

**Genome Evolution:
a study of MHC paralogous genes in the
human genome**

Vikki Rand

**This dissertation is submitted for the degree of
Doctor of Philosophy**

September 2003

**Gonville and Caius College,
University of Cambridge**

**Wellcome Trust Sanger Institute
Wellcome Trust Genome Campus,
Hinxton, Cambridge**

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. This dissertation does not exceed the word limit set by the Biology Degree Committee.

Abstract

One of the interesting findings of the Human Genome Project was that approximately 10% of the genome has arisen by duplication. This is exemplified by the clusters of genes, on chromosomes 1q21-q25, 9q32-q34.3 and 19p13, paralogous to genes located within the Major Histocompatibility Complex (MHC) region, on 6p22.2-p21.3. By definition, paralogues are genes within the same species that have originated through duplication of an ancestral gene. The survey of the human genome identified 82 MHC paralogues based on sequence similarity and conserved gene structure. Analysis of the distribution of the paralogues identified clusters on chromosomes 1q21-q25, 9q32-q34.3 and 19p13 (38/82), and revealed paralogues located elsewhere in the genome (44/82). In total, 44% of the paralogues identified are novel discoveries, of which 89% are located outside the previously known clusters.

Evidence from my phylogenetic analyses indicates that the MHC paralogues located within the regions on 1, 9 and 19 arose by two ancient duplication events, either by duplication of the whole genome or of chromosomal segments, prior to vertebrate emergence. Expansion of paralogous gene families has occurred by additional duplications involving individual loci or chromosomal regions resulting in paralogues outside the clusters. In-depth analysis of the chromosomal region 9q32-q34.3 revealed that the order of paralogues is not conserved and that they are interspersed by other genes, indicating the region has been subjected to genomic rearrangements.

Comparison of the expression profiles of a selected set of MHC paralogues revealed that some have functionally diverged since duplication; with members of the same paralogous gene family being ubiquitously expressed, and others, having an

expression profile restricted to only a few tissues. Evidence of co-expression of paralogues in some tissues suggests a similar function and involvement in the same pathways. This thesis highlights the importance of understanding paralogy, particularly for future investigations of phenotypes associated with paralogous genes.

Acknowledgements

I would like to thank my supervisor Stephan Beck for his tireless enthusiasm, encouragement and support throughout this project. Particular thanks go to my second supervisor at Cambridge University, John Trowsdale. I am also grateful to the past and present members of the Immunogenomics group (a.k.a Team 50) who have helped me throughout my PhD: Karen Novik, Ruth Younger, Roger Horton, Melanie Stammers, Karen Crum, Karen Halls, Jennifer Sambrook, Penny Coggill, Marcos Miretti and Vardhman Rakyan. Thankyou for all your advice, encouragement and help. I am also very grateful for your tolerance – particularly of my moaning over the last few months – and for providing regular distractions in the form of tea breaks, coffee breaks, lunch, pub..... this also goes for the members of team 30, Gavin Wright, Mark Bushall and Nick Bockett, who foolishly share our lab and office!

I am indebted to everyone who has contributed to the mapping and sequencing of the human genome - without this data the work in this thesis would not have been possible. I would also like to thank everyone who has helped me with the different aspects of the project. Special thanks goes to Sean Humphray and the members of the Chromosome 9 Sequencing and Mapping group who helped me with the work presented in chapter 3. Particular thanks go to Andrew Knight for his help with the sub-cloning, Keith Porter for the fingerprinting and the Sanger Institute Cytogenetics laboratory, namely Sheila Clegg and Pawan Dhmi, who helped me with the fluorescent *in-situ* hybridisation (FISH) experiments. I am also very grateful to the members of team 41 and, the now extinct, team 49 who helped me with the sequencing and to Gavin Laird and Adrienne Hunt for their ‘finishing’ tutorials.

Thanks also go to Rosemary Ward for all her assistance with the tissue culture, Alison Coffey for her help with the blots and to Dave Vetrie, Cordelia Langford and the Sanger Institute Microarray Facility for all their help and guidance with the microarray experiments. Thanks also go to Ewan Birney and Michele Clamp for their help with ENSEMBL and to Kevin Howe for maintaining the FINEX database and writing some useful perl scripts. I realise I have been privileged to have worked in an environment where help has been offered so freely.

On a personal note, thanks must go to all my family and friends who have put up with me over the last few months – and years! Particular thanks go to my mam, dad and brother Simon for their pillar-like support and their unwavering belief and confidence in me – without this I don't think I would have made it this far! This also goes for all my friends who have offered encouragement and wine along the way!!! Cheers!

Table of contents

Abstract	ii
Acknowledgements	iv
Table of contents	vi
List of tables	xv
List of figures	xvii
Abbreviations	xxi
Chapter 1 Introduction	1
1.1 Genome evolution	1
1.2 Homologues, paralogues and orthologues	2
1.3 Paralogous genes and the evolution of the human genome	2
1.4 Genome sequencing projects	6
1.5 The Human Genome Project	6
1.6 Analysis of the human genome sequence	8
1.6.1 Gene numbers	8
1.6.2 1-to-4 gene rule	9
1.6.3 Paralogy and the human genome	10
1.6.4 Evolutionary analysis of paralogous gene families	11
1.7 Polyploidy	12
1.8 Mechanisms of gen(om)e duplication	13
1.9 What happens after gen(om)e duplication?	17
1.10 The extended Major Histocompatibility Complex	18
1.10.1 The extended class I region	19

1.10.2	The class I region	20
1.10.3	The class III region	21
1.10.4	The class II region	23
1.10.5	The extended class II region	24
1.11	Origin of the extended MHC	25
1.12	MHC Paralogy	26
1.12.1	Origin of the extended MHC paralogous regions	28
1.13	Thesis aims	30
Chapter 2	Materials and Methods	31
2.1	Materials	31
2.1.1	Solutions, buffers and media	31
2.1.2	Loading dyes	36
2.1.3	Nucleotides	37
2.1.4	Size markers and ladders	37
2.1.5	Sources of DNA and RNA	39
	Methods	40
2.2	Agarose gel preparation and electrophoresis	40
2.3	Sequencing gel	40
	Mapping and sequencing	41
2.4	Restriction Digest Fingerprinting	41
2.4.1	Filterprep isolation of BAC DNA	41
2.4.2	Restriction digest fingerprinting (<i>Hind</i> III) of BAC DNA	42
2.5	Fluorescent <i>in-situ</i> hybridisation (FISH) mapping	43
2.5.1	Labelling of FISH probe using Nick translation	44
2.5.2	Preparation of microscope slides	45

2.5.3	Hybridisation of FISH probes	46
2.6	Production of shotgun libraries for shotgun sequencing	48
2.6.1	Sonication and subfragment end repair of plasmid DNA	48
2.6.2	Selection of suitably sized DNA fragments for subcloning	50
2.6.3	Ligation into pUC18 vector	51
2.6.4	Transformation of pUC18 vector	52
2.7	Shotgun sequencing	53
2.7.1	Vacuum preparation of template DNA in pUC18 vector	53
2.7.2	The sequencing reaction	55
2.7.3	Sequencing instrumentation	56
2.7.3.1	ABI PRISM 373 DNA sequencer set-up	56
2.7.3.2	ABI PRISM 377 DNA sequencer set-up	57
2.7.4	Data analysis of shotgun sequencing reactions and clone assembly	57
2.7.5	Contiguation or ‘finishing’ of a clone	58
2.7.5.1	‘Finishing’ PCR reaction	59
	Expression profile analysis	61
2.8	Design of paralogue specific primers	61
2.9	PCR amplification of paralogue specific PCR products	62
2.10	Total RNA extraction from mammalian cell-lines	62
2.11	DNase treatment of RNA	64
2.12	First strand synthesis cDNA synthesis and amplification of target cDNA using paralogue specific primers	65
2.13	Overview of microarray experiments	66
2.13.1	Description of microarrays used	67
2.13.2	Generation of paralogue specific PCR products with a Universal Adaptor for use on microarrays	68

2.13.3	Generation of fluorescently labelled DNA	69
2.13.3.1	Generation of fluorescently labelled paralogue-specific PCR products using the Cyanine 3-dCTP dye for hybridisation onto the 'Paralogue Microarray'	69
2.13.3.2	Generation of fluorescently labelled single-stranded cDNA target using direct incorporation of Cyanine dyes for hybridisation onto the '10K/Paralogue Microarray'	70
2.13.4	Hybridisation, washing and scanning of microarrays	71
2.13.5	Analysis of microarrays	72
2.14	Overview of blot expression analysis	73
2.14.1	Radioactive labelling of DNA	74
2.14.1.1	Radioactive labelling of paralogue-specific PCR products	74
2.14.1.2	Radioactive labelling of DNA using MegaPrime™ DNA labelling system	74
2.14.2	Probe verification	75
2.14.2.1	Assessment of radiolabel incorporation using thin-layer chromatography	75
2.14.2.2	Measurement of radioactively labelled PCR product concentration	75
2.14.3	Manufacture of Southern Blots	76
2.14.3.1	Restriction digest of human genomic DNA	76
2.14.3.2	Transfer of digested genomic DNA onto filter	76
2.14.4	Hybridisation of radiolabelled PCR product to blots	77
2.14.5	Washing	77
2.15	Computational analysis	79
2.15.1	General programs used in this thesis	79
2.16	Identification of extended MHC paralogous genes in the human genome	81

2.16.1	Identification of extended MHC paralogues based on protein sequence homology	81
2.16.2	Identification of extended MHC paralogues with increasing levels of confidence	83
2.16.2.1	Filter 1: Domain-masking	83
2.16.2.2	Filter 2: FINEX	84
2.17	<i>In-silico</i> expression analysis	85
2.18	Clustering methods	87
2.19	Phylogenetic analysis	88
2.19.1	Protein sequence alignments	88
2.19.2	Estimation of the gamma distribution	89
2.19.3	Bootstrapping and tree-puzzling steps	89
2.19.4	Phylogenetic analysis using distance methods	90
2.19.4.1	PHYLIP	91
2.19.4.2	MEGA2	91
2.19.5	Phylogenetic analysis using the maximum likelihood method	92
2.19.5.1	PHYLIP	92
2.19.5.2	TREE-PUZZLE	93
2.20	Useful web-sites	93
Chapter 3	Characterisation of 9q32-q34.3	95
3.1	Introduction	95
3.2	Results	97
3.2.1	Identification of genes on 9q32-q34.3	97
3.2.2	Mapping of the Olfactory Receptor gene cluster to 9q33.1-q34.12	98
3.2.3	Identification of the Allograft Inflammatory Factor 1 (AIF1) paralogue	101

3.2.4	Problems associated with using mapping data and draft sequence	105
3.2.5	Orientation of contigs containing putative paralogues	107
3.2.6	Current status of 9q32-q34.3	109
3.2.7	Comparison of the MHC paralogous region on 9q32-q34.3 and the MHC region on 6p22.2-p21.3	110
3.2.7.1	Gene and paralogue content	110
3.2.7.2	Genomic landscape	114
3.2.7.3	Evidence of gene and segmental duplication	116
3.2.7.4	Diseases associated with 9q32-q34.3	119
3.3	Discussion	121
Chapter 4	Identification of the extended MHC paralogues in the human genome	124
4.1	Introduction	124
4.2	Strategy used to identify MHC paralogues	125
4.2.1	MHC genes used in the whole-genome survey	125
4.2.2	Identification of MHC paralogues with increasing levels of confidence	128
4.3	Definitions	131
4.3.1	L0-paralogues	131
4.3.2	L1-paralogues	131
4.3.3	L2-paralogues	132
4.3.4	L3-paralogues	133
4.4	Results	134
4.4.1	Identification of MHC paralogues: RXRB as an example	134
4.4.2	Identification of all the MHC paralogues in the human genome	137

4.4.3	Distribution of MHC paralogues in the human genome	142
4.4.4	MHC paralogues located on chromosome 1, 9 and 19	144
4.4.4.1	Chromosome 1 paralogues	146
4.4.4.2	Chromosome 9 paralogues	148
4.4.4.3	Chromosome 19 paralogues	150
4.4.4.4	Putative paralogues not identified in the genome-wide survey	152
4.4.4.5	Comparison of the order of L2- and L3-paralogues located on chromosomes 1, 9 and 19	153
4.4.5	Paralogues located outside the paralogous regions	157
4.4.6	L0- and L1-paralogues	159
4.4.7	Caveats associated with my strategy	160
4.5	Discussion	164
Chapter 5	Phylogenetic analysis of extended MHC paralogous gene families	167
5.1	Introduction	167
5.2	MHC paralogous gene families used in phylogenetic analysis	170
5.3	Results	173
5.3.1	Phylogenetic analysis of the BRD paralogous gene family	174
5.3.2	Phylogenetic analysis of the PBX paralogous gene family	175
5.3.3	Phylogenetic analysis of the NOTCH paralogous gene family	176
5.3.4	Phylogenetic analysis of the complement paralogous gene family	178
5.3.5	Phylogenetic analysis of the RXR paralogous gene family	179
5.3.6	Phylogenetic analysis of the tenascin paralogous gene	181

	family	
5.3.7	Phylogenetic analysis of the AIF paralogous gene family	183
5.3.8	Phylogenetic analysis of the β -tubulin paralogous gene family	184
5.3.9	Phylogenetic analysis of the GPX paralogous gene family	189
5.3.10	Phylogenetic analysis of the CLIC paralogous gene family	191
5.4	Discussion	193
Chapter 6	Expression analysis of extended MHC paralogous gene families	196
6.1	Introduction	196
6.2	Terminology	199
6.3	Results	200
6.3.1	Cross-hybridisation (control) experiments	200
6.3.2	Expression profiling	203
6.3.2.1	<i>In-silico</i> analysis	203
6.3.2.2	Dot-blot analysis	206
6.3.2.3	Northern blot analysis	209
6.3.2.4	Microarray analysis	212
6.3.2.5	Importance of designing specific microarray targets	218
6.3.3	Interpretation of expression data	219
6.3.3.1	Tenascin paralogous gene family	219
6.3.3.2	Microarray expression data	220
6.3.3.3	<i>In-silico</i> expression data	223
6.3.3.4	Dot-blot expression data	225

6.3.3.5	Comparison of the expression profiles of the MHC paralogues located in the paralogous regions on chromosomes 1, 9 and 19	227
6.3.3.6	Comparison of the methods used to generate expression profiles	228
6.4	Discussion	230
Chapter 7 Conclusions and future work		234
7.1	Conclusions	234
7.2	Future work	241
Bibliography		244
Appendices		263
Appendix 1	9q32-q34.3 annotation	264
Appendix 2	Whole-genome survey results	272
Appendix 3	Primers	287
Appendix 4	Primers	289
Appendix 5	<i>In-silico</i> results	291
Appendix 6	Dot blot results	294
Appendix 7	Northern blot results and transcript sizes	297
Appendix 8	Microarray results	299
Appendix 9	Comparison of methods	300

List of tables

Chapter 1 Introduction

1.1	Gene number and genome size for a range of organisms	8
-----	--	---

Chapter 3 Characterisation of 9q32-q34.3

3.1	Summary of the first MHC paralogues identified in three other regions of the genome	96
3.2	Summary of the exon and intron sizes and comparison of splicing phases of the two AIF1 paralogues	103
3.3	Summary of the gene content and sizes of chromosomes 6 and 9 and the paralogous regions	111
3.4	Comparison of the repeat content of the 6p22.2-p21.3 and 9q32-q34.3	116
3.5	Summary of some of the disorders associated with 9q32-q34.3	120

Chapter 4 Identification of the extended MHC paralogues in the human genome

4.1	Distribution of genes in the extended MHC region	127
4.2	Summary of the MHC genes with paralogues with increasing levels of support	137
4.3	Summary of the distribution of MHC paralogues in the human genome	142
4.4	Summary of the L2- and L3-paralogues on chromosome 1	146
4.5	Summary of the L2- and L3-paralogues on chromosome 9	149
4.6	Summary of the L2- and L3-paralogues on chromosome 19	151
4.7	Summary of the putative MHC paralogues not identified in my genome-wide survey	153
4.8	Summary of the MHC paralogues located outside the paralogous regions on chromosomes 1, 9 and 19	158

4.9	Summary of the P-values obtained for the HLA class I-like genes from the BLAST similarity search using HFE, HLA-A, HLA-E, MICA and MICB, and the percentage sequence identities determined from a global sequence alignment	161
Chapter 5 Phylogenetic analysis of extended MHC paralogous gene families		
5.1	Summary of the MHC paralogous gene families used to generate phylogenetic trees	173
5.2	Summary of the TUBB paralogues in the human genome	185
Chapter 6 Expression analysis of extended MHC paralogous gene families		
6.1	Comparison of three methods used to generate the expression profiles for nine MHC paralogous gene families	228

List of figures

Chapter 1 Introduction

1.1	The 2R hypothesis	4
1.2	Distribution of Hox gene clusters in the human genome	5
1.3	Time-line of a range of genome sequencing projects	6
1.4	Progress of the Human Genome Project from the launch in 1990 to its completion in 2003	7
1.5	Models of genome duplication by autotetraploidisation and allotetraploidisation	15
1.6	Karyotype of a male tetraploid <i>Tympanoctomys barrerae</i> from Mendoza, Argentina taken from Gallardo <i>et al</i> (1999)	16
1.7	Schematic representation of the extended MHC class I region	19
1.8	The MHC class I region	21
1.9	The MHC class III region	22
1.10	The MHC class II region.	24
1.11	The extended MHC class II region	25
1.12	Summary of the MHC paralogous regions in the human genome	27

Chapter 2 Materials and Methods

2.1	The 10K/Paralogue Microarray	69
-----	------------------------------	----

Chapter 3 Characterisation of 9q32-q34.3

3.1	FISH analysis of bA465F21	99
3.2	Localisation of the clone bA465F21 to the chromosome 9 tiling path	100
3.3	Computational identification of the AIF1 paralogue	102
3.4	ClustalX sequence alignment of the two AIF1 paralogues	104

3.5	Overview of the gene content of region analysed to identify a putative GPX5 paralogue	106
3.6	Overview of methods used to order and orientate the contigs containing RALGDS and BRD3 putative paralogues	108
3.7	Schematic representation of the status (August 2003) of the MHC paralogous region on 9q32-q34.3	109
3.8	Comparison of the order of paralogues between the MHC region on 6p22.2-p21.3 and the paralogous region on 9q32-q34.3	112
3.9	Evolution of the lipocalin paralogous gene family on 9q34	118
Chapter 4 Identification of the extended MHC paralogues in the human genome		
4.1	Overview of the strategy used to identify MHC paralogues with increasing levels (L0 to L3) of confidence and definitions	126
4.2	Alignment of the exon fingerprints of the extended MHC class I gene RXRB and its paralogues, RXRA and RXRG, identified in the genome survey	129
4.3	Protein sequence alignment of the extended MHC class II encoded protein, RXRB, and its two paralogues, RXRA and RXRG	130
4.4	Summary of the results of the initial (A) and domain-masked (B) TBLASTN search of the human genome using the RXRB protein sequence	135
4.5	Summary of the FINEX search using the RXRA fingerprint	136
4.6	Summary of the results of the whole-genome survey using 128 MHC genes	138
4.7	Summary of the proportion (%) of BLAST hits corresponding to the paralogues with different levels of confidence	139
4.8	Summary of the MHC genes with L0- to L3-paralogues	140
4.9	Summary of the percentage (%) of MHC genes with no, 1, 2, 3, 4 or more L0, L1, L2 and L3-paralogues in the human genome	141
4.10	Distribution of MHC paralogues in the human genome	143
4.11	Summary of MHC paralogues on chromosomes 1, 9 and 19	144

4.12	Comparison of the order of L2- and L3-paralogues on chromosomes 1, 9 and 19	154
4.13	Comparison of the MHC paralogues with copies on all four paralogous regions	156
Chapter 5 Phylogenetic analysis of extended MHC paralogous gene families		
5.1	Summary of the 2R hypothesis	168
5.2	Schematic representation of the effects of two rounds of gene, or genome, duplication on the topology of the phylogenetic tree and the resulting number of paralogues in ‘key’ species	169
5.3	Schematic representation of the ‘ideal’ phylogenetic tree in support of the 2R hypothesis	170
5.4	Summary of the MHC genes and paralogues selected for further investigation	171
5.5	Phylogenetic tree of the BRD paralogous and orthologous family	175
5.6	Phylogenetic analysis of the PBX paralogous gene family	176
5.7	Phylogenetic analysis of the NOTCH paralogous gene family	177
5.8	Phylogenetic analyses showing the relationship of the C4 paralogues and orthologues	179
5.9	Phylogenetic tree showing the evolutionary relationship between the RXRB paralogues and orthologues	180
5.10	Phylogenetic analyses of the TNXB paralogues and orthologues	182
5.11	Phylogenetic tree of the AIF1 paralogues and orthologues	183
5.12	Phylogenetic analysis of the β -tubulin paralogues and orthologues	186
5.13	Phylogenetic tree showing the ancient duplication events that have shaped the present day β -tubulin paralogues and orthologues	188
5.14	Phylogenetic analysis of the GPX family	190
5.15	Phylogenetic analysis of the CLIC family	191

Chapter 6 Expression analysis of extended MHC paralogous gene families		
6.1	Fates of duplicated genes	197
6.2	Verification of probe specificity	202
6.3	Summary of the results of the <i>in-silico</i> expression analysis of the BRD2 gene and its three paralogues	205
6.4	Transcription pattern of the AIF1, AIF1L and β -actin control genes after hybridisation with paralogue-specific probes to the dot blot with RNA from different tissues	207
6.5	Transcription pattern and splice variants of the BRD2, BRD3 BRD4, BRDT and β -actin control genes after hybridisation with specific probes to a Northern blot with eight different tissues	210
6.6	Assessment of the quality of the eleven RNAs used in the expression microarray experiments	213
6.7	Results of a hybridisation with the standard Stratagene RNA to the ‘10K/Paralogue Microarray’	214
6.8	One of the 48 sub-arrays of the ‘10K/Paralogue Microarray’ after hybridisation using the Stratagene standard RNA	215
6.9	Microarray results confirmed by RT-PCR	217
6.10	Comparison of the expression profiles of the paralogue specific PCR products designed in this thesis and those already on the standard Sanger Institute 10K microarray corresponding to GPX4 and BRD3 genes and the key to the tissues and cell-lines used	218
6.11	Expression profile of the TNXB gene indicates that it is adrenal gland specific	220
6.12	Summary of the microarray expression data and the result of applying Hierarchical clustering methods	221
6.13	Clustering of the <i>in-silico</i> expression profile results	223
6.14	Clustering of the dot-blot expression profile results	225
6.15	Comparison of the expression profiles of the paralogues located within the paralogous regions on chromosomes 1, 9 and 19 with the MHC genes using <i>in-silico</i> and dot blot analysis in 28 normal human tissues.	227

Abbreviations

aa	amino acid
AIF	Allograft inflammatory factor
ATP	adenosine 5'-triphosphate
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
bp	base pair
BRD	Bromodomain containing protein
°C	degrees Celsius
cDNA	complementary deoxyribonucleic acid
CLIC	Chloride intracellular chloride channel
CTP	cytidine 5'-triphosphate
dbEST	database of expressed sequence tags
DNA	deoxyribonucleic acid
dNTP	2'-deoxyribonucleoside 5'-triphosphate
DTT	dithiothreitol
EDTA	ethylenediamine tetra-acetic acid
EMBL	European Molecular Biology Laboratory
EST	expressed sequence tag
FISH	Fluorescent <i>in-situ</i> hybridisation
FPC	fingerprinting contig
GPX	Glutathione peroxidase
GTP	guanine 5'-triphosphate
HGMP	Human Genome Mapping Resource Centre
HGP	Human Genome Project
HLA	human leukocyte antigen
IHGSC	International Human Genome Sequencing Consortium
kb	kilobase pairs
l	litre
-L	-like
LB	Luria-Bertani
LINE	long interspersed nuclear element

M	molar
mA	milliamps
Mb	megabase pairs
μg	microgram
μl	microlitre
μM	micromolar
min(s)	minute(s)
MIPS	Munich Information Centre for Protein Sequences
mg	milligram
MHC	Major Histocompatibility Complex
ml	millilitre
mm	millimetre
mM	millimolar
NCBI	National Centre for Biotechnology Information
ng	nanogram
NOTCH	Neurogenic locus Notch homologue
OR	Olfactory receptor
PCR	polymerase chain reaction
PFAM	protein family database
PBX	Pre-B cell leukaemia transcription factor
RNA (mRNA, rRNA, tRNA)	ribonucleic acid (messenger-, ribosomal-, transfer-)
rpm	revolutions per minute
RT-PCR	reverse transcription polymerase chain reaction
RXR	Retinoic acid receptor
SDS	sodium dodecyl sulphate
sec(s)	second(s)
SINE	short interspersed nuclear element
STS	sequence tagged site
TEMED	N, N, N', N'-tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
U	unit
UTR	untranslated region
V	volt