NOTES

Interchange of Abbreviations and Full Generic Names in Computers

MICAH I. KRICHEVSKY,¹ CYNTHIA A. WALCZAK,¹ MORRISON ROGOSA,¹ AND
RAYMOND JOHNSON²

Microbial Systematics Section, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20205¹ and BBL Microbiology Systems, Cockeysville, Maryland 21030²

Previously, we published a list of abbreviations for genus names of bacteria. We now present the guidelines used for abbreviation construction, an expanded list of codes used for parts of genus names, and an improved list of abbreviations. An appendix contains a discussion of some methods for searching lists of abbreviations and an indication of the relative merits of the search methods.

Leveritt and Skerman (6) have suggested two alternatives to our original set of abbreviations for coding the generic names of bacteria (3). The stated main motivation for their alternative sets is that our original set did not fall in the same alphabetic sequence as the generic names themselves. Leveritt and Skerman state that this lack of sequence correspondence renders it difficult and inefficient "to perform changes from generic names to abbreviations and vice versa and to automatically arrange the names or abbreviations in alphabetical order for indexing purposes." However, we contend that the consideration of corresponding sequences is less important than others. Furthermore, there are efficient strategies available for interchange between two lists which do not depend upon identical alphabetic order.

On first inspection, the proposal to make the list of abbreviations fall in the same alphabetic sequence as the list of names would seem to have much merit in allowing simple binary searching (for definition, see Appendix) of both lists. However, there is a serious problem with requiring that the abbreviations fall in the same alphabetic sequence as the generic names. It becomes impossible to insert a new abbreviation between two abbreviations which are already alphabetically contiguous.

Expansion is not practical if the list of abbreviations must always conform in alphabetical sequence to the generic names. Revision of the list will be needed at some point. For example, at least 19 pairs in the four-letter list proposed by Leveritt and Skerman have no room for later insertions (i.e., ACHL-ACHM; ACTM-ACTN; ACTP-ACTQ; ACTQ-ACR; ACR-ACRS; BETB-BETC; BLSB-BLSC; CAUB-CAUC; CHLB-CHLC; CHRA-CHRB; DESM-DESN;

FSOB-FSOC; LACS-LACT; MENB-MENC; MYXB-MYXC; NOSO-NOSP; PLOB-PLOC; SIDM-SIDN; STAF-STAG). The recently published generic names *Actinopolyspora* and *Methanobrevibacter* cannot be inserted in their alphabetic place because of existing contiguous pairs (ACTP-ACTQ; MENB-MENC). This rigidity of sequence requires that the abbreviations be changed if any new genus names are inserted between any of these pairs. Such revision would require extensive (and expensive) editing of existing archival computer files.

Some of the design considerations of an alphabetic coding system of abbreviations are as follows: (i) condensing communication, (ii) standardizing abbreviations, (iii) making searches more efficient, (iv) saving space in computers, (v) allowing internal accuracy checks, (vi) expanding to accommodate new names, (vii) making abbreviations mnemonic, (viii) making abbreviations as dissimilar as possible, and (ix) choosing abbreviation elements that are contained in the word being abbreviated. Some of these points are interrelated.

Letter abbreviations should be meaningful to the user. It would actually be more efficient to use pure numbers as far as computers are concerned. It is our opinion that computers should conform to the requirements of the users, not the converse. Thus, a set of abbreviations should be evocative of the words they represent (i.e., mnemonic). To accomplish this, our original abbreviation list was chosen so that (in order of priority): (i) the first letter of the abbreviation and the generic name was the same; (ii) all letters of an abbreviation were also found in the genus name (except F for PH); (iii) a consistent set of one- and two-letter codes for commonly used parts of generic names was generated and

used to automatically construct many abbreviations; (iv) within the context of i, ii, and iii, abbreviations were constructed to have at least two letters of difference where possible.

We then tested the abbreviations by asking colleagues to deduce the names of genera from the abbreviations without any clue as to the method used to construct them. Within a few trials, most of those asked deduced the codes of iii above and could guess the names of the familiar genera with accuracy. This ease of correlating the abbreviation with the genus name should lead to more accurate coding and editing of data. (For this discussion, the term "abbreviation" will refer to the full abbreviation of a generic name, while "code" will refer to the one- or two-letter elements representing commonly used parts of generic names.)

An example of a code which we did not list in the original article, but which was used, is the use of a final "L" for the suffix "-ella" as in *Bartonella* (BRTL), *Bordetella* (BDTL), *Pasteurella* (PSTL), *Rickettsiella* (RKSL), and *Salmonella* (SLML). The complete amended list of codes is given in Table 1.

Inconsistencies in the use of the code elements to generate abbreviations in the original proposal (3) were induced by conflicts in the logic of using codes to represent parts of names. For example, the abbreviation for *Acetobacter* becomes ACBT. To minimize confusion, the abbreviation for *Achromobacter* was chosen as the first four letters, ACHR, rather than AHBT. The latter (AHBT) would have been only one letter different from ACBT. Most of the inconsistencies in the initial proposal were generated in a similar manner.

The frequency of coding errors is reduced when abbreviations differ by at least two letters. The chance of miscoding two letters is much less than that for a single letter error. Error detection and resolution become simpler with two-letter differences. Therefore, we tended to minimize the number of abbreviations which differed by only one letter.

However, complete adherence to the rule of more than one-letter difference between any two abbreviations was too restrictive if a number of genera began with the same prefix (e.g., 13 genera with THIO as a prefix) and could not always be respected. (Expansion of the abbreviations to five characters does not seem to help much in the example given by Leveritt and Skerman. Many of their five-letter abbreviations still differ by only one letter from each other, e.g., CALBC-CAMBC-CARBC; FLVBC-FLXBC; HAMBL-HAMPL; Menco-MEYCO; MILSO-MIMSO-MIPSO; etc.)

In reviewing our original list, we found a du-

plication of an abbreviation (ATPN) as well as three cases in which improved abbreviation construction was desirable (see below). The revised list of abbreviations is given in Table 2. The list has been expanded to include many generic names published since the previous list. The generic names came from various sources: *International Journal of Systematic Bacteriology* (volumes 26 to 29), *Archives of Microbiology*, *Journal of Bacteriology*, *Journal of Clinical Microbiology*, *Applied and Environmental Microbiology*, *Antonie van Leeuwenhoek Journal of Microbiology and Serology*, and other widely read journals (predominantly in the period of 1976 to 1979); the last two catalogs of the American Type Culture Collection (1, 2); and the nomenclature files of the Bergey's Manual Trust through a personal communication from J. G. Holt. The list includes names not validly published. A new list of all valid names has been

TABLE 1. Codes for frequently used parts of generic names, including prefixes and suffixes

Codes	Name part
AC	ACETO
AR	AERO, AR
AT	ACTINO
AZ	AZOTO, AZO
BC	BACILLUS
BI	BIFID
BS	BLASTO
BT	BACTER, BACTERIUM, BACT, BACTERIO
CE	CELLULO
CO	COCCUS
CS	CYSTIS, CYSTO
DP	DIPLO
EN	ENTERO
F	PH
HL	HALO
HM	HAEMO
L	ELLA
LC	LACTO
LE	LEUCO
LP	LEPTO
ME	METHANO
MI	MICRO
ML	METHYLO
MN	MONAS, MONAD
MX	MYXO
MY	MYCO, MYCES
NI	NITRO
NM	NEMA
NO	NITROSO
OS	OSCILLA, OSCILLO
PD	PEDIO, PEDIA
PE	PELO
PL	PLASMA
PR	PARA, PR
PS	PSEUDO
RO	RHODO
SD	SIDERO
SF	SPHAERA, SPHAERO
SO	SPORO, SPORA, SPORIA, SPORANGIUM
SP	SPIRO, SPIRA, SPIRILLUM
ST	STREPTO
SU	SULPHUR, SULFO
TH	THIO
TM	THERMO, THERMUS
TX	THRIX
VB	VIBRIO
ZY	ZYMO

TABLE 2. Expanded list of abbreviations of generic names of bacteria

Generic name	Abbreviation	Generic name	Abbreviation
Acetobacter	ACBT	Arizona	ARIZ
Acetobacterium	ACBM	Arthrobacter	ARBT
Acetomonas	ACMN	Azomonas	AZMN
Acholeplasma	ACPL	Azospirillum	AZSP
Achromatium	ACMT	Azotobacter	AZBT
Achromobacter	ACHR	Azotomonas	AQMN
Achroonema	ACNM	Bacillus	BC
Acidaminococcus	AACO	Bacterionema	BTNM
Acinetobacter	ANBT	Bacterium	BT
Actinobacillus	ATBC	Bacteroides	BTRD
Actinobifida	ATBI	Bactoderma	BTDA
Actinomadura	ATMA	Bartonella	BRTL
Actinomonospora	ATMO	Bdellovibrio	BDVB
Actinomyces	ATMY	Beggiatoa	BGGT
Actinoplanes	ATPN	Beijerinckia	BEIJ
Actinopolyspora	ATPS	Beneckea	BNKA
Actinopycnidium	ATPC	Betabacterium	BEBT
Actinosclerotium	ATSC	Betacoccus	BECC
Actinosporangium	ATSO	Bifidobacterium	BIBT
Actinosynnema	ATSY	Blastobacter	BSBT
Aegyptianella	AGTL	Blastocaulis	BSCL
Aerobacter	AEBT	Blastococcus	BSCO
Aerococcus	ARCO	Blattabacterium	BLBT
Aeromonas	ARMN	Bordetella	BDTL
Agrobacterium	AGBT	Borrelia	BRRL
Agromyces	AGMY	Brachyarcus	BRAC
Alcaligenes	ALC	Branhamella	BNHL
Alteromonas	ALMN	Brevibacterium	BVBT
Alysiella	ALSL	Brochothrix	BRTX
Amoebobacter	AMBT	Brucella	BRCL
Amorphosporangium	AMSP	Butyribacterium	BUBT
Ampullaria	APUL	Butyrivibrio	BUVB
Ampullariella	APRL	Caldariella	CALL
Anabaena	ANBN	Caldobacter	CABA
Anabaenopsis	ANBP	Calothrix	CATX
Anacystis	ANCS	Calymmatobacterium	CMBT
Anaerobiospirillum	ANSP	Campylobacter	CPBT
Anaeroplasma	ANRP	Capnocytophaga	CPCT
Anaerovibrio	ANVB	Cardiobacterium	CDBT
Anaplasma	ANPL	Caryophanon	CARY
Ancalochloris	ANCA	Caseobacter	CEBT
Ancalomicrobium	ANMI	Catenabacterium	CABT
Angiococcus	AOCO	Caulobacter	CLBT
Aplanobacterium	APBT	Caulococcus	CLCO
Aquaspirillum	AQSP	Cellulomonas	CEMN
Arachnia	ARAC	Cellvibrio	CEVB
Archangium	ACNG	Chaemisiphon	CHSN

published by the International Judicial Commission (7). It should be consulted for the validity of any specific name.

The inconsistency in abbreviation length in our original proposal was not accidental. It was meant either to reinforce the mnemonic aspects of the two-letter code for abbreviations (as in VB for *Vibrio*) or to resolve a similarity in abbreviations where using the first four letters did not help (NOC for *Nocardia* instead of NOCA, since NOCO stands for *Nitrosococcus*). Most computer languages will accept words of varying length as input. In such cases, the use of less than the maximum number of characters in an abbreviation presents no problem. If fixed-length input is a requirement of a particular system, then the abbreviation can be lengthened to four characters by adding nonletter characters

(e.g., VB becomes VB**). Again, the choice is between convenience for the microbiologist and for the computer.

The use of SR for *Sarcina* was not a good choice because of SP for *Spirillum*. SARC as proposed by Leveritt and Skerman clearly is the better choice and was our first change. It should be noted that the list Leveritt and Skerman published (6) as ours was actually a preliminary unpublished version, containing errors. These were corrected in the published version (3).

The second change is the abbreviation for *Salmonella*. We found that miscoding occurred because users transposed SAML into the first four letters of the genus name, SALM. To avoid this error, we changed the abbreviation to SLML.

Last, the abbreviation for *Rochalimaea* was

TABLE 2.—*Continued*

Generic name	Abbreviation	Generic name	Abbreviation
Chainia	CHNA	Eperythrozoon	EPRZ
Chamaesiphonosis	CMSR	Erwinia	ERWN
Chlamydia	CLMD	Erysipelothrix	ESTX
Chlorobium	CLOR	Escherichia	ESCH
Chloroflexus	CLFX	Eubacterium	EUBT
Chlorogloeopsis	CLGL	Excellospora	EXSO
Chloronema	CLNM	Ferribacterium	FRBT
Chloroplanea	CLPN	Ferrobacillus	FRBC
Chloropseudomonas	CSMN	Fischerella	FSEL
Chondrococcus	CNCO	Flavobacterium	FVBT
Chondromyces	CNMY	Flectobacillus	FTBC
Chromatium	CMTM	Flexibacter	FXBT
Chromobacterium	CRBT	Flexigladus	FXGD
Chroococcidiopsis	CRDP	Flexithrix	FXTX
Cillobacterium	CIBT	Francisella	FRCL
Citrobacter	CTBT	Frankia	FRNK
Cladothrix	CLTX	Fusiformis	FSFM
Clathrochloris	CLCH	Fusobacterium	FSBT
Clonothrix	CNTX	Fusocillus	FSCL
Clostridium	CLOS	Gaffkya	GFKY
Cocobacillus	COBC	Gallionella	GLNL
Comamonas	CMMN	Gemella	GMLL
Coprococcus	CPCO	Gemmiger	GEMM
Corynebacterium	CNBT	Geodermatophilus	GDMF
Cowdria	CWDR	Gloeobacter	GLBT
Coxiella	COXL	Gloeocapsa	GLCP
Crenothrix	CRTX	Gloeotheca	GLTC
Cristispira	CRSP	Gluconoacetobacter	GABT
Curtobacterium	CUBT	Gluconobacter	GNBT
Cylindrospermum	CLDR	Gordona	GRDN
Cystobacter	CSBT	Grahamella	GRML
Cytophaga	CTFG	Haemobartonella	HMBL
Dactylosporangium	DTSO	Haemophilus	HMFL
Dermatophilus	DMFL	Hafnia	HAFN
Dermocarpa	DMOC	Haliscomenobacter	HSBT
Derrxia	DRXA	Halobacterium	HLBT
Desmanthos	DMTS	Halococcus	HLCO
Desulfomonas	DSMN	Halomonas	HLMN
Desulfotomaculum	DSMC	Herellea	HERL
Desulfovibrio	DSVB	Herpetosiphon	HPSN
Desulfuromonas	DUMN	Hydrogenomonas	HYMN
Diplococcus	DPCO	Hyphomicrobium	HFMI
Ectothiorhodospira	ETSP	Hyphomonas	HFMN
Edwardsiella	EDWL	Intrasporangium	ITSO
Ehrlichia	EHRL	Janthinobacterium	JNBT
Eikenella	EKNL	Jensenia	JSNA
Elytrosporangium	ELSO	Kineosporia	KNSO
Empedobacter	EMBT	Kingella	KNGL
Enterobacter	ENBT	Kitasatoa	KTST

originally listed as RKLM. Since "K" does not appear in the generic name, this abbreviation was changed to RCLM to conform to priority ii.

Leveritt and Skerman (6) state that a 75% reduction in the number of characters stored in the computer results from a five-letter set of abbreviations. This is a maximum figure derived from the longest generic name and is true only when fixed-length storage is used. As explained below, fixed-length storage is not required. Thus, the average length is a better measure of saving. Since the average length of the generic names in the original list is 11.7 characters, the average saving for four letters is 66% compared to a 57% average saving for five letters.

The majority of computers in use today either

are manufactured by IBM or are IBM-like in internal architecture. These computers contain only four characters per 32-bit word with eight bits per byte or character. If a "word" orientation is used to store abbreviations, a four-letter abbreviation would be efficient on an IBM-like machine, but would leave one byte unused in a five-byte-per-word computer like the PDP-10. A five-letter abbreviation would be efficient on a PDP-10 computer but would require two words on an IBM-like machine and would waste three bytes. However, both the IBM 370 and the PDP-10 can store and access characters individually. They are not necessarily grouped into words, but may be packed. This means that the five-letter abbreviations will take up 20% more space

TABLE 2.—Continued

Generic name	Abbreviation	Generic name	Abbreviation
Klebsiella	KLEB	Methylovibrio	MLVB
Kluyvera	KYVR	Microbacterium	MIBT
Kurthia	KURT	Microbispora	MBSO
Kusnezovia	KNZO	Micrococcus	MICO
Lachnospira	LCSP	Microcylus	MICY
Lactobacillus	LCBC	Microcystis	MICS
Lamprocystis	LMCS	Microechinospora	MISO
Lampropedia	LMPD	Microellobosporia	MLSO
Legionella	LGNL	Micromonospora	MOSO
Leptonema	LPNM	Micropolyspora	MPSO
Leptospira	LPSP	Microscilla	MISL
Leptospirillum	LPSR	Microsporangium	MIRG
Leptothrix	LPTX	Microtetraspora	MTSO
Leptotrichia	LPTC	Microthrix	MITX
Leuconostoc	LEUC	Mima	MIMA
Leucothrix	LETX	Moraxella	MRXL
Levinea	LEVN	Morganella	MGNL
Lieskeella	LSKL	Mukherjeella	MUKL
Lineola	LNOL	Mycobacterium	MYBT
Lingelsheimia	LNGM	Mycococcus	MYCO
Listeria	LIST	Mycoplasma	MYPN
Lophomonas	LOMN	Mycoplasma	MYPL
Lucibacterium	LUBT	Myxobacter	MXBT
Lwoffia	LWFL	Myxococcus	MXCO
Lyngbya	LNBA	Myxomicrobium	MXMB
Lysobacter	LYBT	Nannocystis	NACS
Lyticum	LYTC	Naumannella	NMNL
Macromonas	MAMN	Neisseria	NEIS
Magnococcus	MGOV	Neorickettsia	NERK
Marinovibrio	MRVB	Nevskia	NVSK
Megasphaera	MGSF	Nitrobacter	NIBT
Melittangium	MLNG	Nitrococcus	NICO
Meniscus	MSCU	Nitrosococcus	NOCO
Metallogenium	MLOG	Nitrosolobus	NOLB
Methanobacterium	MEBT	Nitrosomonas	NOMN
Methanobrevibacter	MEVT	Nitrospira	NOSP
Methanococcus	MECO	Nitrospina	NISN
Methanogenium	MEGN	Nocardia	NOC
Methanomicrobium	MEMI	Nocardioides	NOCI
Methanomonas	MEMN	Nocardiopsis	NCDP
Methanosarcina	MESR	Nodularia	NDLA
Methanospirillum	MESP	Noguchia	NGCA
Methylobacillus	MLBC	Nostoc	NSTC
Methylobacter	MLBA	Nostocoida	NSCD
Methylobacterium	MLBT	Obesumbacterium	OBBT
Methylococcus	MLCO	Oceanomonas	OCMN
Methylocystis	MLCS	Oceanospirillum	OCSP
Methylomonas	MLMN	Ochrobium	OCRB
Methylosinus	MLSN	Odontomyces	ODMY

on either type of machine. One does not have an "extra" byte to use on a PDP-10, if space is the overriding consideration.

Leveritt and Skerman's main argument for arranging abbreviations in alphabetical sequence of generic names apparently is based on the erroneous assumption that the number of comparisons in a search method is the overriding determinant of cost. It is true that binary searches involve, on the average, fewer comparisons than sequential searches. This difference can be significant for large lists. However, the present list of genera is *not* considered large in terms of computer searching and is unlikely to grow too large in the foreseeable future.

There are many other factors involved in the time and cost of a search in addition to the number of comparisons. The number of comparisons is overriding when the number of items to be compared is too large to remain in computer main core and must be fetched from a disk (or tape, if a tape-only system is used) before comparisons can be made. This mechanical action is slow and costly. However, this is not a problem with a list of this length in computers of the IBM 370 and DEC PDP-10 class.

For comparative purposes the relative efficiencies of binary searches vis-à-vis sequential searches were determined on the original list of abbreviations. The two types of searches were

TABLE 2.—Continued

Generic name	Abbreviation	Generic name	Abbreviation
Oerskovia	OSKV	Pseudonocardia	PSNC
Oscillaria	OSLR	Ramibacterium	RMBT
Oscillatoria	OSLT	Renobacter	REBT
Oscillospira	OSSP	Rhanella	RHNL
Paracoccus	PRCO	Rhizobium	RHIZ
Paracolobactrum	PRCB	Rhodococcus	ROCO
Paranaplasma	PRPL	Rhodocyclus	ROCY
Pasteurella	PSTL	Rhodomicrobium	ROMI
Pasteuria	PSIA	Rhodopseudomonas	ROPS
Pectinatus	PCTS	Rhodospirillum	ROSP
Pectobacterium	PCBT	Rhodotheca	ROTC
Pediococcus	PDCO	Rickettsia	RKTS
Pedomicrobium	PDMI	Rickettsiella	RKSL
Pelodictyon	PEDT	Ristella	RSTL
Pelonema	PENM	Rochalimaea	RKLM
Peloploca	PEPC	Rothia	ROTH
Pelosigma	PESM	Ruminococcus	RMCO
Peptococcus	PTCO	Runella	RUNL
Peptodiplococcus	PTDP	Saccharobacterium	SABT
Peptostreptococcus	PSCO	Saccharomonas	SAMN
Pfeifferella	PFRL	Saccharomonospora	SASO
Phormidium	PHMD	Saccharopolyspora	SAPP
Photobacterium	PHBT	Salmonella	SLML
Phragmidiothrix	PHTX	Saprospira	SASP
Phyllobacterium	PLBT	Sarcina	SARC
Pilemelia	PIML	Scytonema	SYNM
Pillotina	PITN	Selenomonas	SEMN
Planctomyces	PTHY	Seliberia	SEBR
Planobispora	PBSO	Serpens	SRPN
Planococcus	PNCO	Serratia	SER
Planomonospora	PMSO	Shigella	SHGL
Plectonema	PTNM	Siderobacter	SDBT
Plesiomonas	PLMN	Siderocapsa	SDCP
Polyangium	PONG	Siderococcus	SDCO
Porochlamydia	PCLM	Sideromonas	SDMN
Proactinomyces	POAT	Sideronema	SDNM
Proactinoplanes	PAPN	Siderophacus	SDFC
Prochloron	POCL	Siderosphaera	SDSF
Promicromonospora	PMMS	Simonsiella	SMSL
Propionibacterium	PRBT	Sorangium	SRNG
Prostheco bacter	PTBT	Sphaerophorus	SFFR
Prosthecochloris	PCCL	Sphaerotilus	SFTL
Prosthecomicrobium	PCMI	Spirillospora	SPSO
Protaminobacter	POBT	Spirillum	SP
Proteus	PROT	Spirochaeta	SPCT
Providencia	PRVD	Spiroplasma	SPNM
Pseudanabaena	PSBN	Spiroplasma	SPPL
Pseudobacterium	PSBT	Spiroschaudinnia	SPSD
Pseudomonas	PSMN	Spirosoma	SPSM

performed on both the IBM 370 and the PDP-10 computers. No detectable differences of time were evident. In the PDP-10, increasing the list to 1,000 items resulted in barely measurable increased search times which could be ascribed to the extra comparisons required by sequential searching. The maximum increase was 22 ms for an item not in the list at all. The costs for searches in the milliseconds range are not major since the cost of reading the data from the disk into the computer core and storing the results back on disk will usually be much greater than the search itself. For a more detailed discussion of various methods of searching lists of abbreviations and their relative merits, see the Appendix.

One major use for a list of generic abbreviations might be to aid in retrospective literature searches and abstracting by computer. This will be especially useful now that the list of approved names is available (7). Such a list does not change the need for indices of synonymy to allow perusal of older literature, nor does it eliminate the need to access disapproved names in journals failing to conform in editorial policy to that of the International Judicial Commission. The information contained in such literature may be of considerable value and cannot be ignored. The list of generic abbreviations could find great use in such areas if it were expandable (as originally envisioned by us). Thus, abbreviations for genera of uncertain standing and discarded names

TABLE 2.—Continued

Generic name	Abbreviation	Generic name	Abbreviation
Spirulina	SRUL	Thermopolyspora	TMPO
Sporichthya	SOCH	Thermostreptomyces	TMSM
Sporocytophaga	SOCY	Thermothrix	TMTX
Sporolactobacillus	SOLC	Thermus	TM
Sporosarcina	SOSR	Thiobacillus	THBC
Staphylococcus	SFCD	Thiobacterium	THBT
Stelangium	SLNG	Thiocapsa	THCP
Stella	STEL	Thiocystis	THCS
Stibiobacter	SBBT	Thiodendron	THDD
Stigmatella	SGML	Thiodictyon	THDT
Streptoalloteichus	STAL	Thiomicrospira	THMS
Streptobacillus	STBC	Thiopedia	THPD
Streptobacterium	STBT	Thioploca	THPC
Streptococcus	STCO	Thiosarcina	THSR
Streptomyces	STMY	Thiospira	THSP
Streptopycnidium	STPN	Thiospirillum	THSL
Streptosporangium	STSD	Thiothrix	THTX
Streptothrix	STTX	Thiovulum	THVL
Streptovorticillium	STVC	Toxothrix	TOTX
Succinimonas	SCMN	Treponema	TREP
Succinivibrio	SCVB	Tyzzeria	TYZZ
Sulfobacillus	SUBC	Ureaplasma	URPL
Sulfolobus	SULB	Veillonella	VEIL
Sulfomonas	SUMN	Vertillomyces	VTMY
Sulfospirillum	SUSP	Vibrio	VB
Symbiotes	SMBI	Vitreoscilla	VITR
Synechococcus	SNCO	Waksmania	WKMA
Synechocystis	SNCS	Wolbachia	WLBC
Tectobacter	TTBT	Xanthobacter	XNBT
Tetracoccus	TTCO	Xanthomonas	XNMN
Thermoactinomyces	TMAT	Xenococcus	XECO
Thermoactinopolyspora	TMSP	Xenorhabdus	XERB
Thermoanaerobium	TMAN	Yersinia	YERS
Thermobacterium	TMBT	Zoogloea	ZOGL
Thermomicrobium	TMMI	Zymbobacterium	ZYBT
Thermomonospora	TMMO	Zymomonas	ZYMN
Thermoplasma	TMPL	Zymosarcina	ZYSR

should be included in a complete list of abbreviations. We have included a number of such names in Table 2.

We submit this revised list for consideration. The task is not complete. The list should be reviewed for its utility by such bodies as the International Judicial Commission and the Bergey's Manual Trust. There is no mechanism in the International Code of Nomenclature of Bacteria for official adoption of such a list of abbreviations (5). Since there is no other mechanism available as far as we know, adoption of this abbreviation system will be by informal acceptance and use. The list should be expanded to include all possible genus names, regardless of validity, to allow for synonymy searching, official list building, standardization of literature references, and all the other information transfer tasks wherein abbreviations of generic names are useful.

APPENDIX

Binary searching is the process of searching a list of items which have been presorted into increasing order. Because the list is presorted, one can always tell

whether a particular item is above or below any other item in the list. The search pattern is as follows. First, inspect the middle item in the list. If this item is below the particular item sought, then one need only search the top half of the list. The middle item of the top half is inspected. If it is still below the target item, the top quarter is considered. Thus, each inspection of the list discards one half the remaining items until the target item is found. Binary search is efficient because the number of items to be inspected is, at most, the logarithm to the base 2 of the number of items in the list.

We programmed a sequential search and Leveritt and Skerman's binary search and a more efficient version of binary search (using the techniques of structured programming) in PL/1 and ran them on the IBM 370. We also programmed these algorithms in SAIL, a version of ALGOL, and ran them on the PDP-10. The time was noted before the search for an item began and after it was found. This did not include the time to initially read in and store the list of 315 items or the time to query and read the item to be searched. There was essentially no difference on the IBM 370. Variations in times were random, and we attribute them to being paged in and out of memory in a time-sharing environment. Times were on the order of milliseconds. This was also true on the PDP-10. In the

TABLE 3. *Partial list of generic names and their abbreviations from Table 2 in alphabetic order of the full names*

Generic name	Abbreviation
1. Acetobacter	ACBT
2. Acetobacterium	ACBM
3. Acetomonas	ACMN
4. Acholeplasma	ACPL
5. Achromatium	ACMT
6. Achromobacter	ACHR
7. Achroonema	ACNM
8. Acidaminococcus	AACD
9. Acinetobacter	ANBT

TABLE 4. *Partial list of generic names and their abbreviations in alphabetic order of the abbreviations*

Abbreviation	Generic name
1. AACD	Acidaminococcus
2. ACBM	Acetobacterium
3. ACBT	Acetobacter
4. ACHR	Achromobacter
5. ACMN	Acetomonas
6. ACMT	Achromatium
7. ACNG	Archangium
8. ACNM	Achroonema
9. ACPL	Acholeplasma

TABLE 5. *Partial listing of alphabetized abbreviations with pointer to location in Table 2 of full generic name*

Abbreviation	Location (see Table 2)
1. AACD	8
2. ACBM	2
3. ACBT	1
4. ACHR	6
5. ACMN	3
6. ACMT	5
7. ACNG	47
8. ACNM	7
9. ACPL	4

PDP-10, increasing the list to 1,000 items resulted in barely measurable increased search times which could be ascribed to the extra comparisons required by sequential searching. The maximum increase was 22 ms for an item not in the list at all. The variability is indicated by three identical search requests for an item near the middle of the list resulting in search times of 6, 12, and 13 ms with the same algorithm. The costs for searches in the milliseconds range are not major since the cost of reading the data from the disk into the computer core and storing the results back on disk will usually be much greater than the search itself.

Having abbreviations out of alphabetic sequence does not "require" either of the two specific solutions proposed by Leveritt and Skerman. That is, when only the generic names are in alphabetical sequence and the abbreviations are not (Table 3), Leveritt and Skerman say that to find the full generic name for a given abbreviation, either a sequential search must be performed on column 2 of Table 3 or a new table must be constructed (Table 4) to permit a binary search. This new table has the sorted abbreviations in column 1 and their corresponding full generic names in column 2, thus doubling the storage space used. A more obvious solution is to reconstruct Table 4 as in Table 5 where the second column contains a pointer to the position of the full generic name in Table 3. It is wasteful and unnecessary to store the full generic names again.

Similar logic may be used to solve a closely related, important searching problem in bacterial nomenclature, that of synonym lists. Since synonyms will not usually fall in the same alphabetical position as the currently accepted name (e.g., *Enterobacter* and *Aerobacter*), binary searching cannot be used. Pointers may be used to indicate the position of the synonym. That is, the abbreviation for *Enterobacter*, ENBT, would be followed by a pointer (position 22) indicating synonymy with AEBT (*Aerobacter*) and vice versa.

Sequential search is not the only way to search an unordered list. Furthermore, the list of abbreviation names is not completely unordered. Each abbreviation begins with the same letter as the corresponding genus name. Using this last fact, Table 6 is constructed. This table contains the 26 letters of the alphabet and a pointer to where in the list of genera (Table 2) each letter begins. (Note, no generic names begin with "Q".) Preliminary use of Table 6 restricts the sequential

TABLE 6. *Example of the alphabet and the position in Table 2 of the first occurrence of the letters as the first letter of the generic name*

Letter	Position in Table 2	Letter	Position in Table 2	Letter	Position in Table 2
A	1	J	190	S	360
B	55	K	192	T	419
C	81	L	199	U	449
D	129	M	222	V	450
E	139	N	269	W	454
F	152	O	288	X	456
G	165	P	297	Y	460
H	177	Q	343	Z	461
I	189	R	343		

search of Table 2 to abbreviations that start with the same first letter. For example, given the abbreviation BGGT, to find the full generic name, using Table 6 restricts the search of Table 2 column 2 to those abbreviations in positions 55 to 80. In the worst case, this is 59 comparisons, if the item begins with the letter "S" (average, 30 comparisons). (In computers like the PDP-10, sequential searches of lists on the order of 128 items or less actually are more efficient than any other methods due to specialized built-in hardware instructions.)

It is highly unlikely that repeated searches are random. In fact, according to the "80-20" rule, which has been commonly observed in commercial applications, 80% of the transactions deal with the most active 20% of a file. The same applies to this 20% resulting in 64% of the transactions dealing with the most active 4% of a file. In this case, a very efficient scheme for searching an unordered list is the self-organizing list as described in Knuth (4). In a self-organizing list, items most frequently used reside near the top of the list, whereas the least frequently used are near the bottom. There are many ways to come by this arrangement. First, a use count can be kept with each item. They can then be periodically sorted based on this count. Second, instead of using memory for counts, an item can be moved to the top of the list when it is accessed. Third, (and most efficient), the item accessed can be interchanged with the one above it, if it is not already near the beginning of the list.

Thus, there are many reasonable solutions to the

very common data processing problem of searching lists.

REPRINT REQUESTS

Address reprint requests to: Dr. Micah I. Krichevsky, Building 31, Room 3B-04, Microbial Systematics Section, National Institute of Dental Research, National Institutes of Health, Bethesda, MD 20205.

LITERATURE CITED

1. Gherna, R. L., and P. Pienta (ed.). 1978. Catalogue of strains I, 13th ed. The American Type Culture Collection. Rockville, Md.
2. Gherna, R. L., and P. Pienta (ed.). 1980. Catalogue of strains I, 14th ed. The American Type Culture Collection. Rockville, Md.
3. Johnson, R., M. Rogosa, and M. I. Krichevsky. 1976. Abbreviations of names of genera suggested for coding microbiological data. *Int. J. Syst. Bacteriol.* **26**:278-282.
4. Knuth, D. E. 1975. The art of computer programming, vol. 3. Addison-Wesley Publishing Co., Inc., Reading, Mass.
5. Lapage, S. P., P. H. A. Sneath, E. F. Lessel, V. B. D. Skerman, H. P. R. Seeliger, and W. A. Clark (ed.). 1975. International code of nomenclature of bacteria. American Society for Microbiology, Washington, D. C.
6. Leveritt, W., and V. B. D. Skerman. 1976. Two alternative proposals for abbreviations of names of genera suggested for coding microbiological data. *Int. J. Syst. Bacteriol.* **26**:442-446.
7. Skerman, V. B. D., V. McGowan, and P. H. A. Sneath. 1980. Approved list of bacterial names. *Int. J. Syst. Bacteriol.* **30**:225-420.