Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biomolecular Engineering 21 (2005) 135–144
Biomolecular Engineering
www.elsevier.com/locate/geneanabioeng

Cloneless genomic DNA analysis: an efficient and simple methods for de novo genomic sequencing projects and gap filling

Giang Nguyen[a], Nikolayev Bukanov[a], M. Oshimura[b], Cassandra L. Smith[a,*]

[a]*Molecular Biotechnology Research Laboratory, Departments of Biomedical Engineering, Biology and Pharmacology and Experimental Therapeutics, 36 Cummington Street, Boston, MA 02215, USA*
[b]*School of Life Science, Faculty of Medicine, Tottori University, 86 Nishimachi, Yonago, Tottori 683-0826, Japan*

Received 1 August 2004; accepted 19 August 2004

Abstract

The utility of using genomic DNA directly in agarose, i.e. cloneless libraries, in place of large clone libraries, radiation hybrid panels, or chromosome dissection was demonstrated. The advantage of the cloneless library approach is that, in principle, a targeted genomic resource can be developed rapidly for any genomic region using any genomic DNA sample. Here, a human chromosome 20 *Not* I fragment library was generated by slicing a pulsed field gel lane containing fractionating *Not* I cleaved DNA from a monosomic hybrid cell line into 2 mm pieces. A reliable PCR method using agarose embedded DNA was developed. InterAlu PCR generated unique patterns of products from adjacent slices (e.g. fractions). Further, the specificity of the interAlu products was demonstrated by FISH analysis and in other hybridization experiments to arrayed interAlu products. STS content mapping was used to order the fractions and also demonstrate the unique content of the library fractions.

Keywords: DNA; Clone; Genome; Cloneless; Gap filling; Genomic sequencing; Genomic mapping

1. Introduction

Top-down sequencing of large genomes uses clone libraries that either are completely or partially ordered (e.g. [1–7]) by fingerprinting or cross-hybridization methods [8–10]. In the past, we developed a top-down approach for mapping genomes without the use of clone libraries. Our most extensive mapping experiments effort to date, created a *Not* I restriction map of human chromosome 21 by ordering 60 fragments in 9 different cell lines and locating 11 chromosomal breakpoints using 80 markers [11]. The major focus of these experiments was to link adjacent fragments. Several strategies were used to prove adjacency including linking libraries, partial digest, and polymorphism linkup and overlap (see [12,13] for discussion). Several hundred time-consuming and labor-intensive hybridizations were done to complete this work.

During the course of the chromosome 21 mapping experiments, gap filling was done using interAlu PCR generated probes from *Not* I fragments isolated in gel slices [11,14]. These experiments generated clone libraries from specific genomic restriction fragments. At the same time, we used electrophoretically fractionated DNA fragments in agarose slices for fingerprinting and ordering of ~1700 *Schizosaccharomyces pombe* cosmid clones [15]. Complete cosmid library ordering was done in ~60 hybridization experiments and represented an eight-fold reduction in the amount of work needed to order similar genomic libraries. These experiments demonstrated the utility of using gel slices as fractions as hybridization probes for ordering clone libraries. Here, the utility of using gel slices containing genomic restriction fragments directly, i.e. cloneless libraries, for large scale analysis is further demonstrated in studies on human chromosome 20.

* Corresponding author. Tel.: +1 617 353 8503; fax: +1 617 353 8506.
E-mail address: clsmith@bu.edu (C.L. Smith).

1389-0344/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bioeng.2004.08.002

2. Materials and methods

2.1. Cloneless library generation

Genomic DNA from a monosomic hybrid cell line, A9neo20 cell line [16] was used to generate the cloneless library. This cells line was made from mouse fibroblasts resistant to neomycin. The genomic DNA samples were purified in agarose, digested with the restriction enzyme *Not* I (or others as indicated), and fractionated by pulsed-field gel electrophoresis (PFG) as described (summarized in [17,18]). These procedures involve embedding live cells into molten agarose that has been tested to be free of enzyme inhibitors (e.g. InCert Agarose, FMC Corporation). Low gelling temperature agarose was used in order to avoid a temperature shock to the live cells. The molten agarose was solidified in blocks using a mold. The chromosomal DNA was purified by incubating samples in ESP (0.5 M EDTA (pH 9.0–9.5), 1% sodium lauroyl sarkosyl, and 1 mg/ml proteinase K). Before enzymatic manipulation, proteinase K was inactivated with a phenyl methyl sulfonate treatment and the EDTA and detergent were removed by dialysis with gentle agitation in TE buffer (10 mM Tris–Cl (pH 7.5) and 0.1 mM ethylendiaminetetraacetic acid (EDTA)). The gentle agitation insures rapid diffusion of the molecules into the agarose. For restriction enzyme digests, an extra 160 μg/ml bovine serum albumin (BSA: nuclease-free Roche) was added to each 200 μl total volume reaction that included the 100 μl agarose block and 10 U of enzyme. The sample was incubated at the appropriate temperature with gentle shaking at 37 °C. One sixth of an agarose block was used per gel lane. In some experiments, the distribution of genomic restriction fragments was evaluated by hybridizing an Alu probe to genomic DNA transferred to nylon membranes as described previously [19].

Here, the complete cloneless library was generated from three different PFG fractionations with different windows of optimal resolutions at (a) 0.05–1 Mb, (b) ~1–3 Mb, and (c) ~3–7 Mb. The gels were stained with 1 μg/ml ethidium bromide for 30 min in water and destained 30 min in water. The gel was cooled to 10 °C, and the entire gel lane, including the well and compression region, was cut into ~2 mm thick slices. Exposure to ultraviolet light was minimized and a fresh glass cover slip was used to generate each slice in order to avoid cross contamination between the slices.

The gel slices (~100 μl) were put into 100 μl TE buffer to a final volume 200 μl. The samples were heated to 95 °C for 15 min (to inactivate any nucleases), vortexed, and stored frozen at −20 °C. When needed, the samples were heated to 95 °C for 5–15 min, vortexed and 2–10 μl used in PCR. After testing (see results) of different preservatives, the standard protocol used 20 mM HPLC grade ethanolamine (final concentration) in the agarose DNA stored in TE, and 160 μg/ml DNase free BSA (Roche) in the PCR.

2.2. Long interAlu PCR

Long interAlu products were generated from PFG slices as described [14] except that the long PCR enzyme mixture used was from Roche. The first PCR done in 25 μl contained 1 μM interAlu primer TC65 [20], 3 mM MgCls, 300 μM each of dCTP, dGTP, dATP and dTTP in 10 mM Tris–Cl (pH 8.3), 50 mM KCl and 0.001% gelatin, with 0.6 U of Amplitaq DNA polymerase (Roche). The cycling program was at 94 °C for 4 min, then 35 cycles of 94 °C for 1 min, 55 °C for 1 min, 72 °C for 4 min and a final incubation of 72 °C for 10 min. Two microlitres of the first PCR was added to the second PCR carried out under the same conditions except that the primer concentration was 10 μM.

For hybridization analysis, the PCR product was diluted 10-fold and 5 μl was spotted onto HybondN nylon filters (Amersham). The DNA was denatured DNA and fixed to the filters at 80 °C under vacuum [21]. The interAlu producted were labeled with ^{32}P using a random primer kit (PrimeIt II Labeling Kit, 4 Stratagene). Filters with interAlu products were prehybrized at 68 °C in 6× SSC, 5× Denhardt's solution, 0.5% sodium dodecyl sulfate, and 100 μg/ml salmon sperm DNA for 4 h. Hybridization was carried out under the same conditions with the addition of oligonucleotides complementary to the primer and the repeat sequences on the ends of each product in for 18 h. then the filters were washed in 0.1% SDS, for 1 h at 68 °C after rinsing at room temperature.

2.3. Fluorescent in situ analysis (FISH)

InterAlu products were analyzed by FISH as described [22,23]. The interAlu sample was prepared as follows. About 1 μg (e.g. 1 μl) of the interAlu PCR products were labeled with Dig-11-dUMP (Roche) as described in a nick-translation reaction. The reaction contained 20 μM Dig-11-UTP, 20 μM each dATP, dGTP, and dCTP, 55 mM Tris–Cl (pH 7.8), 5 mM $MgCl_2$, 10 mM *s*-mercaptoethanol, 20 μg/mg BSA, 0.5 mM magnesium acetate, 5% (v/v) glycerol, 7.25 U DNA polymerase I (Gibco), and 0.03375 U DNase I (Gibco) in 50 μl incubated for at 15 °C for 2 h. Two microlitres of the sample was mixed with 5 μg human $C_{ot}1$ DNA, 2 μg salmon sperm DNA, precipitated, dried and resuspended in 3 μl water added to 7 μl hybridization mix [made up by combining 5 ml formamide, 1 g dextran sulfate, and 1 ml 20× SSC]. The sample was denatured at 70 °C for 5–10 min, spun and repeats were blocked by prehybridization at 37 °C for 1 h.

2.4. Sequence-tagged sites (STS)

The sequence-tagged sites, STSs (Table 1), that were used for mapping the *Not* I fragments were taken from lists developed by Gyapay et al. [24] and Hudson et al. [25]. The STSs were chosen to span the entire length of chromosome 20 at approximately equal intervals. Two PCR approaches

Table 1
Ordering of a cloneless *Not* I fragment library of chromosome 20 by STSs mapping

STS	I fragment (kb)	Genetic map position (cM)	Radiation hybrid map position (cR)
D20S103	110		
D20S117	530	3	–
D20S179	200	3	–
D20S105 no signal			
D20S199	353	6	20
D20S113	180	8	21
D20S198	530	9	–
D20S181	100	9	–
D20S193	130	9	26
D20S116	180	11	31
D20S97	430	11	–
D20S95	530	16	–
D20S194	430	18	–
D20S192	940	18	–
D20S115	1630	21	58
D20S175	1950	26	
D20S177	1950	27	
D20S188	2800	31	–
D20S189	140		
D20S186	130	33	–
D20S172	590	35	78
D20S104	1750	38	–
D20S98	710	37	86
D20S114	550	39	87
D20S118	130	39	89
D20S112	430	39	86
D20S182	430	40	
D20S190	180	47	97
D20S180	180	47	
D20S101	300	48	–
D20S184	320	48	101
D20S111	540	49	–
D20S195	630	50	
D20S187	630	50	
D20S191	350	51	–
D20S200	320	50	–
D20S174	1330	54	–
D20S107	1040	55	–
D20S170	410	56	–
D20S99	2000	56	250
D20S108	1600	57	256
D20S96	59	58	–
D20S169	430	58	281
D20S119	49	61	287
D29S197	1010		
D20S178	140	66	310
D20S176	530	71	–
D20S109	140	73	
D20S196	140	74	328
D20S185	130	77	331
D20S854	500		
D20S211	130	79	337
WI-9939	880		–
D20S183	230	79	335
GATA7E09*	790	–	–
WI-6578	500		
AFM22420	500		
D20S913	200	81	–
D20S120	160	82	338
WI-9227	620		
D20S853	620	83	
WI-4228	49	–	337
D20S60	200	–	–
D20S832	500		
D20S469	49	–	–
D20S100	660	83	–
WI-4119	49	–	340
D20S102	2170	86	–
UTR-9681	620		
WI-8810	620		
WI-3773	200	–	340
D20S552	500		
WI-4119	50		
GATAP6309	790		
D20S467	50		
D20S171	130	94	343
D20S173	420	96	346

Total length: 43.5 Mb; minimum tiling path = 39.7 Mb (assuming STSs shown to be adjacent by genetic mapping experiments are on the same fragment when assigned to same sized fraction); total number *Not* I fragments = 67 unique *Not* I fragments (average length = 523 kb); total number of STS = 75.

were used for the STS mapping. The first method used STS primers with the following PCR reaction conditions: 2 μl of molten agarose (∼1 pg template) DNA, 1 mM of each primer, 1× Taq Buffer II (Perkin-Elmer), 160 μg/ml nuclease free bovine serum albumen (Roche), 125 mM dNTP (Pharmacia), 1.5 mM $MgCl_2$, containing DNA and 0.012 U/ml of AmpliTaq (Perkin-Elmer) in a 25 μl final reaction volume. All the PCRs were done in a 96-well U-bottom microtitre plate (Falcon) with heavy mineral oil in the PTC-100 Programmable Thermo Controller (MJ Research Inc.). The PCR cycling conditions were as recommended.

The second approach STS content mapping approach used two sequential PCRs. A degenerate oligonucleotide primed polymerase chain reaction (DOP-PCR) [26] was done first to increase the number of times each slice could be used in a specific PCR experiments. The specific STS PCR was done using 2 μl of the product of the first PCR.

3. Results

The goal of these experiments was to further establish the cloneless top-down for genomic sequencing projects. The experiments described below include methods for the reliable PCR amplification of agarose embedded DNA and a non-specific pre-amplification of template DNA for increasing the cloneless DNA fraction available for analysis. The cloneless libraries was constructed of ∼2 mm slices (aka fractions) of PFG lanes. The uniqueness of DNA contained in the slices was demonstrated by interAlu fingerprinting of individual slices, hybridization of interAlu products to interAlu products from a library of gel slices, FISH analysis, as well as from clone STS content mapping.

3.1. PCR with agarose embedded DNA

Many researchers report inconsistent results with DNA embedded agarose and instead use DNA purified from agarose. In contrast, our past interest in large DNA, lead to the development of simple and reliable methods for purifying and manipulating DNA in agarose. These procedures eased a number of procedures. For instance, simple methods were developed for the purification of intact chromosomal DNA from a wide variety of organism (for summaries see [17,18]). However, variable PCR results were obtained with agarose embedded DNA treated by our standard methods.

Testing of different conditions was done using whole genome DNA embedded in agarose blocks as described in Section 2. The 100 μl agarose block was diluted to 200 μl by the addtion of 100 μl TE buffer. In these experiments, the 200 μl sample melting, then vortexed and a sample removed for the PCR reaction. The 200 μl sample served as the DNA stock. In general, the agarose embedded DNA sample would serve as a template for the first time it was used as a stock. However, subsequent use of this stock led to decreasing yields of product. In these experiments, re-melted of the DNA stock was done at 5–15 min at 95 °C. Hence, it was postulated that the prolonged heating (perhaps some reactive groups from the agarose) was degrading the template DNA.

A number of reagents were tested for their ability to restore maximal PCR yield. For testing purposes, the reaction was divided into two steps, melting of the ''stock'' DNA sample and the PCR reaction itself. The results with ethanolamine, bovine serum albumin (BSA: nuclease free from Roche) and gelatin (Gel) in each steps, independently, is shown in Fig. 1. The best yields through 10 meltings of the agarose embedded template were obtained in the presence of 10–20 mM ethanolamine in the melting step, and 160 μg/ml of BSA in the PCR reaction itself. These conditions were the only ones that led to a significant amount of product after the third use (e.g. melting) of the agarose DNA stock (Fig. 1B). The inclusion of BSA in the PCR reaction itself allowed the production of more product from all samples tested where the DNA had not ''died'' from overcooking. When ethanolamine was present, the initial melting times of 5 and 15 min melting times gave similar results over multiple reuses beyond those shown here. For instance, the same samples of DNA with the same yield of PCR product could be obtained for at least 5–10 reuses of the stock DNA containing ethanolamine. These conditions also work on a number of other DNA samples, PCR primers, and on various DNA samples contained within various concentration of low gelling and low endoosmosis (LE) agarose as would be the case with electrophoretically fractionated DNA (data not shown). In general, there is a broad ethanolamine concentration and a narrow BSA concentration that produces maximal product (Fig. 1C). The presence of ≥20 mM ethanolamine (final concentration) in the stored DNA in TE was necessary to avoid inhibition of PCR from carry over of ethanolamine when large >10 μl amounts of stock DNA was used.

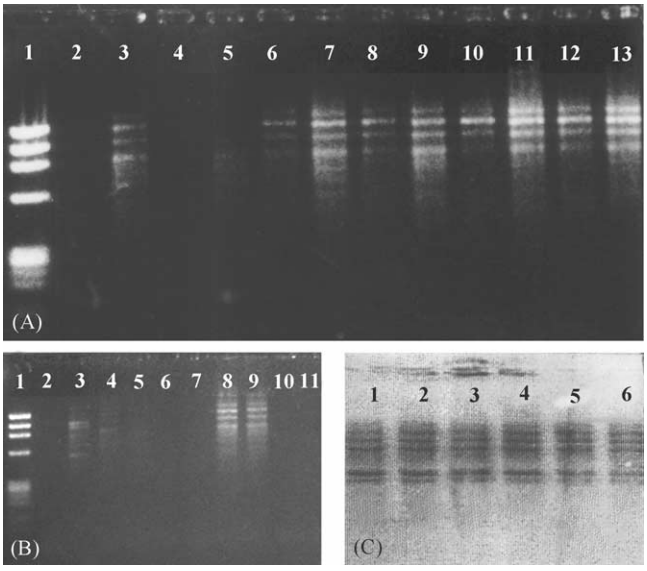


Fig. 1. Reliable and reproducible interAlu PCR directly from agarose embedded A9 neo DNA. These experiments used agarose embedded genomic DNA stored in TE buffer alone [lanes A 2, 3, 4, 5: B6, 7, 10, 11] or in TE containing 100 μg/ml BSA [lanes A6, 7, 8, 9: B2, 3], 20 mM ethanolamine [lanes A10, 11, 12, 13, : B8, 9] or 100 μg/ml gelatin [lanes B4, 5]. The clonesless DNA fractions were heated for 5 [lanes A2, 3, 6,7, 10, 11:B2, 4, 6, 8] or 15 min [lanes A4, 5, 8, 9, 12, 13: B3, 5, 7, 9, 11], vortexed and then 2 μl used in a PCR. DNA template had been melting once before in (A) and (C), and twice before in (B). Hence, the total melting times for the samples in (A) and (C) were 15 min and 45 min in (B). BSA (160 μg/ml) was added to the PCRs in lanes A3, 5, 7, 9, 11, 13 and in all (B) and (C) samples. In (C), ethanolamine was present at 0.5 (lane 1), 1.0 (lane 2), 5 (lane 3), 10 (lane 4), 20 (lane 5), or 40 (lane 6) mM in the DNA embedded DNA. Markers are in A1 and B1.

It should be noted, that inconsistent results were obtained if the DNA-agarose samples were added to the PCR reaction without melting when, for instance, the agarose was diluted enough to allow pipeting without melting (data not shown). In the absence of the ethanolamine, a 15 min melting time appeared to initially produce more PCR product than a 5 min time (data not shown). These types of inconsistencies along with the decreasing yield likely account for the published problems reported for agarose embedded DNA. Also, note that many researchers use stainless steel razor blades rather than glass slides to generate slices. DNA is a strong chelator of iron and iron can damage DNA. In summary, the conditions described here appear to be quite simple and robust and applicable to a wide range of samples.

3.2. Generation of a Not I cloneless library

DNA from a monosomic hybrid cell line containing human chromosome 20 was used for these experiments. First, experiments were carried out to test the suitability of a number of restriction enzymes with infrequently occurring recognite site (Fig. 2). DNA prepared in agarose was cleaved

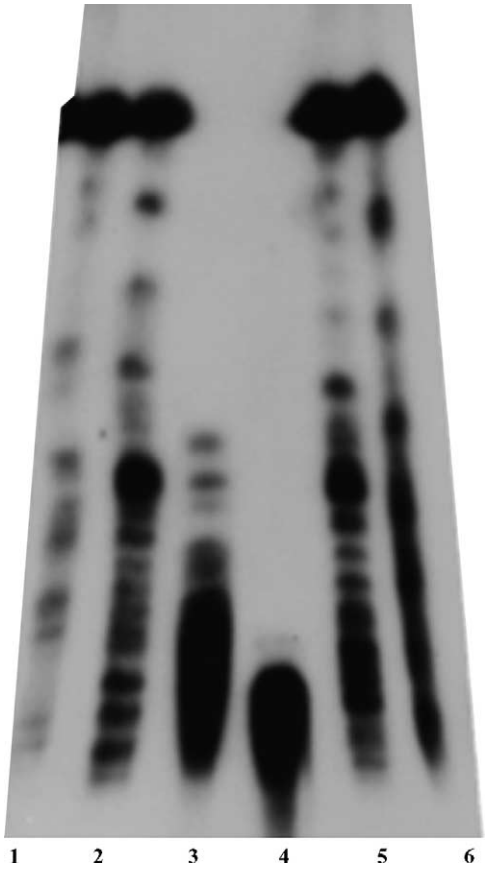


Fig. 2. Distribution of human chromosome 20 restriction fragments. Human hybrid cell line A9-neo DNA was prepared intact in agarose and cleaved with the restriction enzymes. (1) *Asc* I, (2) *Fse* I, (3) *Pme* I, (4) *Pac* I, (5) *Not* I, or (6) *Sgr*A I fractionated by PFG electrophoresis, transfered to nylon membranes and hybridized to the human Alu sequence.

with a variety of enzymes and fractionated by PFG as described previously (see above). In this experiments, the fractionated DNA was transferred to a nylon membrane and hybridized to a human specific Alu probe to visualize the distribution of restriction fragments from chromosome 20. There are ~0.5 million copies of the Alu repeat sequence in the human genome distributed at about 5000 bp intervals; hence, it is expected that most fragments from this chromosome will be visualized by this approach. As expected from earlier experiments on chromosome 21 hybrid cell lines [19], individual *Not* I restriction fragments could be isolated in specific fractions. Further, the lack of smearing between bands indicated the digestion had gone to completion. The restriction enzymes *Asc* I, *Fse* I, *Sgr*A I and *Not* I generated large fragments with, what appeared to be, complete cutting. Use of these enzymes to generate a library of fractions minimizes the confounding effects of partial digest.

Genomic DNA from a monosomic hybrid cell line was used to minimize the number of human fragments of overlapping sizes contained in individual slices. Genomic DNA purified in agarose was cleaved with the restriction enzyme *Not* I and fractionated by PFG using standard conditions reported by us previously. The total size of *Not* I fragments identified, assuming each band contained a single fragment, was 23.4 Mb. The gel was cooled to 10 °C, sliced into ~2 mm pieces placed in 100 μl TE plus 20 mM ethanolamine (final concentration where agarose plus and TE = 200 μl), given a heat treatment to inactivate any contaminated nucleases and stored at 4 or −20 °C. The DNA contained in each slice represents one fraction of the genome. Note that in these experiments only ~1% of the DNA is human.

Three PFG fractionation conditions were used in these experiments (Fig. 3). The conditions were chosen so that the maximal resolution windows were ~0.05–1, ~1–3, or ~3–6 Mb. About 150 slices were generated from each gel. All genomic DNA was on all gels. Hence, a minimum tiling set of 81 fractions or slices with greatest resolution were developed from the complete set of fractions after STS analysis and alignment of the different gels (see below).

During the course of this work, some experiments were done with human chromosome 21 as the robustness of our results (data not shown) could be compared directly to the previously constructed *Not* I restriction map [11]. In all cases, the STS content mapping cloneless library of chromosome 21, similar to that described here for chromosome 20, were in agreement with the previously constructed map. The data that will be described here will focus on the chromosome 20 results.

3.3. InterAlu fingerprinting and analysis of cloneless fractions

An issue when dealing with electrophoretic fractions is whether there is ‘‘bleeding’’ of DNA between different fractions. This was initially tested by interAlu fingerprinting

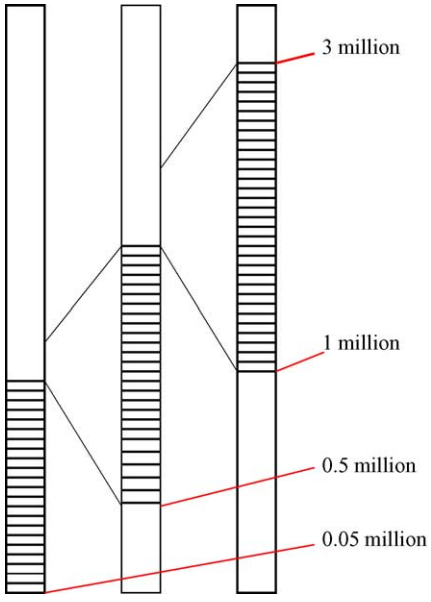


Fig. 3. Alignment of different PFG resolution conditions. The cloneless library was generated from the minimal high resolution slices (fractions) from three gels plus a group of slices extended to the well.

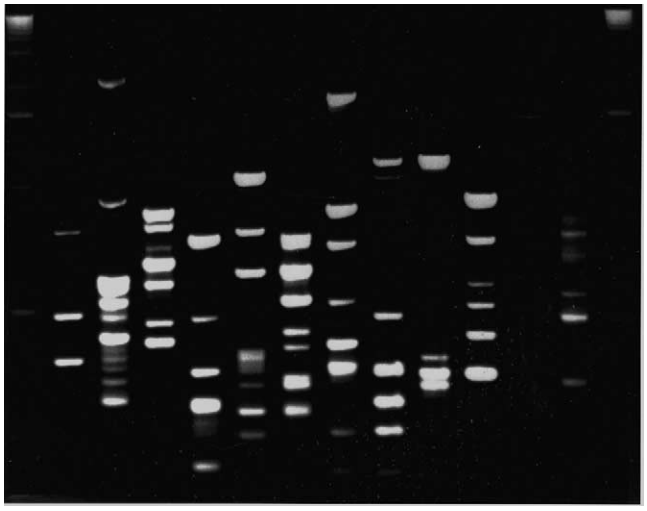


Fig. 4. InterAlu fingerprinting of *Not* I cloneless library fractions. Adjacent gel slices or *Not* I fractions was subjected to interAlu PCR and the products were fractionated by conventional agarose electrophoresis.

of individual slices (Fig. 4). The data shows that unique interAlu fragments (fingerprints) were generated from adjacent fractions. This was not surprising as similar results were obtained in past experiments focused on chromosome 21 [14]. In the past experiments, the interAlu products were used for gap filling with *Not* I fragments having no known STS. In subsequent experiments, interAlu PCR fingerprinting was used as a quality control step and to help align different gel runs.

In another set of hybridization experiments, interAlu PCR products were used to overlap different cloneless and clone libraries (Fig. 5). In these experiments, interAlu products from (a) *Sgr*A I and *Fse* I cloneless fractions (b) YAC or (c) hncDNAs (generated as described in [27]) were arrayed onto nylon membranes and probed with labeled interAlu products generated from the cloneless library. Here, the interAlu products were used so that the experimental sample complexity was reduced and the hybridization signal stronger. The results show that specific fractions were identified that overlapped the *Not* I fragment.

Also, the specificity of the DNA contained in the same *Not* I cloneless library fraction was tested by FISH analysis (Fig. 6). Since most of the DNA in each fraction is mouse DNA, these experiments used the products of an interAlu PCR as a probe to metaphase chromosomes. The library fraction used in these experiments was assigned to 20q13 (a region amplified in some breast and other cancers (e.g. [28–31]) by STS content mapping (see below)). The results show the specificity of the signal in the FISH experiment matched the STS result.

3.4. STS specific PCR on cloneless fractions

The uniqueness of DNA in gel slices was tested by performing STS PCR. Typical results from adjacent gel slices are shown in Fig. 7. In this experiment, two STS

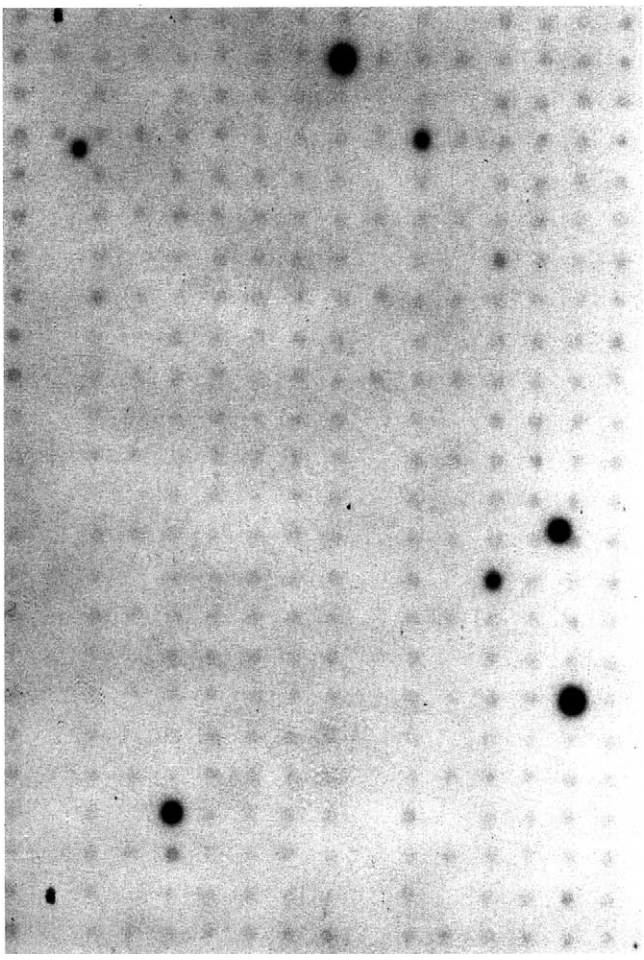


Fig. 5. Hybridization of ^{32}P-labeled interAlu PCR products from cloneless gel fractions to arrayed interAlu products from a clone library. The same interAlu primer was used to generate interAlu products from *Not* I and *Sgr*A I cloneless fraction containing the 20q13.2 region that was labeled with ^{32}P labeled and to interAlu products from a clone library arrayed onto a nylon membrane.

produced product from a single slice, while one STS produced product from adjacent slices.

Experiments were also done to extend the number of PCR reactions that could be done on each slice. In these experiments, the template contained in each fraction was amplified in a nonspecific PCR reaction; then, STS specific PCR was done using the products of the first PCR as template. In the experiment shown in Fig. 8, 8/10 STS gave positive results using the pre-amplification step. This approach increased the number of specific PCRs that could be done by ~1000-fold. Previous experiments mapping the *S. pombe* genome used a different random PCR primers to generation hybridization probes from PCR slices [15,32].

3.5. Ordering of a Not I fractions by STS content mapping

A tiered pooling strategy was used to minimize the number of PCR reactions needed to assign each STS to a

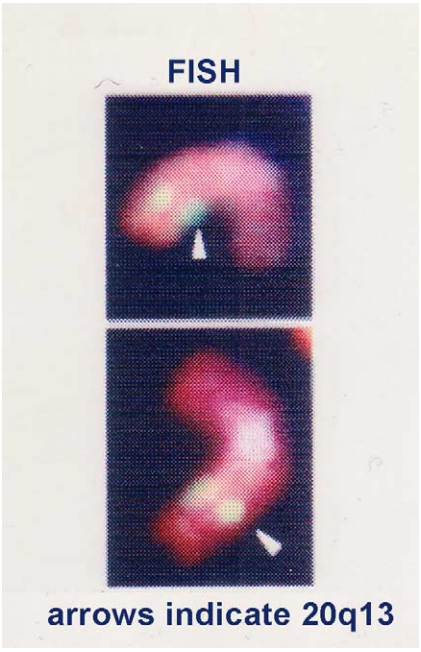


Fig. 6. FISH analysis of long interAlu products. The interAlu products were from the *Not* I fraction containing STS D20S120. The probe hybridized to region 20q13 (indicated with an arrow), with some minor hybridization elsewhere, confirming the STS content mapping.

fraction (Fig. 9). Although the chosen pooling strategy may not be the most efficient, the number of PCR reactions need for an STS assignment was reduced from 81 to 12. The first tier PCR used DNA template from four superpools fractions created from the ∼3–6 Mb gel that corresponded to <1 Mb, ∼1–3 Mb, ∼3–7 Mb, and >7 Mb. Each STS was assigned to one size range. If two pools were positive the STS was assigned to the pools containing the smallest fragments. Each STS was tested in a second PCR using three pools

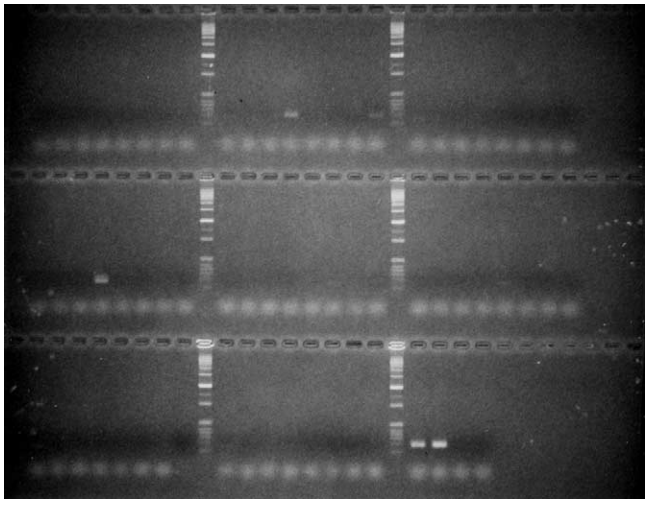


Fig. 7. PCR results with three STSs using *Not* I fractions from a cloneless library. Adjacent gel slices were analysis by PCR for STS content of three different loci.

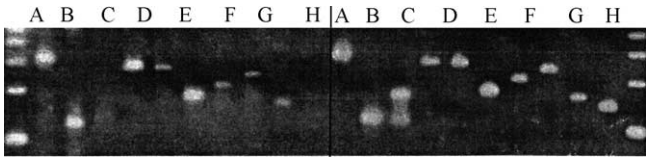


Fig. 8. STS content mapping using a nonspecific preamplification step. Here, the yield of STS products using template from a DOP-PCR (left) or a *Not* I fraction directly (right) can be compared. The end lanes contain marker DNA.

constituted from fractions making up the positive first tier pool. The smallest fragment that contained each STS was established by continue to divide each positive pool into thirds (Table 1 and Figs. 9 and 10). This approach reduced the number of fragments tested by two-thirds in each round. The total length of the assigned *Not* I fragments was 39.7 Mb. This figures assumes that some STS (genetically adjacent and located to the same size *Not* I fragment) are located on the same *Not* I fragment. These experiments located ∼63% of the *Not* I cleaved DNA of chromosome 20 (chromosome size = ∼63.6 Mb). (http://www.ncbi.nih.gov/genome/seq/ (September 2003)). No STSs were assigned to two different regions and only one STS produced no signal. No STS was assigned to a fraction containing fragments >7 Mb. The total *Not* I fragment size in the Alu hybridization experiments was 23.4 Mb (as in Fig. 2). This number is in remarkable close agreement with the actual total seen in the experiments (see above).

Multiple cloneless libraries can be generated from additional PFG fractionations. Hence, it was important to determine the reproducibility of the slicing protocol and STS assignment. The reproducibility was tested by comparing the STS assignments on several cloneless libraries generated by different individuals at different time. The results showed that there was remarkable agreement between the cloneless libraries generated at different times (data not shown).

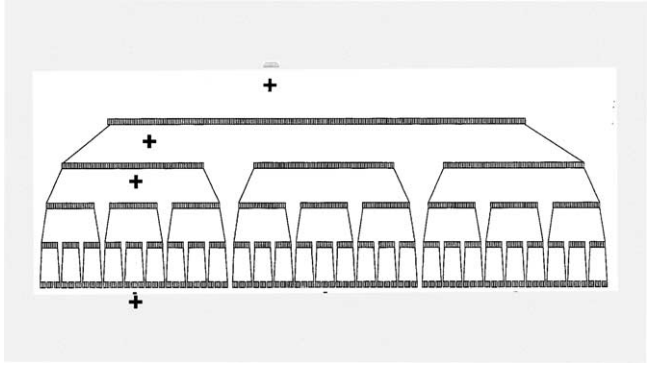
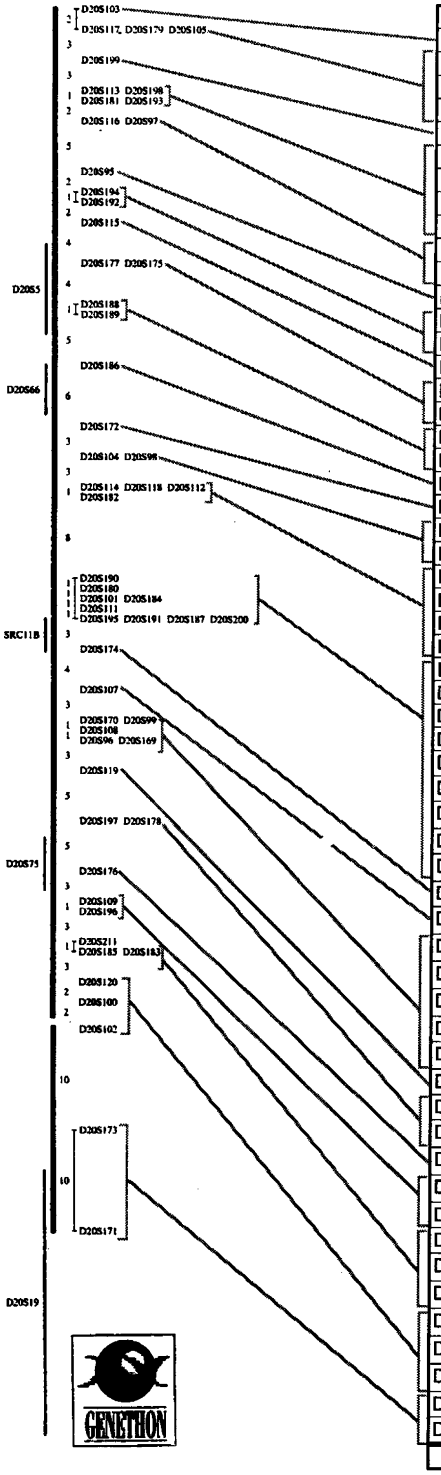


Fig. 9. Scheme for STS content mapping of a *Not* I fragment cloneless library of human chromosome 20. First, STS were mapped to superpools corresponding to maximum resolution windows of >1 Mb, 1–3 Mb or >3 Mb. The pools are successively broken down by thirds until each STS was assigned to a specific fragment.



STS	1st level	2nd level	3rd level	4th level	5th level	NS	Assigned
D20S103	<1000	<279	<120	96-120	110		110
D20S117	1000-3000	431-1496	431-885	431-530	530		530
D20S179	<1000	<279	120-196	177-210	196		196
D20S105	NS						NS
D20S199	<1000	279-561	279-406	295-345	345		345
D20S113	<1000	<279	120-279	177-210	177		177
D20S198	1000-3000	431-1496	431-885	431-530	530		530
D20S181	<1000	<279	<120	96-120	96		96
D20S193	<1000	<279	120-196	130-160	130		130
D20S116	<1000	<279	120-196	177-210	177		177
D20S97	1000-3000	431-1496	431-885	431-530	431		431
D20S95	1000-3000	431-1496	431-885	431-530	530		530
D20S194	1000-3000	431-1496	431-885	431-530	431		431
D20S192	<1000	431-1496	942-1130	942-978	942		942
D20S115	1000-3000	1538-1910	1496-1658	1632-1658	1632		1632
D20S177	1000-3000	1538-1910	1684-1810	1788-1810	1930-2126	1950	1950
D20S175	3000-6000	1930-3130	1930-2126				
D20S188	<1000	279-561	1930-3130	2467-3130			
D20S189	<1000	<279					NS
D20S186	<1000	<279	<120	130			130
D20S172	<1000	279-561	571-613	590			590
D20S104	3000-6000	1538-1910	1684-1810				
D20S98	<1000	279-561	NS	680-743	NS		NS
D20S114	<1000	279-561	524-561	551			551
D20S118	<1000	<279	120-196	130-160	130		130
D20S112	<1000	279-561	406-513	431-471	431		431
D20S182	1000-3000	431-1496	431-885	431-530	431		431
D20S190	<1000	<279	120-196	177-210	177		177
D20S180	<1000	<279	120-196	177-210	177		177
D20S101	<1000	279-561	279-406	295-345	295		295
D20S184	<1000	279-561	279-406	295-406	324		324
D20S111	<1000	279-561	513-561	NS			NS
D20S195	1000-3000	431-1496	431-885	630-729	630		630
D20S191	<1000	279-561	295-406				NS
D20S187	1000-3000	431-1496	431-885	431-530	630		630
D20S200	<1000	279-561	295-406	324			324
D20S174	1000-3000	431-1496	1210-1454	NS			NS
D20S107	1000-3000	>3130	530-1486	942-1130			
D20S170	<1000	279-561	406-513	295-406	406		406
D20S99	1000-3000	431-1496	1930-2126	1950			1950
D20S108	1000-3000	1538-1910	1496-1658	1496-1538	1606		1606
D20S96	<1000	<279	<120	<59	59		59
D20S169	<1000	279-561	406-513	431			431
D20S119	<1000	<279	<120	<59	<48		<48
D20S197	1000-3000	431-1496	942-1130	1014-1050	1014		1014
D20S178	<1000	<279	120-196	130-210	144		144
D20S176	<1000	<279	<120	431-513	529		529
D20S109	<1000	<279	<120	59	130-210	144	144
D20S196	<1000	<279	<120	<59	NS	144	144
D20S211	<1000	<279	120-196	130-160	130		130
D20S185	<1000	<279	120-196	130			130
D20S183	<1000	<279	196-279	225			225
D20S120	<1000	<279	120-196	160			160
D20S100	3000-6000	530-1486	431-885				
D20S102	1000-3000	431-1496	2170-2350	2170			2170
D20S173	<1000	279-561					NS
D20S171	<1000	<279	120-196	130-160	130		130
						45 STSs	22447 kb
						corrected (without ■)	20611 kb

Fig. 10. Summary of STS analysis of a chromosome *Not* I cloneless library. This table tabulates the results obtained with 56 STS content at different tiers as described in Fig. 8 and text.

4. Conclusions

There are a number of reports on the usefulness of DNA directly from agarose. Most involve removal of agarose because of the perceived presence of inhibitors for enzyme manipulations (e.g. [33–37]) that in other cases are not seen (e.g. [38]). In some cases, enhanced sequencing results have been reported for reaction products embedded in agarose [39] and direct sequencing of agarose embedded DNA has been reported [40].

However, besides our own work few researchers choose to analyze DNA in agarose. Most other efforts are focused

on genomic analysis of parasitic organisms of intermediate genome size [1,3,41] because these chromosomes do not condense during cell division and cannot be microscopically visualized. Today's technology allows sequencing of small bacterial genomes in a bottom up approach that avoids the ordered clone libraries because of the expense of creating them. However, efficient ordering strategies, especially those that can be automated could reduce sequencing of small genomes. In contrast, sequencing of larger genomes requires a top down approach where chromosomes are broken into manageable parts.

There is increasing interest in developing inexpensive and efficient approach [42] that allows sequencing of a large number of different genomes. Several researchers are approaches this problem by avoiding clone libraries and instead sequencing individual DNA molecules. Thus far, those approaches appear to be particularly useful for resequencing experiments. In the future decreasing costs in sequencing should enable do novo sequencing.

The most likely application of the cloneless library approach is to reduce the cost of de novo sequencing for new organisms especially with large genomes where no ordered genomic library is available. Also, cloneless library fractions can be cleaved with restriction enzymes for a second size fractionation (and a second dimension cloneless library), fingerprinted, used as hybridization probes, used to create clone libraries, screened by PCR, and even sequenced. Hence, a totally in vitro top-down genomic sequencing project can be envisioned. For instance, low complexity agarose embedded DNA could be digested to sequencing size fragments, ligated to oligonucleotides of known sequences for attachment to beads in a ''polonies'' approach that analyzes PCR generated foci in acrylamide gels [43].

Acknowledgements

CLS would like to thank Jean Weissenbach and Christine Petit for their hospitality at the Institute Pasteur where this work was begun some time ago. Thanks also to David Maas for helping with some of the experiments and Joe Grey and Mary Hintz for testing the specificity of the interAlu products by FISH, and Eugene Sverdlov for providing some of the chromosome 20 clones. This work was support by DOA AIBS2154 and NIH Score HL5501.

References

[1] Foster JM, Kumar S, Ganatra MB, Kamal IH, Ware J, Ingram J, et al. Construction of bacterial artificial chromosome libraries from the parasitic nematode Brugia malayi and physical mapping of the genome of its Wolbachia endosymbiont. Int J Parasitol 2004;34(6): 733–46.

[2] Khorasani MZ, Hennig S, Imre G, Asakawa S, Palczewski S, Berger A, et al. A first generation physical map of the medaka genome in BACs essential for positional cloning and clone-by-clone based genomic sequencing. Mech Dev 2004;121(7–8):903–13.

[3] Myler PJ, Stuart KD. Recent developments from the Leishmania genome project. Curr Opin Microbiol 2000;3(4):412–46.

[4] Schalkwyk LC, Cusack B, Dunkel I, Hopp M, Kramer M, Palczewski S, et al. Advanced integrated mouse YAC map including BAC framework. Genome Res 2001;11(12):2142–50.

[5] Wu CC, Nimmakayala P, Santos FA, Springman R, Scheuring C, Meksem K, et al. Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. Theor Appl Genet 2004;109(5):1041–50.

[6] Wu C, Sun S, Nimmakayala P, Santos FA, Meksem K, Springman R, et al. A BAC- and BIBAC-based physical map of the soybean genome. Genome Res 2004;14(2):319–26.

[7] Sidjanin DJ, Miller B, Kijas J, McElwee J, Pillardy J, Malek J, et al. Radiation hybrid map, physical map, and low-pass genomic sequence of the canine prcd region on CFA9 and comparative mapping with the syntenic region on human chromosome 17. Genomics 2003;81(2): 138–48.

[8] Wong GK, Yu J, Thayer EC, Olson MV. Multiple-complete-digest restriction fragment mapping: generating sequence-ready maps for large-scale DNA sequencing. Proc Natl Acad Sci USA 1997;94(10): 5225–30.

[9] Schein J, Kucaba T, Sekhon M, Smailus D, Waterston R, Marra M. High-throughput BAC fingerprinting. Methods Mol Biol 2004;255: 143–56.

[10] Locke J, Rairdan G, McDermid H, Nash D, Pilgrim D, Bell J, et al. Cross-screening: a new method to assemble clones rapidly and unambiguously into contigs. Genome Res 1996;6(2):155–65.

[11] Wang D, Smith CL. Large-scale structure conservation along the entire long arm of human chromosome 21. Genomics 1994;20(3):441–51.

[12] Bukanov N, Wang D, Smith CL. Physical maps of human chromosomes. In: Meyers RA, editor. Encyclopedia of Molecular Biology and Molecular Medicine, 4. New York: VCH; 1996. p. 215–21.

[13] Smith CL, Wang D, Broude N, Bukanov N, Monastyrskaya GS, Sverdlov E. Parallel processing in genome mapping and sequencing. Methods 1996;9(1):136–44.

[14] Wang D, Zhu Y, Smith CL. A set of inter-Alu PCR markers for chromosome 21 generated from pulsed-field gel-fractionated NotI restriction fragments. Genomics 1995;26(2):318–26.

[15] Grothues D, Cantor CR, Smith CL. Top-down construction of an ordered *Schizosaccharomyces pombe* cosmid library. Proc Natl Acad Sci USA 1994;91(10):4461–5.

[16] Koi M, Shimizu M, Morita H, Yamada H, Oshimura M. Construction of mouse A9 clones containing a single human chromosome tagged with neomycin-resistance gene via microcell fusion. Jpn J Cancer Res 1989;80(5):413–8.

[17] Smith CL, Klco S, Cantor CR. Pulsed field gel electrophoresis and the technology of large DNA molecules. In: Davies K, editor. Genome Analysis: A Practical Approach. Oxford, England: IRL Press; 1988 p. 41–72.

[18] Smith CL, Klco S, Zhang TY, Fang H, Oliva R, Fan JB, et al. Methods in molecular genetics. In: Adolph KW, editor. Methods in Molecular Medicine. San Diego: Academic Press; 1993. p. 155–96.

[19] Sainz J, Pevny L, Wu Y, Cantor CR, Smith CL. Distribution of interspersed repeats (Alu and Kpn) on *Not* I restriction fragments of human chromosome 21. Proc Natl Acad Sci USA 1992;89(3): 1080–4.

[20] Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, et al. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci USA 1989;86(17):6686–90.

[21] Sambrook J, Gething MJ. Protein structure. Chaperones, paperones. Nature 1989;342(6247):224–5.

[22] Kallioniemi A, Kallioniemi OP, Piper J, Tanner M, Stokke T, Chen L, et al. Detection and mapping of amplified DNA sequences in breast

cancer by comparative genomic hybridization. Proc Natl Acad Sci USA 1994;91(6):2156–60.
[23] Stokke T, Collins C, Kuo WL, Kowbel D, Shadravan F, Tanner M, et al. A physical map of chromosome 20 established using fluorescence in situ hybridization and digital image analysis. Genomics 1995;26(1):134–7.
[24] Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, et al. The 1993–1994 Genethon human genetic linkage map. Nat Genet 1994;7(2 Spec No):246–339.
[25] Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, et al. An STS-based map of the human genome. Science 1995;270(5244): 1945–54.
[26] Cheung VG, Nelson SF. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. Proc Natl Acad Sci USA 1996;93(25):14676–9.
[27] Borodin A, Kopatnzev E, Wagner L, Volik S, Ermolaeva O, Lebedev Y, et al. An arrayed library enriched in hncDNA corresponding to transcribed sequences of human chromosome 19: preparation and analysis. Genet Anal 1995;12(1):23–31.
[28] Tanner MM, Tirkkonen M, Kallioniemi A, Collins C, Stokke T, Karhu R, et al. Increased copy number at 20q13 in breast cancer: defining the critical region and exclusion of candidate genes. Cancer Res 1994;54(16):4257–60.
[29] Tanner MM, Tirkkonen M, Kallioniemi A, Isola J, Kuukasjarvi T, Collins C, et al. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. Cancer Res 1996;56(15):3441–5.
[30] Tirkkonen M, Tanner M, Karhu R, Kallioniemi A, Isola J, Kallioniemi OP. Molecular cytogenetics of primary breast cancer by CGH. Genes Chromosomes Cancer 1998;21(3):177–84.
[31] Rennstam K, Ahlstedt-Soini M, Baldetorp B, Bendahl PO, Borg A, Karhu R, et al. Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. Cancer Res 2003;63(24):8861–8.
[32] Grothues D, Cantor CR, Smith CL. PCR amplification of megabase DNA with tagged random primers (T-PCR). Nucleic Acids Res 1993;21(5):1321–2.
[33] Brosius III FC, Holzman LB, Cao X. Purification of PCR products from agarose gels for direct sequencing. Methods Mol Biol 1996;65:11–21.
[34] Denton PH, Cullen JB, Loeb D, Lucas A, Nunes K, Hulett C, et al. Partitioned pulsed-field gel electrophoresis-PCR (PPF-PCR): a new method for pulsed-field mapping for STS and microsatellites. Nucleic Acids Res 1994;22(9):1776–7.
[35] Watanabe M. Rapid and inexpensive recovery method of DNA fragments from agarose and polyacrylamide gels by a cotton-wool column tube. Nucleic Acids Symp Ser 1999;(42):101–2.
[36] Leonard JT, Grace MB, Buzard GS, Mullen MJ, Barbagallo CB. Preparation of PCR products for DNA sequencing. Biotechniques 1998;24(2):314–7.
[37] Yamaguchi Y, Nimbari S, Ookawara T, Oishi K, Eguchi H, Suzuki K, et al. Inhibitory effects of agarose gel and LB medium on DNA sequencing. Biotechniques 2002;33(2):282, 284.
[38] Setterquist RA, Smith GK. Ready to use agarose encapsulated PCR reagents. Nucleic Acids Res 1996;24(8):1580–1.
[39] Vatcher G, Smailus D, Krzywinski M, Guin R, Stott J, Tsai M, et al. Resuspension of DNA sequencing reaction products in agarose increases sequence quality on an automated sequencer. Biotechniques 2002;33(3):532–4, 536, 538–9.
[40] Khorana S, Gagel RF, Cote GJ. Direct sequencing of PCR products in agarose gel slices. Nucleic Acids Res 1994;22(16):3425–6.
[41] Leech V, Quail MA, Melville SE. Separation, digestion, and cloning of intact parasite chromosomes embedded in agarose. Methods Mol Biol 2004;270:335–52.
[42] Shendure J, Mitra RD, Varma C, Church GM. Advanced sequencing technologies: methods and goals. Nat Rev Genet 2004;5(5): 335–44.
[43] Mitra RD, Shendure J, Olejnik J, Edyta Krzymanska O, Church GM. Fluorescent in situ sequencing on polymerase colonies. Anal Biochem 2003;320(1):55–65.