Identification and analysis of MHC-linked olfactory receptor genes

Ruth M Younger

A thesis submitted in partial fulfilment of the requirements of Cambridge University for the degree of Doctor of Philosophy.

Newnham College, Cambridge University.

August 2002

This dissertation does not exceed the length limit set by the Biology Degree Committee.

Abstract/Summary of thesis

Olfactory receptor genes (ORs) are members of the largest known human multigene family of 7-transmembrane receptors; over the past few years, 900-1000 have been identified, existing largely within clusters located on the majority of chromosomes. As their name implies, they are involved in odour perception within the major olfactory epithelium (MOE), although their expression in other tissues, notably the testis, lungs and kidney seems to suggest some non-olfaction related role. The cluster of olfactory receptors located next to the Major Histocompatibility Complex (MHC) were considered to be especially interesting, given that the association between this cluster and the MHC has also been conserved in mouse and rat, and it has been suggested that these genes are involved in detecting odours that control MHC-disparate mate selection.

The aim of this thesis was to identify all olfactory receptor genes located distal to the class I region of the Major Histocompatibility Complex in both the human genome and in the mouse genome, and to study phylogeny, regulation, expression and polymorphism to develop a better understanding of the structure, function and evolution of these genes. In total, 34 MHC-linked ORs were identified in human; within the mouse genome, a larger number of MHC-linked ORs (56) were also identified. Comparing the 2 species, orthologous groups can be identified: groups appear to have undergone both deletions and duplications since mouse-human divergence. An investigation of regulatory elements within the human MHC-linked OR cluster revealed no specific regulatory elements, although a putative locus control region has been identified. Compared to the HLA genes, these OR genes show a limited amount of polymorphism, although existing polymorphisms act to alter the functional repertoire of these genes between individuals and one specific OR gene appears to have an unexpectedly high level of polymorphism.

Acknowledgements

I am indebted to my supervisor at the Sanger Centre, Stephan Beck, whose knowledge, enthusiasm, patience and resourcefulness have seen me through whenever mine have failed. Many thanks also go to my supervisor at Cambridge University, John Trowsdale. I am also very grateful to a number of people with whom I have worked over the course of this project: Armin Volz and Anke Ehlers at Institut für Immungenetik, Universitätsklinikum Charité, Humboldt-Universität, Berlin, Simon Forbes at Cambridge University, and Claire Amadou at CNRS, Toulouse – it has been both a highly informative and a highly enjoyable experience. Special thanks must also go to Andreas Ziegler (Institut für Immungenetik, Universitätsklinikum Charité, Humboldt-Universität, Berlin) – your frequent phone calls were a source of inspiration, motivation and (I admit it, occasionally) fear ! Many thanks also to Kirsten Fischer-Lindahl (HHMI, Dallas, TX) – I have appreciated all your input throughout the course of this thesis.

Moving on to the Sanger Centre, I would like to thank all past and present members of Team 50: Katie Evans, Roger Horton, Melanie Stammers, Vikki Rand, Karen Novik, and Karen Halls – thanks for all the input and advice and the cups of tea and moaning (mine, generally)... I am also very grateful to members of the now (non-existent) Team 33 who did much of the sequencing of this region, as well as members of the Chromosome 6 and mouse project groups at the Sanger Centre. Others at the Sanger Centre who have helped me over the course of this project are too many to name in full - I am privileged to have worked in an environment where help has been offered so freely – but special mention should go to Alex Bateman, Kevin Howe, Matt Craig, Graeme Bethel, Carol Edwards and Andy Mungall. Thanks also to Jill Williamson and Denise Sheer of the Human Cytogenetics Laboratory (ICRF) for their help with the FISH analysis.

On a personal note, thanks must go to the various football teams I have played for and supported whilst doing this thesis – you have provided me with a philosophical approach that being associated with more successful clubs just would not be able to give me and distracted me when I most and least needed it. Thanks to my friends who have allowed me to become even less responsible and cut me huge amounts of slack over the last few months – your understanding and patience is always a lot more than I deserve. I am also grateful to my own source of sunshine, even though it got cloudy, the cloud breaks were great !! Many thanks also go to my parents, Michael and Glynis, for their pillar-like support, unswerving belief and love and the many sandwiches they have made me, especially in the last few weeks. Finally, thanks to the force that has moved in a most mysterious way to shape the big and small events in the universe – I am never grateful enough for everything I have.

Table of contents

List of tabl	es	xiii
List of figu	res	xiv
List of abb	List of abbrieviations	
Publications		xviii
Chapter 1	Introduction	1
1.1	The olfactory system	1
1.2	The major olfactory epithelium: anatomical organization and signal transduction	2
1.3	The molecular basis of the olfactory system	5
1.4	Cloning of further olfactory receptor genes in a variety of species	8
1.5	The genomic organization of olfactory receptor genes	10
1.6	The expression of olfactory receptor genes within the olfactory system	11
1.7	The expression of olfactory receptor genes outside the olfactory system	14
1.8	The regulation of olfactory receptor genes	16
1.9	Ligands of olfactory receptor genes	16
1.10	The accessory olfactory system	18
1.11	The Major Histocompatibility Complex (MHC)	19
1.12	MHC genetic diversity and reproductive selection	23
1 13	Aims of this thesis	26

Cha	pter 2	Materials and Methods	28
2.1		Materials	28
	2.1.1	Solutions	28
	2.1.2	Media	30
	2.1.3	Loading dyes	30
	2.1.4	Buffers	31
	2.1.5	Markers	31
	2.1.6	Sequencing gel	32
	2.1.7	Antibiotics	32
	2.1.8	Web addresses and other in silico resources	33
2.2		Fluorescent fingerprinting	33
	2.2.1	Preparation of DNA for fluorescent fingerprinting	34
	2.2.2	Generation of fluorescent marker	35
	2.2.3	The restriction digest reaction	35
	2.2.4	Data analysis	36
2.3		Fluorescent in situ hybridisation (FISH) mapping	37
2.4		The production of PUC and M13 libraries	39
	2.4.1	Isolation of PAC / BAC DNA using the caesium chloride procedure	39
	2.4.2	Sonication and subfragment end repair of plasmid DNA	41
	2.4.3	Selection of suitably sized fragments for subcloning	43
	2.4.4	Ligation and transformations	44
2.5		Shotgun sequencing	48
	2.5.1	Preparation of template DNA in M13 vector	48
	2.5.2	Preparation of template DNA in pUC18 vector	50
	2.5.3	The sequencing reaction	51
	2.5.4	Sequencing instrumentation	52
2.6		Data analysis of shotgun sequencing reactions and clone assembly	54
	2.6.1	PCR reaction used in clone assembly	55
	2.6.2	Restriction digests used to check veracity of clone assembly	56
2.7		Construction of mouse filters for hybridisation	57
2.8		Construction of mouse OR gene vectors for in situ hybridisations	58
	2.8.1	pGEM T-Easy ligations and transformations	59

	2.8.2	pGEM T-Easy preparation of DNA	60
2.9	2.9.1 2.9.2 2.9.3	Construction of 'olfactory promoter region' pGL3 reporter vectors pGL3 reporter vector restriction digests, ligations and transformations Preparation and sequencing of pGL3 reporter vector DNA Transfections of pGL3 reporter vector DNA	61 62 63 63
2.10		The pGL reporter vector assay	65
2.11		Polymorphism analysis	66
2.12		Mouse RT-PCR	67
2.13		Human RNA dot-blot hybridisations	67
2.14		Titering the olfactory epithelium phage library	69
2.15		PCR amplification of the olfactory epithelium phage library	70
2.16		In silico analysis: gene finding programs	71
2.17		Identification of genes	72
2.18		Identification of olfactory receptor genes	75
2.19		The assembly of DNA clone sequences	75
2.20		Storage of information in sequence feature files	76
2.21		The nomenclature of olfactory receptor genes and the 'ROLF' database	77
2.22		Gene-finding approach for olfactory receptor genes	78
2.23		Analysis of olfactory receptor protein structure	79
2.24		Other programs/scripts used throughout the thesis	79
2.25		Genome browsers	80
2.26	2.26.1	Analysis of regulatory regions within OR clusters 'Promoter Inspector', 'Eponine', and 'Transfac'	80 80

	2.26.2	DNA block aligner ('DBA')	81
2.27		Analysis of comparative regions in the human and mouse cluster: 'Pipmaker'	82
2.28	2.28.1 2.28.2	Phylogenetic analysis Protein alignments Phylogenetic tree production	82 82 83
Chaj	pter 3	The human MHC extended class I region	87
3.1		Introduction	87
3.2		Sequence assembly and gene content of the human MHC extended class I region	88
3.3		Gene clusters in the human MHC extended class I region	91
	3.3.1	The histone cluster	91
	3.3.2	The zinc finger protein cluster	92
	3.3.3	The ribosomal protein gene cluster	92
	3.3.4	The butyrophilin cluster	94
	3.3.5	The tRNA cluster	95
3.4			96
		Related genes within the human MHC extended class I region	
	3.4.1	The glutathione peroxidase loci	96
	3.4.2	The FAT gene and pseudogene	97
	3.4.4	The G-protein coupled receptor (GPCR) loci	97
	3.4.4	The POM121-like loci	98
	3.4.5	The RoRet gene and the Ret Finger Protein (RFP) gene	98
3.5		Other known genes within the human MHC extended class I region	99
3.6		'Novel' genes identified within the human MHC extended class I region	106
3.7		The genomic environment of the human MHC extended class I region	107
3.8		Duplications within the human MHC extended class I region	117
3.9		Conclusions	117

Chapter 4	The human MHC-linked olfactory receptor (OR) genes	120
4.1	Introduction	120
4.2	Identification of OR genes and OR subfamilies	121
4.3	Conservation of amino acids in OR proteins	125
4.4	Human MHC-linked OR pseudogenes	131
4.5	The genomic environment of human MHC-linked OR genes	132
4.6	Local duplications within the human MHC-linked OR cluster	137
4.7	Conclusions	140
Chapter 5	The mouse MHC-linked olfactory receptor (OR) contig and comparative analysis of human and mouse OR gene clusters	142
5.1	Introduction	142
5.2	Constructing the mouse OR contig	144
5.3	Sequence assembly of the mouse OR contig	147
5.4	Identification of mouse MHC-linked OR genes	148
5.5	Identification of mouse MHC-linked OR subfamilies	149
5.6	Conservation of amino acids in mouse MHC-linked OR proteins	152
5.7	Mouse MHC-linked OR pseudogenes	154
5.8	The genomic environment of mouse MHC-linked OR genes	156
5.9	MHC class I and class I-like genes within the MHC-linked OR contig	160
5.10	Other genes located within the MHC-linked OR contig	161
5.11	Local duplications of mouse MHC-linked OR genes	164

5.12	Local duplications of MHC class I and class I-like genes	171
5.13	Identification of MHC-linked OR orthologs in mouse and human	173
5.14	Conservation of sequence outside OR coding regions in mouse and human	175
5.15	Non-orthologous OR genes	180
5.16	Identification of orthologous ORs upstream of the original contig	181
5.17	Conservation of orthologs in non MHC-linked OR clusters	186
5.18	Conclusions	188
Chapter 6	Regulation and expression of MHC-linked OR genes	190
6.1	Introduction	190
6.2	In silico transcript analysis	192
6.3	Experimental analysis of expression of MHC-linked OR genes	200
6.4	Regulation of the generation of alternative transcripts in MHC-linked OR genes	204
6.5	Regulation of expression within the MHC-linked OR cluster	209
6.6	Comparison of upstream regions of MHC-linked OR genes	212
6.7	Comparison of upstream regions of human and mouse MHC-linked OR genes	215
6.8	Conclusions	218
Chapter 7	Polymorphisms of MHC-linked OR genes	221
7.1	Introduction	221
7.2	Alleles of hsM1-17	222

7.3	Alleles of hsM1-20	224
7.4	Alleles of other MHC-linked OR genes	225
7.5	Single nucleotide polymorphisms (SNPs) large scale analysis	229
7.6	SNPs of MHC-linked ORs	233
7.7	Conclusions	236
Chapter 8	Comparison of MHC-linked ORs and other human ORs	238
8.1	Introduction	238
8.2	The human OR database ('ROLF') and the genomic location of OR genes	239
8.3	Description of OR clusters and local duplications	242
8.4	The genomic environment of OR clusters	247
8.5	Phylogenetic analysis of human OR genes	249
8.6	The evolutionary origins of the MHC-linked OR cluster	252
8.7	Paralogous MHC regions and the MHC-linked OR cluster	253
8.8	OR pseudogenes	256
8.9	Conclusions	259
Chapter 9	MHC-linked vomeronasal receptor (VR) genes	262
9.1	Introduction	262
9.2	Identification of VR genes in the human extended MHC class I region	263
9.3	Identification of VR genes in the human genome	264

9.4	Identification of VR genes in the syntenic mouse region	265
9.5	Conclusions	267
Chapter 10	Final discussion and conclusions	269
Bibliography	,	272
Appendices		303
Appendix 1	: Primers used in this thesis	303
Appendix 2:	Example of sequence feature files and example of output from 'olfgrab'	307
Appendix 3	: List of websites used in thesis	309
Appendix 4	: List of scripts and programs used in this thesis	311
Appendix 5:	List of clones assembled to produce MHC extended class I region	316
Appendix 6:	List of human MHC-linked ORs	317
Appendix 7:	Alignment of human MHC-linked ORs	319
Appendix 8:	List of human MHC-linked ORs	322
Appendix 9:	Alignment of human MHC-linked ORs	324
Appendix 10): List of mouse MHC-linked genes from UCSC assembly	328
Appendix 11	: Human 'ROLF' database	329

List of tables

Table 3.1	The distribution of ribosomal protein genes in the extended MHC class I region	94
Table 3.2	tRNA types and the number of genes per type found within the extended MHC class I region	95
Table 3.3	'Novel' genes in the extended MHC class I region	106
Table 3.4	Size and gene density of subsections of the human MHC extended class I region	111
Table 3.5	Isochores of the human MHC extended class I region	113
Table 4.1	Subfamily designations of human MHC-linked OR genes	122
Table 5.1	Clones contributing to mouse contig	148
Table 5.2	Subfamily designations of mouse MHC-linked OR genes	150
Table 5.3	Nucleotide percentage identities within mouse OR subfamily 9	168
Table 6.1	ESTs matching MHC-linked ORs	193
Table 6.2	TRANSFAC matches found in the 'olfactory promoter' region	207
Table 6.3	Results from pGL3 reporter vector assay	208
Table 7.1	Polymorphisms observed in hs6M1-17	223
Table 7.2	Polymorphisms observed in hs6M1-20	225
Table 7.3	Summary of all polymorphisms found within MHC-linked OR genes	227
Table 7.4	SNPs in MHC-linked ORs	234
Table 8.1	The repeat and GC content of OR clusters within the human genome	248

List of figures

Figure 1.1	Anatomy of the olfactory system	2
Figure 1.2	The structure of the major olfactory epithelium	3
Figure 1.3	Sensory transduction in olfactory sensory neurons (OSNs).	5
Figure 1.4a	Schematic diagram of the protein structure of an olfactory receptor	8
Figure 1.4b	Proposed three dimensional structure of an olfactory receptor protein	8
Figure 1.5	OR targeting within the olfactory bulb	13
Figure 1.6	Gene map of the human MHC	23
Figure 1.7	Factors contributing to MHC genetic diversity	25
Figure 2.1	NIX display of AL031983	73
Figure 2.2	AceDB display of AL031983	74
Figure 3.1	Gene content of human MHC extended class I region	90
Figure 3.2	Subsections of the human MHC extended class I region	108
Figure 3.3	Isochores of the human MHC extended class I region	110
Figure 4.1	Phylogenetic tree of human MHC-linked OR genes	124
Figure 4.2	A proposed model of evolution for the MHC-linked OR genes	125
Figure 4.3	Amino acid conservation in human MHC-linked ORs	127
Figure 4.4a	Cysteine conservation in human MHC-linked OR	130
Figure 4.4b	Proposed OR structure showing potential disulphide bridges	130
Figure 4.5	A comparison of the mutations causing frameshifts in 2 MHc-linked ORs	132
Figure 4.6	The genomic organization of the human MHC-linked major and minor OR clusters	134
Figure 4.7	A dot-matrix plot showing the major OR cluster plotted against itself	137
Figure 4.8	Local duplication within the major MHC-linked OR cluster creating 2 new ORs	139
Figure 5.1	Comparative gene map of the human, mouse and rat MHC	143
Figure 5.2	FISH analysis of bM332P19	145
Figure 5.3	Mouse MHC-linked OR contig	146
Figure 5.4	Phylogenetic tree of mouse MHC-linked OR genes	151
Figure 5.5	Conserved amino acids in mouse MHC-linked OR proteins	154
Figure 5.6	A base insertion in mouse pseudogenes mm17M1-30P and mm17M1-16/17P	156
Figure 5.7	The genomic organisation of the mouse MHC-linked OR cluster	157
Figure 5.8	Dot-matrix plot of the mouse MHC-linked OR contig	165
Figure 5.9	Block duplications within the mouse MHC-linked OR cluster	170
Figure 5.10	Duplications associated with H2-M3 loci in the mouse MHC-linked OR cluster	172

Figure 5.11	Orthologous OR genes within the MHC-linked OR clusters in human and mouse	173
Figure 5.12	Dot-matrix plot of the mouse MHC-linked OR cluster against the syntenic human region	175
Figure 5.13a	mm17M1-45 PIP identity plot	177
	9 mm17M1-45 plotted against hs6M1-25P	177
	mm17M1-1 plotted against hs6M1-12 and hs6M1-13P	178
	hs6M1-21 plotted against mm17M1-23 and mm17M1-33	180
	Orthologous ORs within the MHC major OR cluster	183
	Orthologous ORs within the MHC minor OR cluster	183
	Orthologous clusters of OR genes on mouse chromosome 2 and human chromosome 9	187
Figure 5.17b	Orthologous clusters of OR genes on mouse chromosome 2 and human chromosome 11	187
Figure 6.1	EST supported splicing in hs6M1-21	194
Figure 6.2	Alignment of ESTs to hs6M1-16	196
	EST associated with hs6M1-32	197
Figure 6.3b	Alignment of EST N68399 to genomic sequence, including hs6M1-32	197
Figure 6.4a	Dot-blot of hs6M1-16	202
Figure 6.4b	Key to the dot-blot	202
Figure 6.5a	Alternative splice forms of hs6M1-21, -27, -18 and -16	206
Figure 6.5b	Enlarged section showing shared exons and putative promoter sequence	206
Figure 6.6	Promoter prediction within MHC-linked OR clusters	211
Figure 6.7	'DBA' alignment of 4 Kb upstream region of hs6M1-16	214
Figure 6.8	PIP plot of a region of the MHC-linked OR cluster	216
Figure 7.1a	Polymorphisms found in human MHC-linked ORs	230
	Percentage of polymorphisms per regions of the consensus MHC-linked OR protein	230
Figure 7.2	Distribution of SNPs across the major and minor MHC-linked OR clusters	232
Figure 8.1	The distribution of OR genes within the human genome	241
Figure 8.2	Phylogenetic tree of all OR genes within the human genome	251
Figure 8.3	Proposed MHC-linked OR evolution on chromosome 1, 6, 9 and 19	255
Figure 8.4	Distribution of stops and frameshifts in OR pseudogenes	257
Figure 9.1	Phylogenetic tree of VR genes in human MHC extended class I and VR genes on mouse chromosome 13	266

List of abbrieviations

AOB accessory olfactory bulb

(c)AMP (cyclic) adenosine 5'-monophosphate ATP (dATP, ddATP) adenosine 5'-triphosphate (deoxy-, dideoxy-)

BAC bacterial artificial chromosome BLAST basic local alignment search tool

bp base pair
BTN butyrophilin
°C degrees Celsius

cDNA complementary deoxyribonucleic acid

chr chromosome cm centimeter

CNS central nervous system

CpG cytidyl phosphoguanosine dinucleotide
CTP (dCTP, ddCTP) cytidine 5'-triphosphate (deoxy-, dideoxy-)
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

FAT10 HLA-F associated transcript 10 FISH fluorescence 'in situ' hybridisation

FPC fingerprinting contig
GABA gamma-aminobutyric acid
GDP guanine diphosphate
GPCR G-protein coupled receptor
G-protein GTP-binding protein
GPX glutathione peroxidase

GTP (dGTP, ddGTP) guanine 5'-triphosphate (deoxy-, dideoxy-)

HEK human embryonic kidney

HFE Hereditary haemochromatosis locus

HGMP Human Genome Mapping Resource Centre

HGP Human Genome Project
HLA human leukocyte antigen
HMM Hidden Markov Model
HUGO Human Genome Organisation

Ig Immunoglobulin

IHGSC International Human Genome Sequencing Consortium

Kb kilobase pairs

l litre

LB Luria-Bertani

LD linkage disequilibrium

LINE long interspersed nuclear element

M molar

Mb megabase pairs
μg microgram
μl microlitre

 μM micromolar min(s) minute(s) mg milligram

MHC major histocompatibility complex

millilitre ml mm millimetre millimolar mM

MOE major olfactory epithelium

NCBI National Centre for Biotechnology Information

nanogram ng olfactory bulb OB

On-line Mendelian Inheritance in Man **OMIM**

OR Olfactory receptor

OSN Olfactory sensory neuron

P1-derived artificial chromosome PAC **PCR** polymerase chain reaction **PFAM** protein family database RFP ret finger protein

RNA (mRNA, rRNA, tRNA) ribonucleic acid (messenger-, ribosomal-, transfer-)

ribonuclease A RNase A

revolutions per minute rpm RP ribosomal protein

RT-PCR reverse transcription polymerase chain reaction

sodium dodecyl sulphate SDS

second(s) sec(s)

SINE short interspersed nuclear element

small nucleolar RNA snoRNA

SNP single nucleotide polymorphism

sequence tagged site STS

N, N, N', N'-tetramethylethylenediamine **TEMED**

Transmembrane domain TM**TrEMBL** Translated EMBL database

Tris tris(hydroxylmethyl)aminomethane

U unit

UTR untranslated region

ultraviolet uv V volt

v/vvolume/volume **VNO** vomeronasal organ VR pheromone receptor

W watt

weight/volume w/v

WGS whole genome shotgun yeast artificial chromosome YAC

zinc finger protein **ZNF**

Publications

Publications describing parts of this thesis:

Ehlers, A., Beck, S., Forbes, S., Trowsdale, J., Uchanska-Ziegler, B., Volz, A., **Younger, R.**, and Ziegler, A. (2000) MHC-linked olfactory receptor loci exhibit polymorphism and define extended HLA/OR haplotypes *Genome Res* 10(12):1968-78.

Ziegler, A., Ehlers, A., Forbes, S., Trowsdale, J., Uchanska-Ziegler, B., Volz, A., **Younger, R.**, and Beck, S. (2000) *Polymorphic olfactory receptor genes and HLA loci constitute extended haplotypes*. In Kasahara, M. (ed) *Major Histocompatibility Complex: Evolution structure, and function*

Ziegler, A., Ehlers, A., Forbes, S., Trowsdale, J., Uchanska-Ziegler, B., Volz, A., **Younger, R.**, and Beck, S. (2000) Polymorphisms in olfactory receptor genes: a cautionary note. *Hum Immunol* 61(12):1281-4

Younger, R., Amadou, C., Bethel, G., Ehlers, A., Fischer Lindahl, K., Forbes, S., Horton, R., Milne, S., Mungall, A., Trowsdale, J., Volz, A., Ziegler, A., and Beck, S. (2001) Characterisation of clustered MHC-linked Olfactory Receptor Genes in Human and Mouse *Genome Res* 11(4):519-30

Ziegler, A., Beck, S., Ehlers, A., **Younger, R.** and Volz, A. (2002) *Testicular transcripts of HLA-linked olfactory receptor genes exhibit unorthodox features* In Hansen, J.A. and Dupont, B. (ed) *HLA 2002 Immunobiology of the Human MHC*