Journal of Food Protection, Vol. 67, No. 9, 2004, Pages 1967–1971

ComBase: A Common Database on Microbial Responses to Food Environments†

JÓZSEF BARANYI[1]* AND MARK L. TAMPLIN[2]

[1]*Institute of Food Research, Norwich Research Park, NR4 7UA Norwich, UK; and* [2]*Microbial Food Safety Research Unit, U.S. Department of Agriculture, Agricultural Research Service, Eastern Regional Research Center, 600 East Mermaid Lane, Wyndmoor, Pennsylvania 19038, USA*

MS 03-700: Received 25 August 2003/Accepted 12 March 2004

ABSTRACT

The advancement of predictive microbiology relies on available data that describe the behavior of microorganisms in different environmental matrices. For such information to be useful to the predictive microbiology research community, data must be organized in a manner that permits efficient access and data retrieval. Here, we describe a database protocol that encompasses observations of bacterial responses to food environments, resulting in a database (ComBase) for predictive microbiology purposes. The data included in ComBase were obtained from cooperating research institutes and from the literature and are publicly available via the Internet.

Food microbiology research has generated large quantities of microbiological data on bacterial responses to various environments. Such data form the basis of predictive microbiology software packages such as the Pathogen Modeling Program (PMP; U.S. Department of Agriculture [USDA], Agricultural Research Service [ARS], Eastern Regional Research Center; www.arserrc.gov/mfs/pathogen.htm) and the former Food MicroModel (FMM) in the UK, which was replaced by the freely downloadable Growth Predictor in 2003 (see www.combase.cc). These software packages produce predictions of bacterial responses to food environments characterized by controlling factors such as temperature, pH, water activity, atmosphere composition, and food additives.

The predictions are generated by mathematical models that in most instances have been published in the scientific literature *(1–6, 8–10, 12)*. However, the raw data on which the predictions are based are not easily accessible, even when the data are available from the researchers. In this regard, researchers generally have unique methods for recording their data, which makes comparison with model predictions slow and complicated.

Although the PMP models have demonstrated high utility within the food processing community, managers of the PMP software package have sought to increase the transparency of the models by providing access to the raw data with which the models were generated. However, such a task has involved retrieving thousands of data sets from electronic files and in many cases laboratory notebooks. In hindsight, retrieval of the data sets would have been simple had the data been archived in a relational database. Such a database would (i) provide a permanent record of the experimental design and data, (ii) increase the efficiency of data analysis, (iii) permit greater dissemination of the data among end users, (iv) enhance identification of model limitations, and (v) facilitate the application of new modelling techniques.

The ultimate tests for predictive microbiology software are comparisons of model predictions with observations of microorganismal behavior in food. To make these comparisons with large data sets, the data-recording format must be standardized. This standardization refers not to the computational platform (such as the type of spreadsheet used) but rather to the methodology for classifying and formatting microbiological data.

An examination of various experimental designs and associated data revealed that quality control procedures are needed to bring conformity to predictive microbiology information. Without this conformity, any attempt to compile data from various sources would result in a data dump rather than a structured database. Furthermore, a uniform system of physical, chemical, and biological units and associated terminology must be used to facilitate comparisons among data sets.

Here, we describe a database, ComBase, and its formatting protocol that satisfy the above requirements. We demonstrate the utility of this database using data sets produced by the combined efforts of the Institute of Food Research (Norwich, UK), the Food Standards Agency (UK), and the USDA ARS Eastern Regional Research Center (Wyndmoor, Pa.). At the core of ComBase are the data forming the basis of the models in the PMP and FMM software packages. These core data were extended with data submitted by collaborating institutes and with data gleaned from the scientific literature. ComBase was launched on 16 June 2004 at the 4th International Conference of Predictive Modelling in Foods (Quimper, France).

* Author for correspondence. Tel: +44(0)1603255121; Fax: +44(0)1603255288; E-mail: jozsef.baranyi@bbsrc.ac.uk.

† Mention of brand or firm names does not constitute an endorsement by the U.S. Department of Agriculture over other brands or firms of a similar nature not mentioned.

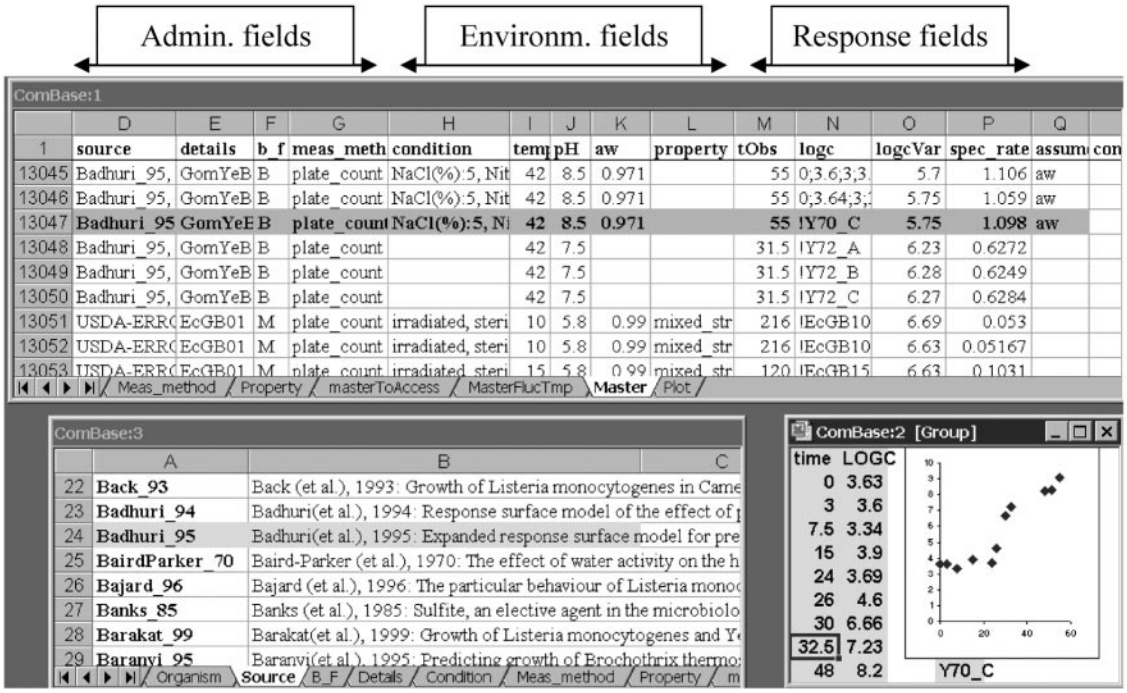

FIGURE 1. *Excerpts from ComBase tables. A sample record is highlighted (shaded) in the master (upper) table. The content of the source field of record 13047 is Badhuri_95, whose definition can be seen in the source definition table (lower left). The value of the logc field is not a single number but rather a table labelled Y70_C, which represents a growth curve (lower right).*

MATERIALS AND METHODS

The data recorded in ComBase can be summarized as a mapping between two multivariate dynamic quantities: environment and microbial response. Because of the inherent complexity of these interactions, one cannot record all aspects of the environment and the associated microbial response. Therefore, a simplification of the experimental results is necessary. To reduce incompatibility among data from different researchers, the syntax and semantics of the database and its structure and format were standardized.

The following questions were identified as essential for determining the structure of ComBase:

1. What components of the environment and the response should be classified and quantified, and how should this information be recorded?
2. How should the quality of the data be determined?
3. What is the structure of the database so that (i) it supports common search requests, (ii) it is user friendly yet sophisticated enough to analyze questions on hazard and risk analysis, and (iii) it is sufficiently flexible to be able to include new types of data?

The database format was established using Microsoft Excel (Microsoft, Redmond, Wash.). Servicing programs for data verification and visualization were written in Visual Basic for Applications (Microsoft) to enhance database functionality.

The data were stored in a typical and simple linear database containing one *master* table and several *definition* tables, where the abbreviations used in the master table were defined. The master table is divided into administrative, environmental, and response fields (Fig. 1). Syntactically, these fields contained numeric or category values. Interpretation regions for numeric values and definition tables for categories were defined for the fields so that incoming data could be tested. This syntax check prevented problems with values outside the interpretation region, such as a water activity value >1.0, being recorded in the database or problems associated with incorrect spelling of organism names (i.e., not as defined in the respective definition table for organisms). To ensure compatibility, doubling times, *D*-values, and other measures of growth or inactivation rates were converted into specific rates. One response (e.g., growth rate or whole growth curve for a quantified environment) became one record (i.e., one row) in the master table of the database.

A unique feature of the database syntax is that it allows recording of both single values and multiple values for variables whose values change with time (i.e., dynamic tables). When the content of a cell in a numeric field is a single value, it is considered the value at the beginning of the observation period. In many instances, this quantity remains the same during the entire observation period. When the value, such as temperature or cell concentration, changes during the observation period, then a name should be given to the dynamic profile that it describes. The profile is recorded in a separate table, and the name of the table will be the content of the cell. This approach is especially advantageous when a whole growth or inactivation curve is recorded rather than just a derived parameter, such as growth or inactivation rate.

A special notation in the database allows the recording of qualitative data in otherwise numeric cells. When the result of a measurement is referenced as N/D (i.e., not detected; more accurately, the value is under the detection limit), then a special number outside the interpretation region of the field can be used as a symbol. For example, the result N/D for the log cell concentration (denoted by logc in ComBase) is denoted by -0.01, which is not a number in this instance but a symbol. Similarly, -0.009999 denotes N/G, for no-growth data (when no log counts, only N/G is reported).

The semantic check of the input data is more difficult than the syntax check and thus cannot always be considered accurate. For example, sometimes it is obvious from the original publication

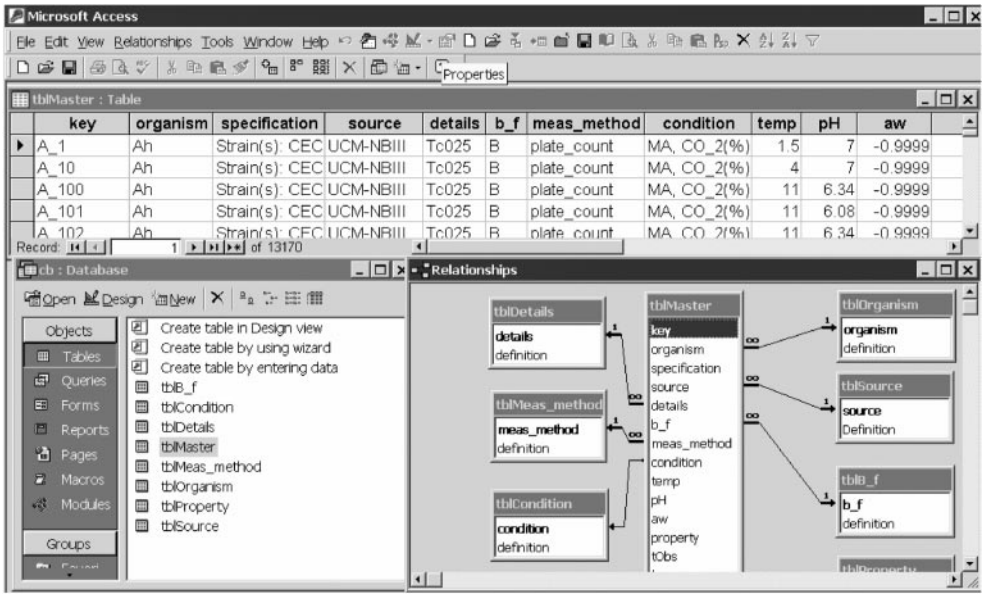

FIGURE 2. *The same structure used in Excel is transposed into Access tables. Access has a larger capacity and more rapid access than Excel.*

that the authors meant to state days and not hours for a certain detection time and thus the published data include a typographical error. In many cases, these errors can be corrected and are not propagated into the database. In addition, by statistical means it is possible to identify outlying data that are obvious errors. However, the semantic verification remains less reliable than the syntactic verification and thus can be affected by subjective judgments.

Excel spreadsheet software was chosen to record the data because this program is widely used among microbiologists, and it is relatively easy to write macros in Excel for data verification. The master and definition tables of the database became one sheet each in the Excel workbook, therefore representing the Excel version of the database. Because its capacity and access speed can be limited for large quantities of data, each of the Excel sheets was converted into corresponding Microsoft Access tables (Fig. 2). During this process, each dynamic table (representing a time-dependent environmental or response quantity) was converted into a single string that provided faster data processing. A browser (Fig. 3) was also written to assist the user in navigating the database. The browser uses the Access version of the database, and the modelling and servicing programs (written in Visual Basic for Excel) uses the Excel version of the database. Both stand-alone and Internet-based versions of the browser have been developed.

The exact technical description and demonstration files can be downloaded from the website (www.combase.cc).

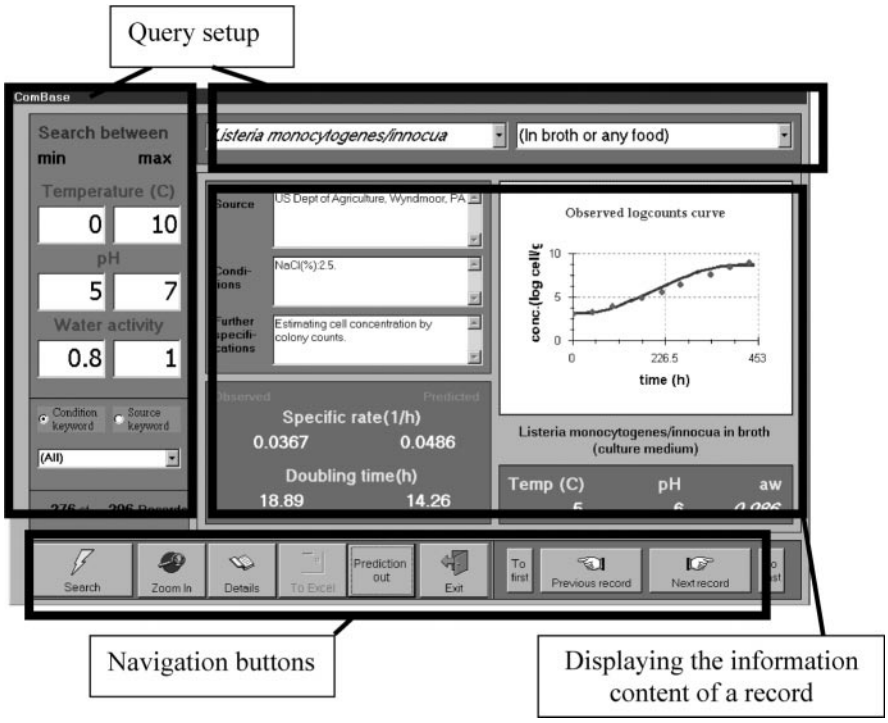

FIGURE 3. *The basic interface structure of the ComBase browser.*

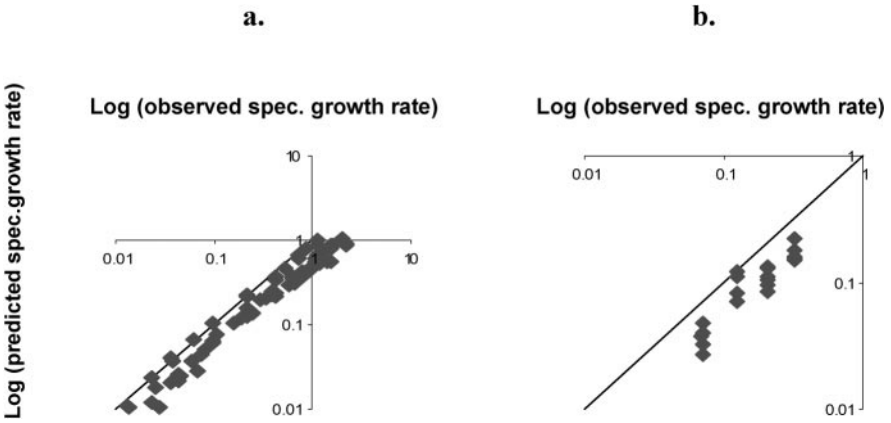

FIGURE 4. *An Excel macro was used to compare ComBase data for the specific growth rates predicted by the PMP software with observations published by Grau and Vanderline (7) (a) and Patterson et al. (11) (b).*

RESULTS

Currently, two types of microbial responses are recorded in ComBase: (i) full growth or survival curves produced by viable count measurements and (ii) specific growth or inactivation rates only, derived from viable count or other measurements (e.g., optical density) as published by various authors in refereed articles.

The majority of the approximately 30,000 full viable count growth curves in ComBase constitute the raw data on which the PMP and FMM software packages were developed. These growth curves were generated in vitro for selected pathogens in various environments. Many of the viable count curves of both pathogenic and spoilage organisms are from supporting institutes around the world. Compilation of these data was funded by the European Commission. Approximately 10,000 records containing mostly specific rates only, not full growth curves, were complied from the scientific literature under the funding of the Food Standards Agency.

One of the useful features of computerized data is that the recorded information can be quickly located, which is a great advantage even when the records do not provide the complete original article or the full description of an experiment. Figure 4 includes two examples where the user can compare between the observed data and predictions. These outputs were prepared using a macro that is part of the kit of Excel add-ins supplied with ComBase. Figure 4a shows a plot comparing the recorded specific growth rates obtained by Grau and Vanderline *(7)* and those predicted by the PMP. The software predictions are based on temperature, pH, and water activity only. Figure 4b shows a comparison between the data obtained by Patterson et al. *(11)* and the PMP predictions. These plots were the result of two select-and-click operations and did not involve the tedious task of collecting data from published literature. Note that the new version of ComBase uses the predictions of Growth Predictor rather than PMP.

When complete viable count growth and survival curves are recorded, those also can be compared with predictions using another Excel macro.

DISCUSSION

ComBase is a system of Microsoft Excel spreadsheets, an Access database, and servicing programs with the following features:

1. ComBase Excel: a Microsoft Excel workbook containing relational worksheets (tables). Input of new data, verification, and modelling is carried out on this version of the database.
2. ComBase Access: the database in Microsoft Access. Its tables are the same as the sheets in the Excel version. After input and verification of the Excel entries, the data are transferred to Access, which has higher capacity and more rapid search capabilities.
3. Maintenance and modelling kit: Excel add-ins working on the Excel version of the database. This kit includes maintenance and statistical macros to help with data verification and the development of predictive models.
4. ComBase Browser. Built on the Access database, this browser navigates in the Access version in a user-friendly manner. An Internet version of this browser has also been developed.

We anticipate that, similar to genomic databases (i.e., gene banks), ComBase will serve as a repository for data that can be accessed by persons seeking to estimate microbial responses to various food environments. ComBase can be used to define data gaps, which will stimulate research needed to bring microbiology information to a critical mass. Such a unified database will also assist in standardizing the work and results of different risk assessors, which could have obvious and positive implications on international trade.

ACKNOWLEDGMENTS

The support of the United Kingdom Food Standards Agency (project FS 3113) and the European Commission, Quality of Life and Management of Living Resources Programme (QoL), Key Action 1 on Food, Nutrition and Health, contract QLK1-CT-2002-30513 (e-ComBase) is gratefully acknowledged.

REFERENCES

1. Bhaduri, S., R. L. Buchanan, and J. G. Phillips. 1995. Expanded response surface model for predicting the effects of temperatures, pH, sodium chloride contents and sodium nitrite concentrations on the growth rate of *Yersinia enterocolitica. J. Appl. Bacteriol.* 79: 163–170.
2. Bhaduri, S., C. O. Turner-Jones, R. L. Buchanan, and J. G. Phillips. 1994. Response surface model of the effect of pH, sodium chloride and sodium nitrite on growth of *Yersinia enterocolitica* at low temperatures. *Int. J. Food Microbiol.* 23:333–343.
3. Buchanan, R. L., L. K. Bagi, R. V. Goins, and J. G. Phillips. 1993.

Response surface models for the growth kinetics of *Escherichia coli* O157:H7. *Food Microbiol.* 10:303–315.
4. Buchanan, R. L., and L. A. Klawitter. 1992. Characterization of lactic acid bacterium, *Carnobacterium piscicola* LK5, with activity against *Listeria monocytogenes* at refrigeration temperatures. *J. Food Saf.* 12:199–217.
5. Buchanan, R. L., and J. G. Phillips. 1990. Response surface model for predicting the effects of temperature, pH, sodium chloride content, sodium nitrite concentration, and atmosphere on the growth of *Listeria monocytogenes. J. Food Prot.* 53:370–376.
6. Buchanan, R. L., J. L. Smith, C. McColgan, B. S. Marmer, M. D. Golden, and B. J. Dell. 1993. Response surface models for the effects of temperature, pH, sodium chloride, and sodium nitrite on the aerobic and anaerobic growth of *Staphylococcus aureus* 196E. *J. Food Saf.* 13:159–175.
7. Grau, F. H., and P. B. Vanderline. 1993. Aerobic growth of *Listeria monocytogenes* on beef lean and fatty tissue: equations describing the effects of temperature and pH. *J. Food Prot.* 56:96–101.
8. Oscar, T. P. 1999. Response surface models for effects of temperature, pH, and previous growth pH on growth kinetics of *Salmonella* Typhimurium in brain heart infusion broth. *J. Food Prot.* 62:106–111.
9. Palumbo, S. A., A. C. Williams, R. L. Buchanan, and J. G. Phillips. 1991. Model for the aerobic growth of *Aeromonas hydrophila* K144. *J. Food Prot.* 54:429–435.
10. Palumbo, S. A., A. C. Williams, R. L. Buchanan, and J. G. Phillips. 1992. Model for the anaerobic growth of *Aeromonas hydrophila* K144. *J. Food Prot.* 55:260–265.
11. Patterson, M. F., A. P. Damoglou, and R. K. Buick. 1993. Effects of irradiation dose and storage temperature on the growth of *Listeria monocytogenes* on poultry meat. *Food Microbiol.* 10:197–203.
12. Zaika, L. L., E. Moulden, L. Weimer, J. G. Phillips, and R. L. Buchanan. 1994. Model for the combined effects of temperature, initial pH, sodium chloride and sodium nitrite concentrations on anaerobic growth of *Shigella flexneri. Int. J. Food Microbiol.* 23:345–358.