

**Piotr Walczak**

Instytut Technologii Fermentacji i Mikrobiologii  
Politechnika Łódzka

## **Komputerowe bazy danych dla Kolekcji Czystych Kultur Drobnoustrojów**

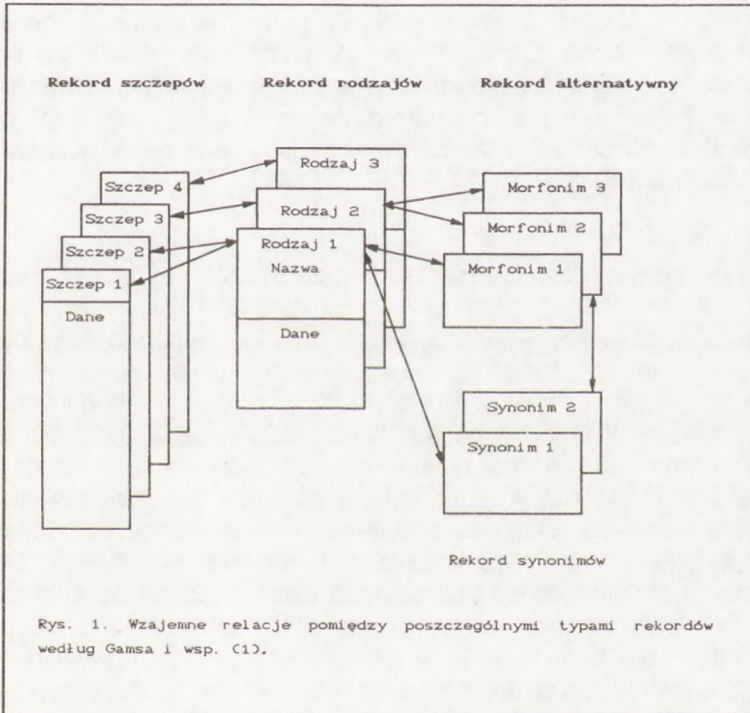
### **Wstęp**

Rozwój techniki mikroprocesorowej i postęp w budowie mikrokomputerów doprowadził do znacznej obniżki cen sprzętu informatycznego, a w konsekwencji do jego upowszechnienia. Pozwoliło to na wykorzystanie komputerów nie tylko do celów obliczeniowo-statystycznych, ale również na bardziej "trywialne" zastosowania, takie jak edycja tekstów oraz tworzenie komputerowych baz danych. Jednym z możliwych zastosowań komputerów w biotechnologii jest tworzenie specjalizowanych baz danych typu bibliograficznego o procesach mikrobiologiczno-biochemicznych, banków genów, sekwencji białek, a także baz danych dla Kolekcji Czystych Kultur Drobnoustrojów. W ostatnim przypadku organizacja komputerowej bazy danych wydaje się bardzo celowa, gdyż znacznie ułatwia dostęp do informacji o przechowywanych szczepach, aktualizację danych, szybkie drukowanie katalogów drobnoustrojów oraz wymianę informacji pomiędzy różnymi Kolekcjami, zwłaszcza w przypadku połączenia poszczególnych komputerów w sieć informatyczną. Dla ułatwienia wymiany informacji należy przyjąć wspólny standard oprogramowania, wspólną strukturę baz danych oraz format przechowywanych informacji. Dotychczas w różnych krajach podejmowano próby tworzenia specjalizowanych baz danych dla Kolekcji Kultur Drobnoustrojów, które działały wyłącznie w obrębie danego kraju. Przykładem może być brytyjski system MiCIS (Microbial Culture Information Service) działający w Laboratory of the Government Chemist, Department of Trade and Industry, lub kanadyjski system, Canadian Federal Collection Database działający w National Research Council of Canada, Halifax. Potrzeba standaryzacji w celu ułatwienia wymiany danych doprowadziła do zorganizowania programu badawczego w krajach Europejskiej Wspólnoty Gospodarczej (Biotechnology Action Programme) dla zaproponowania struktury i organizacji bazy danych, która mogłaby być zaakceptowana przez wszystkie Kolekcje uczestniczące w programie. W ten sposób powstał projekt rozczłonkowanej bazy danych MINE (Microbial Information Network Europe), przechowującej informacje o grzybach i drożdżach (1). Projekt przewiduje zunifikowanie formatu danych i użycie jednakowych komend dla przeszukiwania baz danych mieszczących się w różnych Kolekcjach. Pozwoli to na szybkie i łatwe wydawanie zintegrowanych katalogów drobnoustrojów, a także ułatwi i przyspieszy międzynarodową wymianę infor-

macji. Inną propozycją strukturyzacji danych mikrobiologiczno-biochemicznych jest system MSDN ( Microbial Strain Data Network), który , zaadaptował zaproponowaną przez Rogosę, Krichevsky ego i Colwell (1986) strukturę danych znaną pod nazwą kodu RKC (2). Oprócz baz danych typu bibliograficzno-faktograficznego tworzone są programy wspomagające taksonomię drobnoustrojów, które wykorzystują za podstawę uznane klucze taksonomiczne. Przykładem takiego programu może być YEAST IDENTIFICATION: PC PROGRAM opracowany na podstawie klucza "Yeasts: Characteristics and Identification" przez Barnetta i wsp.(3).

### Charakterystyka ogólna bazy danych MINE

System MINE jest próbą stworzenia bazy danych składającej się z ograniczonej ilości pól (fields), podzielonych na hierarchicznie uporządkowane subpola (subfields), które są wypełniane danymi w postaci pojedynczych symboli lub wyrażane w bardziej lub mniej naturalnym języku. Tworzona w krajach EWG baza danych MINE wykorzystuje pakiet podstawowego oprogramowania "Basis" sprzedawany przez Information Dimensions Inc., Dublin, Ohio, USA. W celu wyboru odpowiedniego sprzętu i oprogramowania należy brać pod uwagę następujące kryteria: (i). duża pojemność i szybkość działania pamięci masowej (dysk twardy), (ii). możliwość tworzenia plików danych różnego typu i ich wzajemnego łączenia, (iii). możliwość używania rekordów danych w postaci pól o zmiennej długości oraz indeksowania rekordów, (iv). łatwą modyfikację struktury bazy danych oraz (v). możliwość zabezpieczenia przed dostępem do danych przez osoby niepowołane. Schemat ogólny podstawowej struktury rekordów oraz ich wzajemne powiązania przedstawia rysunek 1.



Baza składa się z czterech typów rekordów: (i). STR – rekord szczepów (strain record), (ii). SP – rekord rodzaju (species record), (iii). ALT – rekord alternatywny (alternative record) oraz (iv). SYN – rekord synonimów (synonym record). Dla każdego typu rekordów zarezerwowano jedno pole o nazwie RT, które musi zawierać symbol typu rekordu (SP, STR, SYN lub ALT). Każdy rekord posiada unikatowy numer dostępu ACCN (accession number), a wzajemne relacje między rekordami realizowane są przez wpisanie do pola o nazwie APOS (alternate posting) numeru dostępu pliku, który ma być powiązany z danym rekordem. Taka organizacja pozwala na wyszukiwanie i wyświetlanie danych ogólnych o rodzaju jak i pojedynczym szczepie w czasie jednokrotnego przeszukiwania bazy danych. Dla zapewnienia jednoznaczności wprowadzanych danych oraz standaryzacji formatu danych indeksowanych program zawiera 7 słowników zawierających stosowane skróty i oznaczenia, które pozwalają na sprawdzanie poprawności formalnej oraz zamianę nazw skróconych na pełne. Słowniki pozwalają również na zamianę synonimów na nazwy ogólnie akceptowane, a także służą do sprawdzania poprawności wprowadzanych i odczytywanych danych. Standaryzacja danych jest szczególnie ważna dla następujących typów danych: (i). *nazwy gatunkowe grzybów*, (ii). *akronim kolekcji* (iii). *nazwa rośliny żywiciela*, (iv). *pożywki hodowlane* (v). *nazwy związków chemicznych i enzymów* (vi). *nazwy geograficzne*, (vii). *źródła literaturowe*. Szczegółowe informacje na temat źródeł literaturowych uznanych za standardowe dla poszczególnych typów danych podane są w omawianym raporcie.

#### Charakterystyka zestawów danych

W programie MINE rozróżnia się dwa podstawowe zestawy danych, pierwszy MDS, minimalny (minimum data set) i drugi FDS pełny (full data set). Pierwszy zestaw zawiera podstawowe dane o przechowywanym szczepie. Składa się on z 30 pól danych i może być wymieniany pomiędzy poszczególnymi użytkownikami systemu MINE. Zestaw drugi zawiera pełny rekord danych i składa się z 99 pól danych. Katalogi drukowane w systemie MINE zawierają jedynie informacje zawarte w minimalnym zestawie danych.

#### Struktura pól danych

Struktura pól danych powinna być tak zorganizowana aby unikać dwuznaczności jak i powtarzania się danych. Preferowaną praktyką jest rozmieszczanie danych w różnych polach, a nie tworzenie subpól w jednym polu. Niektóre pola są niezbędne ze względu na wewnętrzną strukturę bazy danych, a ich zawartość może być generowana automatycznie przez system, np. data wprowadzenia danych. Dane tego typu są zgrupowane w bloku o nazwie "Informacje wewnętrzne". Dla zwiększenia przejrzystości systemu dane tego samego typu zgrupowane są w odpowiednie bloki informacyjne. Możliwość definiowania subpól pozwala na wprowadzenie kilkunastu danych do jednego pola. Subpola, podobnie jak pola, mogą być również indeksowane, gdyż pozwala na to oprogramowanie podstawowe BASIS. Dla zorganizowania systemu strukturalnego, zastosowano podział pól na subpola przez użycie następujących ograniczników: przecinek, średnik, dwukropek, znak mniejszości, nawiasy itp. Informacje zawarte w bazie danych dotyczą przechowywanych w kolekcjach szczepów i pochodzą ze źródeł publikowanych i nie publikowanych. Baza ta nie ma charakteru taksonomicznego, ale może być również wykorzystana do tego celu.

## Proponowana struktura pól danych

Wspomniano już, że pełny zestaw danych zawiera 99 pól informacyjnych, które ze względu na podobieństwo informacji podzielono na 12 bloków. Zakres tematyczny informacji przypisanych poszczególnym blokom przedstawiono poniżej:

1. Informacje wewnętrzne (Internal administration)
2. Nazwa (Name)
3. Informacje podstawowe o szczepie (Strain administration)
4. Status (Status)
5. Środowisko i historia szczepu (Environment and History)
6. Oddziaływania biologiczne (Biological interactions)
7. Płciowość (Sexuality)
8. Właściwości (Properties)
9. Genotyp i genetyka (Genotype and Genetics)
10. Warunki hodowlane (Growth conditions)
11. Chemia i enzymy (Chemistry and Enzymes)
12. Zastosowanie praktyczne (Practical application)

Poszczególne bloki tematyczne zawierają następujące pola danych:

## 1. Informacje wewnętrzne

ACCN	- Numer dostępu rekordu (Accession number of the record)
APOS	- Przypisanie alternatywne (Alternate posting)
RT	- Typ rekordu (Record type)
EDI	- Data wprowadzenia danych (Data of input)
EDM	- Data aktualizacji danych (Date of update)
SEC	- Kod zabezpieczający (Security code)
ID	- Znak identyfikacyjny operatora (Typist's identification)
OT	- Typ organizmu (Organism type)

## 2. Nazwa

SP	- Nazwa gatunkowa, autorzy (Species name, authors)
SSP	- Nazwa podgatunku, autorzy (Subspecies name, authors)
VAR	- Nazwa odmiany, autorzy (Variety name, authors)
F	- Nazwa formy, autorzy (Form name, authors)
FSP	- Specjalna nazwa formy, autorzy (Special form name, authors)
RACE	- Nazwa rasy, autorzy (Race name, authors)
SERO	- Serotyp, seroodmiana (Serotype, serovar)
MIS	- Nazwa nieużywana, autorzy (Misapplied name, authors)
TAX	- Taksonomia (Taxonomy)
MOR	- Morf (Morph)
ILL	- Ilustracje (Illustrations)
TAXREM	- Komentarz taksonomiczny (Taxonomic remarks)
LIT	- Literatura taksonomiczna i inna (Taxonomic and other literature)

## 3. Informacje podstawowe o szczepie

- STN - Numer szczepu (Strain number)
- OCC - Numery w innych kolekcjach (Numbers in other collections)
- EDA - Data dostępu (Date of accession)
- EDR - Data otrzymania (Date of receipt)
- CORR - Korespondencja (Correspondence)
- SPINF - Dostępność informacji specjalnych (Special information available)
- REM - Uwagi ogólne (General remarks)
- CHK - Wewnętrzne sprawdzenie identyczności przez kolekcję (Internal check of identity at collection)
- PRE - Sposób przechowywania (Mode of preservation/storage)
- PREREM - Uwagi do przechowywania (Remarks on preservation)
- RESTR - Ograniczenia, opieka i środki ostrożności (Restrictions, care and precautions)

## 4. Status

- STAT - Status (Status)

## 5. Środowisko i historia szczepów

- SUSP - Specyficzność substratowa (Substratum specificity)
- SSTR - Substrat oryginalny (Original substratum)
- HAB - Środowisko występowania (Habitat)
- LOC - Pochodzenie/umiejscowienie materiału oryginalnego (Origin/location of original material)
- COLL - Kto zebrał? (Collector)
- HERB - Herbarium, gdzie specimen jest zdeponowany (Herbarium where specimen is deposited) |
- SOL - Kto izolował? (Isolated by)
- ISOM - Źródło i metoda izolacji (Source and method of isolation)
- DET - Kto zidentyfikował? (Identified by)
- DEP - Kto zdeponował? (Depositor)
- HIS - Historia (pomiędzy izolacją i zdeponowaniem) (History (between isolation and deposit))
- NAMCH - Zmiany nazwy/nazwy poprzednie (Name changes/previous names)
- ECOLIT - Literatura na temat ekologii (Literature on ecology)
- ECOREM - Komentarze na temat ekologii i środowiska (Remarks on ecology and environment)

## 6. Oddziaływania biologiczne

- SYMB – Symbioza (Symbiosis)  
 MPAR – Mykoparazytyzm (Mycoparasitism)  
 PATH – Patogenność (Pathogenicity)  
 ALLER – Allergenność (Allergenicity)  
 TOX – Toksyczność dla innych organizmów  
 (Toxicity to other organisms)  
 ANT – Oddziaływania antagonistyczne w stosunku do innych  
 organizmów (Antagonistic activities against other organisms)

## 6. Płciowość

- SEX – Zachowania płciowe (Sexual behaviour) SEXST – Sexual state

## 7. Właściwości

- NUCL – Liczba jąder (Number of nuclei)  
 FSTR – Dane dotyczące struktury wewnętrznej (Fine structure data)  
 WCONS – Skład ściany komórkowej (Wall constituents)  
 CCONT – Zawartość komórki (Cell contents)  
 COQ – Typ koenzymu Q (Coenzyme-Q system)  
 STAIN – Reakcje barwienia (Staining reactions)  
 PIGMENT – Produkcja pigmentu i autofluorescencja  
 (Pigment production and autofluorescence)  
 GRWTH – Charakterystyki wzrostu (Growth characteristics)  
 TKIT – Wyniki oznaczeń przy użyciu zestawów testowych  
 (Results of test-kit)  
 PROPLIT – Literatura na temat właściwości szczepu  
 (Literature on strain properties)

## 9. Genotyp i genetyka

- GENOT – Genotyp (Genotype)  
 GC – Zawartość GC w DNA (Guanine-cytosine content of DNA)  
 DNAD – Hybrydyzacja DNA-DNA z innymi szczepami  
 (Hybridization DNA-DNA with other strains)  
 DNAR – Hybrydyzacja DNA – rRNA z innymi szczepami  
 (Hybridization DNA-rRNA with other strains)  
 RNAD – Hybrydyzacja rRNA-DNA z innymi szczepami  
 (hybridization rRNA-DNA with other strains)  
 MUT – Mutanty (Mutants)  
 MUTMET – Metody indukcji mutacji  
 (Methods by which mutation was induced)  
 HYBR – Hybrydy (Hybrids)  
 PLS – Plazmidy w komórce i poza komórką gospodarza  
 (Plasmid out/in the host)

- KIL – Właściwości zabójcze drożdży (Killer properties of yeasts)  
 GENREM – Komentarze na temat genetyki (Remarks on genetics)  
 GENLIT – Literatura na temat genetyki i mutantów  
 (Literature on genetics and mutants)

## 10. Warunki hodowli

- CONDS – Warunki wzrostu i przechowywania na pożywkach stałych  
 (Conditions for growth and maintenance on solid media)  
 CONDL – Warunki wzrostu w pożywkach płynnych  
 (Conditions for growth in liquid media)  
 CONDSP – Warunki dla owocowania lub sporulacji  
 (Conditions for fruiting or sporulation)  
 CONDGER – Warunki kiełkowania (Conditions for germination)  
 CSOR – Źródła węgla dla wzrostu (Carbon sources for growth)  
 NSOR – Źródła azotu dla wzrostu (Nitrogen sources for growth)  
 CNSOR – Pojedynczy związek badany jako jedyne źródło węgla i azotu  
 (Single compound tested as sole C and N source)  
 NUGR – Wymagania pokarmowe i czynniki wzrostowe  
 (Nutritional requirements and growth factors)  
 DEF – Inne wymagania pokarmowe (Deficiencies)  
 TOL – Tolerancja i wrażliwość (Tolerances and sensitivities)  
 TEMPR – Wpływ temperatury na wzrost  
 (Temperature relationships for growth)  
 TEGR – Temperatury kardynalne dla wzrostu  
 (Cardinal temperatures for growth)  
 TESP – Optymalny zakres temperatury dla sporulacji  
 (Temperature optimum range for sporulation)  
 HEATR – Oporność termiczna (Heat resistance)  
 PHR – Wymagania w zakresie pH dla wzrostu  
 (pH requirements for growth)  
 PHC – Warunki pH (pH conditions)  
 ETHC – Optymalne i maksymalne stężenie etanolu dla wzrostu  
 (Ethanol conditions)  
 SALR – Optymalne zasolenie dla wzrostu (Salinity requirements)  
 SALT – Maksymalne stężenie soli tolerowane przez organizm  
 (NaCl concentrations)  
 SUGC – Optymalne i maksymalne stężenie cukru  
 (Optimum and maximum sugar concentrations)  
 OSM – Osmofilność i kseroofilność (Osmophily and xerophily)  
 WATC – Aktywność wodna środowiska (Conditions of water activity)  
 LC – Warunki oświetlenia (Light conditions)

## 11. Chemia i enzymy

- ENZ – Produkcja enzymów (Enzymes produced)
- DEC – Uzdolnienia do rozkładu biopolimerów  
(Decomposition and deteriorating capacities)
- MET – Produkcja metabolitów (Metabolites produced)
- BIOTR – Biotransformacje (Biotransformations)
- CHREM – Komentarz do danych chemicznych (Remarks on chemical data)
- CHLIT – Literatura na temat wpływu czynników fizycznych i chemicznych  
(Literature on the impact of chemical and physical factors)

## 12. Zastosowanie praktyczne

- APPL – Zastosowanie przemysłowe i ogólne  
(Industrial and general applications)
- TEST – Zastosowanie do testów biologicznych (Application in testing)
- BIOM – Produkcja biomasy, przydatność do spożycia  
(Production of biomass, edibility)
- PATENT – Patenty [Patent(s)]
- CTR – Zwalczanie danego drobnoustroju  
(Control of the microorganism in question)
- APLIT – Literatura na temat zastosowań, patenty itp.  
(Literature on application, patents, etc.)

**System kodowania danych biologicznych**

Całkowicie odmienny sposób organizacji bazy danych przechowującej informacje o drobnoustrojach zaproponowany został przez Rogosę, Krichevsky'ego i Colwell (1986). Polega on na zakodowaniu każdej cechy biologicznej za pomocą sześciocyfrowej liczby. Trzy pierwsze cyfry tej liczby kodują jedną grupę cech podobnych a trzy pozostałe poszczególne cechy w obrębie tej grupy. Zastosowanie liczb sześciocyfrowych pozwala na zakodowanie 999 grup różnych cech oraz 999 pojedynczych cech w obrębie każdej grupy. Poniższy fragment ilustruje sposób kodowania zabarwienia spodniej strony kolonii drobnoustrojów:

KOD	Opis cechy
020009	: Underside of colony is pigmented
020066	: Underside of colony is blue
020067	: Underside of colony is yellow (golden)
020068	: Underside of colony is green
020069	: Underside of colony is red
020070	: Underside of colony is orange
020071	: Underside of colony is violet (purple)
020072	: Underside of colony is brown
020073	: Underside of colony is black



Trzy pierwsze cyfry kodu (020) kodują grupę cech o wspólnej nazwie "Pigmenty i zapachy", a pozostałe trzy cyfry, pojedyncze stwierdzenie np. "Underside of colony is green". Oprócz cech opisowych, system pozwala również na kodowanie cech ilościowych takich jak np. stopień oporności na antybiotyki. W tym przypadku, trzema końcowymi cyframi kodu przyporządkowane są stężenia antybiotyku będące kolejnymi potęgami liczby 2 ( $2^0 - 2^{11}$ ), co odpowiada stężeniom od 1 do 2048 jedn. lub  $\text{mg}/\text{cm}^3$  (tab.1)

Tabela 1

Sposób kodowania ilościowego oporności na antybiotyki według Rogosy i wsp. (2).

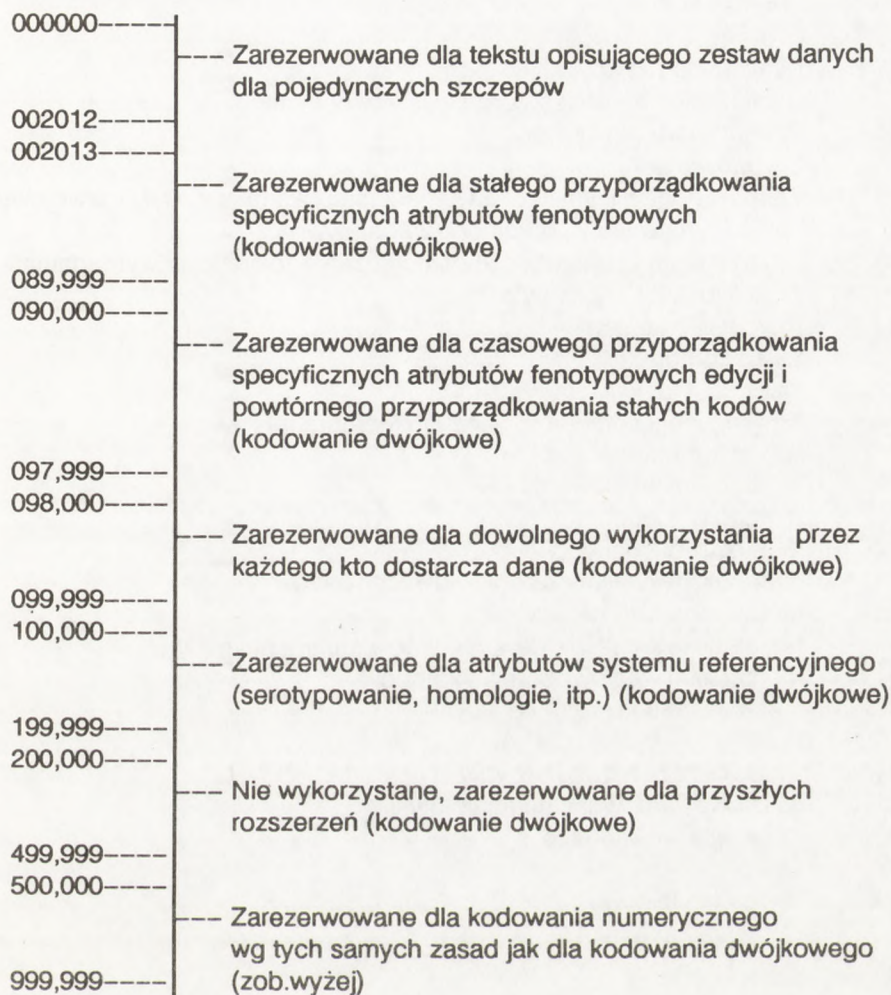
Antybiotyk	Stężenie [ $\text{mg}/\text{cm}^3$ ]	Kod
Ampicylina	1 . . . . .	.019534
	2 . . . . .	.019535
	4 . . . . .	.019536
	8 . . . . .	.019537
	16 . . . . .	.019538
	32 . . . . .	.019539
	64 . . . . .	.019540
	128 . . . . .	.019541
	256 . . . . .	.019542
	512 . . . . .	.019543
	1024 . . . . .	.019544
	2048 . . . . .	.019545

System przewiduje także stosowanie innych wartości kolejnych stężeń antybiotyków.

Sposób strukturyzacji danych kodowanych cyfrowo przedstawia schemat 1. Wynika z niego, że autorzy dotychczas zaproponowali 90 podstawowych grup cech biochemiczno-fizjologicznych i hodowlanych. Przewidywana w tym systemie liczba grup cech kodowanych dwójkowo wynosi 500. Pozostałe numery od 500 do 999 zarezerwowane zostały na kodowanie tych samych cech w sposób numeryczny.

## S c h e m a t 1

## Strukturyzacja danych kodowanych cyfrowo



## Struktura bazy danych typu RKC

1. Informacje ogólne
2. Informacje specyficzne o szczepie
3. Morfologia pojedynczych komórek
4. Wymiary pojedynczych komórek wegetatywnych
5. Nierozpuszczalne wewnątrz- i zewnątrzkomórkowe substancje zapasowe

6. Endospory i cysty
7. Myksospory, sporocysty, ciała owocujące, spory, sporozoity, trofozoity i gametocyty
8. Rozgałęzienia, strzępki i tworzenie spor wegetatywnych
9. Wyrostki (Styliki)
10. Pochewki
11. Otoczki
12. Reakcje barwienia
13. Ruchliwość, urzęsienie i organelle zewnętrzne
14. Sposób podziału komórki
15. Ugrupowania
16. Warunki hodowli, inhibitory, wymagania pokarmowe, Cykl rozwojowy
17. Wpływ temperatury na komórki wegetatywne
18. Chlorek sodu i inne czynniki osmotyczne – tolerancja i wymagania
- 19./40. Wrażliwość na antybiotyki
20. Pigmenty i zapachy
21. Skład komórek
22. Liza
23. Powierzchnia komórki (Sciana lub Membrana)
24. Reakcje metaboliczne
25. Metabolizm węglowodanów
26. Metabolizm alkoholi
27. Metabolizm aldehydów
28. Metabolizm kwasów karboksylowych i estrów
29. Metabolizm aminokwasów
30. Metabolizm amin, amidów, laktamów, puryn, pirymidyn
31. Metabolizm węglowodorów i ketonów
32. Metabolizm tłuszczów i olejów
33. Przechowywanie szczepów
34. Szlaki metaboliczne i enzymy
35. Ilościowa wrażliwość na antybiotyki
36. Organelle wewnętrzne
37. Jądro
38. Kwasy nukleinowe

#### System referencyjny

101. Salmonella
103. Streptococcus
105. Staphylococcus
106. Escherichia
107. Campylobacter

Wprowadzenie ogromnej ilości danych do takiej bazy wymaga zorganizowania sprawnego systemu dostarczającego danych wejściowych i takiej organizacji aby uniknąć błędów i powtórzeń. Dla realizacji tego celu opracowano standardowe formu-

larze, które wypełnia się danymi pochodzącymi z doświadczeń przeprowadzanych na badanych szczepach drobnoustrojów. Przedstawiona przez autorów systemu RKC propozycja struktury bazy danych jest bardzo wygodna z punktu widzenia techniki komputerowej, gdyż pozwala na proste i łatwe jej przeszukiwanie, ale wymaga wiele ludzkiej pracy na etapie kodowania danych do wprowadzenia.

### Baza danych wspomagająca taksonomię drobnoustrojów

W oparciu o klucz "Yeasts: Characteristics and Identification" (3) wykonana została przez autora artykułu, komputerowa baza danych taksonomicznych dla drożdży. Baza zawiera informacje o 469 gatunkach drożdży zgrupowanych w 62 rodzaje. Każdy gatunek opisany jest za pomocą 83 testów biochemicznych podzielonych na następujące grupy: fermentacja węglowodanów (13 testów), asymilacja źródeł węgla (40 testów), asymilacja źródeł azotu (7 testów), wymagania pokarmowe (10 testów) oraz różne (13 testów). Ponadto baza zawiera informacje opisowe o cechach fizjologicznych, takich jak, sposób rozmnażania, sposób pączkowania, barwa kolonii, tworzenie zarodników, ich ilość i kształt, tworzenie grzybni i pseudogrzybni, a także procentową zawartość guaniny i cytozyny w DNA. Dostęp do bazy realizowany jest przez wybór jednej z dziesięciu opcji. Pierwsza z nich wyprowadza na ekran, listę 62 rodzajów drożdży, druga listę 469 gatunków, a trzecia listę 83 testów biochemicznych. Opcja czwarta i piąta pozwala na wyprowadzenie na ekran i drukarkę, wyników testów biochemicznych i cech fizjologicznych dla poszczególnych gatunków, przy czym jako parametr wejściowy może być użyta nazwa rodzajowa i gatunkowa dla danego szczepu lub też jego numer. Wyżej wymienione opcje są częścią informacyjną bazy danych. Część druga, użytkowa, pozwala na wyszukiwanie w bazie danych, nazw rodzajowych i gatunkowych drożdży spełniających określone kryteria. Służą do tego opcje od 6 do 9. Poszukiwanie rozpoczyna się od utworzenia rekordu danych do porównania w postaci pliku tekstowego o długości 83 znaków za pomocą opcji 7. Nie jest konieczne wybranie wszystkich cech z całego zestawu danych. Po utworzeniu tego rekordu, porównuje się jego zawartość z rekordami bazy danych i wyszukuje numery i nazwy szczepów najbardziej podobne dożądanego wzorca. Osobnym problemem w konstruowaniu algorytmu poszukiwań było porównywanie wzorca z niejednoznacznymi danymi wejściowymi dla określonych cech i szczepów. Na przykład, drożdże *Aciculoconidium aculeatum* mogą, według klucza, fermentować glukozę po 24 godz. hodowli, a także dopiero po 3 dobach. Cechę tę zaznaczono w kluczu dwoma znakami +,D, gdzie D oznacza reakcję opóźnioną (delayed). W podstawowej bazie danych dwuznakowa cecha +,D została zakodowana za pomocą znaku Q. Algorytm porównania znaku kodującego fermentację glukozy i wprowadzonego rekordu danych, który może zawierać jeden z sześciu znaków ( +, -, D, S, W, ? ), powinien zarówno dla znaku + jak i D, dawać wynik pozytywny, a dla pozostałych znaków, wynik negatywny. Taki sposób postępowania pozwolił na rozwiązanie problemu tzw. cech niejednoznacznych. Porównanie wprowadzonego rekordu z rekordami bazy danych dokonywane jest za pomocą opcji 6. W wyniku jej działania otrzymuje się numer oraz nazwę rodzajową i gatunkową najbardziej podobnego gatunku, a także stopień podobieństwa do wprowadzonych cech. Opcje 8 i 9 służą do przechowywania wprowadzanych rekordów wzorcowych na nośniku magnetycznym w celu ich późniejszego wykorzystania, a opcja 10 kończy działanie programu. Program został napisany w języku GWBASIC i skompilowany w celu przy-

spieszenia jego działania. Przy posługiwaniu się nim należy zwracać szczególną uwagę na ważność poszczególnych cech biochemicznych w taksonomii, gdyż może to prowadzić do niejednoznaczności. W związku z tym, że program porównuje tylko cechy biochemiczne (a nie uwzględnia cech fizjologicznych) to można się spotkać z sytuacją, iż dane biochemiczne uzyskane w wyniku badania drożdży zarodnikujących są w większym stopniu zgodne z danymi dla innego gatunku drożdży niezarodnikujących i w efekcie porównania wyszukana zostanie nazwa tych drożdży. Wydaje się zatem, że w dalszej pracy nad udoskonaleniem programu należy również uwzględnić cechy fizjologiczne. Praktyka posługiwania się programem wskazuje, że optymalnym rozwiązaniem dla komputerowej taksonomii drożdży byłaby metoda dwustopniowego przeszukiwania bazy danych. W pierwszym etapie należałoby utworzyć rekord cech do porównania o najwyższym priorytecie w taksonomii (fermentacja glukozy, asymilacja etanolu, azotanów, ureaza, barwienie błękitem dwuazoniowym, zarodnikowanie, sposób rozmnażania). Następnie porównać wprowadzone dane z danymi bazy podstawowej i określić rodzaje drożdży najbardziej podobne do wzorca. W kolejnym etapie utworzyć pełniejszy rekord danych uwzględniający najistotniejsze cechy biochemiczne i porównać go z rekordami wybranych rodzajów a nie z całą bazą. Taki sposób postępowania jest zwykle realizowany w praktyce przy określaniu przynależności taksonomicznej nowo wyizolowanych drożdży. Autorzy klucza "Yeasts: Characteristics and Identification" (3) opracowali własny program o nazwie "YEAST IDENTIFICATION: PC PROGRAM", który wykonuje następujące operacje: identyfikuje drożdże oraz wyszukuje gatunki o zadanej charakterystyce. Po wprowadzeniu danych eksperymentalnych, program "listuje" wszystkie gatunki, które całkowicie odpowiadają wprowadzonym danym, oraz te, które w przybliżeniu spełniają założenia podając jednocześnie występujące różnice. Ponadto program\* dostarcza informacji, jakie testy trzeba dodatkowo wykonać aby całkowicie zakończyć identyfikację.

\*Program ten jest dostępny u autorów, a zamówienia należy kierować na adres:  
Dr J.A. Barnett 36 Le Strange Close Norwich NR2 3PW, UK.

## Literatura

1. Gams W., Hennebert G.L., Stalpers J.A., Janssens D., Schipper M.A.A., Smith J., Yarrow D. and Hawksworth D.L., (1988), J. Gen. Microbiol., Structuring Strain Data for Storage and Retrieval of Information on Fungi and Yeasts in MINE, the Microbial Information Network Europe, 134, 1667-1689.
2. Rogosa M., Krichevsky M.I., Colwell R.R., (1986), Coding Microbiological Data for Computers, Starr M.P., Ed. (1986), Springer-Verlag New York-Berlin-Heidelberg-London-Paris-Tokyo.
3. Barnett J.A., Payne R.W., Yarrow D., (1987), Yeast Identification PC Program version 1. based on data (updated and revised) from their book: Yeasts: Characteristics and Identification (1983), Cambridge University Press.