Chapter 3

Using large insert clones to construct contigs:

The development of fluorescent fingerprinting

3.1 Introduction

The construction of sequence-ready maps of overlapping sets of bacterial clones has been one of the central components of large-scale genomic sequencing. Map construction is reliant upon techniques that are able to accurately assemble bacterial clone contigs at a depth that will facilitate the identification of a set of minimally overlapping clones. Restriction digest fingerprinting has proved to be a robust method that can be used as a means of producing accurate fragment patterns to allow the overlapping relationship between genomic clones to be determined accurately, as well being a technique that is amenable to increases in scale and speed for generation of clone data.

The approach of overlapping sets of random cloned DNA was first used to assemble contigs of cosmids or bacteriophage lambda clones of the *Caenorhabditis elegans* (Coulson *et al*., 1986) and *Saccharomyces cerevisae* (Olson *et al*., 1986, Riles *et al*., 1993) genomes, respectively. The basic premise of these projects was to generate overlaps between clones that share a statistically significant number of restriction fragments, where the shared fragments representing the region of overlap. Restriction digest fingerprinting has also been applied to the construction of sequence-ready maps within regions of the human genome (Carrano *et al*., 1989, Heding *et al*., 1992, Taylor *et al*., 1996). The inception of large insert bacterial clones has greatly facilitated the construction of maps of larger chromosomes and genomes. Large insert clones BACs and PACs are maintained in single to low copy number, (like fosmids), which means that a smaller percentage of the clones are likely to be unstable or to lose any cloned material. Additionally, a smaller

fraction of the genome is likely to be unstable in the host *E. coli*. PAC and BAC clones can be manipulated in the same way as cosmids and fosmids and, because of their larger insert size, generate genomic coverage more rapidly. As sequencing projects increase in scale (e.g. from *C. elegans* to human) there is a need to adapt existing fingerprinting techniques to the analyses of genome scale clone resources, to improve upon through-put and to incorporate increased levels of safety and automation.

This chapter outlines the evaluation of the first application of restriction digest fingerprinting to large insert clones by the construction of a 1.4 Mb contig across a region of chromosome 13q12 that includes the breast cancer susceptibility gene *BRCA2*. The chapter also details the development and testing of fluorescent fingerprinting, describes an assessment of the technique in comparison to other fingerprinting methods and discusses its application to the large scale generation of sequence-ready bacterial clone maps.

3.2 Large insert clones

3.2.1 Application of restriction enzyme fingerprinting to large insert clones

Linkage analysis in a series of breast cancer families localised the putative BRCA2 gene to a 3cM interval on 13q12-13. The work described in this chapter therefore contributed to the search for the BRCA2 gene, as well as providing a first opportunity to test the suitability of large insert bacterial clones (specifically PACs) as a resource for map assembly. The clones were identified from a whole genomic library by hybridising the *Alu*-PCR products of five yeast artificial chromosomes (YACs) (Burke *et al*., 1987) from a suggested CEPH YAC tiling path (Cohen *et al*., 1993) (figure 3.1a) and nine genetic markers (from the region) to a total genomic PAC library (see chapter 2.8.2) (initial library screening was done in collaboration with Richard Wooster). All clones were restriction digest fingerprinted (Coulson *et al*., 1986) using the established radioactive *Hin*d III/*Sau*3A I procedure and assembled into contigs using image editing (http://www.sanger.ac.uk/Software/Image) and data manipulation and analysis programs (Soderlund *et al*., 1997). Two separate walking procedures were employed to confirm overlaps within contigs and to bridge the remaining five gaps (in conjunction with Richard Wooster). Clones at the ends of contigs were used either as a template for SP6 or T7 primers vector specific primers (Ragoussis and Olavesen, 1997) or for the production of inter-*Alu* PCR fragments which were hybridised back to the library and to clones previously identified. The identification of new overlaps and the extent of overlap between clones by restriction fingerprinting was confirmed by the hybridisation results. Figure 3.1

displays the complete contig of the critical region containing 49 PACs and the SP6/T7 and *Alu* -PCR hybridisation probes and results. A minimum set of fourteen clones was chosen from the contig, representing the central region D13S260 to sls234, based on hybridisation and fingerprinting overlaps. The absolute size of the region represented by this minimum clone set was later determined to be 1.4 Mb based on complete genomic sequence data

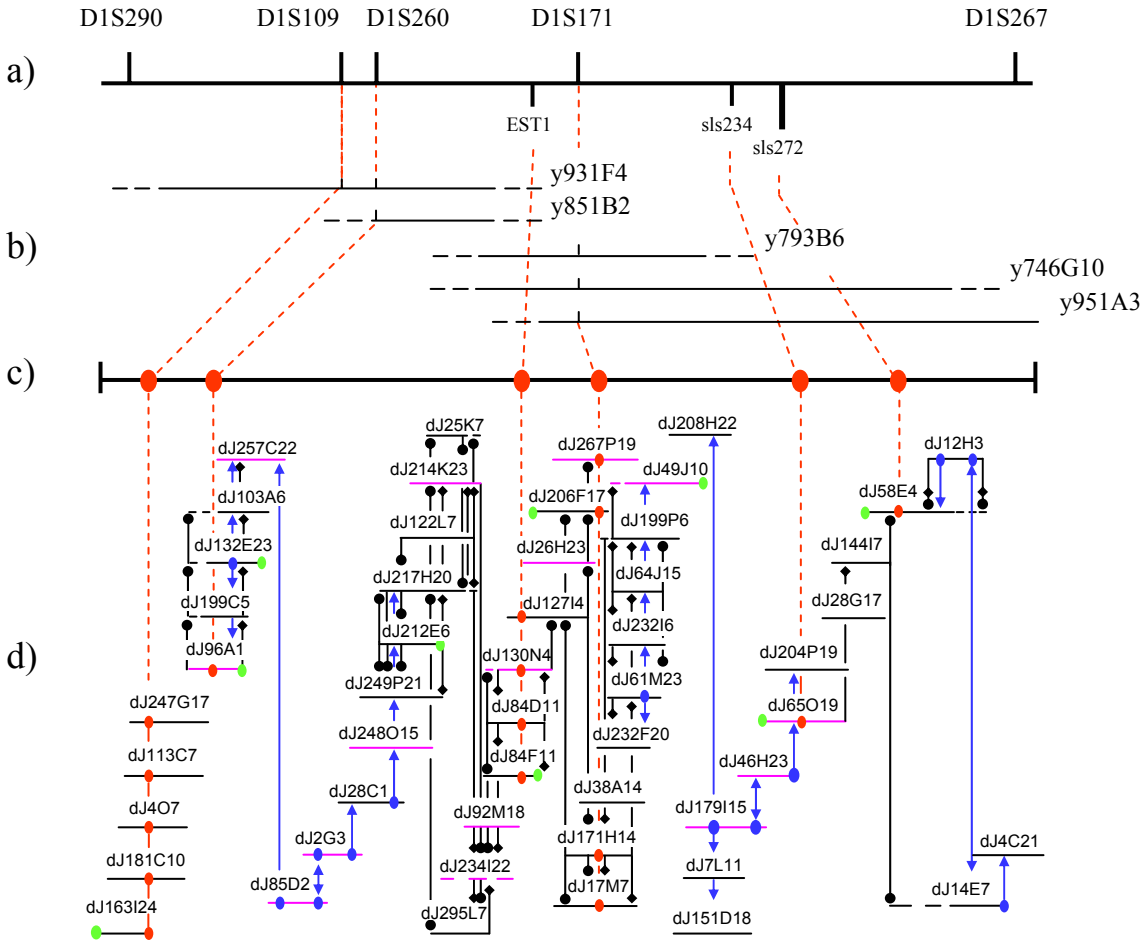

Figure 3.1: A representation of the mapping of BRCA2 region.

a) Genetic markers used in the identification of the putative region; b) YAC minimum set; c) Extent of PAC contig coverage; d) Fingerprint contig across the BRCA2 region; clones isolated by ● *Alu*-PCR hybridisation; ● clones fluorescent *in situ* hybridised; ⬇ T7 🌡 SP6 end probes; ▬ sequencing minimum set (234I22 sequenced as a PCR product)

3.2.2 Validation of PAC inserts

Given that the PACs were a new cloning system it was important to examine the integrity of the large insert PAC clones. Contiguous sequence from the 1.4 Mb contig provided an opportunity to compare the fingerprint patterns obtained empirically with those expected from the finished sequence. An additional check was possible by cross-comparison of fingerprints of regions in common between overlapping clones Four hundred and ninety two ordered *Hin*d III fragments were produced from a 'virtual' digestion of the 1.4 Mb of submitted sequence. The virtual fragments were compared to corresponding bands of the *Hin*d III fingerprints from 37 clones completely overlapping the chosen minimum set (figure 3.2, for example). The resolution of agarose gels limited the accurate sizing of bands to those over 500 bases (figure 3.2b). Occasionally bands of similar size, migrating in the same position of the gel, were not distinguishable. Figure 3.3 displays the band content of the *BRCA2* clones from the contig as compared to the virtual bands derived from the sequence. Bands of less than 500 bp, for which the size could not be accurately determined, accounted for 17% (82/492) of sequence fragments whilst missing bands (see

figure 3.3, red boxes) accounted for 2% (22/1091) of all fingerprint bands. These missing bands do not indicate the presence of large rearrangements within the PACs as no additional bands that could have resulted from the rearrangement were observed in overlapping clones. The absence of bands at the same location of overlapping clones (blue arrow in figure 3.3) may indicate a genuine restriction fragment length polymorphism within the PAC library. The missing bands within 199C5, 257C22 and 103A6 were large enough to be accurately sized on the agarose gel and were contained within overlapping clones.

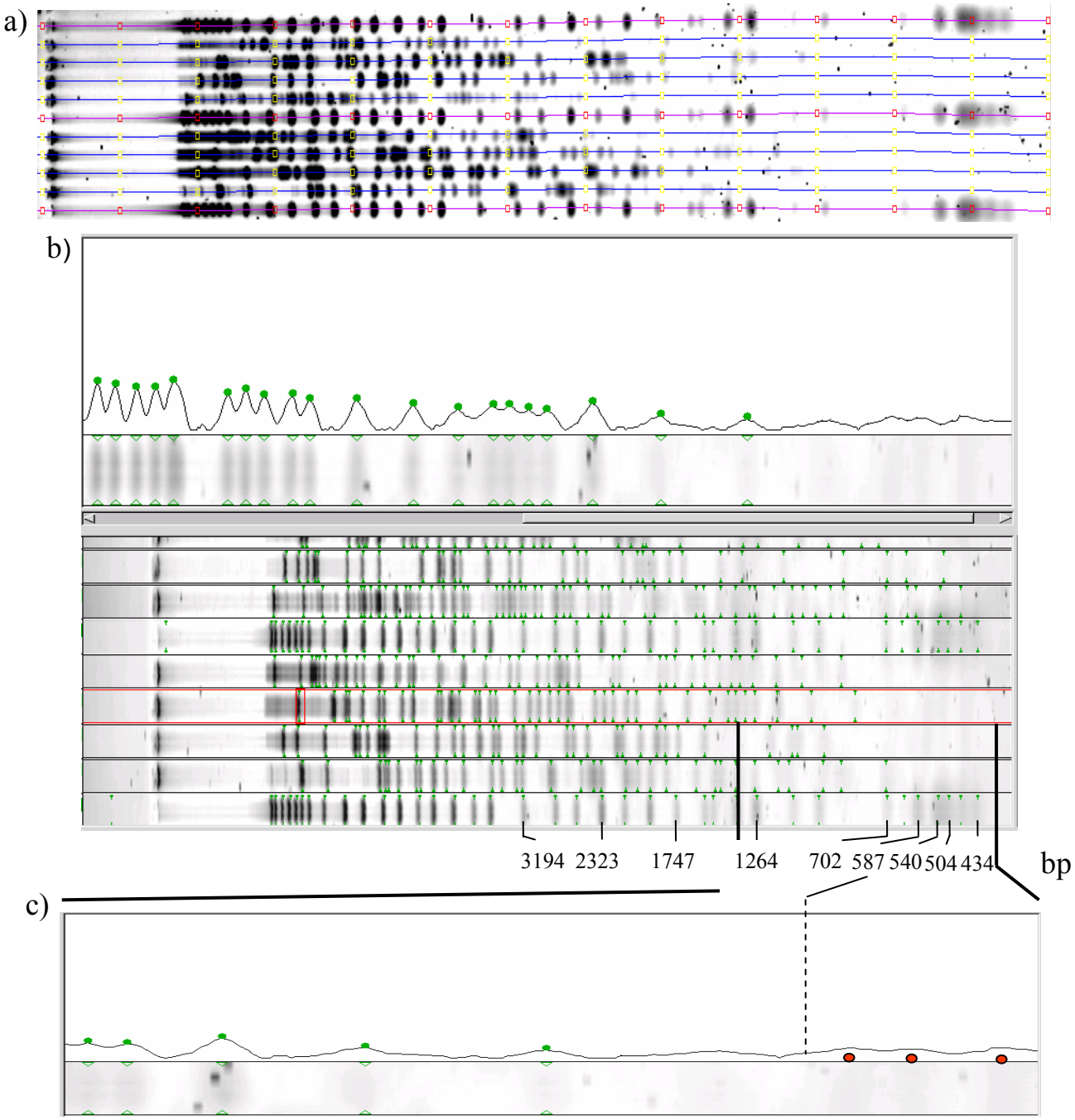

Figure 3.2: An agarose gel fingerprint of large insert bacterial clones in IMAGE. a) depiction of lane lock of 8 clones (yellow boxes) and 3 marker lanes (red boxes). b) data entry phase with edited bands indicated by green pips, also represented are marker lanes

with a sub-set of sized fragments in base pairs. c) zoomed region of clone 3 indicating restriction fragments below 500 bp in size (red dots).

Figure 3.3: A comparison of *Hin*d III fingerprint fragments and genomic sequence for the BRCA2 contig, top left to bottom right. Black boxes indicate when a band contained within the genomic sequence is accounted for by a fingerprint band, grey boxes where a band is ambiguous on the agarose gel, clear boxes where a band derived from the sequence

falls outside of the resolution of the agarose gel, red boxes where the band was absent from the restriction fingerprint. The blue arrow indicates a possible restriction fragment length polymorphism.

3.3 Fluorescent fingerprinting

The development of a hierarchical clone by clone approach by the Human Genome Project (HGP) for the large-scale generation of human genomic sequence (see chapter 4) required the development of a restriction fingerprinting technique that was high throughput, accurate and safe. For these reasons the application of fluorescent sequencing technologies to restriction fingerprinting was investigated.

The first experiments utilized fluorescent dideoxy molecules, used in terminator sequencing, to end label fingerprint fragments. The aim was to emulate the dual restriction digest and end labelling technique developed by Coulson *et al.,* (1986) for the construction of a physical map of the C. elegans genome. In this technique, a radioactive reporter molecule, [α -^{32}P] deoxy-adenosine-tri-phosphate (dATP), is complementarily conjugated via a DNA polymerase (AMV RT) with the first base in the 5' overhang of *Hin*d III restriction digest. A dideoxy guanine-tri-phosphate (ddGTP) is then incorporated at the second base to ensure irreversible completion of the labelling reaction, prior to a second restriction digest with *Sau*3A I. Fingerprint fragments were resolved on a denaturing polyacrylamide gel and migration values of the fragments determined by comparison to a known marker run interposed between fingerprints on the gel. The marker is produced by digesting lambda DNA with *Sau*3A I, filling the first position of the 5' overhang with

dGTP, labelling with radioactive [α -^{33}S] dATP and utilizing ddTTP to prevent exonuclease removal of the reporter molecule.

3.3.1 Fluorescent labelling of cosmid DNA

The first fluorescent fingerprint experiment was designed to mimic the radioactive protocol. DNA minipreps of two chromosome 22 cosmid clones (see chapter 2.4.3) were digested with *Hin*dIII and Sequenase used to incorporate the unlabelled dATP and fluorophore conjugated ddGTP. A phenol/chloroform extraction step was introduced to remove latent fluorescent dye (soluble in phenol) from the labelling mix prior to gel running and data capture on an ABI 373 using ABI Prism sequencing software. The results from this initial experiment showed a stochastic relationship between DNA concentration and signal strength of the fingerprint fragment peaks. Though the data collected from the gel run identified only smaller restriction fragments it determined that fragments had been successfully labelled and identified a noticeable salt front and a large peak corresponding with unincorporated dye (figure 3.4).

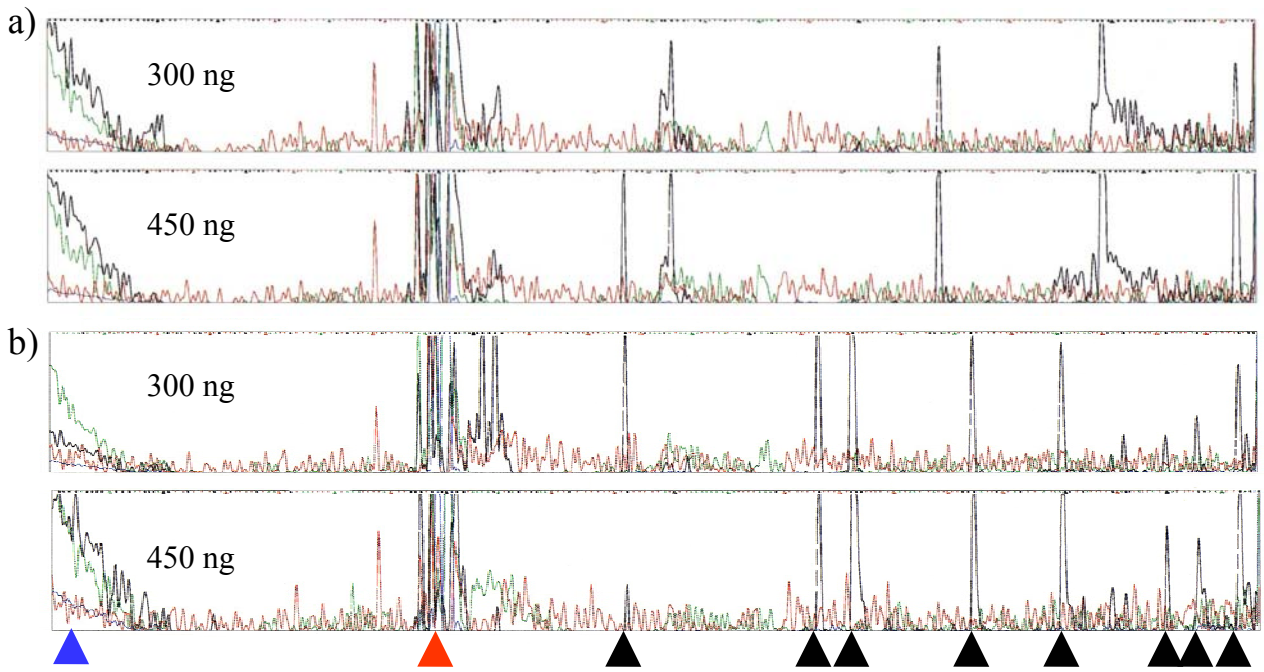

Figure 3.4: The result of the first fluorescent fingerprinting experiment. Two cosmid clones (a and b) of known concentration (shown), were digested with *Hin*dIII and *Sau*3A, labelled with ddGTP dye-terminator run on an ABI 373 sequencing machine. Labelled fragments of different DNA concentrations were collected using Prism sequencing software. The blue arrow denotes salt/buffer front, the red arrow denotes latent dye-terminator present in fingerprint reaction and black arrows identify labelled restriction fragments.

Having proven that it was possible to generate and detect fluorescently labelled fingerprint fragments, the next aim was to repeat the cosmid labelling experiment and to replicate the labelling of the radioactive lambda marker which would enable a comparison to the radioactive cosmid fingerprint to be made and therefore determine if all cosmid restriction

fragments were being collected (figure 3.5). Lambda DNA was digested with *Sau*3A I and labelled by filling the 5' overhangs with dATP, dGTP and fluorescent dideoxy thymine-tri-phosphate (ddTTP) using Sequenase. This protocol necessitated incorporation of the labelled dNTP at the third position of the 5' overhang which would be spectrally distinct to nucleotides used to label cosmids in the first and second positions.

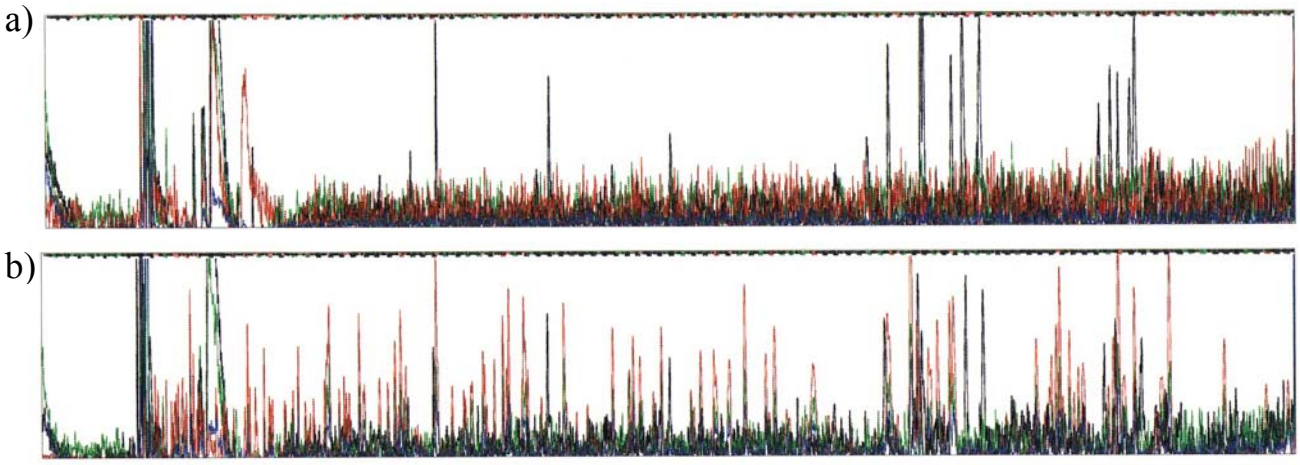

Figure 3.5: Fluorescent fingerprint data collection using an extended run time. a) fragments were labelled as previously but the run time was extended. b) Replications of the labelling of radioactive lambda marker using a spectrally distinct fluorophore to label a Sau3A I digest of lambda DNA.

3.3.2 Residual dye removal

A transition was made to ABI Genescan software which would allow us to separate fingerprint and marker traces from within a single lane on the gel. The presence of latent dye remaining in the fingerprint reaction could affect data production. Small fingerprint

fragments migrating adjacent to the dye front may be masked by dye signal or the dye may affect the resolution of separating fragments. Comparisons were made between techniques to remove the unincorporated dye, namely phenol/chloroform extraction, molecular sieving using Sepharose beads (figure 3.6a and b) and ammonium acetate/ethanol precipitation (figure 3.6c). It was noted that all dye removal techniques resulted in a reduction of fragment signal. Phenol/chloroform extraction resulted in the largest loss of signal (60-70%), ammonium acetate (30-40%) and P10 Sepharose beads (20-30%). Though the P10 beads resulted in the least amount of signal loss, ammonium acetate extraction could more easily be incorporated into the fingerprinting procedure without the introduction of an additional extraction step.

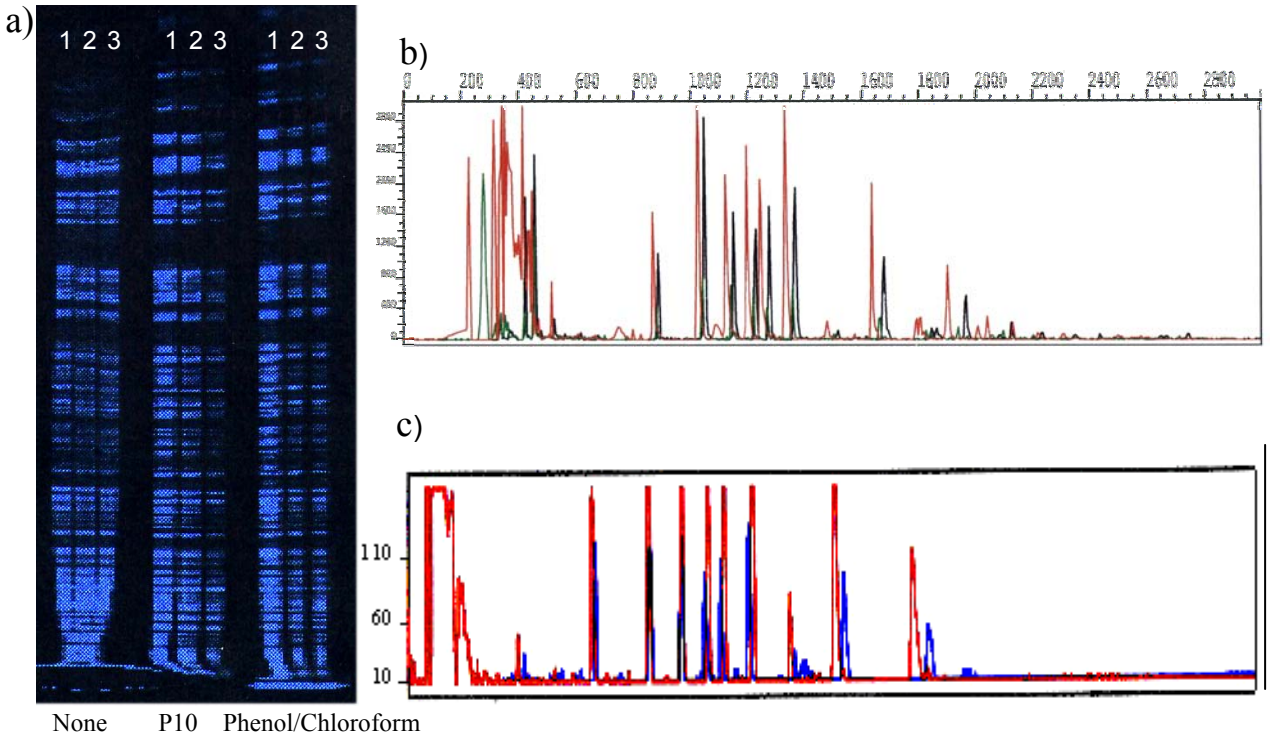

Figure 3.6: A comparison of labelled fragments when investigating removal. a) *Sau*3A I digest of 100 ng (1), 200 ng (2), and 500 ng (3) of lambda DNA. DNA digested with Sau3A I was labelled with ddTTP dye-terminator and data displayed and analysed using Genscan software. b) signal intensities of 500ng sample tracks displaying differences between dye removal techniques. Red traces correspond to fingerprint reaction with no dye removal, black traces using P10 Sepharose beads and green traces phenol/chloroform extraction. c) a comparison of ammonium acetate dye extraction (blue) and fingerprint reaction from which no dye was removed (red).

3.3.3 First Position Labelling

Previously, it was shown that fluorophore labelled dNTPs could be incorporated in positions 1 and 2 of *Sau*3A I digested lambda and position 2 of a *Hin*d III restriction digests of cosmid DNA (figure 3.7a). However, incorporation of a dideoxy-nucleotide reporter molecule in the first position of bacterial clone fingerprints would negate the effect of poor fragment labelling through incomplete dinucleotide incorporation which would prevent efficient labelling with a dideoxy-nucleotide in 2nd or 3rd positions. To test the efficiency of first position labelling, lambda DNA at different concentrations was digested with *Hin*d III and labelled with fluorescent ddATP, figure 3.7. The labelling of restriction fragments with a fluorophore conjugated ddA in the first position of *Hin*d III digest, without the requirement of deoxy-nucleotides, means that DNA could be digested in parallel with *Sau*3A I without these fragments being labelled (figure 3.8a). Results indicated that labelling of relatively low concentrations of DNA with very low background could be achieved using first position labelling. Custom synthesis of ddATP spectrally distinct derivatives, on which one of three differently fluorescing fluorophores were attached (figure 3.7c), permitted multiplexing of three different fingerprint samples within one lane on a gel (yellow, blue and green). Identical sizes of the conjugated fluorophores meant no correction was required during data collection to permit fragment comparison. Lambda DNA digested with *Hin*d III/*Sau*3A I was again used as the control for testing the efficiency of the three ddATP fluorophores.

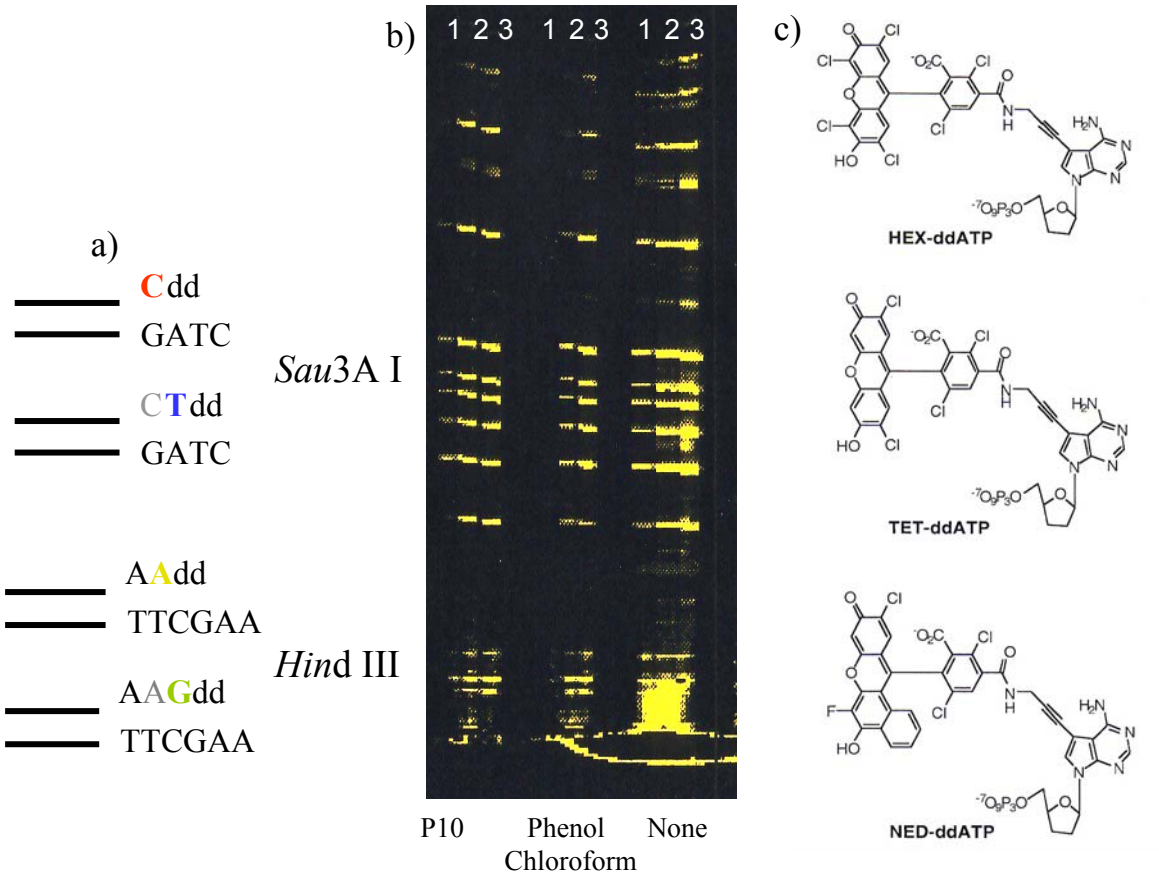

Figure 3.7: Labelling with spectrally distinct fluorophores. a) Labelled nucleotides of spectrally distinct fluorophores attached to the 5' over-hangs of *Sau*3A I and *Hin*d III restriction fragments in positions 1, 2 and 3. Also shown are unlabelled nucleotides (grey) b) First position labelling of lambda DNA. 100ng (1), 200ng (2), and 500ng (3) of DNA are digested with *Hin*d III, labelled using ddATP dye-terminators dye prior to subsequent digestion with *Sau*3A I. Signal strength of labelled fragments is compared between dye extraction techniques, P10 beads and phenol/chloroform. P10 dye extraction of 100 ng of lambda DNA give clear signal above background indicating that first position labelling provides a clear alternative to using dye-terminators and dNTPs for fill-in labelling. c) Spectrally distinct dye-terminators molecules which were custom conjugated onto ddATP and used for separate labelling and multiplexing of fingerprint samples. d) Titration of ddATP dye-terminators using chromosome 22 cosmid DNA.

3.3.4 One-step reaction

The next modification to be made to the fingerprinting protocol was to initiate a one step dual enzyme digest and labelling reaction based on the protocol published by Tang *et al*., (1996). This adaptation would reduce the time required for a fingerprint reaction from $3^{1/2}$ hours to1 hour prior to latent dye extraction. Initial results looked very promising with only a small decrease in the signal of the one-step reaction compared to that of the two-step (figure 3.8). The *Hin*d III/*Sau*3A I combination used to generate fluorescent fingerprints were suitable for the one-step reaction as digest with *Hin*d III results in a T in the first position available to incorporate a fluorescent ddATP whilst a *Sau*3A I digest would only incorporate a ddATP in the second position of in the presence of a dGTP Fingerprints were generated in the presence and absence of unlabelled ddGTP in the reaction mix to determine whether *Sau*3A I fragments were being labelled in the one-step reaction in the presence of residual dGTP (figure 3.9a) or misincorporation of ddATP at first position. No discernable partial or background labelling was noted in either experiment.

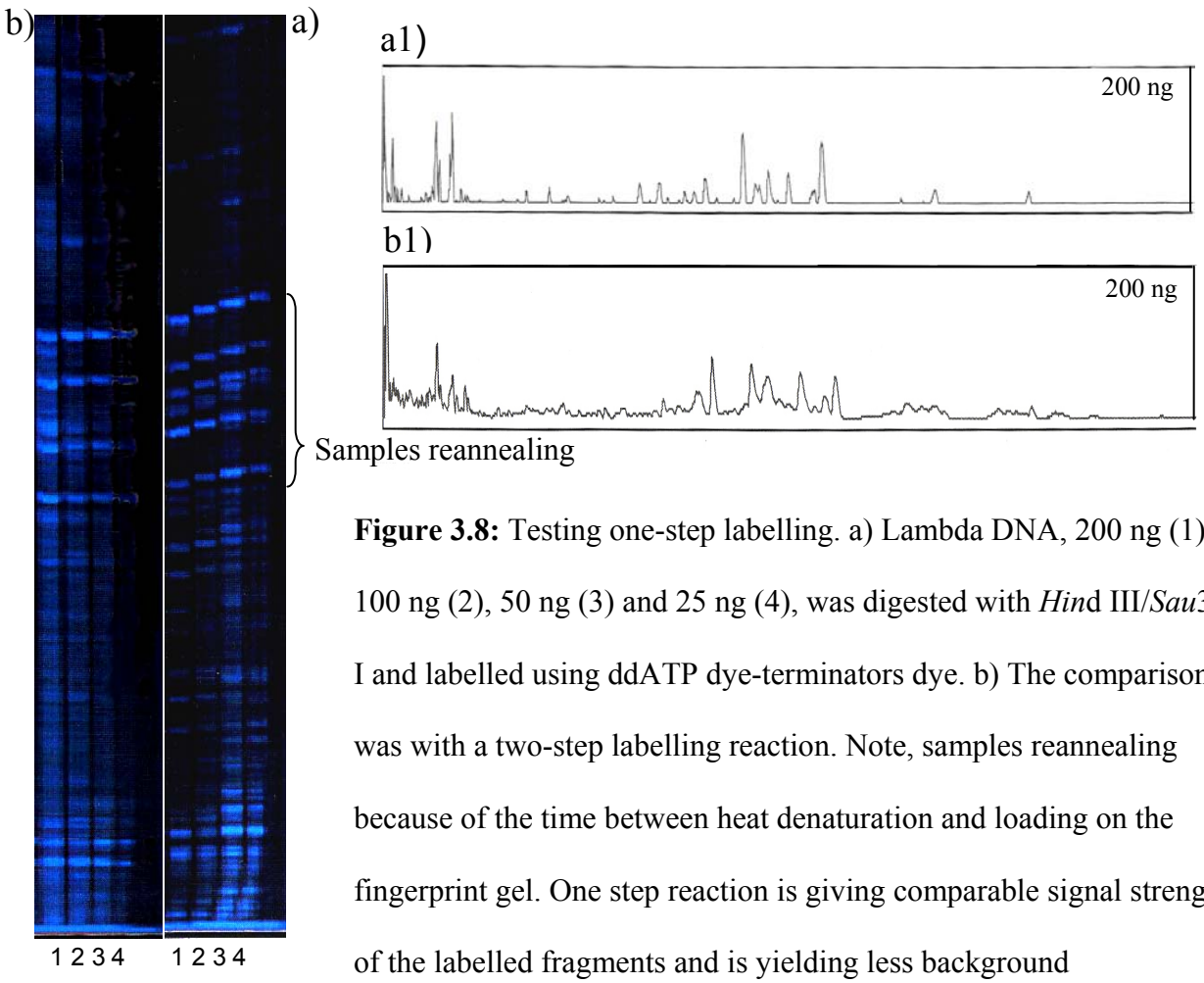

Figure 3.8: Testing one-step labelling. a) Lambda DNA, 200 ng (1), 100 ng (2), 50 ng (3) and 25 ng (4), was digested with *Hin*d III/*Sau*3A I and labelled using ddATP dye-terminators dye. b) The comparison was with a two-step labelling reaction. Note, samples reannealing because of the time between heat denaturation and loading on the fingerprint gel. One step reaction is giving comparable signal strength of the labelled fragments and is yielding less background

3.3.5 DNA prep modifications

The protocol used for isolating cloned bacterial DNA for radioactive and initial fluorescent fingerprinting experiments was based upon the microprep procedure published by Gibson and Sulston (1987). Modifications to this procedure were made to improve DNA yield by utilizing Corning 2 μm filter bottom plates to remove the supernatant from large molecular weight protein precipitated by alkaline-lysis. Adding RNAse to T0.1E before final DNA resuspension replaced the need for RNA removal by lithium chrloride precipitation and in doing so removed an additional DNA precipitation step (figure 3.9b). A moderate increase in signal was observed from the protocol using DNA extracted by the filter prep method. This is presumed to result from a higher DNA yield of the F prep compared to the M prep, attributable to the increased volume of supernatant transferred during alkaline lysis and to the reduction in the number of precipitation steps required before final DNA isolation.

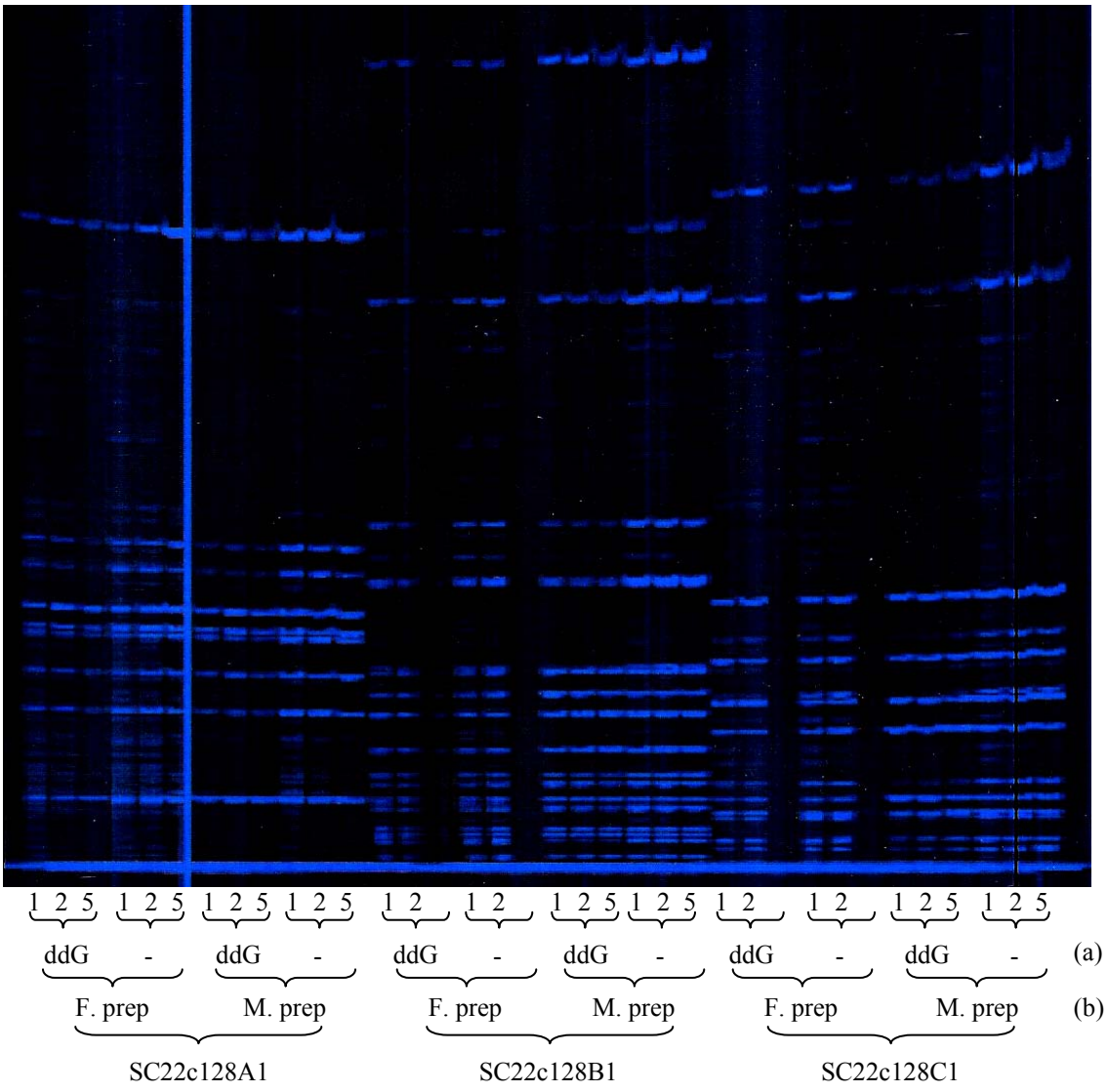

Figure 3.9: Testing one-step labelling and DNA prep protocols. a) Labelling of cosmid DNA (SC22c128A1, SC22c128B1, SC22c128C1) in a one step fingerprinting reaction in the presence and absence of unlabelled 1mM ddG. 1 µl, 2 µl and 5 µl of fingerprint reaction were loaded per lane. No discernable difference was observed in the signal strength or level of background labelling (*Sau*3A I labelling) between one-step and two-step labelling reactions. b) A comparison of filterprep (F) and microprep (M) protocols for the isolation of cosmid DNA samples (see methods). A moderate increase in signal strength above background was observed in the filterprep samples.

3.3.6 New size standard

The separation of radioactive fingerprints is achieved by running samples and lambda *Sau*3A I marker on a 40cm denaturing polyacrylamide gel until the loading dye reaches the bottom of the gel. This form of fragment separation generates an exponential-like distribution of fragment sizes. Lambda *Sau*3A I marker in this context provides even distribution along the length of the gel. Fluorescent fingerprints are also separated on a denaturing polyacrylamide gel but, as fragment data is collected in real time, separation is more linear with increased separation of larger fragments.

Fluorescently labelled lambda *Sau*3A I marker was unsuitable for fingerprint fragment size determination because of uneven distribution of bands along the length of the gel. Uniform distribution of marker bands is necessary to accurately assign migration values to fingerprint bands for overlap analyses. An important additional feature of an appropriate restriction enzyme would be the generation of a G in the first position of the overhangs of the digested fragments, as this would permit incorporation of a red ddC fluorophore allowing a marker to be run in the same lane as green, yellow and blue fluorophore labelled ddA fingerprints. Another important factor in selecting a replacement enzyme is

that its optimum temperature should be compatible with the conditions under which Taq FS polymerase end labels fingerprint fragments.

Two enzymes that fulfil the criteria of having a G in the first position for a labelled ddC nucleotide to anneal, *Bsa*J I and *Taq*α I, were labelled and run with ddT labelled lambda/*Sau*3A I to permit overlap comparison, figure 3.10. *Bsa*J I gave a more uniform distribution of marker fragments along the length of the gel, particularly in the 100 bp to 2000 bp region of the gel in which the majority of fingerprint fragments would migrate. Overlapping the *Sau*3A I with *Bsa*J I marker lanes clearly indicates the distribution differences between the two enzymes (figure 3.10).

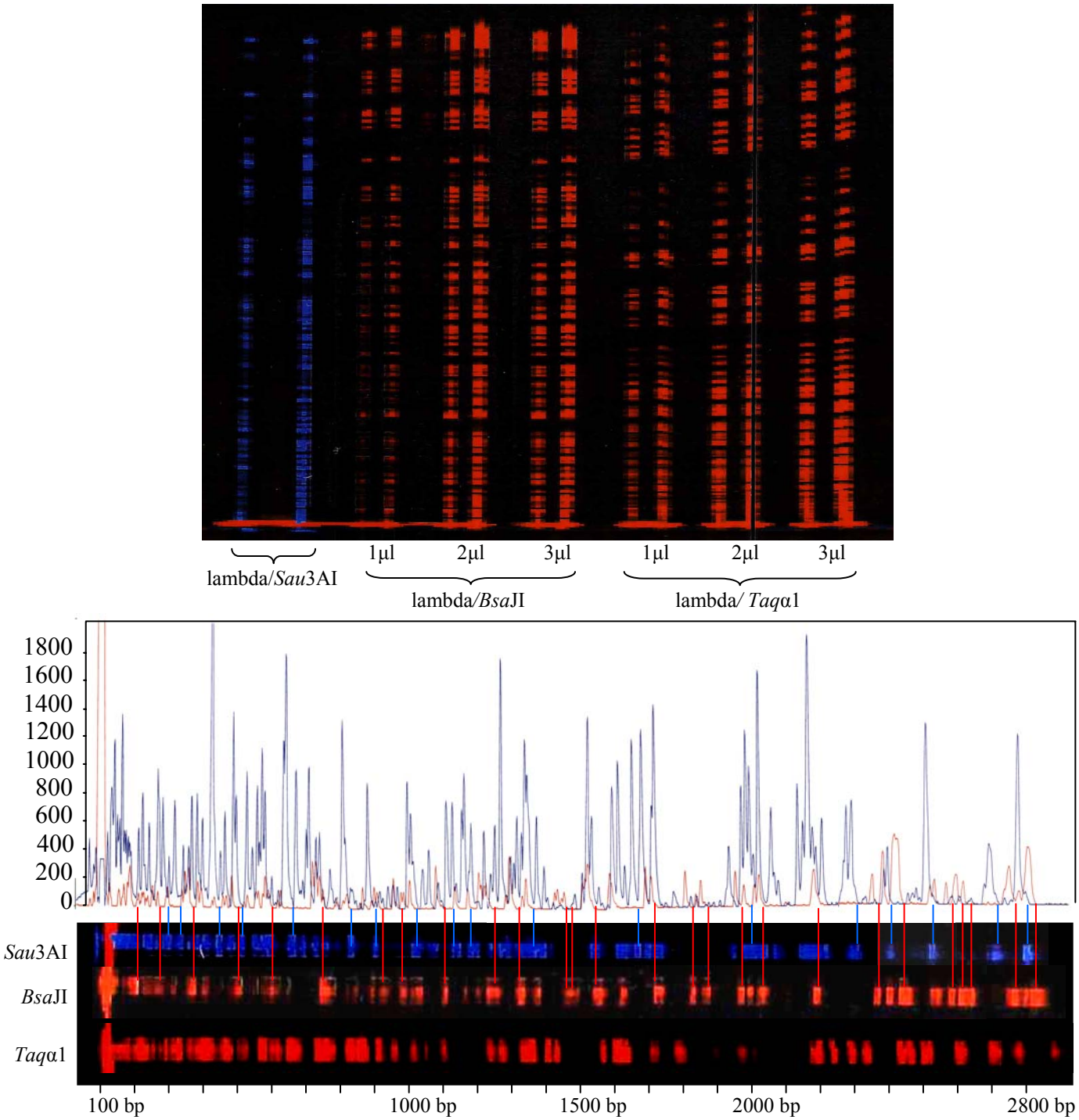

Figure 3.10: A comparison of labelled lambda digest fragments using *Sau*3A I, *Bsa*J I and *Taq*α 1 restriction enzymes. Lambda DNA labelled with fluorescent ddTTP (*Sau*3A I) and ddCTP (BsaJ I and *Taq*α 1) and loaded in different aliquots. b) BsaJ I generates a more uniform distribution of fragments in the 30 bp to 2.8 kb size range of in comparison to *Sau*3A I (which is used for generating the marker standard for radioactive fingerprints) and *Taq*α 1.

3.3.7 Data collection and processing

During protocol development data was collected on an ABI 373 and visualised using Genscan software version 2.0.2. Though the software was useful during the testing phase of the technique it was not designed to interpret the complexity of fingerprint or marker bands produced by the fluorescent fingerprinting. Instead, unprocessed gel images were collected using Prism Collection software (v1.1) running on an ABI 377 and downloaded onto a UNIX workstation. Data was edited using modified version of IMAGE before transfer into Fingerprinted Contigs (FPC) (Soderlund *et al*., 1997) for analysis. This eliminated all of the gel processing, autoradiography and scanning that is associated with radioactive fingerprinting.

3.3.8 Reproducibility

Labelling the marker with a different fluorophore to the fingerprint fragments permits the incorporation of a marker in each fingerprint lane. This facilitates greater accuracy of mobility determination of fingerprint fragments within the 30 bp to 2.8 kb size range than was previously possible. To test the accuracy of fragment analyses and reproducibility of the labelling procedure, a PAC clone, dJ163I24, was fingerprinted 10 times with each of the three ddATP dyes and compared (figure 3.11). All fingerprints were highly

reproducible. The migration values of seven fingerprint bands, representing an even distribution along the length of the gel, were compared between clones. In all cases examined, the variation of migration of bands in 30 independent fingerprints was ±0.26% (maximum observed: ± 11/4500 intervals), which is less than the tolerance used in the analysis.

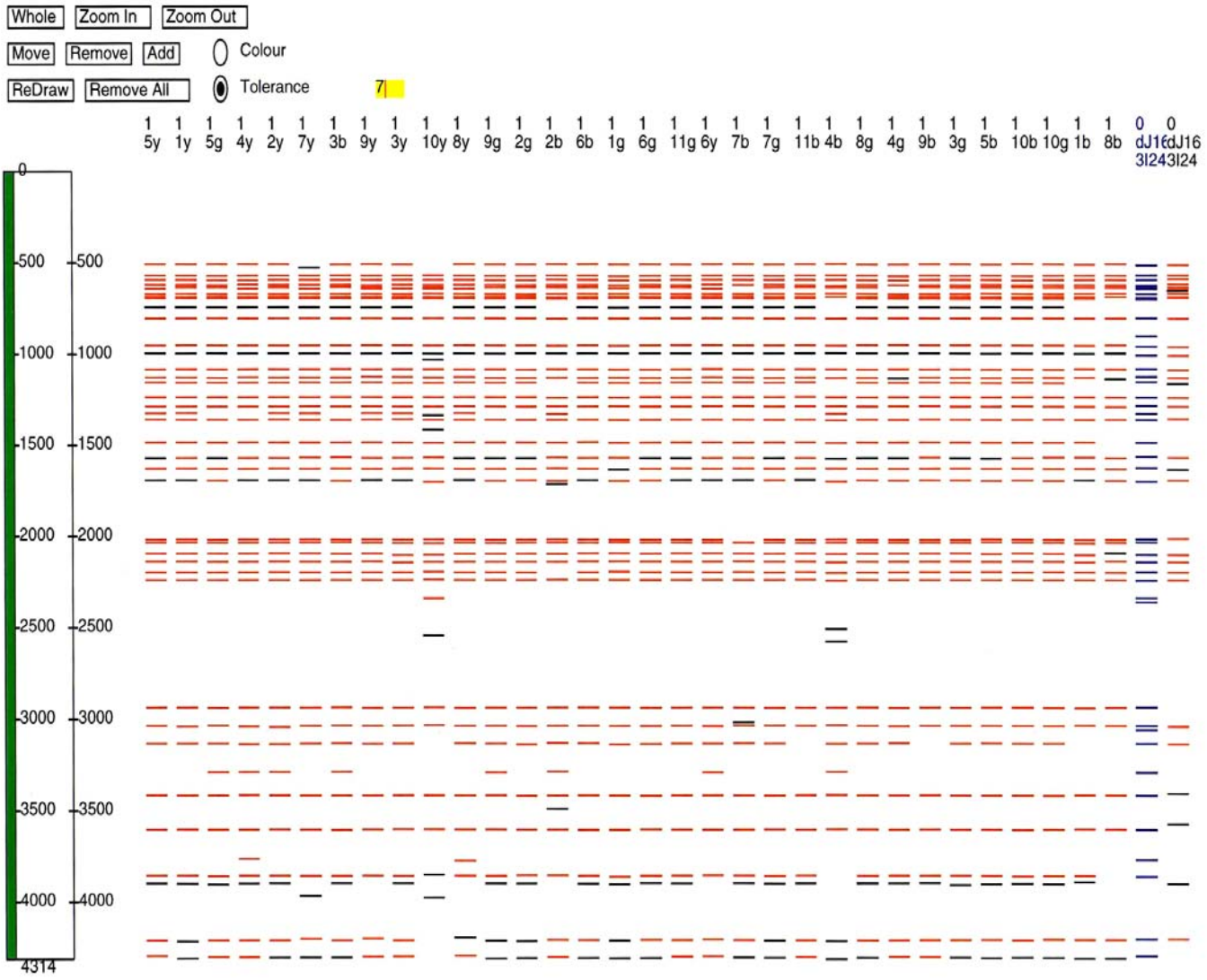

Figure 3.11: An FPC display of band labelling uniformity. 10 PAC, dJ163I24, duplicates were fingerprinted with one of each of the three ddATP fluorophores to test for labelling consistency, clone number followed by suffix b = ddATP-TET, y = ddATP-NED, g =

ddATP-HEX. Horizontal red lines correspond to fingerprint bands that are within the matching tolerance, as discussed in chapter 2.22.2, of the original fingerprint (blue bands).

3.3.9 Validation of fluorescent fingerprinting

Once the labeling and reproducibility of the fluorescent technique was established a comparison of radioactive, fluorescent and non-isotopic fingerprinting was performed using 49 PAC clones from the 1.4 Mb contig constructed as described previously (see figure 3.1). The aim was independently to assemble clones from the contig using the different fingerprinting techniques and compare the results with overlaps defined by 14 clones representing a minimum set of clones for which complete sequence was now available (figure 3.12a). Vertical arrows in (figure 3.12b) represent the overlaps that were not found between clones in each of the four methods. Figure 3.12c represents the contigs as constructed by the fluorescent fingerprinting method. The vertical arrows coincide with those in figure 3.12b, denoting overlaps that were not detected by fingerprinting alone. There was excellent agreement between the contigs assembled using either the radioactive or fluorescent *Hin*d III / *Sau*3A I data. Minor variations were observed where small overlaps between clones fell just outside the probability cut-off used. (These overlaps were originally detected by end-probe hybridisation, and later confirmed by the sequence).

There was also good agreement between the *Hin*d III / *Sau*3A I results and those of the *Hin*d III and *Eco*R I fingerprints resolved on agarose gels, but matches in the latter protocols required a higher statistical probability of overlap, i.e. smaller overlaps were not detected. This can be attributed in part to the fact that each agarose fingerprint pattern is divided into 2000 1 mm intervals along the length of a 20 cm lane as opposed to the 3400 and 4500 intervals of the radioactive and fluorescent fingerprints respectively. A smaller number of intervals implies that there is a greater probability that bands will occur randomly in the same bin and contribute to background matches. However, it should be pointed out that the two approaches are both useful in different ways. The *Hin*d III /*Sau*3A I method can detect smaller overlaps but samples 256 bp / 4096 bp (14%) of the insert DNA of a clone, and provides no information on the remainder of the sequence. By contrast the use of complete *Eco*R I, *Hin*d III or similar digests resolved on agarose gels, although achieving lower resolution, does display bands representing all the DNA of the cloned insert. This data is therefore more appropriate for verifying the integrity of the sequence compared to all genomic clones covering the region. It also confers the ability to look for possible deletions or other rearrangements that result in detectable size alterations in individual restriction fragments.

Where all four methods failed to detect overlaps between adjacent clones (figure 3.12, 1a – c), the overlapping bands were only contained within two clones. This problem would be resolved with greater contig depth. Details of analysis of overlaps which were not detected by individual methods are given in the figure 3.12 legend.

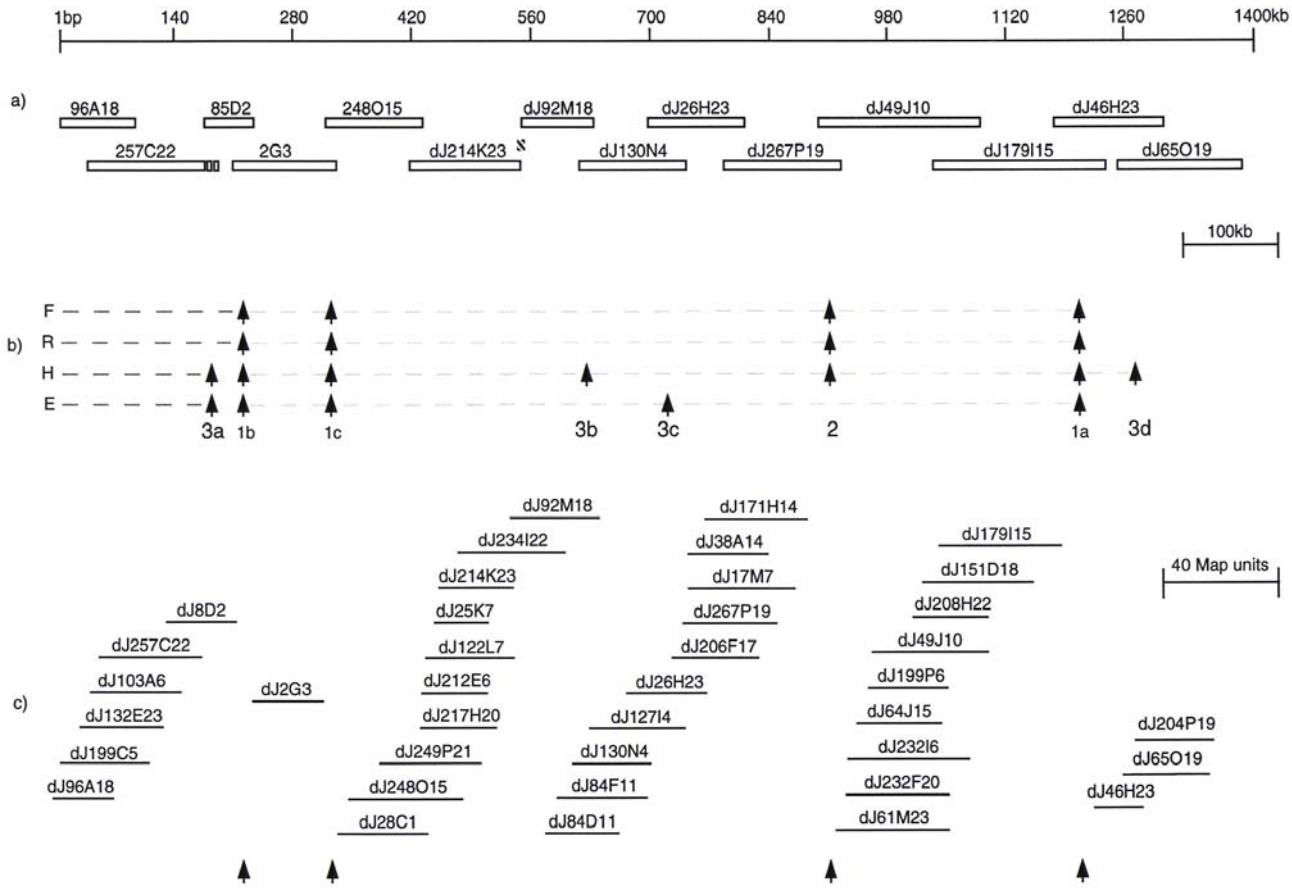

Figure 3.12: A comparison between fingerprinting methods of using 14 clones comprising a minimum tiling path. a) Finished sequence data were used to define the overlaps between

clones and absolute length of each of the clones, represented as clear boxes and scale in kb (except dJ257C22, where data were incomplete). b) Vertical arrows denote overlaps that were not detected by each method: (F) fluorescence, (R) radioactive; (H) *Hin*d III, and (E) *Eco*R I. As expected, there is variation based on recognition sites between the overlaps detected using *Hin*d III (b and d) and those established with *Eco*R I, c). Details of the analysis of undetected overlaps are as follows: (Overlap 1a, 62 kb) Nine *Hin*d III and 11 *Eco*R I fragments are present in this overlap on the basis of DNA sequence. Statistical probability on the basis of these bands shared between dJ179I15 and dJ46H23 is 1 × 10 4 (4) fluorescence, 1 radioactive, 4 *Hin*d III, and 5 *Eco*R I (whose probability cutoffs for all matches are 4, 5, 7, and 7, respectively). (Overlap 1b, 25 kb) Clone dJ2G3 is unattached. Three *Hin*d III or six *Eco*R I fragments shared with clone dJ85D2 are insufficient to score the overlap. (Overlap *1c,* 11 kb) Clone dJ2G3 is unattached. Four *Hin*d III or three *Eco*R I fragments shared with clone dJ248O15 are insufficient to detect the overlap. (Overlap 2, 24 kb) Overlap between dJ267P19 and dJ49J10 was not found by any of the *Hin*d III methods. Six *Hin*d III fragments are not sufficient to detect overlap. Arrows 3a-3d denote where only the agarose methods failed to find overlap. (Overlap 3a) Overlap between dJ257C22 and dJ85D2 could not be determined because of incomplete overlap data and therefore was not analyzed. (Overlap 3b, 18 kb) Four *Hin*d III fragments shared between dJ130N4 and dJ92M18 are not sufficient to detect overlap. (Overlap 3c, 44 kb) Six *Eco*RI fragments shared between dJ130N4 and dJ267P19 are not sufficient to detect overlap. (Overlap 3d, 55 kb) 16 *Hin*d III fragments between dJ46H23 and dJ65O19 should have been significant. The lack of overlap may be attributed to the migration of similar-sized fragments in the same location of the agarose gel and therefore not be representative of the

cloned insert. (*c*) Although all but 1 of the 49 clones (dJ2G3) fell into contigs, only 4 of the contigs that lie within the sequence interval are represented. Arrows denote where the overlaps between clones were not found.

3.4 Discussion

The construction of a 1.4 Mb contig on human chromosome 13q provided the first opportunity to test the application of restriction digest fingerprinting of large insert clones (PACs) for the construction of a physical map and their utilisation as a sequencing resource. Results indicated that they are stable, provide coverage to this extent, and few/no discrepancies between clones suggested little rearrangement. The existing *Hin*d III/*Sau*3A I fingerprinting proved suitable for the generation of PAC fingerprints therefore permitting integration of other types of smaller insert bacterial clones (Dunham *et al*., 1999, Deloukas *et al*., 2000).

The advent of large-insert bacterial clone libraries of PACs and BACs for the construction of high-resolution maps has facilitated sequencing and gene identification within chromosomal regions (Kudoh *et al*., 1997, Hubert *et al*., 1997, Guru *et al*., 1997). Construction of a bacterial clone map in either a regional or a chromosome wide study relies on initial identification of the clones by using hybridisation- or PCR-based markers to screen available libraries. The existence of overlaps between clones may be identified on

the basis of shared marker content, however, as a marker represents one point in the genomic DNA, no measurement of extent of overlap is obtained. In most regions of the human genome marker density is too low to achieve closure of all gaps by marker content alone. Alternative methods to define overlaps between bacterial clones include the hybridisation of end probes to clone arrays (Evans and Lewis, 1989), high resolution fluorescence *in situ* hybridisation (FISH) using DNA fibres (Wang *et al*., 1996) and matching of ends to complete insert sequences (Kupfer *et al*., 1995, Roach *et al*., 1995). End-probe hybridisations do not define the extent of overlap, also a limitation of whole cosmid-cosmid hybridisations (Xie *et al*., 1993), and all hybridisation-based methods employing undefined sequences can yield false positive signals because of cross hybridisation of repeat motifs. The end-sequencing strategy provides precise information on extent of overlap but relies on extensive prior investment in sequencing all clones in the library and also does not permit assembly of contigs prior to sequencing of the inserts. Fibre-FISH, using bacterial clones as probes and yeast artificial chromosomes (YACs) as a target, provides good estimates of both the extent of overlaps and size of gaps between clones but require the existence of a well-define YAC map across the region of interest and has yet to be scaled up. Fingerprinting, however, can detect overlaps where there is no marker available and gives an indication of the extent of each overlap.

Other approaches to fluorescent based fingerprinting have included ligation of fluorochrome-labeled oligonucleotide adaptors to *Eco*R I digests of cosmids (Lamerdin and Carrano 1993) and incorporation of fluorescent dideoxy terminators in 5' overhangs produced by endonuclease digestion (Ding *et al*., 2001). The former technique uses

agarose gels to resolve labeled fragments has been superceded by the VISTRA-green-stained agarose gel fingerprints (Marra *et al*., 1997), in this chapter, and the latter technique results in a complex fragment pattern that is not necessarily amenable to the generation of large scale physical maps.

The fluorescent fingerprinting technique discussed in this chapter utilises three spectrally distinct dye-ddA terminators to label the restriction fragments of three different *Hin*d III-*Sau*3AI endonuclease digests. The combination of three fingerprint samples plus a marker standard, labeled with a dye-terminator in a fourth colour, facilitates multiplexing and enhances the accuracy of data generation and overlap analyses. Improvements in data collection software, permitting the 64 multiplexed samples per gel, resulted in a two fold increase in data collection, increasing throughput to 192 clones per gel run. A requirement for the scale up in the construction of sequence-ready maps is the need to develop non-isotopic forms of contig construction. Fluorescent fingerprinting assists the progression towards large scale mapping by increasing throughput by multiplexing, enhancing the accuracy of the data generated and providing improved safety. Fluorescent fingerprinting has been successfully applied to the construction of a physical map of a region of mouse chromosome X (Mallon *et al*., 2000) as well as large scale sequence-ready bacterial clone maps of human chromosomes 1, 6, 20 (Deloukas *et al*., 2001) and X (Bentley *et al*., 2000) as discussed in the next chapter.