

Technical description of the *ComBase* database

Introduction

The *ComBase* database is a combined database comprising tens of thousands of records of microbial responses to food environments. The home page of the ComBase project is <http://www.combase.cc>, where also the project background is described. The page has a link to the ComBase Browser that navigates in the database, presenting the recorded observations of microbial responses to key environmental factors.

This manual summarises the rationale for a common data format, provides a technical description of the structure of the *ComBase* database, and describes the format required before records can be included in the database. The manual is accompanied by a demo Excel file.

Rationale for a common data format

Microbiological data are presented in various formats. It is necessary to standardize these formats into a uniform one before they can be included in the database. Uniformity facilitates comparison within and between sets of data and allows rapid searching of the database by pre-determined key parameters.

Generalised structure of the ComBase database

The database is basically a system of connected tables, one of which is the main, called **Master** table. The column headings in the **Master** table represent fields and most have a corresponding sheet containing relevant information.

Figure 2. A section from the ComBase **Master** table, in Excel, where the possible values of a field (whose name is in the heading of the respective column) are stored in a linked sheet. (The link is pointing to the sheet tab on this figure).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	key	or	specification	source	detail	b. f.	meas. mc	condition	tem	pH	aw	property	tObs	logc
2	B251_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut			54	6		mixed_strains	1.2	0.7;0.1
3	B252_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, acetic_acid(ppm)			54	5		mixed_strains	0.467	0.7;0.1
4	B253_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, citric_acid(ppm)			54	5		mixed_strains	1.2	0.7;0.1
5	B254_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, lactic_acid(ppm)			54	5		mixed_strains	0.967	0.7;0.1
6	B255_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut			54	6		mixed_strains	1.2	0.7;5.1
7	B256_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, acetic_acid(ppm)			54	5		mixed_strains	0.7	0.7;5.1
8	B257_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, citric_acid(ppm)			54	5		mixed_strains	1.2	0.7;5.1
9	B258_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, lactic_acid(ppm)			54	5		mixed_strains	1.2	0.7;5.1
10	M239_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut			21	6		mixed_strains	24	0.4;8.1
11	M240_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, citric_acid(ppm)			21	5.4		mixed_strains	24	0.4;8.1
12	M241_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, lactic_acid(ppm)			21	5.4		mixed_strains	24	0.4;8.1
13	M242_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, citric_acid(ppm)			21	5		mixed_strains	24	0.4;8.1
14	M243_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, lactic_acid(ppm)			21	5		mixed_strains	24	0.4;8.1
15	M244_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, citric_acid(ppm)			30	5.4		mixed_strains	24	0.4;7.6
16	M245_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, lactic_acid(ppm)			30	5.4		mixed_strains	24	0.4;7.6
17	M246_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, citric_acid(ppm)			30	5		mixed_strains	24	0.4;7.6
18	M247_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, lactic_acid(ppm)			30	5		mixed_strains	24	0.4;7.6
19	M248_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, acetic_acid(ppm)			30	4.7		mixed_strains	24	0.4;7.6
20	M249_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, citric_acid(ppm)			30	4.7		mixed_strains	24	0.4;7.6
21	M250_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut, lactic_acid(ppm)			30	4.7		mixed_strains	24	0.4;7.6
22	O412_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut			21	6.07		mixed_strains	24	0.5;3.1
23	O413_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut			21	5.9		mixed_strains	24	0.5;3.1
24	O414_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut			21	5.7		mixed_strains	24	0.5;3.1
25	O415_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut			30	6.07		mixed_strains	24	0.5;3.1
26	O416_Ec	Ec	Serotype(s): O1 Abdul-Raouf_93	Beef	plate_count	heated_cut			30	5.9		mixed_strains	24	0.5;3.1
ComBase Master Table														
N Field		purc. attrbu		C Source		B. F. Details		Meas. method		Property		Master ID		

The data contained within the database are collected from collaborating institutes and from the scientific literature. They are recorded in Microsoft Excel sheets which are supplied with servicing macros (see the accompanying demo file), to verify and analyse or model incoming data. After verification, data are transferred to Microsoft Access® tables which is the input for the Internet-program, the ComBase Browser.

The remainder of this document describes the Excel version of ComBase, which is used by donors to submit their data.

Fields are divided into three major groups:

1. Administrative fields:

- key:** a unique identifier for each record
- organism:** (category of the) organism
- specification:** specific features of organism
- source:** reference for the data
- meas_method:** measurement method
- tObs:** duration of observation
- details:** materials and methods
- linkID:** an ID linking records on co-culture of different organisms

2. Environment fields:

- temp:** temperature of the environment
- pH:** pH of the environment
- aw:** water activity of the environment
- b_f:** broth (culture medium) or food category
- condition:** secondary environmental factors likely to modify the response significantly
- comment:** field for possible other, uncategorized conditions for which the bacterial response was recorded
- in_on:** details on the food in or on which the micro-organism was inoculated

3. Response fields:

- logc:** logarithm of the cell concentration
- spec_rate:** specific rate, positive for growth, negative for death
- logcVar:** the range in which logc varies
- logc0:** inoculum level

Further subdivision of fields:

All fields can be subdivided into one of two main types; *category* or *numeric*. Category fields contain alphanumeric values (i.e. a *string*, which is a series of characters). This can also be a list of string items. The fields *organism*, *specification*, *source*, *measurement method*, *in_on*, *b_f*, *condition* and *comment* are category fields.

The content of a numeric field can be a number, a series of numbers, or a pointer. The fields *tObs*, *pH*, *temp*, *aw*, *logc*, *logc0* and *spec_rate* are numeric fields. If a numeric field contains letters preceded by an exclamation mark, those letters refer to a numeric dynamic profile which is stored in a separate table. For

example, '!T1' points to an (x;y) table. For all fields, the syntax is important and the content of most fields are checked before inclusion in the database.

Important definitions

record: a row in the **Master** sheet is a record representing a single microbial response: (a parameter or a whole logcount v. time curve).

field: a column in the **Master** sheet is a field. The names of the fields are in the top row of the columns.

list: in some fields multiple data items can be entered to form a list. These are to spare columns in the Excel version of the database. **ATTENTION! For list separator, use comma and, for decimal symbol, use point** (you can select this option in the International setting on the Control panel).

item: an item is the basic element of a list. It must be valid in the respective field. A single item can also be a pointer, which is a reference, preceded by an exclamation mark, for the sheet where the full description can be found.

dynamic profile: both the environment and the response can be dynamic (time-dependent) profiles. For example, the environment is a dynamic profile under fluctuating storage temperature or in a food product which is becoming dry.

List example

The cell content

'co-culture, CO2(%):43, O2(%):39, N2(%):18'

is a list, whose items are

co-culture
CO2(%):43
O2(%):39
N2(%):18

Dynamic profiles can be recorded in two ways:

Dynamic profile (Table-format)

time	temp
0	4
2	4
5	9
9	9

This represents a situation where the temperature was 4°C for two hours, then linearly increased to 9°C during the next 3 hours and remained constant for the next 4 hours. The table can also be entered into the database using the following format:

Dynamic profile (format x1;y1;x2;y2...):

0; 4; 2; 4; 5; 9; 9; 9

The different formats can be converted from one to the other with the servicing macros provided with the demo file.

Description of individual fields and formatting requirements

The syntax of all fields and examples are provided in the following sections.

key

Syntax:

The key is a unique identifier; a string of alphanumeric characters, beginning with a letter. This field must not contain a list.

Typical practice (not compulsory):

It is usual for the person entering the data to use a code that associates the records with him/herself or with the origin of the data.

Examples:

Y_1

B401_160

organism

Syntax:

This field cannot be empty. The possible values are provided in the associated sheet called **Organism**:

organism	Definition
Ac	<i>Aeromonas caviae</i>
Ah	<i>Aeromonas hydrophila</i>
As	<i>Aeromonas sobria</i>
Bac	bacillus spoilage bacteria
Bc	<i>Bacillus cereus</i>
etc	<i>Etc. see the demo file</i>

Typical practice:

It is generally the well-known two-letter abbreviation of the organism, possibly followed by a letter about typical characteristics (like “p” for “proteolytic”, in the case of *C.botulinum*: Cbp). If the response was measured for a mixed culture, a generic name can be given (such as “Ss” for a mixed culture of salmonellae).

Specification

The specification field can contain features about the organisms that the data producer considered important but cannot be put into a category. For example, if the organism field content is **Ss**, the specification can contain also something like this (see the record with key B092_66 in ComBase):

'Species: S.infantis S.thompson S.stanley '

That is: these three formed a mixed culture whose growth was measured.
the organism field is **Ss** for the family of salmonellae).

Syntax:

A comma-separated list of keywords, each followed by a relevant data item (or a space separated list of data items) giving more information on the organism. The following keywords should be used as appropriate:

Group:

Species:

Type(s):

Serotype(s):

Strain(s):

Isolated_from:

The specification after the "Group:" keyword indicates that the information given in the "Organism" field is narrowed down to a smaller group of bacteria.

An example (record Y_100 in ComBase):

'Serotype(s): 9, Strain(s): CECT754'

In this record, the serotype of the organism is 9 and the strain is CECT754. The different items (serotype and strain) of the list are separated by commas.

Specification, Example 1.

'Group: Lactobacilli_mesophili'

In this record, a group of bacteria has been enumerated or studied without distinguishing between the different species. Note that there should be an underscore between "Lactobacilli" and "mesophili", not a space. A space would indicate that two groups of bacteria have been enumerated altogether.

Specification, Example 2.

'Species: Ent.agglomerants Ent.oxytoca Ent.pneumoniae E.coli Ent.morganii, Strain(s): NCFB2071 NCFB2073 NCFB2678 NCFB1010 NCFB555 NCTC235'

This example shows a mixture of species with their corresponding strains. The species and strains are recorded in the same order: *Ent.agglomerants* is studied by the strain NCFB2071, *Ent.oxytoca* by strain NCFB2073, etc. For the last species, *Ent.morganii*, 2 different strains were in the mixture: NCFB555 and NCTC235.

Note that there should be no space between "Ent." and "agglomerants". A space would indicate two different species ("Ent." and "agglomerants").

If all the strains were of the same species, then only one species would be recorded.

source

Syntax:

In the **Master** sheet, the `source` field contains a code for the source of the data. The code comprises the surname of the first author of the paper if the data came from the literature, otherwise the commonly used abbreviation of the institute where the data were generated, followed by the date of publication. For example:

Pin_00
FSA-IFR

The code corresponds to a definition listed in the **source** sheet. The definition should contain the full reference for a publication or can contain email or web addresses.

Typical practice

If the data come from an institute without publication, then the abbreviation of the institute is the pointer to the source. If the data have been published, then the name of the entry is the name of the first author, followed by `_yy` (the yy indicating the year). In case of more publications of the same first author in the same year, a, b, c etc is used after the year.

Source	Definition
FSA-IFR	Food Standards Agency funded data generated at the Institute of Food Research, UK.
Pin_00	Pin(et al.), 2000: Predictive model for the growth of Yersinia enterocolitica under modified atmospheres. Journal of Applied Microbiology 88:521-530

details

Syntax:

A unique code pointing to a text defined in the **details** sheet. For example if the entry is 'T427' then this code refers to a longer description in the **details** sheet, in the row where the first entry is 'T427'. One record in the **details** sheet may be valid for several records in the Master sheet.

Typical practice:

The text describes materials and methods of the recorded response. These details may be brief or extensive. Their purpose is to give an overview of the broad experimental detail to the ComBase user to further facilitate comparison between records. It is useful to provide facts under the subheadings "To determine:", "Inoculum:", "Inoculation and storage:", "Enumeration:", etc. Remember that whole sentences and paragraphs should not be copied from publications protected by copyright!

b_f

Syntax:

A value must be entered. The possible values are given in the **b_f** sheet:

b_f	Definition
Culture_medium	Culture medium
Beef	Beef
Pork	Pork
Poultry	Poultry

Sausage	Sausage
Meat_other	Other or unknown type of meat
Seafood	Seafood
Milk	Milk
Cheese	Cheese
Dairy_other	Other or unknown type of dairy
Egg	Egg or egg product
Dessert	Dessert food
Sauce/Dressing	Sauce/Dressing
Produce	Vegetable or fruit and their products
Bread	Bread
Infant_food	Infant_food
Beverage	Juice, beverage
Water	Water
Other/mix	Other, mixed, uncategorised or unknown type of food

meas_method

Syntax:

An abbreviation of the measurement method. The possible values are given in the **Meas_method** sheet:

meas_method	Definition
plate_count	Measurement by colony counts
MPN	Measurement by Most Probable Number method
od	Measurement by optical density
memb_filtration	Measurement by membrane filtration
electrical	Measurement by impedance, conductance, etc.
microscopy	Measurement by microscopic observation.

condition

Syntax:

Secondary factor(s) of the environment. A list is permitted. If a factor is defined quantitatively, then its syntax is *factor(unit):number*. For example:

NaCl(%): 5, acetic_acid(ppm):1000

is a list of quantitative items in the **condition** field.

The existence of an unquantified condition is denoted by the condition code, without a colon, and a number, such as 'NaCl'.

The field can contain one or a list of these values [possibly with *(unit): number*]:

condition	Definition	unit	default
ALTA	ALTA fermentation product in the environment	(%)	
acetic_acid	Acetic acid (possibly as salt) in the environment	(ppm)	
anaerobic	Anaerobic environment		default: aerobic
ascorbic_acid	Ascorbic acid (possibly as salt) in the environment	(ppm)	
benzoic_acid	Benzoic acid (possibly as salt) in the environment	(ppm)	
<i>etc</i>			

Typical practice:

NaCl is listed as a condition, although for experiments in broth medium, NaCl concentration can be converted to water activity and entered in the `aw` field. In this case the `assumed` field contains the entry 'aw' (see below).

NaCl concentration is frequently not recorded. The default value is NaCl(%):0.5. Note that organic acid concentrations should be given as parts per million, ppm. If they are added as salts, then they should be converted into equivalent ppm acid and it may be noted in the `comments` that they were originally added as salt.

Property

Syntax

Similar to `condition`, but this refers to the *organism(s)* producing the recorded response, not to their environment. Their values can be one or a list of these:

property	Definition	unit
<code>exp_inoculated</code>	Inoculum in the exponential phase of growth.	
<code>stat_inoculated</code>	Inoculated from stationary phase (also if not indicated; default)	
<code>pre_heated</code>	Cells heated before incubation.	
<code>indigenous</code>	Non-inoculated organism (indigenous flora).	
<code>injured</code>	Injured by pre-treatment .	
<code>pre_irradiated</code>	Cells irradiated before inoculation.	(KGy)
<code>mixed_strains</code>	Culture of mixed strains produced the response.	
<code>mixed_species</code>	Culture of mixed species produced the response.	
<code>spore</code>	Spore inoculation.	
<code>pre_aw</code>	Pre-inoculation water activity.	
<code>atr_induced</code>	Acid tolerance response (ATR) induced inoculum	
<code>pre_diacetic</code>	Grown in presence of diacetic acid (possibly as salt) before time 0 of this record	(ppm)
<code>pre_dried</code>	Cells dried before time 0 of this record	
<code>pre_lactic</code>	Grown in presence of lactic acid (possibly as salt) before time 0 of this record	(ppm)
<code>pre_pH</code>	Pre-inoculation pH.	
<code>pre_pressure</code>	Pressurised before time 0 of this record	(Mpa)
<code>pre_temp</code>	Pre-inoculation temperature.	(°C)
<code>pre_tObs</code>	Pre-inoculation observation time	(h)
<code>pre_NaCl</code>	Pre-inoculation NaCl.	(%)

Typical practice:

For example, where pre-boiled rice was inoculated with *Bacillus cereus* cells, the `condition` includes 'heated', and where rice containing *Bacillus cereus* spores was boiled for 20 minutes before a response was recorded (the `property` includes: 'spore, pre_heated').

The default value for the inoculum is stationary phase so it is not listed in the properties.

temp

Syntax:

A numeric value for the temperature, possibly a dynamic list, expressed in °C.
The temperature should always be recorded.

Typical practice:

The temperature recorded here is its value (assumed to be constant) , or a dynamic profile.

If the temperature changes significantly during the observation but no information is provided by the authors (e.g. high-pressure experiments), then only the initial temperature is recorded.

If temperature changes during the response and those changes are documented, they can be recorded as a dynamic list or table.

Dynamic temp profiles can be big therefore the user may want to put these in a sheet dedicated to those specific profiles. Data entered can then be converted to database format using a servicing macro provided with the demo file.

pH

Syntax:

A numeric value, pH at time=0, or possibly a dynamic list, with values between 0 and 14.

Typical practice:

Sometimes pH is ascertained from the descriptions of e.g. food matrix or experimental conditions provided. The pH may change during experiments, however if this information is not available in detail, then only the initial pH value is recorded.

a_w

Syntax:

A numeric value, water activity at time=0, or possibly a dynamic list with values between 0 and 1. Like pH, the initial value is usually recorded but may be not valid throughout a dynamic response.

Typical practice:

Often it is assumed. In most cases in the database, for experiments in broth, a_w has been calculated from NaCl concentration by the formula:

$$a_w = 1 - \text{NaCl} * (5.2471 + 0.12206 * \text{NaCl}) / 1000$$

(Resnik, S.L. and Chirife, J. [1988]. Proposed theoretical a_w values at various temperatures for selected solutions to be used as reference sources in the range of microbial growth. J. Food Prot. 51. 419-423.)

It cannot always be assumed that the water activity is a function of NaCl concentration, other conditions may also influence the water activity, for example sugars.

tObs

Syntax:

Length of observation time or duration of experiment, recorded in hours.

Response fields:

logc

Syntax:

This field contains the 10-based logarithm of the estimated microbial concentration, in terms of cells/ml or cells/g. A growth or death curve is recorded as a dynamic profile of logc, a semi colon separated list (time;logc;time;logc; ...) in the **Master** sheet, or as a table in a separate sheet.

Typical practice:

Full curves are stored as dynamic profiles. From the literature, if the curve is not given, then only the rate is recorded (in the `spec_rate` field). Where only a graph is provided then the full curve is ascertained from the graph or, preferably, raw data provided by the author are recorded. Data entered as tables in separate sheets can be converted by servicing macro to database format. **IMPORTANT:** “Not detected” is denoted by “-0.01”, used here as a symbol, not a number.

logcVar

Syntax:

A numeric field showing the range in which logc during the measurement i.e. the difference between the highest and lowest value. This gives an indication of whether the change in logc is sufficient to conclude a growth or death tendency.

When we do not know the lowest or highest count (e.g. not detected), then *logcVar* is given with a minus (-) sign: it indicates that the variation was greater than a value. For example ‘-3’ means that the variation of the concentration was greater than 3 logs.

Typical practice:

The value of logcVar can be calculated by a macro, if the response is a whole curve.

spec_rate

Syntax:

The maximum specific growth or death rate observed. If a whole logcount curve is available, this is empty. If not, this value has to be provided, since the response is either the logcount curve or only the rate. A ‘no growth’ situation is expressed by -0.009999, used here as a symbol, not a number . This field cannot contain a list.

Typical practice:

Where possible data from the literature has been converted to specific growth or death rate using:

for growth:	spec_rate = 2.3*rate in log counts
	spec_rate = 0.69 / doubling time
for survival and death:	spec_rate = - 2.3 / D-value

Comment fields

assumed

Syntax:

Field name, or a list of names, for which the values in the record are assumed or calculated (not measured).

Typical practice:

The default value is pH 7 in which case it does not need to be recorded in the field pH. If the pH value is ascertained from the food and different from 7, the value of pH should be entered in the pH field and pH should appear in the field assumed. The same applies for aw (default value=0.997).

comment

Syntax:

No special syntax is imposed.

Typical practice:

Specific, unusual conditions that are not indicated in the **condition** sheet, but can significantly affect the response, should be included here. It is recommended that a similar syntax to the condition field is used.

in_on

Syntax:

In: or On: is followed by the name of food in/on which the experiment has been carried out. On: indicates that the inoculation took place on the surface of the food.

Typical practice:

For the experiments in broth, the type can be specified; for instance, 'In: TSYE broth'. An example of inoculation at the surface of a food: 'On: sliced fresh fruit'.

linkID

In the linkID field one can indicate that more than one organisms (i.e. curves / records) have been measured in the same experiment; for example both the total spoilage flora and a pathogen were observed. The environmental fields (temp, pH, aw, condition) of linked records per se must have the same values.

The syntax checking macro validates such integrity of linked records.

logc0

This field is needed only if no curve is given, only the rate, but we have a good estimation on the inoculum size.

Appendix: Servicing macros to enter and collect data in the Excel version of ComBase

The Excel demo file works with macros (CBmacros):

details references:

Opens a window with the detailed explanation of the cell under the cursor. The shown details are either:

1. details about the field in which the selected cell is (e.g. any cell in the `specification` field shows the definition of specification), or
2. details about the first item of the list in categorical field (e.g. in cells H2, it shows the definition of the category NaCl), or
3. the text corresponding to a reference (e.g. in cells E3, it shows the materials and methods referenced Tc027), or
4. a plot a the growth curve in the `logc` field when available or temperature profiles.

syntax check of selection:

Checks the validity of the selected data (whether the syntax is in agreement with the database conventions).

Index file creation / extract data for modelling:

Creates a new Workbook with a sheet **Index0**, containing the selected keys, and a sheet with the corresponding `logc` profiles. The primary conditions (temperature, pH, water activity) and the secondary conditions (from the conditions column) are also given in the **Index0** sheet. The data format is compatible with the use of the DmFit program to create models. Only record with the same micro-organisms can be selected.

replace curves given in string by curvenames:

In the `logc` field, replaces the selected strings by "!key", and saves the data in a table in a worksheet **organism_logc**. The organism name is the same as in the corresponding record.

replace curvenames by values in strings:

Inverse of the macro above: changes the selected curves' names into strings consisting of xy coordinates. The data should be recorded as tables in a sheet named **organism_logc** and the curves' name should be "!key".

compare curve with prediction:

Superposes the predictions of existing model in broth (models of ComBase predictor) to the `logc` data of the selected record. CBPred.xlsm is needed to be present in the same working directory as the demo file

calculate tObs and logcVar:

Creates a new sheet where `tObs` and `logcVar` are estimated from the growth curve recorded in the `logc` field. The records in the field `logc` should be selected.