ELSEVIER

FEMS Microbiology Letters 199 (2001) 229–233

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Description and application of a rapid method for genomic DNA direct sequencing

Evelyne Krin, Florence Hommais *, Olga Soutourina, Saravuth Ngo, Antoine Danchin, Philippe Bertin

Unité de Génétique des Génomes Bactériens, Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France

Received 29 January 2001; received in revised form 3 April 2001; accepted 11 April 2001

First published online 7 May 2001

Abstract

We describe a rapid method for determining nucleotide sequences directly from total genomic DNA. This technique was used to determine genomic DNA sequences in various prokaryotic and eukaryotic microorganisms with a G+C content between 40 and 50%, e.g. *Escherichia coli*, *Vibrio cholerae*, *Bacillus subtilis* and *Saccharomyces cerevisiae*. Furthermore, the method was applied to accurately sequence up to 300 DNA base pairs in *Photorhabdus luminescens*, whose genome sequencing is currently under way. Taken together, these results provide evidence that our technique can be widely used to easily and efficiently determine genomic DNA sequences.

Keywords: *hns*; Transposon insertion; DNA extraction; Unfinished genome

1. Introduction

Genetic techniques, in particular those based on transposon insertion, are commonly used to identify genes specifically involved in various processes. However, the characterization of transposition sites by sequencing usually requires several time consuming steps such as the complete digestion of genomic DNA, cloning into a vector or inverse PCR [1]. Furthermore, they are often somewhat difficult to achieve due to the length of DNA fragments. In the past years, many attempts were made to determine a particular nucleotide sequence directly in its genomic context. For example, the Tn*5*seq1 transposon was constructed to sequence *Escherichia coli* chromosomal DNA after digestion with the *Pst*I restriction enzyme [2]. However, this method requires alkaline denaturation of DNA and reverse transcription and its use is restricted to the localization of a site of transposon insertion. More recently, a protocol was developed to directly sequence up to 700 bp from chromosomal DNA [3]. Nevertheless, this method requires automated sequencing, the Big-Dye kit (Perkin Elmer) and the purification of sequence reactions before loading on gel.

In the present work, we developed an efficient and easy procedure to sequence genomic DNA from different microorganisms. We elaborated a more performant method for genomic DNA preparation and we used the Thermo Sequenase radiolabelled terminator cycle sequencing kit (Amersham Pharmacia Biotech) with a new PCR cycle.

2. Materials and methods

2.1. Strains and primers

The *E. coli* BE1410 strain [4] contains a Tn*5*seq1 transposon in the *hns* coding sequence (GenBank accession number X07688). Nucleotide sequences were determined using either T7 and SP6 primers previously designed from the Tn*5*seq1 coding sequence [2], or HNS primer designed from the *hns* promoter sequence (Table 1). Using genomic DNA from the *Vibrio cholerae* classical Ogawa O395 strain, sequence reactions were performed using VICH primer [5], an oligonucleotide complementary to the *vicH* 5′-end coding sequence [6]. Oligonucleotide GB155 used to sequence genomic DNA from *Saccharomy-*

* Corresponding author. Tel.: +33 (1) 45 68 84 43; Fax: +33 (1) 45 68 89 48; E-mail: fhommais@pasteur.fr

0378-1097/01/$20.00
PII: S0378-1097(01)00186-0

ces cerevisiae FY1679 (Table 1) hybridizes with an intergenic region between *YLR072w* and *YLR073c* ORFs in chromosome XII [7]. Oligonucleotides PHOC and PHON were designed to hybridize respectively to the *flhDC* coding sequence and to its 5′-end. All experiments were performed in accordance with European regulation requirements concerning the contained use of Genetically Modified Organisms of Group I (agreement No. 2735) and Group II (agreement No. 2736).

2.2. Genomic DNA preparation

Bacterial cells were grown in 50 ml Luria broth to stationary phase. Cells were centrifuged at $5000 \times g$ for 10 min at 4°C, resuspended in 9 ml of 50 mM Tris–HCl, 50 mM EDTA pH 8 (in the case of *Bacillus subtilis*, cells were incubated for 10 min with 1 mg ml^{-1} lysozyme at 37°C and the enzyme was then inactivated for 3 min at 65°C). Cells were lysed by the addition of 2.5 mg proteinase K and 1 ml of 10% SDS and gently shaken for 16 h at 37°C. DNA was extracted with an equal volume of phenol saturated with Tris–HCl pH 8 and 2 ml 50 mM Tris–HCl, 50 mM EDTA pH 8 were added. The suspension was gently mixed and incubated at 37°C for 30 min. After centrifugation for 10 min at $12\,000 \times g$ at 15°C, the aqueous phase was extracted twice with an equal volume of phenol: chloroform and then with one volume of chloroform. At each step, the suspension was gently mixed and centrifuged as previously. Genomic DNA was precipitated by addition of one volume of propanol-2 and 1:10 volume of 3 M sodium acetate pH 5.2. Chromosomal DNA was washed four times in absolute ethanol by the use of a Pasteur pipet, air dried for 30 min and resuspended in 0.5 to 2 ml of 10 mM Tris–HCl, 1 mM EDTA pH 7.6. After incubation for 15 min at 65°C, genomic DNA preparation was spectrophotometrically quantified (OD_{260}) and stored at 4°C. Genomic DNA from *S. cerevisiae* was prepared using a zymolase treatment as previously described [8].

2.3. Genomic DNA sequencing

Sequencing reactions were performed using the Thermo Sequenase radiolabelled terminator cycle sequencing kit (Amersham Pharmacia Biotech). In a final volume of 60 μl, 5 μg of genomic DNA (40 μg for *S. cerevisiae*) was mixed with 60 pmol of freshly diluted primer (100 pmol for *S. cerevisiae*) and 6 μl reaction buffer. 3 μl of Thermo-Fidelase enzyme (Fidelity Systems, Inc.) was added to sequence chromosomal DNA of *S. cerevisiae*. The mixture was incubated for 5 min at 95°C in a dried heating block and then allowed to cool down to 30°C in the block before the addition of 3 μl of Thermo Sequenase DNA polymerase. 14 μl of this mixture was added to each termination mix (6 μl nucleotide master mix and 1.5 μl each [α-^{33}P]-ddNTP (Amersham Pharmacia Biotech) respectively). The PCR sequencing program was as follows: 80 cycles of 1 min at 95°C, 30 s at 55°C and 4 min at 72°C. After amplification, reactions were stored at 4°C until the addition of 10 μl of stop solution. After denaturation for 5 min at 95°C in a thermocycler, 7 μl of DNA sequence reactions were loaded on a standard 8% acrylamide:bisacrylamide (29:1 w/w) gel in glycerol tolerant buffer containing 60 μl TEMED and 0.1% ammonium persulfate in 100 ml. Samples were migrated at 65 W for 100 min. The gel was dried and exposed to a phosphor screen (Molecular Dynamics) for at least 5 days or to a Hyperfilm-βmax autoradiography film (Amersham Pharmacia Biotech) for 8 days.

3. Results and discussion

We first used the sequence determination procedure from genomic DNA (see Section 2) to localize precisely the site of transposon insertion in *E. coli* BE1410 strain. This strain contains a Tn*5*seq1 transposon inserted in the *hns* gene located at minute 27 on the chromosome and coding for a pleiotropic regulator [4]. Chromosomal DNA was extracted from the BE1410 strain and sequencing reactions were performed with SP6 and T7 Tn*5*seq1 primers specifically designed for direct sequencing on chromosomal DNA (Table 1). Fig. 1A,B shows the results obtained with SP6 primer and T7 primer, respectively. 89 bp were read from the SP6 primer while 119 bp were read from the T7 primer. Each sequence was compared to the whole genome of *E. coli* (http://bioweb.pasteur.fr/GenoList/Colibri/), using the BLAST program [9]. The sequence obtained from the T7 primer allowed us to localize the insertion site at the 20th codon of the *hns* gene (accuracy > 97%, BLAST P value = 1.9×10^{-19} and BLAST score value = 535). Despite a lower score (accuracy = 90%, BLAST P value = 1.1×10^{-3} and BLAST score value = 242), the Tn*5*seq1 insertion site was surprisingly localized with the SP6 primer at the 17th codon of the *hns* coding sequence without any ambiguity. The higher percentage of error obtained with the SP6 primer can be explained by hybridization to another DNA fragment. Indeed, unlike the T7 primer, we observed that the last 14 bp of SP6 primer, designed before the whole genome sequence was available [2], hybridized with the *yhiS* coding sequence. This provides evidence that primers used for sequencing on genomic DNA must be optimized by comparing their sequence with the whole genome. On the other hand, the discrepancy between the location of the transposon obtained from the T7 primer (20th codon) and that from the SP6 primer (17th) can be explained by the mechanism of Tn*5*seq1 insertion. Indeed, at its insertion site, the transposon is bracketed by a 9-bp duplication of the target sequence (5′-GGCAAGAGA-3′), which is generated during the transposition event [10].

We wanted to know whether our method could be used for sequencing *E. coli* chromosomal DNA from any oli-

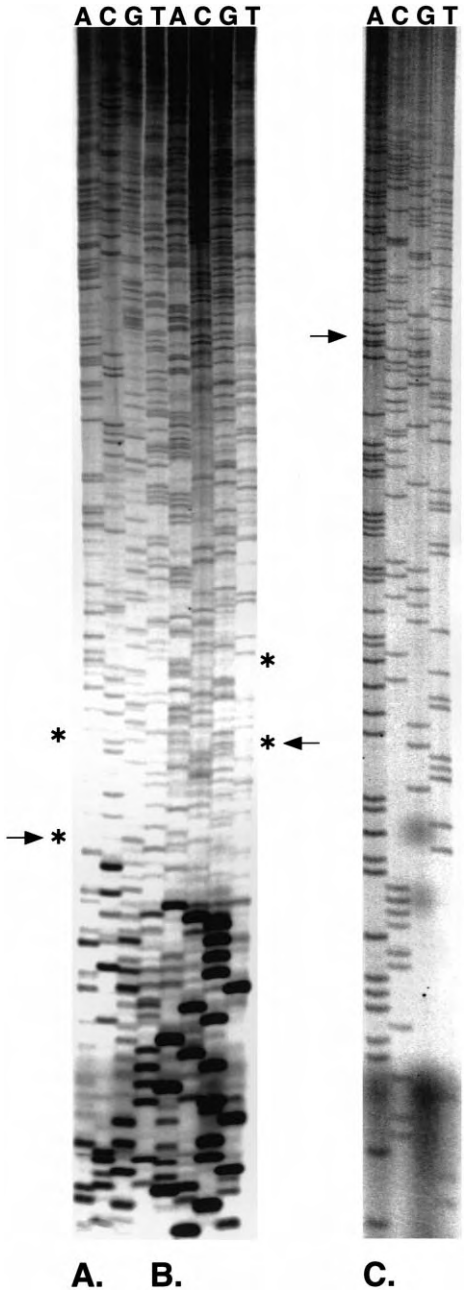

Fig. 1. Nucleotide sequence of *hns* from total *E. coli* chromosomal DNA. Sequencing reactions were performed on the BE1410 strain using primers SP6 (A), T7 (B) and HNS (C). Arrows correspond to the end of the Tn*5*seq1 sequence. Sequences between asterisks correspond to the 9-bp duplication generated during the Tn*5*seq1 transposition event.

gonucleotide. Sequencing reactions were performed using the HNS primer hybridizing solely with the *hns* promoter region in BE1410 strain. The result is shown in Fig. 1C. 152 bp were read and compared with the *E. coli* genome by BLAST program [9]. The sequence corresponded to the *hns* region (accuracy > 97%, BLAST P value = 3.5×10^{-17}, BLAST score value = 485) and confirmed the localization of Tn*5*seq1 in the 20th codon of the *hns* coding sequence.

To know whether our method could be more widely used, we performed sequencing experiments from genomic

Table 1
Strains and primers used

Strains and primers	Description	Source
Strains		
BE1410	*E. coli* FB8 *hns-1001*::Tn*5*seq1	[4]
Ogawa O395	classical *V. cholerae* (wild-type)	C. Parsot
FY1679	*S. cerevisiae*	[7]
TT01	*P. luminescens*	N. Boemare
Primers		
T7	5′-TAATACGACTCACTATAGGG-3′	[2]
SP6	5′-CATACGATTTAGGTGACACTATAG-3′	[2]
HNS	5′-TTGCACAACTGAATTTAAGGCTCTA-3′	This study
VICH	5′-ATGTCGGAAATCACTAAGAC-3′	[5]
GB155	5′-TAACAAACTATCTAGACACC-3′	G. Blandin
PHOC	5′-CATAGAATCGCTAATACCTAAACGG-3′	This study
PHON	5′-TCATTTCTCTATCCCGTCCG-3′	This study

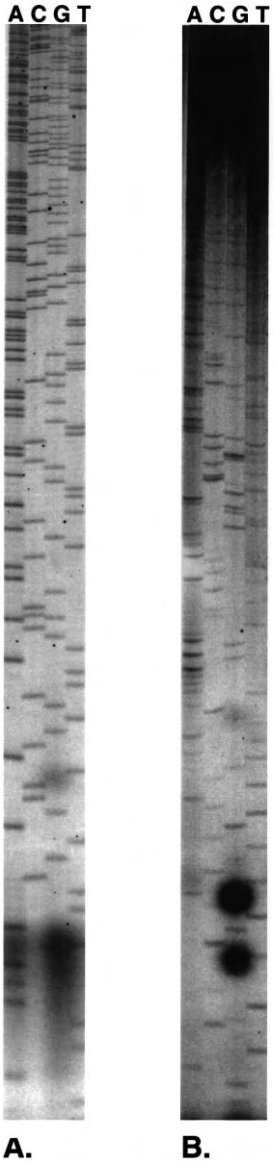

Fig. 2. Nucleotide sequences from total chromosomal DNA of different microorganisms. Sequencing reaction were of *vicH* performed on *V. cholerae* O395 strain using the VICH primer (A) and those of *YLR073c* on *S. cerevisiae* FY1679 strain using GB155 primer (B).

DNA extracted from various bacteria, such as *Bordetella pertussis*, *V. cholerae* and *B. subtilis* (a Gram-positive bacterium), and from the yeast *S. cerevisiae*. From *B. pertussis* genomic DNA, the method did not give any convincing result (data not shown) probably due to the high G+C% content of its genomic DNA. Indeed, the G+C content of the *B. pertussis* DNA ranges from 66 to 70%. This high G+C content could result in very strong secondary structures known to alter enzyme functioning, in particular polymerases used for PCR. In contrast, our method allowed us to sequence *V. cholerae* genomic DNA, using VICH primer hybridizing with the *vicH* 5′-end coding sequence (Fig. 2A). This gene has been recently shown to encode a H-NS-like protein [6]. 182 bp were read and compared with the NCBI BLAST server (http://www.ncbi.nlm.nih.gov/BLAST) to the *V. cholerae* genome

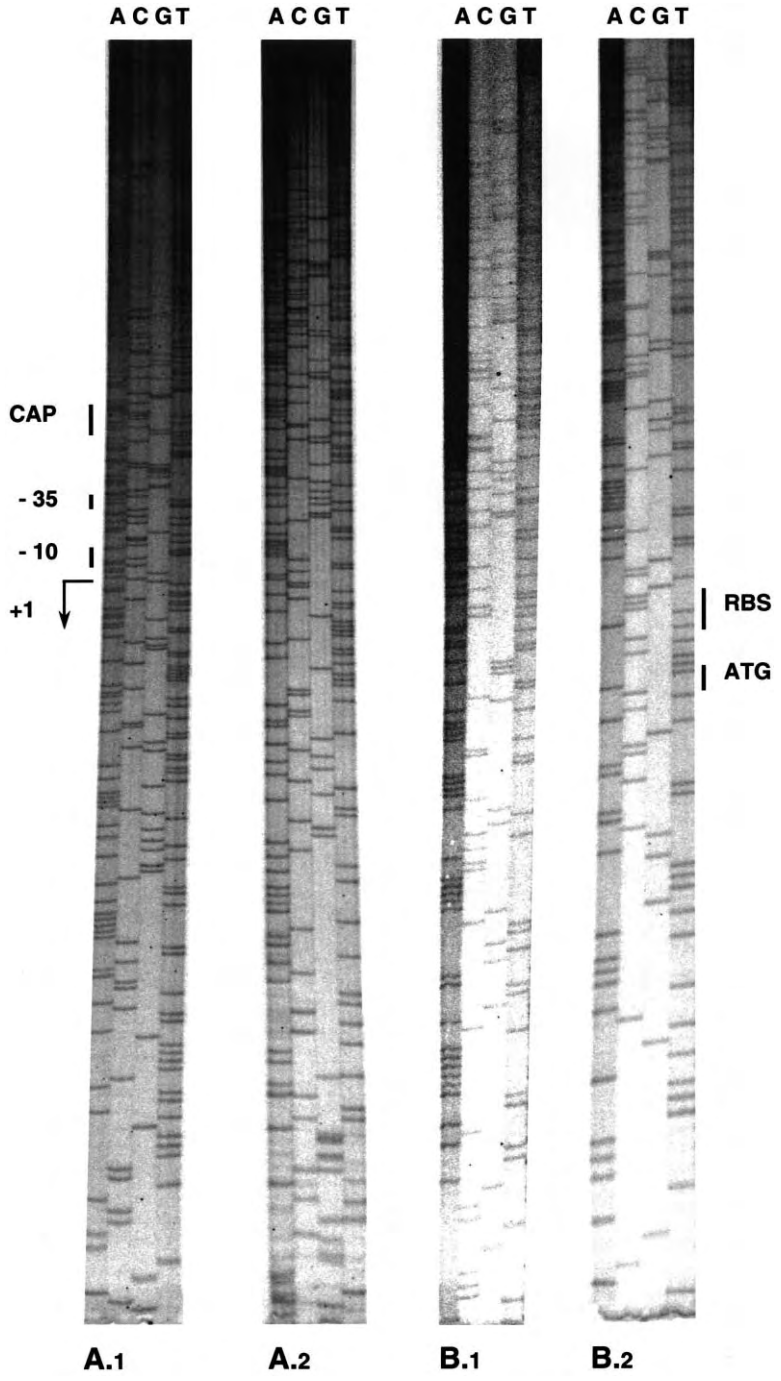

Fig. 3. Nucleotide sequence of the *flhDC*-like operon promoter region from total *P. luminescens* chromosomal DNA. Sequencing reactions were performed on TT01 strain using primers PHON (A) and PHOC (B), with three (1) or two (2) migrations. Positions of the putative RBS, the −10 and −35 hexamers and the CAP binding site consensus sequence are underlined, the putative transcription start point (+1) is indicated by an arrow and ATG corresponds to the translational initiation codon.

sequence. An accuracy of 99%, a BLAST *P* value of 3.5×10^{-35} and a score value of 882 were found with the sequence corresponding to the *hns*-like gene in this organism. Moreover, the technique presented here was successfully used to analyze *cysH*, a gene encoding a putative phospho-adenylsulfate sulfotransferase in *B. subtilis*, after transposon mutagenesis (S. Auger, personal communication). Finally, the sequencing procedure was applied on *S. cerevisiae* genomic DNA using GB155 primer which hybridizes with an intergenic region between *YLR072w* and *YLR073c* ORFs in chromosome XII [7]. The use of ThermoFidelase® (Fidelity Systems, Inc.), an enzyme which binds to DNA and helps polymerase to pass through strong secondary structures, was required to obtain a sequence on yeast genomic DNA. 124 bp were read (Fig. 2B) and were compared with the whole genome of *S. cerevisiae* using the BLAST program [9]. An accuracy of 93%, a BLAST *P* value of 1×10^{-29} and a score value of 133 was found with ORF *YLR073c*.

Photorhabdus luminescens is a Gram-negative bacterium whose genome sequencing is currently under way at the Pasteur Institute (http://www.pasteur.fr/recherche/unites/gmp). During this work, a partial sequence coding for a gene homologous to *flhDC*, the flagellar master operon in *E. coli*, was identified (E. Duchaud, personal communication). Recently, we demonstrated that the *flhDC* promoter in *E. coli* contains a CAP binding site and a large untranslated region, which seems to play a crucial role in the regulation of its expression by H-NS [11]. Therefore, it was of interest to determine the existence of such a regulatory region in *P. luminescens*. Two and three migrations were performed using oligonucleotides PHOC and PHON (Table 1 and Fig. 3). More than 300 bp in the promoter region of the *flhDC*-like genes were read with each oligonucleotide (Fig. 3). Analysis of the nucleotide sequence allowed us to identify −35 and −10 hexamers showing 50 and 67% similarity with the *E. coli* σ^{70} consensus, respectively. These two boxes are separated by a 19-bp spacer and are located 289 and 264 bp upstream from the ATG translational start codon, respectively. In addition, we identified a consensus CAP binding site (tttTGTGAgttatgTCACAtag) centered at position −71.5 with respect to a putative +1 transcriptional start site, as in *E. coli* [11]. This suggests that the regulatory mechanism of the flagellar master operon in *P. luminescens* could be similar to that in *E. coli*.

In summary, we developed a powerful sequencing procedure based on classical techniques of genomic DNA preparation and radioactive sequencing kit. This technique can be easily used to directly sequence total genomic DNA with oligonucleotides designed from the left and the right ends of a transposon or from any genomic DNA fragment, in particular when the whole genomic sequence is not yet available. This simple method could be largely applied to DNA with a G+C content around 40–50%.

Acknowledgements

We wish to thank Sandrine Auger for helpful discussions and technical assistance in *B. subtilis* DNA sequencing, E. Duchaud and F. Kunst for providing us with an unpublished sequence. We also thank C. Parsot, G. Bandin and N. Boemare for the gift of *V. cholerae* classical Ogawa O395 strain, *S. cerevisiae* FY1679 strain and *P. luminescens* TT01 strain, respectively. We are grateful to I. Martin and J.Y. Coppee for critical reading of the manuscript. Financial support came from the Pasteur Institute and the Centre National de la Recherche Scientifique (URA 2171).

References

[1] Quivy, J.P. and Becker, P.B. (1996) Determination of unknown genomic sequences without cloning. In: PCR Sequencing Protocols (Rapley, R., Ed.), pp. 119–131. Humana Press, Totowa, NJ.
[2] Nag, D.K., Huang, H.V. and Berg, D.E. (1988) Bidirectional chain-termination nucleotide sequencing: transposon Tn*5*seq1 as a mobile source of primer sites. Gene 64, 135–145.
[3] Heiner, C.R., Hunkapiller, K.L., Chen, S-M., Glass, J.I. and Chen, E.Y. (1998) Sequencing multimegabase-template DNA with BigDye terminator chemistry. Genome Res. 8, 557–561.
[4] Laurent-Winter, C., Ngo, S., Danchin, A. and Bertin, P. (1997) Role of *Escherichia coli* histone-like nucleoid-structuring protein in bacterial metabolism and stress response: Identification of targets by two-dimensional electrophoresis. Eur. J. Biochem. 244, 767–773.
[5] Tendeng, C., Badaut, C., Krin, E., Gounon, P., Ngo, S., Danchin, A., Rimsky, S. and Bertin, P. (2000) Isolation and characterization of *vicH*, encoding a new pleiotropic regulator in *Vibrio cholerae*. J. Bacteriol. 182, 2026–2032.
[6] Bertin, P., Benhabiles, N., Krin, E., Laurent-Winter, C., Tendeng, C., Turlin, E., Thomas, A., Danchin, A. and Brasseur, R. (1999) The structural and functional organization of H-NS-like proteins is evolutionarily conserved in Gram-negative bacteria. Mol. Microbiol. 31, 319–329.
[7] Goffeau, A. et al. (1997) The yeast genome directory. Nature 387, 1–105.
[8] Fairhead, C., Llorente, B., Denis, F., Soler, M. and Dujon, B. (1996) New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker' recombination. Yeast 12, 1439–1457.
[9] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
[10] Berg, D.E. and Berg, C.M. (1983) The prokaryotic transposable element Tn*5*. Bio/Technology 1, 417–435.
[11] Soutourina, O., Kolb, A., Krin, E., Laurent-Winter, C., Rimsky, S., Danchin, A. and Bertin, P. (1999) Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon. J. Bacteriol. 181, 7500–7508.