Characterization of bacterial operons consisting of two tubulins and a kinesin-like gene by the novel Two-Step Gene Walking method

Martin Pilhofer[1], Andreas Peter Bauer[1], Martina Schrallhammer[2], Lothar Richter[3], Wolfgang Ludwig[1], Karl-Heinz Schleifer[1] and Giulio Petroni[2,*]

[1]Lehrstuhl für Mikrobiologie, Technical University Munich, Am Hochanger 4, D-85354 Freising, Germany, [2]Dipartimento di Biologia, University of Pisa, via A. Volta 4/6, I-56126 Pisa, Italy and [3]Machine Learning and Data Mining in Bioinformatics, Technical University Munich, Boltzmannstraße 3, D-85748 Garching b. München, Germany

ABSTRACT

Tubulins are still considered as typical proteins of *Eukaryotes*. However, more recently they have been found in the unusual bacteria *Prosthecobacter* (*btubAB*). In this study, the genomic organization of the *btub*-genes and their genomic environment were characterized by using the newly developed Two-Step Gene Walking method. In all investigated Prosthecobacters, *btubAB* are organized in a typical bacterial operon. Strikingly, all *btub*-operons comprise a third gene with similarities to kinesin light chain sequences. The genomic environments of the characterized *btub*-operons are always different. This supports the hypothesis that this group of genes represents an independent functional unit, which was acquired by *Prosthecobacter* via horizontal gene transfer. The newly developed Two-Step Gene Walking method is based on randomly primed polymerase chain reaction (PCR). It presents a simple workflow, which comprises only two major steps—a Walking-PCR with a single specific outward pointing primer (step 1) and the direct sequencing of its product using a nested specific primer (step 2). Two-Step Gene Walking proved to be highly efficient and was successfully used to characterize over 20 kb of sequence not only in pure culture but even in complex non-pure culture samples.

INTRODUCTION

Sequencing of DNA using a specific primer requires a template, which must consist of a maximum of target-DNA and a minimum of non-target DNA. Any kind of non-target DNA can lead to unspecific binding of the sequencing primer. The standard methods to shift the ratio towards target-DNA are either to amplify the target-DNA using the PCR technique (1–3) or to clone the target-DNA into a vector (phage or plasmid), which is highly amplified in a bacterial host and subsequently specifically isolated. If the target-DNA is represented by a specific gene, the typical approach to accomplish its characterization is to design two degenerated primers, targeting highly conserved regions of the gene. The intervening region is then amplified, cloned and sequenced. This approach, which is nowadays a routine procedure, rapidly provides sequence data on yet uncharacterized genes belonging to known gene families. The main limit of this approach is that both extremities of the investigated gene remain unknown. Despite several methods have been proposed, the characterization of a sequence adjacent to a known region (gene walking or chromosome walking) still remains a laborious and time-consuming task.

In the past, there were reports that suggested the possibility to perform direct sequencing of genomic DNA to characterize unknown regions adjacent to known ones without any kind of preliminary amplification (4,5). Despite the theoretical advantages of this approach, they never became standard and have been applied only in a very restricted number of cases. The reported main problems concern: (i) the high amount of DNA used (several micrograms) (4,5); (ii) the unspecific binding of the sequencing primer due to the bad ratio of target-DNA over non-target DNA resulting in very short read lengths or in complete failure of the sequencing run (4,5); (iii) the target organisms' genomes, which must possess a $G + C$ ratio in a certain range (5); and (iv) the specific set-up of the sequencing instruments and software, which hampers the use of custom sequence services (4). For these reasons,

*To whom correspondence should be addressed. Tel: +39 050 2211384; Fax: +39 050 2211393; Email: gpetroni@biologia.unipi.it

PCR-based gene walking methods are generally applied and largely favored over direct sequencing of genomic DNA for the characterization of unknown sequences adjacent to known ones.

PCR-based gene walking methods can be divided into three groups (based on their underlying methods): (i) inverse PCR (6–9), (ii) ligation-mediated PCR (10–19) and (iii) randomly primed PCR (20–34). Using inverse PCR and ligation-mediated PCR, as a first step DNA has to be digested with restriction enzymes. Inverse PCR amplifies unknown DNA adjacent to known DNA fragments after its intramolecular circularization, which is performed by a ligase at low DNA concentrations. The region of interest is then amplified using outward pointing primers located at both ends of the known fragment. In ligation-mediated PCR, synthetic adapter-DNAs, which are ligated to restriction fragments, are used as primer annealing sites for PCR. The region of interest is then amplified by a known sequence-specific outward-pointing primer and a primer reverse complementary to the ligated adapter. A disadvantage of ligation-mediated PCR methods is the requirement of relatively high amounts of target DNA. Furthermore, ligation-mediated PCR and inverse PCR are dependent upon time-consuming and labor-intensive enzymatic modifications of the target DNA (e.g. restriction digest and ligation reactions).

Randomly primed PCR methods use the ability of oligonucleotides (primers) to bind unspecifically in the unknown fragment at low stringency conditions (20). There are two possibilities to facilitate priming in the unknown region. Some methods use different types of walking primers to bind in the unknown region (20,22–24,26,27,29,31,33,34). Other experimental approaches utilize only one primer in PCR, which binds specifically in the known sequence stretch and in very low stringency cycles also in the unknown region (21,25,28,30,32). The unspecific binding of these primers causes the facilitation of unspecific PCR products. Therefore, the existing randomly primed PCR methods include cloning and/or enrichment procedures (e.g. immobilization of specific products on paramagnetic beads). The derived PCR products or clones must be intensively screened for specificity, e.g. using modified oligonucleotides in hybridization experiments.

In this study, a straightforward Two-Step Gene Walking method was developed. It consists of a Walking-PCR (step 1) and direct sequencing of the PCR product (step 2). There is no need for enrichment, cloning or screening procedures; the amount of template DNA for Walking-PCR is minimal (50 ng). The new method was developed for the characterization of the genomic environment of bacterial tubulin genes in *Prosthecobacter* and was proven to work even in complex non-pure culture samples.

Tubulins are typical eukaryotic proteins. The first molecular evidences for real tubulin genes in bacteria were reported by Jenkins *et al.* (35) and later confirmed and extended by Pilhofer *et al.* (36) in different *Prosthecobacter* species. Bacterial A and B tubulin genes (*btubAB*) exist as adjacent loci on the genome.

Jenkins *et al.* (35) described the presence of a gene downstream of *btubB* in *Prosthecobacter dejongeii*, which has similarity to kinesin light chain. But to date, no sequence information was published. Therefore, recent intensive biochemical studies on the bacterial tubulin genes of *Prosthecobacter* could not include that third gene (37,38), which may be functionally related to the bacterial tubulin genes. The aim of this study was to characterize and comparatively analyze the genomic organization and the genomic environment of *btub* genes in different *Prosthecobacter* species using the newly developed Two-Step Gene Walking technique.

MATERIALS AND METHODS

Cultures

Cultures of *Prosthecobacter vanneervenii* DSM12252 and *Prosthecobacter debontii* DSM14044 were obtained from DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany) and grown aerobically at 28°C in DSM medium 628. *Euplotidium itoi* with epixenosomes as ectosymbionts was grown according to Rosati *et al.* (39). *Paramecium caudatum* infected with *Caedibacter caryophilus* was cultured in Sonneborn's *Paramecium* medium (ATCC medium 802) inoculated with *Klebsiella planticola*.

DNA Extraction

DNA from *Prosthecobacter* was isolated according to Wisotzkey *et al.* (40). DNA from epixenosomes was extracted using a modified protocol to inhibit high nuclease activity. The lysis of the cells was directly performed in high SDS concentration without previous lysozyme treatment. SDS (10%) was added in ratio 1:1 to the cell resuspension, followed by the addition of Proteinase K (0.625 mg/ml) and RNAse (0.25 mg/ml). The mixture was incubated for 1 h at 55°C. Following steps were as described by Wisotzkey *et al.* (40).

Prior to DNA extraction of *C. caryophilus*, the *Paramecium* culture was starved for ten days to minimize contaminations by food bacteria. Twenty liters of paramecia were harvested with a cream separator (Westfalia Separator AG) and further concentrated with an oil-testing centrifuge for 3 min at 300 g. The concentrated paramecia were mechanically homogenized and the crushed cells were washed in phosphate-buffered saline (PBS). The pellet was resuspended in 1:1 PBS and 90% Percoll (Amersham), and applied on the top of a Percoll gradient. The gradient was build up of 90%, 70% and 50% Percoll (in 1× PBS). Gradient centrifugation was carried out at 10 000 g at 4°C for 20 min. *Caedibacter caryophilus* and other bacteria were recovered from the fraction between 70 and 90% and washed again in PBS. DNA was extracted from the bacteria contained in this fraction according to Wisotzkey *et al.* (40).

Two-Step Gene Walking Procedure

The Two-Step Gene Walking method consists of a Walking-PCR (step 1) and direct sequencing of the
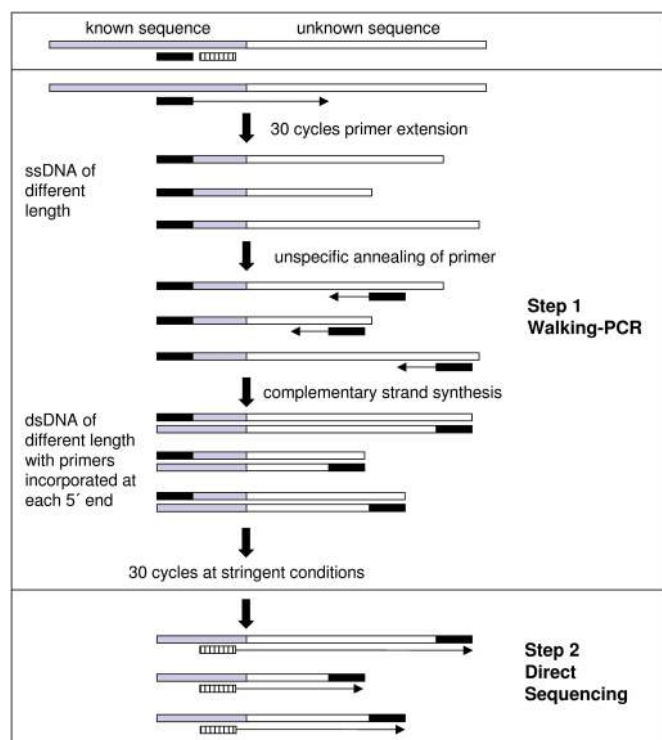
Figure 1. Basic principle of the Two-Step Gene Walking procedure. Known sequence stretch is shown in gray; unknown sequence stretch is shown in white; Walking-PCR primer is shown in black; specific nested sequencing primer is shown striped. In first 30 cycles, the PCR primer binds at stringent conditions; specific ssDNA of different length (caused by different drop off sites of polymerase) is produced. One subsequent cycle at low annealing temperature allows unspecific binding of primer at different sites on ssDNA as reverse primer. dsDNA of different length with primer sequence incorporated at each 5-prime end is produced. Thirty cycles at stringent conditions specifically and exponentially amplify dsDNA. PCR product is sequenced directly by using a specific nested primer.

Table 1. Walking-PCR cycling program

Round	Cycles	Temperature	Time	Action
		94°C	4 min	Primary denaturation
1	30	94°C 56°C 72°C	30 s 30 s 3 min	Specific primer extension; ssDNA synthesis
2	1	94°C 40°C 72°C	30 s 30 s 3 min	Unspecific binding of primer; dsDNA synthesis
3	30	94°C 56°C 72°C	30 s 30 s 3 min	Specific exponential amplification
		72°C	10 min	Final elongation

CHROMAS (Version 2.01, www.technelysium.com. au/chromas.html). Contigs were assembled manually and open reading frames were searched using ORF-Finder (41). Similarities to known protein sequences were searched using protein–protein BLAST (42).

Sequences with similarities to kinesin light chain were analyzed using the program PRINTS in order to detect motifs of the kinesin light chain fingerprint (43). Multiple EM for Motif Elicitation (MEME, http://meme.sdsc. edu/meme/website/meme.html) (44) was used to discover motifs in the bacterial kinesin light chain sequences. The Neuronal Network Promoter Prediction software (45) was used to identify putative promoter sequences. The prediction of Rho-independent terminators was performed with the program FindTerm (www.softberry.com/berry.phtml?topic = findterm&group = programs& subgroup = gfindb).
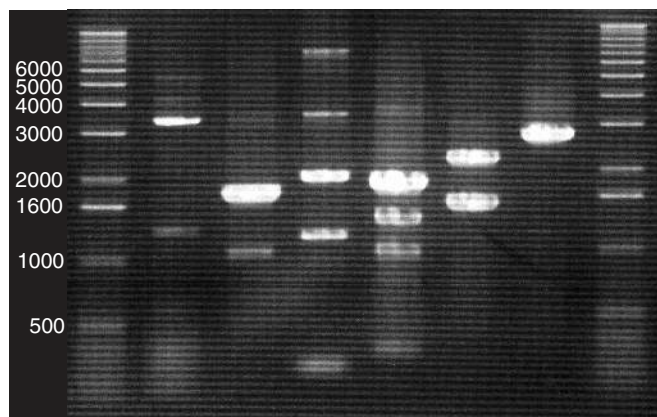
RESULTS AND DISCUSSION

Technical facts and theory of Two-Step Gene Walking

Initially, the authors attempted gene walking in *Prosthecobacter* genomes based on inverse PCR (6), Arbitrary PCR (22), SiteFinding-PCR (34) (randomly primed PCR methods) and ligation-mediated gene walking (13) (data not shown). To overcome the poor results and the low efficiency of these methods and to perform gene walking also in non-pure culture samples, the Two-Step Gene Walking technique was developed. Altogether, over 20 kb of sequence were characterized by Two-Step Gene Walking in very different sample systems.

Over 85% of the performed attempts were successful using Two-Step Gene Walking. In the other cases, there was either no detectable Walking-PCR product or unreadable sequence raw data. But in every case it was possible to keep on walking replacing the Walking-PCR primer, the sequencing primer or both of them.

Walking-PCR primers and sequencing primers that produced unambiguous results meet the following criteria: (i) C-rich sequence at the 3-prime end (preferably CC or CCC), (ii) size of 18 to 25 bp and (iii) melting temperature of 60 to 63°C. Primer design according to these criteria

PCR product (step 2). The basic principle is outlined in Figure 1. Primers were obtained from MWG Biotech AG (Germany) and their sequences are listed in supporting material. Walking-PCRs were performed using 0.25 μl (100 μM) of one specific primer, 0.25 μl Ex-Taq (Takara Bio Inc., Japan), 5 μl Ex-*Taq* buffer (10×), 5 μl dNTP mixture (2.5 mM each), 38.5 μl ultra pure water and 1 μl template DNA (50 ng). PCRs were performed using a Primus 96 Plus (MWG Biotech AG, Germany) and the cycling program shown in Table 1.

Aliquots of Walking-PCR products were examined through agarose gel electrophoresis. The rest was purified using Perfectprep Kit (Eppendorf, Germany) to remove the remaining primers of the Walking-PCR. Typically, 5 μl of purified PCR product were directly sequenced by MWG Biotech AG (Germany) using a specific unmodified nested primer (supporting material) and an ABI 3730 XL system. For partially characterized PCR products, complete sequence was obtained through successive sequencing of the PCR product with newly designed nested primers (primer walking).

Chromatograms were analyzed and sequence stretches overlapping with known sequences were searched using

Figure 2. Typical walking-PCR product patterns. Different Walking-PCR products, which were successfully used for direct sequencing with nested specific primers, are shown. Patterns show big variety based on different templates and different primers. 5 µl aliquots were loaded on a 1% agarose gel and subjected to electrophoresis. 1 kb DNA Ladder (Invitrogen, USA) is shown in first and in last lane. Numbers stand for corresponding fragment lengths of respective bands.

allowed the development of a standard Walking-PCR program (Table 1). The annealing temperature is adjusted to 56°C in specific-annealing cycles and to 40°C in the unspecific-annealing cycle. For confident contig assembly, a minimal distance of 70 bp is required between the priming site of the sequencing primer and the end of the known sequence.

The Walking-PCR products analyzed on agarose gels showed in most cases a pattern representing fragments of different sizes (Figure 2). This can be explained by different drop off sites of the polymerase and due to different binding sites of the primer during the unspecific-annealing cycle. It results in a mixture of specific PCR products, which are heterogeneous in length. Also, a fraction of unspecific fragments is expected, because the primer could also bind unspecifically in other regions of the genome during the low stringency cycle. To estimate the portion of unspecific fragments produced during the Walking-PCR, cloning experiments were performed (data not shown). Some Walking-PCR products, with which gene walking as described above was successful, were selected. The proportion of clones with specific inserts did not exceed the proportion with unspecific inserts. This suggests that there is a significant fraction of unspecific amplification during the Walking-PCR, nevertheless the unspecific products do not interfere with the following step of direct sequencing. Therefore, direct sequencing with a specific nested primer has to be favored over a cloning and sequencing approach.

The chromatograms of successful sequencing attempts showed different typical features. In most cases, the sequence quality was high up to 600–900 bp (Figure 3A). Partially sequenced PCR products could be completely characterized using a newly designed nested primer, which was used for sequencing of the same PCR product (primer walking). In most cases, the Walking-PCR products were long enough to use them in multiple successive sequencing reactions. Thus, there is no need for the performance of a new Walking-PCR before the sequence of a Walking-PCR product is completely characterized through primer walking. This is crucial for the high efficiency of the Two-Step Gene Walking method.

A second typical feature of chromatograms is the type of sequence at the 3-prime end of a Walking-PCR product. Regularly, the reverse complement Walking-PCR primer sequence could be detected at the end of a sequenced fragment (Figure 3B). This is in accordance with the theory, which proposes unspecific binding of the Walking-PCR primer during the unspecific-annealing cycle. Therefore, the reverse complement sequence of the Walking-PCR primer has to be removed at the 3-prime sequence end of a completed PCR product. Further gene walking steps can elucidate the real genomic sequence, to which the Walking-PCR primer bound unspecifically. As expected, the genomic site of unspecific annealing of the Walking-PCR primer was in most cases quite complementary to the Walking-PCR primer (e.g. 12 complement base pair of 21 bp primer length; Figure 3B), especially at the 3-prime end that is fundamental for the proper binding of the polymerase. Also, the Walking-PCR product patterns on agarose gels (Figure 2) support this, showing mainly multiple sharp bands instead of a smear.

Another chromatogram feature was only sporadically observed, but illustrates the mechanism of Two-Step Gene Walking very clearly. Less than 3% of the obtained sequence raw data showed a short stretch (length of the Walking-PCR primer plus 1–2 bp) of clear peaks and double peaks within the sequence (Figure 3C). Afterwards, the chromatogram proceeded with peaks of high quality but with lower amplitudes. The region of ambiguities is produced by the concomitant presence of: (i) the 3-prime end of a short fragment that contains the reverse complement Walking-PCR primer sequence (generally, like in the case of Figure 3C, the higher peaks); (ii) the genomic sequence of the longer PCR products. The decrease of the peak amplitudes after the end of the shorter fragment can be explained by lower template amount. One or two double peaks adjacent to the primer-binding region emerge due to the terminal transferase 'A' activity of the used polymerase.

Genomic organization of bacterial tubulin genes

Initially, Jenkins *et al.* (35) reported the presence of bacterial tubulin (*btub*) genes in three *Prosthecobacter* species. Although no sequence data were released, the authors described the existence of a third gene, located downstream of bacterial B tubulin in *P. dejongeii*, that was referred to as kinesin light chain homologue. Pilhofer *et al.* (36) confirmed the presence of bacterial tubulin genes in *Prosthecobacter vanneervenii* and *P. debontii* and discovered a duplication of *btub* genes in the latter one.

Two-Step Gene Walking was used in this study to characterize the genomic environment of all *btub*-genes in *P. vanneervenii* and *P. debontii*. The EMBL nucleotide database entries AM041148 to AM041150 were extended with the new sequence data. Characterized genes and genomic environments are shown in Figure 4.
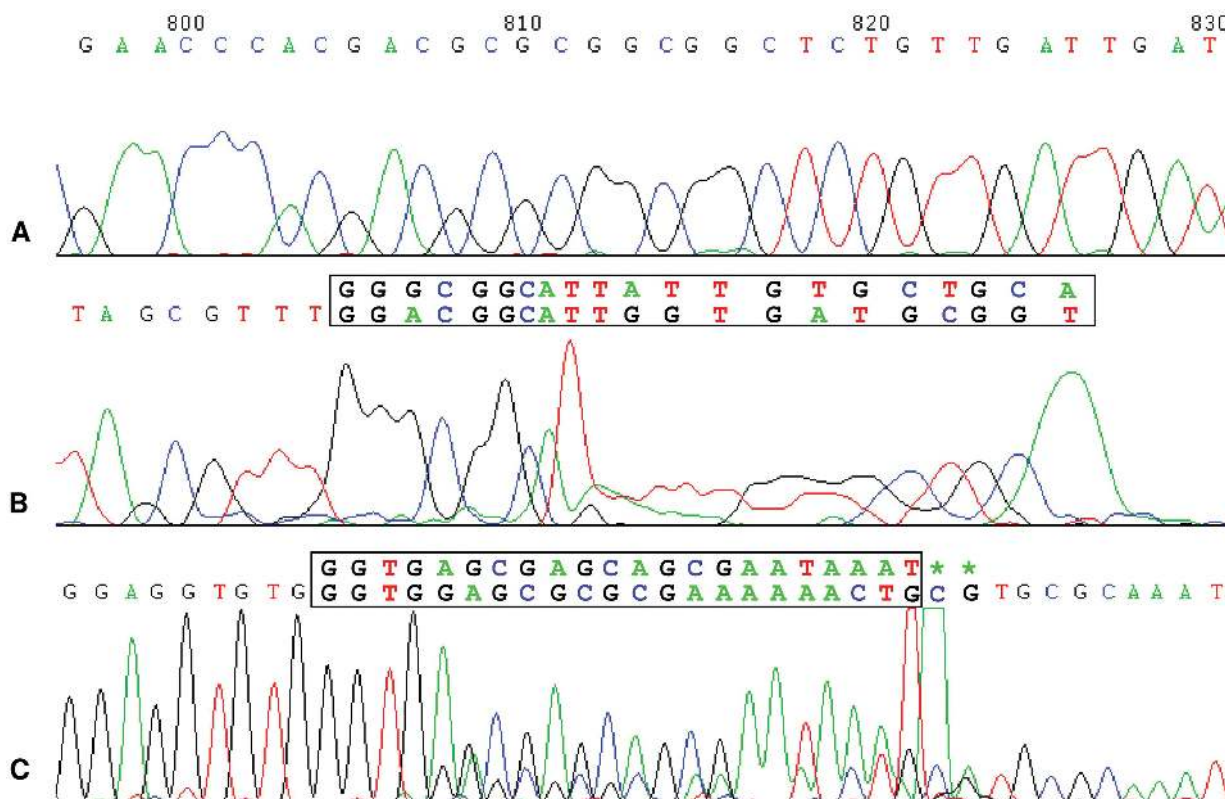
Figure 3. Features of sequence raw data. Panels show sequence raw data of three different Walking-PCR products, sequenced directly with specific nested primers. (**A**) displays high-quality sequence in region 800–830 bp; numbers indicate sequenced base pairs. (**B**) shows the sequence raw data at the end of a Walking-PCR fragment. Boxed sequence highlights Walking-PCR primer-binding region. Upper sequence displays reverse complement sequence of PCR primer, which is mainly recovered by peaks. Lower sequence displays real genomic sequence obtained through another walking step. (**C**) shows raw data of sequence at the end of one specific shorter Walking-PCR fragment and continuing with the sequence of one specific longer Walking-PCR fragment. This feature only occurs in <3% of all sequences. Boxed sequence highlights the binding region of the Walking-PCR primer. Upper sequence displays reverse complement sequence of the Walking-PCR primer; identical to higher peaks. Lower sequence displays real genomic sequence obtained through sequencing of reverse strand of same PCR product; identical to lower peaks. One or two additional double peaks at 3-prime end emerge due to terminal transferase activity (of adenosines) of used polymerase (indicated with asterisk).

All *Prosthecobacter* B tubulin genes characterized by Pilhofer *et al.* (36) are also followed by an open reading frame that shows some similarity to kinesin light chain. According to the original designation (35), this third gene will be referred to as *bklc* for bacterial kinesin light chain. The presence of *bklc* genes downstream of all bacterial B tubulin genes so far characterized clearly suggests that the bacterial kinesin light chain is an essential part of a functional unit, likely an operon, represented by one bacterial A tubulin, one bacterial B tubulin and one bacterial kinesin light chain [in accordance with (35)]. Henceforth, this cluster of genes will be referred to as bacterial tubulin operon (*btub*-operon).

Also gene expression features (e.g. promoters, ribosomal binding sites and terminators) support that these three genes are part of a single typical bacterial operon (see updated features of database entries AM041148 to AM041150) as well as RT–PCR results on *P. dejongeii* (35).

The analysis of the genomic environment of the *btub*-operon shows that the genes upstream of bacterial A tubulins always appear functionally related to those downstream of the bacterial kinesin light chain gene (Figure 4). None of the bordering genes indicated
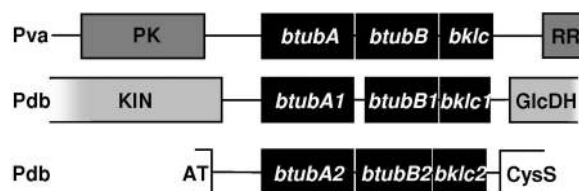
Figure 4. Bacterial tubulin operons including their genomic environment. Besides the same setup of all three bacterial tubulin operons (black), their genomic environments are different and show no functional relationship to cell division, cytoskeleton or typical eukaryotic genes. The two bordering genes of each operon show a functional relationship in any case (genes in dark gray involved in signal transduction mechanisms; genes in light gray involved in carbohydrate transport and metabolism; genes in white involved in amino acid transport and metabolism; PK: serine/threonine-protein kinase; RR: response regulator; KIN: pyruvate/orthophosphate dikinase; GlcDH: glucose/sorbosone dehydrogenase; AT: aspartate aminotransferase; CysS: cysteine synthase). Pva: *Prosthecobacter vanneervenii*, Pde: *Prosthecobacter debontii*

a functional relationship to cell cycle or cytoskeleton. In all three cases the *btub*-operon appears as an insert interrupting functionally related, but always different, genes.

Genes showing low similarities to eukaryotic kinesin light chains have been described in *Bacteria* since 1997 [*Plectonema boryanum*; (46)], but nothing is known about their function. All *Prosthecobacter* bacterial kinesin light chain sequences show very low similarity to eukaryotic and bacterial kinesin light chain sequences detected using BLASTP ($3.7\,e^{-6}$ to $4.4\,e^{-7}$). Fingerprint analysis (43) of verrucomicrobial kinesin light chain sequences could recover only two or three out of the six typical kinesin light chain motifs. A tandem repeat of 4 respectively 5 tetratricopeptide repeat (TPR) domains could be detected in all *Prosthecobacter* Bklc sequences. TPR typically contains 34 amino acids and is found in both bacteria and eukaryotes; it is involved in many functions including protein–protein interactions (47). The function of *bklc* in *Prosthecobacter* still remains to be elucidated, but a functional relationship to *btub*-genes can be supposed based on its genomic organization within the *btub*-operon and the presence of TPR domains.

Application of Two-Step Gene Walking on non-pure culture systems

To our knowledge, the existing gene walking methods were only used for pure culture systems. Two-Step Gene Walking was successfully applied on two different complex non-pure culture sample systems, using the standardized protocol without any optimization.

Epixenosomes are episymbiotic bacteria inhabiting the surface of the hypotrich ciliate *Euplotidium* (48). PCR-analysis on the epixenosome rRNA operon with specific primers targeting 16S and 23S rRNA genes resulted in no amplification product (P. G., unpublished data). This suggested that 16S and 23S rRNA genes are not joined in this organism. Two-Step Gene Walking was applied to characterize the upstream region of the 23S rRNA gene by gene walking. Up to now, these organisms are not available as pure culture and can only be grown together with their host. Therefore, the sample used for DNA preparation also included the host organism *Euplotidium*, as well as several free-living bacteria that were present in the host culture and some of the algae used as food source for the host. Nevertheless, Two-Step Gene Walking proved to be successful to characterize a 1522 bp region upstream of the 23S rRNA gene. An open reading frame coding for a transposase, a tRNA^ile gene and a tRNA^ala gene (from 5-prime to 3-prime) could be detected. The nucleotide sequence was submitted to GenBank with accession number EF650087.

Caedibacter caryophilus are endosymbionts localized in the macronucleus of *P. caudatum* (49,50). *Paramecium caudatum* cannot be grown axenically and, in addition to regularly added food bacteria, a community of diverse free-living bacteria is present in the culture medium. These contaminants complicate each molecular approach. The 16S rRNA gene of *C. caryophilus* shows an unusual insertion of 194 bp within the 5-prime-terminal region. This sequence is not present in mature rRNA and is therefore called internal excised element (IEE) (51,52). Two-Step Gene Walking starting from the IEE was used to characterize the upstream region of the 16S rRNA gene.

Two-Step Gene Walking was successfully used to characterize 936 bp of sequence. The sequence comprised the 5-prime end of the 16S rRNA gene of *C. caryophilus* and 767 bp of the upstream region. An open reading frame coding for a putative soluble lytic murein transglycosylase precursor was detected. The nucleotide sequence was submitted to EMBL nucleotide database with accession number AM743196.

CONCLUSIONS

In comparison to other gene walking techniques, Two-Step Gene Walking has striking advantages, which are detailed below. It proved to be a very simple gene walking method, applicable to very different organisms. The required minimal amount of DNA (50 ng) makes it also feasible to perform gene walking in organisms for which a limited amount of biomass is available, such as slowly growing or even uncultivable organisms. To our knowledge, Two-Step Gene Walking is the first gene walking technique, which was shown to work in complex non-pure culture samples. The method was always processed using the standardized protocol without any optimization. The two-step workflow can be processed within one day, comprising only one Walking-PCR with one specific primer and direct sequencing with one specific nested primer. Over 20 kb of sequence were characterized exclusively by Two-Step Gene Walking, which is more than it is reported for other gene walking techniques. The derived sequence reads showed very high quality up to 900 bp.

Concluding, Two-Step Gene Walking overcomes the major disadvantages of other gene walking techniques. Inverse PCR and ligation-mediated PCR require time-consuming and labor-intensive enzymatic modifications of the target DNA, a relatively high amount of target DNA (several micrograms) and other DNA modifying enzymes in addition to DNA polymerase. The advantages of Two-Step Gene Walking compared to other randomly primed PCR methods are the independence of (i) modified oligonucleotides (18,19,23,27,29,31), (ii) special walking primers (20,23,24,26,31,33,34), (iii) successive PCRs (20,23–27,29,31–34), (iv) cloning steps (21,25,26,28, 29,31,33,34), (v) procedures to screen for specificity or to enrich specific fragments (20,21,25,27–29,31–34) and (vi) other enzymatic modifications than the DNA polymerase (27). Some methods are restricted and optimized to only one walking step starting always from the same known sequence (e.g. an inserted transposon) (28,30). In contrast, the developed method enables theoretically to completely characterize one DNA molecule using successive gene walking steps and primer walking on Walking-PCR products. Potential applications of Two-Step Gene Walking are (i) rapidly obtaining full gene sequences, finding the corresponding promoters and regulatory elements, (ii) characterization of complete operons starting from a small known DNA fragment, (iii) characterization of primer binding sequences (particularly important when the used primers work suboptimal), (iv) identification of transposon integration

sites in known or unknown genomes for gene function studies and (v) gap closure in genome sequencing projects. The latter one could be also performed in a high-throughput scale due to the simple and standardized protocol of Two-Step Gene Walking.

The characterization of the genomic environment of *Prosthecobacter* tubulin genes using Two-Step Gene Walking revealed a typical bacterial operon structure. Strikingly, in addition to *btubA* and *btubB*, the *btub*-operon comprises also a third, kinesin-like gene (*bklc*). The sequence for *bklc* is reported here for the first time.

The *btub*-operons are integrated in a set of genes functionally related among themselves but apparently not with the *btub*-operon. The genomic environments of the different *btub*-operons are always different (Figure 4). These two facts further support the hypothesis of a horizontal gene transfer of *btub*-genes to *Prosthecobacter* (36,37), but the theory has now to be extended to the whole *btub*-operon including also *bklc*. This of course leads to a different view on these peculiar bacterial genes and will influence future functional studies on bacterial tubulins.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

REFERENCES

1. Saiki,R.K., Scharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Erlich,H.A. and Arnheim,N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230**, 1350–1354.
2. Mullis,K., Faloona,F., Scharf,S., Saiki,R., Horn,G. and Erlich,H. (1986) Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. In *Cold Spring Harbor Symposia on Quantitative Biology*, 51 Pt 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, pp. 263–273.
3. Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.
4. 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.
5. Krin,E., Hommais,F., Soutourina,O., Ngo,S., Danchin,A. and Bertin,P. (2001) Description and application of a rapid method for genomic DNA direct sequencing. *FEMS Microbiol. Lett.*, **199**, 229–233.
6. Ochman,H., Gerber,A.S. and Hartl,D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics*, **120**, 621–623.
7. Triglia,T., Peterson,M.G. and Kemp,D.J. (1988) A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res.*, **16**, 8186.
8. Huang,S.H. (1994) Inverse polymerase chain reaction. An efficient approach to cloning cDNA ends. *Mol. Biotechnol.*, **2**, 15–22.
9. Kohda,T. and Taira,K. (2000) A simple and efficient method to determine the terminal sequences of restriction fragments containing known sequences. *DNA Res.*, **7**, 151–155.
10. Shyamala,V. and Ames,G.F. (1989) Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. *Gene*, **84**, 1–8.
11. Rosenthal,A. and Jones,D.S. (1990) Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction. *Nucleic Acids Res.*, **18**, 3095–3096.
12. Siebert,P.D., Chenchik,A., Kellogg,D.E., Lukyanov,K.A. and Lukyanov,S.A. (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.*, **23**, 1087–1088.
13. Diatchenko,L., Lau,Y.F., Campbell,A.P., Chenchik,A., Moqadam,F., Huang,B., Lukyanov,S., Lukyanov,K., Gurskaya,N. *et al.* (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl Acad. Sci. USA*, **93**, 6025–6030.
14. Kilstrup,M. and Kristiansen,K.N. (2000) Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. *Nucleic Acids Res.*, **28**, E55.
15. Drancourt,M., Carlioz,A. and Raoult,D. (2001) rpoB sequence analysis of cultured *Tropheryma whippelii*. *J. Clin. Microbiol.*, **39**, 2425–2430.
16. Yuanxin,Y., Chengcai,A., Li,L., Jiayu,G., Guihong,T. and Zhangliang,C. (2003) T-linker-specific ligation PCR (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends. *Nucleic Acids Res.*, **31**, e68.
17. Rishi,A.S., Nelson,N.D. and Goyal,A. (2004) Genome walking of large fragments: an improved method. *J. Biotechnol.*, **111**, 9–15.
18. Lin,J., Liu,J., Sun,X., Zhou,X., Fei,J. and Tang,K. (2005) An efficient method for rapid amplification of *Arisaema heterophyllum* agglutinin gene using a genomic walking technique. *Prep. Biochem. Biotechnol.*, **35**, 155–167.
19. Nthangeni,M.B., Ramagoma,F., Tlou,M.G. and Litthauer,D. (2005) Development of a versatile cassette for directional genome walking using cassette ligation-mediated PCR and its application in the cloning of complete lipolytic genes from *Bacillus* species. *J. Microbiol. Methods*, **61**, 225–234.
20. Parker,J.D., Rabinovitch,P.S. and Burmer,G.C. (1991) Targeted gene walking polymerase chain reaction. *Nucleic Acids Res.*, **19**, 3055–3060.
21. Parks,C.L., Chang,L.S. and Shenk,T. (1991) A polymerase chain reaction mediated by a single primer: cloning of genomic sequences adjacent to a serotonin receptor protein coding region. *Nucleic Acids Res.*, **19**, 7155–7160.
22. Caetano-Anolles,G. (1993) Amplifying DNA with arbitrary oligonucleotide primers. *PCR Methods Appl.*, **3**, 85–94.
23. Dominguez,O. and Lopez-Larrea,C. (1994) Gene walking by unpredictably primed PCR. *Nucleic Acids Res.*, **22**, 3247–3248.
24. Liu,Y.G. and Whittier,R.F. (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics*, **25**, 674–681.
25. Malo,M.S., Srivastava,K., Andresen,J.M., Chen,X.N., Korenberg,J.R. and Ingram,V.M. (1994) Targeted gene walking by low stringency polymerase chain reaction: assignment of a putative human brain sodium channel gene (SCN3A) to chromosome 2q24-31. *Proc. Natl Acad. Sci. USA*, **91**, 2975–2979.
26. Trueba,G.A. and Johnson,R.C. (1996) Random primed gene walking PCR: a simple procedure to retrieve nucleotide fragments adjacent to known DNA sequences. *Biotechniques*, **21**, 20.

27. Min,G.S. and Powell,J.R. (1998) Long-distance genome walking using the long and accurate polymerase chain reaction. *Biotechniques*, **24**, 398–400.

28. Spector,M.P., DiRusso,C.C., Pallen,M.J., Garcia,d.P., Dougan,G. and Finlay,B.B. (1999) The medium-/long-chain fatty acyl-CoA dehydrogenase (*fadF*) gene of *Salmonella typhimurium* is a phase 1 starvation-stress response (SSR) locus. *Microbiology*, **145**(Pt 1), 15–31.

29. Hermann,S.R., Miller,J.A., O'Neill,S., Tsao,T.T., Harding,R.M. and Dale,J.L. (2000) Single-primer amplification of flanking sequences. *Biotechniques*, **29**, 1176–8, 1180.

30. Karlyshev,A.V., Pallen,M.J. and Wren,B.W. (2000) Single-primer PCR procedure for rapid identification of transposon insertion sites. *Biotechniques*, **28**, 1078, 1080, 1082.

31. Mishra,R.N., Singla-Pareek,S.L., Nair,S., Sopory,S.K. and Reddy,M.K. (2002) Directional genome walking using PCR. *Biotechniques*, **33**, 830–2, 834.

32. Antal,Z., Rascle,C., Fevre,M. and Bruel,C. (2004) Single oligonucleotide nested PCR: a rapid method for the isolation of genes and their flanking regions from expressed sequence tags. *Curr. Genet.*, **46**, 240–246.

33. Gould,V.C., Okazaki,A., Howe,R.A. and Avison,M.B. (2004) Analysis of sequence variation among smeDEF multi drug efflux pump genes and flanking DNA from defined 16S rRNA subgroups of clinical *Stenotrophomonas maltophilia* isolates. *J. Antimicrob. Chemother.*, **54**, 348–353.

34. Tan,G., Gao,Y., Shi,M., Zhang,X., He,S., Chen,Z. and An,C. (2005) SiteFinding-PCR: a simple and efficient PCR method for chromosome walking. *Nucleic Acids Res.*, **33**, e22.

35. Jenkins,C., Samudrala,R., Anderson,I., Hedlund,B.P., Petroni,G., Michailova,N., Pinel,N., Overbeek,R., Rosati,G. *et al.* (2002) Genes for the cytoskeletal protein tubulin in the bacterial genus *Prosthecobacter*. *Proc. Natl Acad. Sci. USA*, **99**, 17049–17054.

36. Pilhofer,M., Rosati,G., Ludwig,W., Schleifer,K.H. and Petroni,G. (2007) Coexistence of Tubulins and *ftsZ* in Different *Prosthecobacter* Species. *Mol. Biol. Evol.*, **24**, 1439–1442.

37. Schlieper,D., Oliva,M.A., Andreu,J.M. and Lowe,J. (2005) Structure of bacterial tubulin BtubA/B: Evidence for horizontal gene transfer. *Proc. Natl Acad. Sci. USA*, **102**, 9170–9175.

38. Sontag,C.A., Staley,J.T. and Erickson,H.P. (2005) In vitro assembly and GTP hydrolysis by bacterial tubulins BtubA and BtubB. *J. Cell Biol.*, **169**, 233–238.

39. Rosati,G., Verni,F. and Lenzi,P. (1993) Epixenosomes - peculiar epibionts of the ciliate *Euplotidium-Itoi* - the formation of the extrusive apparatus and the ejecting mechanism. *Europ. J. Protistol.*, **29**, 238–245.

40. Wisotzkey,J.D., Jurtshuk,P.Jr. and Fox,G.E. (1990) PCR amplification of 16S rDNA from lyophilized cell cultures facilitates studies in molecular systematics. *Curr. Microbiol.*, **21**, 325–327.

41. Wheeler,D.L., Church,D.M., Federhen,S., Lash,A.E., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Sequeira,E. *et al.* (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res.*, **31**, 28–33.

42. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J.H., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

43. Attwood,T.K., Bradley,P., Flower,D.R., Gaulton,A., Maudling,N., Mitchell,A.L., Moulton,G., Nordle,A., Paine,K. *et al.* (2003) PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res.*, **31**, 400–402.

44. Bailey,S.F. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, 28–36.

45. Reese,M.G. (2001) Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput. Chem.*, **26**, 51–56.

46. Celerin,M., Gilpin,A.A., Dossantos,G., Laudenbach,D.E., Clarke,M.W. and Beushausen,S. (1997) Kinesin light chain in a eubacterium. *DNA Cell Biol.*, **16**, 787–795.

47. Blatch,G.L. and Lassle,M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*, **21**, 932–939.

48. Petroni,G., Spring,S., Schleifer,K.H., Verni,F. and Rosati,G. (2000) Defensive extrusive ectosymbionts of *Euplotidium* (Ciliophora) that contain microtubule-like structures are bacteria related to *Verrucomicrobia*. *Proc. Natl Acad. Sci. USA*, **97**, 1813–1817.

49. Schmidt,H.J., Gortz,H.D. and Quackenbush,R.L. (1987) *Caedibacter caryophila* sp. nov., a killer symbiont inhabiting the macronucleus of *Paramecium caudatum*. *Int. J. Syst. Bacteriol.*, **37**, 459–462.

50. Euzeby,J.P. (1997) Revised nomenclature of specific or subspecific epithets that do not agree in gender with generic names that end in –bacter. *Int. J. Syst. Bacteriol.*, **47**, 585.

51. Springer,N., Ludwig,W., Amann,R., Schmidt,H.J., Gortz,H.D. and Schleifer,K.H. (1993) Occurrence of fragmented 16S rRNA in an obligate bacterial endosymbiont of *Paramecium caudatum*. *Proc. Natl Acad. Sci. USA*, **90**, 9892–9895.

52. Schrallhammer,M., Fokin,S.I., Schleifer,K.H. and Petroni,G. (2006) Molecular characterization of the obligate endosymbiont ''Caedibacter macronucleorum'' Fokin and Görtz 1993 and of its host *Paramecium duboscqui* strain Ku4-8. *J. Eukaryot. Microbiol.*, **53**, 499–506.