Chapter Two - Materials and Methods

2 Materials and Methods

Materials

2.1 Chemical reagents

All common chemicals were purchased from Sigma Chemical Co., BDH Chemical Ltd., and Difco Laboratories unless specified below or in the text.

Bio-Rad Laboratories	β-mercaptoethanol
Gibco BRL Life Technologies	ultraPURE™ Ammonium sulphate, enzyme grade, ultraPURE™ agarose
Amresco	EZ Squeeze Gene-PAGE PLUS acrylamide
Amersham Pharmacia Biotech	Dextran sulphate
VWR	Acetic acid
Merck	Acetone
Fluka	Formamide
PE Applied Biosystems	TET, HEX and NED fluorescent nucleotide dyes
Amersham Biosciences	Vistra Green stain
Boehringer	Blocking reagent

2.2 Enzymes and commercially prepared kits

Amersham Pharmacia Biotech	*Sau 3*A1
PE Applied Biosystems	Amplitaq™
	AmplitaqFS
Qiagen	Genomic DNA and DNA gel purification
Sigma	Ribonuclease A
	Deoxyribonuclease I
	DNA polymerase I (10 U/µl)
Ambion	DNA-free DNase treatment kit
Invitrogen Life Technologies	Superscript II cDNA synthesis kit
NEB Biolabs	*Hin*d III (20 U/µl)

2.3 Nucleotides

Amersham Pharmacia Biotech	Redivue™ [α-^{32}P]-dCTP (AA 005) aqueous solution (370 Mbq/ml, 10 mCi/ml)
Amersham Pharmacia Biotech	2'-deoxynucleoside 5'-triphosphates (dATP, dTTP, dGTP, dCTP)
Boehringer	biotin-16-dUTP digoxigenin-11-dUTP

2.4 Solutions

Solutions used in the present study are listed below, alphabetically within each section. Final concentrations of reagents are given for most solutions. Amounts and/or volumes used in preparing solutions are given in some cases. Unless otherwise specified, solutions were made up in nanopure water.

2.4.1 Buffers

10x PCR buffer	670 mM Tris-HCl (pH 8.8) 166 mM $(NH_4)_2SO_4$ 67 mM $MgCl_2$
1x $T_{0.1}E$	10 mM Tris-HCl (pH 8.0) 0.1 mM EDTA

2.4.2 Electrophoresis solutions

6x Glycerol dyes	30% v/v glycerol 0.1% w/v bromophenol blue 0.1% w/v xylene cyanol 5 mM EDTA (pH 7.5)
20x SSC	3 M NaCl 0.3 M Trisodium citrate
10x TBE	890 mM Tris base 890 mM Borate 20 mM EDTA (pH 8.0)

2.4.3 Media

All media were made up in nanopure water and either autoclaved or filter-sterilised prior to use. For agar used for bacterial growth 15 mg/ml bacto-agar were added to the appropriate media. Antibiotics were added to media as appropriate (see Table 2.1) to the following final concentrations: Kanamycin (purchased as a solution, stored at 4°C), 30 μg/ml; Chloramphenicol (stored at 4°C), 12.5 μg/ml, both supplied by Sigma).

Table 2-1 Clones and appropriate antibiotics

Clone type	Library	Antibiotic
Cosmid	LL0XNC01	Kanamycin
PAC	RPCI1,3,4,5, 6	Kanamycin
BAC	RPCI-11, 13, 23, 24. RZPD668 Sminthopsis *macroura* library.	Chloramphenicol

LB	10 mg/ml bacto-tryptone
	5 mg/ml yeast extract
	10 mg/ml NaCl
	(pH 7.4)
2 x TY	15 mg/ml bacto-tryptone
	10 mg/ml yeast extract
	5 mg/ml NaCl
	(pH 7.4)

2.4.4 DNA labelling and hybridisation solutions

Hybridisation buffer	6x SSC
	1% w/v N-lauroyl-sarcosine
	10x Denhardt's
	50 mM Tris-HCl (pH 7.4)
	10% w/v dextran sulphate

2.4.5 General DNA preparation solutions

GTE	50 mM glucose
	1 mM EDTA
	25 mM Tris-HCl (pH 8.0)
3 M K^+/5 M Ac^-	60 ml 5 M potassium acetate (pH 4.8)
	11.5 ml glacial acetic acid
	28.5 ml H_2O
Lysis buffer	50mM glucose
	10mM EDTA
	25mM Tris pH 8.0

2.4.6 FISH solutions

10x nick translation buffer	0.5 M Tris-HCl (pH 7.5)
	0.1 M $MgSO_4$,
	1 mM dithiothreitol,
	500 µg/ml bovine serum albumin
Lysis solution	5 parts 70mM NaOH
DNA fibre preparation	2 parts absolute ethanol
	(made up in distilled H_20)
Hybridisation buffer	50% deionised formamide
	2x SSC
	10% dextran sulphate
	0.1% Tween 20
	10 mM Tris (pH 7.4)
Cadenza wash solution	0.05% Tween 20
	4x SSC
Cadenza blocking buffer	1% w/v blocking reagent
	0.001% sodium azide (in Cadenza wash solution)
	50% formamide
	50% 2x SSC v/v
Cadenza Layer 1 solution	4 µg/ml avidin Texas Red DCS (Vector) (diluted in Cadenza blocking buffer)

Cadenza Layer 2 solution	4 µg/ml biotinylated anti-avidin D plus 1 µg/ml mouse anti-digoxigenin (Boehringer) (diluted in Cadenza blocking buffer)
Cadenza Layer 3 solution	4 µg/ml avidin Texas Red DCS plus 10µg/ml goat anti-mouse FITC conjugate (Sigma) (diluted in Cadenza blocking buffer)

2.5 Size markers

1 kb ladder (1 mg/ml) (Gibco BRL Life Technologies)
Contains 1 to 12 repeats of a 1,018 bp fragment and vector fragments from 75 to 1,636 bp to produce the following sized fragments in bp: 75, 142, 154, 200, 220, 298, 344, 394, 516/506, 1,018, 1,635, 2,036, 3,054, 4,072, 5,090, 6,108, 7,125, 8,144, 9,162, 10,180, 11,198, 12,216.

2.6 Hybridisation membranes and X-ray and photographic film

Amersham	Hybond-N™ Nylon (78 mm x 119 mm)
Polaroid	Polaroid 667 Professional film
Autoradiographs	Fuji RX medical X-ray film

2.7 Sources of genomic DNA

Human placental DNA for pre-reassociation (ready-sheared) was purchased from Sigma Chemical Co.. Human placental DNA for PCR was purchased from Sigma Chemical Co.. DNA from hybrid Clone 2D (Cl2D) that contains the entire X chromosome was kindly provided by Adam Whittacker. Mouse genomic DNA was purchased from Sigma Chemical Co.. Hamster genomic DNA was kindly provided by Frances Lovell and Christine Burrows, and mouse CotI DNA was kindly provided by Ruby Banerjee.

2.8 Sources of RNA

Total RNA was obtained from Ambion and Clontech. Tissue origins are given below in table 2-2.

Table 2-2 Sources of total RNA used for RT-PCR experiments

Supplier	Tissue panel number	Tissue	Supplier	Tissue panel number	Tissue
Clontech	1	Adrenal gland	Clontech	21	Fetal brain
Clontech	2	Bone marrow	Clontech	22	Fetal liver
Clontech	3	Brain (cerebellum)	Ambion	23	Adrenal gland
Clontech	4	Brain (whole)	Ambion	24	Bladder
Clontech	5	Fetal brain	Ambion	25	Brain
Clontech	6	Fetal liver	Ambion	26	Cervix
Clontech	7	Heart	Ambion	27	Colon
Clontech	8	Kidney	Ambion	28	Heart
Clontech	9	Liver	Ambion	29	Kidney
Clontech	10	Lung	Ambion	30	Liver
Clontech	11	Placenta	Ambion	31	Lung
Clontech	12	Prostate	Ambion	32	Ovary
Clontech	13	Salivary gland	Ambion	33	Pancreas
Clontech	14	Skeletal muscle	Ambion	34	Placenta
Clontech	15	Spleen	Ambion	35	Prostate
Clontech	16	Testis	Ambion	36	Skeletal muscle
Clontech	17	Thymus	Ambion	37	Small intestine
Clontech	18	Thyroid gland	Ambion	38	Spleen
Clontech	19	Trachea	Ambion	39	Stomach
Clontech	20	Uterus	Ambion	40	Testis

2.9 Sources of cells for *Sminthopsis macroura* and mouse FISH

Metaphase preparations of a *Sminthopsis macroura* cell line were a kind gift from Dr. Willem Rens (University of Cambridge). Mouse cells from a primary culture derived from the spleen were a kind gift from Dr. Ruby Banerjee (Wellcome Trust Sanger Institute).

2.10 Bacterial clone libraries

2.10.1 Cosmid libraries

Cosmids from the Lawrence Livermore flow-sorted X chromosome cosmid library (LL0XNC01) (prefixed 'cU') were kindly provided by Dave Vetrie and Elaine Kendall. Cosmids from a library constructed from a male with 5 X chromosomes (Holland *et al*., 1993)(prefixed 'cV') were also kindly provided by Dave Vetrie and Elaine Kendall.

2.10.2 PAC and BAC libraries

The RPCI-1, RPCI-3, RPCI-4, RPCI-5 (prefixed 'dJ'), and RPCI-6 (prefixed 'dA') PAC libraries, and the RPCI-11 (prefixed 'bA') and RPCI-13 (prefixed 'bB') BAC libraries were used as a source of human derived PAC clones and BAC clones respectively in this thesis. Mouse-derived BAC clones were obtained from the RPCI-23 (prefixed 'bM') female C57BL/6J and RPCI-24 (prefixed 'bN') male C57BL/6J libraries, and *Sminthopsis macroura*-derived BAC clones were obtained from the RZPD668 library (prefixed 'bF'). The human and mouse libraries were all kindly provided by Pieter de Jong and Joe Catanese (see http://bacpac.chori.org/), and the marsupial library by MA Chapman. Libraries were imported and maintained by the Sanger Institute Clone Resources Group.

2.10.3 cDNA libraries

A range of up to 19 different cDNA libraries were used in this study (see Table 2-3). cDNA libraries were imported and maintained by Jacqueline Bye. Each library contains 500,000 cDNA clones, divided into 25 pools of 25,000 clones. Five pools were combined to form a superpool containing 100,000 clones. Prior to their use in PCR, each superpool was diluted 1:100 and 1:1000 in $T_{0.1}E$.

Table 2-3 cDNA libraries used for SSPCR

cDNA library code	cDNA library description	Vector	Source/ Reference
1. U	(Monocyte NOT activated-from a patient with promonocytic leukaemia) (U937+)	pCDM8	Simmons (1993)
2. H*	Placental, full term normal pregnancy (H9)	pH3M	Simmons (1993)
3. P	Adult brain	pCDNA1	Pfizer
4. DAU	B lymphoma (Daudi)	pH3M	Simmons (1993)
5. FB	Fetal brain	pCDNA1	Invitrogen
6. FL	Fetal liver	pcDNA1	Invitrogen
7. HL	Peripheral blood (HL60)	pCDNA1	Invitrogen
8. SK	Neuroblastoma cells	pCDNA1	Invitrogen
9. T	Testis	pCDM8	Clontech
10. FLU	Fetal lung	pCDNA1	Invitrogen
11. AL	Adult lung	pCDNA1	Clontech
12. UACT*	(Monocyte PMA activated – from a patient with promonocytic leukaemia) (U937act)	pCDM8	Simmons (1993)
13. YT*	HTLV-1+ve adult leukaemia T cell	pH3M	Simmons (1993)
14. NK*	Natural killer cell	pH3M	Simmons (1993)
15. HPB*	T cell from a patient with acute lymphocytic leukaemia (HPBALL)	pH3M	Simmons (1993)
16. BM*	Bone Marrow	pH3M	Simmons (1993)
17. DX3*	Melanoma	pH3M	Simmons (1993)
18. AH	Adult Heart	pcDNA3-Uni	Invitrogen
19. SI **	Small Intestine	pcDNA3	Stammers

* Generously provided by Dr Simmons, Oxford **(**Simmons *et al*., 1993).

** Generously provided by Dr Stammers (Sanger Institute)

2.11 Primer sequences

All primers were synthesised in house by Dave Fraser or externally by Genset. Table 2-4 lists the vector-specific primers and sequences used in SSPCR. Appendices A to D lists the STSs used in this thesis, the sequence of each primer and the expected size in base pairs (bp) of each product, and the optimal annealing temperature (AT – given in °C). Where appropriate, the clones, or genes from which the STSs were derived are also listed.

Table 2-4 Vector-specific primer sequences for primers used in SSPCR

Primer Name	Primer Sequence	Vector
SP6PAC*	ATTTAGGTGACACTATAG	pcDNA3
pH3M1FP	CTTCTAGAGATCCCTCGA	pCDM8, pH3M
pH3M2FP	GCTCGGATCCACTAGTAA	pCDM8, pH3M
pH3M1RP	CTCTAGATGCATGCTCGA	pCDM8, pH3M
pH3M2RP	CGACCTGCAGGCGCAGAA	pCDM8, pH3M
pCDM8.RP	TAAGGTTCCTTCAGAAAG	pcDNA1, pCDM8
T7.2FP	AATACGACTCACTATAG	pCDM8, pcDNA1, pcDNA3

* designed by John Collins (Sanger Institute)

2.12 Key World Wide Web addresses

Baylor College of Medicine Search Launcher	http://searchlauncher.bcm.tmc.edu/
Baylor College of Medicine Sequencing Center	http://www.hgsc.bcm.tmc.edu/
British Columbia Genome Sequence Centre	http://www.bcgsc.bc.ca/
DOTTER	http://www.cgr.ki.se/cgr/groups/sonhammer/Dotter.html
ENSEMBL	http://www.ensembl.org

Genome Sequencing Center, St Louis	http://www.ibc.wustl.edu/cgm/jcgm.html
Genome Sequencing Center, Jena	http://genome.imb-jena.de/
Genome Sequencing Center, Naples	http://hpced.area.na.cnr.it/grsl/
INTERPRO	http://www.ebi.ac.uk/interpro/scan.html
MPIMG, Berlin (X sequencing)	http://www.mpimg-berlin-dahlem.mpg.de/~xteam/
National Centre for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
OMIM	http://www3.ncbi.nlm.nih.gov/Omim/
RepeatMasker	http://ftp.genome.washington.edu/RM/webrepeatmaskerhelp.html
The Institute for Genome Research	http://www.tigr.org/
The Wellcome Trust Sanger Institute	http://www.sanger.ac.uk/
X Chromosome Mapping Project at the Sanger Institute	http://www.sanger.ac.uk/HGP/ChrX/
European Bioinformatics Institute	http://www.ebi.ac.uk/
Human Genome Mapping Project-Resource Centre (HGMP-RC)	http://www.hgmp.mrc.ac.uk/
Mouse Genome Informatics	http://www.informatics.jax.org/
UCSC genome browser	http://genome.cse.ucsc.edu/

Methods

2.13 Isolation of bacterial clone DNA

2.13.1 Miniprep of BAC DNA

1. Ten ml of 2 x TY media, supplemented with 10 µl of (12.5 mg/ml) chloramphenicol, were inoculated with a scraping from a frozen bacterial glycerol stock. The culture was incubated overnight at 37°C with shaking.
2. A bacterial pellet was formed from the culture by centrifugation at 2000 g (Sorvall RT7, Du Pont Company Sorvall, Delaware US) at 4°C for 10 minutes. The supernatant was discarded into hycolin and the tube inverted on tissue to drain.
3. 200 µl lysis buffer were added to the pellet and pipetted gently to resuspend the pellet. The resulting suspension was transferred to a 1.5 ml microfuge tube and left to stand at room temperature for 10 minutes.
4. 400 µl of fresh 0.2 M NaOH/1% SDS were added to the cells and the lysate mixed gently by inversion. The sample was then incubated on ice for 5 minutes.
5. 300 µl of 3 M sodium acetate (pH 5.2) were added and the lysate mixed gently by inversion. The sample was then incubated on ice for 10 minutes.
6. The tube contents were pelleted by centrifugation in a microfuge at 13,000 rpm for 5 minutes in an Eppendorf microfuge. The clear supernatant was transferred to a fresh microfuge tube. This procedure was repeated twice (if the supernatant still was not clear, the sample was incubated on ice for another 10-30 minutes, and the centrifugation repeated).
7. 600 µl of isopropanol were added to the supernatant, the tube contents were mixed gently, and incubated at -70°C for 10 minutes or longer.
8. The sample was then subjected to centrifugation at 13,000 rpm in an Eppendorf microfuge for 5 minutes, the supernatant removed and the tube inverted on tissue to drain.
9. The pellet was resuspended in 200 µl 0.3 M sodium acetate (pH 7).
10. 200 µl of 50/50 v/v phenol/chloroform were added to the sample. The sample was then vortexed, and subjected to centrifugation for 3 minutes at 13,000 rpm in an Eppendorf microfuge. 150 µl of the aqueous phase (upper layer) were then transferred to a new tube.

11. 50 µl of 0.3 M sodium acetate (pH 7) were added to the remaining organic layer, the sample was vortexed and subjected to centrifugation in an Eppendorf microfuge for 2 minutes at 13,000 rpm. 50 µl of the aqueous phase were pooled with the aqueous layer from step 10.
12. 200 µl of isopropanol were added to the pooled aqueous phases, and mixed by inversion. The sample was then incubated at -70°C for 10 minutes.
13. The sample was subjected to centrifugation in an Eppendorf microfuge for 8 minutes at 13,000 rpm, the supernatant was removed, and the tube inverted on tissue to drain.
14. 500 µl of ice-cold, 70% ethanol were added, taking care not to disturb the pellet. The sample was subjected to centrifugation in an Eppendorf microfuge for 5 minutes at 13,000 rpm, the supernatant removed by aspiration, and the pellet dried at 37°C for approximately 30 minutes.
15. 50 µl of $T_{0.1}E$ containing 200 µg/ml RNAseA were added, and the tube "flicked" to resuspend the DNA pellet. The sample was then incubated at 55°C for 15 minutes.
16. Two µl of the sample were analysed by electophoresis on a 1% agarose gel. The remainder of the sample was stored at -20°C until required.

2.13.2 Microprep of BAC DNA for restriction digest fingerprinting

1. 500 µl of 2 x TY containing chloramphenicol (see Table 2-1) were added to a 96-well deep-well microtitre plate (COSTAR).
2. Each well was inoculated from a glycerol stock with either a 96-well inoculating tool, or a sterile cocktail stick. A plate sealer (Dynax) was placed on top of the plate to seal the wells, and the culture grown for 18 hours at 37°C with gentle shaking.
3. For each well, 250 µl of the overnight growth were transferred to a clean microtitre plate. The cells were collected by centrifugation (Sorvall RT7, Du Pont Company Sorvall, Delaware US) at 1550 g for 4 minutes.
4. For each well, the supernatant was removed and the pellet resuspended in 25 µl of GTE, by vortexing gently (a cocktail stick was used for resuspending pellets still attached to the plate).
5. 25 µl of GTE were added to each well and gently mixed. 25 µl of freshly prepared 0.2 M NaOH/1% SDS were added, mixed and left to stand for 5 minutes at room temperature.

6. 25 µl of 3 M K^+/5 M Ac^- (pH 5.0) were added, mixed and left at room temperature for 5 minutes. A plate sealer was placed on top of the plate and the plate was vortexed gently for 10 seconds.
7. A microtitre plate containing 100 µl of isopropanol was taped to the bottom of 2 µm filter-bottomed plate (Millipore cat. no. MAGVN2250). The total well volume of the sample was transferred to the filter-bottomed plate and the sample was filtered by centrifugation at 1550 g for 2 minutes at 20°C.
8. The filter-bottomed plate was removed and the microtitre plate was left at room temperature for 30 minutes, before being centrifuged at 1500 g for 20 minutes at 20°C.
9. The supernatant was removed and the DNA was dried by inverting the plate on clean tissue paper, ensuring no disruption of the pellet.
10. 100 µl of 70% ethanol were added to the dried DNA, mixed gently, and DNA precipitated by centrifugation at 1500 g for 10 minutes at 20°C. The wash was repeated for restriction digest fingerprinting. The supernatant was removed and the DNA dried as described in step 9.
11. 5 µl of freshly prepared $T_{0.1}E$ / 1 µg/ml RNase were added and mixed gently to resuspend the DNA. Samples were stored at -20°C.

2.14 Bacterial clone *Hin*d III/*Sau* 3A 1 fluorescent fingerprinting

1. For one 96-well microtitre plate of sample DNAs, three digest premixes were prepared, one for each fluorescent label, in three 1.5 ml microfuge tubes labelled TET, HEX and NED. Each premix contained 25.5 µl $T_{0.1}E$, 24.5 µl NEB2 buffer (as supplied by the manufacturer), 5.0 µl *Hin*d III (20 U/µL), 8.0 µl *Taq* FS (32 U/µl) and 3.0 µl *Sau* 3A I (30 U/µl), 4.0 µl of the appropriate ddA-dye. Each premix was mixed prior to being aliquotted.
2. 2 µl of the TET premix were added to wells A1-H4 of the microtitre plate containing sample DNAs using a Hamilton repeat dispenser. Similarly, 2 µl of the HEX premix were added to wells A5-H8, and 2 µl of the NED premix were added to wells A9-H12. The plate was covered with a plate sealer and the reaction mixed by gentle agitation on a vortex. In order to ensure the sample was in the bottom of the wells

the plate was centrifuged at 150 g for 10 seconds (Sorvall RT7, Du Pont Company Sorvall, Delaware US).

3. The reaction was incubated for 1 hour at 37°C.
4. To precipitate the DNA, 7 µl 0.3 M sodium acetate and 40 µl 96% ethanol were added to each well. For multiplexing the samples, rows 5 and 9 were added to row 1, rows 6 and 10 were added to row 2, rows 7 and 11 were added to row 3, and rows 8 and 12 were added to row 4 respectively, using a multichannel pipette.
5. The samples were incubated at room temperature for 30 minutes in the dark.
6. The samples were subjected to centrifugation at 1550 g for 20 minutes at 20°C to pellet the DNA.
7. The supernatants were discarded and the pellets dried by tapping the plate face down onto tissue paper.
8. The pellets were washed by adding 100 µl of 70% ethanol to each well, mixed gently by tapping the plate. The samples were subjected to centrifugation at 1550 g for 10 minutes at 20°C.
9. The supernatants were discarded and the pellet dried as described in step 7.
10. The DNAs were resuspended in 5 µl $T_{0.1}E$.
11. Prior to loading, 2 µl of the marker DNA (kind gift from Frances Lovell, Wellcome Trust Sanger Institute, see Section 2.5) were added to each sample using a Hamilton repeat dispenser. The samples were denatured for 10 minutes at 80°C. 1.00 µl of each sample were loaded on a 5% denaturing acrylamide gel and resolved on an ABI377 Automated DNA sequencer. Data were collected using the ABI Prism Collection Software v1.1.
12. After data collection, the gel image was transferred to a UNIX workstation for entry into IMAGE.

2.15 Agarose gel preparation and electrophoresis

1. Agarose gels were prepared in 1x TBE containing 250 ng/µl ethidium bromide and the appropriate percentage of agarose according to the size of fragments being separated: 2.5 % agarose gels were used for electrophoresis of fragments below 1 kb; 1.0 % agarose gels were used for analysis of larger fragments.

2. Electrophoresis was performed at 50 - 90 V for 15 - 45 minutes depending on the separation required.
3. Products were visualised by UV illumination.

2.16 Applications using the polymerase chain reaction

2.16.1 General primer design

Primers were designed using Primer3 (Rozen and Skaletsky, 2000), or manually using the following guidelines:

As far as possible, sequences chosen were 18 - 25 bp in length.

Sequences were chosen to avoid areas of simple sequence and obvious repetitive sequence i.e., runs of single nucleotide (e.g. TTTT) or double nucleotide (CGCGC) motifs.

Sequences were chosen to exclude palindromes which will form inhibitory secondary structure, especially at the 3' ends (e.g. GACGTC).

As far as possible, sequences were chosen with a GC content of at least 50%.

Sequences were chosen to avoid complementarity between pairs of primers, especially at the 3' end, which could result in primers annealing to each other and forming primer dimers.

If possible, sequences were chosen which would generate products of at least 100 bp in length.

2.16.2 Oligonucleotide preparation

All oligonucleotides used were synthesised in house by David Fraser or supplied as working dilutions from Genset. The concentration of the primer in ng/µl was determined by measuring the absorbance at 260 nm (Abs260) and multiplying this by 33 and any necessary dilutionfactor.

2.16.3 Amplification of genomic DNA by PCR

1. 1-10 ng/µl of genomic DNA were amplified in a reaction volume of 15 to 50 µl as required. Reactions contained approximately 1.3 µM of each oligonucleotide primer, 67 mM Tris-HCl (pH 8.8), 16.6 mM $(NH_4)_2S0_4$, 6.7 mM $MgCl_2$, 0.5 mM of each deoxyribonucleoside triphosphate (dATP, dCTP, dGTP, dTTP), 0.6 U of AmplitaqTM (Cetus Inc.). 10 mM β-mercaptoethanol and 170 µg/ml of BSA (Sigma Chemical Co., A-4628) were added to the reactions from freshly made stock solutions as the reactions were set up.

2. Unless specified otherwise, cycling conditions were as follows: all reactions were preceeded by an initial denaturing step of 5 minutes at 94°C, followed by 35 cycles of: 93°C for 30 seconds, [primer-specific annealing temperature] for 30 seconds, and 72°C for 30 seconds; followed by a final extension step of 5 minutes at 72°C. Primer-specific annealing temperatures are given for each primer pair in the text or in Appendices A-D.

3. PCR products were separated on 2.5% agarose minigels as described in Section 2.15 and visualised by ethidium bromide staining.

2.16.4 Colony PCR of STSs from bacterial clones

1. Colony PCR on bacterial clones was performed by touching a sterile toothpick onto the surface of a colony and stirring this into 200 µl of $T_{0.1}E$, and using 5 µl of the resulting suspension in a 15 µl final volume PCR (as described in Section 2.15.3).

2. PCR products were separated on 2.5% agarose minigels as described in Section 2.15 and visualised by ethidium bromide staining.

2.17 Radiolabelling of DNA probes by direct incorporation

PCR products were radiolabelled essentially as described in (Bentley *et al*., 1992).

1. 10-15 µl of PCR product were separated on a 2.5% agarose minigel and visualised by ethidium bromide staining.
2. The gel was rinsed in deionised water to remove excess buffer. The desired band was excised from the gel and placed in 100 µl of $T_{0.1}E$ at 4°C overnight.
3. 5 µl of the $T_{0.1}E$ were used as template in the 15 µl PCR-labelling reaction containing 1.3 µM of each primer, 1.5 µl of 10x PCR buffer, 0.5 µl of [α-32P]-dCTP (3,000 Ci/mmol), 0.12 U of *Taq* polymerase (Cetus) and 0.5 mM each of dATP, dTTP and dGTP. Reactions were performed in a 0.5 ml microfuge tube and overlaid with mineral oil (Sigma) in a DNA thermal cycler (Perkin Elmer, USA).

4. PCR cycling conditions were as follows: 94°C for 5 minutes; followed by 20 cycles of: 93°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds; followed by 72°C for 5 minutes.
5. For marsupial BAC screening, unincorporated nucleotides were then removed by elution on a Sephadex G50 column (Pharmacia Biotech). Probe was diluted to 400 µl in $T_{0.1}E$ and applied to the column. Five further additions of 400 µl $T_{0.1}E$ were made to the column, and the corresponding fractions were collected. Each fraction was counted using a scintillation counter (Easicount4000, Scotlab, UK). The labelled probe was eluted in fraction 2 which was used for subsequent hybridisations. The percentage incorporation of [α-32P]-dCTP into the labelled probe was calculated (count of fraction2/total counts for all fractions).
6. All probes were boiled for 5 minutes and snap-chilled on ice prior to use.

2.18 Hybridisation of radiolabelled DNA probes

2.18.1 Hybridisation of DNA probes derived from STSs to same-species BAC filters

1. Filters were pre-hybridised tightly rolled in 15 ml Sterilin tubes or flat in sandwich boxes for 3 hours in sufficient hybridisation buffer to cover the filters at 65°C with gentle shaking.
2. Radiolabelled probe was added and hybridised to the filters for greater than 16 hrs at 65°C in a shaking incubator.
3. Filters were washed twice at room temperature in 2x SSC for 5 minutes, twice at 65°C in 0.5 x SSC, 1% sarkosyl for 30 minutes. Filters were rinsed at room temperature in 0.2x SSC prior to draining the excess liquid, wrapping in Saran wrap (Dow Chemical Co.) and exposing to autoradiograph film.

2.18.2 Hybridisation of DNA probes derived from human STSs to Sminthopsis macroura BAC filters

1. Filters were prehybridised tightly rolled in 15 ml Sterilin tubes or flat in sandwich boxes for 3 hours in sufficient hybridisation buffer to cover the filters at 58°C with gentle shaking.
2. Radiolabelled probe was added and hybridised to the filters for greater than 16 hours at 58°C with gentle shaking.

3. Filters were washed twice at room temperature in 2x SSC for 5 minutes, once at 58°C in 1.5 x SSC, 1% sarkosyl for 30 minutes, and once at 58°C in 1 x SSC, 1% sarkosyl for 30 minutes (unless as directed in the text). Filters were rinsed at room temperature in 0.2x SSC prior to draining the excess liquid, wrapping in Saran wrap (Dow Chemical Co.) and exposing to autoradiograph film.

2.18.3 Stripping radiolabelled probes from hybridisation filters

Filters were washed in 0.4 M NaOH for 30 minutes at 42°C followed by 30 minutes in 0.2 M Tris-HCl (pH 7.4), 0.1x SSC, and 0.1% w/v SDS at 42°C with gentle shaking. Successful removal of radiolabelled probe was assessed by autoradiography.

2.19 Clone library screening

2.19.1 cDNA library screening by PCR

1. Nineteen different cDNA libraries were subdivided into 25 subpools of 20,000 clones, which were then combined to produce 5 superpools of 100,000 clones by J. Bye and S. Rhodes. Details of the cDNA libraries are given in Table 2.3.
2. Aliquots of the superpools of each library were arranged in a microtitre plate to facilitate subsequent manipulations and gel-loading post PCR with a multi-channel pipetting device.
3. In the primary screen, 5 µl of each superpool were used as template in a 15 µl final volume PCR using buffer and PCR conditions as described in Section 2.16.3.
4. PCR products were loaded on 20 cm x 20 cm 2.5% agarose horizontal slab gels using an 8-way multi-channel pipetting device, separated by electrophoresis and visualised by ethidium bromide staining.
5. In the secondary screen, 5 µl of each of the 5 subpools of 20,000 clones corresponding to the superpool that were positive in the first round, were screened by PCR with the same primer pair as used in step 2. PCR products were separated by electrophoresis through 2.5% agarose gels and visualised by ethidium bromide staining.

2.19.2 Single-sided specificity PCR (SSPCR) of cDNA

The principle of SSPCR (Huang *et al*. 1993) is illustrated in Figure 2-1.

1. SSPCR was performed on the subpools of the cDNA libraries, each containing 20,000 clones. Prior to their use in PCR, the subpools were diluted 1:10 in $T_{0.1}E$ and boiled. Dilutions were stored at -20°C until required. On removing from -20°C, tubes were centrifuged briefly in a microfuge to settle the contents and then mixed carefully when thawed.
2. In the first round, PCR was performed using 1 µl of the diluted subpools as template in a 15 µl final volume using buffer conditions as described in Section 2.16.3. The primer combinations used are given in Table 2-5.
3. PCR was performed in microtitre plates in a DNA thermocycler (Omnigene) using hot-start. Cycling conditions were as follows: an initial denaturation step of 5 minutes at 95°C, followed by 25 cycles of: 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3 minutes; followed by a final step of 10 minutes at 72°C.
4. For the second round of PCR, products from the first round were diluted 1 in 50 and 1 in 500 in $T_{0.1}E$. 5 µl of each dilution were used as template in 15 µl final volume PCR using buffer conditions as described in Section 2.16.3. Cycling conditions were as described in step 3. The primer combinations used are given in Table 2-5.

Table 2-5 Primer combinations used in SSPCR*

First round SSPCR	Second round SSPCR
Specific primer A and FP vector primer	Specific primer B and FP vector primer
Specific primer A and RP vector primer	Specific primer B and RP vector primer
Specific primer C and FP vector primer	Specific primer D and FP vector primer
Specific primer C and RP vector primer	Specific primer D and RP vector primer

*Primer sequences are given in Table 2-4.

5. 5 µl of the second-round PCR products were separated by electrophoresis through either 1% or 2.5% agarose gels depending on product size and visualised by ethidium bromide staining. Products were gel purified using the Qiaquick gel extraction kit (Qiagen™) prior to sequencing directly.

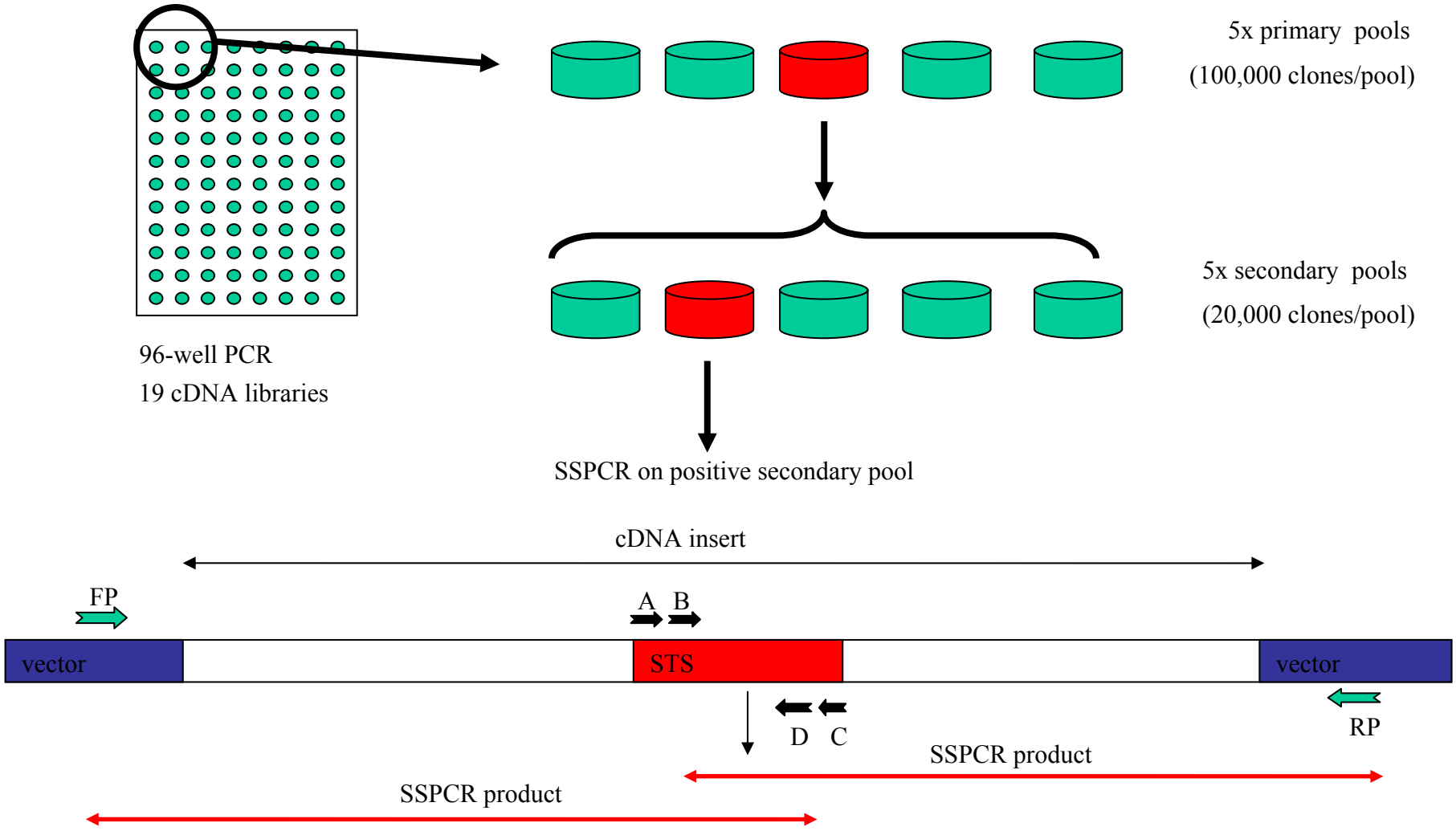

Figure 2-1 Diagram illustrating the SSPCR procedure described in Section 2.19.2.

2.20 RT-PCR expression profiling using total RNA samples

2.20.1 Generation of cDNA from total RNA

2.20.1.1 RNA treatment

1. RNA obtained from Clontech was subjected to DNase treatment as described in Section 2.20.1.2 (Ambion RNA was supplied pre-treated with DNase). RNA was stored at -70°C until required.
2. cDNA synthesis reaction omitting RT was performed as described in Section 2.20.1.2 for RNA samples 1-20 and tested by PCR using various non intron-spanning STSs.

2.20.1.2 DNase treatment of total RNA samples

1. Total RNA samples were treated with DNase to remove contaminating genomic DNA as required, using a DNA-Free kit (Ambion), as per the manufacturers' instructions.

2.20.1.3 Generation of single-stranded cDNA from total RNA

1. Single-stranded cDNA was generated from 0.5 µg–2 µg of total RNA using a Superscript II cDNA synthesis kit (Invitrogen Life Technologies) as per the manufacturers' instructions.
2. The resultant cDNA was diluted in $T_{0.1}E$ to give 200 µl/ug, and stored at -20°C. 5µl of the cDNA were then used per reaction in subsequent PCR.

2.20.2 Screening of cDNA samples by PCR

1. PCR was performed in a total volume of 15 µl using 5 µl of single-stranded cDNA as a template (Section 2.20.1.2) in a 96-well format essentially as described in Section 2.16.3. Primer annealing temperatures used are given in Appendix C. Unless otherwise noted in the text, 35 cycles of PCR were performed.

2.20.2.1 Visualisation of PCR products by agarose gel electrophoresis and Vistra Green post-staining

1. PCR products were separated by agarose gel electrophoresis as described in Section 2.15, except ethidium bromide was omitted and electrophoresis tanks were rinsed with distilled H_20 prior to use.
2. Following electrophoresis, gels were stained in Vistra Green (50 µl Vistra Green, 5 ml 1M Tris-HCl, 0.5 mM 0.1 M EDTA made up to 500 ml in doubly-distilled H_20) for ~30-45 minutes with gentle rocking, rinsed with distilled H_20 and scanned on a FluorImager SI.

2.21 FISH of BAC-derived probes to Sminthopsis macroura metaphase preparations

2.21.1 Metaphase slide preparation

1. A coplin jar containing freshly prepared 3:1 methanol/acetic acid fixative was prepared. Microscope slides were soaked in 96% ethanol and dried using lint-free tissue then lain horizontally on a humid-tray (containing tissue soaked in DDW).
2. The *Sminthopsis macroura* cell line suspension was mixed and the appropriate numbers of drops were added to the slide using a pasteur pipette (most slides were prepared using two single drops placed side-by-side).
3. The slides were examined under a phase-contrast microscope and the limits of the cell-spot noted on the slide using a pencil. The slides were placed in a coplin jar containing 3:1 methanol/acetic acid fixative for 30-60 minutes, then air-dried.
4. Slides were then passed through a 70%, 70%, 90%, 90%, 100% ethanol series (one minute in each) by dipping in a series of coplin jars and then air-dried.
5. After fixing in acetone for 10 minutes, slides were kept at room temperature in a sealed box containing dessicant for at least 24 hours before use.

2.21.2 Nick-translation labelling of BAC clones

1. 25 µl reactions were prepared for each clone. Approximately 1 µg of BAC clone DNA, was used per reaction.

2. The following reagents were added to a 1.5ml microfuge tube on ice: approximately 1 µg BAC clone DNA, 2.5 µl 10x nick translation buffer, 1.9 µl 0.5 mM dNTPs, 0.7 µl 1 mM hapten-conjugated dUTP, 1 µl DNAseI working solution (titrated to give fragment smears of 200-700 bp under appropriate incubation times), 0.5 µl DNA polymerase (10 U/µl) and sterile distilled water to give a final volume of 25µl.
3. The tube contents were gently mixed by flicking, briefly centrifuged in a microfuge and then incubated at 14°C for the appropriate length of time determined by DNAseI titration.
4. 2.5 µl of 0.5 M EDTA were added to the reactions to inactivate the enzymes. 2.5 µl of 3M sodium acetate (pH 7) and 1 ml ice-cold absolute ethanol were added and mixed. The reactions were incubated at -70°C for 30 minutes to precipitate the DNA.
5. The reactions were subjected to centrifugation at 13,000 rpm for 10 minutes in a microfuge, and the supernatant removed. 500 µl ice-cold 70% ethanol were added to the pellet, and the tubes were then centrifuged at 13,000 rpm in a microfuge for 2 minutes.
6. The supernatant was removed carefully, and the pellets air-dried. 10 µl of $T_{0.1}E$ were added to each pellet, and incubated on ice for 10 minutes. The tube was then flicked to resuspend the pellet, and 2 µl were removed for analysis by electrophoresis on a 1 % agarose gel. Samples were stored at -20°C until required.

2.21.3 Hybridisation of FISH probes to Sminthopsis macroura metaphase slides

1. For each hybridisation, the following were added to a 0.5 ml microfuge tube: 0.5 µl each nick-translated probe (prepared as described in Section 2.21.2), competing DNA as required, FISH hybridisation buffer to give a final volume of 13 µl. The contents of the tubes were then mixed by vortexing and briefly subjected to pulsed centrifugation in a microfuge.
2. Probes were denatured by incubating the reactions at 65°C for 10 minutes. Pre-annealing was then performed by incubating the reactions at 37°C for 40 minutes.

3. Metaphase slides, prepared as described in Section 2.21.1, were denatured by incubation in 70% formamide at 65°C for the appropriate length of time (1 minute and 5 seconds unless directed otherwise in the text).
4. Slides were immediately transferred to ice-cold 70% ethanol, then passed through a 70%, 70%, 90%, 90%, 100% ethanol series (one minute in each) by dipping in a series of coplin jars and air-dried.
5. Each probe mix was pipetted onto a 22 x 22 mm coverslip (one per cell spot contained on the slide as described in Section 2.21.1), and placed onto a slide. The coverslip was then sealed in place with rubber cement and the slide incubated at 37°C overnight in a humidified incubator.

2.21.4 Washing of Sminthopsis macroura metaphase slides and detection of signal

1. After removal of rubber cement, the slides were soaked in 2x SSC for 15 minute to remove the coverslips.
2. The slides were then washed as follows: 2x SSC at 42°C for 5minutes, 50% formamide at 42°C for 5minutes, a second wash of 50% formamide at 42°C for 5minutes, a second wash of 2x SSC at 42°C for 5minutes.
3. Slides were then loaded into a Cadenza rack for staining (See Section 2.21.5).
4. Following staining, slides were washed briefly with 2x SSC, then stained in 0.08 µg/ml DAPI solution for 3 minutes. Slides were rinsed in 2x SSC, and passed through a 70%, 70%, 90%, 90%, 100% ethanol series (one minute in each) by dipping in a series of coplin jars and air-dried.
5. For each slide, 3 drops of antifade (Citifluor AF1) were applied to a 22 x 50 mm coverslip, which was then overlaid onto the slide and sealed in place using nail varnish. Slides were left at 4°C for at least 30 minutes before image capture. Slides were stored at 4°C until required.
6. Images were obtained using a CoolSnap HQ camera (Photometrics) on an Axioplan 2 microscope (Zeiss), and analysed using SmartCapture 2 (Digital Scientific).

2.21.5 Two-colour detection using a Cadenza instrument

1. Rubber cement was removed from slides and coverslips soaked off at room temperature in 2x SSC for 15 minutes.
2. Slides were incubated in 2x SSC at 42^0C for 5 minutes in a Coplin jar.
3. Slides were incubated in 50% formamide (in a Coplin jar) at 42^0C for 5 minutes.
4. Slides were incubated in 50% formamide (in a second Coplin jar) at 42^0C for 5 minutes.
5. Slides were washed in 2x SSC at 42^0C for 5 minutes.
6. The Cadenza unit reservoir was topped up with Cadenza wash solution, the slides were mounted to Cadenza coverplates, and 2xSSC was pipetted into each the top of each Cadenza coverplate/slide reservoir to check mounting.
7. The reagent carousel was loaded: Position 1 – layer 1 solution, position 2 – layer 2 solution, position 3 – layer 3 solution, position 4 – blocking buffer (100 µl solution per slide plus 500 µl excess was prepared for each solution).
8. The detection run was then performed using the Cadenza 3-layer detection protocol.

2.22 Fibre-FISH of BAC-derived probes to Mus musculus DNA fibres

2.22.1 <u>Mus musculus</u> DNA fibre slide preparation

1. 2-3 mls cell suspension were taken by disposable pasteur pipette and subjected to centrifugation at 1200 rpm for 5 minutes.
2. Supernatant was removed and the pellet resuspended in PBS, then subjected to centrifugation at 1200 rpm for 5 minutes. This step was repeated twice.
3. Supernatant was removed and the pellet resuspended in 1 ml PBS.
4. An aliquot of cells were counted using a haemocytometer, and the cell suspension was diluted or concentrated accordingly to give a concentration of approximately 2-3x10^6 cells/ml.
5. 10 µl cell suspension were spread over a 1 cm area on the upper part of a clean microscope slide.
6. The slides were left to air dry at room temperature for 30 minutes.

7. The slides were fitted to Cadenza coverplates (cell spot uppermost) and clamped in a vertical position in a rack. 150 µl lysis solution were then added to the top of the slide.
8. Once the level of lysis solution dropped below the upper frosted region of the slide, 150 µl 96 % ethanol were added to the top of the slide, and allowed to drain.
9. Once the meniscus stopped falling, the slide and Cadenza unit were removed from the rack. The top of the slide was pulled gently away from the Cadenza coverplate, allowing the meniscus to move down the slide.
10. Slides were air-dried, and then fixed in acetone in a Coplin jar for 10 minutes.
11. Limits of the fibres were marked on the slide by pencil, and slides were left at room temperature for at least 24 hours before hybridising.

2.22.2 Hybridisation of FISH probes to Mus musculus DNA fibre slides

1. Probe master mix was prepared for each probe by mixing 4 µl labelled *Mus musculus* BAC clone (see section 2.21.2), 6 µl mouse CotI DNA and 40 µl FISH hybridisation buffer.
2. 5 µl of each probe master mix were then combined and the volume adjusted to 15 µl with FISH hybridisation buffer. Hybridisation was then performed as described in Section 2.21.3, step 2. Washing and two-colour detection were performed as described in Sections 2.21.4 and 2.21.5.

2.23 Mapping and sequence analysis software and databases

2.23.1 IMAGE

All processing of fingerprinting gels was carried out using IMAGE. IMAGE processed gels from fluorescent fingerprinting and extracted a normalised band pattern for each lane on a gel. Several procedures were run on each gel in turn:

Lane tracking – a grid was superimposed on the gel image and the grid manually edited to ensure it exactly matched the lanes on the gel.

Band calling – an analysis module traced the band pattern along the lanes and tried to identify the bands. Manual editing ensured the correct bands are chosen.

Marker locking – in order to compare band patterns from one gel to another all band positions were normalised to one master gel. A set of DNA fragments of known length or migration distance was loaded as a marker lane. Manual editing ensured the standard pattern matched to the pattern from the master gel.

Normalisation – once the marker lane patterns were locked onto the standard lane, the band positions of the sample lanes were normalised so that each lane appeared to have been run on the master gel with all distortions cancelled out. IMAGE finally generated a 'Bands' file for each gel containing normalised migration distances for all selected bands in each clone lane.

2.23.2 FPC

All contig construction and visualisation described in this thesis was performed in FPC (Soderlund *et al*., 1997). FPC took as the input a set of clones and their restriction fragments (called Bands) from IMAGE. Each fingerprint pattern for each clone is compared to the fingerprint patterns of all other clones in the database. The relationship between two clones was reported as a probability of coincidence, i.e. the probability that two clones overlap by chance. Two variables can be set to filter the reported overlaps:

Cut off – a match between two clones will only be reported if the probability of coincidence is less than or equal to the cut off. When analysing matches between *Sminthopsis macroura* BAC clones, the tolerance was set to 1e –04, and when analysing larger insert *Mus musculus* BAC clones, the tolerance was decreased to 1e –14.

Tolerance – two bands are considered as their migration distances differ by less than tolerance. For the analysis carried out in this thesis the tolerance was set to 7.

Overlapping clones were identified automatically and contigs were constructed manually using the available editing tools provided by FPC. A minimum set of clones for sequencing was chosen where appropriate based on a combination of shared bands and shared marker data.

2.23.3 Xace and other custom mapping databases

Human X chromosome data and annotation generated in this thesis were stored in Xace, a chromosome-specific implementation of ACeDB. Other ACeDB implementations

were used to store mouse and marsupial mapping and annotation data, as described in Chapter 4 and 6. ACeDB was originally developed for the *C. elega*ns genome project (Durbin and Thierry-Mieg, 1991). Documentation code and data available from anonymous FTP servers at lirmm.lirmm.fr,cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov).

ACeDB works using a system of windows and presents data in different types of windows according to the type of data. All windows are linked in a hypertext fashion, so that clicking on an object will display further information about that object. For example, clicking on a region of a chromosome map will highlight landmarks mapping to that part of the chromosome; clicking on a landmark will display information about that landmark including landmark-clone associations, etc.

All PAC, BAC and cosmid library filters and polygrids are represented graphically in Xace where each square on the grid represents an individual clone. Hybridisation data were entered directly via the grid. Data were then saved in the database establishing landmark-to-clone associations that can be displayed as text windows relating to either the landmark or the clone. Data can also be entered via text windows or via an internal web page. PCR library pool screening and colony PCR results were entered via the text windows.

In addition to the data generated by the X chromosome mapping group, Xace also contains displays of published X chromosome maps. Genomic sequence data is also displayed in ACeDB along with the collated results from the computational sequence analysis performed by the Sanger Institute Human Sequence Analysis Group.

Xace can be accessed by following the instructions at:
http://www.sanger.ac.uk/HGP/ChrX.

2.23.4 BLIXEM

Individual matches identified as a result of similarity searches using the BLAST algorithm, or matches between sequences of cDNA clones or PCR products amplified from genomic DNA generated as part of the project, were viewed in more detail using

BLIXEM. BLIXEM, (Blast matches In an X-windows Embedded Multiple alignment) is an interactive browser of pairwise Blast matches displayed as a multiple alignment. Either protein or DNA matches can be viewed in this way at either the amino acid or nucleotide level respectively. BLIXEM contains two main displays: the bottom display panel shows the actual alignment of the matches to the genomic DNA sequence, and the top display shows the relative position of the sequence being viewed within the context of the larger region of genomic DNA. A program "EFETCH" retrieves the record from an external database (e.g. EMBL, SWISSPROT).

2.23.5 RepeatMasker

Human repeat sequences were masked using RepeatMasker, a program that screens DNA sequence for interspersed repeats and low complexity DNA sequence (Smit & Green RepeatMasker at http://ftp.genome.washington.edu/RM/RepeatMasker.html). The output of the program is a detailed annotation of the repeats that are present in the query sequence and a version of the sequence with repeats masked by "N" characters. Sequence comparisons are performed by the program cross_match, an implementation of the Smith-Waterman-Gotoh algorithm developed by P. Green. The interspersed repeat databases screened by RepeatMasker are based on the repeat databases (Repbase Update (Jurka, 2000)) copyrighted by the Genetic Information Research Institute.

2.24 Alignment of nucleic acid and protein sequences and phylogenetic analysis

2.24.1 Alignment of nucleic acid and protein sequences

Nucleic acid and protein sequences were aligned using the program ClustalW (Thompson *et al*., 1994) via a web-based server at the EBI (http://www.ebi.ac.uk/), or ClustalX installed locally on a PC unless otherwise noted in the text (Thompson *et al*., 1997). User-defined parameters were left at their default settings unless directed otherwise in the text. Alignments were then manually edited and presented using the program GeneDoc (Nicholas, 1997).

2.24.2 Calculation of sequence identities and similarities

Nucleic acid and protein sequences were aligned as described in section 2.24.1. The "statistics report" function of GeneDoc was then used to calculate and display sequence identities and similarities.

2.24.3 Phylogenetic analysis of nucleic acid and protein sequences

Nucleic acid and protein sequence alignments were subjected to various phylogenetic analyses to estimate their order of relationship. In each case, alignments produced as described in Section 2.24.1 were manually edited as necessary to minimise the number of gaps, and the most reliably aligned region of the alignment was then used for the respective phylogenetic analyses. Any columns within the alignment containing gaps were removed prior to phylogenetic analysis.

2.24.3.1 Neighbour-joining distance analysis

Phylogenetic analyses were performed using the program MEGA2 (Kumar *et al*., 2001) installed locally on a PC. This package combines various phylogenetic analysis methodologies in a straightforward interface. For distance-based phylogenetic analysis, a Neighbour-Joining method was employed (Saitou and Nei, 1987). The distance measure used for peptide sequences was the poisson-corrected measure, to attempt to account for multiple substitutions at sites when distantly related sequences were compared. For nucleotide comparisons, the Kimura 2-paramater measure was used, considering both transitions and transversions. In all analyses, 1000 bootstrap replicates were selected to assess robustness of the tree produced.

2.24.3.2 Maximum likelihood analysis

Maximum-likelihood analysis was performed using the package PIE via a web-server at the MRC Rosalind Franklin Centre for Genomics Research. For protein sequences, the program PROTML was used, using the JTT substitution model. For nucleotide sequences, the program DNAML was used. For each type of analysis, a search for the best tree was performed, 100 bootstrap replicates were chosen, and sites were not weighted.

2.25 Comparative sequence analysis

2.25.1 PIPmaker

PIP plots were generated using PIPmaker as per the authors instructions (Schwartz *et al*., 2000). Text files were generated containing relevant sequences in fasta format, and an annotation file was generated as per the authors instructions. The annotation file was also used to generate an underlay file as per the authors instructions. The base sequence (human unless otherwise specified) was masked for repeats using RepeatMasker. Most program parameters were as default, except sequences were searched on both strands, and chaining was employed. Chaining reports only those matches occurring in the same order in the different species, and avoids build-up of matches due to repetitive sequence occurring throughout the sequences. Chaining assumes that the order of matches should be conserved.

The "high sensitivity" setting was also employed.

Key to PIP annotations (reproduced from authors web-site):

White pointed box – L1 repeat, Light grey triangle – SINEs, Black triangle – MIR, Black pointed box – LINE2, Dark grey triangle and pointed boxes – other interspersed repeats (e.g. LTRs, DNA transposons), Short dark grey box – CpG island with CpG/GpC > 0.75, Short white box - CpG island with CpG/GpC between 0.6-0.75.

2.25.2 VISTA

VISTA plots were generated as per the authors instructions (Mayor *et al*., 2000). Input files were the same as those generated for PIPmaker as described above. The Fugu and Human Xp sequences were reverse-complemented. Unless otherwise specified in the text, a window level of 100 bp was used with a conservation level of 75%, minimum conservation level of 50% and maximum difference was left as default. Percentage identity was plotted on the y-axis.