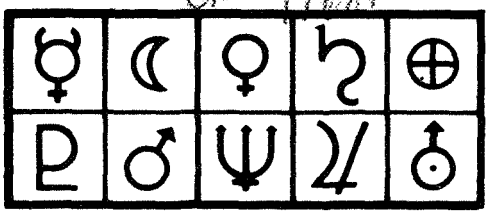


PLANETARY QUARANTINE

SC-RR-70-779
November 1970

AN APPROACH TO COMPUTERIZED BACTERIAL IDENTIFICATION

Richard T. Dillon, Sr.
Planetary Quarantine Systems
Studies Division 1741
Diane B. Holdridge
Dynamic Analysis Division 5162

SANDIA LABORATORIES

OPERATED FOR THE UNITED STATES ATOMIC ENERGY COMMISSION BY SANDIA CORPORATION | ALBUQUERQUE, NEW MEXICO. LIVERMORE, CALIFORNIA

Issued by Sandia Corporation
a prime contractor to the
United States Atomic Energy Commission

SC-RR-70-779

AN APPROACH TO COMPUTERIZED BACTERIAL IDENTIFICATION*

Richard T. Dillon, Sr.
Planetary Quarantine Systems
Studies Division 1741
Diane B. Holdridge
Dynamic Analysis Division 5162
Sandia Laboratories, Albuquerque 87115

Program Completed -- November 1970

ABSTRACT

This report describes the analysis and design of a general approach to bacterial identification which readily lends itself to computerization. This approach has been applied to the microbial sampling data of four Apollo spacecraft. These data were supplied by the PHS Spacecraft Bioassay Laboratory who obtained them in connection with NASA's lunar planetary quarantine responsibility. The approach described here was found to agree quite well with the PHS identifications assigned in the laboratory.

The computer program used in this application is described in detail, and some analysis of the results of the use of this computer program with the Apollo data is given. Other possible applications of the general approach to bacterial identification described here are mentioned.

*This work was conducted under Contract No. W-12,853, Bioscience Division, Office of Space Science and Applications, NASA Headquarters, Washington, D. C.

CONTENTS

FIGURES

TABLES

AN APPROACH TO COMPUTERIZED BACTERIAL IDENTIFICATION

Introduction

Interest in the use of a computer to identify microorganisms arose during microbial sampling of Apollo spacecraft. This sampling was done in response to the lunar planetary quarantine responsibility of NASA which requires that an inventory of probable post landing biological contamination levels at each manned and automated landing site (1) be maintained, and that microbial life be identified, quantified, and insofar as possible located on the lunar surface (2). The U. S. Public Health Service (PHS) collected, cultured, identified and stored for future reference the microbial samples obtained from the spacecraft. A Lunar Information System, designed by Sandia Laboratories (3), is now being used to store, categorize and analyze the data collected by the Public Health Service so that the above information is continuously available to the NASA Planetary Quarantine Officer (4).

Inasmuch as the data collected by the PHS, as well as the results of their identification, were already being stored in the Lunar Information System, it was questioned whether it was feasible for the Lunar Information System to "standardize" the identification process so that it could be accomplished within the information system. In fact, this possibility was earlier considered in the Lunar Information System concept (3). The feasibility of computerized bacterial identification was demonstrated earlier (5) using the Apollo 10, 11, and 12 data that were then available in the Lunar Information System. Feasibility here was based upon agreement with the PHS identification, and it was found that a gross computerized identification scheme agreed with PHS identification in excess of 87 percent of the time. Based on these results, it was proposed that a program be incorporated into the Lunar Information System which would utilize the laboratory data provided by the PHS and, in fact, make the identification. Such a program has been written and its efficiency has been determined by use of previous data from Apollo missions 10, 11, 12, and 13. The data are shown in Figure 1. Some of these data have previously been published (6). It was felt that a total of 7,049 samples would provide a good data base for comparison of the two approaches.

		Apollo			
		10	11	12	13
A.	Catalase positive, gram positive, aerobic cocci				
	001 Staphlococcus, Subgroup I	142	28	1	0
	002 " " II	261	355	334	152
	003 " " III	38	6	29	2
	004 " " IV	131	242	228	51
	005 " " V	163	215	152	34
	006 " " VI	134	154	216	48
	007 Micrococcus, Subgroup 1	142	207	160	124
	008 " " 2	85	103	82	80
	009 " " 3	68	36	77	75
	010 " " 4			1	0
	011 " " 5	20	26	20	4
	012 " " 6			3	0
	013 " " 7	94	104	283	220
	014 " " 8	4	1		0
	Totals Scheme A	1282	1477	1566	790
B.	Catalase negative, gram positive, aerobic cocci				
	015 Entrococcus	2			
	016 Lactic				
	017 Viridans	15	5		1
	018 Pyogenic	3			
	Totals Scheme B	20	5	0	1
C.	Gram positive, spore forming, aerobic rods				
	019 B. alvei	3		2	
	020 B. badius		2	5	
	021 B. brevis	1		1	
	022 B. cereus	2	4	4	4
	023 B. circulans	13	1	8	5
	024 B. coagulans		2	7	4
	025 B. firmus	2	4	3	4
	026 B. laterosporus	1			
	027 B. lentus	7	2	1	4
	028 B. licheniformis	6	8	2	1
	029 B. macerans	3	1		
	030 B. megaterium	3	1	1	
	031 B. pathothenicus	1	20	3	3
	032 B. polymyxa	2	12	4	2
	033 B. pulvifaciens	2	1	3	2
	034 B. pumilus	1	7		0
	035 B. sphaericus	1	1	4	2
	036 B. subtilis			3	1
	Totals Scheme C	45	66	51	32
D.	Aerobic and anaerobic non-spore forming gram positive rods, anaerobic gram negative cocci and anerobic gram negative rods: Test reactions only.				
	037 Aerobic, non-sporeforming gram positive rods	160	183	195	159
	038 Anaerobic, non-sporeforming, gram positive rods				
	039 Anaerobic gram negative cocci				
	040 Anaerobic gram negative rods				
	Totals Scheme D	160	183	195	159
E.	Aerobic gram negative cocci:				
	041 Neisseria, spp.				
	Totals Scheme E	0	0	0	0
F.	Aerobic gram negative rods:				
	042 Achromobacter, spp.		2		
	043 Alcaligenes, spp.	3	9	2	
	044 Flavobacterium, spp.	1	1	7	
	045 Pseudomonas, spp.	1			
	Totals Scheme F	5	12	9	0

Figure 1. PHS Identifications for Apollo 10, 11, 12, and 13

			Apollo 10	11	12	13
G.	046	Gram positive, catalase negative, anaerobic cocci				
		Totals Scheme G	0	0	0	0
H.	047	Gram positive, catalase negative anaerobic cocci				
		Totals Scheme H	0	0	0	0
I.		Anaerobic sporeforming gram positive rods				
	048	C. acetobutylicum				
	049	C. aerofoetidum				
	050	C. amylolyticum				
	051	C. amylosaccharobutylprophylicum				
	052	C. aurantibutyricum				
	053	C. bejerinckii				
	054	C. bifermentans				
	055	C. botulinum				
	056	C. butylicum				
	057	C. butyricum				
	058	C. capitovale				
	059	C. caproicum				
	060	C. *carnis*				
	061	C. cochlearium				
	062	C. difficile				
	063	C. fallax				
	064	C. felsineum				
	065	C. feseri				
	066	C. flavum				
	067	C. haemolyticum				
	068	C. histolyticum				
	069	C. innocuum				
	070	C. iodophilum				
	071	C. lactoacetophilum				
	072	C. lentoputrescens				
	073	C. novyi				
	074	C. parabotulinum				
	075	C. paraputrificum				
	076	C. pasteurianum				
	077	C. pectinovorum				
	078	C. perfringenes				
	079	C. roseum				
	080	C. septicum				
	081	C. sordelli				
	082	C. phenoides				
	083	C. sporogenes				
	084	C. subterminale				
	085	C. tertium				
	086	C. tetani				
	087	C. tetanomorphum				
		Totals Scheme I	0	0	0	0
J.	088	Actinomycetes	2	7	1	2
		Totals Scheme J	2	7	1	2
K.	089	Streptomycetes				1
		Totals Scheme K	0	0	0	1
L.	090	Yeasts	5	8	13	18
		Totals Scheme L	5	8	13	18
M.	091	Molds	37	34	36	14
		Totals Scheme M	37	34	36	14
	092	Atypical cocci	9	57	22	2
	093	Aerobacter	1	0	0	0
	094	Atypical bacillus	17	13	2	0
	095	Nogro Differential Media	409	173	76	1*
		Total Apollo	1992	2020	1997	1020
		*Lost in process	Total = 7049 samples			

Figure 1. (Cont.)

When discussing the "identification" of bacteria, the meaning of the word must be agreed upon as accurately as possible. It would be nice if there were a well defined set containing all bacteria, each of which could be characterized by the results of some standard tests. Then any unknown could be subjected to the tests and the test outcomes compared to those describing the above categories until a match was obtained yielding an identification. Unfortunately, such an idealistic view of the identification of microbes is, to date, impractical. The existing knowledge about the set of all microorganisms is too incomplete, and a theoretical or practical basis for c pletely subdividing the set of those we know about has serious limitations. Some of the difficulties are as follows:

1. A precise workable definition of bacteria is not available. For example, a recent text (7) uses a modified definition due to Cohn, 1872, which is totally inadequate in view of recent knowledge.

2. Even if well-defined, the class of all bacteria would be open ended. Naturally occurring or artifically induced mutations are being formed constantly, many with newly observed properties.

3. The existing classification used in the U. S., Bergey's Manual (8), is based at least in part on old categories which were devised by means of old lists, some of which are now regarded as inconclusive or, at best, subjective.

4. The use of all "comprehensive" keys, including Bergey's, for identification requires the frequent interpretation of statements qualified as: rarely, frequently, some, and usually - which leads to overlapping categories.

The above difficulties reflect those differences between the mechanical way a computer operates and the manner in which people function. The computer requires well-defined categories and tests with a set of rules to relate them. It is difficult to include subjective judgments such as: "since these other tests came out that way and this is variable, its importance is minimal since it looks like_____ anyway". It might seem that a mechanized (or standardized) identification would be difficult, if not impossible, which is probably a fair conclusion if one operates within the above traditional framework. However, simple efficient schemes, capable of computerization, can be devised for special purposes. It should be noted that an identification scheme need only be adequate for the task at hand. There should be no objection to

having such a scheme suitable for the PHS Apollo task, another for clinical diagnostic work, another for work in the problem of bacterial classification or taxonomy, and yet others suitable for research with particular strains.

For NASA's Apollo Program, the U. S. PHS had devised an identification scheme tailored to that particular need (9). Categories appropriate to this need were defined in terms of test outcomes without mandatory reference to traditional categories of bacteria. Appendix 1 contains the current Data Reporting Form and list of categories being used. There was a backup inasmuch as cultures of the sampled microbes were to be stored for future reference. In addition, all laboratory data and the associated identification were stored and available to a computer as part of the Lunar Information System. Thus, it was possible to simulate the identification process and conveniently compare results with those of the PHS. Moreover, the large amount of data available in the NASA Lunar Information System provided an excellent basis for comparison of the two methods.

Every effort was made during the development of the system to retain as much flexibility as possible in the hope that it would have wider applications than the Apollo system for which it was designed. It was felt, in particular, that it should be constructed so that it could easily be modified by a microbiologist without reprogramming. This was accomplished by arranging the program so that it performed operations only on the data subject to the particular identification scheme which it had read. That is, not only were the data input, but the identification process itself became input. The form of the PHS identification scheme suggested a manner in which this could easily be done. The process that the PHS uses in order to identify a sample with the data they have collected was abstracted so that the computer would make the same (or similar) decisions. Thus, with some limitations, almost any scheme to be used can be handled by the program. The program will be described first by using this scheme as an example and then by showing the results and potential applications.

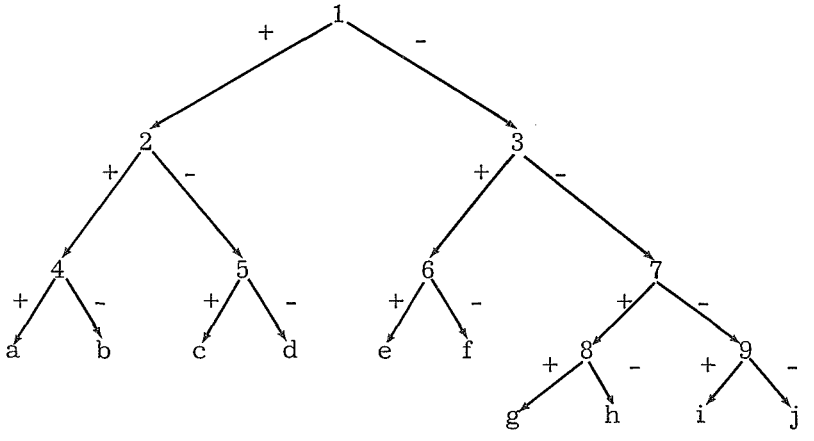
The Abstract Approach

Ideally, a microbiological identification system would consist of a (closed) class, M, of microbiological categories and a set, T, of tests which are sufficient to span M, that is, a set of tests with the property that for every pair of categories in M there is

at least one test which distinguishes between these. Then the problem of identification of a particular sample would reduce to taking the sequence of test results arising from the sample and comparing those with the test results associated with each category in M until an "identification" is made. Indeed, something approaching this ideal is necessary if one is to think in terms of a computerized identification scheme. However, as indicated in the Introduction, the class of all microorganisms as traditionally defined is not closed and, in fact, is continually changing, as is the set of tests that are used. Furthermore, there is no well-defined and generally accepted manner by which one can proceed from a specified series of test outcomes to a definite microorganism category. Rather, there are numerous different approaches to the problem of microbial identification. Some of these have been developed for specific groups of microorganisms, for particular application, and, in some cases, more as a method of research documentation. Some of the various types of approaches will be discussed.

Every approach to identification of organisms in some class M relies on an underlying classification of the organisms belonging to M. A classification is merely a specification of the abstract relationship between organism categories (in M) and test results. There are two basic means of presenting such a classification scheme, both of which from an informational point of view, are essentially equivalent, but which have totally different implications to practicing microbiologists when used for identification purposes.

First, there is the more traditional tree form of a classification scheme, [as exemplified in Bergey's Manual (8), discussed earlier]. In this approach, the tests occur as nodes of a tree, the end branches of which represent organism categories. An abstract example of this is shown below:



In this figure, there are organism categories a, b, c, d, e, f, g, h, i, and j. In the tree, nine tests are needed to distinguish among these. Although this tree indicates each test has only one of two outcomes (+ or -), this need not be the case in general classification schemes [e. g., Bergey's Manual (8)]. When used for identification purposes, this approach to classification has the implied disadvantage that the performance of tests must be done in a sequential manner with time. This results in running the fewest number of tests, but not necessarily in obtaining identification in the least amount of time, or indeed, with the least man-time in repeated routine identification programs.

The second approach to the presentation of classification schemes is through the use of matrices relating test results to organism categories. An example of such a matrix is shown in the figure below:

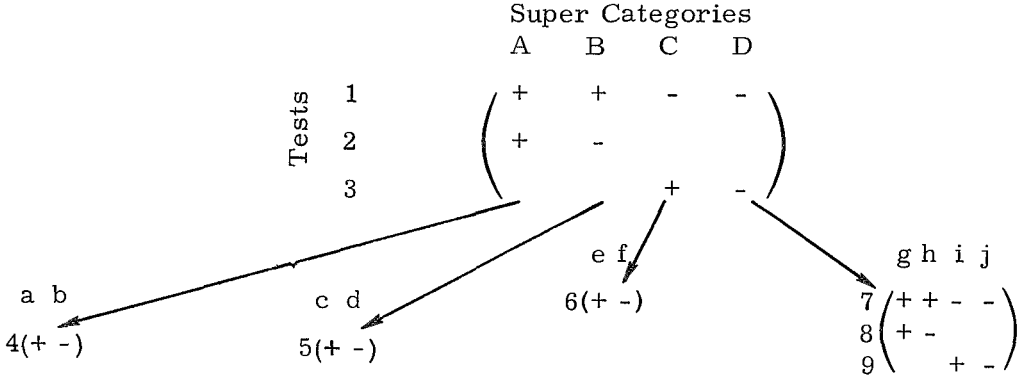
		Organism Categories									
		a	b	c	d	e	f	g	h	i	j
Tests	1	+	+	+	+	-	-	-	-	-	-
	2	+	+	-	-						
	3					+	+				
	4	+									
	5			+							
	6					+					
	7							+	+		
	8							+	-		
	9									+	

Here the rows represent tests, the columns represent organism categories, and an entry in the matrix represents the test outcome for the row (in which it occurs) and for the organism category of the column in which it occurs. Thus, for example, the (5, c) entry indicates that test 5 yields a + result when run on organisms in category c. The matrix form of a classification scheme implies in effect that one must run all tests prior to attempting an identification. This is not to say that all are needed a posteriori, but they must be known since the organism being identified is not known. This has the disadvantage of frequently taking much unnecessary data and requiring unnecessary time, but has the advantage that the scheme itself imposes no sequential time constraints like those found with a tree.

It was asserted earlier that these two approaches (trees and matrices) have essentially the same abstract informational value. This may be seen by comparing the above tree and matrix. They both use precisely the same test results to characterize each organism category. For example, in both cases, to characterize organism g, one needs only specify that tests 1 and 3 be negative and that tests 7 and 8 be positive. Indeed, any column of the matrix in which entries occur is obtained from the path of the tree leading to the organism category corresponding to that column.

From the point of view of taking laboratory data in identification situations, neither the tree nor the matrix approach has an abstract clear-cut advantage as regards "efficiency". Any such advantage would depend highly upon the categories and tests. For example, if some tests occur in all paths of the tree (that is, their results must be known in identifying all the organism categories), then these tests might just as well be represented by a matrix.

From such considerations, and the consideration of the PHS Apollo scheme, a very general approach to classification has been devised which possesses the advantages of both the matrix and tree approaches and minimizes their practical disadvantages. In simple terms this is accomplished by generalizing the notion of a tree in which nodes represent tests, to a tree in which nodes represent matrices. The "categories" of the matrices at the upper levels of the tree are supersets of the original categories (i. e., if a, b are two original categories, one possible outcome for a matrix at the upper levels of the tree is A ={a, b}, meaning A is a or b). To see how this concept might work, we will consider the above examples of a tree and matrix. This may be redrawn as follows:

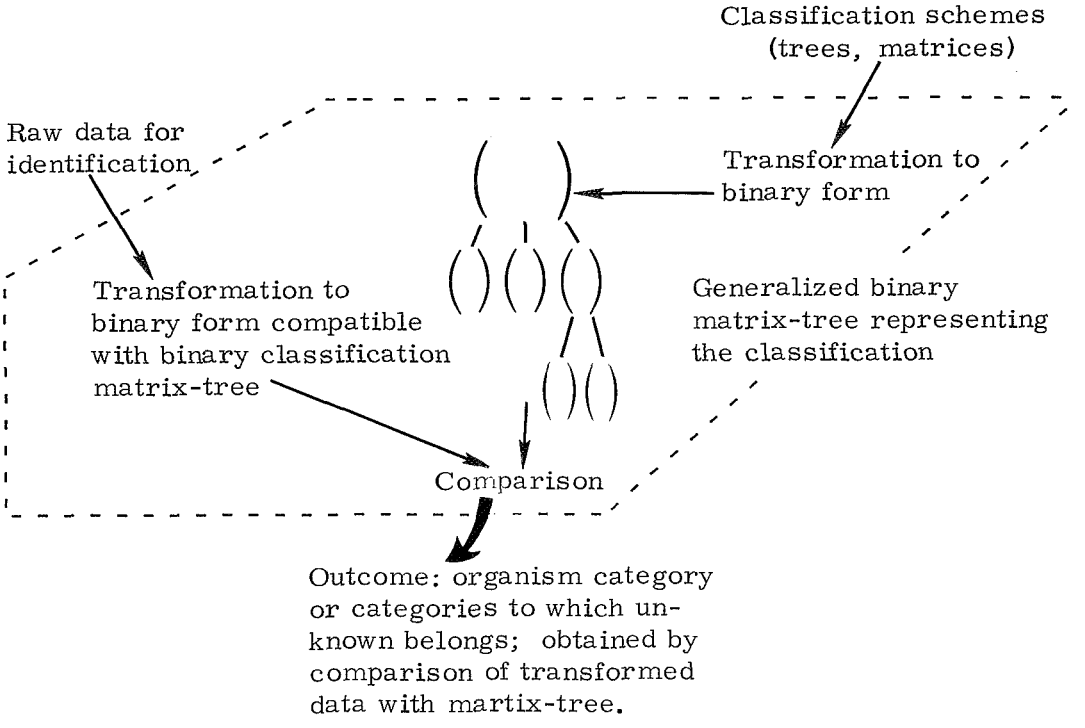


A = {a, b}, B = {c, d}, C = {e, f}, D = {g, h, i, j}

The practical implications of this figure are that tests 1, 2 and 3 will be run simultaneously (or as nearly so as possible) at the outset. The unknown will then belong to supercategory A, B, C or D. From this knowledge, the appropriate test or tests needed to determine the category a, b, c, d, e, f, g, h, i or j will then be run. This represents a compromise between a strict sequential running of tests and running all tests concurrently. This is particularly advantageous if, for example, tests 1, 2 and 3 are particularly easy to run.

There is another advantage to this general approach when computerization is contemplated. All "decisions" are made using matrices. This is advantageous for two reasons. First, this represents a particularly simple computer operation; and second, all operations are of exactly the same type and hence may be regarded as essentially the same computer operation, leading to considerable computer program efficiency (as will be seen later).

One other point in abstracting classification and identification schemes needs to be addressed. In the above examples, the test results were all binary in character (+, -). As stated earlier, this is not always the case. At the same time there are several reasons why it is best that the system be binary for computerization purposes. First, if multiple outcomes occur in a matrix, then the general utility of the program is somewhat lessened since the program and its inputs become considerably more complex. Additionally, decisions about identification become complicated (as will be seen later) due to the increased possibility of "multiple" identifications. Since, as a general precept, we may, through analysis, always regard multiple outcomes in one scheme as binary, in another (by redesigning the former) it has been elected to consider classification and identification as preferably a binary process. This necessitates the introduction of the concept of "transformation" whereby both classification schemes and test data may be transformed from multiple test outcomes to essentially binary outcomes. This is not precisely so since "variable" outcomes are allowed as a third possibility. To explain this, "transformations" of data are distinguished from "transformations" of classification schemes. Abstractly, a description of these transformations is complex. Because of this, only their existence and desirability are discussed here. Later, specific examples will be given in profusion. Thus, the total abstract approach may be summarized in the following diagram:



The part within the dashed lines is performed in the computer.

In application, it was relatively simple to convert the trees in use by the Public Health Service into matrices. Next, these matrices are particularly easy to deal with on modern computers. This choice provided a uniform format which made possible an exceedingly simple computer program. Furthermore, selecting this option made it easy to distinguish between the variability of the tests used and the judgment or decisions of a microbiologist in recording the data he had obtained in the PHS program. The utilization of matrices as the standard identification system permitted the establishment of normal (or standardized) entries in the matrices through transformations that will be described later. This general approach permitted the particular identification scheme (tree, key, etc.) or transformation to be included, not as part of the program, but as input to the program which the microbiologist or other user could change at will, since any use of computers to assist microbiologists in the interpretation of data to make an identification should allow easy modification of both test interpretation and the actual identification scheme itself. To achieve the desired flexibility the computer program should be quite independent of the particular identification scheme in use. It was possible to accomplish this with the PHS problem by analyzing the entire process and separating it into a number of separate operations.

These operations then might be categorized as (1) those associated with the "standard" general identification process (or form of identification), (2) those operations particular to the identification scheme in use, and finally (3) those operations which reflect the decisions or judgments of a microbiologist. To show how the preceding abstract approach was conceived, and as a base for describing the precise computerization of the PHS problem, a description of the latter as it was first seen, is presented below.

It was noted earlier that a microbial identification scheme need only be adequate for some particular purpose. In keeping with this, categories of interest for the Apollo program were established. To classify or identify samples taken from a spacecraft as belonging to some of these categories, appropriate sets of tests were used. Hence, a collection of test sets both necessary and sufficient to classify a sample in the appropriate category was developed. Hopefully such a test set is also nearly minimal in the sense of requiring the least number of tests for this purpose. The identification system used for the Apollo program was developed by the U.S. PHS (9). This system will be described in some detail as a basis for the approach to incorporating it in a computer. At first it would appear that their identification system consists of two parts. The first, or general scheme, takes the form of a flow diagram or tree which uses the results of five tests to enter the appropriate place in the second part. This flow diagram is reproduced as Figure 2.

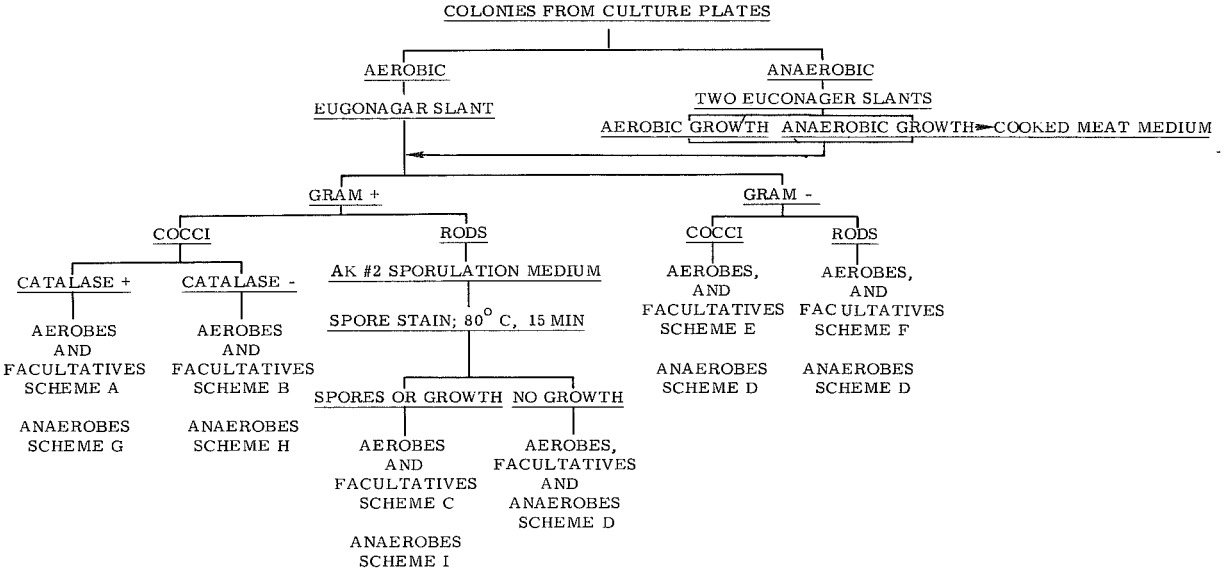


Figure 2. Flow Diagram of General Scheme for PHS Identification

The second part of the system consists of a number of individual schemes (in Figure 2 - A through I) which are reached after stepping through the first part of the identification system. In general, the schemes of this second part allow an identification to be made (i.e., assignment to a category) on the basis of certain specified tests. This relationship of test to microorganism category is usually presented in matrix form. For example, the key for Scheme A of Figure 2 is given as the following matrix.

	Group I Staphylococcus Rosenbach						Group II Micrococcus Cohn							
Subgroup:	I	II	III	IV	V	VI	1	2	3	4	5	6	7	8
Pink Pigment	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Acid from glucose:														
(1) aerobic	+	+	+	+	+	+	+	+	+	+	+	+	±	±
(2) anaerobic	+	+	+	+	+	+	-	-	-	-	-	-	-	-
Coagulase	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Phosphatase	+	+	+	-	-	-	-	-	-	-	-	+	-	-
Acetyl-methyl-carbinol	+	+	-	+	+	+	+	+	+	+	-	-	-	-
Acid from:														
(1) arabinose	-	-	-	-	-	-	-	-	-	+	v	+	-	-
(2) lactose	+	+	v	-	+	v	-	+	v	+	+	+	-	-
(3) maltose	+	+	-	v	+	v	v	+	+	+	+	+	-	±
(4) Mannitol	+	-	-	-	-	+	-	-	+	+	+	+	-	-

± = weak or negative, v = variable

Note: Staphylococcus subgroup I corresponds to S. aureus and subgroups II-VI to S. epidermidis; Micrococcus subgroup 7 corresponds to M. luteus and subgroup 8 to M. roseus.

Figure 3. Key for Scheme A (Baird-Parker)

The key for scheme A may be regarded as typical although there are keys which are more complicated, i.e. consisting of two matrices (one indicating entry to another) as well as keys which admit only a single identification or classification. Scheme A is taken from Baird-Parker (10) and may be regarded as typical of the key approach to identification. This particular key will be discussed in more detail later.

In practice, then, the system consists of a list of microbial categories and a set of tests with rules or procedures relating them. For example, five (or fewer) test results, in conjunction with Figure 2, indicate a scheme to be used. The key for this scheme specifies a number of tests to be performed whose results will permit classification of the sample into a category. A computer is to be programmed so that, given the data, it can simulate the process of identification; that is, it can duplicate the steps the microbiologist takes after the data is collected.

These steps include first interpreting each test result associated with an unknown as positive, negative, or inconclusive. Then the interpreted results are compared with those specified by the key (or matrix) for the scheme being used in order to determine if a suitably close match can be obtained for an identification. The system, and hence the computer, must simulate these steps. In addition, it must allow for differing degrees of variability in the interpretation of the data. The operation might best be seen, in a general way, in terms of the matrix for Scheme A above.

In the case of Scheme A the operation is that the results of a number of tests are compared with the entries in a matrix to obtain a match. This, for example, is done with a matrix constructed to represent the initial tree in order to determine the appropriate final scheme. Then appropriate data is compared with the entries of the matrix for that scheme to determine the sample identification. The fact that the same operations are repeatedly performed with various matrices allows for an extremely simple program. This suggests that the matrices and appropriate indices could be provided as input in the program rather than being part of the program itself. This allows the particular identification scheme being used to be easily changed at any time. There is, however, still the problem of variable or indeterminate test results referred to above.

The way this difficulty was addressed was to incorporate a set of transformations for each of the matrices which would translate the data into a standard form. These transformations are based upon previous experience and are easily modified as required.

The indeterminate nature of the test is taken care of at two levels, thus allowing more flexibility. That indeterminancy which is largely inherent in the scheme is incorporated in the corresponding matrix or matrices, while the indeterminancy due to interpretation of the test is considered in the data transformation. Hopefully this

separation permits the computer to more closely simulate microbiological practice. The precise manner in which this is accomplished will be described in the detailed system description.

In summary, the objective was to write a program (as general as possible) which would be able to make an identification on the basis of recorded laboratory data. This was accomplished by analyzing the PHS identification system and abstracting from this the operations that are performed in making an identification, the schemes used as a basis for the data and identification, and finally the data interpretations that are made by the microbiologist (subject to the system that he is using). This was done in as simple a manner as could be accomplished so that the program could easily be modified to accommodate modifications in the process used by the PHS or even in entirely new identification schemes. There are only a few subroutines which need to be changed for a completely new identification system. The program will be described in detail in the next section.

Identification by Computer

As indicated in the preceding section, bacterial identification can be simple. One takes the results of a number of tests and compares these with those attributed to various categories of bacteria. When the observed results agree with the known results, an identification is made. Practically, there are a number of difficulties as noted before. For computerized identification it is important to approach the above ideal as closely as possible. This was one of the major motivations for the particular approach taken, namely transforming the data so that they could be compared with matrices representing individual schemes in a uniform manner. Through this device, which has other advantages, a relatively simple program can be written since similar operations are iteratively performed. That is, different transformed test results are compared with a matrix representing a different scheme, but the same operations are repeatedly performed.

This is accomplished by representing each scheme by a matrix having as entries only +1, -1, and 99 which represent respectively a positive, negative, or inconclusive test result. Thus, to make a comparison, transformations must be

specified which will convert reported test results into this form. In setting up these transformations it is possible to incorporate some of the judgment that a microbiologist used in making an identification at the bench. In conjunction with the interpretation of test results as used in the matrices, considerable flexibility is allowed. To make the above concepts specific the PHS identification system will be used in some detail. In the preceding section it was noted that the identification procedure begins with a flow diagram which uses the results of five tests to determine which of nine schemes to enter. These latter, in general, were given in a matrix form that related test results to organism categories.

The first part of the Public Health Service system, namely the flow diagram (Figure 2), does not at first appear to fit the above description of a binomial tree. However, during the feasibility analysis, as has been previously mentioned, it was demonstrated that the flow diagram could be regarded as a binomial tree which used the outcomes of five tests to determine the appropriate scheme to enter. This tree is reproduced as Figure 4.

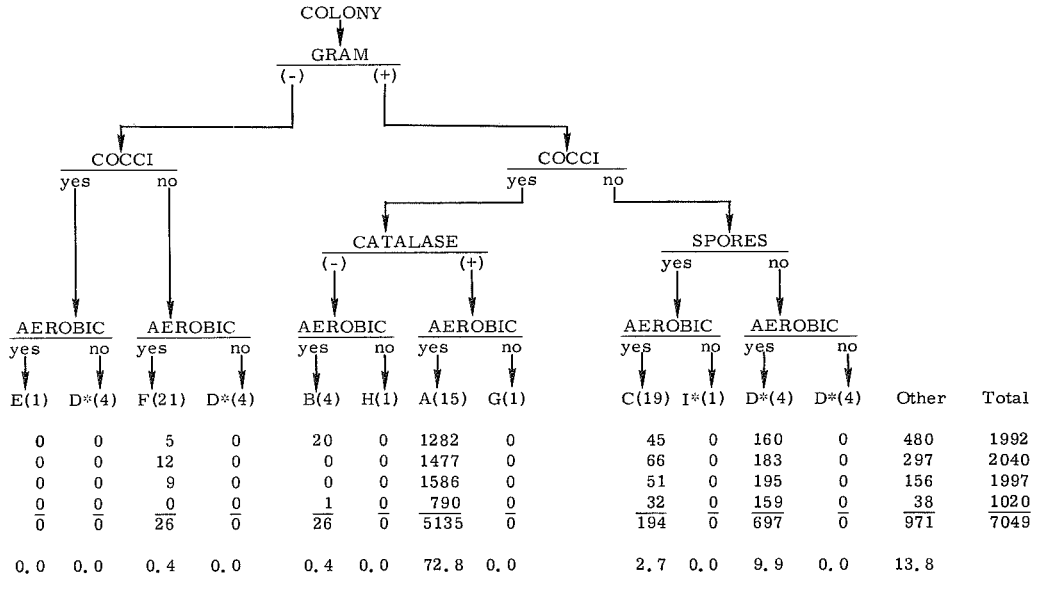


Figure 4. Flow Diagram of General Scheme in Yes/No Form

In Figure 4, schemes A through I are indicated by the appropriate letter with the number in parentheses designating the number of categories that the scheme leads to. In general, schemes themselves consist of matrices although there are exceptions. For example, it was sufficient for the Public Health Service Apollo purpose to record in Scheme D only the results of the four determining tests, e. g. anaerobic, non spore-forming, gram positive rods. Thus scheme D occurs in four places in the above tree, each containing only one identification category (itself) which depends solely upon the five previous tests. Thus, to proceed through the flow diagram (hereafter referred to as the tree), it is only necessary to know the results of five tests which have a yes/no nature. The fact that these outcomes are yes/no in nature is somewhat arbitrary and based upon the Public Health Service requirements for this particular identification process. For example, morphology is reported in a number of categories (see Appendix 1) but decisions are made based upon whether the sample was coccoidial or not. Similarly, spore forming ability is somewhat open ended in that it depends on the ability to detect a spore in a number of ways.

For a computer to be able to deal with the results of biological tests it has to be coded some way. In the Public Health system, the test results are recorded in a column on a card. Thus there is a natural numbering associated with the test, namely, the number of column of the card in which the result was recorded. The tests of primary interest for the tree are as follows:

Table 1. Primary Yes/No Test Results for General Scheme

Test	Col. No.
1. Gram Stain	18
2. Cocci	17
3. Catalase Reaction	41
4. Spore Forming Ability	17
5. Aerobic Growth	15

With the more or less natural conventions indicated above where +1, -1, and 99 were allowed as matrix entries, we have clear interpretations of the tests in Table 1. That is, catalase reaction positive and gram stain positive will be regarded as +1, as are the other three properties listed in Table 1, with negative reactions being -1. An inspection of Figure 4 reveals that not all of the five tests are always required. For example, a gram negative result indicates that the catalase reaction and the spore forming ability does not influence the outcome. Continuing with the conventions

indicated in the beginning of this section, we can use a 99 as a matrix entry where this situation prevails. With these interpretations the tree in Figure 4 can be written as a matrix as follows:

Scheme (Supercategory)

		A	B	C	D1	D2	D3	D4	E	F	G	H	I
		1	2	3	4	5	6	7	8	9	10	11	12
Card	18	1	1	1	-1	-1	1	1	-1	-1	1	1	1
Column	17	1	1	-1	1	-1	-1	-1	1	-1	1	1	-1
	41	1	-1	99	99	99	99	99	99	99	1	-1	99
(Test)	17	99	99	1	99	99	-1	-1	99	99	99	99	1
	15	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1

Figure 5. Matrix Form of General Identification Scheme

In Figure 5 the rows represent the numbered test possibilities, while the columns represent the scheme which is to be used for the next stage in the identification. Thus, what originally appeared as the flow diagram and is now referred to as the tree can be represented as a matrix just as all of the other schemes involved in the Public Health Service identification system. This similarity allows each of the stages in the Public Health Service system to be handled by a computer in a uniform manner, as will be seen. Not only the tests but the schemes need to be indexed by some number readable by a computer. This can be arbitrary or could follow some preconceived pattern. In the case of the Public Health Service system there is a matrix which represents the tree resulting from the flow diagram itself. This is treated somewhat separately inasmuch as it is the first entry into any identification process. The remainder of the schemes are then numbered in their natural order, thus allowing for repetition; that is, A = 1, B = 2, C = 3, and D = 4 (note that D appears four times because it can appear four times from the matrix representing the flow diagram), E = 5, etc. There is the possibility that test results for an unknown are either incomplete or inconsistent with the entries in the matrix in Figure 4. Hence we have the possibility of no identification. No identification can occur in the case that there is no outcome consistent with the test results from the tree or in the case that there is an outcome where there is difficulty matching the matrix. This latter point will be elaborated in more detail later.

The process of converting the schemes themselves to a matrix for the computer identification is relatively simple compared to that of converting the flow diagram to a matrix. These schemes are more typical of those used by microbiologists in general; where a result is positive, +1 is entered, -1 is entered for a negative result, and 99 is used for a weak or variable result. Thus Scheme A (Baird-Parker) of Figure 3 appears as follows:

Identification Matrix A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
14	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
37	1	1	1	1	1	1	1	1	1	1	1	1	99	99
37	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1
42	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
43	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1
53	1	1	-1	1	1	1	1	1	1	1	-1	-1	-1	-1
36	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	99	1	-1	-1
38	1	1	99	-1	1	99	-1	1	99	1	1	1	-1	-1
39	1	1	-1	99	1	99	99	1	1	1	1	1	-1	99
40	1	-1	-1	-1	-1	1	-1	-1	1	1	1	1	-1	-1

Figure 6. Computer Version of Scheme A Matrix

In Figure 6 the column headings indicate the organism category as used by the Public Health Service or Baird-Parker, while the left hand column indicates the entry on the card where the test result will be recorded by the Public Health Service. Similarly this is used for the other schemes. There are a few instances where this conversion procedure is not as straightforward as indicated above. Examples of these occurrences will be provided and discussed later.

With all of the identification procedure represented uniformly as matrices of the above type, it is, in principle, quite simple to make an identification. The appropriate data for an unknown are selected and compared with the columns of the matrix representing the tree to determine the next step in the process. There are, however, a few difficulties remaining. One such difficulty is that the data being compared with the columns in the above matrices must be of the same form as those entries, e.g. +, - or 99 representing indeterminate. The data required are in general either not recorded or given some value in the range 0, 1, ... 9. Some

means must be provided for the computer to determine, for example whether an entry of 3 in the Baird-Parker mannitol test should be counted as +, - or 99. To accomplish this, a series of transformations were provided which would convert such entries to the appropriate value for comparison in the above matrices. It should be noted that this conversion is of a different nature than the previous conversion of test results to entries in the matrices. Previously, an established outcome of a test was being interpreted as a +, - or an indeterminate test for classification (or matrix generation) purposes, whereas now observed data are being translated for comparison with the classification. In this latter case, an attempt is made to simulate the judgment that a microbiologist uses in taking, observing and interpreting the data.

Because of this fundamental difference the transformations were treated in as general and flexible a manner as possible. As part of the input to the program, each test that will be used in the identification process is listed on a card. On the same card each possible outcome that could be recorded by the Public Health Service microbiologists is listed, followed by its interpretation as +1, -1, or 99 as it is to be used in comparison with the matrices. Thus, it is an extremely simple matter to change an interpretation to improve the performance of the identification program. Since occasionally the same test is used in different portions of the identification process and interpreted in a different manner, the transformations have to be associated with the appropriate stage in the identification process.

An example of the above occurs in the tree where column 17 on the PHS card (Cellular Morphology) is used for a number of purposes. In the tree we need the results of 17 to determine cocci or not, that is the second row in the tree matrix; also it is possible that the PHS has checked spores, No. 1 in column 17. Thus, column 17 is used twice as an entry into the tree matrix. It is a bit more complicated than that because there are other tests, namely 65, 66 and 67 which are also used to determine the ability to sporulate. Thus, if, during the morphological examination, a spore is detected, 1 is checked in column 17 and the issue insofar as sporulation is concerned is settled. If something other than 1 is checked, tests 65, 66, and 67 are run. 67 involves a sporulation media. If there is growth on this media, No. 2 is checked and it is concluded that the sample is capable of sporulation. Similarly, in 66 (heatshock) the checking of a 2 indicates sporulation, and in 67 the checking of a 2 also indicates sporulation. Thus 17 really provides the entry for a

series of conditional inquiries to determine whether sporulation is possible. If so, a +1 is recorded as a result of the transformation. The figure below indicates this sequence as it is incorporated into the program.

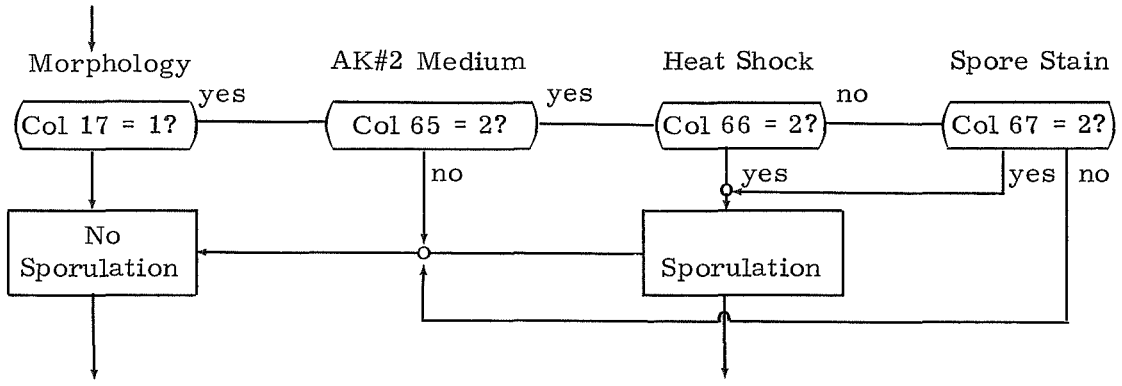


Figure 7. PHS Sequence for Detecting Sporulation

It should be noted that although the logic, illustrated by the above diagram, is incorporated into the program this could have been taken care of a number of other ways. For example, there could have been a separate column for sporulation. The same series of tests could have been run by the Public Health Service and recorded as + or -, in which case this particular bit of programming to fit the actual sequence of operation might have been avoided. However, there is no real problem and it can be regarded as part of the transformation of the data into a number which indicates whether sporulation was observed or not. It is, nevertheless, a place where the program departs from complete generality in simply comparing transformed data with the appropriate matrices, since the logic illustrated above is incorporated as part of the program.

There is one other area in which the program departs from the generality that has been indicated thus far. This also involves the entry in column 17. Column 17 is used to record the occurrence of yeasts, molds, actinomycetes, and streptomycetes which are observed during the morphological examination. Thus, in the event that one of the items 3 through 6 appears in column 17, an immediate identification of the appropriate category is made rather than proceeding through the matching of matrices with data which would not have been taken inasmuch as the sample had already been identified.

Having set up the above procedure to simulate the identification process as carried out by the Public Health Service, it would seem a simple matter to identify a sample. One simply takes the data, transforms them appropriately, makes the comparison until a match is found in the matrix representing the flow diagram, and then proceeds to the appropriate scheme where he selects a new set of data, transforms that, and obtains a match with the right column in that scheme to make an identification. Unfortunately, biological systems are highly variable and it is optimistic to expect exact agreement with the average of a collection of preceding experience for a given sample. As a result, rules have to be developed which would allow for this variation which is usually expressed as a number of disagreements in the above comparisons. The rules adopted were those that have been developed by the Public Health Service Spacecraft Bioassay Laboratory at Cape Kennedy in performing the Apollo sampling and identification. Before the rules are examined, it should be noted that a disagreement between the transformed data and a scheme may happen in a number of ways. First, it may happen that the transformed data disagree with any column in the matrices in <u>one</u> place or test. This can also happen for two, three or more distinct columns, indicating that a number of microorganism categories are equally close (only one disagreeing test) to the "category" represented by the test data. Similarly it may turn out that the test data disagree in two test places with one or more microorganism categories. The rules finally adopted by the Public Health Service were that an identification was made when there was an exact match, or there were not more than three single disagreements (with no exact match) — that is, there was a choice between three microorganism categories which disagreed at one place with the test results, or there was a single microorganism category which disagreed in two places and none that disagreed in fewer places. These possibilities are summarized in Table 2.

No. of Test Disagreements	No. of Categories	Computer Statement of RESULT
0		EXACT IDENT
1	1	ONE DISAGREEMENT
1	2	2 SINGLE DISAGREEMENTS
1	3	3 SINGLE DISAGREEMENTS
2	1	TWO DISAGREEMENTS
Other		NO ID POSSIBLE

Table 2. Computer Statement for Possibilities of Classification

This, of course, allowed for the possibility of multiple identifications, that is, as many as three equally likely microorganism categories. Thus there is the final problem of choice between these categories.

In the event that more than one equally likely microorganism category is arrived at by the computer, there are a number of options for choosing between these multiple identifications. One such option would be to use the statistics accumulated during the identification of some 7,049 samples by the Public Health Service and to select among the indicated possibilities in the same manner that the Public Health Service has previously used. In certain instances it develops that there is a natural choice between the alternatives based upon an analysis of the tests and the associated scheme. There are, of course, other methods of making a decision between the selected categories. However, the two mentioned above are those used in this particular system. At present the program simply prints out all possibilities in a manner that will be illustrated in the next section. The following section will discuss in detail the above two proposals for making a decision.

In this section, then, the PHS scheme was analyzed for incorporation into the abstract approach previously discussed. In addition, the rules and conventions adopted by the PHS and to be used in computerized identifications were defined and discussed.

The Program Itself - BUGID

In the preceding two sections, an abstract approach to the problem of simulating the identification process has been outlined and means described for applying it to the problem of identifying organisms taken from Apollo missions. It was seen that the Public Health Service Identification system lends itself very well to the abstract approach. It is largely composed of keys which can be represented by matrices of a standard form. Furthermore, in the process of using their system, the Public Health Service has developed conventions which are easily implemented in a program. Most of the conventions that are to be used have been discussed. The system that has been programmed and checked out with the Apollo 13 data is called BUGID for Bug Identification.

Initially, it was hoped that a program could be written which was entirely independent of the particular identification scheme to which it was applied. In practice this turned out to be not quite possible although it was very nearly achieved. There are a few instances (as will be seen) where BUGID is dependent upon the Public Health Service system. These instances involve, first, the possibility of an immediate identification, e.g. yeasts, molds, actinomycetes, and streptomycetes, where the program looks for the right digit in the appropriate column before proceeding. The other major instance where scheme instructions are not read into the program occurs in a routine which simulates the Public Health Service search for sporulation.

The heart of BUGID consists of two sub-routines called SETUP and MATCH. SETUP is the routine that selects the appropriate data entries and transforms them for comparison with the identification matrix being used. MATCH (which normally immediately follows SETUP) is the sub-routine which takes the vector consisting of the data transformed by SETUP (IVEC) and compares it with the matrix in the hopes of obtaining a suitably close match. Here, suitably close match means one that satisfies the conditions of Table 2.

The flow diagram of BUGID is presented as Figure 8. Figure 9 has the flow diagram of the sub-routine SETUP, while Figure 10 contains that of MATCH.

In Figure 8 the two exceptional parts of the main program BUGID are indicated. Before the program is run through, a check for immediate identification is made as indicated. The numbers in this portion of the program flow chart are actually part of the program and would have to be modified if any change in the procedures were made. Although this modification is trivial, it is a place where parameters of the identification scheme are not supplied as generalized inputs. Similarly, the section of the program indicated by sporulation is peculiar to the program, and the program would have to be modified internally if there were any change in these procedures. Neglecting these exceptional portions, one can see from Figure 8 that the program consists of repeatedly calling SETUP and then MATCH. As can be imagined, SETUP simply selects the appropriate data and applies the transformation to the data that the operator supplied to the program as input. MATCH simply compares the transformed vector with the columns of the correct matrix in order to determine if there is a sufficient agreement in order to make an identification. Again, this matrix is part of the input to the program. The manner in which the transformations and matrices are

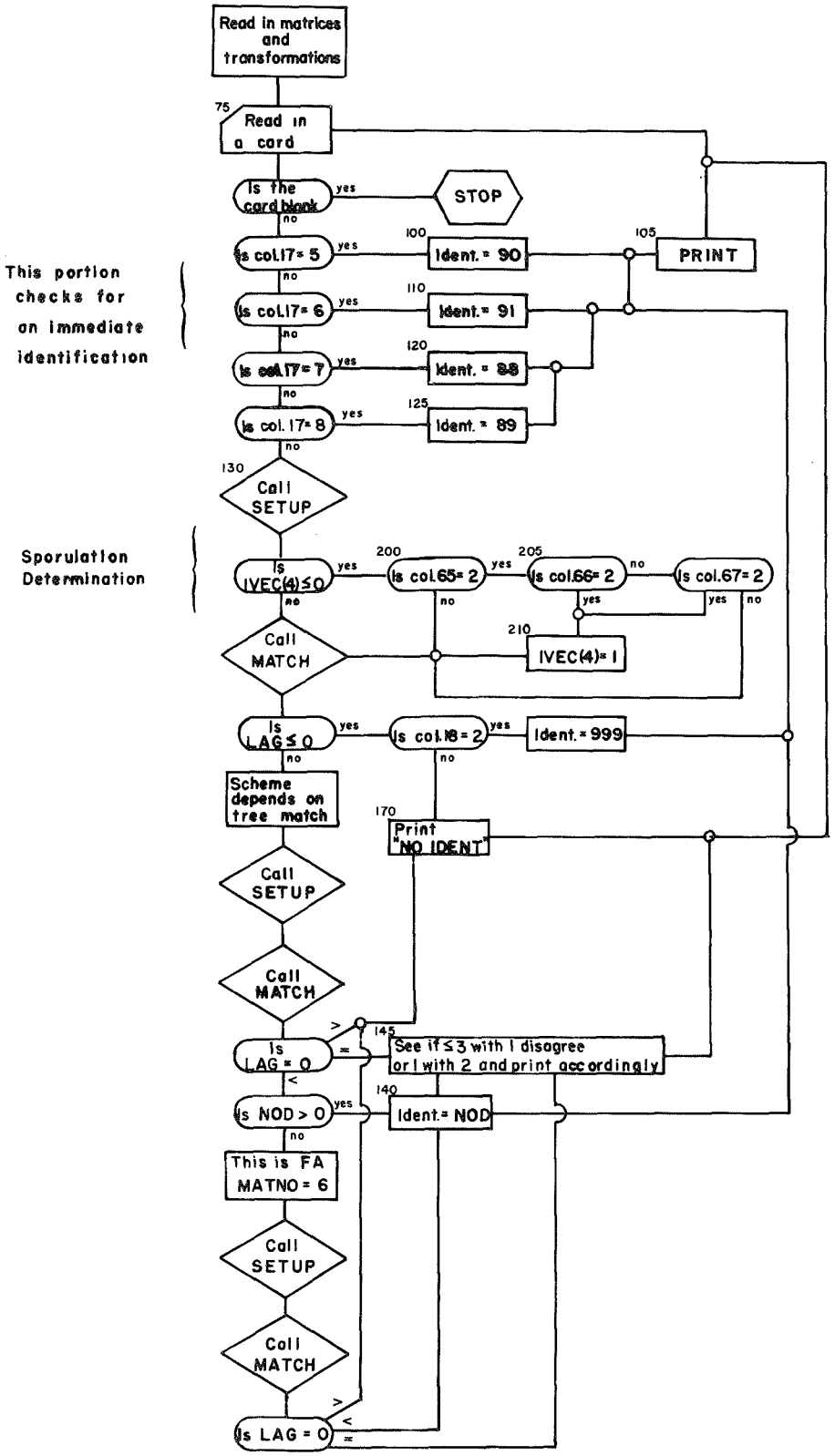


Figure 8. Flow Diagram - BUGID, Main Program

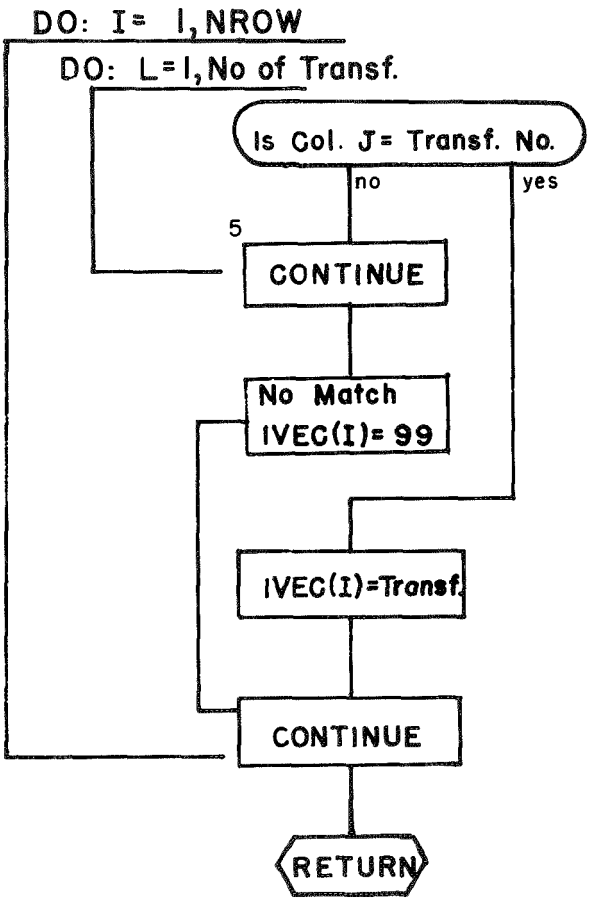


Figure 9. Flow Diagram - Sub-routine SETUP

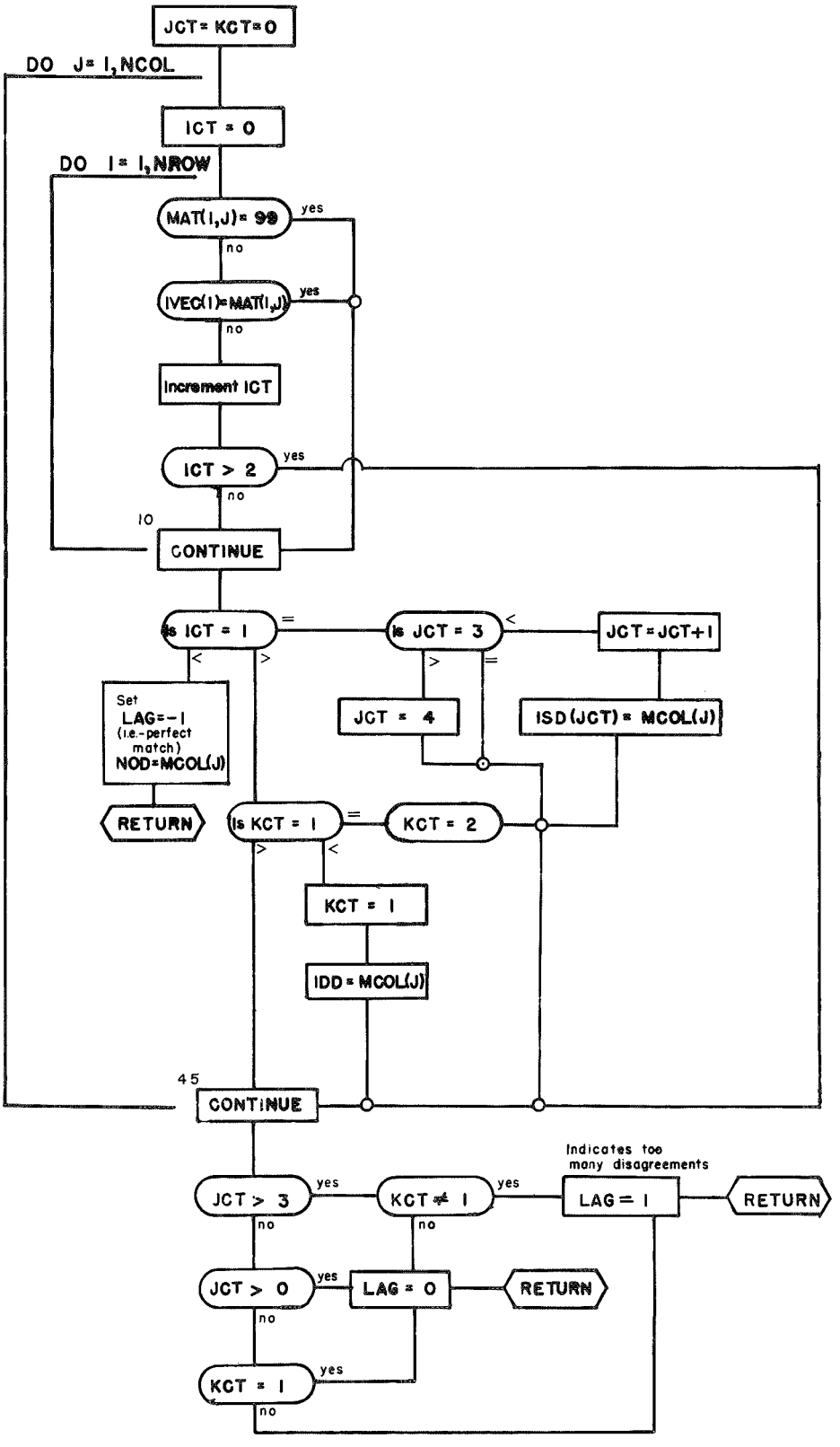


Figure 10. Flow Diagram - Sub-routine MATCH

fed into the program will be covered later. In Figure 8 IVEC is the vector obtained after SETUP has applied the transformations to the raw data. LAG is a flag or index used by the program to determine as it proceeds whether an identification is possible. The program is so arranged that LAG = -1 if an exact identification has been determined at some point, LAG = 0 if there are a number of single disagreements or exactly one double disagreement occurring during the matching process. Finally, LAG = +1 whenever the computer determines that the above condition (the conventions of Table 2) cannot be satisfied, in which case no identification is possible. NOD is an index which has been set to record the column of the Matrix in which sufficient agreement has been obtained with IVEC.

Before a discussion of the fundamental sub-routines, SETUP and MATCH, an idea of the format of the matrices and transformations provided as input to the program will be required. The listing of these for Scheme A is presented as Table 3.

```
   10   14                                                       NR, NC   MA
14 37 37 42 43 53 36 38 39 40
 1  2  3  4  5  6  7  8  9 10 11 12 13 14                        COLS     MA
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1  1                        ROW1     MA
 1  1  1  1  1  1  1  1  1  1  1  1 99 99                        ROW2     MA
 1  1  1  1  1  1 -1 -1 -1 -1 -1 -1 -1 -1                        ROW3     MA
 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1                        ROW4     MA
 1  1  1 -1 -1 -1 -1 -1 -1 -1 -1  1 -1 -1                        ROW5     MA
 1  1 -1  1  1  1  1  1  1  1 -1 -1 -1 -1                        ROW6     MA
-1 -1 -1 -1 -1 -1 -1 -1 -1  1 99  1 -1 -1                        ROW7     MA
 1  1 99 -1  1 99 -1  1 99  1  1  1 -1 -1                        ROW8     MA
 1  1 -1 99  1 99 99  1  1  1  1  1 -1 99                        ROW9     MA
 1 -1 -1 -1 -1  1 -1 -1  1  1  1  1 -1 -1                        ROW 10   MA
   9  6  1  1 -1  2 -1  3 -1  4 -1  5 -1  7 -1  8 -1  9 -1       TR1      MA
   6  3  1  0 -1  1 -1  2 -1  4  1  5 -1                         TR2      MA
   6  4  1  0 -1  1 -1  2 -1  3 -1  5 -1                         TR3      MA
   2  2  1  1 -1                                                 TR4      MA
   2  2  1  1 -1                                                 TR5      MA
   2  2  1  1 -1                                                 TR6      MA
   5  3  1  4  1  1 -1  2 -1  5 -1                               TR7      MA
   5  3  1  4  1  1 -1  2 -1  5 -1                               TR8      MA
   5  3  1  4  1  1 -1  2 -1  5 -1                               TR9      MA
   5  3  1  4  1  1 -1  2 -1  5 -1                               TR10     MA
```

Table 3. Identification Input Data for Scheme A

For each matrix or associated scheme, the first line contains two digits providing the number of rows and the number of columns as labeled on the right hand side. The number of the column in the data card from which the data are required by the corresponding row of the matrix is listed in order on the next line. For example, the fourth row for the matrix of Scheme A requires the data from column 42 as punched by the PHS. The next row labels the columns of the matrix. These numbers are, in general, the identification number of the microorganism category used by the Public Health Service. For example, column 9 is coded 009 Micrococcus Subgroup 3. The transformations to which the raw data are subjected are listed immediately below the matrix. One transformation is given for each entry for the corresponding row of the matrix. Thus, row 10 of the matrix for Scheme A uses the data from column 40 on the card, which will be the result of the Baird-Parker mannitol test. Transformation 10 (found in the right hand column) indicates by the first number in the row that there are five possible outcomes that can be recorded. The subsequent integers taken pair-wise indicate that 3 is a positive result and should be regarded as +1. Similarly, 4 is positive, while the recorded result of 1, 2 or 5 would be negative and transformed to a -1. It should be noted that the number of transformations must equal the number of rows of the associated matrix.

SETUP is a sub-routine which performs the transformation on the data required for identification in a particular matrix. It reads the appropriate columns on the data card, applies the indicated transformations to this data, and obtains a resulting IVEC for the sample. This IVEC will consist of a series of numbers, one for each row of the matrix. The numbers will either be +1, -1, or 99. From an inspection of Figure 9 it can be seen that if an entry in a specific column is not an admissible one, that is, not one listed in the transformation for that column, a 99 will be inserted in IVEC. Thus a missing or incorrect test will be recorded as 99. As previously mentioned, a 99 in IVEC is treated differently than a 99 in the matrix where it represents an indeterminate or variable test rather than incomplete or erroneous information as it does in IVEC. This point will be illustrated by example later.

The other major sub-routine of the identification program, MATCH, is diagrammed in Figure 10. Essentially the function of MATCH is to compare IVEC as prepared by SETUP with the columns of the associated matrix. Hopefully a match

with one of the columns is obtained or at least sufficiently few disagreements (subject to the conventions of Table 2) are obtained so that an identification can be made.

As MATCH is called, three flags, ICT, JCT, and KCT, are set equal to zero. ICT is used during the comparison of IVEC with the column of the matrix to see whether an identification with the column is possible. JCT is used to keep track of the score of IVEC against each column with respect to single disagreements. KCT is similarly used to score the comparison for double disagreements. No index is required to keep track of exact agreement with a column since, in a well designed key, this can only occur for a unique column. Having set the flags, the first part of MATCH compares IVEC with the columns of the matrix MAT then being considered by running down the rows of MAT. As can be seen, if the entry in MAT under consideration is 99, this counts as an agreement. Then it simply continues to the next comparison. If, on the other hand, it is not 99, the corresponding entry in IVEC must equal the entry in MAT for agreement. If not, ICT is incremented. After being incremented, ICT is checked to see whether it is worthwhile proceeding with the remainder of the column. Recalling the columns which are candidates for possible identification must have two or fewer disagreements, a candidate is rejected whenever ICT is greater than two. If so, the sub-routine proceeds to the next column as a candidate. If not, it continues until all rows have been examined or the column is ruled out as a potential identification. Previously it was mentioned that 99s in a matrix were treated differently than 99s in the transformed data IVEC. From Figure 10 and the above it can be seen that, during the comparison process, if a 99 occurs in the matrix it does not count as a disagreement. On the other hand, if the matrix entry under consideration is not 99, exact agreement is required. That is, a 99 appearing in IVEC will count as a disagreement unless the corresponding entry in MAT is also 99 in which case it didn't make any difference. Thus the program has the ability to cope with weak or indeterminate tests appearing in a key while, at the same time, discriminating against missing or inadequate laboratory data.

The next portion of MATCH keeps score on the columns of MAT that have been or are being compared with IVEC. This is done by examining the value of the flag ICT as each column comparison is completed. From the above it can be seen that ICT will be equal to zero only if no disagreements occurred with IVEC in the column being compared. If this occurs, an exact identification is made and MATCH is exited.

Since ICT = 1 would indicate a single disagreement of which three possibles are allowed by convention, such a column is a candidate unless there are more than three. Thus, for ICT = 1, the value of JCT is determined relative to three. If it is not more than three, it is incremented and the sub-routine proceeds to the next column. Finally, there is the possibility of a single column having two disagreements. When a column has two disagreements, ICT is greater than one and less than three, namely two. In this case, KCT is examined to see if it is one. If it is already one, there are now two columns having two disagreements and an identification is not possible. On the other hand, if KCT is still zero, it is incremented to one and the sub-routine proceeds through the remainder of the columns.

In the event that an identification has not been made but the sub-routine has compared the transformed IVEC with each column of the matrix MAT, there is a final decision process of whether an identification can be made or not. The manner in which this is done is indicated in the bottom of Figure 10. Essentially, flags JCT and KCT are examined. In the event that JCT is greater than 3 (more than three single disagreements) and JCT is not equal to 1 (no single double disagreement), the flag is set equal to 1, which indicates too many disagreements with no identification possible. The other possibilities are similarly examined, and the sub-routine returns to the main program with the appropriate identification or identifications of the IVEC representing data from a sample.

The printed output of the program consists of the identification matrices that were read in and a tabulation of the identifications that have been made. A sample of this tabulated output is provided in Figure 11.

As can be seen from Figure 11, the basic identification line contains the sample identification, the date and the serial number from the data card, and the organism category (ORGANISM) that the program has assigned the sample to. Finally, the identification that was made by the Public Health Service and recorded on the card is listed for comparison purposes. Above this line are one or two lines which list the vector that was used in the matrix or matrices for the identification; that is, IVEC, as defined above, is printed. Thus, we see for Sample #4, which is listed as D0004, that the IVEC that resulted from the transformed data which was used for comparison in the tree matrix consisted of [1, 1, 1, -1, 1] as shown in entry D0004.

ID	DATE SN	ORGANISM	PHS
VECTOR	1 -1 1 -1 1		
X0001	032770 2	37	-0
VECTOR	1 1 1 -1 1		
VECTOR	-1 1 1 -1 -1 -1 -1 -1 -1 1		
X0002	032770 2	6	-0
VECTOR	1 1 1 -1 -1		
VECTOR	-1 1 -1 -1 -1 -1 -1 1 1 -1		
X0003	032770 1	8 11	-0
VECTOR	1 1 1 -1 1		
VECTOR	-1 1 -1 -1 -1 -1 -1 -1 1 -1		
X0004	032770 1	7 13 14	-0
VECTOR	1 1 1 -1 1		
VECTOR	1 -1 -1 -1 1 1 -1 -1 1 -1		
X0005	032770 1	14	-0
VECTOR	1 1 -1 -1 1		
VECTOR	99 99 99 99 99 99 99 99		
X0006	032770 1	NO ID. B	-0
VECTOR	1 -1 1 -1 1		
D0002	032770 1	37	37
VECTOR	1 -1 1 -1 1		
D0003	032770 1	37	37
VECTOR	1 1 1 -1 1		
VECTOR	-1 1 1 -1 1 1 -1 1 1 -1		
D0004	032770 1	2	2
VECTOR	1 -1 -1 -1 1		
D0005	032770 1	37	37

Figure 11. BUGID Sample Output

This leads to Scheme A, and the IVEC used for comparison with the matrix for Scheme A reads [-1, 1, 1, -1, 1, 1, -1, 1, 1, -1]·. Finally, the identification of organism Category 2 was made, which agrees with that of the Public Health Service. As mentioned above, it was desirable to retain information about the level at which an identification was made (that is, the number of disagreements). This was done by printing in the first column under ORGANISM the identifications that were made by an exact identification. Three further columns were reserved for those identifications made with one disagreement; this identification is one of three which were possible. Finally, there is a column between these last three columns and the column in which the Public Health Service identification is printed that is reserved for an identification by means of one double disagreement. In the case that no identification was possible, a NO ID is printed in the ORGANISM column. It is convenient for checkout purposes to analyze why no identification was possible. As a result,

the computer also prints the letter designating the last matrix that it was using when it concluded that there was no possibility of an identification. These possibilities can be seen by an examination of Figure 11.

The above has explained the general operation of the program with a detailed description of the flow through the basic program. Similarly, the major sub-routines, SETUP and MATCH, have been described. A complete list of the program with the matrices and transformations presently used as input will be found in Appendix 2. To further illustrate the operation of the program, a sample data card will be used so that the steps taken by the program to assign an identification can be followed in detail.

For this illustration, Sample No. 4 of the Apollo 13 data will be used. A listing of this data card is given below as Figure 12 with a heading which makes it convenient to read the columns.

```
         10        20        30        40        50        60        70        80
             A TTT                   AAAAATAA          A          (TTT)
12345678901234567890123456789012345678901234567890123456789012345678901234567890
D00040327701182122                   14331212          2                    0021
```

Figure 12. Listing of Sample No. 4, Serial D0004

Referring to Figure 8, it is seen that the first operations after reading in the program are to see if the card is blank, and then, if it is not, to see if column 17 has a 5, 6, 7 or 8. If column 17 has such an entry, an immediate identification is possible. As can be seen, the entry in column 17 for Sample No. 4 is 2, so the program then calls SETUP in order to construct IVEC for comparison with the tree matrix. Here, SETUP will use the transformation associated with the tree matrix. This matrix and the transformations are reproduced as Figure 13, and the columns on the data card used are designated "T" above.

```
   5   12                                                    NR, NC    TREE
18 17 41 17 16                                               ROWS      TREE
 1  2  3  4  5  6  7  8  9 10 11 12                          COLS      TREE
 1  1  1 -1 -1  1  1 -1 -1  1  1  1                          ROW 1     TREE
 1  1 -1 -1 -1 -1 -1 -1 -1  1  1 -1                          ROW 2     TREE
 1 -1 99 99 99 99 99 99 99  1 -1 99                          ROW 3     TREE
99 99  1 99 99 -1 -1 99 99 99 99  1                          ROW 4     TREE
 1  1  1 -1 -1  1 -1  1  1 -1 -1 -1                          ROW 5     TREE
   2  2  1  1 -1                                            TR1       TREE
   4  2  1  1 -1  3 -1  4 -1                                TR2       TREE
   2  2  1  1 -1                                            TR3       TREE
   4  3  1  1 -1  2 -1  4 -1                                TR4       TREE
   2  1  1  2 -1                                            TR5       TREE
```

Figure 13. Input Identification Data for General Scheme or Tree

As previously described, the second row of Figure 13 lists in order the data card columns (or tests) associated with the rows of the matrix and transformations. As can be seen, in this case, there are five rows of transformations and columns listed, although one column (17) is listed twice. SETUP proceeds to construct IVEC as follows: First, column 18 of the data in Figure 12 is selected. The entry is a 2; the transformation (TR 1 of Figure 13) indicates there are two possible entries in this column, namely 2 and 1, where 2 is to be changed to a 1 and a 1 is regarded as a negative result and, hence, is changed to a -1. Thus SETUP replaces the 2 in column 18 by 1 in IVEC. SETUP then proceeds, similarly, with column 17. The entry in column 17 is a 2 which, in this case, is replaced by a +1. By continuing through all five cases, the results indicated in Table 4 are obtained.

(i)	Col. No.	Data Entry	IVEC (i)
1	18	2	1
2	17	2	1
3	41	2	1
4	17	2	-1*
5	16	1	1

*Note: Although 2 in column 17 was a positive result for TR 2, it was a negative result in TR 4.

Table 4. Result of SETUP Applied with Tree Transformations to Data of Figure 9

After the program has constructed IVEC by means of SETUP for the tree, (see Figure 8) it queries "Is IVEC(4) ≤ 0?". In this instance, it is, and hence column 65 is examined to see if it is 2. From Figure 12, where columns 65, 66 and 67 are shown (TTT), one can see that it is not, and hence the program proceeds to call MATCH.

Since MATCH simply searches for a sufficiently close match between the columns of a matrix and IVEC, it is convenient to illustrate the situation by Figure 14.

	A	B	C	D1	D2	D3	D4	E	F	G	H	I	IVEC
	1	2	3	4	5	6	7	8	9	10	11	12	
18	1	1	1	-1	-1	1	1	-1	-1	1	1	1	1
17	1	1	-1	1	-1	-1	-1	1	-1	1	1	-1	1
41	1	-1	99	99	99	99	99	99	99	1	-1	99	1
17	99	99	1	99	99	-1	-1	99	99	99	99	1	-1
16	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	1

Figure 14. Tree Matrix and IVEC for Sample No. 4

MATCH first sets the flags, JCT, KCT and ICT, all equal to 0 and proceeds to make a comparison down the columns with the matrix MAT. Proceeding with the first column in Figure 14, it will ask in the first row if the entry MAT (1, 1) = 99. If not, as in this case, it asks "Is IVEC(1) = MAT (1, 1)?". In this case, it continues since 1 = 1. It will do so throughout the entire first column, the only difference being the 99 appearing in row 4 which does not count as a disagreement. Completing the first column, MATCH inquires whether ICT has been incremented. Since ICT has not been and is still 0 < 1, it will set LAG = -1 and return to the main program having determined the next step. When LAG was set equal to -1, NOD was also set equal to MCOL(1), the column in which the perfect match was obtained. Thus, it is known which scheme matrix is used for the next iteration, in this case the first or A.

Since, in the main program, LAG is < 0, the program proceeds to decide on the scheme based on the match with the tree vector. It now considers Scheme A whose matrix and transformations are reproduced as Figure 15. It now calls SETUP to prepare IVEC as before but now for Scheme A. Again, the columns on the data card of interest for Scheme A are listed in the second row of Figure 15. These are also

```
  10   14                                                          NR, NC   MA
14 37 37 42 43 53 36 38 39 40
 1  2  3  4  5  6  7  8  9 10 11 12 13 14                          COLS     MA
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1  1                         ROW1     MA
 1  1  1  1  1  1  1  1  1  1  1  1 99 99                         ROW2     MA
 1  1  1  1  1  1 -1 -1 -1 -1 -1 -1 -1 -1                         ROW3     MA
 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1                         ROW4     MA
 1  1  1 -1 -1 -1 -1 -1 -1 -1 -1  1 -1 -1                         ROW5     MA
 1  1 -1  1  1  1  1  1  1  1 -1 -1 -1 -1                         ROW6     MA
-1 -1 -1 -1 -1 -1 -1 -1 -1  1 99  1 -1 -1                         ROW7     MA
 1  1 99 -1  1 99 -1  1 99  1  1  1 -1 -1                         ROW8     MA
 1  1 -1 99  1 99 99  1  1  1  1  1 -1 99                         ROW9     MA
 1 -1 -1 -1 -1  1 -1 -1  1  1  1  1 -1 -1                         ROW10    MA
   9  6  1  1 -1  2 -1  3 -1  4 -1  5 -1  7 -1  8 -1  9 -1        TR1      MA
   6  3  1  0 -1  1 -1  2 -1  4  1  5 -1                          TR2      MA
   6  4  1  0 -1  1 -1  2 -1  3 -1  5 -1                          TR3      MA
   2  2  1  1 -1                                                  TR4      MA
   2  2  1  1 -1                                                  TR5      MA
   2  2  1  1 -1                                                  TR6      MA
   5  3  1  4  1  1 -1  2 -1  5 -1                               TR7      MA
   5  3  1  4  1  1 -1  2 -1  5 -1                               TR8      MA
   5  3  1  4  1  1 -1  2 -1  5 -1                               TR9      MA
   5  3  1  4  1  1 -1  2 -1  5 -1                               TR10     MA
```

Figure 15. Input Identification Data for Scheme A

shown as "A" columns in Figure 2. These tests are labeled as rows for matrix A. Exactly as in the case of the tree, these columns are read, transformed by the above transformation and listed as the appropriate component of IVEC.

(i)	Col. No.	Data Entry	IVEC (i)
1	14	8	-1
2	37	4	1
3	37	4	1*
4	42	1	-1
5	43	2	1
6	53	2	1
7	36	1	-1
8	38	3	1
9	39	3	1
10	40	1	-1

*Note: Test 37 is used two times in this scheme.

Table 5. Result of SETUP Applied with Scheme A Transformations to Data of Figure 9

Having formed IVEC for Scheme A, MATCH returns to the main program which calls for MATCH for Scheme A.

Identification Matrix A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	IVEC
14	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1
37	1	1	1	1	1	1	1	1	1	1	1	1	99	99	1
37	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1
42	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
43	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1
53	1	1	-1	1	1	1	1	1	1	1	-1	-1	-1	-1	1
36	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	99	1	-1	-1	-1
38	1	1	99	-1	1	99	-1	1	99	1	1	1	-1	-1	1
39	1	1	-1	99	1	99	99	1	1	1	1	1	-1	99	1
40	1	-1	-1	-1	-1	1	-1	-1	1	1	1	1	-1	-1	-1

Figure 16. Scheme A Matrix and IVEC for Sample No. 4

As before, it is convenient to illustrate the situation by means of a figure, Figure 16. MATCH then commences to compare IVEC with the entries in MAT column by column by running down the rows. As can be see by an inspection of Figure 16, IVEC(3) $\neq$ MAT(1, 3) and hence ICT will be incremented at this point. Again, ICT will be incremented when comparing IVEC(9) with MAT(1, 9). This is a possible candidate for a single double disagreement. As MATCH completes the first column, ICT = 2. Referring again to Figure 16, it can be seen that there will be no disagreements during the comparison with the second column. Thus ICT for this column will be $0 < 1$, and again MATCH will assign LAG = -1 [for an exact agreement, NOD = MCOL(2)] and return to the main program.

In the main program, since LAG < 0 at this point, the program proceeds to inquire whether NOD > 0, and since NOD = $2 > 0$, the program prints the identification of NOD which equals 2.

For comparison, the output of BUGID for Sample No. 4 is listed below as Figure 17.

```
                                                        ORGANISM    PHS
VECTOR   1  1  1 -1  1
VECTOR  -1  1  1 -1  1  1 -1  1  1 -1
D0004    032770  1                                        2          2
```

Figure 17. BUGID Output for Sample No. 4

The vectors from the computer output can be compared with the corresponding IVECs determined above. As can be seen, both the computer and the Public Health Service identified this sample as belonging to microorganism category 2. The computer's identification was based on an exact match as would be expected from the above.

Lunar Information System

As mentioned in the Introduction, the interest in a computerized approach to bacterial identification arose through the Apollo sampling program. Such a possibility was envisioned (3) during the original design of the Lunar Information System. Feasibility had been established using the existing Lunar Information System and operating the identification system with the results of the sub-routine, QUALSUM, in the information system. After the feasibility had been demonstrated with this approach, it was hoped that a program could be written which would fit in the Lunar Information System, perhaps as originally conceived. The result of this consideration was the approach of the present report and the associated program which could be inserted into that portion of the Lunar Information System known as QUAL.

Between Apollo 12 and Apollo 13 there were many changes to the procedures used by the Public Health Service in identifying microorganisms. These changes were based upon previous experience in the program which indicated the types of organisms likely to be found. As a result of this experience, many new categories were added under Aerobic Gram Negative rods, Scheme F, and a new identification scheme written. In addition, since no Clostridium species had ever been found, Scheme I (the anaerobic, spore forming, gram positive rods) was condensed to "Clostridium species". These and the other changes necessitated revisions in the reporting format. The form included in Appendix 1 is the one presently being used. These changes greatly influenced the design of the present program in motivating the retention of sufficient flexibility to be able to incorporate any such future changes.

In addition to not operating from the Lunar Information System program QUALSUM, which requires an identification to be made before its utilization, the present program has numerous other advantages. One of these is the above mentioned capability of easy change; aside from providing a means for keeping up with

the evolving programs such as the Lunar Information System, this capability has many other uses. One of the minor modifications, the keys, can be incorporated with ease in order to ascertain their relative effectiveness. Also, since indeterminate tests and test results are being treated somewhat differently in the program, there is the option of modifying either of these at will in an effort to obtain a more efficient or accurate system. The program itself is sufficiently general that it also has application to entirely different systems. It could, for example, be modified with little effort to use a system developed specifically for other applications such as clinical microbiology.

Inasmuch as the feasibility program ran off QUALSUM, it was very dependent upon the Schemes A through I since QUALSUM had already sorted the data into these schemes based upon the PHS assignment of identification categories. In fact, it searched QUALSUM for a scheme and then performed the identification based upon the fact that the Public Health Service had identified the sample as belonging to a category in Scheme A. The present program, on the other hand, takes the raw data and makes the identification. In the case of the data for Apollo 13, the Public Health Service had made an identification, so it was possible to make a comparison between the computer and the Public Health Service results. It is of interest to compare this with the results obtained during the feasibility study. Table 6 presents such a statistical summary. Here, when multiple identifications were made by the computer, agreement was found if the identification of the PHS was included among them.

From the inspection of Table 6, it is apparent that there has been much improvement between Apollo 11 and Apollo 13. It should be noted that a higher percentage of the Apollo 13 identifications are made with no disagreements and there is much less choosing between a number of equally likely categories. In addition, there has been an improvement in the performance of the computerized identification. With Apollo 11 there was disagreement in 187 cases, or 12.7 percent. With Apollo 13, for the Scheme A data, there was agreement in 735 out of 793 samples identified by the Public Health Service, indicating disagreement in 58 cases, or 7.4 percent.

One point of interest is the large fraction of those samples which were identified as belonging to category No. 7 (Micrococcus sub-group 1) by means of three single disagreements, one each between 7, 13, and 14. It is further of interest that there are no instances of two single disagreements. From the distribution

illustrated in Table 6, the possible existence of a distinct category of Micrococcus might be postulated. It is of interest that test results that give rise to this occurence of 7, 13, 14 in the Baird-Parker scheme happen in a unique manner, and this is discussed in some detail later.

Apollo 11 Data - Scheme A
Organism Category

No. of: Disagree	Categ.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	996	Total	
0		12	261	4	59	78	77	41	11	14		13		11	0		581	
1	1	7	0	0	157	128	40	16	2	1		1		49	1		412	
1	2	3	82	0	18	2	30	13	67	18		9		5	0		247	
1	3	5	1	2	3	0	1	108	0	0		0		0	0		120	
2	1	0	0	0	2	0	0	2	2	1		2		2	0		11	
No ID		0	3	0	2	1	5	26	19	2		0		27	0		85	
Not A		1	8	0	1	6	1	1	2	0		1		0	0		21	
Total		28	355	6	242	215	154	207	103	36		26		104	1		1477	
Disagree		4	1	1	23	5	3	20	1	10		2		11	0		81	5.5%
No ID		0	3	0	2	1	5	26	19	2		0		27	0		85	5.7%
Not Scheme A		1	8	0	1	6	1	1	2	0		1		0	0		21	1.4%
Total		5	12	1	26	12	9	47	22	12		3		38	0		187	12.7%

Apollo 13 Data - Scheme A
Organism Category

No. of: Disagree	Categ.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	996	Total
0		0	142	0	8	17	21	31	20	45	0	3	0	116	0		403
1	1	0	0	0	20	17	24	12	2	2	0	0	0	58	0		135
1	2	0	7	1	17	0	2	0	50	12	0	0	0	21	0		110
1	3	0	0	1	3	0	0	71*	0	8	0	0	0	0	0		83
2	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0		1
																2	2
No ID A																	0
Total		0	149	2	48	34	47	114	73	67	0	3	0	195	0	2	734

*Note: Of these 71, 70 were 7, 13, 14 combinations.

Table 6. Summary of Computerized Identification of Apollo 11 and 13 Data

The incorporation of the present program in QUAL as part of the Lunar Information System would provide a number of advantages to the program. One of these

advantages would be the uniformity that is inherent in a computer operation. For example, every sample would be treated in exactly the same manner. Data would be interpreted in a repeatable manner each time. In the event of an inquiry, the process whereby an identification was made can be recovered and gone over step by step for review. There is a price paid for this, of course, in that the resolution of marginal cases by microbiological judgment is minimized. This occurs in the present largely in the case of equally likely categories determined by the computer which has to determine one particular one to record. Fortunately, the Apollo experience of these occurrences follows a pattern which can be simulated by the computer. That is, the computer would simulate previous experience in the Apollo Lunar inventory program. These decisions can be written down and many are straightforward. For example, in the event that a double disagreement with 8 and 11 occurs, 8 is selected. Similarly, between 2 and 4 the choice has always been category 4. Also, in the situation where a single disagreement occurs between category 7, 13, and 14, in every case 7 is selected by the Public Health Service.

Another advantage to having such a program as an integral part of the Lunar Information System would be the reduction of load placed on the microbiologists of the Public Health Service. They could devote themselves to taking the appropriate data, and the computer would then appropriately categorize them. More time could thus be spent developing more efficient schemes or systems for identification which could easily be adopted into the Lunar Information System.

There are, on the other hand, some restrictions inherent in using such a system. For such a program to work, data must be taken and recorded on a card before the program can even get started. Implicit in this is the fact that the appropriate data must be taken. That is, the microbiologist must have some rough idea of the identification in order to perform the appropriate tests. It would appear that rules of thumb could be developed so that the appropriate data would be taken with a minimal amount of effort. In fact, an inspection of the listing of the Apollo data tends to indicate that such rules have already been developed on the basis of experience and are currently being used.

To summarize, the incorporation of the identification program in QUAL would provide the Lunar Information System and Apollo sampling program with a number of advantages. Some of these would include the uniform manner in which identifications

are made along with the ability to reconstruct the data and decisions involved. It would also give great flexibility to the microbiologist in modifying, improving, or adding new identification procedures, and at the same time provide means by which their utility might be assessed. The system could possibly be applied to a variety of similar situations including possible non-Apollo applications. Another advantage might be the possibility of a natural extension of the Lunar Information System to planetary considerations.

Applications

The system that has been discussed was developed for Apollo applications and the program written for inclusion in the Lunar Informations System. Thus, at least one application was in mind during its development. However other applications became apparent as it evolved. As has been mentioned, the program is a useful tool for microbiologists in addition to its utility in the Apollo program. It could be used to determine key efficiency (that is, to assess how efficient tests are in an identification process) in addition to other possible applications. This section will consider some of these alternative uses of this approach to identification.

To illustrate some of these, the results obtained in the Apollo 13 data for the Baird-Parker scheme, or Scheme A, will be used. That part of Table 6 which contains the results is reproduced as Table 7 below.

Apollo 13 Data

No. of: Disagree	Categ.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	996	Total
0		0	142	0	8	17	21	31	20	45	0	3	0	116	0		403
1	1	0	0	0	20	17	24	12	2	2	0	0	0	58	0		135
1	2	0	7	1	17	0	2	0	50	12	0	0	0	21	0		110
1	3	0	0	1	3	0	0	71*	0	8	0	0	0	0	0		83
2	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0		1
																2	2
No ID A																	0
Total		0	149	2	48	34	47	114	73	67	0	3	0	195		2	734

*Note: Of these 71, 70 were 7, 13, 14 combinations

Table 7. Statistical Summary of BUGID Results with Apollo 13

Table 7 compares the results of the computerized identification with those of the Public Health Service for the Baird-Parker isolates on Apollo 13. The nomenclature in the left hand column has already been discussed. It is simply the listing of the ways in which the computer can obtain a result, as was presented in Table 2. In addition to obtaining information as to how well the computer agreed with the Public Health Service, other facts can be gathered from Table 7. For example, although the Public Health Service data give only the number of the identification made, the computer data give the distribution of identifications in terms of how close they were to the exact test sequence hoped for for the organism category. If the microorganisms were well behaved, the categories were well defined and very good tests were available, it might be hoped, in the best of all possible worlds, that, given a microorganism, appropriate tests would be run and they would agree exactly with those specified by the scheme for identifying this class of microorganisms. In reality this does not occur for a number of reasons. It is common for at least one test to differ from that expected for the microorganism at hand. This can be seen in Table 7 by inspecting the row listing the number of times that an organism was identified even though it disagreed in one test with the category in which it was identified. It is seen that there were 135 such occurrences. This particular situation does not unduly complicate the identification process, but when it is extended so that there is one disagreement with each of two or three categories, a choice must be made between the categories. The extent to which this reaching occurs in order to obtain an identification provides an indication of how convincing the identification is. Furthermore, having done this for a large number of samples, the results, as tabulated in Table 7, provide a measure of how well the identification process is performing. Looking at the totals listed in the right hand column of Table 7, it can be seen that most of the identifications were made when there was an exact agreement with the test results, a somewhat fewer number where there was one test neglected, a still fewer number where one test was neglected and a choice had to be made between two categories, and so forth. This would seem to indicate that the Baird-Parker scheme is quite good at identifying the Staphlococcus and Micrococcus into the categories assigned. It would be possible to use large collections of data to make judgments about the efficiencies of schemes and their associated tests by using reasoning similar to that above.

The totals examined above provide a measure of the scheme as a whole. If the body of Table 7 is examined, a few departures from the above pattern are noticed.

One in particular stands out as an anomoly. This occurs in category No. 7 (Micrococcus, sub-group 1) where the distribution of sample identification for the 114 samples in this category begins with 31 exact agreements with the expected test results, and then drops to 12 identifications made with one single disagreement (one test result was not as anticipated). There are no double disagreements. Then there are an unusually large number of 71 samples identified on the basis of three single disagreements. For these last 71, then, a choice had to be made between three alternative categories, each of which had one test result which differed from that sequence of tests provided by the sample. This unusual change in the trend discussed above seemed worthy of further analysis for, if nothing else, it may have been a peculiarity in the program. Upon inspection, it developed that the vast majority of these 71 (in fact 70) with three single disagreements were choices between categories 7, 13, 14. Figure 18 presents the test results expected for these three categories as given in the Baird-Parker scheme which was presented in Figure 1.

Organism category →	7	13	14	Vector	Disagrees
Baird-Parker subgroup →	1	7	8		
Data entry					
14	-	-	+	- ←	w 14
37	+	±	±	+	
37	-	-	-	-	
42	-	-	-	-	
43	-	-	-	-	
53	+	-	-	- ←	w 7
36	-	-	-	-	
38	-	-	-	-	
39	V	-	±	+ ←	w 13
40	-	-	-	-	

Figure 18. Pertinent Excerpts from the Baird-Parker Scheme with Commonly Occurring Vector Assigned as Category 7

It can be seen that a single disagreement with three test sequences as tabulated in Figure 18 can occur in only one way. The test results giving the vector (denoted "Vector") which disagrees in this manner are also tabulated along with an indication of which category disagreement occurs for a given test. It should be noted that the Baird-Parker glucose test is used twice. This test result is recorded in column 37 of the Public Health Service reporting form, and, hence, data are read from that

column twice although a different transformation is applied to the two readings in order to obtain the Baird-Parker results.

The 70 data cards representing these 7, 13, 14 combinations were extracted from the Apollo 13 data and listed. This listing is given as Table 8. In Table 8 the columns used for the tree are headed by a T and, similarly, those used by Baird-Parker Scheme A by an A. As can be seen from an inspection, the only variation in the relevant data is in column 14 which is used to record the pigment coloration. There, 8 (white) or 9 (yellow) is recorded with the single exception of one 5 (orange). Thus, not only is IVEC unique, but the raw data are remarkably uniform for those samples which give rise to the possibility 7, 13, 14. It was possible that the transformations converted the data in some peculiar manner so as to give rise to such a high incidence of samples falling in this particular category having a multiple choice. It appears from Table 8 that this is not the case and that, moreover, some 59.8 percent of the samples identified as belonging to Microorganism category 7 (Baird-Parker Micrococcus sub-group 1) in fact do not quite fit the Baird-Parker classification.

It is of interest that Baird-Parker (10) indicates that although a number of schemes have been proposed for dividing the general Staphlococcus and Micrococcus into species, "groups" or "subgroups", at present none of the classifications is acceptable to all taxonomists. Considering the Micrococcus genus, it is generally agreed there are two well-defined species M. luteus and M. roseus, which correspond to his subgroup 7 and 8 respectively or the Public Health Service categories 13 and 14. Baird-Parker suggests "Undoubtedly other species will be recognized eventually and these will be given specific names when they have been adequately defined." In view of the two kinds of samples occurring in the Apollo data and the high incidence of the one discussed above, it might be worth considering these as separate groups. A worthy investigation would be the development of a test to further distinguish between the two types of Micrococcus sub-group 1.

Thus, some applications, in addition to the immediate Apollo usage, have been indicated. This approach provides a tool for the microbiologists which, in addition to simulating the identification process, can be used to form judgments about the efficiency of a particular scheme or process. It, furthermore, can be used to ascertain how well a classification scheme fits a collection of data.

```
12345678901234567890123456789012345678901234567890123456789012345678901234567890
          A TTT                    AAAAATAA          A          (TTT)
D00110327701182122                 13131211          1                        0071
D00240327702182122                 13131211          1                        0071
D00280327702182122                 13131211          1                        0071
D00300327702182122                 13131211          1                        0071
D00320327702152122                 13131211          1                        0071
D00350327702182122                 13131211          1                        0071
D00720327703182122                 13131211          1                        0071
D00730327703182122                 13131211          1                        0071
D00780327703182122                 13131211          1                        0071
D00820327703182122                 13131211          1                        0071
D00840327703182122                 13131211          1                        0071
D01180327703182122                 13131211          1                        0073
D01480327703182122                 13131211          1                        0074
D01660327701192122                 13131211          1                        0075
D01670327701192122                 13131211          1                        0075
D02000327702192122                 13131211          1                        0076
D02440330702182122                 13131211          1                        0072
D02560330702192122                 13131211          1                        0072
D02750330702182122                 13131211          1                        0072
D02780330702192122                 13131211          1                        0072
D03150330703192122                 13131211          1                        0072
D03280330703182122                 13131211          1                        0072
D03350330703192122                 13131211          1                        0072
D03400330703182122                 13131211          1                        0072
D03710403701192122                 13131211          1                        0071
D03850403703192122                 13131211          1                        0071
D03860403703192122                 13131211          1                        0071
D03870403703192122                 13131211          1                        0071
D03890403703192122                 13131211          1                        0071
D03940403703192122                 13131211          1                        0071
D03980403703192122                 13131211          1                        0071
D04010403703192122                 13131211          1                        0071
D04080403703192122                 13131211          1                        0071
D04110403703192122                 13131211          1                        0071
D04140403703192122                 13131211          1                        0071
D04160403703192122                 13131211          1                        0071
D04170402703192122                 13131211          1                        0071
D04200403703192122                 13131211          1                        0071
D04230403703192122                 13131211          1                        0071
D04250403703192122                 13131211          1                        0071
D04260403703192122                 13131211          1                        0071
D04300403703192122                 13131211          1                        0071
D04310403703192122                 13131211          1                        0071
D04320403703192122                 13131211          1                        0071
D04430403703192122                 13131211          1                        0072
D04860403703152122                 13131211          1                        0072
D04870403703192122                 13131211          1                        0072
D05140403703192122                 13131211          1                        0072
D05300403703192122                 13131211          1                        0072
D05310403703192122                 13131211          1                        0072
D05500403703152122                 13131211          1                        0072
D05830403702182122                 13131211          1                        0073
D06070403702182122                 13131211          1                        0073
D06190403701191122                 13131211          1                        0074
D06270403701192122                 13131211          1                        0074
D07190407701182122                 13131211          1                        0072
D07450407701192122                 13131211          1                        0072
D07540407701192122                 13131211          1                        0072
D07670407701182122                 13131211          1                        0072
D07770407701182122                 13131211          1                        0072
D08380407703192122                 13131211          1                        0072
D08560407701182122                 13131211          1                        0073
D09820410701182122                 13131211          1                        0071
D09920410701192122                 13131211          1                        0071
D10020410701192122                 13131211          1                        0071
D10040410701192122                 13131211          1                        0071
D10100410701192122                 13131211          1                        0071
D10470410703182122                 13131211          1                        0071
D10480410703182122                 13131211          1                        0071
D10500410703182122                 13131211          1                        0071
12345678901234567890123456789012345678901234567890123456789012345678901234567890
          A TTT                    AAAAATAA          A          (TTT)
```

Table 8. Listing of Apollo 13 Data Cards for 7, 13, 14 Combinations Identified as 7

Conclusions

In the process of writing a program to perform identification for the Public Health Service Apollo Program, a novel approach to the simulation of the identification process was developed. In the development of this approach, a system was designed which could easily be used by practicing microbiologists and easily modified without major changes to the system itself. The design was such that the system can incorporate the spectrum of traditional identification techniques ranging from the binary selection process (such as trees) to keys which are typically presented as test-organism matrices, and it can encompass a variety of other methods.

The resulting program is intended to be included as part of the Lunar Information System. It obtains agreement with the Public Health Service on 92 percent of their data and will have a natural and easy extension to similar programs concerned with NASA planetary missions. In addition to relieving the Public Health Service of some of the burden of microbial identification, it provides a tool which can be used to improve this aspect of their tests.

The system can be used to assess existing or proposed microbiological identification procedures and to indicate directions in which they might be improved. It has the property of formalizing many of these procedures so that standard criteria can be established.

In addition to the above largely NASA-oriented applications, it has potential if applied to microbiological research. The above formal "standardized" rules or agreements about the level or criteria for identification could be used as a measure of grouping similar microorganisms. An example of this is Micrococcus Cohn, sub-group I, discussed above.

Not only does this approach have immediate application to NASA problems and demonstrated possibilities in microbiological research, but it may also find a wide variety of other applications in use. Interest has been expressed on the part of the Public Health Service in broader applications within their program.

References

1. NASA Policy Directive, NPD 8020.8A.

2. NASA Policy Directive, NPD 8020.7.

3. Roark, A. L. and Wyer, A. L., An Interactive Computer Information System for Planetary Quarantine for Lunar Programs, SC-RR-68-545, Sandia Laboratories, July 1968.

4. Roark, A. L., and Gavin, W. R., Users Manual for the Planetary Quarantine Lunar Information System, SC-M-70-604, Sandia Laboratories, November 1970.

5. Dillon, R. T., Letter to Dr. Donald G. Fox on Computerized Identification Scheme, December 11, 1969.

6. Puleo, John R., Oxborrow, Gordon S., Fields, Norman D., and Hall, Herbert E., "Quantitative and Qualitative Microbiological Profiles of the Apollo 10 and 11 Spacecraft," Applied Microbiology, September 1970.

7. Thimann, Kenneth V., The Life of Bacteria, The MacMillan Company, New York, 1968.

8. Breed, Robert S., et al, Bergey's Manual of Determinative Bacteriology, The Williams & Wilkins Company, Baltimore, 1957.

9. Identification Schemes for Microorganisms Isolated from Apollo Space-Craft, U.S. Public Health Service, April 3, 1970.

10. Gibbs, B. M. and Skinner, F. A., Identification Methods for Microbiologists, Academic Press, New York, 1966.

Appendix 1

Organism Categories and Data Recording Form

KEY NUMBER OF ORGANISMS

IBM CODES FOR COLUMNS 77 - 79 ON MICROBIAL IDENTIFICATION DATA SHEET

A	Catalase positive, gram positive, aerobic cocci
001	Staphylococcus, Subgroup I
002	Staphylococcus, Subgroup II
003	Staphylococcus, Subgroup III
004	Staphylococcus, Subgroup IV
005	Staphylococcus, Subgroup V
006	Staphylococcus, Subgroup VI
007	Micrococcus, Subgroup 1
008	Micrococcus, Subgroup 2
009	Micrococcus, Subgroup 3
010	Micrococcus, Subgroup 4
011	Micrococcus, Subgroup 5
012	Micrococcus, Subgroup 6
013	Micrococcus, Subgroup 7
014	Micrococcus, Subgroup 8
996	Atypical Micrococcaceae
B	Catalase negative, gram positive, aerobic cocci
015	Enterococcus
016	Lactic
017	Viridans
018	Pyogenic
C	Gram positive, sporeforming, aerobic rods
019	B. alvei
020	B. badius
021	B. brevis
022	B. cereus
023	B. circulans
024	B. coagulans
025	B. firmus
026	B. laterosporus
027	B. lentus
028	B. licheniformis
029	B. macerans
030	B. megaterium
031	B. pantothenicus
032	B. polymyxa
033	B. pulvifaciens
034	B. pumilus
034	B. spaericus
036	B. subtilis
998	Atypical Bacillus spp.

KEY NUMBER OF ORGANISMS (Cont.)

D		Aerobic and anaerobic non sporeforming gram positive rods, anaerobic gram negative cocci, and anaerobic gram negative rods: test reactions only
	037	Aerobic, non sporeforming gram positive rods
	038	Anacrobic, non sporeforming gram positive rods
	039	Aerobic gram negative cocci
	040	Anaerobic gram negative rods
E		Aerobic gram negative cocci
	041	Neisseria spp.
F		Aerobic gram negative rods
	045	Pseudomonas, Group I
	049	Pseudomonas, Group II
	050	Pseudomonas, Group III
	051	Pseudomonas, Group IV
	052	Aeromonas spp.
	053	Bibrio spp.
	044	Flavobacterium spp.
	042	Achromobacter spp.
	043	Alcaligenes spp.
	054	Escherichia
	055	Klebsiella-Enterobacter
	056	Salmonella
	057	Arizona
	058	Citrobacter
	059	Shigella
	060	P. vulgaris
	061	P. morganii
	062	P. mirabilis
	063	P. rettgeri
	064	Providencia
	065	Unidentified gram negative rods
G	046	Gram positive, catalase positive anerobic cocci
H	047	Gram positive, catalase negative anaerobic cocci
I	048	Clostridium spp.
J	088	Actinomycetes
K	089	Streptomycetes
L	090	Yeasts
M	091	Molds
N	999	Lost in process

MICROBIAL IDENTIFICATION

(Circle Code Number that Applies) | AREA:

(1-5) Identification Number

(6-11) Date

(12) Sample Number

(13) Sample Treatment
Aerobic 1
Anaerobic 2
Heat Sh. Aerobic 3
Heat Sh. Anaerobic 4
Blood Agar Aerobic 5
Blood Agar Anaerobic 6
MacConkey 7
Mycophil 8

(14) Pigment
Black 1 Pink 6
Brown 2 Red 7
Green 3 White 8
Grey 4 Yellow 9
Orange 5

(15) Translucent 1
Opaque 2
Soluble pigment 3

(16) Growth characteristics
Aerobic 1
Anaerobic 2

(17) CELLULAR MORPHOLOGY
RODS 1 SPORES 3 YEASTS 5 ACTINOM. 7
COCCI 2 PLEO. 4 MOLDS 6 STREPTOM. 8

(18) Gram Stain
– 1 + 2

(19) P.R. Aesculin
ND 0 A 3
NC 1 AG 4
NG 2 ALK 5

(20) P.R. Arabinose
ND 0 A 3
NC 1 AG 4
NG 2 ALK 5

(21) P.R. Glucose
ND 0 A 3
NC 1 AG 4
NG 2 ALK 5

(22) P.R. Glycerol
ND 0 A 3
NC 1 AG 4
NG 2 ALK 5

(23) MALONATE BROTH
ND 0 + 3
NC 1
NG 2

(24) P.R. Lactose
ND 0 A 3
NC 1 AG 4
NG 2 ALK 5

(25) P.R. Maltose
ND 0 A 3
NC 1 AG 4
NG 2 ALK 5

(26) P.R. Mannitol
ND 0 A 3
NC 1 AG 4
NG 2 ALK 5

(27)

(28) NITRATE (ANAEROBIC)
ND 0
– 1
+ 2

(29) GLUCOSE (ANAEROBIC)
ND 0 GAS 3
NG 1
G 2

(30) P.R. Sucrose
ND 0 A 3
NC 1 AG 4
NG 2 ALK 5

(31) P.R. Xylose
ND 0 A 3
NC 1 AG 4
NG 2 ALK 5

(32) O.F. Glucose
ND 0 AO 3
NC 1 AF 4
NG 2 ALK 5

(33) Gas – 1 + 2

(34) O.F. Maltose
ND 0 AO 3
NC 1 AF 4
NG 2 ALK 5

(35) Gas – 1 + 2

(36) B.P. Arabinose
ND 0 AO 3
NC 1 AF 4
NG 2 ALK 5

(37) B.P. Glucose
ND 0 AO 3
NC 1 AF 4
NG 2 ALK 5

(38) B.P. Lactose
ND 0 AO 3
NC 1 AF 4
NG 2 ALK 5

(39) B.P. Maltose
ND 0 AO 3
NC 1 AF 4
NG 2 ALK 5

(40) B.P. Mannitol
ND 0 AO 3
NC 1 AF 4
NG 2 ALK 5

(41) Catalase
ND 0 – 1 + 2

(42) Coagulase
ND 0 – 1 + 2

(43) Phosphatase
ND 0 – 1 + 2

(44) Hemolysis
ND 0 Beta 2
Alpha 1 Gamma 3

(45) Litmus Milk
ND 0 Acid 2
NC 1 ALK 3

(46) Stormy – 1 + 2

(47) Soft 1 Hard 2 No curd 3

(48) Red. Litmus – 1 + 2

(49) Peptonization – 1 + 2

(50) Gelatin ND 0 – 1 + 2

(51) Starch ND 0 – 1 + 2

(52) Oxidase ND 0 – 1 + 2

(53) V. P. ND 0 – 1 + 2

(54) M. R. ND 0 – 1 + 2

(55) Indol ND 0 – 1 + 2

(56) Citrate ND 0 – 1 + 2

(57) Nitrate ND 0 – 1 + 2

(58) TSI
ND 0 S+ 2
S– 1 ALK 3

(59) B– 1 B+ 2 ALK 3

(60) H2S– 1 H2S+ 2

(61) No gas 1 Gas 2

(62) Optochin ND 0 – 1 + 2

(63) Urease ND 0 – 1 + 2

(64) Motility (SIM)
ND 0 – 1 + 2

(65) AK#2 ND 0 NG 1 G 2

(66) Heat Shock S– 1 S+ 2

(67) H– 1 H+ 2

(68) Peptone Broth
ND 0 Gas 2
NG 1 No Gas 3

(69) 10°C ND 0 NG 1 G 2

(70) 45°C ND 0 NG 1 G 2

(71) 6.5% NaCl
ND 0 NG 1 G 2

(72) 50°C
ND 0 NG 1 G 2

(73) 0.1% M.B. Milk
ND 0 NG 1 G 2

(74) MacConkey
ND 0 P 2
NG 1 G 3

(75) Flagella
ND 0 – 1 + 2

(76) Capsule
ND 0 – 1 + 2

(77-79) Key Number

(80) Portion of Spacecraft Sampled
CMI 1 IU 5
LAI 2 S4B 6
LAE 3 SLA 7
LDE 4

REMARKS

HSM 8.27 (NCDC)
4-69

Appendix 2

Listing of BUGID and Identification Input Cards

```
BUGID,T100,CM100000.                          DICK DILLON   BOX 624
ACCOUNT,E9935,D1741,A0003010,G1700,RT,KUNC.
FUN,S,,,,,,37777.
MAP,ON.
REDUCE.
PRESET,NIND.
LGO.
'
      PROGRAM BUGID(INPUT,OUTPUT)
      DIMENSION IAGR(100),IDOWN(5)
      DIMENSION IAGR(5,100)
      COMMON/A/NROW(6),MROW(6,20),ICARD(80),NTRAN(6,20),MTRAN(6,20,10),
     1 MTRA2(6,20,10),IVEC(20)
      COMMON/B/MAT(6,20,20),NCOL(6),MCOL(6,20),LAG,ISD(3),IDD,NOD,JCT
     1 ,KCT
      DIMENSION IDMT(6)
      DATA IDMT/4HTREE,4HA   ,4HB   ,4HC   ,4HF   ,4HFA  /
      NMT=6
      NRT=20
      NCT=20
      DO 4 I=1,5
      DO 4 J=1,100
 4    IAGR(I,J)=0
C     LEFT INDEX---1 IS EXACT AGREEMENT
C                  2 IS ONE SINGLE DISAGREEMENT
C                  3 IS TWO SINGLE DISAGREEMENTS
C                  4 IS THREE SINGLE DISAGREEMENTS
C                  5 IS ONE DOUBLE DISAGREEMENT
C                  1 IS USED ALSO TO TALLY THE NO IDS.
C     RIGHT INDEX--1-91 ARE REGULAR OUTCOMES
C                  92 IS 996
C                  93 IS 998
C                  94 IS 999
C                  95 IS NO ID IN TREE MATRIX
C                  96 IS NO ID IN MATRIX A
C                  97 IS NO ID IN MATRIX B
C                  98 IS NO ID IN MATRIX C
C                  99 IS NO ID IN MATRIX F
C                  100 IS NO ID IN MATRIX FA
 5    FORMAT(14I5)
C     NM IS NUMBER OF MATRICES TO SET UP.  THE FIRST MATRIX IS FIRST TREE.
      READ 5,NM
      IF(NM .LE. NMT) GO TO 15
 10   FORMAT(*1NUMBER OF MATRICES IS*I3*--*I3* ALLOWED.  CHANGE DIMENSIO
     1NS OF MAT,NCOL,NROW,MCOL, MROW, NTRAN AND MTRAN.  ALSO CHANGE NMT.
     2*)
      PRINT 10,NM,NMT
      STOP
 15   DO 65 IM=1,NM
      READ 5,NR,NC
      IF(NR .LE. NRT) GO TO 25
 20   FORMAT(*1NUMBER OF ROWS FOR MATRIX*I2* IS*I3*--*I3* ALLOWED.  CHAN
     1GE DIMENSIONS OF MAT, IVEC, MROW, NTRAN AND MTRAN.  ALSO CHANGE NR
     2T.*)
```

```
      PRINT 20,IM,NR,NRT
      STOP
 25   IF(NC .LE. NCT) GO TO 35
 30   FORMAT(*1NUMBER OF COLUMNS FOR MATRIX*I2* IS*I3*--*I3* ALLOWED.  C
     1HANGE DIMENSIONS OF MAT AND MCOL.  ALSO CHANGE NCT.*)
      PRINT 30,IM,NC,NCT
      STOP
 35   NROW(IM)=NR
      NCOL(IM)=NC
      READ 40,(MROW(IM,IR),IR=1,NR)
      READ 40,(MCOL(IM,IC),IC=1,NC)
 40   FORMAT(36I3)
      DO 45 IR=1,NR
      READ 40,(MAT(IM,IR,IC),IC=1,NC)
 45   CONTINUE
 50   FORMAT(*1IDENTIFICATION MATRIX *A4//)
      PRINT 50,IDMT(IM)
      IF(IM-1) 52,52,54
 51   FORMAT(10X,1HA,3X,1HB,3X,1HC,2X,2HD1,2X,2HD2,2X,2HD3,2X,2HD4,
     1 3X,1HE,3X,1HF,3X,1HG,3X,1HH,3X,1HI)
 52   PRINT 51
 53   FORMAT(7X,20I4)
 54   PRINT 53,(MCOL(IM,IC),IC=1,NC)
 55   FORMAT(1X,I4,2X,20I4)
      DO 60 IR=1,NR
      PRINT 55,MROW(IM,IR),(MAT(IM,IR,IC),IC=1,NC)
 60   CONTINUE
      DO 62 IR=1,NR
 61   FORMAT(I5,20I3)
      READ 61,NTR,(MTRAN(IM,IR,I),MTRA2(IM,IR,I),I=1,NTR)
      NTRAN(IM,IR)=NTR
 62   CONTINUE
 65   CONTINUE
      IDB=1
      IDE=5
      IDTB=6
      IDTE=11
      ISN=12
      ICM=17
      IDIS=0
      MANY=0
 69   FORMAT(5H1  ID,4X,4HDATE,2X,2HSN,40X,8HORGANISM//)
      PRINT 69
 70   FORMAT(A1,75I1,I3,I1)
 75   READ 70,(ICARD(I),I=1,78)
      MANY=MANY+1
C     A BLANK CARD WILL STOP PROGRAM
      IF(ICARD(2)) 76,76,85
 76   IF(ICARD(3)) 77,77,85
 77   IF(ICARD(4)) 78,78,85
 78   IF(ICARD(5)) 79,79,85
 79   IF(ICARD(6)) 80,80,85
 80   PRINT 81
 81   FORMAT(1H1,4X,5HEXACT,2X,8H1 SINGLE,2X,8H2 SINGLE,2X,8H3 SINGLE,
```

```
     1 2X,8H1 DOUBLE,5X,5HTOTAL)
       DO 88 J=1,100
       ISUM=0
       DO 87 I=1,5
 87    ISUM=ISUM+IAGR(I,J)
 88    IACR(J)=ISUM
       DO 92 I=1,5
       ISUM=0
       DO 91 J=1,100
 91    ISUM=ISUM+IAGR(I,J)
 92    IDOWN(I)=ISUM
 82    FORMAT(1X,I4,I5,5I10)
       DO 83 J=1,91
       PRINT 82,J,(IAGR(I,J),I=1,5),IACR(J)
 83    CONTINUE
       J=996
       PRINT 82,J,(IAGR(I,92),I=1,5),IACR(92)
       J=998
       PRINT 82,J,(IAGR(I,93),I=1,5),IACR(93)
       J=999
       PRINT 82,J,(IAGR(I,94),I=1,5),IACR(94)
 84    FORMAT(1X,A4,I5,5I10)
       DO 86 J=95,100
       JI=J-94
       PRINT 84,IDMT(JI),(IAGR(I,J),I=1,5),IACR(J)
 86    CONTINUE
       ISUM=0
       DO 94 J=1,100
 94    ISUM=ISUM+IACR(J)
       JSUM=0
       DO 96 I=1,5
 96    JSUM=JSUM+IDOWN(I)
       IF(JSUM-ISUM) 98,99,98
 97    FORMAT(32H TOTALS DO NOT AGREE--ACROSS IS ,I5,9H  DOWN IS,I5)
 98    PRINT 97,ISUM,JSUM
       GO TO 102
 93    FORMAT(5X,I5,5I10)
 99    PRINT 93,(IDOWN(I),I=1,5),ISUM
 101   FORMAT(31H0THE NUMBER OF DISAGREEMENTS IS,I5)
 102   PRINT 101,IDIS
       MANY=MANY-1
 103   FORMAT(*0NUMBER OF INPUT CARDS IS*I6)
       PRINT 103,MANY
       CALL EXIT
 85    IF(ICARD(ICM)-4) 130,130,90
 90    IF(ICARD(ICM)-6) 100,110,115
 95    FORMAT(1X,A1,4I1,2X,6I1,2X,I1,40X,I3,14X,I3)
C      PERFECT MATCH
 100   ICT=90
 105   IF(ICT-ICARD(77)) 117,106,117
 106   IF(ICT-91) 114,114,107
 107   IF(ICT-996) 109,108,109
 108   LCT=92
       GO TO 116
```

```
 109  IF(ICT-998) 112,111,112
 111  LCT=93
      GO TO 116
 112  IF(ICT-999) 117,113,117
 113  LCT=94
      GO TO 116
 114  LCT=ICT
 116  IAGR(1,LCT)=IAGR(1,LCT)+1
      GO TO 118
 117  IDIS=IDIS+1
 118  PRINT 95,(ICARD(I),I=IDB,IDE),(ICARD(I),I=IDTB,IDTE),
     1 ICARD(ISN),ICT,ICARD(77)
      GO TO 75
 110  ICT=91
      GO TO 105
 115  IF(ICARD(ICM)-7) 110,120,125
 120  ICT=88
      GO TO 105
 125  ICT=89
      GO TO 105
C     SET UP TREE VECTOR
 130  CALL SETUP(1)
      MATNO=1
      IF(IVEC(4)) 200,200,220
 200  IF(ICARD(65)-2) 220,205,220
 205  IF(ICARD(66)-2) 215,210,215
 210  IVEC(4)=1
      GO TO 220
 215  IF(ICARD(67)-2) 220,210,220
 220  CALL MATCH(1)
      IF(LAG) 225,221,221
 221  IF(ICARD(18)) 222,222,170
 222  ICT=999
      GO TO 105
C     PERFECT MATCH
 225  IF(NOD-1) 235,230,235
C     SCHEME A
 230  MATNO=2
      GO TO 135
 235  IF(NOD-2) 245,240,245
C     SCHEME B
 240  MATNO=3
      GO TO 135
 245  IF(NOD-3) 255,250,255
C     SCHEME C
 250  MATNO=4
      GO TO 135
 255  IF(NOD-9) 265,260,265
C     SCHEME F
 260  MATNO=5
      CALL SETUP(MATNO)
      CALL MATCH(MATNO)
      IF(LAG)261,145,170
  261 IF(NOD)262,262,140
```

```
C       SCHEME FA
  262   MATNO=6
        GO TO 135
 265    IF(NOD-8) 275,270,275
C       E
 270    ICT=41
        GO TO 105
 275    IF(NOD-11) 285,280,285
C       H
 280    ICT=47
        GO TO 105
 285    IF(NOD-10) 295,290,295
C       G
 290    ICT=46
 295    IF(NOD-12) 305,300,305
C       I
 300    ICT=62
        GO TO 105
 305    IF(NOD-4) 315,310,315
C       D--39
 310    ICT=39
        GO TO 105
 315    IF(NOD-5) 325,320,325
C       D--40
 320    ICT=40
        GO TO 105
 325    IF(NOD-6) 335,330,335
C       D--37
 330    ICT=37
        GO TO 105
 335    IF(NOD-7) 170,340,170
C       D--38
 340    ICT=38
        GO TO 105
 135    CALL SETUP(MATNO)
        CALL MATCH(MATNO)
        IF(LAG) 140,145,170
C       PERFECT MATCH
 140    ICT=NOD
        GO TO 105
 145    IF(JCT-3) 146,146,155
 146    IF(JCT) 147,147,160
 147    IF(KCT) 170,170,155
 150    FORMAT(1X,A1,4I1,2X,6I1,2X,I1,52X,I3,2X,I3)
 155    IF(IDD-ICARD(77))158,156,158
 156    IF(IDD-91) 157,157,158
 157    IAGR(5,IDD)=IAGR(5,IDD)+1
        GO TO 151
 158    IDIS=IDIS+1
 151    PRINT 150,(ICARD(I),I=IDB,IDE),(ICARD(I),I=IDTB,IDTE),
       1 ICARD(ISN),IDD,ICARD(77)
        GO TO 75
 159    FORMAT(1X,A1,4I1,2X,6I1,2X,I1,43X,3I3,5X,I3)
C       MCT=1 MEANS AGREEMENT, SO TALLY
```

```
C       MCT=0 MEANS NO AGREEMENT
 160    MCT=0
        DO 352 I=1,JCT
        IF(ISD(I)-ICARD(77)) 352,351,352
 351    MCT=1
 352    CONTINUE
        IF(MCT) 357,357,353
 357    IDIS=IDIS+1
        GO TO 356
 353    LCT=ICARD(77)
        IF(LCT-91) 354,354,356
 354    IAGR(JCT+1,LCT)=IAGR(JCT+1,LCT)+1
 356    GO TO(163,164,166),JCT
 161    FORMAT(1X,A1,4I1,2X,6I1,2X,I1,43X,2I3,8X,I3)
 162    FORMAT(1X,A1,4I1,2X,6I1,2X,I1,43X,I3,11X,I3)
 163    PRINT 162,(ICARD(I),I=IDB,IDE),(ICARD(I),I=IDTB,IDTE),ICARD(ISN),
       1 (ISD(I),I=1,JCT),ICARD(77)
        GO TO 75
 164    PRINT 161,(ICARD(I),I=IDB,IDE),(ICARD(I),I=IDTB,IDTE),ICARD(ISN),
       1 (ISD(I),I=1,JCT),ICARD(77)
        GO TO 75
 166    PRINT 159,(ICARD(I),I=IDB,IDE),(ICARD(I),I=IDTB,IDTE),ICARD(ISN),
       1 (ISD(I),I=1,JCT),ICARD(77)
        GO TO 75
 165    FORMAT(1X,A1,4I1,2X,6I1,2X,I1,40X,7HNO ID. ,A4,6X,I3)
  170   GO TO (380,350,380,360,380,380),MATNO
C       SCHEME A WHICH DOES NOT MEET CRITERION
 350    IF(ICARD(17)-2)380,355,380
 355    ICT=996
        GO TO 105
C       SCHEME C WHICH DOES NOT MEET CRITERION
 360    IF(ICARD(17)-1) 365,375,365
 365    IF(ICARD(17)-3) 370,375,370
 370    IF(ICARD(17)-4) 380,375,380
 375    ICT=998
        GO TO 105
 380    PRINT 165,(ICARD(I),I=IDB,IDE),(ICARD(I),I=IDTB,IDTE),ICARD(ISN),
       1 IDMT(MATNO),ICARD(77)
        IAGR(1,94+MATNO)=IAGR(1,94+MATNO)+1
        GO TO 75
        END
```

```
      SUBROUTINE MATCH(MATNO)
      COMMON/A/NROW(6),MROW(6,20),ICARD(80),NTRAN(6,20),MTRAN(6,20,10),
     1 MTRA2(6,20,10),IVEC(20)
      COMMON/B/MAT(6,20,20),NCOL(6),MCOL(6,20),LAG,ISD(3),IDD,NOD,JCT
     1 ,KCT
C     LAG=-1 MEANS PERFECT AGREEMENT--LAG=0 MEANS IDENT.--LAG=1 MEANS NO IDENT.
      NC=NCOL(MATNO)
      NR=NROW(MATNO)
      JCT=KCT=0
      DO 45 J=1,NC
      ICT=0
      DO 10 I=1,NR
      IF(MAT(MATNO,I,J)-99)4,10,4
 4    IF(IVEC(I)-MAT(MATNO,I,J)) 5,10,5
 5    ICT=ICT+1
      IF(ICT-2) 10,10,45
 10   CONTINUE
      IF(ICT-1) 40,15,25
C     SINGLE DISAGREEMENT
 15   IF(JCT-3) 21,20,45
 20   JCT=4
      GO TO 45
 21   JCT=JCT+1
      ISD(JCT)=MCOL(MATNO,J)
      GO TO 45
C     DOUBLE DISAGREEMENT
 25   IF(KCT-1) 35,30,45
 30   KCT=2
      GO TO 45
 35   KCT=1
      IDD=MCOL(MATNO,J)
      GO TO 45
C     PERFECT MATCH
 40   LAG=-1
      NOD=MCOL(MATNO,J)
      RETURN
 45   CONTINUE
      IF(JCT-3) 60,60,50
 50   IF(KCT-1) 55,65,55
C     TOO MANY DISAGREEMENTS
 55   LAG=1
      RETURN
 60   IF(JCT) 70,70,65
 65   LAG=0
      RETURN
 70   IF(KCT-1) 55,65,55
      END
```

```
      SUBROUTINE SETUP(MATNO)
      COMMON/A/NROW(6),MROW(6,20),ICARD(80),NTRAN(6,20),MTRAN(6,20,10),
     1 MTRA2(6,20,10),IVEC(20)
      N=NROW(MATNO)
      DO 15 I=1,N
C     J IS THE NUMBER OF COLUMN ON CARD.
      J=MROW(MATNO,I)
      K=NTRAN(MATNO,I)
C     MAKE TRANSFORMATION
      IS=ICARD(J)
      DO 5 L=1,K
      IF(IS-MTRAN(MATNO,I,L)) 5,10,5
 5    CONTINUE
C     EXIT HERE MEANS NO MATCH--THEREFORE 99
      IVEC(I)=99
      GO TO 15
C     MATCH
 10   IVEC(I)=MTRA2(MATNO,I,L)
 15   CONTINUE
 20   FORMAT(* VECTOR*20I3)
      PRINT 20,(IVEC(I),I=1,N)
      RETURN
      END
```

```
    6
    5   12
18 17 41 17 16
 1  2  3  4  5  6  7  8  9 10 11 12
 1  1  1 -1 -1  1  1 -1 -1  1  1  1
 1  1 -1  1 -1 -1 -1  1 -1  1  1 -1
 1 -1 99 99 99 99 99 99 99  1 -1 99
99 99  1 99 99 -1 -1 99 99 99 99  1
 1  1  1 -1 -1  1 -1  1  1 -1 -1 -1
    2  2  1  1 -1
    4  2  1  1 -1  3 -1  4 -1
    2  2  1  1 -1
    4  3  1  1 -1  2 -1  4 -1
    2  1  1  2 -1
   10   14
14 37 37 42 43 53 36 38 39 40
 1  2  3  4  5  6  7  8  9 10 11 12 13 14
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1  1
 1  1  1  1  1  1  1  1  1  1  1  1 99 99
 1  1  1  1  1  1 -1 -1 -1 -1 -1 -1 -1 -1
 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
 1  1  1 -1 -1 -1 -1 -1 -1 -1 -1  1 -1 -1
 1  1 -1  1  1  1  1  1  1  1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1 -1  1 99  1 -1 -1
 1  1 99 -1  1 99 -1  1 99  1  1  1 -1 -1
 1  1 -1 99  1 99 99  1  1  1  1  1 -1 99
 1 -1 -1 -1 -1  1 -1 -1  1  1  1  1 -1 -1
    9  6  1  1 -1  2 -1  3 -1  4 -1  5 -1  7 -1  8 -1  9 -1
    6  3  1  0 -1  1 -1  2 -1  4  1  5 -1
    6  4  1  0 -1  1 -1  2 -1  3 -1  5 -1
    2  2  1  1 -1
    2  2  1  1 -1
    2  2  1  1 -1
    5  3  1  4  1  1 -1  2 -1  5 -1
    5  3  1  4  1  1 -1  2 -1  5 -1
    5  3  1  4  1  1 -1  2 -1  5 -1
    5  3  1  4  1  1 -1  2 -1  5 -1
    8    4
69 70 71 73 47 45 44 44
15 16 17 18
 1  1 99 -1
 1 -1 99 -1
 1 -1 -1 -1
 1  1 -1 -1
 1  1  1  1
-1  1 -1 -1
99 99 -1  1
99  1 99 99
    2  2  1  1 -1
    2  2  1  1 -1
```

```
NR,NC TREE
ROWS  TREE
COLS  TREE
ROW 1 TREE
ROW 2 TREE
ROW 3 TREE
ROW 4 TREE
ROW 5 TREE
TR1   TREE
TR2   TREE
TR3   TREE
TR4   TREE
TR5   TREE
NR,NC   MA

COLS    MA
ROW1    MA
ROW2    MA
ROW 3   MA
ROW4    MA
ROW5    MA
ROW6    MA
ROW7    MA
ROW8    MA
ROW9    MA
ROW10   MA
TR1     MA
TR2     MA
TR3     MA
TR4     MA
TR5     MA
TR6     MA
TR7     MA
TR8     MA
TR9     MA
TR10    MA
NR,NC   MB
ROWS    MB
COLS    MB
ROW1    MB
ROW2    MB
ROW3    MB
ROW4    MB
ROW5    MB
ROW6    MB
ROW7    MB
ROW8    MB
TR1     MB
TR2     MB
```

```
   2   2   1   1  -1                                        TR3     MB
   2   2   1   1  -1                                        TR4     MB
   3   1   1   2   1   3  -1                                TR5     MB
   3   2   1   1  -1   3  -1                                TR6     MB
   3   2   1   1  -1   3  -1                                TR7     MB
   3   3   1   1  -1   2  -1                                TR8     MB
  10    18                                                  NR,NC   MC
  72  29  28  26  51  50  56  53  57  55                    ROWS    MC
  30  22  28  36  34  24  20  25  27  32  29  19  23  26  33  21  31  35    COLS    MC
  -1  -1   1   1  -1   1   1  -1  -1  -1  -1  -1  -1  -1  -1   1  -1  -1    ROW1    MC
  -1   1   1  -1  -1   1  -1  -1  -1   1   1   1   1   1   1  -1   1  -1    ROW2    MC
  -1   1   1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1   1  -1  -1  -1  -1    ROW3    MC
   1  -1   1   1   1  99  -1  99  99   1   1  -1  99   1   1   1  -1  -1    ROW4    MC
   1   1   1   1  -1   1  -1   1   1   1   1   1   1  -1  -1  -1   1  -1    ROW5    MC
   1   1   1   1   1  -1   1   1  -1   1   1   1   1   1   1   1   1   1    ROW6    MC
   1   1   1   1   1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1   1   1  -1    ROW7    MC
  -1   1   1   1   1   1  -1  -1  -1   1  -1   1  -1  -1  -1  -1  -1  -1    ROW8    MC
  -1   1   1   1  -1  -1  -1   1  -1   1   1  -1  99   1   1  99  99  -1    ROW9    MC
  99  99  99  99  99  99  -1  -1  -1  -1  -1   1  -1   1  -1  -1  -1  -1    ROW10   MC
   2   2   1   1  -1                                        TR1     MC
   3   2   1   3   1   1  -1                                TR2     MC
   2   2   1   1  -1                                        TR3     MC
   4   3   1   4   1   1  -1   5  -1                        TR4     MC
   2   2   1   1  -1                                        TR5     MC
   2   2   1   1  -1                                        TR6     MC
   2   2   1   1  -1                                        TR7     MC
   2   2   1   1  -1                                        TR8     MC
   2   2   1   1  -1                                        TR9     MC
   2   2   1   1  -1                                        TR10    MC
  12    10                                                  NR,NC   MF
  64  52  15  14  14  32  32  32  32  32  32  32            ROWS    MF
  42  43  44  45  49  50  51  52  53  00                    COLS    MF
  -1  -1  -1   1   1   1   1   1   1   1                    ROW1    MF
  99  99  99   1   1   1   1   1   1  -1                    ROW2    MF
  99  99  99   1  -1  -1  -1  99  99  99                    ROW3    MF
  -1  -1   1   1  -1  99  99  99  99  99                    ROW4    MF
  99  99  99   1  99  99  99  99  99  99                    ROW5    MF
  99  99  99  99  99  99  99   1  -1  99                    ROW6    MF
   1  -1  99  -1  -1  -1  -1  -1  -1  99                    ROW7    MF
  -1   1  99  -1  -1  -1  -1  -1  -1  99                    ROW8    MF
  -1  -1  99   1   1  -1  -1  -1  -1  99                    ROW9    MF
  -1  -1  99  -1  -1  -1  -1  -1   1  99                    ROW10   MF
  -1  -1  99  -1  -1   1  -1  -1  -1  99                    ROW11   MF
  -1  -1  99  -1  -1  -1   1  -1  -1  99                    ROW12   MF
   2   2   1   1  -1                                        TR1     MF
   2   2   1   1  -1                                        TR2     MF
   3   3   1   1  -1   2  -1                                TR3     MF
   9   1   1   2   1   3   1   4   1   5   1   6   1   7   1   9   1   8  -1    TR4     MF
   9   3   1   1  -1   2  -1   4  -1   5  -1   6  -1   7  -1   8  -1   9  -1    TR5     MF
   2   2   1   1  -1                                        TR6     MF
   4   3   1   4   1   2  -1   5  -1                        TR7     MF
   4   2   1   5   1   3  -1   4  -1                        TR8     MF
   4   3   1   2  -1   4  -1   5  -1                        TR9     MF
   4   4   1   2  -1   3  -1   5  -1                        TR10    MF
```

```
   4  5  1  2 -1  3 -1  4 -1
   5  1  1  2 -1  3 -1  4 -1  5 -1
  12   11
23 38 53 54 55 56 58 59 61 60 63 64
54 55 56 57 58 59 60 61 62 63 64
-1  1 -1  1 -1 -1 -1 -1 -1 -1 -1
 1  1 -1  1  1 -1 -1 -1 -1 -1 -1
-1  1 -1 -1 -1 -1 -1 -1 -1 -1 -1
 1 -1  1  1  1  1  1  1  1  1  1
 1 -1 -1 -1 -1 99  1  1 -1  1  1
-1  1 99  1  1 -1 99 -1 99  1  1
 1  1 -1 99 99 -1  1 -1  1 -1 -1
 1  1  1  1  1  1  1  1  1  1  1
 1  1  1  1  1 -1  1  1  1  1  1
-1 -1 99  1  1 -1  1 -1  1 -1 -1
-1 99 -1 -1  1 -1  1  1  1  1 -1
 1 99  1  1  1 -1  1  1  1  1  1
   3  3  1  1 -1  2 -1
   5  3  1  4  1  5  1  1 -1  2 -1
   2  2  1  1 -1
   2  2  1  1 -1
   2  2  1  1 -1
   2  2  1  1 -1
   2  2  1  3 -1
   2  2  1  3 -1
   2  2  1  1 -1
   2  2  1  1 -1
   2  2  1  1 -1
   2  2  1  1 -1
```

TR11	MF
TR12	MF
NR,NC	FA
ROWS	FA
COLS	FA
ROW1	FA
ROW2	FA
ROW3	FA
ROW4	FA
ROW5	FA
ROW6	FA
ROW7	FA
ROW8	FA
ROW9	FA
ROW10	FA
ROW11	FA
ROW12	FA
TR1	FA
TR2	FA
TR3	FA
TR4	FA
TR5	FA
TR6	FA
TR7	FA
TR8	FA
TR9	FA
TR10	FA
TR11	FA
TR12	FA

Appendix 3

Sample BUGID Results for Apollo 13

D0002	032770	1	37					37
D0003	032770	1	37					37
D0004	032770	1	2					2
D0005	032770	1	37					37
D0006	032770	1	37					37
D0007	032770	1	37					37
D0008	032770	1	37					37
D0009	032770	1	37					37
D0010	032770	1	13					13
D0011	032770	1		7	13	14		7
D0012	032770	1	37					37
D0013	032770	1	37					37
D0014	032770	1		8	11			8
D0015	032770	1	37					37
D0016	032770	1		3	13			13
D0017	032770	1	37					37
D0018	032770	1	37					37
D0020	032770	1	37					37
D0021	032770	2	37					37
D0022	032770	2		13				13
D0023	032770	2	2					2
D0024	032770	2		7	13	14		7
D0025	032770	2		6				6
D0026	032770	2	996					13
D0027	032770	2	37					37
D0028	032770	2		7	13	14		7
D0029	032770	2		6				6
D0030	032770	2		7	13	14		7
D0032	032770	2		7	13	14		7
D0033	032770	2	37					37
D0034	032770	2	37					37
D0035	032770	2		7	13	14		7
D0036	032770	2		7	13			13
D0037	032770	2		6				6
D0038	032770	2	2					2
D0039	032770	2	13					13
D0040	032770	2	37					37
D0041	032770	2	2					2
D0042	032770	2		8	11			8
D0043	032770	2	2					2
D0044	032770	2		6				6
D0045	032770	2		6				6
D0046	032770	2	2					2
D0048	032770	2	37					37
D0049	032770	2	2					2
D0050	032770	2	2					2
D0051	032770	2	37					37
D0052	032770	2	37					37
D0053	032770	2	13					13
D0054	032770	2		6				6
D0055	032770	2		3	4	13		4
D0056	032770	2	13					13
D0057	032770	2		6				6
D0058	032770	2		8	11			8

D0059	032770	2			6	7	9		7
D0060	032770	2			6				6
D0061	032770	2		37					37
D0062	032770	2			6				6
D0063	032770	2		2					2
D0064	032770	2			6				6
D0065	032770	2			3	4	13		4
D0066	032770	2			6				6
D0067	032770	2			8	11			8
D0068	032770	2		37					37
D0069	032770	2		37					37
D0070	032770	2		2					2
D0071	032770	3			8				8
D0072	032770	3			7	13	14		7
D0073	032770	3			7	13	14		7
D0074	032770	3			13				13
D0075	032770	3			6				6
D0076	032770	3			9	11			7
D0077	032770	3			4				4
D0078	032770	3			7	13	14		7
D0079	032770	3			8	11			8
D0080	032770	3			8	11			8
D0081	032770	3		2					2
D0082	032770	3			7	13	14		7
D0083	032770	3		13					13
D0084	032770	3			7	13	14		7
D0085	032770	3		7					7
D0086	032770	1			13				13
D0087	032770	1			2	4			4
D0088	032770	1		13					13
D0089	032770	1			13				13
D0090	032770	1		2					2
D0091	032770	1			2	3			2
D0092	032770	1		996					13
D0093	032770	1		NO	ID.	B			13
D0094	032770	1		996					13
D0095	032770	1		37					37
D0096	032770	1			3	13			13
D0097	032770	1			7	13			13
D0098	032770	1		13					13
D0099	032770	1			13				13
D0100	032770	1			13				13
D0101	032770	1			13				13
D0102	032770	1			13				13
D0103	032770	1		13					13
D0104	032770	1			13				13
D0105	032770	1			13				13
D0106	032770	2		37					37
D0107	032770	2		37					37
D0108	032770	2		13					13
D0109	032770	2		37					37
D0110	032770	2		37					37
D0111	032770	3		37					37
D0112	032770	3			13				13

D0113	032770	3			13				13
D0114	032770	3	13						13
D0115	032770	3	9						9
D0116	032770	3			13				13
D0117	032770	3	9						9
D0118	032770	3			7	13	14		7
D0119	032770	3	7						7
D0120	032770	3			13				13
D0121	032770	1			8	11			8
D0122	032770	1			8	11			8
D0123	032770	1			8	11			8
D0124	032770	1			8	11			8
D0125	032770	1	2						2
D0126	032770	1	996						8
D0127	032770	1	9						9
D0128	032770	1	9						9
D0129	032770	1	9						9
D0130	032770	1	9						9
D0131	032770	2	13						13
D0132	032770	2	37						37
D0133	032770	2	37						37
D0134	032770	2	37						37
D0135	032770	2	37						37
D0136	032770	2	37						37
D0137	032770	2	7						7
D0138	032770	2	37						37
D0139	032770	2	90						90
D0140	032770	2	37						37
D0141	032770	2	37						37
D0142	032770	2			13				13
D0143	032770	2	37						37
D0144	032770	2	37						37
D0145	032770	2			13				13
D0146	032770	3	13						13
D0147	032770	3	37						37
D0148	032770	3			7	13	14		7
D0149	032770	3			13				13
D0150	032770	3	37						37
D0151	032770	3			13				13
D0153	032770	3	2						2
D0155	032770	3	37						37
D0156	032770	3	996						13
D0157	032770	1	37						37
D0158	032770	1	998						25
D0159	032770	1	5						5
D0160	032770	1	5						5
D0161	032770	1	88						88
D0162	032770	1			8	11			8
D0163	032770	1	2						2
D0164	032770	1	37						37
D0165	032770	1	37						37
D0166	032770	1			7	13	14		7
D0167	032770	1			7	13	14		7
D0168	032770	1			8	11			8

D0169	032770	1	37								37
D0170	032770	1	37								37
D0171	032770	1	37								37
D0172	032770	2	37								37
D0173	032770	2	90								90
D0174	032770	2	37								37
D0175	032770	2		13							13
D0176	032770	2	13								13
D0177	032770	2	37								37
D0178	032770	2		13							13
D0179	032770	2	89								89
D0180	032770	2	998								998
D0181	032770	3	13								13
D0182	032770	3	37								37
D0184	032770	3	4								4
D0185	032770	3	998								28
D0187	032770	1	998								22
D0188	032770	1		7							7
D0189	032770	1							25		25
D0190	032770	1		2	4						4
D0191	032770	1	996								8
D0192	032770	1		2	4						4
D0193	032770	1		13							13
D0194	032770	1	37								37
D0195	032770	1	91								91
D0196	032770	2	37								37
D0197	032770	2	996								8
D0198	032770	2	2								2
D0199	032770	2		13							13
D0200	032770	2		7	13	14					7
D0201	032770	2		8	11						8
D0202	032770	2		7	13						13
D0203	032770	2		7							7
D0204	032770	2	91								91
D0205	032770	2	91								91
D0206	032770	3	2								2
D0207	032770	3	2								2
D0208	032770	3		24							23
D0209	032770	3	90								90
D0210	032770	3	37								37
D0211	032770	3		9							9
D0212	032770	3	2								2
D0213	032770	3	5								2
D0215	032770	3		8	11						8
D0216	033070	1		5							5
D0217	033070	1	2								2
D0218	033070	1	5								5
D0219	033070	1		8	11						8
D0220	033070	1		8	11						8
D0221	033070	1	996								4
D0222	033070	1		4							4
D0223	033070	1	37								37
D0224	033070	1		5							5
D0225	033070	1	37								23

D0226	033070	1	998						25
D0227	033070	1	2						2
D0228	033070	1		5					5
D0229	033070	1	9						9
D0230	033070	1		8	11				8
D0231	033070	2		5					5
D0232	033070	2	13						13
D0233	033070	2	13						13
D0234	033070	2	13						13
D0235	033470	2	13						13
D0236	033070	2		9	11				7
D0237	033070	2	13						13
D0239	033070	2	13						13
D0240	033070	2	13						13
D0241	033070	2	13						13
D0242	033070	2	2						2
D0243	033070	2	37						37
D0244	033070	2		7	13	14			7
D0245	033070	2		5					5
D0246	033070	2		6					6
D0247	033070	2		6	11				6
D0248	033070	2	7						7
D0249	033070	2		6					6
D0250	033070	2	9						9
D0251	033070	2	13						13
D0252	033070	2		6					6
D0253	033070	2		13					13
D0254	033070	2	7						7
D0255	033070	2	9						9
D0256	033070	2		7	13	14			7
D0257	033070	2	9						9
D0258	033070	2		13					13
D0259	033070	2	6						6
D0260	033070	2	6						6
D0261	033070	2	13						13
D0262	033070	2	5						5
D0263	033070	2	2						2
D0264	033070	2		8	11				8
D0265	033070	2		7					7
D0266	033070	2		5					5
D0267	033070	2	13						13
D0268	033070	2		6					6
D0269	033070	2		2	3				2
D0270	033070	2	13						13
D0271	033070	2	13						13
D0272	033070	2	13						13
D0273	033070	2		4					4
D0274	033070	2		8	11				8
D0275	033070	2		7	13	14			7
D0276	033070	2	2						2
D0277	033070	2		6					6
D0278	033070	2		7	13	14			7
D0279	033070	2		6					6
D0280	033070	2	13						13

D0281	033070	2		8 11	8
D0282	033070	2	13		13
D0283	033070	2		3	4
D0284	033070	2	2		2
D0285	033070	2	996		4
D0286	033070	2		13	13
D0287	033070	2	996		7
D0288	033070	2	13		13
D0289	033070	2	2		2
D0290	033070	2	2		2
D0291	033070	2		13	13
D0292	033070	2	4		4
D0293	033070	2	13		13
D0294	033070	2		13	13
D0295	033070	2		2 4	4
D0296	033070	2	2		2
D0297	033070	2		13	13
D0298	033070	2	5		5
D0299	033070	2	13		13
D0300	033070	2		8 11	8

DISTRIBUTION:

NASA, Code SC
Grants and Contracts
400 Maryland Avenue, SW
Washington, D. C. 20546 (25)

L. B. Hall, NASA
Code SB
400 Maryland Avenue, SW
Washington, D. C. 20546 (25)

B. W. Colston
Director, Space & Special Programs
Division
Office of Operations
U. S. Atomic Energy Commission
Albuquerque, New Mexico 87115

L. P. Daspit, Jr.
Viking Project Quarantine Officer
Viking Project Office, NASA
Langley Research Center
Hampton, Virginia 23365

University of California, LRL
P. O. Box 808
Livermore, California 94551
Attn: Tech. Info. Div.
For: Report Librarian

Los Alamos Scientific Laboratory
P. O. Box 1663
Los Alamos, New Mexico
Attn: Report Librarian

Richard G. Bond
School of Public Health
College of Medical Science
University of Minnesota
Minneapolis, Minnesota 55455

John H. Brewer
Star Route 2
Brownwood, Texas 76801

Edward B. Kasner
Director of Research Services
Graduate College
University of New Mexico
Albuquerque, New Mexico

Frank B. Engley, Jr.
Chairman, Department of Microbiology
School of Medicine
University of Missouri
Columbia, Missouri

Gilbert V. Levin
Biospherics, Inc.
4928 Wyaconda Rd.
Rockville, Maryland 20853

Irving J. Pflug
Professor of Environmental Health
University of Minnesota
College of Medical Sciences
Minneapolis, Minnesota 59455

Gerald Silverman
Department of Nutrition and Food Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

John A. Ulrich
School of Medicine
University of New Mexico
Albuquerque, New Mexico

Samual Schalkowsky
Exotech Incorporated
525 School Street, SW
Washington, D. C. 20024

Boris Mandrovsky
Aerospace Technology Division
Library of Congress
Washington, D. C.

Mark A. Chatigny
Research Engineer
Naval Biological Laboratory
Naval Supply Center
University of California, Berkeley
Oakland, California 94625

Richard G. Cornell
Associate Professor of Statistics
Department of Statistics
Florida State University
Tallahassee, Florida

DISTRIBUTION (cont.):

Martin S. Favero
Department of Health, Education
and Welfare
CDC-Phoenix Field Station
4402 North 7th Street
Phoenix, Arizona 85014

F. N. LeDoux
Head, Structural & Mechanical
Applications Section
Goddard Space Flight Center
Greenbelt, Maryland

Q. Ussery
Code NC3, Quality Assurance Branch
Manned Spacecraft Center, NASA
Houston, Texas

F. J. Beyerle
George C. Marshall Space Flight Center
Manufacturing Engineering Laboratory
Code R-ME-MMC
Huntsville, Alabama 35812

J. Gayle
Code SO-PLN, Rm 2123, HQS. Bldg.
Kennedy Space Center, NASA
Cape Canaveral, Florida

Murray Schulman
Division of Biology and Medicine
Headquarters, AEC
Washington, D. C. 20545

N. H. MacLeod
Space Biology Branch
Code 625, Bldg. 21, Rm 161
Goddard Space Flight Center
Greenbelt, Maryland 20771

J. E. Campbell
U. S. Public Health Service
222 E. Central Parkway
Cincinnati, Ohio 45202

G. Rotariu
Process Radiation Staff
Division of Isotopes Development
Headquarters, AEC
Washington, D. C. 20545

Martin G. Koesterer, Microbiologist
Bioscience Operation
General Electric
P. O. Box 8555
Philadelphia, Pennsylvania 19101

Carl Bruch
Chief, Bacteriology Branch
Division of Microbiology
Food and Drug Administration
3rd and C., SW, Rm 3876
Washington, D. C. 20204

John W. Beakley
Department of Biology
University of New Mexico
Albuquerque, New Mexico

Loren D. Potter, Chairman
Department of Biology
University of New Mexico
Albuquerque, New Mexico

Loris W. Hughes
Department of Biology
New Mexico State University
University Park, New Mexico

Richard W. Porter
Corporate Engineering Staff
General Electric Company
570 Lexington Avenue
New York, New York

Fred L. Whipple
Smithsonian Astrophysical Observatory
Cambridge, Mass. 02138

J. J. McDade
Biohazards Group
Pitman-Moore Company
Dow Chemical Company
P. O. Box 10
Zionsville, Indiana 46077

Otto Hamberg
Aerospace Corporation
Building A2, Rm 2019
2350 East El Segundo Blvd.
El Segundo, California

DISTRIBUTION (cont.):

Lawrence P. Chambers
NASA Headquarters
Office of Manned Space Flight
Code MLR
Washington, D. C. 20546

Arthur H. Neill
Code SB
400 Maryland Avenue, SW
Washington, D. C. 20546

Richard Green
Jet Propulsion Laboratory
4800 Oak Grove Dr.
Pasadena, California 91103

Rudy Puleo
Public Health Service
Spacecraft Bioassay Laboratory
Drawer Y
Cape Canaveral, Florida 32900

USAEC, Division of Technical
Information
P. O. Box 62
Oak Ridge, Tennessee 37830
Attn: Reference Branch
P. E. Postell

Carl Sagan
Cornell University
Center for Radiophysics and Space
Research
Space Science Building
Ithaca, New York 14850

Document Library
Lovelace Foundation for Medical
Education and Research
5200 Gibson Blvd., SE
Albuquerque, New Mexico 87108

Martin S. Tierney
Group J-10
Los Alamos Scientific Laboratory
Los Alamos, New Mexico

E. C. Pollard
Professor of Biophysics
Pennsylvania State University
618 Life Sciences Building
University Park, Pennsylvania 16802

Robert Angelotti
Deputy Director
Division of Microbiology
Food and Drug Administration
Health, Education and Welfare
200 C. Street, SW
Washington, D. C. 20546

Vance I. Oyama, Chief
Life Detection Systems Branch
NASA, Ames Research Center
Moffett Field, California 94035

Byron W. Brown, Jr.
Department of Community and
Preventive Medicine
Stanford University School of Medicine
Stanford University Medical Center
Stanford, California 94305

Don G. Fox
Sterility Control Officer
NASA Headquarters, Code SB
400 Maryland Avenue, SW
Washington, D. C. 20546

A. A. Rothstein
Manager, Planetary Quarantine
Martin Marietta Corporation
Mail No. 8401
Denver, Colorado 80201

Hellel S. Levinson
U. S. Army Natick Laboratory
Natick, Massachusetts

A. Anellis
U. S. Army Natick Laboratory
Natick, Massachusetts

B. S. Schweigert, Chairman
Department of Food Science
College of Agriculture
Michigan State University
East Lansing, Michigan 48823

H. O. Halvorson
Biochemistry Department
St. Paul Campus
University of Minnesota
St. Paul, Minnesota

DISTRIBUTION (cont.):

Jack Kaye
11607 Georgetowne Court
Patomic, Maryland 20854

H. W. Johnson, LTC
U.S. Army Medical Research and
Development Command
Washington, D.C. 20314

Donald A. Kautter
Dept. of HEW
Food and Drug Administration
Div. of Microbiology
BF-135
200 C Street S.W.
Washington, D.C. 20204

Mr. James Martin
Viking Project Engineer
Langley Research Center, NASA
Langley Station
Hampton, Virginia 23365

Dr. Richard C. Corlett
Dept. of Mechanical Engineering
University of Washington
Seattle, Washington 98105

J. A. Hornbeck, 1
J. M. Wiesen, 100
W. J. Howard, 1000
D. B. Shuster, 1200
W. A. Gardner, 1500
H. E. Lenander, 1600
T. M. Burford, 1700
C. Winter, 1710
D. R. Morrison, 1720
J. W. Worrell, Jr., 1721
D. P. Peterson, 1724
R. G. Clem, 1730
H. D. Sivinski, 1740 (35)
A. A. Lieber, 1750
R. W. Henderson, 2000
C. B. McCampbell, 2310
B. H. Van Domelen, 2345
S. J. Buchsbaum, 5000
A. Narath, 5100
G. W. Gobeli, 5110
O. E. Jones, 5130
J. E. Schirber, 5150
W. Herrmann, 5160
L. Berthoff, 5162
D. B. Holdridge, 5162
L. C. Hebel, 5200
A. W. Snyder, 5220
R. M. Jefferson, 5221
J. E. McDonald, 5300
L. M. Berry, 5500
D. W. Ballard, 7361
G. A. Fowler, 9000
J. H. Scott, 9200
A. Y. Pope, 9300
L. A. Hopkins, Jr., 9500
R. S. Gillespie, 3411
G. C. McDonald, 3416 (3)
Calla Ann Crepin, 3424-1
W. K. Cox, 3422-1 (15)