Computer-assisted identification and clustering for regulated phytopathogenic bacteria: construction of a reference database and development of a computer system*

F. Poliakoff[1], M. Fischer-Le Saux[2], G. Hunault[3], C. François[1], C. Rivoal[1], S. Bonneau[2], S. Perret[2], H. Soubelet[2] and D. Caffier[1]

[1]*DRAF-SRPV Pays de la Loire, LNPV Unité Bactériologie, 7 rue J. Dixméras, 49044 Angers Cedex 1 (France); e-mail: francoise.poliakoff@agriculture.gouv.fr*
[2]*UMR Pathologie Végétale, INRA, BP 57, 42, rue Georges Morel, 49071 Beaucouzé Cedex (France)*
[3]*Université d'Angers, UFR Sciences, 2 Bd Lavoisier, 49045 Angers Cedex 01 (France)*

CABIQ is a specific computer-assisted identification system for the reliable and rapid identification of the main regulated phytopathogenic bacteria. It is based on phenotypic and genomic properties of bacteria. About 500 reference strains have been used to initiate the database, including conventional phenotypic tests and the Biotype 100 (BioMérieux) galleries. The CABIQ system, with its database and reference matrices, is a guide on the tests to be done when identifying new isolates. The modules dealing with phenotypic and Biotype100 characteristics are already finalized. Results on repetitive PCR will soon be added to the system. This opens interesting perspectives for improving the reliability of computer-assisted identification.

Introduction

The phytosanitary regulations of the European Union, in EU Directive 2000/29 (EU, 2000), require all member states to survey for certain quarantine pests and to prohibit their introduction and spread into the EU. For regulated bacteria, an accurate and reliable system is therefore needed for diagnosis and identification at specific and infraspecific levels. In France, a project has jointly been elaborated by INRA (the National Agricultural Research Institute), the bacteriology laboratory (LNPV-UB) of the French NPPO, and the University of Angers, under the overall control of the Ministry of Agriculture, to create a computer tool for reliable identification of the main regulated phytopathogenic bacteria. After isolation and selection of the most likely pathogenic colonies, the determination of the genus, species or pathovar can be achieved by adequate simple tests, with an automated identification system using a genomic and phenotypic database. This tool will be released as a Windows-based system called CABIQ ('Classification Automatique Bactéries Identification Quarantaine') incorporating the database. This tool is innovative because it combines traditional identification tests and molecular methods in a single system to provide a more rapid and reliable identification.

*Paper presented at the EPPO Conference on Quality of Diagnosis and New Diagnostic Methods for Plant Pests (Noordwijkerhout, NL, 2004-04-19/22).

Development of the system

The project was developed through the following steps:

- creation of a large culture collection of strains of regulated Gram-positive and Gram-negative bacteria, together with representative non-regulated phytopathogenic bacteria, from the French Collection of Phytopathogenic Bacteria (CFBP)
- optimization and standardization of known phenotypic tests in tubes or Petri dishes, of Biotype 100 galleries and of repetitive PCR
- selection of the most pertinent phenotypic tests to differentiate genera and species, based on published data and on the expertise of the project partners
- screening of the culture collection with all selected tests
- entry of data into the databases via specific menus, plotting and interpretation of the results.

Culture collection and conditions

Fourteen taxa of regulated bacteria (Table 1) were selected from the EU list, along with representative species of close taxonomic groups, to create a collection of nearly 500 reference strains. For each regulated taxon, 15 strains on average were selected including the type strain. Strains were selected from ecological and geographical origins which were as different as possible, so that the collection should be representative of potential natural diversity. Strains from the CFBP, kept dried in phials, were revived and spread on YPGA nutrient medium in Petri dishes. The incubation temperature depended on each bacterium.

Table 1 List of regulated bacteria studied for the creation of the database

Gram group	Genus or family	Regulated taxon
+	*Clavibacter*	*C. michiganensis* subsp. *insidiosus*
		C. michiganensis subsp. *michiganensis*
		C. michiganensis subsp. *sepedonicus*
+	*Curtobacterium*	*C. flaccumfaciens* pv. *flaccumfaciens*
–	*Pseudomonas*	*P. syringae* pv. *persicae*
–	*Xanthomonas*	*X. arboricola* pv. *pruni*
		X. axonopodis pv. *citri*,
		X. axonopodis pv. *aurantifolii*
		X. axonopodis pv. *citrumelo*
		X. fragariae
		X. arboricola pv. *fragariae*
–	*Ralstonia*	*R. solanacearum*
–	*Enterobacteriaceae*	*Pantoea stewartii* subsp. *stewartii*
		Erwinia amylovora

Optimization and evaluation of methods for phenotypic characterization

Numerous biochemical characterization tests exist for bacteria (Lelliott & Stead, 1987; Schaad *et al.*, 2001). Initial identification tests for common genera, and discriminating tests for species and subspecies, were selected. For each strain of the collection, two different galleries of tests were defined and realized: (1) the CARGEN gallery, composed of general tests (mainly phenotypic tests), orientation tests in tubes or Petri dishes (adapted to each Gram group), and identification tests related to each genus and later species; (2) the BIOTYPE gallery, in the form of Biotype 100 from BioMérieux, a standardized method for establishing nutrient profiles. The resultant set of phenotypic characters combines all the morphological, growth, physiological, enzymatic, metabolic and biochemical characteristics. Identification is based on all these properties. Relatively few differential characters are required to identify the commonly isolated genera of plant pathogenic bacteria.

General tests

All tests were carried out on young cultures (24–48 h) grown on YPGA or on King's medium B and incubated at the optimum temperature for the bacterium concerned. The tests were as follows:

- Gram determination: Gram-staining is always done first when identifying an unknown bacterium. Gram-positive and Gram-negative bacteria must be differentiated. Because of its good correlation with Gram staining for plant pathogenic taxa, Gram type was determined by the test of solubility in 3% KOH solution (Suslow *et al.*, 1982).
- Colour and mucosity of colonies
- Production of diffusible pigments: indigoïdine is a blue extracellular, water-insoluble pigment produced on glucose yeast extract medium containing calcium carbonate. Pyoverdines are green, water-soluble fluorescent pigments, produced on iron-deficient media, revealed by fluorescence under UV irradiation at a wavelength of 365 nm. For elicitation of this property, strains are spread onto King's medium B (King *et al.*, 1954) and incubated for 48 h at the optimum temperature for the bacterium concerned.
- Host
- Symptom(s) of the disease
- Hypersensitivity reaction (HR) on tobacco: bacterial suspension of approx. 10^9 cfu mL^{-1} prepared from 24-h colonies is infiltrated into the leaf parenchyma (Klement, 1963). The result is observed after 24 h. Three results are possible: no reaction (HR negative), necrosis of the infiltrated parenchyma (HR positive) and disease on tobacco (the strain is then pathogenic to tobacco).

Orientation tests within each Gram group (Tables 2 and 3)

- Glucose metabolism on Hugh and Leifson medium (oxidative (O) or fermentative (F) ways)
- Enzymatic activities
- Production of acid from different carbon sources: published data generally give contradictory information about assimilation of carbon sources. This can be explained by the use of different media and by various testing procedures. In this study, use of carbohydrates by Gram-positive bacteria was tested in two different media: ARJ (Ayers *et al.*, 1919) and medium C of Dye (Dye, 1968). ARJ medium is a mineral medium, while medium C, which contains complex substrates, is more commonly used for bacteria with complex trophic needs. Based on our results,

Table 2 Rules to determine the probable genus of a Gram-negative bacterium according to the results of orientation tests

Test results	Probable taxon
Fermentative on Hugh and Leifson (+)	*Enterobacteria* (*Erwinia*/*Pantoea*)
Non-fermentative on Hugh and Leifson (–), fluorescence on King B (+)	*Pseudomonas*
Non-fermentative on Hugh and Leifson (–), fluorescence on King B (–), nitrate reductase (+), oxidase (+)	*Ralstonia*
Non-fermentative on Hugh and Leifson (–), fluorescence on King B (–), nitrate reductase (–), no oxidase (–), cream colonies	*Pseudomonas*
Non-fermentative on Hugh and Leifson (–), fluorescence on King B (–), nitrate reductase (–), no oxidase (–), yellow colonies, mucoid culture	*Xanthomonas*

Table 3 Rules to determine the probable genus of a Gram-positive bacterium according to the results of orientation tests

Test results	Probable genus
Urease (+), no esculin (–), orange colonies on YPGA	*Rhodococcus*
Urease (–), esculin (+), acid from erythritol (+), acid from ribose (+), acid from adonitol (+)	*Curtobacterium*
Urease (–), esculin (+), acid from erythritol (–), acid from ribose (–), acid from adonitol (–)	*Clavibacter*/*Rathayibacter*

Table 4 Identification tests used within each probable genus of Gram-negative bacteria to discriminate strains at specific and subspecific levels

Code of test	*Erwinia*/*Pantoea*	*Pseudomonas*	*Xanthomonas*[1]	*Ralstonia*[1]
1	Pectin hydrolysis on Sutton's medium	Pectin hydrolysis	Esculin	Arginine (Thornley)
2	Mobility in water	Levan production	Gelatinase (Frazier)	Gelatinase (Frazier)
3	Growth at 36 °C	Arginine (Thornley)	Starch hydrolysis	Levan production
4	Growth at 39 °C	Acid from sucrose	Tween 80 hydrolysis	Esculine
5	Reducing compounds from sucrose		Motility Mannitol	Tween 80 hydrolysis
6	Gelatinase (Frazier)		Urease	Starch hydrolysis
7	Acid from D(+) arabitol		Catalase	Growth in NaCl 1%
8	Acid from α-methyl glucoside		Acid from adonitol	Growth in NaCl 2%
9	Acid from melibiose		Acid from arabinose	Growth at 40 °C
10	Acid from maltose		Acid from erythritol	Urease
11	Acid from L-Rhamnose		Acid from fructose	Tryptophane desaminase
12	Acid from Sorbitol		Acid from glucose	Tyrosinase
13	Malonate utilization		Acid from glycerol	Acid from cellobiose
14			Acid from lactose	Acid from dulcitol
15			Acid from maltose	Acid from erythritol
16			Acid from mannitol	Acid from galactose
17			Acid from ribose	Acid from glucose
18			Acid from sucrose	Acid from inositol
19			Acid from sorbitol	Acid from lactose
20				Acid from maltose
21				Acid from mannitol
22				Acid from mannose
23				Acid from ribose
24				Acid from sucrose
25				Acid from sorbitol
26				Acid from trehalose

[1]For this probable genus, studies are in progress and all the tests performed are presented

medium C of Dye was retained as the most appropriate for Gram-positive bacteria. ARJ medium was used for Gram-negative bacteria.

Identification tests for each genus

Various phenotypic tests were performed in Petri dishes or tubes for all the strains of the culture collection. Numerical analysis of the results allowed us to select those showing greatest discrimination in identifying regulated taxa. These selected tests are presented in Tables 4 and 5.

Table 5 Identification tests used within each probable genus of Gram-positive bacteria to discriminate strains at specific and subspecific levels

Code	*Rhodococcus*	*Curtobacterium*	*Clavibacter*/*Rathayibacter*
1	(none)	Acid from lactose	Acid from lactose
2		Acid from sorbitol	Acid from sorbitol
3		Host	Acid from mannitol
4			Gelatinase (Frazier)
5			NaCl 5% tolerance (21 ± 3 °C)
6			Levan production
7			Host

Characterization using the Biotype 100 gallery

Biotype 100 (Biomérieux) is a research product for the establishment of nutritional profiles, generally used for Gram-negative bacteria. The Biotype 100 strip contains 100 tubes for the study of the metabolism of carbon sources using assimilation tests. Tubes are inoculated with an assimilation medium adapted to the nutritional requirements of the organism to be examined: liquid Biotype Medium 1 (containing 16

growth factors) for most Gram-negative bacteria, or liquid Biotype Medium 2 (31 growth factors) for more fastidious bacteria. After incubation, reactions are read visually, either as a cloudy appearance or as the development of a colour.

The method had first to be optimized, particularly for Gram-positive bacteria. Biomedium, concentration of bacterial suspensions, gallery incubation temperatures and durations were studied to determine the appropriate combination. We found that Biotype Medium 1 is suitable for Gram-negative bacteria and Biotype Medium 2 for Gram-positive bacteria.

A suspension with a turbidity of 3 on McFarland's scale is prepared from a pure culture of the bacterium under investigation. 2 mL of this suspension is transferred into Biotype Medium. After homogenization, the tubes are filled with suspension and incubated under aerobic conditions, at 28 °C for Gram-negative bacteria, or at 26 °C for Gram-positive bacteria. Strips are read after 2, 4 and 6 days for Gram-negative bacteria, and after 5, 10, 15 and 20 days for Gram-positive.

Genomic identification and characterization

Three DNA families unrelated at the DNA sequence level (REP, ERIC and BOX[1]) amplified by PCR give products that reflect the number and distribution of repetitive sequences or the genomic fingerprinting of bacteria (Louws *et al.*, 1994). This approach (rep-PCR) offers a highly sensitive level of discrimination at the species level, and even to some extent at the pathovar, biovar or subspecies levels. REP-, ERIC- and BOX-PCRs were used on both the Gram groups. The most suitable rep-PCR in the case of Gram-positive bacteria was BOX-PCR, carried out with primer BOX A1R. In the case of Gram-negative bacteria, REP-PCR or ERIC-PCR were more convenient, using primers REP1R-I and REP2-I, or ERIC 1R and ERIC2.

BOX-PCR was carried out in 25-µL volumes containing 1× DNA polymerase buffer, 1.5 mM $MgCl_2$, 2 µM BOXA1R primer, 0.8 mM dNTP and 2 U of DNA polymerase (Taq Goldstar, Eurogentec) with an initial denaturation step at 95 °C for 7 min followed by cycles of 94 °C (60 s), 53 °C (60 s), 65 °C (8 min). After 30 cycles, a final extension was done at 65 °C for 16 min.

REP and ERIC-PCR were carried out with the same kind of mix but at 8 µM concentration of each primer (REP1R-I and REP2-I, or ERIC 1R and ERIC2). DNA amplification consisted of an initial denaturation step at 95 °C for 5 min followed by cycles of 94 °C (60 s), 40 °C (60 s) for REP-PCR and 52 °C (60 s) for ERIC-PCR, 65 °C (5 min). After 30 cycles, a final extension was done at 65 °C for 15 min. PCR fragments were separated by electrophoresis on agarose gel (Seakem or Eurobio 1.5%) at 6 V per cm, stained with ethidium bromide, and photographed under UV light. The technique, and the imaging of gels, needed to be carefully standardized before the repetitive profiles could be entered into the CABIQ database, to ensure that results coming from different laboratories could be compared. The interface with CABIQ already existed.

[1]REP = Repetitive Extragenic Palindromic, ERIC = Enterobacterial Repetitive Intergenic Consensus, Box = Repetitive Elements Box.

Numerical analysis of phenotypic features with the CABIQ software

Microbiological tests are subject to experimental errors, even under conditions of strict standardization. The main advantage offered by numerical computerized techniques is their repeatability and objectivity in handling and analysing large sets of data. CABIQ proposes a simple and user friendly window interface to:

- introduce, read and manage data, together with other files linked to galleries, bacterial characters (reference strains and isolates) as well as identification matrices
- obtain statistics sorted by genus, species and gallery within the reference database
- identify isolates and validate the obtained identifications, thanks at the primary level to the Jaccard-Sneath distance between pairs of bacterial strains and ordered pair-wise comparisons for preliminary diagnosis and selection of specific identification tests, and at the secondary level to the Willcox identification score (also called Normalized Likelihood) with its validation coefficient known as Modal Likelihood Fraction (Willcox *et al.*, 1980)
- import and export strains for automatic clustering using the UPGMA (Unweighted Pair Group Method with Average) leading to dendrograms and CDC (coefficients of diagnosis capacity) computations, considering bacterial strains as Operational Taxonomic Units.

CABIQ software is moreover able to import and export data using classical formats (Dbase, Excel, text, Phylip).

Selection of specimens and characters

A character is defined as any property that can vary between operational taxonomic units and can assume two or more possible values. The choice of a large panel of characters allows coverage of a representative part of the total genome. Correlated characters, meaningless characters and non-reproducible features are not used.

Test reactions, character states, coding of characters and data matrix

Character states can generally be qualitative (e.g. the presence or absence of reaction, colour, shape) or quantitative. In CABIQ software, tests are binary-coded (a positive reaction is coded as 1, a negative one as 0). All phenotypic data for each strain is stored in a computer database but some information (colour of colonies, diffusible pigment …) is stored as such, being used only to give extra information about the strain. All tests that can be scored as binary variables can then be coded in a positivity list.

All phenotypic data obtained for each strain is saved in the computer database. The character states are coded entities

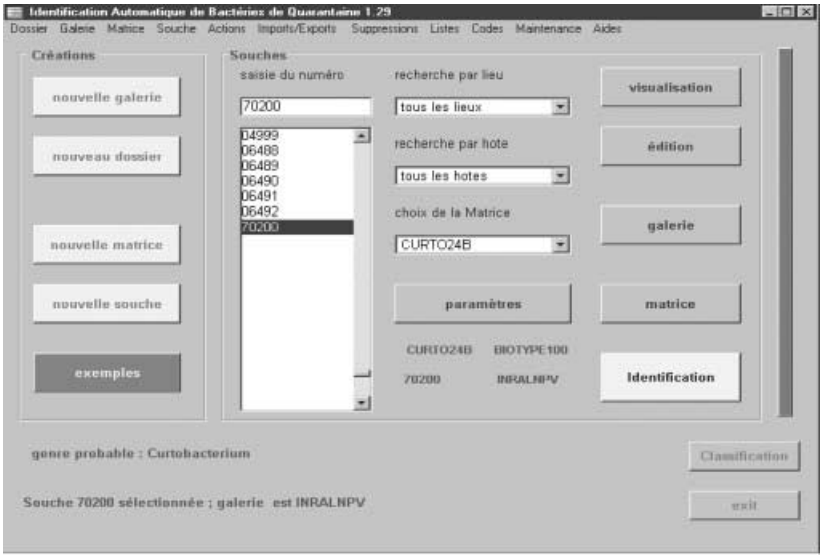

Fig. 1 Identification interface of CABIQ.

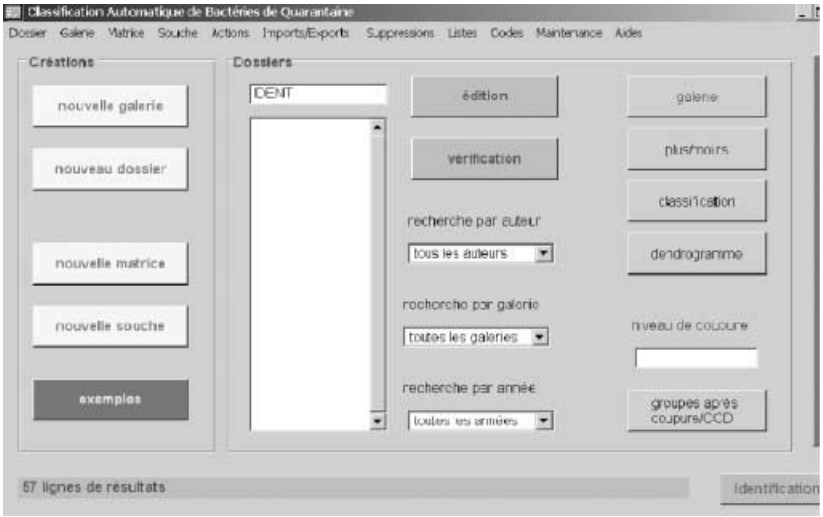

Fig. 2 Clustering interface of CABIQ.

determined by morphological, physiological, biochemical or other tests. They allow the creation of reference frequency matrices used for strain identification by numerical analysis. The primary aim is to assess and measure phenotypic resemblances between the individual bacterial strains and to construct taxonomic groups including those strains revealing the greatest degree of overall similarity.

Differentiating features and identification

The identification of an unknown isolate (not already entered into the database) is the process of determining whether it belongs to one of the established taxa in the database. The following sequence can be followed:

1. general tests including colour of colonies, presence of diffusible pigment, mucosity, host, disease symptoms, Gram reaction, HR on tobacco when Gram-negative
2. orientation tests (Tables 2 and 3) according to Gram reaction
3. according to the results of orientation tests, definition of probable genus according to rules which depend on Gram reaction (Tables 2 and 3)
4. specific identification tests, according to the probable genus (Tables 4 and 5).

CABIQ offers two interfaces. The first deals with the identification computations (Fig. 1) and the second with the clustering computations (Fig. 2). In order to identify the unknown isolate, CABIQ helps to choose:

- one or more galleries to be used (use of both orientation, identification and Biotype 100 tests is ideal)
- list of reference bacteria needed for identification. This pool of bacteria should be homogeneous from the taxonomic point of view. For example, the genera *Pseudomonas* or *Xanthomonas* may be chosen, or the group of Gram-positive bacteria. In case of doubt, identification trials can be done for

Subspecies	Oxidase	Levan	Glucose	Erythritol	Esculin
Clavibacter michiganensis subsp. *insidiosus*	0.14	0.10	0.43	0.10	0.90
Clavibacter michiganensis subsp. *michiganensis*	0.10	0.10	0.81	0.10	0.90
Clavibacter michiganensis subsp. *nebraskensis*	0.10	0.90	0.90	0.10	0.90
Clavibacter michiganensis subsp. *sepedonicus*	0.10	0.20	0.67	0.10	0.90
Clavibacter michiganensis subsp. *tessellarius*	0.10	0.80	0.90	0.10	0.90
Unknown isolate	0	1	1	0	1

Table 6 Comparison of an unknown isolate with the positivity frequency matrix of the different subspecies of *Clavibacter michiganensis* given by CABIQ

5 taxa and 19 conventional tests	Score	Modal likelihood fraction
Clavibacter michiganensis subsp. *insidiosus*	5.9%	7.3%
Clavibacter michiganensis subsp. *michiganensis*	9.0%	11.1%
Clavibacter michiganensis subsp. *nebraskensis*	98.7%	100.0%
Clavibacter michiganensis subsp. *sepedonicus*	9.0%	11.1%
Clavibacter michiganensis subsp. *tessellarius*	1.0%	1.2%
Unknown isolate	Successful identification	

Table 7 Identification score of Lapage and Wilcox and modal likelihood fraction given by CABIQ (example of validated identification)

5 taxa and 19 conventional tests	Score	Modal likelihood fraction
Clavibacter michiganensis subsp. *insidiosus*	0.0%	0.0%
Clavibacter michiganensis subsp. *michiganensis*	0.0%	0.0%
Clavibacter michiganensis subsp. *nebraskensis*	14.7%	4.2%
Clavibacter michiganensis subsp. *sepedonicus*	82.1%	0.3%
Clavibacter michiganensis subsp. *tessellarius*	3.1%	1.5%
Unknown isolate	Non validated identification	

Table 8 Identification score of Lapage and Wilcox and likelihood fraction given by CABIQ (example of non-validated identification)

several reference taxa (*Pseudomonas*, *Xanthomonas*, *Ralstonia* …)
- reference probability matrix to be used, depending on the selected tests.

Measurement of resemblance

A paired comparison method may be used to check which reference strains are most similar to the unknown isolate. This can be done for all galleries. It ensures that the chosen identification tests are appropriate for the identification. (e.g. if identification tests for '*Clavibacter*' have been selected, one can check that reference *Clavibacter* strains are indeed the closest relatives; Table 6). After these comparisons, CABIQ calculates the interrelation between the unknown isolate and each taxon through likelihood coefficient, normalized likelihood coefficient (also called identification score) or Willcox's probability, and relative likelihood. For the example given in Table 7, the unknown isolate is identified as *Clavibacter michiganensis* subsp. *nebraskensis*, since this taxon has the highest identification score.

The indicated taxon is then validated by calculation of the global modal likelihood fraction for each taxon (maximum likelihood coefficient) and modal probability. The modal likelihood fraction proves the absolute degree of affinity to the species in the matrix. In the case of Table 8, the unknown isolate looks like *Clavibacter michiganensis* subsp. *sepedonicus* but this identification is not validated.

The clustering automatically provided by CABIQ before or after the identification enables the user to locate the position of the isolate within the dendrogram of all chosen strains or taxa (Fig. 3) using the UPGMA clustering method applied to the matrix of Jaccard-Sneath distances built from the positivity lists. The dendrogram is then automatically drawn using GNUPLOT. The validation of the clustering uses prior and conditional probabilities to compute the CCD indicators (Descamps & Véron, 1981). Entering a cut-off value derived from the analysis of the dendrogram enables CABIQ to provide strain classes and their most frequent, and thus most discriminating, characters.

Conclusion

At the moment, the CABIQ system is a probabilistic identification tool running with phenotypic data only. Recently, rep-PCR profiles were obtained for most of the strains of the collection. DNA patterns are currently being analysed using GelCompar II software (Applied Maths). The genomic data matrix will be exported and integrated into CABIQ to improve

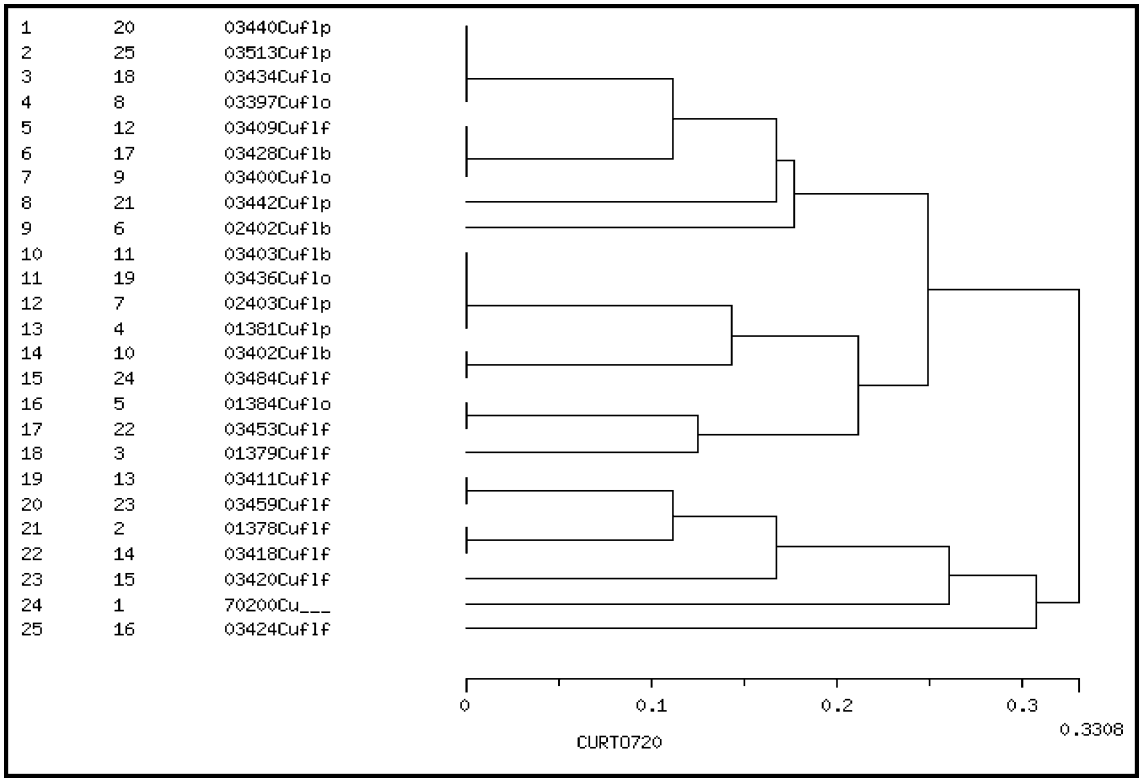

Fig. 3 Clustering of an unknown isolate 70200 in the dendrogram of subspecies of *Curtobacterium flaccumfaciens* given by CABIQ.

the quality and flexibility of the project identification system. In association with the reference database of INRA-LNPV, CABIQ provides a taxonomic expert system for identification of regulated phytopathogenic bacteria and the relatives.

Acknowledgements

L. Gardan and L. Sutra are two of those who initiated this project. Louis Gardan has now retired, after long and highly valuable professional activities in bacteriology and especially taxonomy. Laurent Sutra was a young and promising researcher, who prematurely passed away. He input all his energy, enthusiasm, freshness and kindness in the team, and openly shared his wide knowledge and skills. We miss him.

Identification informatisée et classification des bactéries phytopathogènes réglementées: constitution d'une base de référence et création d'un logiciel d'aide à la décision

Un logiciel spécifique destiné à une identification fiable et rapide des principales bactéries phytopathogènes réglementées, appelé CABIQ, est décrit. Il est basé sur les propriétés phénotypiques et génomiques des bactéries. Environ 500 souches de références ont été utilisées pour initier la base de données, qui comprend des tests phénotypiques conventionnels et des profils mutritionnels via les galeries du Biotype 100 (BioMérieux). Le logiciel CABIQ, avec la base de données et les matrices de référence, est un guide pour réaliser des essais d'identification de nouveaux isolats. Les modules traitant des caractéristiques phénotypiques et de Biotype 100 sont déjà terminés. A court terme, des résultats de rep-PCR compléteront système. Ces perspectives intéressantes amélioreront la fiabilité de l'identification assistée par ordinateur.

Компьютерная идентификация и группирование регулируемых фитопатогенных бактерий: создание справочной базы данных и разработка компьютерной системы

CABIQ представляет собой специфическую компьютерную систему для надежной и быстрой иденти-фикации основных регулируемых фитопатогенных бактерий. Система основана на фенотипных и геномных свойствах бактерий. Для начала в ней использованы около 500 справочных штаммов, включая обычные фенотипные тесты, а также галереи Биотипа 100 (BioMérieux). Система CABIQ с ее базой данных и справочными матрицами служит руководством по анализам, необходимым при выявлении новых изолятов. Модули, связанные с фенотипными характеристиками и Биотипом 100 уже созданы. В ближайшее

время к системе будут добавлены результаты при повторном анализе ПЦР. Эта система открывает интересные перспективы для увеличения надежности компьютерной идентификации.

References

Ayers SH, Rupp P & Johnson WT (1919) *A Study of the Alkali Forming Bacteria in Milk. US Department of Agriculture Bulletin*, no. 782. USDA, Washington (US).

Descamps P & Véron M (1981) Une méthode de choix des caractères d'identification basée sur le théorème de Bayes et la mesure de l'information. *Annales de l'Institut Pasteur, Microbiologie* **132B**, 157–170.

Dye DW (1968) A taxonomic study of the genus *Erwinia*. 1. The '*Amylovora*' group. *New Zealand Journal of Science* **11**, 590–607.

Hugh R & Leifson E (1953) The taxonomic significance of fermentative versus oxidative metabolism of carbohydrates by various gram-negative bacteria. *Journal of Bacteriology* **66**, 24–26.

King EO, Ward MK & Raney DE (1954) Two simple media for the demonstration of pyocyanin and fluorescein. *Journal of Laboratory Medicine* **44**, 301–307.

Klement Z (1963) Rapid detection of the pathogenicity of phytopathogenic P*seudomonas*. *Nature* **199**, 299–300.

Lelliott RA & Stead DE (1987) *Methods for the Diagnosis of Bacterial Diseases of Plants*, Vol. 2. Blackwell, Oxford (GB).

Louws FJ, Fulbright DW, Stephens CT & de Bruijn FJ (1994) Specific genomic fingerprints of phytopathogenic *Xanthomonas* and *Pseudomonas* pathovar and strains generated with repetitive sequences and PCR. *Applied and Environmental Microbiology* **60**, 2286–2295.

Schaad NW, Jones JB & Chun W (2001) *Laboratory Guide for Identification of Plant Pathogenic Bacteria*, 3rd edn. APS Press, St Paul (US).

Suslow TV, Schroth MN & Isaka M (1982) Application of a rapid method for Gram differentiation of plant pathogenic and saprophytic bacteria without staining. *Phytopathology* **72**, 917–918.

Willcox WR, Lapage SP & Holmes B (1980) A review of numerical methods in bacterial identification. *Antonie Van Leeuwenhoek* **46**, 233–299.