

## Comparative evaluation of three commercial software packages for analysis of DNA polymorphism patterns

A. Rementería<sup>1</sup>, L. Gallego<sup>2</sup>, G. Quindós<sup>2</sup> and J. Garaizar<sup>1</sup>

<sup>1</sup>Departamento de Inmunología, Microbiología y Parasitología, Facultad de Farmacia, Vitoria-Gasteiz, Spain and

<sup>2</sup>Facultad de Medicina y Odontología, Bilbao, Universidad del País Vasco-Euskal Herriko Unibertsitatea, Spain

**Objective** In the present study we have compared three commercial software packages, GelCompar, Molecular Analyst Fingerprinting, and BioImage, to determine if the results generated by the programs were comparable and correlated adequately with visual interpretation of electrophoretic gels, in the analysis of several well characterized incidents of infections.

**Methods** Infections caused by *Pseudomonas aeruginosa*, *Candida dubliniensis*, *C. albicans*, and serotypes of *Salmonella* were characterized by restriction endonuclease analysis, macrorestriction analysis of genomic DNA with pulsed-field gel electrophoresis, and random amplified polymorphic DNA. The genotypes were visually detected based on band presence or absence in the different gels. The similarity values of DNA profiles were computed using Dice coefficient and were presented in dendrograms by UPGMA. The concordance or agreement between the number of genotypes obtained and their clustering, using the computerized programs, was determined.

**Results** In general, agreement in number of genotypes obtained visually and by using the commercial DNA analysis software was achieved, but discrepancies were also denoted between the systems. The concordance between the visual and the computerized analysis ranged from 72% to 100%.

**Conclusion** In our experience, although the programs evaluated in the present study performed acceptably well, such programs may be used as an aid in the analysis of complex banding patterns, and they do not provide an indisputably correct analysis in genotype definition.

**Keywords** DNA patterns, computerized analysis, GelCompar, molecular analyst, BioImage

**Accepted** 26 February 2001

*Clin Microbiol Infect* 2001; 7: 331–336

### INTRODUCTION

Electrophoresis-based DNA fingerprinting methods are now widely used for epidemiological typing of microorganisms. These methods include restriction endonuclease analysis (REA), macrorestriction analysis of genomic DNA with pulsed-field gel electrophoresis (PFGE), and PCR-based methods, such as random amplified polymorphic DNA (RAPD) [1–5].

Recent efforts have aimed at international standardization of the procedures of the molecular typing methods [2,4]. Nevertheless, a general consensus on the ways of interpretation of gels

has not yet been achieved. Although a visual interpretation of an electrophoretic gel based on the presence or absence of bands is the reference method for pattern recognition at present, the use of computerized programs that analyze band patterns automatically is increasing world-wide. The computerized programs not only analyze electrophoretic gels but have the possibility of creating databases or libraries of profiles as well, allowing for a quick identification of new profiles and the analysis of the relatedness or similarity of band patterns using mathematical coefficients. Furthermore, harmonization in the use of methods based in computerized analysis of band patterns could make possible in the near future the comparison of large numbers of DNA patterns generated in different laboratories.

In spite of the interest of this subject, few data on comparative analysis of commercial software packages against visual interpretation are available in the literature at present [3–8]. A practical study of identical gels by several computer programs, with fixed parameters such as band definition, mathematical coefficients or clustering methods, could be a step forward in

Corresponding author and reprint requests: J. Garaizar, Departamento de Inmunología, Microbiología y Parasitología, Facultad de Farmacia, Universidad del País Vasco-Euskal Herriko Unibertsitatea, Apartado 450, E-01080 Vitoria-Gasteiz, Spain

Tel: +34 945 013912

Fax: +34 945 130756

E-mail: oipgacaj@vc.ehu.es

the standardization of computerized band analysis. In the present study, we have compared three systems, GelCompar, Molecular Analyst, and BioImage, to determine if results generated by these programs correlated adequately with visual interpretation of DNA patterns in the analysis of several well characterized incidents of infections.

## MATERIALS AND METHODS

### REA of *Pseudomonas aeruginosa*

Ten *P. aeruginosa* isolates were recovered from blood cultures from 10 different bacteremic patients in Seville, Bilbao, and Madrid, Spain. The O-serological typing and phage typing were performed in the Central Public Health Laboratory, Colindale, London, UK. Chromosomal DNA was isolated and digested with *SalI* restriction endonuclease as described previously [9]. The digested DNA was subjected to horizontal electrophoresis in 0.4% agarose in TBE running buffer for 13 h at 1.5 V/cm at room temperature. High molecular weight DNA markers (Gibco-BRL, Madrid, Spain) were used as molecular weight standards.

### RAPD of *P. aeruginosa*

Twelve *P. aeruginosa* isolates were recovered from patients with bronchiectasis without cystic fibrosis attending the Santa Marina Hospital (Bilbao, Spain). Total DNA was extracted and amplified as described previously [10]. Primer RD1 (5'-AGCGGGCCAA-3') were purchased from Gibco-BRL. The amplified products were electrophoresed in 1.3% agarose gels with TBE running buffer. A molecular weight marker 100 bp DNA ladder (Gibco-BRL) was included in the gel.

### RAPD of *Candida* isolates

Eighteen *Candida* isolates recovered from patients with presumed epidemiologically unrelated candidiasis were studied. The isolates, including 12 *C. albicans* from 11 patients and six *C. dubliniensis* from four patients, were identified by conventional mycological identification methods and by their carbohydrate assimilation patterns on ID 32C strips (bioMérieux, Marcy-l'Etoile, France). *C. dubliniensis* isolates identities were confirmed by immunofluorescence as previously described [11,12]. Polymerase chain reactions were performed according to the method described by Alonso et al. [13], using the primer AB1-12 (5'-CCTTGACGCA-3') purchased from Advanced Technologies Ltd. (Leatherhead, Surrey, UK). The amplification products were analyzed by gel electrophoresis on a horizontal 2% agarose gel, in TBE running buffer. A 123-bp DNA ladder (Sigma-Aldrich, Madrid, Spain) was used to flank the samples.

### RAPD of *Salmonella* isolates

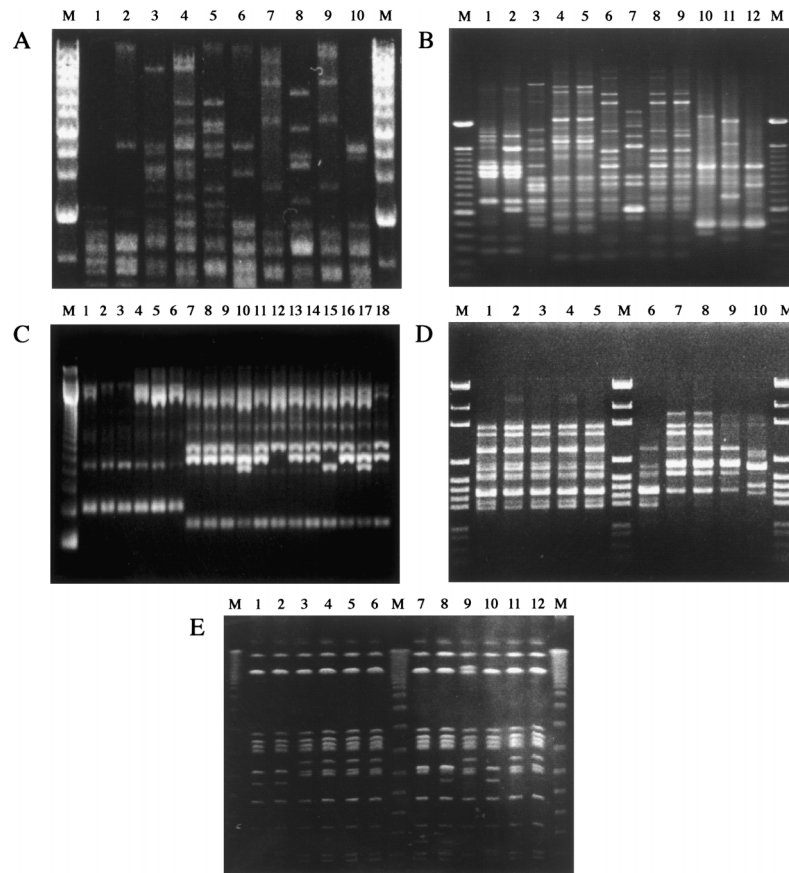
Ten *Salmonella* isolates (including six *S. enteritidis*, two *S. typhimurium*, one *S. virchow*, and one *S. arizonae*) recovered from clinical, foods, and environmental sources were analyzed. Serotyping was performed in the Public Health Laboratory (Bilbao, Spain). Phage typing (PT) was performed at the Central Public Health Laboratory (Colindale, London, UK), and Instituto de Salud Carlos III (Majadahonda, Madrid) [14]. DNA extraction for RAPD was performed as described previously [15]. Oligonucleotide OPS-19 (5'-GAGTCAGCAG-3') was purchased from Operon (Alameda, CA, USA). The amplified PCR product was run in 2% agarose gels for 90 min at 100 V. pGEM (Promega, Barcelona, Spain) was used as molecular weight standard.

### PFGE of *Salmonella* isolates

Twelve epidemiologically unrelated *S. enteritidis* isolates recovered from human, animal and environmental sources in Denmark, and belonging to phage types 1, 4, 6, and 8, were analysed. Macrorestriction analysis of genomic DNA was performed by *XbaI* restriction endonuclease (Boehringer Mannheim, Mannheim, Germany) and subjected to PFGE as described previously [16]. The DNA size standard used was lambda ladder concatamers (Bio-Rad, Hercules, CA, USA).

### Computerized gel analysis

All the evaluated gels were stained with ethidium bromide and photographed under UV light (Figure 1). In order to standardize the quality of images, photographs of DNA patterns obtained by the different molecular typing methods were scanned with an Arcus II Scanner (Agfa, Mortsel, Belgium), digitalized images were saved in TIFF format files and were sent by E-mail to the three personal computers for performing the computerized analysis of images. The three programs used to analyze the digitalized data were GelCompar version 4.0 (Applied Maths, Kortrijk, Belgium), Molecular Analyst Fingerprinting version 1.0 (Bio-Rad) and BioImage version 3.2 (BioImage Corporation, Ann Arbor, MI, USA). The GelCompar program was running in a Pentium II-350 MHz personal computer under Microsoft Windows 95. The Molecular Analyst Fingerprinting was running in a PowerPC Macintosh 8600/200 under MacOS 8.0 operating system. The BioImage was running on a SPARC Station IV with a SunOs 3.5 operating system. Two of us (A.R. and J.G.) agreed on the definition of the bands according to densitometric curves and the accompanying hard-copy photograph, and took these data as reference in comparison between visual and computerized analysis. For the purposes of this study, the similarity values of DNA profiles were computed on the basis of band positions by using the Dice coefficient in GelCompar (corresponding to



**Figure 1** Isolates of *P. aeruginosa* analyzed by REA (A). Isolates of *P. aeruginosa* analyzed by RAPD (B). Isolates of *C. dubliniensis* (no.1–6) and *C. albicans* (no. 7–18) analyzed by RAPD (C). Isolates of *S. enteritidis* PT 4 (no.1–5), and PT 7 (no. 6), *S. typhimurium* PT 169 (no. 7), and PT 96 (no. 8), *S. virchow*, and *S. arizonae* analyzed by RAPD (D). Isolates of *S. enteritidis* PT 1 (no. 1, 2, 9, 10), PT 4 (no. 3, 4), PT 6 (no. 5, 6, 11, 12), and PT 8 (no. 7, 8) analyzed by PFG (E). Figure 1C is shown with permission of Revista Iberoamericana de Micología.

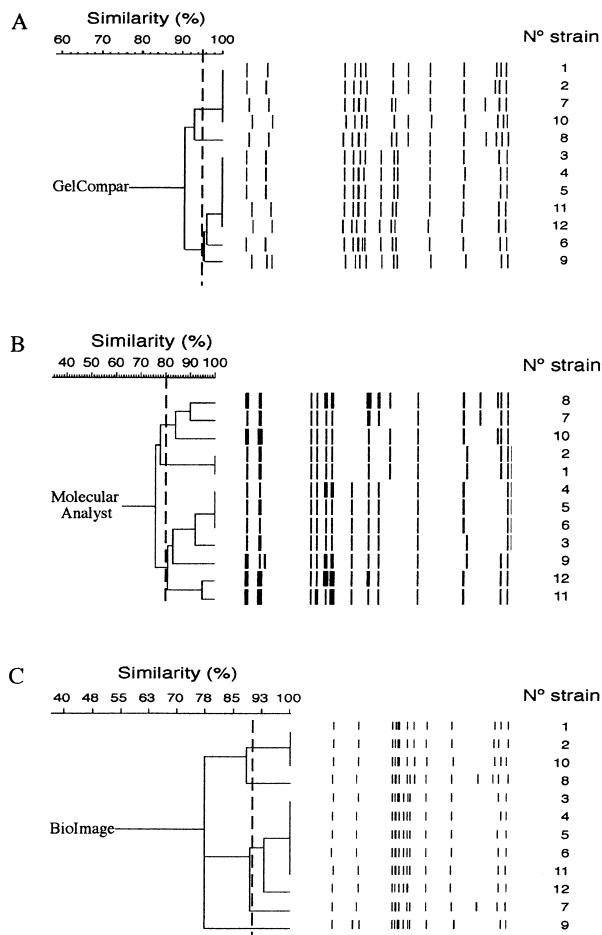
similarity coefficient no. 2 in BioImage and similarity coefficient no. 1 in Molecular Analyst). The similarity values were then used to build a dendrogram by the unweighted pair group method using arithmetic averages (UPGMA). Cut-offs in the dendrograms for defining genotypes were adapted to each program and were maintained throughout the study at 80% similarity for Molecular Analyst, 90% similarity or BioImage, and 95% similarity for GelCompar. At these percentages of band similarity, the DNA patterns of the molecular weight standards that were allocated for most of the gels were correctly grouped in a branch with the computer programs. One example of the dendrograms generated by the three programs analysing the PFGE of *S. enteritidis* is showed in Figure 2.

## RESULTS

Although the software systems analyzed in this study allowed automated band detection, preliminary readings of some gels

showed discrepancies between computers due to differences in sensitivity of the band detection thresholds. Therefore, in order to standardize the number of bands present, these were initially evaluated on the basis of visual inspection and the defined band patterns were considered as standards for the computerized approaches. Table 1 shows the number of different genotypes visually detected according to the general criterion of band presence or absence in the different gels, and the concordance or agreement between the number of genotypes and their clustering in groups obtained both visually and using the computerized programs.

When 10 *P. aeruginosa* isolates were processed by REA, a combination of visual analysis of REA with O-serology and phage typing defined nine pheno-genotypes, and possible cross-infection with identical combined typing results was detected in two unrelated patients in a single hospital. The 10 isolates were grouped in nine different genotypes by GelCompar and Molecular Analyst systems. The concordance between automatic



**Figure 2** Dendrograms of DNA pattern similarity obtained by GelCompar (A), Molecular Analyst (B), and BioImage (C) analysing isolates of *S. enteritidis* by macrorestriction and PFGE. Cut-offs in the dendrograms were established at 95% for GelCompar, 80% for Molecular Analyst, and 90% for BioImage.

reading with both programs and visual reading was complete. BioImage was able to distinguish as many genotypes as isolates studied, not grouping in the same genotype two isolates from two different patients that had identical serotype and phage types.

The 12 isolates of *P. aeruginosa* analyzed by RAPD were recovered from patients with bronchiectasis in a single hospital in Spain. The study characterized the strains by an antibiotic susceptibility test and PCR fingerprinting techniques, with three objectives: to assess the diversity of strains colonizing patients over time, to determine how frequently antimicrobial resistance was related to the acquisition of new clones, and to estimate the degree of variability of the genomic distance among identified clones. Visually, 10 different genotypes were observed, two patients were colonized by different clones at a time, and cross-infection or common-source exposure did not seem to have occurred. The same number as obtained by

Molecular Analyst and BioImage programs. GelCompar distributed the isolates into nine genotypes. The concordance between visual and computerized reading was 92% for the three systems evaluated, due to differences in the grouping of some isolates.

When 18 *Candida* isolates were studied by RAPD, six genotypes were visually observed. Interestingly, two band profiles were detected for the recently defined new species *C. dubliniensis*. Patients with sequential isolates showed identical DNA profiles. *C. albicans* isolates were also differentiated into several groups. GelCompar showed seven genotypes, with 94% concordance with visual reading. Eight genotypes were detected by the Molecular Analyst and BioImage and lower concordance was observed (72% and 89%, respectively).

RAPD analysis of *Salmonella* isolates showed five different genotypes by visual reading. Each *Salmonella* serotype showed a characteristic DNA profile. The group of five isolates of *S. enteritidis* serotype Enteritidis with PT 4 showed an identical DNA profile, meanwhile the isolate with PT 7 showed a distinguishable profile. The profile of two isolates from serotype Typhimurium were visually indistinguishable, although it showed a different phage type. The isolates of *S. virchow* and *S. arizonae* showed unique genotypes. The readings by GelCompar and Molecular Analyst gave the same number of genotypes and their concordance with visual grouping of isolates was complete. Molecular Analyst showed more genotypes and a decrease of the concordance up to 80%.

When the collection of 12 epidemiologically unrelated *S. enteritidis* isolates belonging to four phage types were typed by PFGE, five different genotypes were visually observed, although the combination of both techniques resulted in further subdivisions. In this case, only BioImage gave the same number of genotypes as visual reading and the concordance between both methods was complete. Conversely, Molecular Analyst and GelCompar recognized only three genotypes, with a concordance of 75% and 83%, respectively. The dendrograms of similarity of the macrorestriction patterns obtained using the three programs is shown in Figure 2.

## DISCUSSION

Some authors have outlined the numerous factors that may influence the reliability of molecular typing methods, one of them being the interpretation of the DNA band patterns in electrophoretic gels and the grouping of isolates in different genotypes [2,4]. The visual interpretation of gels based on band presence or absence is the current standard for DNA pattern definition, although this approach is hampered by the inherent difficulties of the number of gels that could be analyzed, the time-consuming process of the mathematical analysis, and the manual construction of dendrograms. Computerized gel analyses have the advantage of rapid mathematical analysis and

**Table 1** Concordance among visual and computerized analyses of DNA polymorphism patterns

Typing method	DNA patterns analysis method			
	Visual reading	GelCompar	Molecular Analyst	BiolImage
REA-P <sup>a</sup>				
No. of genotypes	9	9	9	10
Concordance (%) <sup>b</sup>	–	100	100	90
RAPD-P <sup>c</sup>				
No. of genotypes	10	9	10	10
Concordance (%)	–	92	92	92
RAPD-C <sup>d</sup>				
No. of genotypes	6	7	8	8
Concordance (%)	–	94	72	89
RAPD-S <sup>e</sup>				
No. of genotypes	5	5	7	5
Concordance (%)	–	100	80	100
PFGE-S <sup>f</sup>				
No. of genotypes	5	3	3	5
Concordance (%)	–	83	75	100

REA-P<sup>a</sup>: Restriction endonuclease analysis of *Pseudomonas* isolates; Concordance<sup>b</sup>: proportion of strains that are well allocated in genotype groups taking visual reading of electrophoretical gels as the standard; RAPD-P<sup>c</sup>: Random amplified polymorphic DNA of *Pseudomonas* isolates; RAPD-C<sup>d</sup>: Random amplified polymorphic DNA of *Candida* isolates; RAPD-S<sup>e</sup>: Random amplified polymorphic DNA of *Salmonella* isolates; and PFGE-S<sup>f</sup>: Pulsed field gel electrophoresis of *Salmonella* isolates.

construction of dendrograms; on the other hand, they are expensive and require adequate training in the handling of electronic data. Recently, we have successfully used the Gel-Compar program for the construction of three libraries of PFGE patterns of *S. enteritidis* [17]. Such libraries could help in the comparison of DNA profiles and the surveillance of these microorganisms, if laboratories agree on the standardization and the use of such a computer program.

In this work we have designed a practical approach for the measurement and quantification of the discrepancies between the visual analysis of gels and three computerized programs used world-wide for DNA pattern analysis. The images for the analysis were selected from our research in epidemiological typing of incidents of infections, as representatives of each of the methods evaluated and the general variability of electrophoretical gels [9,10,13,15,16]. We have maintained fixed parameters which some authors denoted as important in the comparison of computerized programs [6–8]. These parameters included the quality of images, the visual definition of bands, the mathematical approach to the assessment of similarities and clustering of data, and the cut-offs for the interpretation of groups or genotypes in the dendrograms. We have observed concordance and agreement using the programs and visual interpretation. Although some discrepancies arose as expected, not fixing the parameters described above could have resulted in even more discrepancies.

All programs required the user to make decisions at various steps of the analysis, as in the selection of the parameters and matching of patterns. To emphasize a critical point, before a gel

analysis is performed, it has to pass through a so-called normalization step, in order to compensate for the differences in the run length of the gels. Rigorous normalization was considered necessary for the three programs, and the inclusion of a set of molecular size markers or standard strains run in every assay for the assessment of acceptable intergel reproducibility, was considered to be important. Even so, the normalization process had to be supervised, as automatic association methods often produced incorrect alignments.

A broad consensus has to be reached in future for the comparison of DNA patterns and exchange of data among laboratories using computer programs. The combination of the data gathered from other epidemiological typing techniques, with high quality epidemiological information on the incidents of infection, would result in more accuracy in strain characterization than simply an inexact grouping in a dendrogram. Again, we would stress that computer programs may be used as an aid in the analysis of complex banding patterns, although they do not provide an indisputably correct analysis, and that strong conclusions about genotypes obtained by computerized analyses should always be accepted with caution.

## ACKNOWLEDGMENT

This work was funded by grants from the University of Basque Country (EA 123.93, EA 125.96, TA 131.97, and EB 157.99). This work was presented as communication at the 5th International Meeting on Bacterial Epidemiological Markers, held from 6 to 9 September 2000, at Noordwijkerhout, The

Netherlands. We thank colleagues from Central Public Health Laboratory (Colindale, UK), Public Health Laboratory (Bilbao, Spain), and Instituto de Salud Carlos III (Majadahonda, Spain), for performing serotyping and phage typing of the strains, and Itziar Marquiegui (UPV/EHU) for revising the manuscript.

## REFERENCES

1. Struelens M, Gerner-Smidt P, Rosdahl V, *et al.* Guidelines for appropriate use and evaluation of microbial epidemiological typing systems. *Clin Microbiol Infect* 1996; 2: 2–11.
2. van Belkum A, Kluytmans J, van Leeuwen JW *et al.* Multicenter evaluation of arbitrarily primed PCR for typing of *Staphylococcus aureus* strains. *J Clin Microbiol* 1995; 33: 1537–47.
3. Soll DR. The ins and outs of DNA fingerprinting the infectious fungi. *Clin Microbiol Rev* 2000; 13: 332–70.
4. van Belkum A, van Leeuwen W, Kaufmann ME, *et al.* Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of *Sma*I macrorestriction fragments: a multicenter study. *J Clin Microbiol* 1998; 36: 1653–9.
5. Vila J. Current and future prospects of molecular biology techniques in the epidemiologic study of nosocomial infections. *Enferm Infecc Microbiol Clin* 1996; 14: 341–4.
6. Seward RJ, Ehrenstein B, Grundmann HJ, Towner KJ. Direct comparison of two commercially available computer programs for analysing DNA fingerprinting gels. *J Med Microbiol* 1997; 46: 314–20.
7. Gerner-Smidt P, Graves LM, Hunter S, Swaminathan B. Computerized analysis of restriction fragment length polymorphism pattern: comparative evaluation of two commercial software packages. *J Clin Microbiol* 1998; 36: 1318–23.
8. Cardinali G, Martini A, Tascini C, Bistoni F. Critical observations on computerized analysis of banding patterns with commercial software packages. *J Clin Microbiol* 1999; 37: 876–7.
9. Garaizar J, Latorre M, López-Molina N *et al.* Computerized restriction endonuclease analysis compared with O-serotype in the epidemiological fingerprinting of *Pseudomonas aeruginosa*. *Clin Microbiol Infect* 1997; 3: 220–8.
10. Pujana I, Gallego L, Martín G, López F, Canduela J, Cisterna C. Epidemiological analysis of sequential *Pseudomonas aeruginosa* isolates from chronic bronchiectasis patients without cystic fibrosis. *J Clin Microbiol* 1999; 37: 2071–3.
11. Lehmann PF, Lin D, Lasker BA. Genotypic identification and characterization of species and strains within the genus *Candida* by using random amplified polymorphic DNA. *J Clin Microbiol* 1992; 30: 3249–54.
12. Bikandi J, San-Millán R, Moragues MD *et al.* Rapid identification of *Candida dubliniensis* by indirect immunofluorescence based on differential localization of antigens on *C. dubliniensis* blastospores and *C. albicans* germ tubes. *J Clin Microbiol* 1998; 36: 2428–33.
13. Alonso R, Garaizar J, Pontón J, Quindós G. Utility of random amplified polymorphic DNA in the discrimination between *Candida albicans* and *C. dubliniensis*. *Rev Iberoam Micol* 2000; 17: 10–3.
14. Ward L, de Sa RJDH, Rowe B. A phage-typing scheme for *Salmonella enteritidis*. *Epidemiol Infect* 1987; 99: 291–4.
15. López-Molina N, Laconcha I, Rementeria A, Audicana A, Perales I, Garaizar J. Typing of *Salmonella enteritidis* of different phage types of PCR fingerprinting. *J Appl Microbiol* 1998; 84: 877–82.
16. Laconcha I, Baggesen DL, Rementeria A, Garaizar J. Genotypic characterisation by PFGE of *Salmonella enterica* serotype *enteritidis* phage types 1, 4, 6, and 8 isolated from animal and human sources in three European countries. *Vet Microbiol* 2000; 75: 155–65.
17. Garaizar J, López-Molina N, Laconcha I *et al.* Suitability of PCR fingerprinting, infrequent-restriction-site PCR, and pulsed-field gel electrophoresis, combined with computerized gel analysis, in library typing of *Salmonella enterica* serovar *Enteritidis*. *Appl Environ Microbiol* 2000; 66: 5273–81.