International Journal of Systematic and Evolutionary Microbiology (2000), **50**, 665–677 Printed in Great Britain

Comparison of AFLP and rep-PCR genomic fingerprinting with DNA–DNA homology studies: *Xanthomonas* as a model system

Jan L. W. Rademaker,[1]† Bart Hoste,[5] Frank J. Louws,[1,3,4] Karel Kersters,[6] Jean Swings,[5,6] Luc Vauterin,[7] Paul Vauterin[7] and Frans J. de Bruijn[1,2,3]

Author for correspondence: Frans de Bruijn. Tel: +1 517 353 2229. Fax: +1 517 353 9168. e-mail: debruijn@pilot.msu.edu

[1,2,3] MSU-DOE Plant Research Laboratory[1], Department of Microbiology[2] and NSF Center for Microbial Ecology[3], Michigan State University, East Lansing, Michigan 48824, USA

[4] Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina 27695, USA

[5,6] BCCM/LMG Bacteria Collection[5] and Laboratorium voor Microbiologie[6], Rijksuniversiteit Gent, Ledeganckstraat 35, B-9000 Ghent, Belgium

[7] Applied Maths BVBA, Derbystraat 331, B-9051 Ghent, Belgium

The genus *Xanthomonas* contains a large number of strains, which have been characterized by a variety of phenotypic and genotypic classification methods. The *Xanthomonas* collection constitutes one of the largest groups of bacteria that have been characterized phylogenetically by DNA–DNA homology studies and genomic fingerprinting. Presently, a total genomic DNA–DNA homology value of 70% represents an internationally accepted criterion to define bacterial species levels. However, the complexity of DNA–DNA reassociation kinetics methods precludes the rapid analysis of large numbers of bacterial isolates, which is imperative for molecular microbial diversity studies. Therefore, the aim of this study was to compare more facile PCR-based genomic fingerprinting techniques, such as repetitive-sequence-based (rep)-PCR and AFLP genomic fingerprinting, to DNA–DNA hybridization studies. Using three different primer sets, rep-PCR genomic fingerprint patterns were generated for 178 *Xanthomonas* strains, belonging to all 20 previously defined DNA–DNA homology groups, and one *Stenotrophomonas maltophilia* strain. In addition, AFLP genomic fingerprints were produced for a subset of 80 *Xanthomonas* strains belonging to the 20 DNA–DNA homology groups and for the *S. maltophilia* strain. Similarity values derived from rep-PCR- and AFLP-generated fingerprinting analyses were calculated and used to determine the correlation between rep-PCR- or AFLP-derived relationships and DNA–DNA homology values. A high correlation was observed, suggesting that genomic fingerprinting techniques truly reveal genotypic and phylogenetic relationships of organisms. On the basis of these studies, we propose that genomic fingerprinting techniques such as rep-PCR and AFLP can be used as rapid, highly discriminatory screening techniques to determine the taxonomic diversity and phylogenetic structure of bacterial populations.

Keywords: rep-PCR, AFLP, *Xanthomonas*, DNA–DNA homology studies, molecular phylogeny

† **Present address:** The Netherlands Culture Collection of Bacteria (NCCB), PO Box 80056, 3508 TB Utrecht, The Netherlands.

Abbreviations: ARDRA, amplified rDNA restriction enzyme analysis; BER, individual BOX-, ERIC- and REP-PCR generated genomic fingerprints combined linearly using the 'combined gel' option of the GelCompar program (BER profile); BOX-PCR, PCR based on primers targeting the highly conserved repetitive DNA sequences of the BOXA subunit of the BOX element of *Streptococcus pneumoniae*; DAF, DNA amplification fingerprinting; ERIC, enterobacterial repetitive intergenic consensus; RAPD, random amplified polymorphic DNA; rep-PCR, repetitive-sequence-based PCR; REP, repetitive extragenic palindromic sequence; UPGMA, unweighted pair group method using averages.

01227 © 2000 IUMS

Table 1. DNA–DNA homology groups and strains analysed by AFLP and rep-PCR genomic fingerprinting

Taxon	Host	Geographical origin	LMG strain no.*
Group 1			
X. fragariae	*Fragaria ananassa*	Greece, USA	**706**, **708**[T]
Group 2			
X. hortorum pv. *hederae*	*Hedera helix*	USA	**733**†
X. h. pelargonii	*Pelargonium* sp.	New Zealand, France, Belgium	**7314**†, **7356**, **7585**, 7712
X. h. vitians	*Lactuca sativa*	Zimbabwe	**938**
Group 3			
X. populi	*Populus canadensis*	France	**974**, **5743**[T], 5753
Group 4			
X. arboricola corylina	*Corylus* sp.	UK, USA	**688**, **689**†, **8658**, **8660**
X. a. juglandis	*Juglans regia*	New Zealand, The Netherlands	747†, **8047**
X. a. poinsettiicola	*Euphorbia pulcherrima*	New Zealand	**5403**
X. a. populi	*Populus euramericana*	The Netherlands	**12141**†
X. a. pruni	*Prunus* sp.	New Zealand	**852**†, **8680**
Group 5			
X. cassavae	*Manihot esculenta*	Malawi, Rwanda	670, **673**†, **5264**
Group 6			
X. codiaei	*Codiaeum variegatum*	USA	**8677**, **8678**
Group 7			
X. bromi	*Bromus* sp.	France, New Zealand	(**947**), **8269**, 8272
Group 8			
X. cucurbitae	*Cucurbita maxima*	New Zealand	**690**†, **8662**
Group 9			
X. axonopodis alfalfae	*Medicago sativa*	Australia, Sudan	497†, 8079
X. a. axonopodis	*Axonopus scoparius*	Colombia	**538**[T], 539
X. a. bauhiniae	*Bauhinia racemosa*	India	548
X. a. begoniae	*Begonia* sp.	Belgium, New Zealand, UK	**551**, 7178, 7188, 7226, **7303**†, 7304, 7601
X. a. cajani	*Cajanus cajan*	India, Sudan	558†, 7387t1
X. a. cassavae	*Manihot esculenta*	Niger	8049
X. a. cassiae	*Cassia tora*	India	675†
X. a. citri (A)	*Citrus aurantifolia*, *C. latifolia*, *C. limon*, *C. reticulata*	Brazil, India, New Zealand, Pakistan, USA	681, (**682**†), 8650, 8654, 8657, 9176, 9321, 9665, 9671
X. a. aurantifolii (*citri* B)	*Citrus limon*	Argentina	9179, 9183
X. a. aurantifolii (*citri* C)	*Citrus aurantifolia*	Brazil	8655, 9181, 9658
X. a. aurantifolii (*citri* D)	*Citrus aurantifolia*	Mexico	9182
X. a. citrumelo (*citri* E)	*Poncirus trifoliata* × *Citrus sinensis*, *Citrus paradisi*	USA	9160, 9167, 9172, 9175, 9325
X. a. clitoriae	*Clitoria biflora*	India	9045†
X. a. coracanae	*Eleusine coracana*	India	686†, 7476
X. a. cyamopsidis	*Cyamopsis tetragonolobus*	India	691†
X. a. desmodii	*Desmodium dichotomum*	India	692†
X. a. desmodiigangetici	*Desmodium gangeticum*	India	693†
X. a. desmodiilaxiflori	*Desmodium laxiflorum*	India	9046†
X. a. desmodiirotundifolii	*Desmodium styracifolium*	India	694†
X. a. dieffenbachiae	*Anthurium* sp., *Dieffenbachia* sp.	Brazil, USA	**695**†, **7399**, 8664
X. a. erythrinae	*Erythrina variegata*	India	698†
X. a. glycines	*Glycine max*	Sudan	712†
X. a. lespedezae	*Lespedeza* sp.	USA	757†
X. a. malvacearum	*Gossypium* sp.	Sudan, Uganda	**761**†, 7429
X. a. manihotis	*Manihot esculenta*	Brazil, Nigeria	771, 773†, 778, 784†
X. a. patelii	*Crotalaria juncea*	India	811†
X. a. phaseoli	*Phaseolus vulgaris*	Romania	**7455**†, **8014**
X. a. phaseoli var. *fuscans*	*Phaseolus vulgaris*	Germany, South Africa, USA	**837t1**, 7511, 8036
X. a. phyllanthi	*Phyllanthus niruri*	Sudan	**844**†
X. a. poinsettiicola	*Euphorbia pulcherrima*	India	849†
X. a. rhynchosiae	*Rhynchosia memnonia*	Sudan	8021†
X. a. ricini	*Ricinus communis*	Ethiopia, Hong Kong, India, Zimbabwe	**861**†, 862, **7441**, 7443

***Table 1** (cont.)*

Taxon	Host	Geographical origin	LMG strain no.*
X. a. sesbaniae	*Sesbania sesban*	Unknown	867†
X. a. tamarindi	*Tamarindus indica*	India	869, 955†
X. a. vasculorum	*Saccharum officinarum*, *Thysanolaena maxima*	Australia, Mauritius	895, **899**, **901**†, 903
X. a. vesicatoria	*Capsicum* sp., *Lycopersicon esculentum*	Morocco, USA	**905**, **910**, **929t1**
X. a. vignaeradiatae	*Vigna radiata*	Sudan	936†
X. a. vignicola	*Vigna unguiculata*	India, Sudan, USA	828, 8139, 8752†
X. a. vitians	*Lactuca* sp.	Colombia, USA	937†
Group 10			
X. oryzae oryzae	*Oryza sativa*	Philippines	795, **5047**^{T}, 6518
X. o. oryzicola	*Oryza sativa*	Malaysia, Philippines	**665**, 793, 797†
Group 11			
X. vasicola holcicola	*Sorghum* sp., *Holcus* sp.	New Zealand, USA	**736**†, **7416**, **7489**
X. v. vasculorum	*Saccharum officinarum*, *Zea mays*	Malagasy Republic, South Africa	900, 902, 8284
Group 12			
X. pisi	*Pisum sativum*	Japan	**847**†
Group 13			
X. melonis	*Cucumis melo*	Brazil	**8670**†, **8672**
Group 14			
X. vesicatoria	*Lycopersicon esculentum*	Italy, New Zealand	**911t1**†, **920t1**
Group 15			
X. campestris aberrans	*Brassica oleracea* L. var. *capitata* L.	Australia	9037†
X. c. armoraciae	*Iberis* sp., *Armoracia lapathifolia*	Tanzania, USA	**535**†, **7383t2**
X. c. barbareae	*Barbarea vulgaris*	USA	**547**†, **7385**
X. c. campestris	*Brassica oleracea*, *Raphanus sativus*	France, Malawi, Burundi, New Zealand, UK	567, **568**^{T}, 571, 573, 583, 7514, 8032
X. c. incanae	*Matthiola* sp.	USA	**7421**, **7490**†
X. c. raphani	*Raphanus sativus*	USA	860†, **7505**, **8134**
Group 16			
X. translucens arrhenatheri	*Arrhenatherum elatius*	Switzerland	**727t1**†
X. t. cerealis	*Bromus inermis*, *Hordeum vulgare*	Canada, USA	**679**†, 880
X. t. graminis	*Lolium perenne*, *Dactylis glomerata*	Belgium, Switzerland	713, 726†
X. t. hordei	*Hordeum vulgare*, *Dactylis glomerata*	Canada, India, Japan, New Zealand	737†, 879, **882**, **8279**
X. t. phlei	*Phleum pratense*	Belgium, Norway	716, **730**†
X. t. phleipratensis	*Phleum pratense*	USA	**843**†
X. t. poae	*Poa trivialis*	Switzerland	594, **728**†
X. t. secalis	*Secale cereale*	Canada	**883**†, 7507
X. t. translucens	*Hordeum vulgare*, *Secale cereale*	USA	**876**†, 5259, 5260
X. t. undulosa	*Triticum* sp.	Canada	885, 888, **892**†
Group 17			
X. hyacinthi	*Hyacinthus orientalis*	The Netherlands	739†, **742**, **8041**
Group 18			
X. theicola	*Camellia sinensis*	Japan	**8684**†
Group 19			
X. sacchari	*Saccharum officinarum*	Guadeloupe	**471**, **476**
Group 20			
X. albilineans	*Saccharum officinarum*	Mauritius, Fiji	842, **494**^{T}
Group 21			
Stenotrophomonas maltophilia	Human with mouth cancer	Unknown	**958**^{T}

* Strains indicated in bold were also used in AFLP analysis, strains in parentheses were not used in rep-PCR genomic fingerprinting. t1, t2 and t3 indicate several stable colony types.

† Pathovar reference strain.

INTRODUCTION

Polyphasic taxonomy is increasingly being accepted as a comprehensive approach to microbial systematics (Colwell, 1970; Vandamme *et al*., 1996). However, the determination of total genomic DNA–DNA homology values has persisted as a dominant component of taxonomic analyses (Wayne *et al*., 1987; Murray *et al*., 1990). In fact, it has been referred to as the 'gold standard' for the definition of bacterial species (Stackebrandt & Goebel, 1994), and 70 % DNA–DNA homology values are considered to be the 'species limit' (Wayne *et al*., 1987; Murray *et al*., 1990). DNA–DNA homology-based approaches to classification are clearly impractical when analysing large collections of environmental isolates, due to the experimental limitations of analysing large numbers of strains by DNA–DNA reassociation methods in a pair-wise fashion. The analysis of 16S or 23S rDNA genes by DNA sequence analysis (Woese, 1987) or restriction enzyme digestion (ARDRA; Vaneechoutte *et al*., 1993) have become more useful approaches to assess the phylogenetic and taxonomic diversity of bacterial isolates (Heyndrickx *et al*., 1996; Moyer *et al*., 1996). However, the power of rDNA-based protocols resides at a low phylogenetic or taxonomic level of resolution, which is valuable for classifying bacteria from the genus to even the kingdom level but which is insufficient to classify bacteria at the (sub)species level (Woese, 1987; Fox *et al*., 1992; Stackebrandt & Goebel, 1994; Hauben *et al*., 1997). Therefore, we postulated that rapid molecular genomic fingerprinting methods would be valuable, high-resolution alternative approaches to the classification of bacteria, especially on the species-, subspecies- or strain-specific level (Louws *et al*., 1994, 1995, 1997; Janssen *et al*., 1996; Lin *et al*., 1996; Rademaker & de Bruijn, 1997; Rademaker *et al*., 1997, 1998; Bragard *et al*., 1997).

Recently described genomic fingerprinting methods such as repetitive-sequence-based (rep)-PCR (Versalovic *et al*., 1991, 1994, 1998; de Bruijn, 1992), random amplified polymorphic DNA (RAPD) or DNA amplification fingerprinting (DAF) (Caetano-Anollés *et al*., 1991) and AFLP genomic fingerprinting (Vos *et al*., 1995) analyses have been suggested as accurate approaches to determine taxonomic and/or phylogenetic relationships between bacteria (Janssen *et al*., 1996; Huys *et al*., 1996; Clerc *et al*., 1998). However, the correlation between genomic fingerprinting and DNA–DNA homology studies has not been investigated on a sufficiently large scale to validate the former approaches as an alternative to, or a pre-screen precursor of, DNA–DNA homology-based bacterial speciation methods.

In this study, we chose the genus *Xanthomonas* as a model system to examine the relationship between phylogenetic relationships in bacteria based on DNA–DNA homology studies (Vauterin *et al*., 1995) and complemented by a variety of phenotypic methods (Vauterin *et al*., 1991, 1995, 1996; Yang *et al*., 1993) to those derived from genomic structure (fingerprinting) analyses based on rep-PCR or AFLP genomic fingerprinting approaches.

The rep-PCR-generated genomic fingerprints of 177 *Xanthomonas* strains, representing all 20 DNA–DNA homology groups (Vauterin *et al*., 1995), and one *Stenotrophomonas maltophilia* strain were employed to derive dendrograms of genetic relatedness using the GelCompar program (Vauterin & Vauterin, 1992). In addition, a subset of 80 *Xanthomonas* strains, representing the same 20 DNA–DNA homology groups, was analysed by AFLP analysis. On the basis of the results obtained, we suggest that the genomic structure of a bacterium, as deduced from its genomic fingerprint, represents an accurate reflection of its taxonomic and phylogenetic position based on total genomic DNA–DNA hybridization values.

METHODS

Bacterial strains. The 180 strains used in this study are listed in Table 1. Totals of 178 and 80 strains were selected for rep-PCR genomic fingerprinting and AFLP analysis, respectively (Table 1). The *S. maltophilia* type strain was included in the study as a control. The *Xanthomonas* strains selected comprise 20 genospecies and over 80 pathogenic variants, originating from 40 different countries (Table 1), and were obtained from the BCCM/LMG culture collection of the Laboratorium voor Microbiologie, Universiteit Gent, Ghent, Belgium.

DNA–DNA hybridization analysis. The global DNA–DNA homology matrix of 791 values, derived as described by De Ley *et al*. (1970) from the same 180 strains used for genomic fingerprinting, was obtained from Vauterin *et al*. (1995). Each DNA–DNA homology value represents the mean of between two and eight independent experiments (Vauterin *et al*., 1995).

Genomic fingerprinting. DNA was extracted from the bacterial strains as described by Rademaker & de Bruijn (1997). BOX-PCR (based on primers targeting the highly conserved repetitive DNA sequences of the BOXA subunit of the BOX element of *Streptococcus pneumoniae*), ERIC-PCR (based on primers targeting the highly conserved enterobacterial repetitive intergenic consensus) and REP-PCR (based on primers targeting the repetitive extragenic palindromic sequence) genomic fingerprints were obtained as described by Rademaker & de Bruijn (1997) and Rademaker *et al*. (1998). AFLP fingerprints were generated as described by Janssen *et al*. (1996), using restriction enzymes *Eco*RI and *Taq*I and one 3′ cytosine as selective base in both primers. Electrophoresis was performed in TBE buffer at 120 W and 50 °C for 2 h (Janssen *et al*., 1996).

Computer-assisted analysis of genomic fingerprints. Computer-assisted analysis of the genomic fingerprints was performed by using the commercially available GelCompar version 4.1 software program (Applied Maths) (Vauterin & Vauterin, 1992), as described previously (Rademaker & de Bruijn, 1997; Rademaker *et al*., 1997, 1999). Briefly, rep-PCR genomic fingerprints were analysed using a resolution of 400 and AFLP profiles of 1200 data points. Using the GelCompar software, the similarity between pairs of separate or linearly combined BOX-, ERIC- and REP-PCR

genomic fingerprints was calculated by using the product-moment correlation coefficient (*r* value; Pearson, 1926), applied to the whole densitometric curves of the gel tracks (Häne *et al.*, 1993; Rademaker & de Bruijn, 1997; Rademaker *et al.*, 1999). Four full 178-by-178 similarity matrices of the respective rep-PCR fingerprints and one 80-by-80 matrix of *r* values for AFLP patterns were obtained (31 684 and 6400 values, respectively). Cluster analysis of the pairwise similarity values was performed by using the UPGMA (unweighted pair group method using averages) algorithm (Sneath & Sokal, 1973).

Regression and other statistical analyses. The set of 791 DNA–DNA homology values derived from pairs of 180 strains (Vauterin *et al.*, 1995) was compared with the similarities between pairs of AFLP and separate or linearly combined BOX-, ERIC- and REP-PCR genomic fingerprints generated from the same 180 strains. The resulting data were analysed by using the BioNumerics (Applied Maths) and spss 7.5 (SPSS Inc.) software packages. Product-moment and Kendall's tau correlation values, scatterplots and second-degree regression curve fits were obtained for each combination of genomic fingerprint similarity values (*r* values) and the DNA–DNA homology values.

RESULTS AND DISCUSSION

The primary goal of this work was to determine whether clusters derived from genomic fingerprints corresponded to DNA–DNA homology-derived clusters and to determine the correlation between the primary similarity matrix values derived from genomic fingerprint patterns with actual DNA–DNA homology values (Fig. 1).

Cluster analysis of rep-PCR genomic fingerprints versus DNA–DNA homology groupings

First, BOX-, ERIC- and REP-PCR genomic fingerprints were generated from purified chromosomal DNA of 177 xanthomonad strains used previously for DNA–DNA hybridization experiments (Vauterin *et al.*, 1995), including more than 80 pathovars of 20 genomic species and one strain classified as *S. maltophilia* (Table 1). The rep-PCR-amplified bacterial genomic fragments were resolved on agarose gels, resulting in complex fingerprint patterns with multiple bands of distinct intensities (Fig. 2). These patterns were analysed with the GelCompar software package (Vauterin & Vauterin, 1992), as described by Rademaker *et al.* (1999). Pearson's product-moment correlation (*r*) values were calculated separately for the BOX-, ERIC- and REP-PCR genomic fingerprints and for the three linearly combined patterns (BOX-ERIC-REP; referred to as BER; Fig. 2). Twenty-five distinct and coherent clusters were observed after computer-assisted pattern analysis of the separate BOX-, ERIC- and REP-PCR genomic fingerprints and, to an even more precise extent, of the three linearly combined profiles (Fig. 2; boxes numbered 1 to 8, 9-1 to 9-6 and 10 to 20). Groups 1–8 and 10–20 (Fig. 2) were found to correspond precisely to those identified by DNA–DNA homology studies (DNA homology groups 1–8 and 10–20; Vauterin *et al.*, 1995).

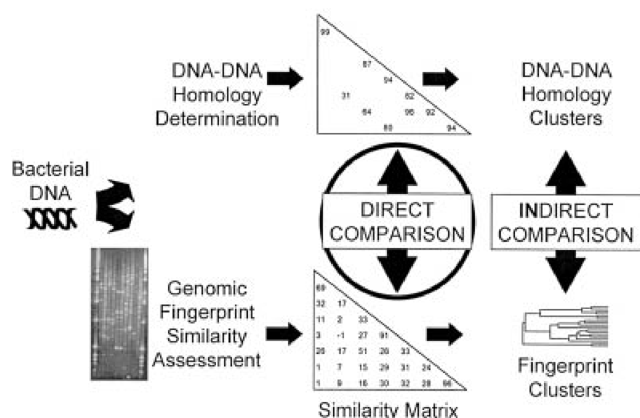

Fig. 1. Scheme illustrating direct and indirect comparison of genomic fingerprint analyses and DNA–DNA homology studies.

Group 9 (*Xanthomonas axonopodis*), as identified by Vauterin *et al.* (1995), contains the strains of the collection with the lowest DNA–DNA homology values and was also found to be heterologous on the basis of rep-PCR genomic fingerprinting. Group 9 includes six genetic clusters on the basis of rep-PCR genomic fingerprints (9.1, 9.2 etc.; Fig. 2). Individual BOX-, ERIC- or REP-PCR genomic fingerprint data sets tended to group strains into similar subclusters with a few exceptions. A more detailed rep-PCR genomic fingerprinting analysis, using a much larger collection of strains from DNA–DNA homology group 9, will be presented elsewhere.

Cluster analysis of AFLP genomic fingerprints versus DNA–DNA homology groupings

AFLP fingerprints were generated from 82 xanthomonad strains belonging to the collection of strains analysed by DNA–DNA homology studies (Vauterin *et al.*, 1995) and rep-PCR genomic fingerprinting (this study), including the pathotype of each of the 20 genomic species (Table 1) and the *S. maltophilia* strain. The AFLP genomic fingerprint patterns obtained were found to be similar in complexity to the rep-PCR-generated patterns (Fig. 3), and were analysed by using the GelCompar software. Similarity values between the AFLP fingerprints were calculated. Subsequent cluster analysis yielded 24 groups, 19 of which corresponded to clusters identified by DNA–DNA homology studies (Fig. 3; boxes numbered 1 to 8, 9-1 to 9-5 and 10 to 20).

The AFLP clusters were found to be virtually identical to those generated by rep-PCR genomic fingerprinting (Figs 2 and 3). Strains classified within all but two DNA–DNA homology groups formed the same coherent clusters in AFLP and rep-PCR genomic fingerprinting analyses, the exceptions being DNA–DNA homology groups 9 and 4. Strains classified within

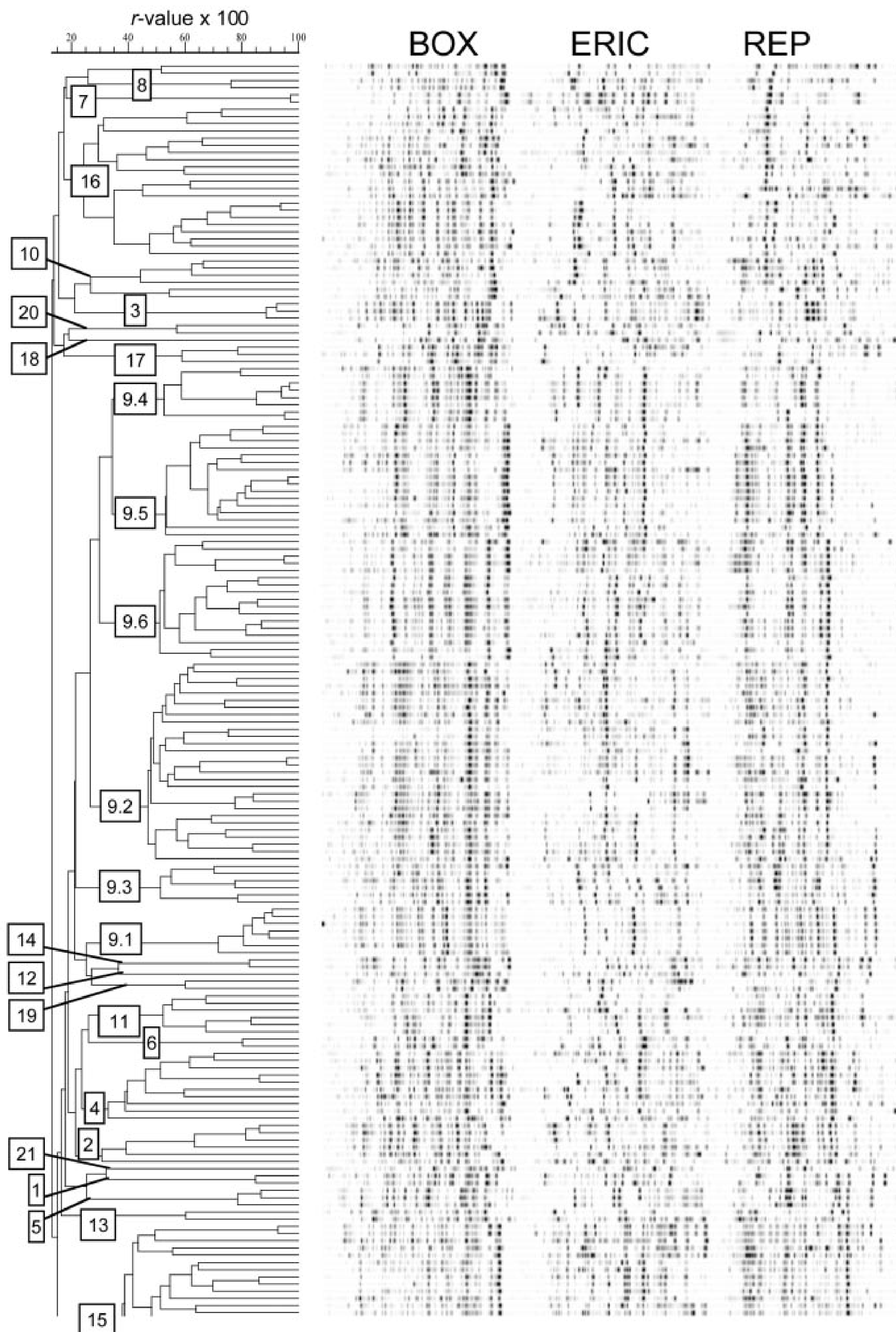

Fig. 2. UPGMA/product-moment cluster analysis of 178 linearly combined BOX-, ERIC- and REP-PCR-generated genomic fingerprints of the *Xanthomonas* strains listed in Table 1. The DNA–DNA homology groups are those determined by Vauterin *et al*. (1995) and subclusters within group 9 are as indicated.

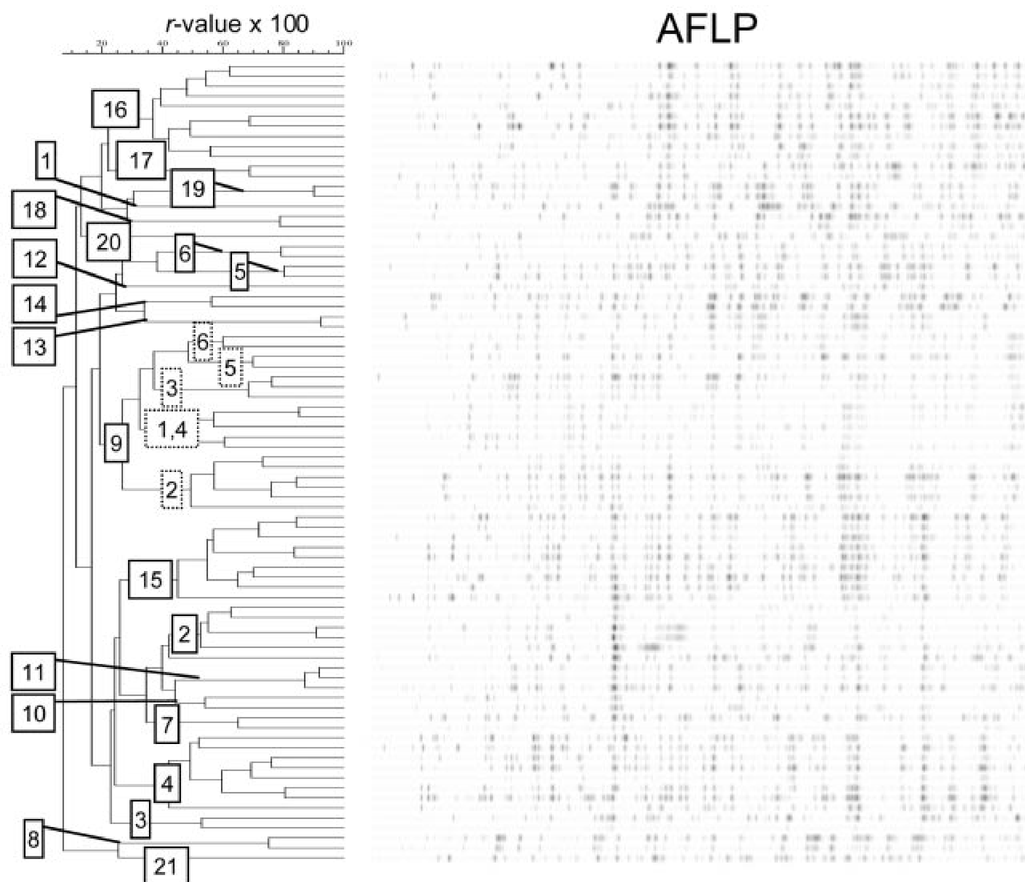

Fig. 3. UPGMA/product-moment cluster analysis of 80 AFLP genomic fingerprints of the *Xanthomonas* strains listed in Table 1. DNA–DNA homology groups (Vauterin *et al.*, 1995) are shown. Moreover, subgroups of DNA–DNA homology group 9, as determined by rep-PCR genomic fingerprinting (see Fig. 2) are indicated in boxes with dotted lines.

DNA–DNA homology group 9 were found to be more similar to each other than to strains in any other group, but at a substantial lower level than within any other DNA–DNA homology group. Moreover, some of the subclusters found by AFLP analysis belonging to DNA–DNA homology group 9 corresponded to those observed by rep-PCR analysis (Figs 2 and 3; boxes numbered 9-1 to 9-6). *Xanthomonas arboricola* pv. *populi* strain LMG 12141 was separated from the other members of DNA–DNA homology group 4 in the cluster analysis of the AFLP patterns. This observation is supported by similar results obtained by analysing BOX-PCR fingerprinting patterns (data not shown; Rademaker *et al.*, 1997).

Correlation between similarity matrices using the product-moment correlation coefficient

The topology of the dendrograms obtained by rep-PCR and AFLP genomic fingerprinting analyses was found to be in excellent agreement with the groups obtained by DNA–DNA homology studies. However, groupings derived from fingerprint similarity and DNA–DNA homology values could be regarded as 'secondary data', while similarity matrices derived from fingerprint patterns and initial DNA–DNA homology matrices constitute 'primary data', which are more suitable for direct comparison (Fig. 1). Therefore, a correlation study of the relevant primary data was carried out using the estimates of genetic similarity from rep-PCR or AFLP genomic fingerprints and DNA–DNA homology values as input.

First, the Pearson's product-moment correlation coefficient was employed, since it permitted a direct comparison of our results with those from other studies (see below). Subsequently, the non-parametric Kendall's tau coefficient was used.

Product-moment (r) values were calculated between the DNA–DNA homology matrix and the corresponding similarity matrices of the respective genomic fingerprints from BOX-, ERIC- and REP-PCR-, BER- and AFLP-based analyses (see above). A highly significant correlation was observed between the

Table 2. Pearson's product-moment and non-parametric Kendall's tau correlation coefficients between and among the DNA–DNA homology matrix and similarity matrices of AFLP or rep-PCR genomic fingerprints

The observed correlation values are all significant at the $P = 0.001$ level (two-tailed). n is the number of values used in each analysis. DNA indicates the DNA–DNA homology matrix.

Matrix	BOX	ERIC	REP	BER	AFLP
Pearson's product-moment					
ERIC	0·712				
REP	0·774	0·693			
BER	0·909	0·849	0·890		
AFLP	0·722	0·680	0·713	0·781	
DNA	0·779	0·669	0·777	0·808	0·838
Kendall's tau					
ERIC	0·404				
REP	0·478	0·374			
BER	0·670	0·537	0·625		
AFLP	0·312	0·321	0·340	0·353	
DNA	0·522	0·436	0·549	0·566	0·522
n					
ERIC	732				
REP	732	732			
BER	732	732	732		
AFLP	294	294	294	294	
DNA	732	732	732	732	322

DNA–DNA homology matrix and corresponding fingerprint similarity matrices (Table 2). This comparison was carried out on the primary data, without interference or bias introduced by clustering methods or classification schemes. The observed correlations of fingerprint similarities with DNA–DNA homology values were found to be high and statistically significant ($P = 0.0001$, two-tailed test), ranging from 0·669 for ERIC to 0·838 for AFLP genomic fingerprint analyses.

The correlation between duplicate DNA–DNA homology series ($n = 751$) was found to be 0·942, with the same level of significance ($P = 0.0001$; for a survey of the literature on experimental error in DNA–DNA pairing, see Hartford & Sneath, 1990). The correlation of the combined BER fingerprint similarities with DNA–DNA homology values (0·808) was found to be higher than those found with individual BOX (0·779), ERIC (0·669) or REP (0·777) data sets. This probably reflects the larger number of fragments included in the pattern analysis of BER versus single fingerprints. It may also reflect a possible uneven distribution of BOX, ERIC or REP primer-annealing sites around bacterial genomes. It has been observed in *Escherichia coli*, for example, that ERIC and REP sequences can occur in clusters and that, in specific areas of the genome, too few closely arrayed ERIC or REP elements may be present to generate REP-element-to-REP-element or ERIC-element-to-ERIC-element amplification products (Lupski & Weinstock, 1992).

The high correlation observed when using the primary data was found to be in concordance with the corresponding topologies of the dendrograms (Figs 2 and 3 and data not shown). Correlation values between the rep-PCR fingerprinting methods ranged from 0·693 (ERIC/REP) to 0·909 (BOX/BER); r values between AFLP and rep-PCR fingerprinting methods ranged from 0·680 (ERIC/AFLP) to 0·781 (BER/AFLP) (Table 2). This is in contrast with the results obtained by Gonzalez *et al.* (1998), who found a very low correlation value (0·15) between AFLP (four combinations) and RAPD (10 primers) fingerprinting of *Colletotrichicum lindemuthianum*. These authors excluded the possibility that their results were based on the complexity differences (number of bands) generated by the two fingerprinting techniques.

The high correlation we observed between rep-PCR-generated and AFLP-based genomic fingerprinting results may be due to the fact that the different techniques are based on either naturally occurring repetitive elements or restriction sites, in contrast to RAPD analysis, which is based on the occurrence of random sequences that may have less biological significance. However, Clerc *et al.* (1998) found that genetic distance similarity values derived from RAPD data versus a simplified AFLP protocol (each with 12 primers) correlated well in a study of '*Pseudomonas tomato*' strains ($r = 0.76$). These authors reported that the two methods provided essentially the same results and are, therefore, both useful for exploring genetic diversity. They also concluded that AFLP analysis delineated genospecies more effectively than did RAPDs. Moreover, Clerc *et al.* (1998) suggested that the AFLP method but not the RAPD method could be used as a rapid and efficient alternative to DNA–DNA homology experiments to verify the nature of new isolates belonging to the genospecies '*P. tomato*'.

Correlation between similarity matrices using Kendall's tau coefficient

The congruence between two different taxonomic or phylogenetic methods is traditionally estimated by using the Pearson product-moment correlation coefficient, which has some disadvantages. Firstly, the product-moment correlation only evaluates linear relationships. Therefore, a non-linear relationship (see Fig. 5) will lower the Pearson correlation, even if the relationship is without scatter of data points. The relationship between the similarity values of two different methods is often non-linear. Secondly, the Pearson correlation, being a parametric test, is very sensitive to the presence of outlying data points and, therefore, a few very aberrant data points can influence the final results substantially. Thirdly, the product moment is an inadequate statistical parameter for

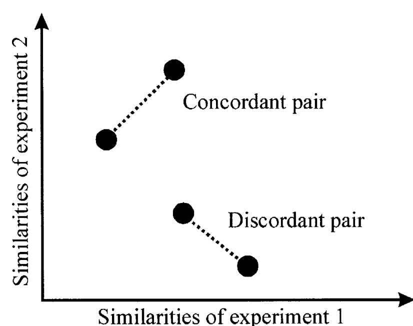

Fig. 4. Concordant and discordant pairs of similarity values (for explanation see text).

deciding whether the correlation is significant up to a certain confidence level. Unjustifiable assumptions may have to be made for the application of the product moment, e.g. that the data set follows a two-dimensional Gaussian distribution function.

For these reasons, a non-parametric statistic is a valuable alternative. Here, we applied Kendall's tau coefficient (Kendall, 1970). This coefficient ranks and considers all possible pairs of similarity values and counts the number of concordant and discordant pairs. A pair is called concordant if both techniques agree upon the ordering of the similarities between the corresponding organisms and discordant if they do not (Fig. 4).

Kendall's tau value is defined as (concordant − discordant)/(concordant + discordant) values (Kendall, 1970). This value will be high (close to one) in the case of a positive relationship, positive meaning that both values are increasing in the same direction. Because it only involves the relative orderings of similarity values, it is relatively insensitive to 'outliers'. Moreover, the statistical significance of the relationship can be calculated directly without any further assumptions. The observed congruence of the genomic fingerprinting methods with DNA–DNA homology values using Kendall's tau coefficient was found to be highly significant ($P = 0{\cdot}0001$, two-tailed test; Table 2). The closest relationships were observed when comparing DNA–DNA homology values with BER (0·566) or REP (0·549) genomic fingerprint similarity values. The comparison between AFLP or BOX-PCR fingerprint similarity and DNA–DNA homology values yielded a Kendall's tau coefficient of 0·522 (Table 2). The lowest Kendall's tau coefficient value (0·436) was observed with ERIC-PCR genomic fingerprint similarities versus DNA–DNA homology values, as observed with the product moment coefficient (see above). Among genomic fingerprints, the lowest Kendall's tau value was found when comparing AFLP and rep-PCR fingerprint similarity matrices (Table 2). Among the individual rep-PCR fingerprint similarity matrices, the largest Kendall's tau coefficient (0·478) was observed when comparing REP- and BOX-PCR genomic fingerprint matrices.

Regression analysis

Since DNA–DNA homology values and genomic fingerprint similarity values between pairs of strains lead to highly similar groupings, we attempted to describe their relationship. While the previous section dealt with overall levels of correlation between the DNA–DNA homology matrix and genomic fingerprint similarity matrices and among the fingerprint similarity matrices, we also explored their specific relation.

DNA–DNA homology values were plotted against the product-moment-generated similarities between pairs of BOX-, ERIC- and REP-PCR, BER and AFLP genomic fingerprints (Fig. 5). A second-degree polynomial was fitted to the data, subject to the constraints of being monotonically increasing and passing through the (100 %, 100 %) position. The relationships between DNA–DNA hybridization and each of the fingerprinting techniques were found to be comparable. The regression analysis highlighted important relationships among the different data sets (Fig. 5). For example, in the case of BOX-PCR-generated fingerprints, essentially all similarity (r) values above 0·5 correlated well with DNA–DNA hybridization values of more then 50 %, r values above 0·7 correlated with homology values above 70 % and the product-moment values above 0·8 correlated with DNA–DNA homology values above 85 % (Fig. 5a). The regression curves between DNA–DNA homology values and ERIC- or REP-PCR similarity values were found to be very similar to those obtained for BOX-PCR similarity values (curves not shown).

The relationship between DNA–DNA homology values and the r values of BER analysis is shown in Fig. 5(b). The scatter in this plot is lower than that observed in plots of the individual rep-PCR fingerprint similarities and is, in fact, directly comparable with the fit between duplicate DNA–DNA homology values (data not shown). Essentially all BER fingerprint similarity values above 0·40 were found to correlate with DNA–DNA homology values exceeding 50 %, above 0·5 with values exceeding 70 % and above 0·8 with values exceeding 90 %.

The relationship between DNA–DNA homology values and the r values of AFLP fingerprints is shown in Fig. 5(c). In essence, AFLP similarity values above 0·50 correlated with DNA–DNA homology values exceeding 50 %, above 0·65 with values exceeding 70 % and above 0·8 with values exceeding 85 %. The trend line for AFLP fingerprint similarities versus DNA–DNA homology values is based on a smaller number of data sets (322), but was still found to be comparable to the regression curves based on the collection of 732 similarity values (Table 2). Even an analysis of only 160 pairs of similarity values resulted in trend lines

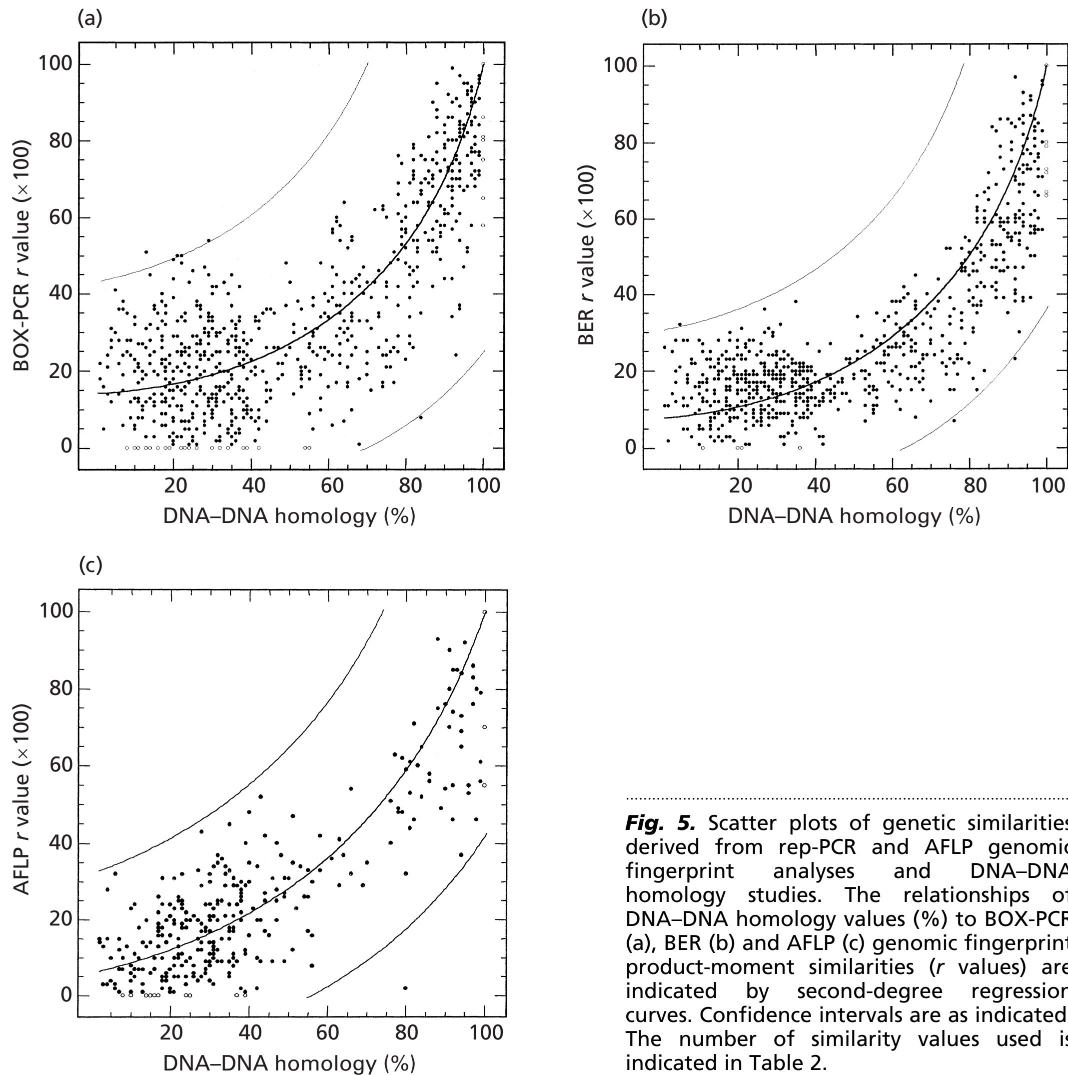

Fig. 5. Scatter plots of genetic similarities derived from rep-PCR and AFLP genomic fingerprint analyses and DNA–DNA homology studies. The relationships of DNA–DNA homology values (%) to BOX-PCR (a), BER (b) and AFLP (c) genomic fingerprint product-moment similarities (*r* values) are indicated by second-degree regression curves. Confidence intervals are as indicated. The number of similarity values used is indicated in Table 2.

essentially identical to those generated based on a larger number of similarity values, not only for AFLP but also for all rep-PCR fingerprint types presented here (data not shown).

The trend lines between the different genomic fingerprinting techniques were also determined (data not shown). The product-moment correlation between BER and the individual rep-PCR fingerprints was found to be high (0·909, 0·849 and 0·890 for BER versus BOX-, ERIC- and REP-PCR fingerprints). The linear relationship was also found to have an intercept close to zero and a slope of approximately one, suggesting that the corresponding similarity values obtained from these data sets are almost identical. The lowest scatter was found with BOX-PCR fingerprint similarity values, supporting earlier observations that BOX-PCR may be the most robust of the three rep-PCR methods analysed here. The highest data point scatter was observed with ERIC-PCR fingerprint similarity values, which are more sensitive to disturbances.

Conclusions

rep-PCR- or AFLP-generated genomic fingerprint analyses yield results that are in close agreement with DNA–DNA homology studies, as determined by cluster analyses (indirect), as well as by direct comparison of primary similarity values. The observed high correlation between DNA–DNA homology studies and rep-PCR and AFLP genomic fingerprint analyses suggests that genomic fingerprint methods can function as core techniques in polyphasic taxonomy, regardless of the statistical model employed. The conclusion from the comparison of rep-PCR genomic fingerprinting with DNA–DNA homology studies is supported by the results of a smaller-scale study comparing rep-PCR genomic fingerprinting and

dot-blot-mediated DNA–DNA homology studies (Nick *et al.*, 1999). With regard to AFLP genomic fingerprinting versus DNA–DNA homology studies, our conclusions are also supported by other studies on a variety of bacteria (Huys *et al.*, 1996; Janssen *et al.*, 1996, 1997; Janssen & Dijkshoorn, 1996; Clerc *et al.*, 1998; Pedersen *et al.*, 1998), although some controversy exists about this issue (Esteve, 1997).

Studies carried out by our laboratory and others on a variety of different bacterial genera and species have revealed that, at fine taxonomic resolution, phylogenetic trees derived from BOX-, ERIC- and REP-PCR genomic fingerprinting are not always identical. This is to be expected, since different numbers of bands may be generated with each primer set, the annealing conditions vary between primers or sets, and the prevalence/distribution of the target repetitive elements in question may vary. A similar variation may occur with restriction-enzyme-dependent fingerprinting methods like RFLP, ARDRA and AFLP, depending on the restriction enzyme that is used. However, the phylogenetic 'trees' derived using each primer set separately are only slightly different. Moreover, our studies have shown that, regardless of the individual primer set used for rep-PCR genomic fingerprinting, the correlation of the phylogenetic trees with those derived from DNA–DNA homology studies is highly significant (Table 2). However, we want to emphasize here that the most highly significant and consistent results were obtained by combining the data obtained from BOX-, ERIC- and REP-PCR fingerprints (BER). This is logical, since the total number of data points (bands/peaks) is greatly increased. Moreover, the genome is more extensively covered, since certain regions may have more (properly spaced) copies of a particular element than others (Lupski & Weinstock, 1992). Informative rep-PCR and/or AFLP genomic fingerprint patterns have been generated from all Gram-negative and several Gram-positive bacteria analysed so far in studies that support our conclusions [see Versalovic *et al.* (1994) and Rademaker *et. al.* (1998) for lists of bacterial species tested]. To our knowledge, the *Archaea* have not yet been examined by rep-PCR genomic fingerprinting.

Similarities between genomic fingerprints, as well as DNA–DNA homology values, represent indirect measurements of sequence homologies and are dependent on the specific experimental conditions and analysis methods employed, and it is recognized that a hierarchy exists in terms of general applicability and standardization. Total DNA–DNA hybridization experiments are likely to be the most general, followed by AFLP and rep-PCR fingerprinting analyses. However, we stress in this paper that, in the bacterial species where significant comparative studies have been carried out, the results are comparable. Most important, however, is the applicability of the methods used to the analysis of large(r) collections of strains. There, an opposite (inverse) hierarchy exists, which demonstrates clearly that rep-PCR genomic fingerprinting is the least experimentally demanding and possibly the most discriminatory method of choice.

In any case, rep-PCR and AFLP genomic fingerprinting techniques appear to reflect the genotypic, phylogenetic and taxonomic relationships of organisms and, therefore, we propose that these genomic fingerprinting techniques can be used as a rapid means of determining taxonomic diversity and phylogenic structure, especially of large collections of bacterial isolates.

ACKNOWLEDGEMENTS

The development of the rep-PCR genomic fingerprinting method for the analysis of plant-associated and soil microbes has been supported by the DOE (DE FG 0290ER20021), the NSF Center for Microbial Ecology (DIR 8809640), Heinz, Roger Seeds and Procter & Gamble and by the Consortium for Plant Biotechnology Research (DE – FC05-92OR22072) and the states of Michigan and North Carolina. We also gratefully acknowledge Maria Schneider, Jessica Wallace, Robert Jaros, Junko Yasuda (MSU), Lysiane Hauben, Sylvie VanEygen (RUG, Belgium) and Jim Lupski (Baylor College of Medicine, Houston) for technical support and many useful discussions on rep-PCR genomic fingerprinting.

REFERENCES

Bragard, C., Singer, E., Alizadeh, A., Vauterin, L., Maraite, H. & Swings, J. (1997). *Xanthomonas translucens* from small grains: diversity and phylopathological relevance. *Phytopathology* **87**, 1111–1117.

de Bruijn, F. J. (1992). Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergenic consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Appl Environ Microbiol* **58**, 2180–2187.

Caetano-Anollés, G., Bassam, B. J. & Gresshoff, P. M. (1991). DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Biotechnology* **9**, 553–557.

Clerc, A., Manceau, C. & Nesme, X. (1998). Comparison of randomly amplified polymorphic DNA with amplified fragment length polymorphism to assess genetic diversity and genetic relatedness within genospecies III of *Pseudomonas syringae*. *Appl Environ Microbiol* **64**, 1180–1187.

Colwell, R. R. (1970). Polyphasic taxonomy of bacteria. In *Culture Collections of Microorganisms*, pp. 421–436. Edited by H. Iizuka & T. Hasegawa. Tokyo: University of Tokyo Press.

De Ley, J., Cattoir, H. & Reynaerts, A. (1970). The quantitative measurement of DNA hybridization from renaturation rates. *Eur J Biochem* **12**, 133–142.

Esteve, C. (1997). Is AFLP fingerprinting a true alternative to the DNA-DNA pairing method to assess genospecies in the genus *Aeromonas*? *Int J Syst Bacteriol* **47**, 245–246.

Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., Jr (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**, 166–170.

González, M., Rodríguez, R., Zavala, M. E., Jacobo, J. L., Hernández, F., Acosta, J., Martínez, O. & Simpson, J. (1998). Characterization of Mexican isolates of *Colletotrichum lindemuthianum* by using differential cultivars and molecular markers. *Phytopathology* **88**, 292–299.

Häne, B. G., Jäger, K. & Drexler, H. G. (1993). The Pearson

product-moment correlation coefficient is better suited for identification of DNA fingerprinting profiles than band matching algorithms. *Electrophoresis* **14**, 967–972.

Hartford, T. & Sneath, P. H. (1990). Experimental error in DNA-DNA pairing: a survey of the literature. *J Appl Bacteriol* **68**, 527–542.

Hauben, L., Vauterin, L., Swings, J. & Moore, E. R. B. (1997). Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int J Syst Bacteriol* **47**, 328–335.

Heyndrickx, M., Vauterin, L., Vandamme, P., Kersters, K. & De Vos, P. (1996). Applicability of combined amplified ribosomal DNA restriction analysis (ARDRA) patterns in bacterial phylogeny and taxonomy. *J Microbiol Methods* **26**, 247–259.

Huys, G., Coopman, R., Janssen, P. & Kersters, K. (1996). High-resolution genotypic analysis of the genus *Aeromonas* by AFLP fingerprinting. *Int J Syst Bacteriol* **46**, 572–580.

Janssen, P. & Dijkshoorn, L. (1996). High resolution DNA fingerprinting of *Acinetobacter* outbreak strains. *FEMS Microbiol Lett* **142**, 191–194.

Janssen, P., Coopman, R., Huys, G., Swings, J., Bleeker, M., Vos, P., Zabeau, M. & Kersters, K. (1996). Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology* **142**, 1881–1893.

Janssen, P., Maquelin, K., Coopman, R., Tjernberg, I., Bouvet, P., Kersters, K. & Dijkshoorn, L. (1997). Discrimination of *Acinetobacter* genomic species by AFLP fingerprinting. *Int J Syst Bacteriol* **47**, 1179–1187.

Kendall, M. G. (1970). *Rank Correlation Methods*, 3rd edn. London: Charles Griffin.

Lin, J. J., Kuo, J. & Ma, J. (1996). A PCR-based DNA fingerprinting technique: AFLP for molecular typing of bacteria. *Nucleic Acids Res* **24**, 3649–3650.

Louws, F. J., Fulbright, D. W., Stephens, C. T. & de Bruijn, F. J. (1994). Specific genomic fingerprints of phytopathogenic *Xanthomonas* and *Pseudomonas* pathovars and strains generated with repetitive sequences and PCR. *Appl Environ Microbiol* **60**, 2286–2295.

Louws, F. J., Fulbright, D. W., Stephens, C. T. & de Bruijn, F. J. (1995). Differentiation of genomic structure by rep-PCR fingerprinting to rapidly classify *Xanthomonas campestris* pv. *vesicatoria*. *Phytopathology* **85**, 528–536.

Louws, F. J., Schneider, M. & de Bruijn, F. J. (1997). Assessing genetic diversity of microbes using repetitive sequence-based PCR (rep-PCR). In *Nucleic Acid Amplification Methods for the Analysis of Environmental Microbes*, pp. 63–94. Edited by G. Toranzos. Lancaster, PA: Technomic Publishing.

Lupski, J. R. & Weinstock, G. M. (1992). Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J Bacteriol* **174**, 4525–4529.

Moyer, C. L., Tiedje, J. M., Dobbs, F. C. & Karl, D. M. (1996). A computer-simulated restriction fragment length polymorphism analysis of bacterial small-subunit rRNA genes: efficacy of selected tetrameric restriction enzymes for studies of microbial diversity in nature. *Appl Environ Microbiol* **62**, 2501–2507.

Murray, R. G. E., Brenner, D. J., Colwell, R. R., De Vos, P., Goodfellow, M., Grimont, P. A. D., Pfennig, N., Stackebrandt, E. & Zavarzin, G. A. (1990). Report of the ad hoc committee on approaches to taxonomy within the Proteobacteria. *Int J Syst Bacteriol* **40**, 213–215.

Nick, G., Jussila, M., Hoste, B., Niemi, M., Kaijalainen, S., de Lajudie, P., Gillis, M., de Bruijn, F. J. & Lindström, K. (1999). Rhizobia isolated from root nodules of tropical leguminous tree characterized using DNA-DNA dot-blot hybridisation and rep-PCR genomic fingerprinting. *Syst Appl Microbiol* **22**, 287–299.

Pearson, K. (1926). On the coefficient of racial likeness. *Biometrika* **18**, 105–117.

Pedersen, K., Verdonck, L., Austin, B. and 9 other authors (1998). Taxonomic evidence that *Vibrio carchariae* Grimes *et al*. 1985 is a junior synonym of *Vibrio harveyi* (Johnson and Shunk 1936) Bauman *et al*. 1981. *Int J Syst Bacteriol* **48**, 749–758.

Rademaker, J. L. W. & de Bruijn, F. J. (1997). Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer assisted pattern analysis. In *DNA Markers: Protocols, Applications and Overviews*, pp. 151–171. Edited by G. Caetano-Anollés & P. M. Gresshoff. New York: John Wiley.

Rademaker, J. L. W., Louws, F. J., Schultz, M. H., Rossbach, U., Vauterin, L., Swings, J. & de Bruijn, F. J. (1997). Molecular systematics of xanthomonads by rep-PCR genomic fingerprinting and computer-assisted pattern analysis. *Phytopathology* **87**, S81.

Rademaker, J. L. W., Louws, F. J. & de Bruijn, F. J. (1998). Characterization of the diversity of ecologically important microbes by rep-PCR genomic fingerprinting. In *Molecular Microbial Ecology Manual*, supplement 3, chapter 3.4.3, pp. 1–26. Edited by A. D. L. Akkermans, J. D. van Elsas & F. J. de Bruijn. Dordrecht: Kluwer.

Rademaker, J. L. W., Louws, F. J., Rossbach, U., Vinuesa, P. & de Bruijn, F. J. (1999). Computer-assisted pattern analysis of molecular fingerprints and database construction. In *Molecular Microbial Ecology Manual*, supplement 4, chapter 7.1.3, pp. 1–33. Edited by A. D. L. Akkermans, J. D. van Elsas & F. J. de Bruijn. Dordrecht: Kluwer.

Sneath, P. H. A. & Sokal, R. R. (1973). *Numerical Taxonomy*. San Francisco: Freeman.

Stackebrandt, E. & Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**, 846–849.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* **60**, 407–438.

Vaneechoutte, M., De Beenhouwer, H., Claeys, G., Verschraegen, G., De Rouck, A., Paepe, N., Elaichouni, A. & Portaels, F. (1993). Identification of *Mycobacterium* species by using amplified ribosomal DNA restriction analysis. *J Clin Microbiol* **31**, 2061–2065.

Vauterin, L. & Vauterin, P. (1992). Computer-aided objective comparison of electrophoresis patterns for grouping and identification of microorganisms. *Eur Microbiol* **1**, 37–41.

Vauterin, L., Swings, J. & Kersters, K. (1991). Grouping of *Xanthomonas campestris* pathovars by SDS-PAGE of proteins. *J Gen Microbiol* **137**, 1677–1687.

Vauterin, L., Hoste, B., Kersters, K. & Swings, J. (1995). Reclassification of *Xanthomonas*. *Int J Syst Bacteriol* **45**, 472–489.

Vauterin, L., Yang, P., Alvarez, A., Takikawa, Y., Roth, D. A., Vidaver, A. K., Stall, R. E., Kersters, K. & Swings, J. (1996). Identification of non-pathogenic *Xanthomonas* strains associated with plants. *Syst Appl Microbiol* **19**, 96–105.

Versalovic, J., Koeuth, T. & Lupski, J. R. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **19**, 6823–6831.

Versalovic, J., Schneider, M., de Bruijn, F. J. & Lupski, J. R. (1994).

Genomic fingerprinting of bacteria using repetitive sequence based PCR (rep-PCR). *Methods Mol Cell Biol* **5**, 25–40.

Versalovic, J., de Bruijn, F. J. & Lupski, J. R. (1998). Repetitive sequence-based PCR (rep-PCR) DNA fingerprinting of bacterial genomes. In *Bacterial Genomes: Physical Structure and Analysis*, pp. 437–454. Edited by F. J. de Bruijn, J. R. Lupski & G. M. Weinstock. New York: Chapman & Hall.

Vos, P., Hogers, R., Bleeker, M. and 8 other authors (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**, 4407–4414.

Wayne, L. G., Brenner, D. J., Colwell, R. R. and 9 other authors (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221–271.

Yang, P., Vauterin, L., Vancanneyt, M., Swings, J. & Kersters, K. (1993). Application of fatty acid methyl esters for the taxonomic analysis of the genus *Xanthomonas*. *Syst Appl Microbiol* **16**, 47–71.