

Vector-hexamer PCR isolation of AII insert ends from a YAC contig of the mouse Igh locus

Christopher D. Herring, Christophe Chevillard, Sean L. Johnston, Peter J. Wettstein, Roy Riblet

► **To cite this version:**

Christopher D. Herring, Christophe Chevillard, Sean L. Johnston, Peter J. Wettstein, Roy Riblet. Vector-hexamer PCR isolation of AII insert ends from a YAC contig of the mouse Igh locus. Genome Research, Cold Spring Harbor Laboratory Press, 1998, 8 (6), pp.673-681. <hal-01593096>

HAL Id: hal-01593096

<https://hal-amu.archives-ouvertes.fr/hal-01593096>

Submitted on 6 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Vector–Hexamer PCR Isolation of All Insert Ends from a YAC Contig of the Mouse *Igh* Locus

Christopher D. Herring,¹ Christophe Chevillard,¹ Sean L. Johnston,²
Peter J. Wettstein,³ and Roy Riblet^{1,4}

¹Medical Biology Institute, La Jolla, California 92037 USA; ²Department of Immunology and ³Departments of Surgery and Immunology, The Mayo Foundation, Rochester, Minnesota 55905 USA

We have developed a simple PCR strategy, termed vector–hexamer PCR, that is unique in its ability to easily recover every insert end from large insert clones in YAC and BAC vectors. We used this method to amplify and isolate all insert ends from a YAC contig covering the mouse *Igh* locus. Seventy-seven ends were amplified and sequenced from 36 YAC clones from four libraries in the pYAC4 vector. Unexpectedly, 40% of the insert ends of these YACs were LINE1 repeats. Nonrepetitive ends were suitable for use as probes on Southern blots of digested YACs to identify overlaps and construct a contig. The same strategy was used successfully to amplify insert ends from YACs in the pRML vector from the Whitehead Institute/MIT-820 mouse YAC library and from BACs in pBeloBAC11. The simplicity of this technique and its ability to isolate every end from large insert clones are of great utility in genomic investigation.

[The nucleotide sequence data reported in this paper are accessible in GenBank under accession nos. B07512–B07598.]

One of the primary goals of genome research is the production of genetic and physical maps of whole genomes. Such maps are a necessary resource for the identification of all functional genes and the localization of specific genes associated with disease. The main tools to build physical maps are large genomic insert libraries in yeast and bacterial artificial chromosome vectors (YACs and BACs, respectively) (Burke et al. 1987; Shizuya et al. 1992). Such libraries consist of multiple genome equivalents in tens of thousands of clones.

To begin the construction of the physical map of a genomic region YAC and BAC libraries are screened to identify clones carrying markers mapped to the region. This can be accomplished by either hybridization on densely gridded filters or PCR on hierarchical or multidimensional pools (Green and Olson 1990). Next, both ends of each clone are isolated for use as hybridization probes and sequenced to design new PCR markers. These probes and PCR markers are used to assess the overlap of the clones and to “walk” if needed, that is, to

rescreen the libraries to isolate additional clones that overlap the end and extend the developing contig. The walking strategy is effective but time consuming, making the choice of end isolation strategy critical. The method should be simple, rapid, and completely efficient, successful at every end.

Early methods to isolate insert ends involved subcloning into phage or plasmid vectors and screening subclone libraries with YAC vector arms to identify the junction fragments. Simpler, faster PCR-based methods are more generally used, including IRS PCR, inverse PCR, and vectorette PCR. Inter repeat sequence (IRS)–PCR takes advantage of the frequent occurrence of repeat elements in the genome (*Alu* for human and B1 for mouse). PCR amplification is performed between *Alu* (or B1) consensus sequences and the YAC vector sequence (Nelson et al. 1989; Breukel et al. 1990). However, it fails in many cases because it relies on the presence of a repeat element within a short distance from the end. Another problem is the generation of extraneous non-end products from amplification between repeats. Inverse PCR relies on restriction enzyme cleavage at sites within the vector and the genomic insert (Triglia et al. 1988; Silverman

⁴Corresponding author.
E-MAIL riblet@cerfnet.com; FAX (619) 554-0684.

et al. 1989; Silverman 1993). The fragments produced are then self-ligated and the end is amplified using divergent vector primers. Although conceptually simple, this strategy is limited to the use of a small number of restriction enzymes with appropriate YAC arm sites, and frequently it failed in our hands. Vectorette PCR involves the ligation of a "bubble" linker to digested total YAC DNA and subsequent recovery of the insert end by amplification between this linker primer and nested YAC arm primers (Riley et al. 1990). This is a very effective strategy but somewhat complex in operation.

We have developed a method for the isolation of insert ends from YACs that uses PCR exclusively, involves a minimum number of manipulations, and does not require adapters, linkers, or ligations. On the basis of restriction-site PCR (Sarkar et al. 1993; Johnston et al. 1995), and drawing on degenerate oligonucleotide primed (DOP)-PCR (Wesley et al. 1990; Telenius et al. 1992), this generalized strategy, called vector-hexamer PCR, uses hexamer primers designed by adding a 6-bp sequence to the 3' end of an arbitrary "anchor" sequence. Three rounds of PCR are performed using a nested series of vector arm primers in conjunction with one hexamer primer. It is similar to the DOP-vector PCR (Wesley et al. 1994; Wu et al. 1996) and thermal asymmetric interlaced (TAIL)-PCR (Liu and Whittier 1995) methods described for cosmid, P1, YAC, and BAC end isolation but is more general in that it incorporates a variety of alternative primers and is less complex than either method. It is completely effective, recovering every end that we have attempted.

In this paper we demonstrate the use of vector-hexamer PCR for the isolation of insert ends from BACs and two varieties of YAC vector clones from the mouse immunoglobulin heavy chain (*Igh*) locus located near the telomere of chromosome 12. Whereas the human IGH locus has been cloned recently in YACs and characterized (Cook et al. 1994), knowledge of the organization of the mouse locus is still based mainly on genetic and deletion mapping (Tutter et al. 1991; Mainville et al. 1996). It is comprised of >100 gene segments spread over more than a centiMorgan of recombination distance. To investigate this complex locus we identified a series of YACs containing all *Igh* gene elements. To assemble a contig the insert ends were isolated from all 36 YAC clones by vector-hexamer PCR to characterize YAC overlap and to detect chimeric clones. A description of this simple, rapid, and efficient method is presented here.

RESULTS AND DISCUSSION

The vector-hexamer technique to recover clone insert ends uses a nested series of vector arm primers in conjunction with a hexamer primer that anneals to multiple sites randomly located in the clone insert (Fig. 1a). The primers are shown in Table 1. The hexamer primers correspond to a variety of restriction enzyme recognition sites and frequently occurring sequences; in principle, the hexamer can be any useful sequence. For these short sequences to act as primers, five cycles of low annealing temperature (35°C) PCR are used to begin the amplification (Fig. 1b). In addition to the hexamer sequence, the hexamer primers contain an anchor sequence at the 5' end; for most of them this is the T7 promoter sequence. The incorporation of the anchor sequence into the products from the first five cycles of PCR allows the annealing temperature of subsequent cycles of PCR to be increased to 60°C (Fig. 1c-e).

In the pYAC4 vector the cloning site is in the *SUP4* nonsense suppressor tRNA gene (Fig. 1a). Because there are related tRNA genes in the yeast host, primers that are designed in the area immediately adjacent to the cloning site amplify host sequences along with the YAC insert end. In early experiments, when the initial primer was L-II or R-II, yeast background bands were isolated and sequenced in 5 of 21 PCR products. When L-I and R-I primers, located outside of the *SUP4* gene (Fig. 1a) were used for the first round of PCR, yeast products were incorrectly recovered in 6 of 87 cases. Additional measures to minimize and identify amplification of yeast background include amplifying and analyzing many YACs at once with the same hexamer primer to identify common yeast fragments and using an "empty YAC" negative control (AB1380 transformed with linearized pYAC4).

First round PCR products (Figs. 1c and 2a) typically show either a pattern of yeast bands present in every YAC or nothing at all. Second-round, nested PCR products (Figs. 1d and 2b) typically consist of faint YAC-specific end products among persistent yeast bands. These products can be isolated and sequenced, but not reliably. Third round, further nested, amplification (Figs. 1e and 2c) is necessary to generate enough insert end product to sequence routinely. These PCR products consist of several YAC-specific bands between ~100 and 2000 bp, often with one or two very intense bands (estimated 200–2000 ng). As seen in Figure 2, c and d, yeast and vector bands are often present in the third-round product but are identified by their presence in mul-

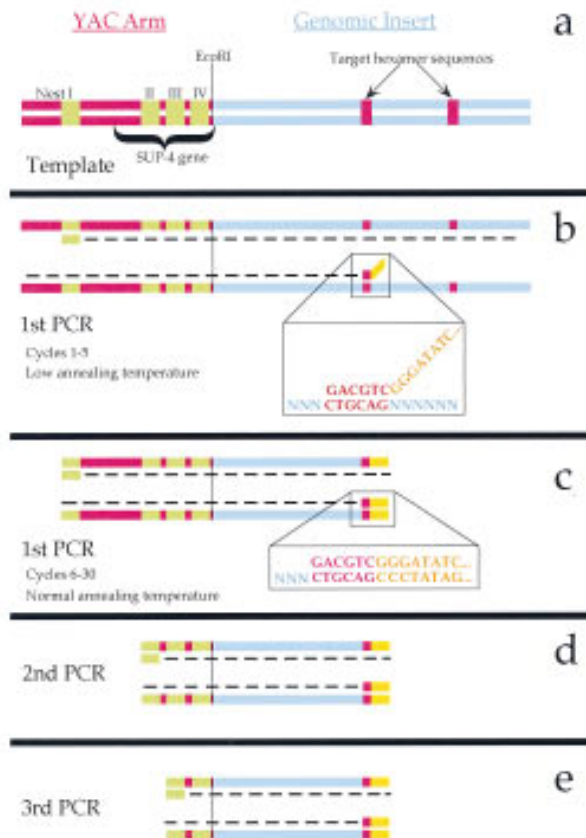


Figure 1 Schematic representation of vector-hexamer PCR. (a) The YAC vector/insert junction (right or left arm) before amplification. The YAC vector is shown in red (*left*) with nested vector priming sites in green. The first nested priming site is located outside the *SUP4* gene. The mouse genomic insert is shown in blue (*right*) with two particular hexamer sequences shown in red. (b) The initial cycles of amplification at low annealing temperature. The vector primer (in green) anneals to its complement in the YAC arm, from which a new strand is synthesized (broken line). The hexamer primer is shown annealing to a target site in the mouse insert, its hexamer portion shown in red and its 20-bp anchor sequence in orange. The low annealing temperature (35°C) allows pairing and priming by the 3' hexamer alone (shown in expanded view). (c) The later cycles of the first round of PCR are performed at a higher annealing temperature so that the entire hexamer primer stably anneals only to products from the initial cycles. (d,e) The second and third rounds of PCR reamplify the vector/insert junction products from the first PCR with nested vector primers to increase the specificity and yield of the amplification. The fourth nested vector primer allows reduction of vector arm contribution to the PCR product for use as a hybridization probe.

multiple lanes and the control. The importance of multiple hexamer primers is illustrated in Figure 2d. The

YAC RAD52-y99A9 yields a difficultly large, possibly yeast product with the *BanI* primer (left), apparently only yeast-derived products with the *PstI* primer (center), but a useful 200-bp product with the *SduI* primer (right).

Vector-hexamer PCR was used to amplify and isolate 77 ends from 36 *Igh* region YACs using 9 different hexamer primers. More than two ends were isolated from some clones because five of them contained more than one YAC. Table 2 shows the size and intensity of the best product from each vector-hexamer amplification that was performed. Also listed are the GenBank accession numbers for each of the 77 insert end sequences. Useful, medium to strong intensity end products were readily obtained, usually with several hexamer primers. In two instances five, and even six, restriction site hexamer primers failed, but in these cases use of the two DOP primers succeeded.

Apparently two DOP primers were the most uniformly successful in comparing the different hexamer primers; however, these were incorporated in the primer panel after the technique had been refined and have been less extensively tested. Highly successful primers include *DdeI*+G, *PstI*, and *SduI*. The efficiency of the *PvuII* primer differed greatly between the left and the right arm in this study.

We were able to examine the annealing specificity of hexamer primers in cases in which two separate end products of different lengths were sequenced from the same YAC end. As seen in Table 3, almost every example shows at least one mismatch from the hexamer sequence, usually in the 5' half. This result was reproducible—in cases in which two different length end products were generated from the same vector-hexamer reaction, the smaller fragment could be regenerated by reamplifying the gel-purified larger fragment, even without low annealing temperature cycles.

Sequence Analysis

The PCR insert end products were sequenced with the third or fourth nested vector arm primer. Sometimes the hexamer primer was used as well to sequence the reverse strand. For many ends several products made with different hexamer primers were sequenced. All end sequences were submitted to the BLAST e-mail server (Altschul et al. 1990). Of the 77 ends, 12 were similar to known sequences from the *Igh* locus, usually sequences flanking coding regions. Known genes other than *Igh* or repeat elements were identified by five YAC ends, but they

Table 1. Primers

		Annealing temp.	Ref.
pYAC4 Left Arm: (containing CEN4, ARS1, and TRP1)			
L-I	CACCCGTTCTCGGAGCACTGTCCGACCGC	60°C	a
L-II	TATACTCTTTCTTCAACAAT	50°C	
L-III	ACTCTCGGTAGCCAAGTTGG	55°C	
L-IV	GGCGCAAGACTTTAATTTAT	55°C	
pYAC4 Right Arm: (containing URA3)			
R-I	ATATAGGCGCCAGCAACCCGACCTGTGGCG	60°C	a
R-II	GTCTGGGAAGTGAATGGAGA	60°C	
R-III	AAAACAAAAAATCTCCGG	55°C	
R-IV	AGTCGAACGCCGATCTCAA	55°C	
Hexamer Primers			
	[<----- T7 ----->] [-RS-]		
BarI	TAATACGACTCACTATAGGGGGYRCC		b
BarIII	TAATACGACTCACTATAGGGATCGAT		b
DdeI+G	TAATACGACTCACTATAGGGCTCAGG		b
NspHI	TAATACGACTCACTATAGGGRCATGT		b
PstI	TAATACGACTCACTATAGGGCTGCAG		b
PvuII	TAATACGACTCACTATAGGGCAGCTG		b
SduI	TAATACGACTCACTATAGGGGDGCHC		b
DOP-6MW	CCGACTCGAGNNNNNATGTGG		c
DOP-JUN1	CCCAAGCTTGCATGCGAATTCNNNNCAGG		d
pRML1 (Left Arm)			
L-I	GCTCAATGTTGACTAGCCAGC	55°C	g
L-II	TCATGTTTGACAGCTTATCATCG	55°C	g
L-III (T3)	ATTAACCCCTACTAAAGGGA	55°C	h
pRML2 (Right Arm)			
R-I	CACAGGATTTTCGTGTGTGG	55°C	g
R-II	GGCATTAATACCTGAGAGCAGG	55°C	g
R-III (T7)	TAATACGACTCACTATAGGG	50°C	h
pBeloBAC11 (Sp6 arm)			
Amplification	CAGCTATGACCATGATTACG	60°C	e
Sequencing	GATTACGCCAAGCTATTTAGGTGACACTATAGAATAC	50°C	f
pBeloBAC11 (T7 arm)			
Amplification	T7 promoter, TAATACGACTCACTATAGGG	55°C	
Sequencing	T7 promoter, TAATACGACTCACTATAGGG	50°C	

Ambiguous nucleotides are denoted with standard IUPAC/IUB codes. References for previously described primers are (a) primers 1089 and 1091 in (Riley et al. 1990), (b) (Johnston et al. 1995), (c) (Telenius et al. 1992), (d) (Weier et al. 1994), (e) R. Maatman; Jackson Laboratory, (f) (Wang et al. 1994). (g) The pRML1 (L-I and L-II) and pRML2 (R-I and R-II) nesting primers were designed using primer 0.5 (Lincoln et al. 1991). (h) The pRML1 L-III and pRML2 R-III primers are the T3 and T7 promoter sequences flanking the insert in the assembled YAC (Spencer et al. 1993).

appear to be the result of YAC chimerism or other artifact as they hybridized only to the YACs from which they were isolated, or they matched sequences from *Escherichia coli*.

Repetitive sequences were identified in 33 of 77 ends (43%) by BLAST analysis. Twenty-nine ends belonged to the LINE1 or L1 family of high copy number repetitive sequences, and the others included a B1 repeat, an intracisternal A-particle (IAP) *env* gene, a CT microsatellite repeat, and a flanking sequence of a presumed novel transposon (Bultman et al. 1994). The incidence of L1 elements at YAC ends is markedly higher than the 5% incidence observed in a YAC contig on chromosome 11 (Nehls et al. 1995) but more similar to the 25% seen in the *Igk* locus (Schupp et al. 1997) and the even higher incidence in the MHC locus (Jones et al. 1995). Clearly the frequencies of L1 elements at different genomic sites varies considerably. This is difficult to assess properly, as they are repeats and therefore, most were not mapped in our study or in the other studies. These YACs are potentially chimeric, and accurate comparison is not possible.

Hybridization Results

Fifty-six vector-hexamer PCR products were used as hybridization probes on YAC or other mapping blots. Six were unsuitable because of repetitive sequence content, but this did not always make an end product un-

Figure 2 Vector-hexamer PCR products. (a) Products of the initial PCR using the outermost YAC vector right arm primer R-I with the *PvuII* primer. Only nonspecific products are visible. (b) Products of the second round of PCR using the first nested right arm primer R-II with the *PvuII* primer. Both nonspecific and a few specific products are visible. (c) Products of the third round using the second nested right arm primer R-III with the *PvuII* primer. Bright YAC-specific products are apparent in many lanes. (d) Third-round PCR products from several different YACs made with the left arm L-III primer in conjunction with three different hexamer primers. The center gel demonstrates a background band amplified from the yeast host near the position of the largest size marker band (ϕ X174 digested with *HincII*) in all four YACs and the control YAC.

suitable. Nine ends that contained repeats were successfully used as probes, in most cases when the similarity to the repeat was low. Useful probes were generated when similarity was high if nonrepetitive sequence was detected beyond a repeat element and

could be isolated. For example, in one case a new PCR primer was designed to amplify the unique region, and in another the PCR product was digested at a restriction site near the end of the repeat and gel purified.

Of the 50 useful YAC insert end probes that were generated, 37 hybridized to other *Igh* YACs, whereas 13 probes hybridized only to their source YACs. In the latter cases, the overlap of other YACs ruled out incomplete coverage and indicated that these ends are not present in the *Igh* locus. Two of these probes were from clones containing more than one YAC; the remaining 11 ends from 36 YACs indicate chimerism.

Extension of Vector-Hexamer PCR to Other Vectors

The new WI/MIT-820 YAC library is intended to be the basis for development of a complete physical map of the mouse genome (Haldi et al. 1996). This library is constructed in a newer vector system, pRML1/2, and required design of new nested vector primers, shown in Table 1. In an initial test they were used in conjunction with only the two DOP hexamer primers to attempt recovery of the ends of 14 *Igh* region YACs from this library. Twenty-two ends were amplified and sequenced (data not shown), and we expect that the few remaining ends can be recovered using other hexamer primers. Consistent with our findings in *Igh* clones from the other YAC libraries 8 of the 22 ends (36%) were repeat sequences, and 6 of the repeats were L1 sequences.

We have also used vector-hexamer PCR to isolate and sequence the ends of BAC clones in the pBeloBAC11 vector (primers in Table 1). All 28 ends were readily obtained from the 14 BAC clones studied (data not shown). It is an advantage that vector-hexamer PCR can be used to amplify

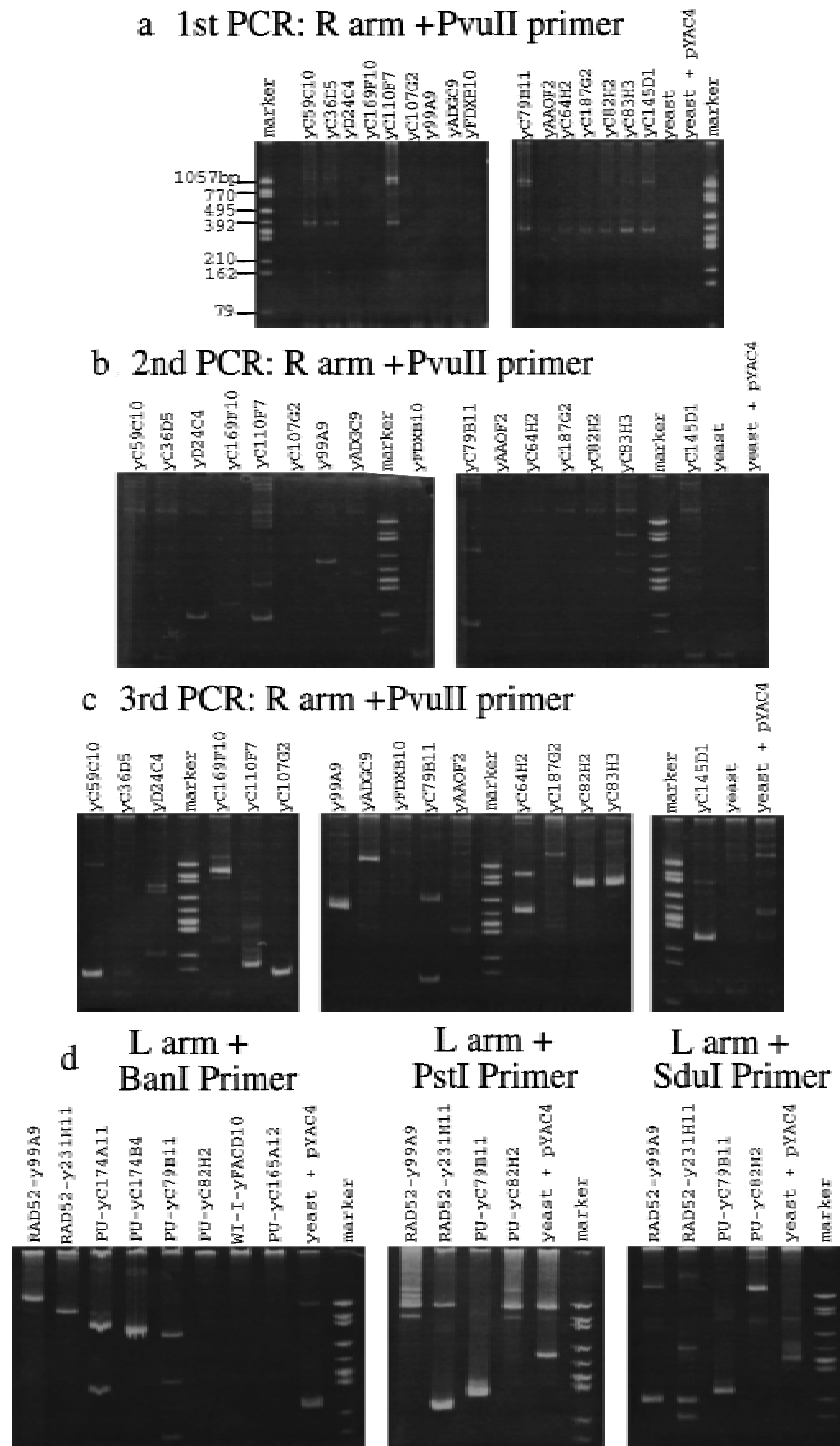


Figure 2 (See facing page for legend.)

Table 2. YAC Insert End Recovery with Vector-Hexamer PCR

Left Arm

YAC	BanI	BanIII	DdeI+G	NspHI	PvuII	PstI	SduI	DOP-6MW	DOP-JUN1
C36D5			280 ●	X	X	500 ○	260 ● B07590		
C59C10		X		350 ●	420 ○	780 ● B07592	860 ●		
D24C4			310 ● B07513		280 ●				
C110F7	240 ● B07594	270 ○	360 ●	950 ●	630 ●	130 ○	360 ● B07594	320 ● B07594	1100 ● B07594
C169F10	X	1200 ● B07553		X	X	700 ●	250 ○		
FEIH2				240 ● B07525					
C107G2		X	X	X	520 ●	480 ○	350 ● B07555		
99A9	1100 ●					1200 ●	210 ● B07519		
231H11	800 ●					220 ● B07546	210 ●		
ADGC9						1150 ● B07528	650 ○	170 ●	330 ●
C174B4	600 ● B07557	650 ●	370 ● B07557	500 ●	X	150 ●	1150 ●	370 ● B07571	350 ● B07572
C174A11	650 ●			190 ● B07575		X	X		
AAOF2						>1100 ●	330 ● B07530		
FDXB10						560 ● B07532	700 ●		
FEEH5						360 ● B07534	550 ● B07535		
52F12			>1100 ○		X	520 ● B07549	X		
116D10			160 ●		470 ● B07585	X	1150 ● B07586		
ABD10						300 ● B07577	1100 ●		
17A2	240 ● B07551	550 ○	>1100 ● B07552	>1100 ● B07552	X	X	100 ●		
84G12	X	X	>1100 ○	X	390 ●	X	X	1100 ● B07559	1100 ● B07559
FCDB1			160 ● B07537		X	500 ○	360 ● B07537		
C143D1			90 ●	X	X	150 ● B07579	100 ●		
C82H2	800 ○	X	1150 ●	X	X	X	1200 ● B07581		
FEQG5			X	X	X	340 ○	750 ○ B07539		
C117B7			650 ○	290 ● B07565	X				
C165A12	X	X	90 ●	X		350 ●	850 ○	340 ●	400 ● B07567
F3A5				500 ○		220 ○ B07569	220 ●		
C79B11	520 ● B07516					280 ● B07517	220 ● B07517		
FACD10	X	X	70 ●	X	X	750 ● B07542	1000 ● B07541		
I22G4			280 ● B07588		450 ○				
FCLA12			1150 ●	X	X	500 ○	340 ● B07544		
IB3			>1100 ○	290 ● B07561	X				
C64H2			270 ● B07520		X	750 ○	X		
C187G2						560 ● B07598	450 ●	700 ●	500 ●
C199B5			750 ● B07523	320 ● B07523	280 ○				
92C11			1150 ● B07597		240 ○	210 ○	210 ● B07597		

Right Arm

YAC	BanI	BanIII	DdeI+G	NspHI	PvuII	PstI	SduI	DOP-6MW	DOP-JUN1
C36D5	350 ● B07584		>1100 ○		550 ○	140 ●			
C59C10			700 ● B07591	100 ○	150 ●	X	110 ○ B07591		
D24C4	220 ● B07512		340 ● B07512		600 ○	120 ●			
C110F7	190 ● B07593		600 ● B07593		170 ●	210 ○	X		
C169F10			500 ●		900 ●	230 ● B07595	240 ○		
FEIH2	450 ● B07584					340 ●			
C107G2			750 ○	320 ● B07554	150 ●	370 ○	X		
99A9		390 ● B07518	900 ●	360 ● B07518	410 ●				
231H11	160 ●					650 ● B07545			
ADGC9			130 ○		1150 ● B07526	140 ● B07526	140 ●	370 ●	1150 ●
C174B4		X		270 ● B07556					
C174A11		1000 ○		400 ●		450 ● B07573	450 ●		
AAOF2			350 ● B07529		450 ●	>1100 ● B07574	1200 ●		
FDXB10	250 ●	400 ● B07531	530 ●	X	X	X	X		
FEEH5		X		340 ●		850 ● B07533	800 ○		
52F12			120 ●		360 ● B07147	140 ○	310 ●		
116D10			150 ●		210 ● B07562	1200 ●	1150 ● B07563		
ABD10	X	X	100 ● B07536	X	100 ○	440 ● B07576	X		
17A2	X	350 ● B07550	380 ●	300 ○	X	210 ○	90 ●		
84G12	220 ○	350 ○	200 ●	420 ●	X	1150 ●	350 ● B07538		
FCDB1	X	X	1200 ○	410 ● B07536	X	X			
C143D1			X		250 ● B07578				
C82H2			1100 ●		800 ● B07580				
FEQG5			380 ● B07538		240 ○				
C117B7		480 ● B07564	X	800 ●	X				
C165A12			640 ○		500 ● B07566			400 ●	130 ○
F3A5			800 ●		350 ● B07568				
C79B11			140 ● B07514		140 ● B07513				
FACD10			1150 ●		550 ● B07540				
I22G4			540 ● B07587		350 ○				
FCLA12			1150 ●		220 ● B07543				
IB3			330 ● B07560		X				
C64H2		X	320 ●	X	420 ● B07570	210 ○	X		
C187G2	X	X	X	X	320 ○	X	X	500 ● B07521	480 ●
C199B5			160 ● B07522		230 ○				
92C11	360 ○	X	1150 ●	800 ● B07589	X	360 ●			

(See facing page for footnote.)

Table 3. Hexamer Priming Fidelity

Accession no.	Priming site	Primer	Primer sequence
B07512	<u>GGCCCC</u>	BanI	GG (C, T) (G, A) CC
B07593	<u>TGCACC</u>	BanI	GG (C, T) (G, A) CC
B07594	<u>ACCACC</u>	BanI	GG (C, T) (G, A) CC
B07537	<u>CTCAGG</u>	DdeI+G	CTCAGG
B07557	<u>CCCAGG</u>	DdeI+G	CTCAGG
B07576	<u>TTAAGN</u>	DdeI+G	CTCAGG
B07589	<u>ACTTGT</u>	NspHI	(A, G) CATGT
B07526	<u>CTGCAG</u>	PstI	CTGCAG
B07529	<u>CTTCAG</u>	PstI	CTGCAG
B07515	<u>CACCTG</u>	PvuII	CAGCTG
B07517	<u>GGTCCC</u>	SduI	G (G, A, T) GC (C, A, T) C
B07591	<u>TGGCAC</u>	SduI	G (G, A, T) GC (C, A, T) C
B07594	<u>AAATGG</u>	DOP-6MW	ATGTGG
B07594	<u>CAGG</u>	DOP-JUN1	CAGG

Sequences from short YAC end PCR products in which the hexamer primer was discernible at the end were aligned to sequences from longer PCR products generated from the same YAC end with the same or a different hexamer primer. The sequence of the hexamer primer incorporated into the short product was compared to the corresponding sequence in the longer product (priming site). Mismatched nucleotides in the priming site are underlined.

BAC ends directly from bacterial colonies, and usually with just one round of PCR, simplifying the procedure and eliminating the expense of BAC DNA purifications.

METHODS

YAC and BAC Libraries

YAC and BAC clones were obtained from the following libraries. The Princeton University C57BL/6J mouse YAC library (Burke et al. 1991; Rossi et al. 1992) can be screened by PCR using DNA pools obtainable from S.M. Tilghman (Howard Hughes Medical Institute, Princeton University, NJ). The ICRF C3H mouse YAC library (Larin et al. 1991) is accessed through H. Lehrach at the Resource Center/Primary Database (RZPD) of the German Human Genome Project. See the World Wide Web information server of the RZPD at <http://www.rzpd.de>. The first Whitehead Institute (WI-I) C57BL/6J mouse YAC library (Kusumi et al. 1993) can be screened by PCR using pools available from Research Genetics (Huntsville, AL). The second Whitehead (WI/MIT-820) C57BL/6J mouse YAC library (Haldi et al. 1996) can be screened by PCR using pools obtained from

Research Genetics (see <http://www-genome.wi.mit.edu>). The Saint Mary's Hospital Medical School RAD52 C57BL/10 mouse YAC library (Chartier et al. 1992) is available for PCR screening using DNA pools obtained from Dr. S.D.M. Brown at the Mouse Genome Center, Harwell, UK (see <http://www.mgc.har.mrc.ac.uk>). The 129/Sv mouse BAC library was prepared by Dr. Bruce Birren at the Whitehead Institute Genome Center and can be screened by PCR or high-density filter hybridization; DNA pools and filters are available from Research Genetics (Huntsville, AL).

Vector-Hexamers PCR

This is a heminested strategy and is diagrammed in Figure 1; the primers used are listed in Table 1. First-round reactions consisted of 2 μ l 10 \times PCR buffer [500 mM KCl, 100 mM Tris (pH 8.0), 15 mM MgCl₂, 1 mg/ml Knox gelatin], 0.4 μ l dNTP solution (2.5 mM each dNTP), 0.08 μ l *Taq* polymerase (5 U/ μ l), 1 μ l (40 ng) of the outermost left arm or right arm nested vector primer (L-I or R-I; Table 1), 1 μ l (40 ng) hexamer primer, 1 μ l YAC DNA (25 ng/ μ l, as prepared in Ausubel et al. (1995), and water to a total volume of 20 μ l. The PCR program was five cycles (94°C for 30 sec, 35°C for 60 sec, 72°C for 3 min), immediately followed by 25 cycles (94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min) in an Ericomp Twinblock Easycycler (San Diego, CA). Multiple YACs and a negative control YAC lacking insert (see below) were amplified at the same time to allow identification of yeast background products. For the second round of nested PCR the same conditions as the first round PCR were used, except that the template was 1 μ l of first-round PCR product and the vector primer was the second nested vector arm primer, L-II or R-II. The PCR program was 30 cycles [94°C for 30 sec, 50°C (pYAC4 left arm) or 60°C (pYAC4 right arm) for 30 sec, 72°C for 3 min]. For the third round 1 μ l of second-round product was used as template in a 40- μ l total volume reaction with the third nested vector primers. Volumes of all reagents (except template) were doubled. The PCR program was 30 cycles (94°C for 30 sec, 55°C for 30 sec, 72°C for 3 min). A negative control was a YAC clone identical to the experimental YACs but lacking insert. This control was created by transforming yeast host AB1380 in lithium acetate with linearized pYAC4 (both available from ATCC, nos. 20843 and 67379, respectively), as described in Ausubel et al. (1995).

Table 2. (Continued from facing page.) The 36 *Igh* YAC clones in this study are listed vertically in the *top* half for the left-arm ends and in the bottom half for the right-arm ends. Hexamer primers are listed across the *top*. Results of each attempt to recover a YAC insert end with vector-hexamers PCR are indicated. Reactions that did not produce any visible YAC end product are denoted X; blank cells are combinations of YAC end and hexamer primer that were not tried. For each successful reaction the size (in bp) and intensity are given for the most readily sequenceable product (those of the greatest intensity and closest to 400 bp). The yield of the PCR is indicated as intense product (~0.5–2.0 μ g) (●); medium intensity (◐); and faint product (○). Sequencing was usually successful with medium- and high-intensity fragments but only rarely with faint products. The size listed is the third-round product, which includes ~75 bp of vector sequence. Where an end product was successfully sequenced the GenBank accession no. is given to the *right* of the intensity indicator. Where GenBank entries are the compilation of several sequences obtained with different hexamer primers the same accession no. appears several times in a row. In certain cases, the sequence from the vector primer did not overlap the reverse sequence from the hexamer primer and multiple accession numbers are listed for one YAC end amplification.

The pYAC4 plasmid was isolated from bacterial culture with a standard alkaline lysis procedure and linearized with *Bam*HI before being used to transform the yeast.

BAC insert ends were recovered in a similar manner using as initial template either ~100 ng of BAC DNA purified on Qiagen columns according to the manufacturer's instructions or 1–2 μ l of a crude "boil prep" made by picking a bacterial colony into 20 μ l of water, boiling for 5 min, and centrifuging particulate matter. For the SP6 end different primers were used for amplification and sequencing as shown in Table 1. For the T7 end amplification and sequencing were done with the same T7 promoter primer. Multiple BACs and a negative control (pBeloBAC11 without insert) were amplified at the same time to allow identification of bacterial background products. The BAC PCR program was 35 cycles [94°C for 30 sec, 60°C (Sp6 amplification primer) or 55°C (T7 amplification primer) for 30 sec, 72°C for 1 minute].

Gel Analysis

Thirty-five microliters of third round PCR products was electrophoresed on 10 \times 8 \times 0.15-cm vertical 10% 37.5:1 polyacrylamide (AMRESCO, Solon, OH) gels. Gels were stained with ethidium bromide. Bands ranging in size from 120 bp to ~1500 bp (optimally 300–500 bp) were isolated if they were reasonably bright and did not appear in other YAC lanes or the negative control using the same hexamer primer. Gel segments measuring ~2 \times 5 \times 1.5 mm were excised with a scalpel and placed in individual yellow 200- μ l pipette tips. These tips were put into 0.65-ml microcentrifuge tubes with the pipette tips protruding through holes that had been punched in the bottom of the tubes with an 18-gauge needle. The 0.65-ml tubes were placed into 2-ml microcentrifuge tubes and spun for 1 min in a microcentrifuge at 12,000g. The gel fragments were extruded through the yellow tips into the bottom of the 2.0-ml tubes as a fine paste. The tip and 0.65-ml tubes were discarded and 30 μ l of water were added. They were protected from light and agitated overnight to elute the DNA.

Sequencing

Insert ends were sequenced with the ABI Prism dye terminator cycle sequencing kit and ABI373 automated sequencer (Perkin Elmer, Foster City, CA). The gel fragment homogenates were centrifuged briefly and up to 11 μ l of eluate were removed from the top and used as template with 1 μ l of L-III or R-III primer (20 ng/ μ l). Use of this primer resulted in the presence of the vector/insert junction (*Eco*RI site) in the sequence confirming that the correct product was obtained. Temperature cycling was 25 cycles (96°C for 30 sec, 50°C for 15 sec, 60°C for 4 min) in an Ericomp Twinblock Easycycler. The excess terminators were removed by phenol:chloroform extraction and ethanol precipitation or by ethanol precipitation alone. The same methodologies were used for BAC insert end sequencing with the SP6 and T7 sequencing primers in Table 1.

ACKNOWLEDGMENTS

We are grateful for the generous assistance and access to YAC libraries and screening pools provided by D. Koos, G. Guan,

and S. Tilghman; F. Chartier, G. Argyropoulos, and S.D.M. Brown; and E. Brundage. This work was supported by National Institutes of Health grant AI23548.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.
- Ausubel, F.M., R. Brent, R. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl. 1995. *Current protocols in molecular biology*. John Wiley, New York, NY.
- Breukel, C., J. Wijnen, C. Tops, H. Van der Klift, H. Dawerse, and P.M. Khan. 1990. Vector-Alu PCR: A rapid step in mapping cosmids and YACs. *Nucleic Acids Res.* 18: 3097.
- Bultman, S., M.L. Klebig, E.J. Michaud, O. Sweet, M.T. Davisson, and R.P. Woychik. 1994. Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (at) and white-bellied agouti (aw) reveals alternative forms of agouti transcripts. *Genes & Dev.* 8: 481–490.
- Burke, D.T., G.F. Carle, and M.V. Olson. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236: 806–812.
- Burke, D.T., J.M. Rossi, J. Leung, D.S. Koos, and S.M. Tilghman. 1991. A mouse genomic library of yeast artificial chromosome clones. *Mamm. Genome* 1: 65.
- Chartier, F.L., J.T. Keer, M.J. Sutcliffe, D.A. Henriques, P. Mileham, and S.D.M. Brown. 1992. Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. *Nature Genet.* 1: 132–136.
- Cook, G.P., I.M. Tomlinson, G. Walter, H. Riethman, N.P. Carter, L. Buluwela, G. Winter, and T.H. Rabbitts. 1994. A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. *Nature Genet.* 7: 162–168.
- Green, E.D. and M.V. Olson. 1990. Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci.* 87: 1213–1217.
- Haldi, M., C. Strickland, P. Lim, V. VanBerkel, X.-N. Chen, D. Noya, J.R. Korenberg, Z. Husain, J. Miller, and E.S. Lander. 1996. A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome. *Mamm. Genome* 7: 767–769.
- Johnston, S.L., M. Strausbauch, G. Sarkar, and P.J. Wettstein. 1995. A novel method for sequencing members of multi-gene families. *Nucleic Acids Res.* 23: 3074–3075.

- Jones, E.P., H. Xiao, R.A. Schultz, L. Flaherty, Z. Trachtulec, V. Vincek, Z. Larin, H. Lehrach, and K.F. Lindahl. 1995. MHC class I gene organization in >1.5-Mb YAC contigs from the *H2-M* region. *Genomics* 27: 40-51.
- Kusumi, K., J.S. Smith, J.A. Segre, D.S. Koos, and E.S. Lander. 1993. Construction of a large-insert yeast artificial chromosome library of the mouse genome. *Mamm. Genome* 4: 391-392.
- Larin, Z., A.P. Monaco, and H. Lehrach. 1991. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci.* 88: 4123-4127.
- Lincoln, S.E., M.J. Daly, and E.S. Lander. 1991. *PRIMER: A computer program for automatically selecting PCR primers*. MIT Center for Genome Research and Whitehead Institute for Biomedical Research, Cambridge, MA.
- Liu, Y.G. and R.F. Whittier. 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.
- Mainville, C., K. Sheehan, L.D. Klamon, C.A. Giorgetti, J.L. Press, and P.H. Brodeur. 1996. Deletional mapping of fifteen mouse Vh gene families reveals a common organization for three Igh haplotypes. *J. Immunol.* 156: 1038-1046.
- Nehls, M., K. Lüno, M. Schorpp, D. Pfeifer, S. Krause, U. Matysiak-Scholze, H. Dierbach, and T. Boehm. 1995. YAC/P1 contigs defining the location of 56 microsatellite markers and several genes across a 3.4 cM interval on mouse Chromosome 11. *Mamm. Genome* 6: 321-331.
- Nelson, D.L., S.A. Ledbetter, L. Corbo, M.F. Victoria, R. Ramirez-Solis, T. Webster, D.H. Ledbetter, and C.T. Caskey. 1989. Alu polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci.* 86: 6686-6690.
- Riley, J., R. Butler, D. Ogilvie, R. Finniear, D. Jenner, S. Powell, R. Anand, J.C. Smith, and A.F. Markham. 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* 18: 2887-2890.
- Rossi, J.M., D.T. Burkle, J.C. Leung, D.S. Koos, H. Chen, and S.M. Tilghman. 1992. Genomic analysis using yeast artificial chromosome library with mouse DNA inserts. *Proc. Natl. Acad. Sci.* 89: 2456-2460.
- Sarkar, G., R.T. Turner, and M.E. Bolander. 1993. Restriction-site PCR: A direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. *PCR Methods Appl.* 2: 318-322.
- Schupp, I.W., T. Schlake, T. Kirschbaum, H.G. Zachau, and T. Boehm. 1997. A yeast artificial chromosome contig spanning the mouse immunoglobulin kappa light chain locus. *Immunogenetics* 45: 180-187.
- Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300 kilobase-pair fragments of human DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 89: 8794-8797.
- Silverman, G.A. 1993. Isolating vector-insert junctions from yeast artificial chromosomes. *PCR Methods Appl.* 3: 141-150.
- Silverman, G.A., R.D. Ye, K.M. Pollock, J.E. Sadler, and S.J. Korsmeyer. 1989. Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3. *Proc. Natl. Acad. Sci.* 86: 7485-7489.
- Spencer, F., G. Ketner, C. Connelly, and P. Hieter. 1993. Targeted recombination-based cloning and manipulation of large DNA segments in yeast. *Methods: Companion Methods Enzymol.* 5: 161-175.
- Telenius, H., N.P. Carter, C.E. Bebb, M. Nordenskjold, B.A. Ponder, and A. Tunnacliffe. 1992. Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics* 13: 718-725.
- Triglia, T., M.G. Peterson, and D.J. Kemp. 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res.* 16: 8186.
- Tutter, A., P. Brodeur, M. Shlomchik, and R. Riblet. 1991. Structure, map position, and evolution of two newly diverged mouse Ig VH gene families. *J. Immunol.* 147: 3215-3223.
- Wang, M., X.-N. Chen, S. Shouse, J. Manson, Q. Wu, R. Li, J. Wrestler, D. Noya, Z.-G. Sun, J. Korenberg, and E. Lai. 1994. Construction and characterization of a human chromosome 2-specific BAC library. *Genomics* 24: 527-534.
- Weier, H.U., D. Polikoff, J.J. Fawcett, K.M. Greulich, K.H. Lee, S. Cram, V.M. Chapman, and J.W. Gray. 1994. Generation of five high-complexity painting probe libraries from flow-sorted mouse chromosomes. *Genomics* 21: 641-644.
- Wesley, C.S., M. Ben, M. Kreitman, N. Hagag, and W.F. Eanes. 1990. Cloning regions of the *Drosophila* genome by microdissection of polytene chromosome DNA and PCR with nonspecific primer. *Nucleic Acids Res.* 18: 599-603.
- Wesley, C.S., M.P. Myers, and M.W. Young. 1994. Rapid sequential walking from termini of cosmid, P1 and YAC inserts. *Nucleic Acids Res.* 22: 538-539.
- Wu, C., S. Zhu, S. Simpson, and P.J. de Jong. 1996. DOP-vector PCR: A method for rapid isolation and sequencing of insert termini from PAC clones. *Nucleic Acids Res.* 24: 2614-2615.

Received July 14, 1997; accepted in revised form February 20, 1998.