Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 811 (2004) 263–268
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Short communication

# Identification and separation of PCR products based on their GC content by denaturing high-performance liquid chromatography

Eugeni Belda, Vicente Sentandreu, Francisco J. Silva*

*Departament de Genètica, Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartat 22085, 46071 Valencia, Spain*

Received 26 May 2004; accepted 14 September 2004
Available online 13 October 2004

**Abstract**

We show that denaturing high-performance liquid chromatography is a suitable method for the separation of DNA molecules of similar sizes but with different GC contents. A mixture of homologous molecules coming from different bacterial species may be obtained when PCR with degenerate primers is used for the amplification of a specific gene from an environmental sample. We have observed that, by selecting an appropriate temperature for the partial denaturation of the molecules, we are able to separate them according to the GC content of each DNA product. This allows us to determine if one or several types of molecules are amplified in the course of a PCR reaction. In the latter case it is possible, even with minority products, to isolate them by collecting the eluted volumes, followed by cloning, sequencing or reamplifying them by PCR, depending on the DNA concentration. We have applied this analysis to the amplification of a fragment of the *ribA* gene in the bacterial endosymbionts of insects, obtaining a high correlation coefficient (0.978) between retention time and the GC content of the molecules.


*Keywords:* Denaturing high-performance liquid chromatography; GC content; Bacterial endosymbionts; *Buchnera aphidicola*

## 1. Introduction

PCR is a potent technique that allows the amplification of DNA fragments from microorganisms without the need to isolate and culture each species. When degenerate primers are used to amplify a specific gene from bacterial species present in environmental samples, several situations may result. First, it is frequent to amplify more than one species simultaneously, which will force the cloning and screening of the clones, prior to sequencing in order to detect the different amplified products. If the major component of the PCR product comes from one bacterial species, but there are others with minor frequencies, a large screening will be required to isolate the minority clones. On the other hand, we may be interested in the amplification of a fixed bacterial species, but primers may amplify the gene from other unwanted species which are also present in the sample. Direct PCR product sequencing would be very fast and convenient but the putative presence of several kinds of DNA fragments will again require the cloning and screening of the cloned PCR products prior to the sequencing step.

A decade ago, the development of denaturing HPLC became the most sensitive screening method for the detection of point mutations or variations in the DNA [1,2]. Based on the temperature used in the analysis, short DNA molecules migrate through the chromatographic column at different speeds, depending on their interactions with the non-polar stationary phase and the ion-pair reagent. The ionic interactions of the negatively charged phosphates and the hydrophobic interactions of the nucleobases are the main DNA factors responsible for the retention time. Nucleobase hydrophobic interactions require the single strand state, and it is know that this state depends on the temperature, the nucleotide composition and the distribution of the bases along the DNA molecule.

* Corresponding author. Tel.: +34 963543650; fax: +34 9634543670.
*E-mail address:* francisco.silva@uv.es (F.J. Silva).

1570-0232/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2004.09.010

We have applied DHPLC to the separation and identification of PCR products in an attempt to determine if there is a correlation between the GC content of the molecule and its retention time [3]. We have applied this method to the analysis of the bacterial symbionts of insects. Most of these bacterial species are uncultured and, for that reason, PCR based on degenerate primers is a potent tool for the isolation of genes using DNA extracted from the whole insect body. This total DNA not only includes the insect one but that coming from the different symbiont associated bacterial species.

Aphids are an ideal group of insects for this study because most of the species contain an obligate endosymbiont, *Buchnera aphidicola*, which after an original infection up to 200 million years ago, has coevolved in parallel, always being transmitted vertically from parents to the offspring [4,5]. Many changes have been experienced by the genome of *Bu. aphidicola* due to its special lifestyle. One of them was the high increase in the AT content [6], while others were the loss of genes and the genome size reduction [7]. After its association with the ancestor of aphids, the endosymbiont of each aphid lineage experienced independent gene losses [7–10], the reason why the searched for gene is not always amplified in every *Bu. aphidicola* strain.

In addition, aphid species may carry facultative endosymbionts, which are often called secondary symbionts [11–14]. They belong to different taxonomic groups, but some of them are related to *Bu. aphidicola*, belonging to gamma-Proteobacteria. One important characteristic of the secondary endosymbionts is that they present intermediate AT contents in relation to many closely related free-living bacterial species such as *E. coli* or *Salmonella* spp. [13].

Here we show a method based on DHPLC that permits the identification of DNA molecules based on a strong correlation between their GC contents and retention times.

## 2. Experimental

### 2.1. DNA extraction

Insects from ten aphid species were used in this study (see Table 1) and were collected in several Spanish locations. The total aphid DNA, including that coming from its bacterial endosymbionts, was obtained as previously described [15].

### 2.2. PCR amplification

Two degenerate oligonucleotide primers were initially designed for the amplification of a 280 bp fragment of the *ribA* gene which encodes GTP cyclohydrolase II, an enzyme involved in riboflavin metabolism. The primers were designed on regions of the RibA protein with high amino acid conservation in gamma-Proteobacteria. Forward primer (ribAdF1, 5′-GAAGAAGCTATATTACCWACWYVWTKKGG-3′) and reverse primer (ribA-dR1, 5′-TARACCDATRTTNCKNCCYTCYTG-3′).

Table 1
Aphid species used for the extraction of total DNA

| Aphid species | Subfamily | Bacterial symbiont | |
|---|---|---|---|
| | | *Buchnera aphidicola* | Secondary symbionts |
| *Acyrthosiphum pisum* | Aphidinae | + | |
| *Baizongia pistaciae* | Pemphiginae | + | |
| *Cinara cedri* | Lachnidae | | + |
| *Eriosoma lanuginosum* | Pemphiginae | + | |
| *Panaphis juglandis* | Myzocallidinae | | + |
| *Pemphigus bursarius* | Pemphiginae | + | |
| *Periphyllus bulgaricus* | Chaitophorinae | | + |
| *Pterochloroides persicae* | Lachnidae | | + |
| *Rhopalosiphum padi* | Aphidinae | + | |
| *Tetraneura ulmi* | Pemphiginae | + | + (2) |

Subfamilies of the Aphididae which they belong. The symbol (+) indicates that the PCR amplification of the *ribA* gene was from *Bu. aphidicola* or a secondary symbiont. The *ribA* gene fragment from the endosymbiont of *Pe. bulgaricus* was not sequenced. The number 2 in parenthesis indicates that in *T. ulmi* we detected two different secondary endosymbionts.

PCR reaction mixture (50 μl) comprised 0.3 μl of *Taq* polymerase (5 U/μl) (Amersham, Pharmacia), 5 μl of 10× reaction buffer, 2 μl of each 10 μM primer solution, 2 μl of 5 μM dNTPs and DNA. PCR reactions were performed in GeneAmp PCR system 2400 (Perkin-Elmer) in the following conditions: Initial detanuration step of 2 min at 94 °C, 30 cycles of 20 s at 94 °C, 30 s at 48 °C and 20 s at 72 °C and a final step of 4 min at 72 °C.

### 2.3. DHPLC analysis

PCR product analyses were performed by DHPLC on a WAVE DNA Fragment Analysis System (Transgenomic). WAVEMaker software was initially used to predict the appropriate range of temperature to produce partial denaturation of *Bu. aphidicola* PCR products. Injection volumes were generally in the range of 5–10 μl. An analytical column DNASep (Transgenomic) was used. The mobile phase was composed of a binary gradient of buffer A (0.1 M triethylammonium acetate, TEAA) and buffer B (0.1 M TEAA/25% acetonitrile) (from 53 to 61% buffer B) at a flow rate of 0.9 ml/min. The products were detected by absorbance at 260 nm. Oven temperature was 56 °C for most of the analyses. A fraction collector CFC203B (Gilson) was used to collect the eluted volumes with time intervals of 15 s between 2 and 6 min of elution. In the case of *Tetraneura ulmi*, 10 μl of several fractions between 2.5 and 5.5 min were used for PCR reamplification with primers ribAdF1 and ribAdR1. The products of these PCR reactions were analysed by DHPLC (see Section 3). Two independent measurements of the retention time were determined for each type of DNA fragment.

### 2.4. DNA sequencing

PCR products were purified with the QIAquick PCR purification kit (Qiagen). Fractions eluted from DHPLC

were directly used for sequencing or previously concentrated with a SpeedVac (Savant), depending on the DNA concentration estimated from the peak area. Primers ribAF1 and ribAR1 were used for sequencing. Because of some problems with the forward primer, two additional internal forward primers were designed. One, expecting to amplify this gene from *Bu. aphidicola* strains (ribAdF4, 5′-AAAATCACRTTGCWYTWRTWTATGG-3′) and the other from species related to the already known secondary symbionts (ribAdF3, 5′-TTTGAAGAATTAGCKACKGGRMAYGA-3′). The inability of the latter to sequence the majority *T. ulmi* secondary symbiont forced the cloning of this PCR product in a pGEM T-easy plasmid (Promega) and the sequencing of it with plasmid primers SP6 and T7. Sequencing was carried out on ABI 3700 equipment (Applied Biosystems) with a BigDye v3.1 Cycle sequencing kit (Applied Biosystems).

### 2.5. DNA sequence analyses

STADEN package software (http://www.mrc-lmb.cam.ac.uk/pubseq/) was used for the assembly and visual inspection of the sequences. In order to make a primary identification of our sequences, they were compared with sequences present in public databases using BLAST [16]. A phylogenetic reconstruction by maximum likelihood was performed with the program TREEPUZZLE version 5.0. First and second nucleotide codon positions were used for the analysis. The quartet puzzling method [17] implemented in this program, was used to obtain the support for each internal branch. Accession numbers of the EMBL database for the nucleotide sequences reported in this work are AJ634362–AJ634640.

## 3. Results and discussion

PCR amplification with primers ribAdF1 and ribAdR1 from ten aphid species produced, in all cases, a single band of around 280 bp. This was the expected size for the fragment of the *ribA* gene based on the previously known sequences of several gamma-Proteobacterial species (273–276 bp). DHPLC analyses at temperatures ranging from 48 to 60 °C were performed with a linear gradient of 48–53% B in 0.1 min followed of 53–61% B in 4 min. The temperature of 56 °C was selected for further analyses because non complete denaturation was produced for all of the analysed fragments. At this temperature we expected that all molecules to be partially denatured, but to a different extent depending on their base composition.

The DHPLC profiles obtained with the different samples were highly variable, with elution time peaks ranging from about 2.5 to 5.0 min (Fig. 1). A single wide peak of short retention time (ca. 2.5 min) was observed in some samples. In contrast, other DHPLC profiles were composed of two or more peaks. The two-peak profile, or the profile with a peak and a shoulder, was observed in 7 out of the 10 samples. To determine the DNA molecules generating such peaks, a fraction collector was used to recover the DNA from the two peaks of a sample. The collected fractions were sequenced with the two flanking primers. This revealed that the two DNA peaks corresponded exactly to the same sequence. We could only detect a difference in two nucleotide positions at the 3′ end of the forward primer whose degenerations consisted of two K (G or T). While both were G in the sequence of the high retention time peak, in the low retention time peak both G and T were observed in the sequencing electropherograms. We have observed that WAVEMaker software predicts that

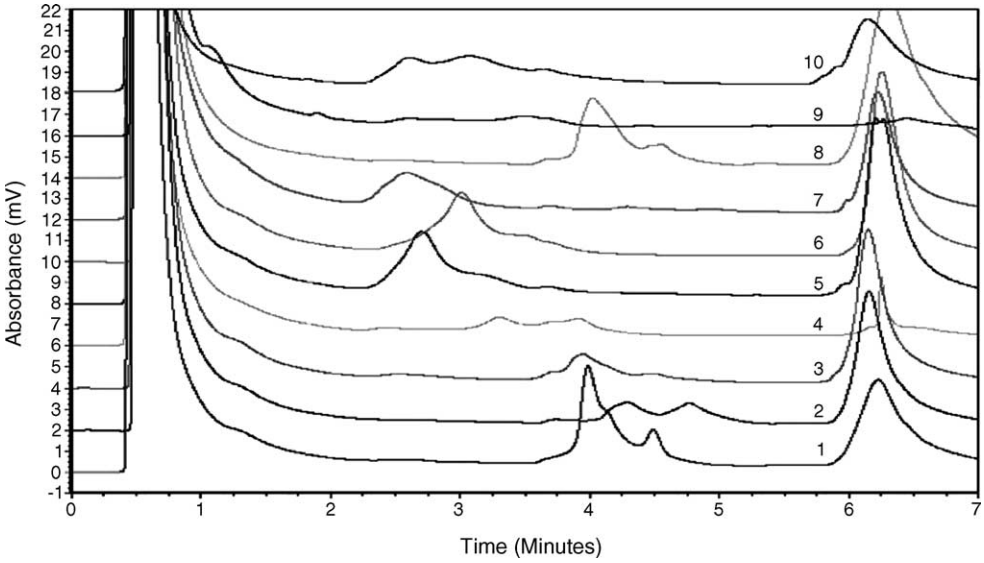


Fig. 1. DHPLC profiles of PCR products of a fragment of the *ribA* gene from bacterial endosymbionts of aphids: *C. cedri* (1), *Pa. juglandis* (2), *Pe. bulgaricus* (3), *A. pisum* (4), *P. bursarius* (5), *E. lanuginosum* (6), *T. ulmi* (7), *Pt. persicae* (8), *B. pistaciae* (9) and *R. padi* (10). Column, DNASep® Cartridge matrix of PS-DVB particles, 2.2 μm, 4.6 mm × 50 mm length; mobile phase, 0.1 M TEAA, pH 7.65, 25% acetonitrile; flow rate, 0.9 ml/min; temperature, 56 °C; detection, UV, 260 nm.

the region near the primer ribAdF1 is not denatured at 56 °C when these two G are present in the molecule, but that it is almost completely denatured if two T are present, with intermediate situations for one T and one G. For that reason, we consider that the different retention time peaks observed for a single PCR product are related to these two partially denatured DNA molecules. For our next analyses we will only compare the values of the shorter retention time peaks, which in most cases are the majority ones.

The DHPLC profile of the *T. ulmi* PCR product was more complex with a major peak at 2.6 min and two small peaks at 4.4 and 5.1 (marked with arrows at Fig. 2A); A system peak at 3.7 min is observed in these conditions in all DHPLC chromatograms. In order to separate the apparently different two DNA products observed in this chromatogram, we performed a new PCR with eluted fractions collected at different times. While the PCR reaction from fractions of around 3 min amplified the DNA with the shorter retention time, the PCR from samples of around 5 min produced an enrichment of peaks at 4–5 min compared with the 2.6 peak (Fig. 2B). After direct sequencing of the reamplified PCR products, we were able to identify the presence of three species. One was AT rich (peak at 2.6 min) and belonged to *Bu. aphidicola*. The peak at 4.3 min was shown to be majoritarily composed of a PCR product of 48% GC content but with a minor product of 43% GC. The existence of the minor product was revealed because one internal primer used for sequencing was unable to anneal the 48% GC DNA molecule and rendered the sequence of the minority product. In order to determine the exact retention time of the more GC rich product, we separated the 4–5 min enriched PCR product with DHPLC at higher temperatures, finding that at 63 °C we were able to observe their partial separation (Fig. 3A, arrow marks the peak corresponding to the partially denatured DNA molecule from a minority secondary endosymbiont from *T. ulmi*), while at 64.5 °C we produced the complete isolation of the majority product because, in contrast to the minority, it was already eluting as a partial denatured DNA molecule (Fig. 3B). We collected, cloned and sequenced the DNA from this fraction and estimated the exact retention time at 56 °C.

The remaining PCR products, except the one from *Periphyllus bulgaricus*, were sequenced and compared with

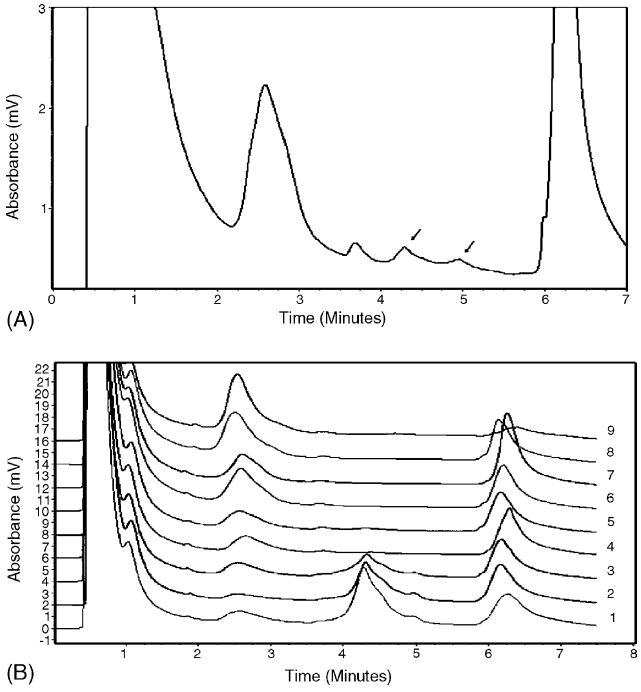


Fig. 2. DHPLC isolation of DNA molecules belonging to different bacterial endosymbionts of a single aphid. (A) Elution profile of the PCR product from *ribA* gene fragment from bacterial endosymbionts of *T. ulmi*. Peak at 2.6 min is produced by *Bu. aphidicola* DNA. Minority peaks marked with arrows correspond to secondary endosymbiont DNA. A system peak at 3.7 min is observed in these conditions in all DHPLC chromatograms. (B) DHPLC profiles after PCR reamplification of fractions of the eluted volumes from chromatogram A. Elution times at fraction collector of 2.5–2.75 min (9), 2.75–3.0 min (8), 3.0–3.25 min (7), 3.25–3.5 min (6), 4.25–4.5 min (5), 4.5–4.75 min (4), 4.75–5.0 min (3), 5.0–5.25 min (2) and 5.25–5.5 min (1). Elution times at absorbance detection cell were at least 0.5 min smaller. Column, DNASep® Cartridge matrix of PS-DVB particles, 2.2 μm, 4.6 mm × 50 mm length; mobile phase, 0.1 M TEAA, pH 7.65, 25% acetonitrile; flow rate, 0.9 ml/min; temperature, 56 °C; detection, UV, 260 nm.

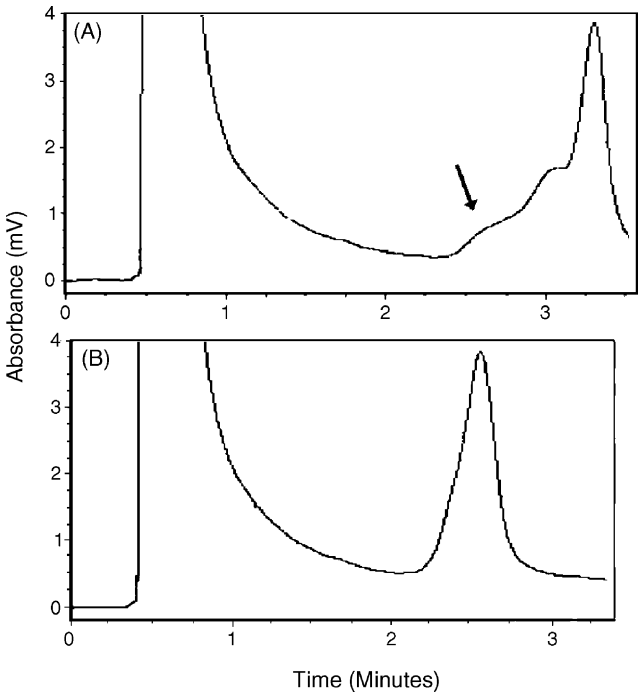


Fig. 3. Chromatograms of the PCR reamplification product from fraction 1 (see Fig. 2B) detected by DHPLC at 63 °C (A) and 64.5 °C (B). Arrow marks the peak corresponding to the partially denatured DNA molecule from a minority secondary endosymbiont from *T. ulmi* (A) which become completely denatured at 64.5 °C (B). The highest peak (A and B) corresponds to the DNA molecule from the majority secondary endosymbiont. Column, DNASep® Cartridge matrix of PS-DVB particles, 2.2 μm, 4.6 mm × 50 mm length; mobile phase, 0.1 M TEAA, pH 7.65, 25% acetonitrile; linear gradient, 48–53% B in 0.1 min, 53–61% B in 4 min; flow rate, 0.9 ml/min; temperature, 56 °C; detection, UV, 260 nm.

the already published sequences of the *ribA* gene from the *Bu. aphidicola* strains BAp and BBp, endosymbionts of *Acyrthosiphon pisum* [18] and *Baizongia pistaciae* [9], respectively, and with the sequences of several free-living gamma-Proteobacterial species. In spite of the shortness of the sequence, the phylogenetic reconstruction permitted the identification of the *Bu. aphidicola* sequences because they formed a monophyletic group with high support (Fig. 4). The rest of the sequences were identified as secondary symbionts belonging to different gamma-Proteobacterial species. They probably belong to the U, T, R and V types described earlier [12,14]. A summary of the taxonomic classification of the sequences is shown (Table 1). The absence of a *Bu. aphidicola* sequence in four aphid species belonging to the subfamilies Lachninae, Myzocallidinae and Chaitophorinae would be associated to the loss of the *ribA* gene in the *Bu. aphidicola* strain from these lineages or to the inability of the degenerate primers to amplify this sequence.

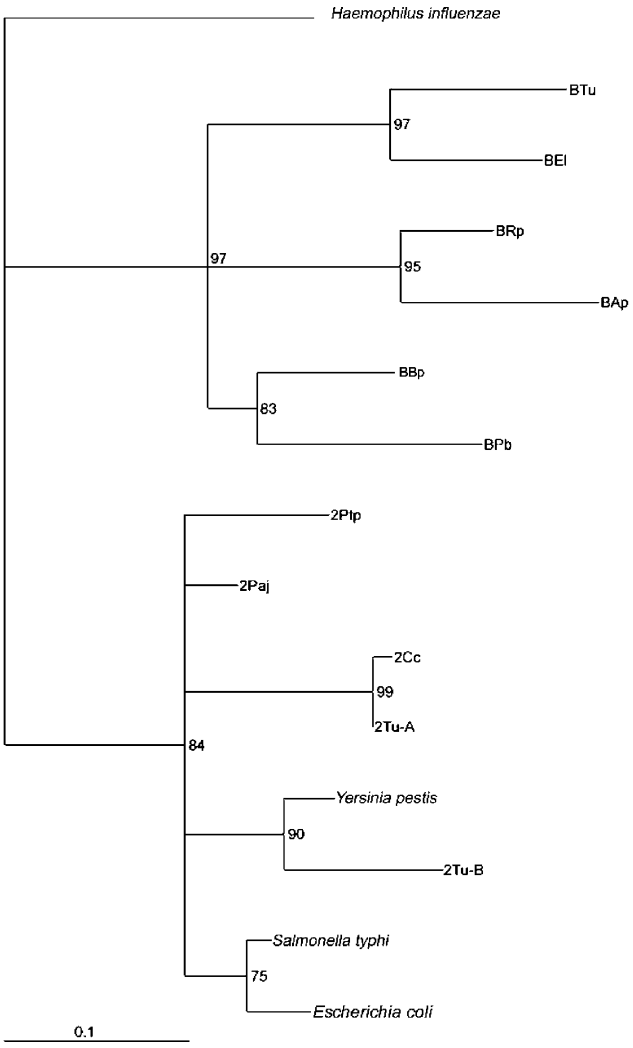


Fig. 4. Phylogenetic tree obtained by maximum likelihood from a nucleotide alignment of a fragment of the *ribA* gene. Gene names from free-living bacterial species: *Haemophilus influenzae* (HI1224) *Yersinia pestis* (YPO2227) *Salmonella typhi* (STY1344) and *E. coli* (B1281) Gene names and insect host names from completely sequenced *Bu. aphidicola* genomes: BAp (BU271, *A. pisum*) and BBp (BBp252, *B. pistaciae*). Abbreviations and host names for sequences obtained in this work: (a) *Bu. aphidicola* strains: BTu (*T. ulmi*), BEl (*E. lanuginosum*), BRp (*R. padi*), and BPb (*P. bursarius*); (b) secondary endosymbionts: 2Ptp (*Pt. persicae*), 2Paj (*Pa. juglandis*), 2Cc (*C. cedri*), 2Tu-A (*T. ulmi*, minority endosymbiont), and 2Tu-B (*T. ulmi*, majority secondary endosymbiont).

Because of the primer associated degeneration of the ends of the PCR products, the GC contents of each fragment were determined for the DNA sequence composed between the two primers. These values are expected to be only slightly smaller than those obtained if the average of all possible sequences for a fixed PCR product is estimated (data not shown). Our set of bacterial sequences showed values ranging between 24.5 and 49.0% GC. This range may be subdivided for the *Bu. aphidicola* strains (24.5–31.8), and for secondary endosymbionts (42.1–49.0). When the GC content percentage of each molecule was plotted against its retention time at 56 °C, a high positive correlation was observed (Fig. 5). The high correlation coefficient (0.978) showed that the main factor contributing to the retention time was the GC content of the DNA product. Other factors such as the distribution of the bases along the DNA fragment or the length of the DNA product seem to have a minor effect on this molecule. In fact the *ribA* fragments from *Bu. aphidicola* from *T. ulmi* and *Eriosoma lanuginosum* are slightly larger than the others (294 and 285 bp versus 273–276 bp) with practically no effect on the correlation coefficient.

Although we cannot discard the possibility that other molecules were affected by the base distribution in a larger extension than our *ribA* PCR product, we propose than the use of DHPLC as a second dimension for the separation of DNA molecules may be applied to the analyses of other DNA fragments. The choice of an appropriate temperature that

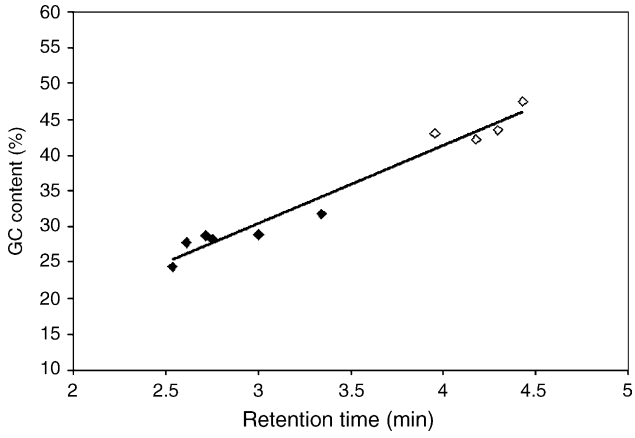


Fig. 5. Relationship between retention time obtained in DHPLC analyses at 56 °C and GC content of the *ribA* PCR product sequences. Open and close diamonds correspond to secondary endosymbionts and *Bu. aphidicola* strains, respectively. Column, DNASep® Cartridge matrix of PS-DVB particles, 2.2 μm, 4.6 mm × 50 mm length; mobile phase, 0.1M TEAA, pH 7.65, 25% acetonitrile; linear gradient, 48–53% B in 0.1 min, 53–61% B in 4 min; flow rate, 0.9 ml/min; temperature, 56 °C; detection, UV, 260 nm. Correlation coefficient, $R = 0.978$ ($n = 10$).

permits their migration in a partial denatured stage will also be an important factor.

## 4. Conclusions

Our results indicate that DHPLC may be used to determine if we have amplified one or more DNA molecules after a PCR reaction. In addition, we can determine the approximate GC content of the molecule and, in this way, determine if this DNA molecule may belong to the searched for species. However, it is crucial to select an appropriate temperature to produce the partial denaturation of the molecules. In our study we have separated molecules from 25 to 49% GC. Molecules smaller than 25% would probably become completely denatured at 56 °C, while others with higher GC contents would migrate as non-denatured dsDNA according to their sizes. We have also shown that the use of the DHPLC permits the isolation, or at least enrichment, of a specific DNA molecule from a mix of PCR products. The collected fractions may be used later for PCR reamplification or directly for sequencing or cloning.

## Acknowledgments

This work was supported by grant BMC2003-00305 from the *Ministerio de Ciencia y Tecnología* (Spain) and grant Grupos03/204 from *Generalitat Valenciana* (Spain).

## References

[1] W.Z. Xiao, P.J. Oefner, Hum. Mutat. 17 (2001) 439.
[2] P.J. Oefner, C.G. Huber, J. Chromatogr. B, Anal. Technol. Biomed. Life Sci. 782 (2002) 27.
[3] P.J. Oefner, P.A. Underhill, Current Protocols in Human Genetics, Wiley & Sons, New York, 1998.
[4] N.A. Moran, M.A. Munson, P. Baumann, H. Ishikawa, Proc. R. Soc. Lond. B: Biol. Sci. 253 (1993) 167.
[5] P. Baumann, C.-Y. Lai, L. Baumann, D. Rouhbakhsh, N. Moran, M.A. Clark, Appl. Environ. Microbiol. 61 (1995) 1.
[6] N.A. Moran, Cell 108 (2002) 583.
[7] F.J. Silva, A. Latorre, A. Moya, Trends Genet. 17 (2001) 615.
[8] I. Tamas, L. Klasson, B. Canback, A.K. Naslund, A.S. Eriksson, J.J. Wernegreen, J.P. Sandstrom, N.A. Moran, S.G.E. Andersson, Science 296 (2002) 2376.
[9] R.C. van Ham, J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J.M. Fernandez, L. Jimenez, M. Postigo, F.J. Silva, J. Tamames, E. Viguera, A. Latorre, A. Valencia, F. Moran, A. Moya, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 581.
[10] F.J. Silva, A. Latorre, A. Moya, Trends Genet. 19 (2003) 176.
[11] D.Q. Chen, A.H. Purcell, Curr. Microbiol. 34 (1997) 220.
[12] J.P. Sandstrom, J.A. Russell, J.P. White, N.A. Moran, Mol. Ecol. 10 (2001) 217.
[13] A. Moya, A. Latorre, B. Sabater-Muñoz, F.J. Silva, J. Mol. Evol. 55 (2002) 127.
[14] J.A. Russell, A. Latorre, B. Sabater-Muñoz, A. Moya, N.A. Moran, Mol. Ecol. 12 (2003) 1061.
[15] A. Latorre, A. Moya, F.J. Ayala, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 8649.
[16] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (1997) 3389.
[17] K. Strimmer, A. von Haesseler, Mol. Biol. Evol. 13 (1996) 964.
[18] S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, H. Ishikawa, Nature 407 (2000) 81.