are separated by a comma. O = organic compounds (organotrophic): METH = methanol; etc. IO = inorganic compounds (lithotrophic): NH3 = ammonia; NO2 = nitrite; Fe = iron (ferro); H2 = hydrogen-gas; H2S = hydrogen sulfide; S = sulfur; S2O3 = thiosulfate; etc. Example: IO : H2S,S,S2O3
SALR	Salinity relationships: I In SP and STR. NH = non-halophilic (best growth in media containing less than 1% NaCl); SH = slightly halophilic ("marine") (best growth in media containing 1 to 5% NaCl); MH = moderately halophilic (best growth in media containing 5 to 10% NaCl); EH = extreme halophilic (best growth in media containing more than 10% NaCl); HT = halotolerant (non-halophilic which can tolerate salt).	ENERS	Energy source In SP and STR. P = phototrophic; C = chemotrophic
SALC	Salinity conditions Only in STR. Name of compound(s) or ion(s) and minimum, optimum and maximum concentration (in that order). The val-	ELAC	Special electron acceptors In SP and STR. Abbreviations can be used; for example: NO3 = nitrate; NO2 = nitrite; SO4 = sulfate; CO2 = carbonate; S = sulfur; FUM = fumarate; etc.
		CONDREM	Remarks on growth conditions In SP and STR.
		CONDLIT	Literature on growth conditions In SP and STR.

H. Chemistry and Enzymes

- ENZ** Enzymes produced: I
In SP and STR. Name of enzyme(s). Subfields are separated by “;”. Remarks after “<” may include optimal conditions for production and the information whether the enzyme is inducible. Prefixes (1, 2, D, L, alpha, beta) to be entered after the name, separated by “/”. No abbreviations in the database. Abbreviations or EC numbers for input will be organized through a thesaurus (based on Enzyme Nomenclature, 1984).
Example: urease; xylosidase/beta; amylase
- DEC** Decomposition, degradation and utilization capacities: I
In SP and STR. Organic substrates, like cellulose, xylan, peptone, lignin, chitin, keratin. Subfields separated with “;” without spaces. Remarks after “<” may include “g” (growth observed), “u” (utilization without observed growth), “d” (degradation and decomposition without utilization or growth), AE (under aerobic conditions), AN (under anaerobic conditions).
NOTE: chemically defined compounds tested as sole source of carbon, nitrogen or carbon + nitrogen for growth are entered in CSOR, NSOR or CNSOR resp.; heterogeneous substrates are entered in DETER.
- DETER** Biodeterioration: I
In SP and STR. Indicated by the deteriorated substrates. Subfields separated with “;” without spaces.
Examples: paper
 optical lenses
 stone monuments
 paintings
- MET** Metabolites: I
In SP and STR. Includes antibiotics, toxins, fermentation products, also “acid”, “gas” etc.). Name of metabolite, possibly substrate (especially in cases of biotransformation). No abbreviations in the database. Abbreviations or CAS (Chemical Abstracts Service) numbers for input will be organized through a thesaurus. Subfields separated with “;” without spaces. Prefixes (1, 2, D, L, alpha, beta) to be entered after the name, separated by “/”. After “<” optimal conditions for production and origin of information may be entered in each subfield.
Examples: lactic acid
 hydroxylation/12-beta from ergosterol
 streptomycin
- LEACH** Leaching
In SP and STR.
- BIOLUM** Bioluminescence
In SP and STR.
- HAEMOL** Haemolysis
In SP and STR. Type of blood and results.
Example: sheep, alpha
- MISPROP** Miscellaneous properties
In SP and STR. For not classifiable metabolic properties, like accumulation of metallic iron or elementar sulfur.
- CHREM** Remarks on chemical data and metabolic properties
In SP and STR.
- CHLIT** Literature on chemical data and metabolic properties
In SP and STR. Literature on the impact of chemical and physical factors.
- I. Biological Interactions*
- SYMB** Symbiosis or synthrophy: I
In SP and STR. Type of symbiosis, Latin name of symbiont or group of symbionts, separated by “:”.
Example: actinorhiza : Alnus kefir
- PATH** Pathogenicity and virulence: I
In SP and STR. Name and/or symptoms of disease, susceptible group of organisms, separated by “:”. If possible use Latin names of organisms. Subfields for different diseases are separated by “;”. Remarks can be entered after “<”. The expression “causing dysentery in man” is entered as: dysentery : man.
Example: crownrot: Rheum rhaponticum
- ALLER** Allergenicity: I
In SP and STR. Type of allergy, susceptible group of organisms (usually “man”, which is not entered).
- TOX** Toxicity: I
In SP and STR. Name of disease or symptoms followed by a semicolon (no spaces!) and susceptible group of organisms, a comma, and then organs (the toxin involved may be given between parentheses) and degree of toxicity. Toxic metabolites should also be listed in field MET.
The word “causing” must not be entered here, like in PATH. After the comma the field contents are not indexed. Use singular for the organism. Subfields are separated by “;”.
Example: skin lesions : man
- ANT** Antagonistic activity: I
In SP and STR. Action and name of susceptible organism, separated by “:”, responsible mechanism, separated by “<”.
Example: inhibition : gram-positive bacteria
- BINTLIT** Literature on biological interactions
In SP and STR.
- BINTREM** Remarks on biological interactions
In SP and STR.

J. Practical Applications

- APPL General and industrial application: I
In SP and STR. Subfields separated by “;”.
Keyword, description.
- TEST Applications in testing and assays: I
In SP and STR. Subfields separated with “;”
without spaces. Assay for amino acids, anti-
biotics, vitamins, fungicides, toxicity, test-
ing materials. The expression “used for” ap-
plying to all entries must not be entered.
Example: biotin assay.
- CTR Procedures for control
in SP and STR.
- PATENT Patent(s): I
In STR. Information in this field is rarely
made public, only upon instruction by pa-
tent holder. It contains patented applica-
tion, country, number. The word “patent”
is to be entered.
Example: U.S. patent 4,033,823, lysocellin
production
- APLIT Literature on applications, patents
In STR.
- APREM Remarks on applications and patents
In STR.

Acknowledgements

The development of data bases for culture collections would not have been possible without the financial support given by the

national resources and the European Commission in its Biotechnology Action Programme (BAP). The Dutch MINE node owes strong support to the Dutch Ministry of Education and Sciences, the Royal Dutch Academy of Sciences, and the Innovation-oriented Programme in Biotechnology (Ministry of Economic Affairs). The Belgian node is indebted to the Belgian Science Policy Office (project “Belgian Coordinated Collections of Microorganisms”) and to the Bureau M. van Dijk, particularly Mr L. Caeymex. The German node thanks the German Ministry of Research and Technology.

The format was developed in close collaboration with the culture collections participating in MINE.

References

- Environmental Resources Ltd.: EEC Information System for microorganisms. Report prepared for the Commission of the European Communities 1984
- Enzyme Nomenclature: Recommendations (1984) of the Nomenclature Committee of the International Union of Biochemistry. Orlando, Academic Press 1984
- Gams, W., Hennebert, G. L., Stalpers, J. A., Janssens, D., Schipper, M. A. A., Smith, J., Yarrow, D.: Structuring strain data for storage and retrieval of information on fungi and yeasts in MINE, the Microbial Information Network Europe. *J. Gen. Microbiol.* 134, 1667–1689 (1988)
- McGowan, V. F., Skerman, V. B. D. (eds.): World Directory of Collections of Cultures of Microorganisms, 2nd ed. Brisbane 1982
- Pridham, T. G.: Micro-organism culture collections: acronyms and abbreviations. *Agricultural Reviews and Manuals ARM-NC-14*. Agriculture Peoria, U.S. Dept. 1981

Dr. J. A. Stalpers, Centraalbureau voor Schimmelcultures, P.O. Box 273, 3740 AG Baarn, The Netherlands