4524–4529 Nucleic Acids Research, 1993, Vol. 21, No. 19 © 1993 Oxford University Press

HAPPY mapping of a YAC reveals alternative haplotypes in the human immunoglobulin V_H locus

Gerald Walter[1,2], Ian M.Tomlinson[1,2], Graham P.Cook[1,2], Greg Winter[1,2], Terry H.Rabbitts[2] and Paul H.Dear[1,2,*]

[1]MRC Centre for Protein Engineering and [2]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Received June 18, 1993; Revised and Accepted August 17, 1993

ABSTRACT

We have identified and sequenced 14 human immunoglobulin V_H segments cloned in a yeast artificial chromosome, and have used a rapid PCR-based technique (HAPPY mapping, 12) to derive the order and approximate distances between them. The sequences mapped comprise thirteen germline V_H segments and one rearranged V_H3 gene. Comparison of our map with other data suggests the existence of at least two distinct haplotypes, differing in the presence or absence of the consecutive genes DP-78, DP-46 and DP-64, and in the duplication of segments DP-49 and DP-65. Screening of ten individuals confirms the existence of both haplotypes, and indicates that both are common amongst the population.

INTRODUCTION

The variable portions of the immunoglobulin heavy-chain genes are generated during B-cell development by the recombination of V_H, D and J_H segments. The human germline V_H locus contains approximately 120 V_H segments, of which 40–50 appear functional (1,2). V_H segments can be divided into six families on the basis of nucleotide homology (3–7, 10). The majority of segments are located on the long arm of chromosome 14 (8, 9), and members of the different families are randomly interspersed throughout a region of ~1Mb upstream of the major D locus (11).

We have previously used family-specific primers to amplify and sequence germline V_H segments from genomic DNA of a single individual using the polymerase chain reaction (1). Using this information, we set out to map the human V_H segments on chromosome 14. The high degree of homology between V_H segments of the same family restricts the use of mapping techniques which rely mainly on hybridisation. We have therefore used a novel method (HAPPY mapping, 12) in which gene-specific PCR primers can be used to exploit minor sequence differences. The method is analogous to classical linkage mapping, in which meiosis breaks the DNA (crossover) and segregates the recombined fragments into gametes, linked markers tending to cosegregate. We replace these steps with *in vitro* analogues: DNA is broken physically and 'segregated' by dilution into many aliquots containing, on average, none or one copy of each marker. Linked sequences tend to lie on the same DNA fragment, and hence tend to occur together amongst the aliquots (i.e., they 'cosegregate'). Since we break the DNA far more frequently than in meiosis, sequences can be mapped with far higher resolution.

We report here the mapping of 14 V_H segments from a 280kb yeast artificial chromosome (YAC) spanning part of the human V_H locus. DNA from the YAC-containing yeast was broken by shearing, and divided into 120 aliquots, each containing ~0.5 yeast haploid genome equivalents (and hence, on average, ~0.5 copies of each V_H segment). V_H segments present in each aliquot were amplified in 'nested' PCRs (the first phase with a set of primers designed to amplify all 14 segments, followed by a series of second-phase amplifications with segment-specific primers) and detected by gel electrophoresis. Linkages between pairs of segments were then calculated from the frequency with which they were found together in aliquots (12). From the linkages between all possible pairs of segments, a map giving segment order and approximate distances has been constructed.

MATERIALS AND METHODS

Characterisation of yeast artificial chromosome IGH_2

A YAC library (13) derived from GM1416 (48, 4X, a human lymphoblastoid cell line) was screened using PCR primers which amplify a 250bp fragment spanning the BglII site of the human IgH enhancer (ref. 14; forward primer 5'-GGG TCT CCG GCA CCC ACA GCA GGT G-3'; back primer 5'-GCT AAG ATT GGA AAT GAA ATT CAG A-3'; ~10ng template DNA; 30μl total volume; 30 cycles, each 90°C×2min, 55°C×2min, 72°C×2min, followed by a final step of 72°C×10min; other conditions as for second-phase PCRs [below]). A 280kbp YAC (IGH_2), giving a PCR product of the correct size, was isolated and shown by Southern hybridisation to contain a V-D-J rearrangement (data not shown). V_H segments on IGH_2 were amplified using family-specific primers (table 1(a); annealing

* To whom correspondence should be addressed

temperature 55°C, otherwise as described for second-phase PCRs [below]), and the PCR products cloned and sequenced according to ref. 1.

Preparation of yeast DNA

Yeast cells (*Saccharomyces cerevisiae*, strain AB1380) containing the YAC IGH_2 were grown at 30°C to mid-logarithmic phase in rich medium lacking uracil, spheroplasted and encapsulated in agarose microbeads at a final density of $\sim 10^6$cells/ml of agarose (15). Microbeads were washed 6 times at intervals of 30 minutes in 10mM Tris (pH8.0), 1mM Na_2EDTA, 1% w/v lithium dodecyl sulphate to remove cellular proteins, and stored at 4°C in TE (10mM Tris.HCl pH 8.0, 1mM Na_2EDTA).

Shearing and sonication of DNA

Microbeads were melted in TE to give a final concentration of ~ 0.5 haploid equivalents/μl (~ 6fg/μl), and the DNA dispersed by inverting the tube10 times. DNA was either sheared by passing once through a 200μl pipette tip, or sonicated using a Heat Systems model XL sonicator (10 seconds, 95% duty cycle, maximum amplitude, tip diameter 2mm). To estimate the size of the resulting fragments, samples containing a higher concentration ($\sim 0.2\mu$g/μl) of yeast DNA were treated similarly and analysed using either pulsed-field or conventional electrophoresis for sheared and sonicated material respectively, with suitable size standards. (Earlier experiments [12] have shown that the size-range in the dilute samples used for mapping can

Table 1. PCR primers

a)

Family		Primer sequence
V_H1	F:	GCT CTA GAT STG GKT TYT CAC ACT GTG
	B:	CCC AAG CTT CCA TGG ACT GGA YYT GGA G
	B:	CCC AAG CTT GRA RGR GAT TKD KTC CAG
V_H2	F:	GCT CTA GAC CAC AGG GTC CAT GTT GG
	B:	CCC AAG CTT CCT TGA RGG AGT CTG GTC C
V_H3	F:	GGA ATT CMT GRC YTC CCC TCR CTS TG
	B:	CCC AAG CTT GTW TGC ARG TGY CCA GTG T
V_H4	F:	GGA ATT CAC TCA CCT CCC CTC ACT GTG
	B:	CCC AAG CTT CTG TTC ACA GGG GTC CTG TC
J-region	F:	GCT CTA GAT GAG GAG ACG GTG ACC AGG GT

b)

Segment		Primer sequence	Product size/bp	Temp/ °C
DP-5	F:	GCT CTA GAC ACW GTA ATA HDY GGC HGY RTC	229	60
	B:	GGT TTC CGG ATA CAC CC		
DP-9	F:	GAC AAA CCG TCC TGT GA	167	60
	B:	CTG GAG CCT CAT TGA AG		
DP-26	F:	TCC TTG GAG ATG GTG AG	129	60
	B:	TGC TAG AAT GGG TGT GA		
DP-46	F:	GCT CTA GAG TRA RTC KGC CYT TCA CRG	138	60
	B:	CCT TCA GTA GCT ATG CT		
DP-47	F:	CTC TGG AGA TGG TGA AC	215	62
	B:	GGT GCA GCT GTT GGA GT		
DP-49	F:	GGA ATT CAT TRC TTC CAT CAT ATG	96	56
	B:	CCT TCA GTA GCT ATG GC		
DP-50	F:	GGA ATT CAT TAC TTC CAT CAT ACC	96	56
	B:	CCT TCA GTA GCT ATG GC		
DP-55	F:	GGC TGT TCA TTT GCA GAT GA	216	62
	B:	CTG CGT GGT CCC CGA GA		
DP-57	F:	CTG TTG TCC CAC CAT TA	106	60
	B:	GGA TTC ACC TTC AGT TA		
DP-64	F:	GGT TCT TGG ACC TGT CT	227	60
	B:	GCT GCA GGA GTC CGG CT		
DP-65	F:	GGG AGA ACT GGT TCT TA	136	60
	B:	AGC TGG ATC CGC CAG CA		
DP-68	F:	TGG ACG TGT CTA CTG AC	137	56
	B:	CAG TAG TAA CTG GTG GG		
DP-78	F:	TAC ACG GCC GTG TCT GC	190	56
	B:	GTG ATT ACT ACT GGA GT		
DP-77(R)	F:	AGC TAG TAG GGG TTG AT	235	60
	B:	GGA TTC ACC TTC AGT AT		

F=forward primer, B=back primer; all primer sequences are shown 5′ to 3′ **(a)** External (family-specific) primers. To amplify V_H1 segments present on IGH_2, two family-specific back primers were used in combination; the J-region primer, in conjunction with the V_H3 back primer, amplifies the rearranged V_H3 gene DP-77(R). **(b)** Internal (segment-specific) primers; the annealing temperature used is indicated in the last column. One-letter ambiguity codes used to indicate mixed bases: S=G/C; K=G/T; Y=C/T; R=G/A; M=A/C; W=A/T; D=G/A/T; H=A/C/T. Underlined bases are restriction sites added to facilitate cloning, and do not correspond to bases in the target sequence.

be assumed to be similar to that in these more concentrated samples.)

PCR primers

Primer sequences are given in table 1. Family-specific primers were modified from ref.1; segment-specific primers were designed using the program Oligo 4.0-s (National Biosciences). All primers were tested for efficiency according to the criteria described in ref.12. Due to the similarity between many V_H sequences, all gene-specific (internal) primers were tested for specificity based on their ability to yield a product of the correct size from a mixture of M13 clones carrying all V_H sequences present in IGH_2 ($\sim$10 copies of each clone), whilst yielding no product from a mixture of clones from which their cognate V_H segment alone was omitted. These PCRs were performed using conditions identical to those used in mapping experiments (below). All gene-specific primer pairs met these criteria.

Marker detection

Two-phase 'nested' PCRs were used to detect V_H segments in the aliquots. External family-specific primers (below) were used to amplify all V_H sequences present in each aliquot in a single multiplex PCR. Sub-fractions of the products of this reaction were then re-amplified using specific nested primers for groups of two or three V_H segments at a time.

First-phase (external) PCRs contained 20mM Tris–HCl (pH 8.3 at 20°C), 50mM KCl, 2mM $MgCl_2$, 200μM each dNTP (Pharmacia), 2μM each external primer, 0.25U Taq polymerase (Cetus) and $\sim$0.5 haploid-equivalents of sheared or sonicated DNA prepared as above, in a total volume of 5μl, overlaid with $\sim$30μl mineral oil (Sigma), and were performed in polycarbonate microtitre plates (Techne). Amplification was performed in a Techne PHC-3 thermal cycler (93°C×5min, followed by 25 cycles of 95°C×20s, 50°C×30s, 72°C×1min). 95μl of water were added to each well, and the plate stored at −20°C pending second-phase PCRs.

Second-phase (gene-specific) multiplex PCRs were performed for two or three V_H sequences simultaneously (combinations of gene-specific primer pairs which could be multiplexed in this way were dictated by the requirement that multiplexed primer pairs have identical annealing temperatures and give products of different sizes which can be resolved by agarose gel electrophoresis). Second-phase PCRs (10μl total volume) comprised Tris–HCl, KCl, dNTPs and $MgCl_2$ as above, 1μM each relevant gene-specific primer, 0.5U Taq polymerase and 4μl of the diluted first-phase PCR product, and were performed as for first phase PCRs other than the cycle conditions (93°C×5min, followed by 33 cycles of 94°C×10s, annealing [see table 1]×30s, 72°C×1min).

8μl of loading dyes (25mM EDTA, 10mg/ml Ficoll 400, 0.1mg/ml bromophenol blue in TBE (16)) were added, and 10μl samples resolved on horizontal 4% agarose gels followed by ethidium bromide staining and visualisation by UV light.

Contamination control

Rigorous precautions were taken to exclude contaminating DNA, particularly before second-phase PCRs. All first-phase PCRs were prepared using dedicated equipment in a laminar flow hood in a room remote from that used for second-phase PCRs and product analysis. Aerosol-resistant pipette tips (Continental Laboratory Products) were used for all liquid handling other than gel loading. The contamination rate seen in negative controls was <0.05% (i.e. one false-positive in 2184).

Data entry and analysis

Data entry and analysis were performed using software written by P.H.D. essentially as described in ref. 12. The linkage between all pairwise combinations of V_H segments was determined as a LOD score (reflecting the certainty of linkage between segments) and theta (reflecting the distance between them). The 100 likeliest segment orders were determined and, for each of these candidate orders, the optimum segment spacing and total LOD score of the candidate map was determined. The difference in LOD scores for various possible segment orders then reflects their relative likelihoods (12).

Amplification and sequencing of V_H segments from genomic DNA

Individuals lacking the segment DP-64 were identified as follows. DNA was prepared from peripheral blood lymphocytes of ten anonymous, normal blood donors (17). 10ng of each DNA was amplified with DP-64 primers (table 1(b)),which are specific for DP-64 in genomic DNA, using conditions for second-phase PCRs [as above].

This approach could not be used to screen individuals for DP-78, as close sequence homology prevents the design of primers which are specific in genomic DNA. A two-step approach was therefore used. Genomic DNA was amplified using the DP-78 forward and V_H4 back primers (table 1; annealing temperature 69°C). PCR products were then cloned (Stratagene pCR-Script system), and clones probed with a ^{32}-P labelled DP-78-specific oligonucleotide (5′-TAG TAA TCA CCA CTG CT-3′; wash temperature 59°C, TMACl wash solution, as described in ref. 1). The combination of PCR primers and this probe are specific for DP78.

Table 2. Equivalence between V_H exon sequences on IGH_2 and those from other sources

IGH_2	Equivalent	Nucleotide differences	Reference
DP-5	1–24P	0	2
DP-9	1–27P	0	2
DP-26	2–26	0	2
DP-46	3d216	0	19
" "	56p1	0	18
DP-47	3–23	0	2
DP-49	3–30	1	2
" "	3d28	0	19
" "	1–9III	0	10
DP-50	3–33	4	2
" "	3d277	0	19
DP-55	3–25P	2	2
DP-57	3–22P	2	2
DP-64	3d216d	0	20,19
DP-65	4–31	1	2
" "	3d277d	1	19
DP-68	4–28	0	2
" "	3d28d	0	19
" "	1–9II	0	10
DP-78	3d230d	0	20
DP-77(R)	3–21	4	2

In each case, the number of nucleotide differences between the IGH_2 sequence and its equivalent(s) is given. The IGH_2 sequence DP-77(R) is a rearranged form of the germline segment 3–21.

RESULTS

V_H segments present on IGH_2

Thirteen germline V_H segments (two V_H1s [DP-5, DP-9], one V_H2 [DP-26], six V_H3s [DP-46, DP-47, DP-49, DP-50, DP-55, DP-57] and four V_H4s [DP-64, DP-65, DP-68 and DP-78]) and one rearranged V_H3 gene (DP-77(R)) were identified amongst the cloned PCR products from IGH_2 (nomenclature as in ref. (1)). The sequences of the germline segments are identical to those seen in the unrelated individual D.P. (1), suggesting that there is little sequence polymorphism at the level of individual V_H segments. The rearranged gene (DP-77(R), rearranged to J_H4b) differs by 4 nucleotides from the germline segment DP-77 (1); these differences are probably due to somatic mutation associated with expression, since little germline polymorphism has been found in this segment (unpublished). Of the thirteen germline segments on IGH_2, ten correspond closely to sequences lying on a recent YAC contig map of this region (2); the remaining three do not correspond to any of the sequences on this contig map. The relation between V_H segments on IGH_2 and other mapped sequences is given in table 2.

Mapping of V_H segments

A total of 120 aliquots of sheared DNA were screened for the fourteen V_H segments present on IGH_2 as described. Figure 1 shows an example of one of the second-phase multiplex PCRs. Maximum LOD scores and the corresponding theta values between all pairs of segments are shown in table 3 (Theta is defined as the probability of a break occurring between two segments, and is analogous to the recombination fraction in classical linkage mapping.)

Assuming that the YAC IGH_2 carries a single copy of each segment, each should be found in a similar number of aliquots. This was the case for all segments with the exception of DP-49 and DP-65, which were found in 56 and 61 aliquots respectively (mean for all other segments=43.7 occurrences out of 120 aliquots; standard deviation of number of occurrences=5.3). These frequencies are higher than expected by chance ($p<0.01$ and $<.0006$ respectively), and suggest a duplication of these segments (see also below). These segments were therefore omitted from the initial analysis, and are considered later.

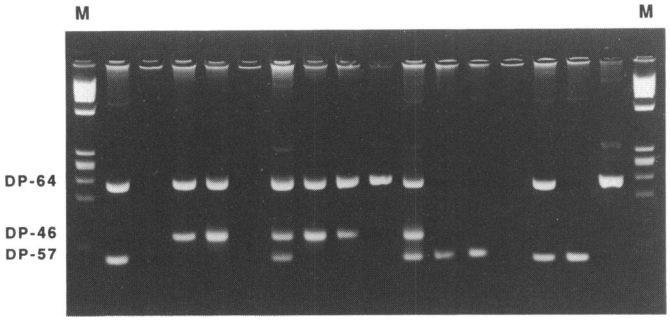


Figure 1. Example of multiplex PCR detection of V_H segments. Sixteen tracks are shown, corresponding to 16 of the 120 aliquots of sheared DNA analysed. Each aliquot was amplified for all V_H segments in a first-phase PCR as described in the text, then sub-fractions were re-amplified in a second-phase multiplex PCR for segments DP-46, DP-57 and DP-64. Markers (M) were HaeIII-digested ϕX-174 DNA.

For the remaining twelve segments, candidate gene-orders were evaluated as described (Materials and Methods), and the relative likelihood of these orders determined. The most probable order is shown in figure 2(a). The second and third likeliest orders were identical but for the exchange of positions of DP-9 with DP-68, or of DP-47 with DP-57, with likelihoods of 1/590 and 1/720 respectively, relative to the most probable order; all other orders had relative likelihoods below 1/5000.

Map distances are initially derived relative to the size of the fragments from which the aliquots are taken; a mapping function, based on the size distribution of these fragments, allows absolute distances to be calculated (12). The fragments of sheared DNA were found by pulsed-field gel electrophoresis to lie in the range 25−125kbp, with the greatest concentration of fragments at ~70kbp (results not shown). From this size distribution, the mapping function and physical distances were determined essentially as described in (12). Error estimates (95% confidence level) for these distances are +/−3.5kbp for all segments except DP46 and DP64 (+/−5kbp). These error estimates, however, do not take into account any systematic errors which may tend to distort the map (see below).

Effect of sonication

120 aliquots of sonicated DNA (mean size ~5kb) were analysed in the same way as for sheared DNA. Again, all V_H segments were found in a similar number of aliquots, with the exception of DP-49 and DP-65 which showed an even greater excess than in the sheared DNA (89 and 86 occurrences respectively; mean for all other segments=55.5 occurrences). With regard to the other segments, the effect of sonication of the DNA (prior to division into aliquots) should be to abolish linkage between all segments separated by more than a few kilobase pairs. (Sonication thus serves as a control, confirming that the linkages seen in sheared DNA are not artefactual). This was seen to be the case: only those segments most tightly linked in the sheared DNA remained linked (at a lower level) after sonication (for example, LOD score between DP-9 and DP-68=2.81, compared with 15.57 in sheared DNA).

Placement of DP-49 and DP-65

As stated, both of these segments were found with anomalously high frequencies amongst the aliquots of sheared DNA, and at even higher frequencies when using sonicated DNA. The simplest explanation—PCR contamination—is eliminated by the absence of contamination seen in negative controls, and would not account for the effects of sonication. A second explanation—that the internal primers for DP-49 and DP-65 are cross-reacting with the other V_H segments on IGH_2—is also eliminated by testing primers for specificity against a panel of M13 clones (see Materials and Methods).

This suggests that DP-49 and DP-65 are each duplicated; indeed, duplication of DP-49 has recently been reported (18). In the sheared DNA, the two copies of any one segment show linkage to one another, and hence frequently appear as a single copy; sonication breaks apart the two copies of each segment, so that they are detected independently and hence more frequently. By comparing the numbers of occurrences of these segments before and after sonication, we calculate that the two copies of DP-49 lie approximately 30−40kb apart, as do the two copies of DP-65.

These duplications complicate efforts to rigorously place DP-49 and DP-65 on the map, because the inability to distinguish

between the duplicate copies degrades the data. However, the linkages seen between DP-49 and DP-65 and other segments are as one would expect if one copy of each lay between DP-64 and DP-68, and the second copy of each between DP-50 and DP-78. We therefore tentatively locate these segments as shown in figure 2(a). This model provides a mechanism for the deletion of the three segments DP-78−DP-46−DP-64 by homologous recombination between the two copies of the DP-65−DP-49 region, to produce the alternative haplotype seen in figure 2(b).

Alternative haplotypes in the population

In addition to the duplications of DP-65 and DP-49, our map contains a 3-segment insertion (DP-78−DP-46−DP-64) relative to that recently published (2). To verify that this insertion represents an alternative haplotype of this region, rather than a YAC cloning artefact, genomic DNA from a panel of ten normal individuals was screened by PCR for DP-64. Seven individuals having DP-64 and three lacking DP-64 were identified. One individual with, and one without DP-64 were screened for DP-78 using PCR, cloning and probing (Materials and Methods). Of 250 V_H4 clones from the individual having DP-64, 67 carried DP-78, the remainder representing other genomic sequences cross-reacting with the DP-78 primers. Conversely, of 250 clones from the individual lacking DP-64, none contained DP-78.

Hence we conclude that two haplotypes (DP-64$^+$/DP-78$^+$ and DP-64$^-$/DP-78$^-$) exist in this region. DP-46 has been found adjacent to DP-64 in a bacteriophage lambda clone (19), and deletions have previously been observed involving DP-46 (18) or DP-64 and DP-78 (20). This supports our evidence that DP-78, DP-46 and DP-64 lie consecutively in one haplotype of this region, and are deleted *en bloc* in another. All three of these segments have been seen to be functional (21).

Table 3. Maximum LOD scores and theta values (italics) between V_H segments on the yeast artificial chromosome IGH_2, based on the analysis of 120 aliquots of sheared DNA as described in the text.

	DP 9	DP 26	DP 46	DP 47	DP 49	DP 50	DP 55	DP 57	DP 64	DP 65	DP 68	DP 78	DP 77R
DP5	1.59 *0.70*	3.17 *0.59*	0.39 *0.84*	11.09 *0.31*	0.02 *0.96*	0.02 *0.97*	7.90 *0.38*	7.04 *0.42*	0.11 *0.92*	0.00 *1.00*	1.90 *0.67*	0.00 *1.00*	6.67 *0.42*
DP9	-	8.02 *0.39*	2.24 *0.65*	1.61 *0.71*	6.130 *0.48*	0.29 *0.87*	4.16 *0.54*	2.38 *0.64*	4.09 *0.55*	0.04 *0.95*	15.57 *0.21*	0.86 *0.77*	0.79 *0.78*
DP26		-	1.00 *0.76*	3.47 *0.58*	1.67 *0.71*	0.26 *0.88*	7.11 *0.42*	2.84 *0.62*	1.57 *0.71*	0.00 *1.00*	4.06 *0.54*	0.11 *0.91*	0.73 *0.79*
DP46			-	0.32 *0.86*	7.27 *0.43*	2.65 *0.63*	0.44 *0.83*	0.55 *0.82*	15.57 *0.21*	2.08 *0.67*	2.93 *0.60*	11.27 *0.29*	0.12 *0.91*
DP47				-	0.29 *0.88*	0.01 *0.98*	6.61 *0.44*	12.86 *0.27*	0.31 *0.87*	0.00 *1.00*	1.04 *0.76*	0.00 *1.00*	8.59 *0.37*
DP49					-	3.23 *0.61*	0.04 *0.95*	0.50 *0.84*	10.35 *0.35*	3.85 *0.58*	6.78 *0.45*	8.24 *0.40*	0.02 *0.96*
DP50						-	0.00 *1.00*	0.00 *1.00*	1.55 *0.71*	19.00 *0.18*	0.22 *0.89*	2.60 *0.63*	0.00 *1.00*
DP55							-	3.35 *0.58*	1.23 *0.73*	0.00 *1.00*	3.20 *0.58*	0.00 *1.00*	7.57 *0.69*
DP57								-	0.83 *0.78*	0.00 *1.00*	2.46 *0.64*	0.00 *1.00*	7.57 *0.40*
DP64									-	2.02 *0.68*	5.47 *0.48*	7.88 *0.39*	0.00 *1.00*
DP65										-	0.09 *0.93*	3.87 *0.56*	0.00 *1.00*
DP68											-	0.73 *0.79*	1.42 *0.71*
DP78												-	0.02 *0.97*

Comparison with the physical map

With the exception of the insertion, our map agrees well with the physical map shown in figure 2. Segment orders are identical and most distances agree well, with the exception of those involving DP49 and DP65 (which are error-prone due to the complicating effect of the duplication). Apart from any systematic errors (see below), distance estimates become more accurate as more samples are analysed. In practice, however, the number of samples analysed is generally determined by the need to define segment order unambiguously (12), the accuracy of the distance estimates being of secondary consideration.

The greatest distance errors occur toward the ends of our map, which are contracted relative to the physical map. This contraction is probably due to two factors. First, our model assumes that, in preparing the DNA samples, breakage by shearing occurs uniformly throughout the YAC. In practice, large molecules shear by repeatedly breaking into approximately equal halves (22), and hence breaks occur less frequently toward the ends of the YAC, leading to an underestimate of distances in these regions. Second, our method of data analysis may lead to a further contraction of the ends of the map: in effect, linkages to segments in the centre of the map tend to 'pull' the terminal segments inward, whilst there are no counterbalancing linkages tending to pull them

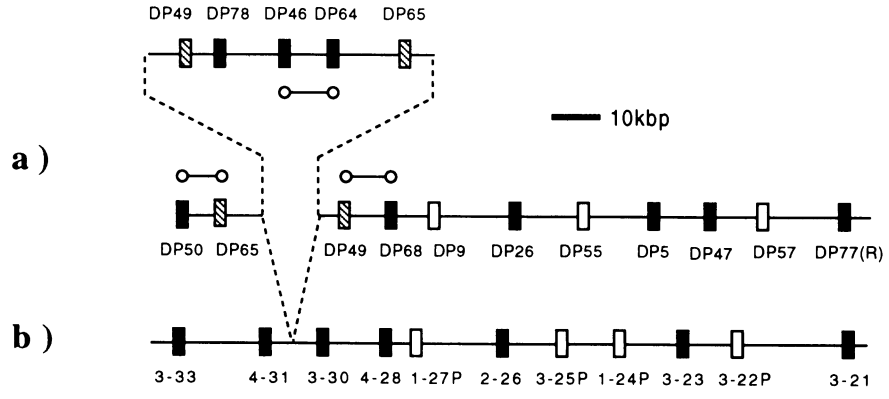


Figure 2. (**a**) Map of IGH_2, derived as described in the text. (**b**) Corresponding region from ref. 2. In our map, segments DP-49 and DP-65 (hatched) are duplicated; the indicated positions of these are approximate. Dotted lines delineate an insertion present in (a) and indicate the point of insertion on (b). Open boxes indicate pseudogenes; open circles joined by lines indicate phage lambda clones (4,19). The 14q telomere lies to the left and the immunoglobulin heavy-chain constant region genes to the right of the maps. Segments 3−32P and 3−29P have been omitted from (b); these are highly mutated pseudogenes not detected by the primers used to construct our map. DP-5 has a viable coding region and has been seen in rearranged form (G. Walter, unpublished); it is therefore shown here as a functional gene; the identical segment 1−24P on map (b) was reported as a pseudogene (2).

outward. Indeed, we have previously observed this phenomenon when mapping markers in genomic DNA broken by irradiation (12).

DISCUSSION

These results demonstrate the use of HAPPY mapping for rapidly ordering V_H segments within a yeast artificial chromosome. The most striking feature of the resulting map is the finding of an alternative haplotype, differing from that seen previously (2) by the insertion of the three consecutive segments DP-78–DP-46–DP-64. Three out of ten randomly chosen individuals appear to be homozygotes lacking this insertion, implying that both haplotypes are of roughly comparable frequency in the population. DP-46 has been seen in rearranged form in foetal liver (23), and in the synovial tissue of patients with rheumatoid arthritis (24). Furthermore, the insertion appears to be flanked by a duplication of two segments, DP-49 and DP-65; these are amongst the V_H segments most commonly seen in rearranged form (see (1) for references), perhaps a consequence of their duplication.

As improved techniques lead to the cloning of larger inserts in YACs, so it becomes easier to construct YAC contigs but correspondingly more difficult to map sequences within a single YAC. HAPPY mapping should prove a useful addition to the available repertoire of mapping techniques, particularly where close homologies between related sequences impede the use of other methods. Moreover, since the method can be applied directly to genomic DNA over a wide range of sizes (12), we believe that HAPPY mapping will prove valuable in investigating and monitoring suspected chimerism in YACs.

Finally, we note that the dynamic nature of the V_H locus during evolution, and its consequent insertion/deletion polymorphism, suggests that it may be necessary to construct maps of a number of haplotypes before this region of the genome is fully understood.

ACKNOWLEDGEMENTS

We thank Michael Neuberger for providing unpublished information, Rakesh Anand of ZENECA pharmaceuticals for screening their YAC library and providing the clone IGH_2, and K.Dear for proof-reading the manuscript. This work was supported by the MRC human genome mapping programme.

REFERENCES

1. Tomlinson,I.M., Walter,G., Marks,J.D., Llewelyn,M.B. and Winter,G. (1992) J. Mol. Biol., **227**, 776–798.
2. Matsuda,F., Shin,E.K., Nagaoka,H., Matsumura,R., Haino,M., Fukita,Y., Taka-ishi,S., Imai,T., Riley,J.H., Anand,R., Soeda,E. and Honjo,T. (1993) Nature Genet., **3**, 88–94.
3. Kodaira,M., Kinashi,T., Umemura,I., Matsuda,F., Noma,T., Ono,Y. and Honjo,T. (1986) J. Mol. Biol., **190**, 529–541.
4. Lee,K.H., Matsida,F., Kinashi,T., Kodaira,M. andHonjo, T. (1987) J. Mol. Biol., **195**, 761–768.
5. Shen,A., Humphries,C., Tucker,P. and Blattner,F. (1987) Proc. Natl. Acad. Sci. USA., **84**, 8563–8567.
6. Humphries,C.G., Shen,A., Kuziel,W.A., Capra,J.D., Blattner,F.R. and Tucker,P.W. (1988) Nature, **311**, 446–449.
7. Buluwela,L. and Rabbitts,T.H. (1988) Eur. J. Immunol., **18**, 1843–1845.
8. Croce,C.M., Shander,M., Martinis,J., Cicurel,L., D'Ancona,G.G., Dolby,T.W. and Koprowski,H. (1979) Proc. Natn. Acad. Sci. U.S.A., **76**, 3416–3419.
9. Hobart,M.J., Rabbitts,T.H., Goodfellow,P.N., Solomon,E., Chambers,S., Spurr,N. and Povey,S. (1981) Ann. Hum. Genet., **45**, 331–335.
10. Berman,J.E., Melliks,S.J., Pollock,R., Smith,C.L., Suh,H., Heinke,B., Kowal,C., Surti,U., Cantor,C.R., and Alt,F.W. (1988) EMBO J., **7**, 727–738.
11. Walter,M.A., Surti,U., Hofker,M.H. and Cox,D.W. (1990) EMBO J., **9**, 3303–3313.
12. Dear,P.H. and Cook,P.R. (1993) Nucleic Acids Res., **21**, 13–20.
13. Anand,R., Riley,J.H., Butler,R., Smith,J.C. and Markham,A.F. (1990) Nucleic Acids Res., **18**, 1951–1956.
14. Rabbitts,T.H., Forster,A., Baer,R. and Hamlyn,P.H. (1983) Nature, **306**, 806–809.
15. Dear,P.H. and Cook,P.R. (1991) Biochem. J., **273**, 695–699.
16. Sambrook,J., Fritsch,E.F., and Maniatis,T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour.
17. Perry,D.J. and Carrell,R.W. (1989) Mol. Biol. Med., **6**, 239–243.
18. van Dijk,K.W., Milner,L.A., Sasso,E.H. and Milner,E.C.B. (1992) Proc. Natl. Acad. Sci. U.S.A., **89**, 10430–10434.
19. Sasso,E.H., van Dijk,K.W., Bull,A., van der Maarel,S.M., and Milner,E.C.B. (1992) J. Immunol., **149**, 1230–1236.
20. van der Maarel,S., van Dijk,K.W., Alexander,C. M., Sasso,E.H., Bull,A. and Milner,E.C.B. (1993) J. Immunol., **150**, 2858–2868.
21. Marks,J.D., Tristrem,M., Karpas,A. and Winter,G. (1991) Eur. J. Immunol., **21**, 985–991.
22. Levinthal,C. and Davison,P.D. (1961) J. Mol. Biol., **3**, 676–693.
23. Schroeder,H., J., Hillson,J. L. and Perlmutter,R. M. (1987) Science, **238**, 791–793.
24. Pascual,V., Randen,I., Thompson,K., Sioud,M., Forre,O., Natvig,J. and Capra,J. D. (1990) J. Clin. Invest., **86**, 1320–1328.