Gene 500 (2012) 186–193

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Comprehensive DNA copy number profile and BAC library construction of an Indian individual

Sanjiban Chakrabarty [a,1], Reena R. D'Souza [a,1], Ravishankara Bellampalli [a], Harish Rotti [a], Abdul V. Saadi [a], Puthiya M. Gopinath [a], Raviraja V. Acharya [b], Periyasamy Govindaraj [c], Kumarasamy Thangaraj [c], Kapaettu Satyamoorthy [a,*]

[a] Division of Biotechnology, Manipal Life Sciences Centre, Manipal University, Manipal 576104, Karnataka state, India
[b] Department of Medicine, Kasturba Medical College, Manipal University, Manipal 576104, Karnataka state, India
[c] Centre for Cellular and Molecular Biology, Hyderabad 500007, Andhra Pradesh state, India

ARTICLE INFO

Article history:
Accepted 13 March 2012
Available online 21 March 2012

Keywords:
Bacterial artificial chromosome
BAC
Array based CGH
Copy number variation
Taqman assay
Fluorescent *in situ* hybridization

ABSTRACT

Bacterial artificial chromosomes (BACs) are used in genomic variation studies due to their capacity to carry a large insert, their high clonal stability, low rate of chimerism and ease of manipulation. In the present study, an attempt was made to create the first genomic BAC library of an anonymous Indian male (IMBL4) consisting of 100,224 clones covering the human genome more than three times. Restriction mapping of 255 BAC clones by pulse field gel electrophoresis confirmed an average insert size of 120 kb. The library was screened by PCR using *SHANK3* (SH3 and multiple ankyrin repeat domains 3) and *OLFM3* (olfactomedin 3) specific primers. A selection of clones was analyzed by fluorescent *in situ* hybridization (FISH) and sequencing. Fine mapping of copy number variable regions by array based comparative genomic hybridization identified 467 CNVRs in the IMBL4 genome. The IMBL4 BAC library represents the first cataloged Indian genome resource for applications in basic and clinical research.

1. Introduction

Translational genomic research requires development of evidence-based diagnostic, therapeutic and prognostic clinical protocols for the practice of personalized clinical medicine (Collins, 2011; Kumar, 2011). The advent of high-throughput technologies now permits investigators to interrogate the genome, transcriptome, proteome and metabolome in a systematic manner. The availability of the bacterial artificial chromosome (BAC) libraries with the integrated genetic information is a valuable resource for a wide variety of genomic applications because of their large insert carrying capacity, high clonal stability and low rate of chimerism (Shizuya et al., 1992; Shizuya and Kouros-Mehr, 2001). In the past, six human BAC libraries have been constructed (Asakawa et al., 1997; Kim et al., 1995, 1996, 2009; Osoegawa et al., 2001; Park et al., 2006; Terabayashi et al., 2011). With the availability of end sequenced human BAC clones, it has now become an important tool for identification of genes involved in single as well as multigene disorders (Osoegawa et al., 2001; Park et al., 2006). Human BAC clones are prospective candidates for further utilization as reagents for downstream sequence based characterization, functional cell based assays and for other validation purposes. One among many of the applications of these BAC clones is for the precise mapping of genes at breakpoints of chromosomal rearrangements and defining commonly deleted/amplified regions involved in various cancers and genetic disorders. It encompasses a microarray based comparative genomic hybridization (aCGH) technique, which has revolutionized detection of genomic aberrations in the form of copy number gains and losses (Bastian et al., 1998; Pinkel et al., 1998). The utility of BAC clones in array based CGH was established since its inception in the late nineties owing to its advantage of being the most reproducible platform (Hester et al., 2009; Wicker et al., 2007). Sequenced BAC clones can be used for construction of targeted mini arrays as a diagnostic tool for detection of submicroscopic deletions/duplications and chromosomal aneuploidies in clinical laboratories (Boone et al., 2010; Dhar et al., 2010; Stankiewicz et al., 2010). In addition, BACs that are identified in the arrayCGH (aCGH) experiment can subsequently be used as probes for hybridization to interphase/metaphase spreads for further validation of copy number changes (Ligon et al., 1997; Weise et al., 2010). BAC arrays constructed with known clinical loci, redundancy over each region, and minimal polymorphisms provide the greatest clinical utility (Brunet et al., 2009; De Preter et al., 2008; Park et al., 2010). However, BAC array has its

Abbreviations: BAC, bacterial artificial chromosome; PFGE, pulse field gel electrophoresis; CNV, copy number variation; arrayCGH, array based comparative genomic hybridization; PCR, polymerase chain reaction; SHANK3, SH3 and multiple ankyrin repeat domains 3; OLFM3, olfactomedin 3; FISH, fluorescent *in situ* hybridization.

* Corresponding author at: Manipal Life Sciences Centre, Manipal University, Manipal 576104, Karnataka state, India. Tel.: +91 820 2922058; fax: +91 820 2571919.
E-mail addresses: sanjiban.c@manipal.edu (S. Chakrabarty), reena.dsouza@manipal.edu (R.R. D'Souza), ravishankar.bsa@gmail.com (R. Bellampalli), vasu.harish@manipal.edu (H. Rotti), saadi.a@manipal.edu (A.V. Saadi), gopinathpm@yahoo.com (P.M. Gopinath), rvacharya@manipal.edu (R.V. Acharya), pgovinth@ccmb.res.in (P. Govindaraj), thangs@ccmb.res.in (K. Thangaraj), ksatyamoorthy@manipal.edu (K. Satyamoorthy).
[1] These authors contributed equally to this work.

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2012.03.054

limitations while addressing small structural variations and complex chromosomal rearrangements (Neill et al., 2010; Ylstra et al., 2006).

In the present study, we have constructed the first Indian human BAC library with the annotated copy number profile of the donor using genome wide microarray based aCGH platform and further established the utility of the cataloged BAC clones. In the implementation of the screening strategy pursued here, individual BAC clones from 384 well plates have been pooled in a hierarchical fashion and screened for *SHANK3* (SH3 and multiple ankyrin repeat domains 3) and *OLFM3* (olfactomedin 3) genes followed by end sequencing and fluorescent *in situ* hybridization (FISH). We have delineated the structural variation of the IMBL4 genome using high-density genome wide array based CGH platform and further validated the findings using real time SYBR green and Taqman polymerase chain reaction (PCR) assays. Our study explores the opportunity to delineate the genome of a male of Indian ethnic origin and facilitates the availability of characterized BAC clones. It also provides a resource of annotated copy number variation (CNV) data for other personal genome studies and is a significant addition to previously established human BAC libraries.

2. Results

2.1. Library construction

We have constructed human BAC library from the DNA of an anonymous Indian male donor of normal karyotype covering the haploid human genome more than three times. Epstein–Barr virus (EBV) immortalized lymphoblastoid cell line was also generated from the lymphocytes of the donor as a stable source of DNA for future studies. Immortalization was confirmed by flow cytometric analysis and the amplification of the EBNA2 gene from the DNA obtained from the cell line. The library consists of 100,224 BAC clones arrayed into two hundred sixty one 384 well microtiter plates. To facilitate long-term maintenance, the library was replicated and used for PCR screening application.

2.2. Insert size distribution

Restriction mapping using *Not* I enzyme followed by pulse field gel electrophoresis (PFGE) showed an average insert size of 120 kb with insert length ranging from 50 to 250 kb in size (Figs. 1a and b). To characterize the clones of the BAC library, we have end sequenced BAC clones using Sanger based di-deoxy sequencing with T7 and M13 reverse universal sequencing primers (Fig. 1d). We have observed 5.1% non-recombinant BAC clones (13 out of 255) and the empty well percentage was less than 1% confirming 94.9% recombinant BAC clones in the library.

2.3. Identification of CNV regions in IMBL4 genome

We used high-density CGH and single nucleotide polymorphisms (SNP) genotyping arrays to map regions of relative gains and losses in IMBL4 genome. The Agilent 244K aCGH platform identified 467 CNV segments (Fig. 2). Out of 467 CNV segments, 80 regions represented by multi probe CNVRs with 387 CNVRs detected by single probes accounting for 182 gains and 285 losses ranging in size from 238 bp to 3.6 Mb (Fig. 3, Supplementary Table 1). When comparing

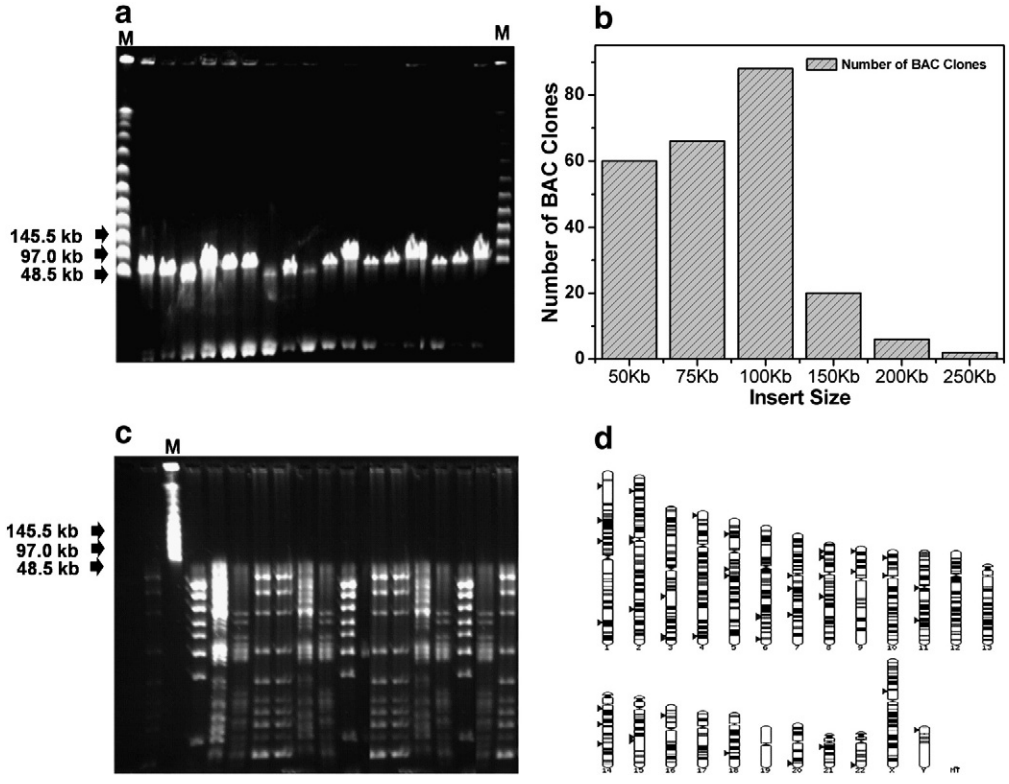

Fig. 1. Restriction mapping and end sequencing of recombinant BAC clones from the IMBL4 BAC library. (a) A typical agarose pulsed-field gel electrophoresis image showing *Not* I digested BAC clones from the library. The vertical axis describes the pulse field gel lanes with the first and last lane loaded with a lambda DNA PFGE high molecular weight marker with a size range of 50–1000 kb (Sigma Aldrich, St. Louis, MO) denoted as M and the rest of the lanes with *Not* I digested BAC clones from the human BAC library. (b) Insert size distribution of the Indian human male BAC library. The vertical axis refers to the number of clones and the horizontal axis refers to the size of the insert DNA. (c) BAC clones digested with *Hind* III and resolved on 1% pulsed-field gel electrophoresis. The vertical axis describes the pulse field gel lanes with second lane is loaded with a lambda DNA PFGE high molecular weight marker with a size range of 50–1000 kb (Sigma Aldrich, St. Louis, MO) denoted as M and rest of the lanes with *Hind* III digested BAC clones from the human BAC library. (d) Chromosomal distribution of end sequenced BAC clones from IMBL4 BAC library. The vertical axis that describes BAC clones identified by end sequencing are highlighted as a black arrow on the chromosomal ideogram according to their location in the genome.