THE EVOLUTION OF
GENE ARRANGEMENTS AND
GENE FAMILIES IN TAMMAR WALLABY

Amir Mohammadi (M.Sc.)

A thesis submitted for the degree of Doctor of Philosophy of
The Australian National University

July 2011
DECLARATION

Except where specific reference is made to other sources, the work presented in this thesis is the work of the author. It has not been submitted, in whole or in part for any other degree.

Amir Mohammadi

14/07/2011
ACKNOWLEDGEMENT

This PhD thesis could have not been completed without the helps and supports I received from my supervisors, friends and family members. First of all I wish to deeply thank Prof. Jennifer A.M. Graves for accepting me into her lab and letting me to be part of the exciting Kangaroo genome project. Jenny you are the best mentor I could have ever imagined and your supports were always bigger than the obstacles I had on my road to this work. I also would like to express my deepest and most honest thanks to my supervisor, Dr. Margaret L. Delbridge for guiding me through this research and providing her endless supervision to me. Margaret, you might not know but every five minutes discussion with you was bringing me out of weeks of darkness and was a big encouragement for me to continue the work. Dr. Paul D. Waters and Dr. Janine E. Deakin have also been very influential and helpful during the course of this research. Not only as my advisors but also as good friends you guys were always there and very fast responsive to my questions or proof readings. Apart from the courage I learned many of my lab techniques from you two.

My family, especially my father and my mother cannot be thanked enough for their patience and supports during my studies. Coming to Australia and doing a PhD introduced a lot of pressure and difficulties to the life of each of you, but your kindness and smile was the only think I received in return. I’m especially and warmly grateful to my brother and my best friend, Majid, for being a fantastic replacement at home and also for his infinite supports. Buddy! I’m so honored to have you in my life and warmly wishing you all the best in life.

The list of friends and colleagues who helped me in one way or another is just difficult to compile but I’m grateful to former CGG students and personnel including but not limited to Shafagh Al Nadaf, Veronica Murtagh, Vidushi Patel, Hardip Patel, Dennis O’Meally, Ali Livernois, Tariq Ezaz, Barbara Harriss, Ke-Jun Wei, Jason Limnios, and my Canberra friends Hooman Salari, Khashaiar Kamali, Sean McKibbin and many more whom always will be my friends.

The ARC Centre of Excellence for Kangaroo Genomics and The Australian National University are also thanked for providing the scholarship and research resources.
ABSTRACT

As a representative of Australian marsupials, the recently sequenced genome of a model kangaroo, the tammar wallaby \textit{(Macropus eugenii)} provides unique opportunities to understand the organization and evolution of the genome in marsupials, and in Class Mammalia in general. Comparisons with the fully sequenced genome of the Brazilian short-tailed opossum \textit{Monodelphis domestica} allow me to compare the genomes of American and Australian marsupials, which have been evolving separately in different environments for the last 70 million years. The general aims of this thesis were to examine the extent to which part of the genome has been conserved in marsupials and in therian mammals, as well as to explore the organization and evolution of the largest gene family in mammals, whose members code for olfactory receptors.

As part of the KanGo’s (ARC Centre of Excellence for Kangaroo Genomics) task of establishing a map of the tammar genome, I undertook the comparative mapping of the long arm of chromosome 6 in the tammar wallaby. Firstly I identified segments conserved between opossum and human that I expected from chromosome painting to lie in this region, and isolated large insert clones from a tammar BAC library that contained conserved genes that defined these regions. Then I established their locations and order on the long arm of chromosome 6 in tammar wallaby. I found that there are only few rearrangements between tammar wallaby and the opossum in this part of the genome. However, the genomic parts orthologous to tammar wallaby 6q reside on several chromosomes in human, dog, and chicken, suggesting that the fusion occurred in the marsupial ancestors and remained conserved during marsupial evolution.

I then developed a strategy to explore the olfactory receptor gene (ORG) family in the tammar wallaby. Sequences corresponding to ORGs were extracted from the first assembly of the tammar wallaby genome and sequences classified into families and subfamilies. BACs bearing conserved mammalian ORG clusters were isolated and physically mapped in tammar wallaby. Comparison with the opossum OR repertoire revealed that these two distantly related marsupials share a very similar ORG superfamily. Conserved features include the total numbers of genes, families, and
subfamilies, gene distribution across the families and subfamilies, patterns of expansions and contractions in families and subfamilies and genomic location of major ORG clusters.

I then examined in detail the genomic organization of a highly conserved ORG cluster that lies near the MHC locus in several mammals. I made a BAC contig over the entire chromosome region. I found that this cluster is conserved in tammar wallaby and carries almost the same genes as in the opossum. Preliminary analysis of platypus ORGs dates the origin of this cluster back to the common ancestor of therian and monotreme mammals more than 166 million years ago, and provides examples of both conservation and adaptation of some genes in this cluster.

My general conclusion is that the two distantly related marsupial species have retained very similar genomes since their divergence 70 million years ago. This conservation is reflected both at the level of genome arrangement, and at the organization and evolution of gene families. This conservation is in marked contrast to the variability observed between eutherian groups, both in gross gene arrangement and in the constitution of the ORG family, suggesting that marsupial genomes have been evolving more slowly than other mammals, possibly due to some unique features of their physiology and way of life.
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>GABBR1</td>
<td>Gamma-Aminobutyric Acid (GABA) B Receptor, 1</td>
</tr>
<tr>
<td>HEXDC</td>
<td>Hexosaminidase (glycosyl hydrolase family 20, catalytic domain) Containing</td>
</tr>
<tr>
<td>hsp70</td>
<td>Heat Shock Protein 70</td>
</tr>
<tr>
<td>MOG</td>
<td>Myelin Oligodendrocyte Glycoprotein</td>
</tr>
<tr>
<td>NARF</td>
<td>Nuclear Prelamin A Recognition Factor</td>
</tr>
<tr>
<td>P2RX3</td>
<td>Purinergic Receptor P2X, Ligand-Gated Ion Channel, 3</td>
</tr>
<tr>
<td>PRNP</td>
<td>Prion Protein</td>
</tr>
<tr>
<td>PTPRJ</td>
<td>Protein Tyrosine Phosphatase, Receptor Type, J</td>
</tr>
<tr>
<td>RBMX</td>
<td>RNA Binding Motif Protein, X-linked</td>
</tr>
<tr>
<td>RFX</td>
<td>Regulatory Factor X</td>
</tr>
<tr>
<td>SRZ</td>
<td>Seven-Pass Receptor Family Z</td>
</tr>
<tr>
<td>UBD</td>
<td>Ubiquitin D</td>
</tr>
<tr>
<td>WDR45L</td>
<td>WD Repeat Domain 45 Like</td>
</tr>
<tr>
<td>ZFY</td>
<td>Zinc Finger Protein, Y-linked</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS

DECLARATION ............................................................................................................. ii

ACKNOWLEDGEMENT ................................................................................................. iii

ABSTRACT ..................................................................................................................... iv

LIST OF ABBREVIATIONS .............................................................................................. vi

TABLE OF CONTENTS ................................................................................................. vii

Chapter 1. Introduction ................................................................................................. 1

Mammals and Animal Evolution ................................................................................. 2
  Definition of mammals ............................................................................................... 2
  Evolution of mammals ............................................................................................... 2
  Higher order classification of mammals .................................................................... 3
  The genome of placental mammals ............................................................................ 4
  The evolution of mammalian genomes ..................................................................... 4

Gene Families: One Source of Genome Diversity ......................................................... 6
  Definition of gene families ....................................................................................... 6
  Evolution of gene families ....................................................................................... 7
  Duplication in the evolution of gene families ............................................................ 9
  Adaptive evolution in the evolution of gene families ................................................. 10
  Olfactory receptors: a dynamic evolution ..................................................................12

Marsupials: The Other Mammals ................................................................................. 12
  Definition of marsupials ......................................................................................... 12
  Evolution of marsupials ........................................................................................... 13
  Classification of marsupials ..................................................................................... 14
  The marsupial genome ............................................................................................ 14
  The value of marsupials in genomic studies ............................................................. 15
  Our model marsupials ............................................................................................. 16

The Thesis .................................................................................................................... 18

CHAPTER 2: MATERIALS AND METHODS ............................................................... 20

Agarose Gel Electrophoresis ....................................................................................... 20

Ethanol Precipitation .................................................................................................. 20

Culture Media ............................................................................................................... 21
  Liquid culture .......................................................................................................... 21
  Agar plates .............................................................................................................. 21
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agar stabs</strong></td>
<td>21</td>
</tr>
<tr>
<td><strong>Bacterial Artificial Chromosome (BAC) Clone Analysis</strong></td>
<td>21</td>
</tr>
<tr>
<td>BAC DNA extraction</td>
<td>21</td>
</tr>
<tr>
<td>BAC library screening</td>
<td>22</td>
</tr>
<tr>
<td>Selection of genes for mapping</td>
<td>22</td>
</tr>
<tr>
<td>Design and preparation of overgo probes</td>
<td>22</td>
</tr>
<tr>
<td>Filter prehybridization</td>
<td>23</td>
</tr>
<tr>
<td>Labelling overgo probes</td>
<td>23</td>
</tr>
<tr>
<td>Washing BAC filters</td>
<td>24</td>
</tr>
<tr>
<td>Developing X-ray films</td>
<td>24</td>
</tr>
<tr>
<td>BAC clone identification</td>
<td>24</td>
</tr>
<tr>
<td><strong>Dot-Blot Analysis</strong></td>
<td>25</td>
</tr>
<tr>
<td><strong>BAC End Sequencing</strong></td>
<td>25</td>
</tr>
<tr>
<td><strong>BAC DNA Fingerprinting</strong></td>
<td>26</td>
</tr>
<tr>
<td><strong>Fibroblast Cell Culture</strong></td>
<td>26</td>
</tr>
<tr>
<td>Thawing cells</td>
<td>26</td>
</tr>
<tr>
<td>Splitting cells</td>
<td>27</td>
</tr>
<tr>
<td>Changing the medium</td>
<td>27</td>
</tr>
<tr>
<td><strong>Metaphase Chromosome Preparation</strong></td>
<td>27</td>
</tr>
<tr>
<td>Dropping cells on slides</td>
<td>28</td>
</tr>
<tr>
<td><strong>Fluorescent In Situ Hybridization</strong></td>
<td>28</td>
</tr>
<tr>
<td>Nick translation labelling of BAC DNA</td>
<td>28</td>
</tr>
<tr>
<td>BAC DNA precipitation</td>
<td>29</td>
</tr>
<tr>
<td>Hybridizing BAC DNA to slides</td>
<td>29</td>
</tr>
<tr>
<td>Washing slides</td>
<td>30</td>
</tr>
<tr>
<td>Microscopy examination</td>
<td>30</td>
</tr>
<tr>
<td>Cot=1 DNA preparation</td>
<td>30</td>
</tr>
<tr>
<td><strong>Bioinformatics</strong></td>
<td>31</td>
</tr>
<tr>
<td>Sequence alignments</td>
<td>31</td>
</tr>
<tr>
<td>Phylogenetic tree analysis</td>
<td>32</td>
</tr>
<tr>
<td><strong>CHAPTER 3: THE PHYSICAL MAP OF TAMMAR WALLABY 6q</strong></td>
<td>33</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
<td>34</td>
</tr>
<tr>
<td><strong>Results and Discussion</strong></td>
<td>38</td>
</tr>
<tr>
<td><strong>Conclusion</strong></td>
<td>43</td>
</tr>
<tr>
<td><strong>CHAPTER 4: THE OLFACTORY RECEPTOR GENE SUPERFAMILY OF THE</strong></td>
<td>45</td>
</tr>
<tr>
<td><strong>KANGAROO MACROPS EUGENII (THE TAMMAR WALLABY)</strong></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5: THE MARSUPIAL MHC-LINKED OLFACTORY RECEPTOR GENE CLUSTER; CLUES TO THE EVOLUTION AND FUNCTION

Abstract ........................................................................................................... 78
Introduction ....................................................................................................... 79
Materials and Methods ........................................................................................ 81
  Probes .............................................................................................................. 81
  BAC end sequencing ....................................................................................... 81
  BAC library screening and Fluorescence in situ hybridization (FISH) .......... 83
  BAC DNA fingerprinting .................................................................................. 83
  Phylogenetic analysis ....................................................................................... 84
Results .................................................................................................................. 85
  MHC-linked ORG cluster in tammar wallaby .................................................. 85
  BAC DNA fingerprinting and contig assembly ............................................... 86
  MHC-linked ORG cluster in platypus .............................................................. 88
  Genomic structure of MHC-linked ORG clusters in different mammals .... 88
  Evolution of the subfamilies of the MHC-linked ORG cluster ...................... 90
Discussion .............................................................................................................. 95
  The MHC-linked ORG cluster of tammar wallaby ......................................... 95

ix
Functional relationship of ORG cluster and MHC loci in mammals ............................................. 96
Conclusion ............................................................................................................................................. 97
Additional Files.................................................................................................................................. 99

CHAPTER 6: DISCUSSION .................................................................................................................. 101
Genome stability at the chromosomal level .................................................................................... 102
Genome change in the olfactory receptor gene family ............................................................. 107
Genome changes in the MHC-linked ORG cluster ................................................................. 109
Conclusions ....................................................................................................................................... 111
Future directions ............................................................................................................................. 115

REFERENCES ...................................................................................................................................... 117

APPENDIX 1: THE OLFACTORY RECEPTOR GENE FAMILY OF MARSUPIALS
............................................................................................................................................................ 135

APPENDIX 2: PRESENTATIONS ........................................................................................................ 158
Oral Presentations: ............................................................................................................................ 158
Poster Presentations: ......................................................................................................................... 158
Chapter 1. Introduction

During 3.7 billion years of biological life on earth, organisms have evolved different forms, shapes and sizes in response to different environments; an alga on a rock in the sea, a bird flying high in the sky or a land mammal grazing plants. Sizes of extant organisms range from a few nanometres to 30 metres, from a few nanometres to hundreds of tonnes. Based on a most commonly quoted estimate there are 30-50 million species of biological organisms currently living on Earth (Erwin 1997). This, although an overestimate and controversial, is still a very small fraction (1-5%) of all the species that ever evolved, lived and became extinct during the last 3.7 billion years. This diversity is made possible by changes in life’s backbone, DNA or RNA, selected to adapt to an environment shaped by geological and climatic changes.

Among different forms of vertebrate life, marsupials evolved to adapt to their environment and succeed in survival and reproduction. Since the first encounter with these amazing mammals, biologists have been interested in genome changes that underlie their unique yet successful lifestyle. Comparing the genomes of marsupials with those of their sister group, eutherians (“placental” mammals), especially humans and rodents, is a particularly fruitful way of investigating the differences in their biochemistry, physiology and reproduction.

At a fundamental level, this task is carried out by comparative genomics. Comparing the genomes of mammals with each other and the genomes of other classes of vertebrates at the level of gene content and arrangement in conserved blocks has revealed the evolutionary process of mammal speciation and adaptations, with more yet to be discovered. In the view of comparative genomics, a species is best understood when its genomic connections (similarities and differences) with close and distant relatives are examined. Thus marsupials are usually compared with eutherian mammals from one side of the phylogeny, and fish and birds from the other side. Genetic and genomic comparisons complement a wide range of biological, anatomical, physiological, and behavioral studies.

Genetic and genomic studies can be conducted at different levels of resolution. At the lowest level are genomic features such as packaging of genetic material into
chromosomes, and the arrangements of the genes on these chromosomes in different species, which represents the way that a genome (hence a new species) has diverged from an ancestral genome. At a higher level of resolution are analyses of gene families and their evolution. Since most of the protein-coding genes of a typical metazoan genome can be put into gene families, understanding how they are formed and how they are evolving will help to understand the strategies adopted by different organism for adapting to its constantly changing surroundings, and how these adaptations are reflected in its genome. At the highest level of resolution is the DNA sequence of individual genes.

This thesis describes my attempts to understand the evolution of gene arrangements and gene families, using mapping and sequencing data from a model marsupial. Any unique or common feature of their genome is only definable when it is compared to characteristics of the placental mammal models, principally humans and mice, which are widely used as the reference model in comparative studies. To provide a background for comparisons, I review here the relationship and genome characteristics of the groups of mammals.

**Mammals and Animal Evolution**

**Definition of mammals**

The term mammal is used to name a class of vertebrates whose females possess mammary glands and nurse their young with milk. These are the characteristics used to define and distinguish mammals from other vertebrates (Wilson & Reeder 2005). Other features unique to mammals that are found in males as well as females are the presence of either fur or hair, and of three bones in the middle ear. All mammals are warm-blooded, air-breathing animals, and their brains feature a neocortex for processing higher functions (sensory perception, and language in humans) (McKenna & Bell 1997).

**Evolution of mammals**

The immediate ancestors of mammals were mammal-like reptiles (synapsids) that lived from the late Carboniferous period that extended from 360 million years ago
(MYA) to the very early Triassic period (250 MYA) (Ahlberg & Milner 1994). The first mammal, as we know them today, evolved from the synapsids around 200 MYA during the late Triassic period (Kielan-Jaworowska 2007). For some 140 million years mammals and dinosaurs coexisted; during this period most of the mammals were small, nocturnal and insect-eating forest-dwellers. It has long been speculated that the extinction of dinosaurs in the early Paleocene (65 MYA) provided a big opportunity for mammalian species to thrive and diversify, and large and even giant mammals appeared later and occupied most of the vacant ecological niches of land, and even some in water and air (Wible et al. 2007). This widely accepted theory was challenged recently when it was shown that most of the current day superorder and order of mammals diverged around 89 MYA, long before the mass extinction of dinosaurs, but their diversification remained slow for a long time (Bininda-Emonds et al. 2007). The sudden expansion was attributed to those mammalian orders that are mostly extinct now.

**Higher order classification of mammals**

Mammals range in size, shape, habitat, and diet, so that every corner of the planet and there are at least a few species in every ecological environment. This makes mammals one of the most successful forms of life that ever appeared on earth.

The widely accepted vertebrate classification system groups the 5400 species of mammals into 1200 genera, 153 families, and 29 orders (Vaughan et al. 2000) based on DNA analysis. Mammals are divided into two Subclasses, Prototheria (monotremes) and Theria (live-bearing mammals). Therian mammals include infraclasses Metatheria (marsupials) and Eutheria (placental mammals). There is considerable debate about when these three groups diverged and a wide range of 65 to 180 MYA has been suggested by different studies. The major discrepancies are rooted from different methods of estimation. While the archaeological methods, by relying on the age of the first appeared fossils, date the divergence time around the Cretaceous-Tertiary boundary at around 55-65 MYA (Alroy 1999, Benton 1999, Archibald et al. 2001, Wible et al. 2007), the molecular studies push this time further into the Cretaceous and assumes a divergence time of 100-180 MYA (Kumar & Hedges 1998, Waddell et al. 1999, Cao et al. 2000, Waddell et al. 2001, Springer et al. 2003). The most comprehensive study, however by including more than 4500
mammalian species in a meta-analysis concludes that monotremes (the egg-laying mammals such as platypus) diverged from therians 166 MYA and marsupials diverged from eutherians 147 MYA (Bininda-Emonds et al. 2007). Here I will use these dates since the platypus genome paper adopted them.

The genome of placental mammals

In every living organism, a single copy of DNA bearing the entirety of the heredity information is called its genome. In mammals, as in most other animals, the entire genomes (except for mitochondrial DNA) are assembled into chromosomes in the cell nucleus.

The size of every genome is very consistent across eutherian lineages, at approximately 3 billion base pairs (Gregory 2005). However, the number, size and shape of the chromosomes (karyotype) varies greatly, from three enormously large chromosomes in Indian muntjac (Muntiacus muntjak) (Wurster & Benirschke 1970) to 42 small ones in black rhinoceros (Diceros bicornis). The karyotype may be very consistent, or very variable in different mammalian orders and families. For example, in cats, species that have been independently evolving for 30 million years still share an obviously similar karyotype, whereas the sister orders, such as dogs and bears, have very diverse karyotypes, and even closely related rodents may have quite different karyotypes (for a review see O'Brien et al. 1988).

These karyotype variations imply different frequencies of genome rearrangements in different lineages. Overall, however, mammalian genomes are conserved (Ferguson-Smith & Trifonov 2007). DNA and gene count, gene blocks and gene orders within these syntenic blocks represent three levels of genome conservation in mammals.

The evolution of mammalian genomes

As more mammalian genomes are being sequenced and annotated, it becomes increasingly evident that all mammalian species have approximately the same number and virtually the same sets of protein-coding genes. This feature is so conserved that the annotated metatherian opossum (Monodelphis domestica) genome lacks only 1% of the coding sequences that are orthologous with the eutherians human, mouse, and dog (Mikkelsen et al. 2007). More remarkably, the conserved gene count in other
vertebrates and even invertebrates is not much lower than in mammals. For instance, it was established a long time ago that the nematode *C. elegans* has roughly the same number of genes (20,100) (Claverie 2001) as the human genome (20,500) (Clamp et al. 2007).

Mammalian chromosomes contain large conserved blocks of DNA, containing the same genes, often in the same order. Genes that occupy the same chromosome are known as syntenic (literally “same thread”). Syntenic blocks with orthologues in different species are said to share conserved synteny, and blocks in which gene order is conserved, and even base sequence can be aligned, are called homologous regions. Orthology is defined as identity by descent between species, and paralogy as homology between different regions within the same species (Comparative Genome Organization Workshop; Graves J.A.M corresponding author 1996).

Quite large homologous regions may be shared between different species, even very distantly related ones (O'Brien et al. 1999). Syntenic blocks can vary in size from a few genes (≈1 Mbp) to a whole chromosome or chromosome arm. The differences in the chromosome number and morphology between different mammals is due to how these segments are assembled together to make up the karyotypes (Murphy et al. 2005). The blocks can fuse together in one species or split onto two or more chromosomes in another species, causing decrease or increase in the number of chromosomes (Murphy et al. 2005).

The first evidence for conservation of chromosome gene content between mammalian species was established for the X chromosome more than 40 years ago. Ohno and colleagues (Ohno et al. 1964) proposed, on the basis of X linkage of isozyme variants, that all mammalian X chromosomes share a similar gene content. Extensive G-banding, gene mapping and cross-species chromosome painting (ZOO-FISH) studies, starting in the early 1990s, confirmed that chromosome segments are conserved on the mammalian X chromosomes.

Knowledge of considerable conservation was extended to autosomes by the development of methods to map genes in mammals, first using somatic cell genetics, and more recently *in situ* hybridization (Murphy et al. 2001, Richard et al. 2003, Wienberg 2004). The extent of genome rearrangements is different in different
mammalian orders. For example, primate chromosomes show a low level of rearrangement even between human and baboon species, that diverged 30 million ago (Dutrillaux et al. 1978), whereas rodent genomes, with a much shorter divergence time, have experienced more rearrangements (Stanyon et al. 1999). It is calculated that human genome has suffered fewer than two rearrangements every 10 million years, while the mouse genome has had one rearrangement every million years (Ehrlich et al. 1997). However, there are exceptions like the highly rearranged karyotype of the gibbon (Roberto et al. 2007).

Translocations and inversions are the main causes of reshuffling the syntenic blocks (Ferguson-Smith & Trifonov 2007). Centric fusions (Robertsonian translocations) are the most common form of rearrangements, occurring in one in 1,000 human babies (Therman et al. 1989). Other forms of rearrangement including duplications, inversions, and tandem fusions are also possible but they are less common than Robertsonian translocations, and are more deleterious in heterozygous state (Griffiths et al. 1999). These rearrangements are not randomly distributed throughout the mammalian genome, but may occur independently at conserved locations between syntenic blocks called evolutionary breakpoints or “hot spots” (Ruiz-Herrera et al. 2006).

It is suggested that the production of acrocentric chromosomes is more favored during karyotype evolution in some lineages (Qumsiyeh 1994). Based on the minimum interaction hypothesis, an increase in the number of acrocentric chromosomes minimizes the risk of deleterious rearrangements, as chromosomes will have a lower chance of interaction (Imai et al. 1986). A higher number of chromosomes leads to a higher rate of re-assortment, which increases variability within the species, while decreased recombination allows mutations to be fixed, and hence seeds speciation events (Schubert 2007).

**Gene Families: One Source of Genome Diversity**

**Definition of gene families**

A typical metazoan genome contains some 20,000 protein-coding genes (Ponting 2008). Based on their sequence homology most of them can be put into gene families
(Dayhoff 1976), although they may differ in size, sequence, function, and regulatory elements. By definition, a gene family (also referred to as a multigene family) is a group of genes that descended from one common ancestor, share the same functional domains and have >50% amino acid similarity. A superfamily (supergene family) is made up of related multigene families with sequence similarities of less than 50% (Fryxell 1996).

Gene families can be as small as only a few genes (e.g., 3 genes in both calmodulin and enolase gene families) or as big as a few thousand genes (olfactory receptor and histone gene families). The number of members may be strictly conserved between species, or vary extensively. For example, in all mammalian species, the RFX gene family of transcription factors has only seven genes, while the SRZ chemosensory gene family comprises 25-113 genes even in closely related species of Caenorhabditis (Thomas et al. 2005)

**Evolution of gene families**

For the last fifty years the evolution of mammalian gene families has received a great deal of attention and been the ground for much controversy. “Divergent evolution” was the first theory to be proposed for the evolution of gene families. Based on this model, the similarities between members of the same family are due to their divergence from the same ancestor. However after speciation, orthologous members (copies of the same gene that was duplicated before the speciation) of the gene families start gaining mutations independently that would gradually alter the final product of gene, thus producing interspecies diversity (Nei & Rooney 2005).

Although divergent evolution theories were capable of explaining the evolution of small and highly conserved gene families (with the same number of members), it failed to provide a satisfying explanation for the evolution of bigger gene families like the hemoglobin and myoglobin families (Ingram 1961). In these gene families the sequence similarities of members within species is greater than between species, which contradicts the divergent evolution theory and supports a model in which new paralogues (copies of the same gene that was duplicated after the speciation) are being created.
By studying the ribosomal RNA (rRNA) gene family of Xenopus, a new model of concerted evolution was proposed that could address the shortcomings of the divergent evolution theory (Brown et al. 1972). The concerted evolution theory predicts that all members of a gene family are evolving together, and any mutation that occurs in one member after speciation, will spread to all other members of the family. This will keep the paralogous members in one species more similar than their orthologs in another species that diverged from the same ancestor. Unequal crossing over and gene conversion are two mechanisms by which these mutations are traded between family members (Jeffreys 1979). The composition of most gene families of tandemly arrayed similar genes makes the homogenization of gene families by these two mechanisms more likely. Unequal crossing over occurs randomly between the members of the same gene family and homogenizes the sequences by changing the size of gene family (Smith 1976). Although the details of the molecular mechanism of gene conversion have only recently been described (Chen et al. 2007), its role in homogenizing gene families was established few decades ago (Nagylaki 1984).

Concerted evolution was later proposed to be a universal biological phenomenon that exists in all forms of life from bacteria to mammalian species (Liao 1999). By the late 1980’s, in addition to the rRNA genes, several other gene families were shown to follow the concerted mode of evolution, including small nuclear RNA (snRNA) genes (Pavelitz et al. 1995), the histone gene family (Coen et al. 1982), heat shock proteins such as hsp70 (Leigh Brown & Ish-Horowicz 1981), polyubiquitin genes (Nenoi et al. 1998) and many others (for a review see Liao 1999).

Nevertheless this model was shown to be not inclusive enough to accommodate every gene family. The major histocompatibility complex (MHC) genes were the first gene family to show a deviation from concerted evolution (reviewed in Nei & Rooney 2005). MHC molecules are involved in triggering the immune response by binding to self and non-self peptides and presenting them to T lymphocytes (Klein & Horejsi 1997). The high degree of polymorphism in classical class I MHC molecules (Ia) could not be simply explained by unequal crossing over events. On the other hand, polymorphic Ia MHC genes that are located in the same region are less polymorphic than Ib MHC genes in the human genome. This close proximity rules out gene conversion as the mechanism to produce polymorphism only for Ia genes (Martinsohn et al. 2000).
“Birth-and-Death” evolution was proposed to be the mechanism by which MHC molecules evolved (Hughes & Nei 1989, Nei et al. 1997). This model of evolution is a combination of the two models of gene family evolution described above. As for the concerted evolution theory, this model predicts that a gene family in two species derived from the same gene family in their common ancestor, and some genes are more similar within species than between species. However in birth-and-death model, every gene could experience a quite different fate after speciation. Some genes could duplicate once or several times in one species (birth), or the same gene could be lost (death) in other species (Nei & Rooney 2005). This ongoing process of duplication and gene loss leads to a huge diversity in family size as well as sequence homology. Extensive phylogenetic analyses have indeed confirmed that MHC genes in some mammalian orders or families are not true orthologues, and although some genes could be shared by some orders, there are some genes that are specific to only one or few orders of mammals (Friedman & Hughes 2004).

Most of the gene families that are involved with immunity and chemosensory perception show a birth-and-death pattern of evolution. The MHC gene family is the best-studied gene family that is consistent with this model. Immunoglobulin gene families, the olfactory receptor superfamily in vertebrates, and mammalian vomeronasal receptors are other examples of gene families that follow a birth-and-death pattern of evolution.

**Duplication in the evolution of gene families**

Duplication of ancestral gene(s) provides the start point for evolution of a gene family, no matter by which mechanism the evolution occurs. There are several forms of gene duplication, including whole genome duplication (WGD), tandem duplication and segmental duplication. In WGD, the entire genome content of an organism is duplicated, resulting in two copies of each gene in the duplicated genome. Occurrence of WGD once or more leads to polyploidy. Tandem duplication describes the case in which a gene is duplicated and the new duplicate resides next to the seeder gene in a tandem array. Segmental duplications are duplicated blocks of genomic DNA typically ranging in size from 1–200kb that may be located in other regions or on other chromosomes.
Duplications that lead to the genesis of gene families involve only a gene or a segment of the genome, and do not result in drastic change in genome size of the organism. Gene duplication may occur by any of three mechanisms; unequal crossing over, retrotransposition, and segmental (chromosomal or genome) duplication (Ohno et al. 1968). Unequal crossing over is the most prevalent cause by which duplicate gene(s) are generated close to the original copy and an array of similar genes is produced as a result (tandem duplication) (Smith 1976). This arrangement is common to most of the mammalian gene families that are composed of cluster(s) of duplicated genes (e.g. alpha globins). Duplication by unequal crossing over could involve only a part of the gene, a complete gene, or a bigger segment of the DNA (segmental duplication) (Samonte & Eichler 2002).

Retrotransposition occurs when a messenger RNA (mRNA) is reverse transcribed into complementary DNA (cDNA) and then re-inserted into the genome (Long 2001). In contrast to tandem gene duplication, the newly retrotransposed gene lacks intron and regulatory sequences (Long et al. 2003). In theory, segmental and tandem duplications can occur at any genomic location but genomic regions rich in transposable elements (TEs) undergo relatively high frequencies of unequal crossing over which in turn leads to a high frequency of gene duplication (Hancock 2005).

**Adaptive evolution in the evolution of gene families**

Regardless of the mechanisms by which they were generated, duplicate genes play an important role in the evolution of genomes and organisms. It is estimated that almost 10-20% of a typical metazoan genome is made up of duplicate genes (Moore & Purugganan 2003). Because gene duplications occur randomly, it is expected that, in different species, the duplication rate and the type of duplicates will be different. These differences may provide a source of phenotypic differences between even closely related species. For example, although human and chimpanzee share on average 98% similarity in their coding DNA, there are 1,418 (6.5% of all genes) in human that are the products of differential duplications and have no true orthologs in chimp genome (Demuth et al. 2006). These are almost entirely the genes of big gene families like G protein-coupled and immune gene families in which new duplicates could gain a slightly different function, or expression pattern, from that of the original or ‘seeder’ genes, and hence contribute toward producing a different phenotype.
There are four possible fates of duplicate genes; retaining the same function as the original gene and contributing toward producing twice as many transcripts (conservation of function), independently acquiring mutations and producing products with slightly different functions (subfunctionalization), turning into pseudogene by collecting deleterious mutations (pseudogenization), or gaining a totally new function compare to the original gene (neofunctionalization).

Genes that undergo subfunctionalization may code for products with differences in function that are the basis of selection for the acquisition of new functions by existing genes. Pseudogenization can occur when one of the duplicated copies is silenced and starts accumulating deleterious mutations that would turn it into an untranslateable pseudogene (Ohno 1970). In gene families with ‘birth-and-death’ evolution, pseudogenization is very common in duplicate genes. For instance, in different mammalian species, 20-60% of the olfactory receptor genes (that are products of repeated tandem duplications) are pseudogenes (Niimura & Nei 2005a).

Neofunctionalization can occur when a small change in one of the duplicated genes, such as a few base changes that lead to amino acid substitutions in the protein, or changes in the regulatory sequences, results in a new function that is beneficial for the better survival of the organism (Walsh 1995). The red and green opsins genes are an example of duplicate genes that have evolved a new beneficial function (Asenjo et al. 1994). In some cases when a larger amount of a gene product is beneficial, gene duplication provides extra copy of the same gene for more transcription. A well-known example where multiple gene copies are transcribes are the ribosomal RNA (rRNA) genes in vertebrates.

In general, gene duplications provide evolution with the raw material for creating new functions (thus new phenotypes) that may help the organism to adapt to the ever-changing environment. This adaptive evolution is more obvious in families with ‘birth-and-death’ evolution. A large and varied MHC repertoire will produce more diverse and novel MHC peptides which is advantageous to host organism survival. In the same way, olfactory perception of the environment is also affected by the number and variety of receptors possessed by each species. A dynamic and varied olfactory gene family will ensure the maximum utilization of environmental resources by organism in its constant battle for survival.
Olfactory receptors: a dynamic evolution

Vertebrates have evolved a family of G protein-coupled receptors for acquiring information about the chemical content of their environment (Freitag et al. 1998). These olfactory receptors (ORs) provide a good example of birth-and-death evolution, which is evident in the diversity of the number of genes and duplication patterns between lineages (Nei et al. 2008). Up to a 10-fold difference in the number of olfactory receptor genes (ORGs) between fish and mammalian species are shown to be the product of few rounds of duplication of the entire ORG family. However, there is still a remarkable difference in the number of genes between different mammalian lineages. Even for species that diverged only few million years ago and have almost the same number of ORGs, the number of unique ORGs is very high (Niimura & Nei 2005b). Many ORGs that are functional in one species are pseudogenized (or lost in some cases) in another, and some of them have experienced several duplications only in one species. These species-specific expansions (birth) and contractions (death) in the ORG family size and content provide a bigger opportunity for animal to detect a wider range and more important odorant signals in its environment. A comprehensive introduction to mammalian ORGs and their evolution is given in chapter 4.

Marsupials: The Other Mammals

As gene families include most of the protein-coding genes in mammalian genomes, understanding their evolution provides a valuable background to understanding the evolution of species and their adaptation for a successful lifestyle. Gene families (especially those with many members) can serve as small-scale models of genome evolution that would reflect the overall mechanisms and forces by which the genomes have evolved. Marsupial gene families, by this definition, can be studied in order to unravel the evolutionary history of their genomes.

Definition of marsupials

DNA analyses complement the long-studied anatomical and physiological characteristics that classify mammals into three groups (Kielan-Jaworowska 2007). There are two Subclasses; Prototherian mammals (order Monotremata, containing the
platypus and four echidna species) diverged from the Theria (marsupial and eutherian mammals) around 166 MYA (Bininda-Emonds et al. 2007). Theria are divided into two Infraclasses; Metatherians (marsupials such as opossum and kangaroos) diverged from Eutheria (often known as the ‘placental’ mammals, including humans and rodents) around 147 MYA (Phillips & Penny 2003, Woodburne et al. 2003, van Rheede et al. 2006). Although there is a debate about the divergence time of these three mammalian infraclasses, in this thesis the aforementioned dates will be used. The eutherians are characterized by giving birth to young that are nourished in utero by a well-developed placenta. In contrast, the term marsupial was originally used to describe the group of mammals that gave birth to a very immature young that developed further externally to the mother’s womb in her pouch (marsupium). Although it is known today that many marsupials, such as opossum, do not bear a pouch, the name is still universally used.

**Evolution of marsupials**

Two theories have put forward about the origin and travel routes of marsupials. In 1924, Harrison proposed that marsupials originated in modern-day China and reached Australia through Southeast Asia (Harrison 1924). Harrison’s theory was supported by the discovery of the oldest marsupial fossil in China (125 million years old) (Luo et al. 2003) and the observation that different groups of Australian marsupials are more closely related to each other phylogenetically than they are to the American marsupials (Spotorno et al. 1997). However, in 1977 Clemens (Clemens 1977) proposed that marsupials diverged from eutherian mammals were originally scattered throughout the North America, later traveled to South America. One branch reached Australia via Antarctica before the Gondwana land split, and was then isolated when South America separated from Gondwana. A comprehensive comparison of more than 200,000 retroposon-containing loci has recently supported this theory (Nilsson et al. 2010). Four Australian Marsupial orders were shown to share a same origin with the only living member of the order Microbiotheria (*Dromiciops gliroides*) that lives in South America. This study also places the South American opossums (Didelphimorphia) as the first branch of the marsupial tree, supporting a clear divergence between South American (Superorder Ameridelphia) and Australasian (Australidelphia) marsupials. The divergence between these two superorders was estimated to be 55 MYA from the time at which South America separated from
Gondwanaland, but has been more reliably estimated at 69 MYA from molecular data (Nilsson et al. 2004).

### Classification of marsupials

About 334 species of marsupials have been classified into 10 orders (Wilson & Reeder 2005). Two orders, containing 96 species, are confined to South America while a single order containing only one species is found in North America. The remaining 7 orders containing 237 species are found in Australia and New Guinea. Morphological and molecular data put the North American Marsupial as the sister group to all marsupials (Horovitz et al. 2009, Nilsson et al. 2010) It is recently shown that there has been only one migration of marsupials from South America to Australia via Antarctica. This sets the divergence time of Australian marsupials (like tammar wallaby) and American marsupials (like opossum) before the split of Australia from Gondwanaland that happened 55 MYA.

Australian marsupials range in size from small mouse-like members of the Dasyuridae family to large kangaroos. Different marsupial species are insectivorous, herbivorous or carnivorous and occupy different ecological niches. Kangaroos and wallabies belong to the family Macropodidae that is one of the most diverse marsupial families with 47 species. Macropods are herbivores and own the ability of moving on the two hind legs as well as on all four limbs. This enables them to travel long distances at high speed without great energy expenditure, which is crucial to their evolutionary success on a continent of low soil fertility and low, unpredictable average rainfall.

### The marsupial genome

Opossum has a genome roughly the same size as eutherian mammals (Mikkelsen et al. 2007), and the tammar wallaby perhaps a little smaller at 2.9 Gb (Renfree et al. 2011). The major difference between marsupial genomes and the well-characterized eutherian genomes is that marsupial genomes are arranged in only a small number of relatively large chromosomes (Samollow 1998, Graves & Westerman 2002a, Ferreri et al. 2004). Karyotype conservation among marsupials is striking and has been documented for many years (for a review see Rens et al. 2003). The range of
chromosome numbers vary from 2n=10 in coppery ringtail possum (*Pseudocheirus cupreus*) and female swamp wallaby (*Wallabia bicolor*), to 2n=32 in rufous bettong (*Aepyprymnus rufescens*). However, ~90% of the 211 metatherian species studied so far have a karyotype of 2n=14 or 2n=22. These observations led to the proposition of 2n=14 as the metatherian ancestral karyotype (Rofe & Hayman 1985). However, Svartman and colleague (Svartman & Vianna-Morgante 1999a) made a good case for a larger number of chromosomes (22) in the ancestral karyotype on the grounds that South American species with 2n=14 have interstitial telomere sequences that are clear signs of chromosome fusion. The sequence data that puts American opossums at the base of the phylogeny supports this view. The 2n=14 karyotype was clearly basal to the Australian radiation, and is shared by the South American groups from with the Australidelphia diverged.

The ancestral karyotype was made up of conserved segments that are easily identifiable in every marsupial genome by G-banding (Rofe & Hayman 1985), and subsequently 19 were identified by chromosome painting (Rens *et al.* 1999). Deviations from the ancestral karyotypes in marsupials can be explained by a small number of rearrangements (fusions and fissions) in the way that these conserved segments are assembled into chromosomes (Spotorno *et al.* 1997, De Leo *et al.* 1999).

**The value of marsupials in genomic studies**

Marsupials are interesting subjects for genomic studies. They have a distinctive reproduction strategy and have some basic anatomical and physiological differences from eutherians (Tyndale-Biscoe 2005). They occupy a “sweet spot” in the vertebrate phylogeny. Humans and birds (and fish) are too distantly related (more than 310 million years) for many comparisons because many sequences cannot be aligned. At the other extreme, sequences of different eutherians can be too similar to provide informative variation. Including the marsupial genomes in comparative genomic studies halves the long gap between the divergence of birds and humans (Koina *et al.* 2006), and allows us to detect significant conservation and differences at an intermediate level.
Many evolutionary questions about different mammalian characteristics and their origins (such as the origin of X inactivation, lactation and immunity genes) can be answered by comparing metatherians with eutherians. Marsupials can also be useful as biomedical models. The immature metatherian newborns have been used for studies of sexual differentiation, hormone-behavior interactions during reproduction, estrus cycling, ontogeny and evolution of placental structure and function, and the earliest stages of embryonic development (Samollow 2006).

**Our model marsupials**

Among marsupials, the Australian tammar wallaby and Brazilian gray short-tailed opossum are the most widely used model animals. Their ability to be bred in captivity, their small size and the relative ease of handling them has made them the subject of intensive genetic and genomic, reproduction and medical experiments since the 1970s (Tyndale-Biscoe & Renfree 1987, Tyndale-Biscoe 2005).

The opossum is a small member of the family Didelphidae, belonging to the superorder of American marsupials (Ameridelphia). The tammar wallaby is a small (approximately 8kg) member of the kangaroo family in the order Diprotodontia. The opossum was the first marsupial mammal to be subject to genome sequencing (Mikkelsen et al. 2007). This work, along with several companion papers, has contributed significantly to our understanding of genome evolution and adaptation in mammals. For example, the opossum was shown to have almost the same number of genes as human, and there were only eight opossum genes with no identifiable counterparts in the human genome (Samollow 2008). Therefore the phenotypic differences observed between even these distantly related mammals must be due to a source other than gene number. Changes in base sequence that are expressed as an altered amino acid sequence are the most obvious source of variation. Non-coding DNA also provides candidate for interspecies variation, as opossum lacks almost 20% of the non-coding regulatory elements conserved among eutherians.

Opossum autosomes have a somewhat lower G+C content compared to eutherian autosomes (37% compare to 41%) while the CpG content of opossum X is higher than autosomes, in contrast to other mammals. Interestingly the length of the female linkage map is only about half that of the male map in opossum, suggesting that
recombination is much less frequent in oogenesis than in spermatogenesis, leading to
a sexual dimorphism in recombination rate. This finding corroborates earlier findings
from linkage analysis in *Smynthopsis crassicaudata* (Bennett *et al.* 1989) and suggests
that low female recombination might be a marsupial trait. In eutherian mammals,
however, either the recombination rate is similar between the two sexes or is slightly
higher in female (Samollow 2008). In addition, the finding that opossum genome has
much less segmental duplication (1.7% of the total genome sequence) compared to
mouse (5.2%), and human (10.4%), might provide an explanation for the
conservation and stability of karyotypes in marsupials compared to eutherians
(Hayman *et al.* 1982).

Genomic comparisons between marsupials and other mammals (as well as other
vertebrates) have helped in finding new protein-coding genes (for example, the
human candidate mental retardation gene *RBMX*; Delbridge 2010), assessing the
credentials of candidate genes (e.g. the putative mammal sex determining gene *ZFY;
Sinclair *et al.* 1988), as well as identifying potential regulatory sequences (for
example those 5' to the prion protein gene, *PRNP*; Premzl *et al.* 2004). The
availability of opossum sequence has greatly advanced this work; for instance the
discovery of a new human alpha-like globin gene (Cooper *et al.* 2006). Comparison
of marsupial opossum sequences with those of other mammals and other vertebrates
has also helped in finding the evolutionary roots of some highly conserved
mammalian genes such as the genes involved in milk production and secretion
(Lemay *et al.* 2009), and production of inflammatory mediators (Reimer *et al.* 2010).

Sequencing a second marsupial would provide greater opportunities to explore all
these questions by enabling comparisons between marsupials as well as between
marsupials and eutherians. The tammar wallaby, which has been used for decades as
a representative species of Australian marsupials, was selected for whole genome
sequencing by a joint consortium of the Baylor College of Medicine in Houston,
USA, and the Australian Genome Research Facility (AGRF). The sequence of the
tammar wallaby has just been submitted for publication (Renfree *et al.* 2011), but the
assembly reported therein was not available when I completed my thesis work.

As part of collaboration with this consortium, the ARC Centre of Excellence for
Kangaroo Genomics (KanGO) was responsible for building a physical map of the
tammar wallaby genome, and comparing it to the opossum genome and the human genome. The identification of regions of conserved gene order between opossum and human genomes allowed a virtual map of the tammar genome to be constructed (Deakin 2010), in order to provide a scaffold for anchoring the sequenced contigs onto chromosomes.

The Thesis

The aim of this work was to examine the evolution of gene arrangements, and the evolution of a gene family in the Australian model marsupial, the tammar wallaby, as part of the effort to construct a map of the tammar genome, and to use it to explore the evolution of gene families. To achieve this goal, different levels of tammar wallaby genome structure and organization were examined, reported in chapters 3, 4, and 5 of this thesis, in order to build up a larger view of how gene rearrangements have contributed to the evolution of gene families and chromosomes.

To examine whether the gene order and content of the tammar wallaby chromosomes was well conserved as G-banding and chromosome painting had previously suggested, I contributed to the construction of a physical map of the tammar wallaby genome by determining the location and order of genes from opossum chromosome 7, using cytogenetic techniques, that opossum chromosome 7 genes mapped to the long arm of tammar wallaby chromosome 6q. In this way I was able to discover gene arrangements on the tammar wallaby chromosomes, and determine the location and order of gene blocks in the tammar wallaby known to be conserved between opossum and eutherians (Chapter 3).

In order to look more closely at how gene families are evolving in the tammar wallaby, the complete repertoire of the large olfactory receptor gene family was extracted from the first assembly of the sequence of the tammar wallaby genome. Classification and phylogenetic comparisons with opossum and other mammals were carried out to establish how this large gene family evolved in tammar wallaby, and determine if the changes that have occurred in this family have occurred within the conserved gene blocks identified between tammar wallaby and humans, or have occurred independently of the conserved gene blocks (Chapter 4).
I then examined in detail the genomic structure of a particularly interesting olfactory gene cluster linked to the MHC locus. This cluster has been well characterized in other mammalian species because of interest in the interaction of olfaction and MHC in mate choice. The gene order within the MHC-associated olfactory receptor gene cluster was characterized in the tammar wallaby, which provides a particularly interesting test case because the MHC Class I genes are dispersed over several chromosomes. This in-depth look at how this olfactory gene cluster was conserved showed that, although the entity of the cluster is conserved in tammar wallaby, there are a few minor rearrangements in its organization and location in this marsupial genome compared to the opossum (Chapter 5).
CHAPTER 2: MATERIALS AND METHODS

This chapter outlines procedures routinely used throughout this thesis. This includes general molecular biology methods, and bioinformatics. All commonly used buffers and solutions (such as SSC, EDTA, SDS, TAE, TBE, PBS, LB) were prepared according to established protocols (Sambrook & Russell 2001).

Agarose Gel Electrophoresis

Electrophoresis was used to separate different sized fragments of DNA on a 0.8-2% agarose (Gibco, BRL) gel, buffered with 0.5 x TBE (0.045M Tris-borate (ICN;BDH), 0.001M EDTA (Ajax), pH8) run at a maximum of 6 volts/cm. to visualize bands under UV light, 0.2μg/ml ethidium bromide (Sigma) was added to the agarose. Gels were photographed with GEL LOGIC 100 Imaging System (Kodak). For marker DNA, 100bp and 1kb DNA ladders (NEW ENGLAND BioLabs) were used. The 6x loading dye (0.25% bromophenol blue, 0.25% xylene cyanol FF, and 15% ficoll) was prepared at and stored at room temperature.

Ethanol Precipitation

Whenever needed, extracted DNA or PCR products were subjected to ethanol precipitation for removing the reaction components and concentrating the DNA. A 2.5x volume of 100% ethanol and 1/10 volume of 3M sodium acetate (pH5.2) were added to the DNA solution, and the mixture was incubated at -80°C for more than 20 minutes or at -20°C for more than 4 hours. The precipitated DNA was collected by centrifugation at 13,000g for 20 minutes and the pellet was washed with 70% ethanol. Pellets were re-suspended in a minimum volume of Milli-Q water to give a DNA concentration of desired.
Culture Media

Liquid culture

For DNA extraction a high volume of BAC clones was needed. Therefore, a single colony of each BAC clone was transferred to 50ml tubes containing 15ml of LB medium (10g/L tryptone, 5g/L yeast extract, 5g/L NaCl) supplemented with 20µg/ml chloramphenicol. Tubes were incubated overnight at 37 °C with shaking. The following day, two stocks (1ml) of the enriched clones were prepared in glycerol (15% final concentration) and stored at -80°C. The rest of culture was used for DNA extraction.

Agar plates

Single colonies of BAC clones were generated by streaking cultures on LB agar plates (1.5% agar (Amresco) in LB medium). Agar plates were streaked for single colonies with BAC clones and grown overnight at 37 °C.

Agar stabs

BAC clones were sent to British Columbia Cancer Agency (Vancouver, BC V5Z 4E6) for BAC fingerprinting by inoculating one single colony of the BAC into a 2ml screw-tap tube containing 1ml agar LB supplemented with 20µg/ml chloramphenicol. Tubes were put in the 37 °C incubator overnight to ensure that clones were viable and were sent off the next day.

Bacterial Artificial Chromosome (BAC) Clone Analysis

BAC DNA extraction

The Wizard® Plus SV Miniprep DNA Purification System (Promega, Madison, WI 53711 USA) was used to extract the DNA from BAC clones cultured overnight at 37°C in LB. The manufacturer’s protocol was followed except that it was scaled up to extract DNA from 15mls of overnight culture. The quality and quantity of the extracted DNA was assessed by agarose gel electrophoresis and kept at -20 °C until required.
BAC library screening

A male genomic tammar wallaby BAC (Eisenbach & Giojalas) library was obtained from Arizona Genomics Institute (Tucson, AZ 85721, USA). The library filters, on which the BAC library was arrayed, were screened for BAC clones containing specific gene sequences using radioactively labeled overlapping oligonucleotide (overgo) probes as described in the following sections.

Selection of genes for mapping

The first step toward localizing genes in another species is to choose appropriate genes from the reference genome and design overgo probes for those genes. Overgos are then used to isolate BAC clones containing those genes that are then localised onto the chromosomes of the unknown genome. The ultimate concern was that the designed probes would be as specific as possible, so there would not be any cross-hybridization with paralogous sequences.

In general, the genes that were initially selected had only one orthologue in the reference and human genomes. An exon of the gene that had the lowest protein homology to known protein domains was targeted for the synthesis of a probe. The whole DNA sequence of the exon was compared to the trace archives of the tammar wallaby database (Macropus eugenii – WGS) in NCBI using MegaBLAST. If unique sequences were found, when the sequences were compared back to the tammar wallaby database and the reference genome, an overgo was designed from the unique tammar wallaby sequence. Alternatively a new gene was selected for screening. Genes were targeted at approximately 20 Mbp intervals in the reference genome to clone and map homologues in the tammar wallaby genome.

Design and preparation of overgo probes

Overlapping oligonucleotide (overgo) probes consist of two oligonucleotides (24 bp) that complement each other over 8bp at the 3’ end. After annealing the oligonucleotides, single-stranded portions of the overgo are end-filled with Klenow DNA polymerase in the presence of $^{32}$P-tagged deoxynucleotide triphosphates, leading to a radioactive, double-stranded DNA fragment of 40 bp. Single copy
tammar wallaby trace sequences were submitted to Overgo Maker (http://bioinf.wehi.edu.au/cgi-bin/overgomaker) for the design of overgo probes.

Probes were checked for cross-hybridization by a BLAST against tammar wallaby trace archives. Oligonucleotide primers were synthesized at Geneworks (Hindmarsh, SA 5007, Australia) and rehydrated in Milli-Q water to a concentration of 100µM in the original vial. Working stocks were prepared from original vials by diluting 20µl of each probe in 80µl Milli-Q water (final concentration: 20µM).

**Filter prehybridization**

Library filters (3 at a time) were removed from their container and meshes were put between them for better exposure to the hybridization solution. The filters-mesh combination was rolled and put in a 100ml tube. 25ml of 60 ºC pre-warmed hybridization buffer (7% SDS, 0.5M sodium phosphate pH 7.2, 1mM EDTA) was added and tubes were incubated for one hour (one night in the case of new filters) at 60 ºC in a rotating incubator.

**Labelling overgo probes**

From the working stocks of overgos, 0.5µl of each forward and reverse primer along with 6µl of Milli-Q water were added to a 0.5ml PCR tubes and mixtures were heated in a PCR machine at 80 ºC for 5 minutes, and allowed to anneal for 10 minutes at 37 ºC before incubating on ice during preparation of the reaction. Bovine serum albumin (1mg/ml), oligo ligation buffer (Han et al. 2000), α-32P-dATP (50µCi), α-32P-dCTP (50µCi), and the Klenow fragment of DNA polymerase I (2 units) were combined with the ice-cold primer mixture. The reaction was allowed to proceed at room temperature for at least one hour to allow the incorporation of radioactively labeled nucleotides.

To removing unincorporated nucleotides, 40µl of STE buffer (supplied with the columns) was added to each tube and the reactions were centrifuged through Illustra ProbeQuant™ G-50 Micro Columns (GE Healthcare UK Ltd, Buckinghamshire HP7 9NA England) for 2 minutes at 3000g. The liquid collected through the column was transferred to a new 0.5ml tube and denatured for 10 minutes at 95 ºC in a PCR
machine. After denaturation, tubes were put on slushy ice for 2 minutes and then were added to the hybridization tubes along with 25ml of hybridization solution (1mM EDTA, 7% SDS, 0.5M sodium phosphate buffer, pH7.2). Tubes were incubated in a rotating incubator overnight at 60 °C.

**Washing BAC filters**

After one night of hybridization, tubes were removed from the incubator and hybridization solution was replaced with 100ml washing solution (1mM EDTA, 1% SDS, 40mM sodium phosphate buffer, pH7.2) pre-warmed to 60°C, and incubated in the rotating incubator for another 30 minutes. This process was repeated with the second washing solution (1.5x SSC, 0.1% SDS), and third washing solution (0.5x SSC, 1% SDS), for 20 minutes each. After the last wash, filters were removed from the tubes and were unrolled in a tray containing 500ml of 2x SSC solution. Each filter was taken out of the tray and patted dry, and wrapped in plastic wrap to prevent them from drying out. Filters were put in autoradiography cassettes with their face up and fixed in their position by tapes. An x-ray film was marked for orientation, and was placed on top of the filter and cassettes were left in -80°C freezers for 1-10 days.

**Developing X-ray films**

After exposure, the films developed in developer solution (ILFORD, South Yarra Australia) for 5 minutes, then fixed for 5 minutes in the fixer solution (ILFORD, South Yarra Australia).

**BAC clone dentification**

Applying the address grid provided by the BAC membrane producer identified positive BAC clones, which have produced signals on the film. Positive BACs were picked up from the stock kept in -80°C and proceed for DNA extraction and Dot-Blot analysis (Deakin *et al.* 2008) after enriching.
Dot-Blot Analysis

Dot-Blot analysis was used to sort out which of the pooled probes in the initial library screens hybridised specifically to each individual BAC clone, as up to 20 overgo probe pairs were used together in each hybridization tube, in the initial screens. Positive BAC clones were spotted onto multiple positively charged nylon transfer membranes (Amersham Hybond™ –N+; GE Healthcare, UK) incubated on LB plates at 37°C overnight. The following day membrane filters were removed from the plates and placed on Whatmann paper soaked in 10% SDS for 10 minutes to sharpen the final signals. Filters were transferred to Whatmann papers soaked in denaturing solution, neutralizing solution (Sambrook & Russell 2001), and 0.4M NaOH for 5, 10, and 20 minutes, respectively. NaOH fixes the DNA to filters. Filters were then transferred in a tray containing 6X SSC and cell debris were removed by wiping filters with a kimwipe paper. Filters were allowed to dry at room temperature and were used in hybridization with separate overgo probe pairs according to the hybridization protocol.

BAC End Sequencing

BAC DNA was extracted as described. For each BAC clone, 2-4 extractions were pooled together and subjected to standard ethanol precipitation. Sequencing reactions were set up containing 400ng purified BAC DNA, 4µl Big Dye Terminator (BDT) v3.1 (Applied Biosystems, Mulgrave, 3170 Australia), 2µl 5x BDT Buffer (400mM Tris, 10mM MgCl2), 5pmol primer, and Milli-Q water in a final volume of 10µl. Sequencing reactions were carried out in 0.2ml PCR tubes for the following PCR program: 95°C for 5 minutes, followed by 99 cycles of denaturation at 96°C for 10 seconds, annealing at 50°C for 10 seconds, and extension at 60°C for 4 minutes.

Following the completion of the PCR reactions, sequencing products were cleaned up as follows. Milli-Q water was added to each tube to make the volume up to 20µl (i.e. 10µl for 10µl reaction) and transferred to 1.5mL tubes for a modified ethanol precipitation. 4µl of 3M sodium acetate (pH4.8), and 50µl of ice-cold 100% ethanol were added to each tube. Tubes were inverted to mix and immediately centrifuged at full speed on the bench-top centrifuge for 30 minutes at 4°C. After the removal of the
supernatant, 200µl of ice cold 70% ethanol was added and mixed gently, before centrifuging at full speed for 15 minutes. The liquid supernatant was removed carefully and the DNA pellets were air dried for at least 5 minutes. Dried DNA was sent off to Australian Genome Research Facility (St Lucia QLD, Australia) for sequencing.

**BAC DNA Fingerprinting**

Restriction enzyme fingerprinting of BAC clones was carried out by Genome Science Centre (GSC) at the British Columbia Cancer Agency (Vancouver, Canada). For this purpose, agar stabs of BAC clones were sent to the agency. Briefly, this company extracted BAC clone DNA from overnight liquid cultures, and digested the DNA with two restriction endonuclease enzymes. The DNA fragments were analysed by gel electrophoresis and the overlapping DNA fragments were assembled into contigs using automated assembly pipeline tools. The results were viewable in Java-based Internet Contig Explorer (iCE v3.5), which was available for download from http://www.bcgsc.ca/platform/bioinfo/software/ice.

**Fibroblast Cell Culture**

Cell culture from fibroblasts was carried out according to standard protocols in order to generate rapidly dividing cells for the preparation of metaphase chromosome spreads. These chromosome spreads were dropped on to slides for use in fluorescent in situ hybridization experiments, which physically located BAC clones on chromosomes.

**Thawing cells**

The vial containing cells was removed from the -80°C freezer and thawed directly in a 37°C water bath. 5ml of pre-warmed 10% DMEM media (Invitrogen, Mulgrave, 3170, Australia) was added together with 2ml of the thawed cells into a 10ml tube and centrifuged at 1500 rpm for 5 minutes. The supernatant was carefully removed and the cell pellet was re-suspended in 5ml of fresh 10% DMEM. The cells in media
were transferred to a 25cm² culture flask. For 2 minutes, CO₂ was injected to the flask and the lid was tightened and flask was incubated overnight at 37°C.

**Splitting cells**

If the cells grew enough to cover almost 80% of the surface area of the flask, they needed to be split into more flasks. For splitting cells from a 25 cm² flask or 75 cm² flasks, firstly the medium was sucked out of the flask to be divided. PBS (2ml) was used to rinse out the foetal calf serum (FCS) and dead cells. This step is necessary as FCS stops trypsin from working. 1ml trypsin was put in the 25 cm² flask (2mls for 75 cm² flasks). Cells were given enough time to lift off from the flask’s wall and the flask was shaken on the bench to facilitate and speed up the process. In the meantime, new flasks were prepared and enough medium was out in them. Cells streaming off were checked by tilting the flask and checking under the microscope. Free cells were sucked from the original flask and divided between the new flasks by using a Pasteur pipette. Enough medium and CO₂ gas was added to the original tube and all flasks were put in the incubator overnight. The next day medium of all flasks was changed.

**Changing the medium**

All flasks were checked every 2-3 days and if the flasks were non-confluent, their media were changed. For changing the medium, the entire old medium was sucked off and 5ml medium at room temperature was added to the 25 cm² flasks (10mls for 75 cm² flasks). Flasks were gassed for 1 minute and put back in the incubator. Cells were grown until they reached 60-80% confluency, and then harvested for treatment with colcemid for metaphase chromosome preparations.

**Metaphase Chromosome Preparation**

Colcemid was added to the cells to be harvested to a final volume of 50-100ng/ml and flasks were returned to the incubator for between 45 minutes and one hour depending on the species of the fibroblasts grown in culture. Media were transferred into a 10ml centrifuge tube and flasks were rinsed with 3-5mls of PBS. The PBS was added to the tubes. Trypsin (Invitrogen, Mulgrave 3170 Australia) was added (3-5 mls) to the flasks and dividing cells were carefully detached and transferred to a centrifuge tube.
Flasks were rinsed one more time with 3-5 mls of PBS and added to the collected cells. Multiple tubes were prepared and the cell pellets were collected by centrifugation at 1500 rpm for 5 minutes. Cell pellets were resuspended in a small amount of PBS and pooled together, before washing again in PBS. PBS was poured off and the cell pellets were resuspended gently in 250µl of PBS. Warm (37°C) 0.075M KCl was slowly added to the tubes to a final volume of 2mls and tubes were incubated in the 37°C waterbath for 30-45 minutes.

Cells were fixed by the slow addition of -20°C fresh fix (3:1 methanol; acetic acid) to a final volume of 10mls. After centrifuging at 1500 rpm for 5 minutes, the supernatant was poured off and cell pellets were re-suspended in fresh 10mls of fix. Washes were repeated three times and then pellets were re-suspended in 1ml of fix and ready for dropping on to slides.

**Dropping cells on slides**

New microscopic slides (76mm x 26mm) were cleaned with 100% ethanol and were dried by wiping with tissues. Cells were dropped on to the slides (10-20µl for each slide) from a height of 50cm to facilitate spreading of the chromosomes. One slide was dropped and examined under the microscope. If a sufficient number of well spread metaphase chromosomes were present, more chromosome spreads were dropped onto slides. Slides were then left to dry on the bench and were stored in slide boxes at -80°C.

**Fluorescent In Situ Hybridization**

Fluorescence *in situ* hybridization (FISH) was used to visualize the location of specific BAC clones on the chromosome spreads.

**Nick translation labelling of BAC DNA**

BAC DNA was labeled with a fluorochrome using a nick translation procedure. The following reagents (25µl in total) were added, in order, to a 0.5ml PCR tube: 1X NT buffer (50mM TrisHCl, 0.1mM DTT, 10mM MgSO₄, 0.05mg/ml BSA solution), dNTP mix (38µM), dUTP (28µM of SpectrumOrange-dUTP (Vysis) or
SpectrumGreen-dUTP (Abbott Molecular Inc., IL, USA), DNAse I (0.01 U), DNA polymerase I (2.5 U), and BAC DNA (1µg). After mixing, the reactions were incubated for 1-2 hour at 15°C and tubes were then transferred on ice to stop the reactions. 4µl of the reaction mix was electrophoresed on a 1% agarose gel to check the size of the nick translation products. If there was a smear in the range of 200-700bp, then reactions were stopped, otherwise an extra (1µl) DNAse I was added and tubes were incubated for another 30-45 minutes and were checked again on the gel. Reactions were stopped by adding 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol.

**BAC DNA precipitation**

After labelling, 1µg of labelled BAC DNA was combined with C₀=1 DNA (1µg per slide), glycogen (20µg), and Milli-Q water (23µl) and 100% ethanol. The DNA was precipitated by incubating overnight at -20°C or at least 2 hours at -80°C. The probe mixture was then centrifuged for 30 minutes at maximum speed in 4°C centrifuge. The DNA pellet was dried for 10 minutes at 37°C. 20µl of pre-warmed hybridization buffer (50% v/v deionised formamide, 10% v/v dextran sulfate, 2x SSC, 40mM sodium phosphate buffer, pH7.2, 1x Denhardt’s solution) was added to the pellet and allowed to dissolve at least 30 minutes at 37°C. Probes were then processed for hybridization.

**Hybridizing BAC DNA to slides**

Before proceeding to hybridization, slides (maximum two at a time) were taken out of the -80°C and allowed to dry at room temperature. Slides were denatured for exactly 35 seconds in denaturation solution (70% v/v deionised formamide, 2x SSC) pre-warmed to 70°C. Denaturation was stopped by transferring the slides to 70% ice-cold ethanol for 5 minutes, and the slides were dehydrated through an ethanol series (3 minutes in each of 70%, 90% and 100% ethanol) and allowed to dry at room temperature. Precipitated probes in hybridization solution were removed from 37°C and denatured by incubating for 10 minutes at 70°C. Tubes containing probes were then placed immediately on ice for 2 minutes to stop the denaturation and were incubated for 30-45 minutes at 37°C to allow C₀=1 DNA to pre-anneal. A drop (18µl) of the denatured probe was put on a cover slip (22mm × 32mm) and a
Denatured slide was carefully lowered onto it. Slides were sealed with rubber cement so the probe mix did not dry out and incubated overnight at 37°C in hybridization chamber containing a wet floor.

**Washing slides**

Fresh wash solution I (0.3% Tween 20, 0.4x SSC) was pre-warmed to 60°C for at least 30 minutes. Slides were taken out of the hybridization chamber in the dark and the rubber cement seal was removed carefully. Coverslips were removed and the slides were immersed in the wash solution I, agitated for 3 seconds and incubated for 2 minutes at the 70°C. Slides were then transferred to wash solution II (0.1% Tween 20, 2x SSC) at room temperature, agitated for 5 seconds and incubated for 1 minute. Washed slides were incubated for 30 seconds in a tank containing DAPI solution (4',6-diamidino-2-phenylindole, 10µg/ml) at room temperature, rinsed in distilled water and air-dried at room temperature. One drop of Vectashield was put on a cover slip (22mm × 32mm) and dried slide was lowered onto it. The slides were then ready for microscopy.

**Microscopy examination**

Slides were analyzed using an 83000-filter set (Chroma Technology Corp Rockingham, VT, USA) mounted on a Zeiss AxioPlan2 fluorescence microscope fitted with a 100-W mercury lamp and a SPOT RT Monochrome CCD camera (Diagnostic Instruments Inc Sterling Heights, MI, USA). Images were captured and enhanced using IPLab software (Scanalytics Inc Fairfax, VA, USA).

**C₀t=1 DNA preparation**

C₀t=1 DNA was used in FISH hybridization to suppress the repetitive sequences especially in the centromeric regions. To make C₀t=1 DNA, 1mg of tammar wallaby male genomic DNA (1mg/ml) was boiled for 45 minutes to shear the DNA into 300-600bp fragments. Fragment size was checked on an agarose gel (1%) and if there were smears bigger than 600 bp, boiling was continued for another 15-30 minutes to reduce the size of the DNA fragments. The concentration of sheared DNA was measured using spectrophotometer and diluted to 1mg/ml with TE. Pre-warmed 5M
NaCl (65°C) was added to the DNA to give a final concentration of 0.3M NaCl. The DNA was denatured by incubating for 15 minutes at 97°C. DNA was then allowed to re-anneal at 65°C until $C_{ot} = 1$ (i.e. 5.92/concentration in mg per ml = time in minute). An equal volume of ice-cold S1 nuclease buffer (0.5M NaCl, 2mM ZnSO$_4$, 60mm sodium acetate, 10% sterile glycerol, pH4.6) and S1 nuclease (1 unit per µg DNA) were added and incubated for 30 minutes at 37°C. Reactions were stopped by adding 1/10 total volume of 0.5M EDTA pH 8.0 and incubating at 65°C for 10 minutes. An equal volume of isopropanol was added and the DNA was precipitated for at least 20 minutes at room temperature. After centrifugation at 13000 g for 30 minutes, the supernatant was discarded and the DNA pellet was re-suspended in TE to a final concentration of 1-2 mg/ml and incubated for 10 minutes at 50°C to dissolve the DNA. The concentration was checked and $C_{ot}$=1 DNA was stored at -20°C.

**Bioinformatics**

Sequence alignments

*Single sequence alignment and search*

Different BLAST services from NCBI (blastx, tblastn, tblastx, trace archives), UCSC Genome Browser and from Ensebmbl (BLAT) were used online in searching for sequence homologies.

*Multiple sequence alignment*

ClustalX version 2.1 (Larkin *et al.* 2007) was used for producing multiple sequence alignments. The software was downloaded from http://www.clustal.org/#Download and alignments were performed locally. In most cases the standard parameters of the software were applied. One of the advantages of ClustalX is its ability to handle big number of sequences with using a small amount of computer memory. This ability was very crucial when working with aligning hundreds of OR sequences. For editing the multiple sequence alignments, Jalview version 2 was used (Waterhouse *et al.* 2009).
**Phylogenetic tree analysis**

Multiple sequence alignments were exported from ClustalX as x.aln files and imported to Geneious 4.8.2 software for further analysis. Geneious is the user-friendliest phylogenetic package that provides a substantial number of sequence analysis tools with graphic interface. This software has the capacity to construct phylogenetic trees based on Neighbor-joining method with different bootstrap repeats. For visualizing the trees and further editing the nodes, branches, or taxa (like color-coding), FigTree v1.3.1 ([http://tree.bio.ed.ac.uk/software/figtree/](http://tree.bio.ed.ac.uk/software/figtree/)) was used.
CHAPTER 3: THE PHYSICAL MAP OF TAMMAR WALLABY 6q

Authors:
Mohammadi, A., Delbridge, M.L., Waters, P.D., and Graves, J.A.M.

Current status of paper (circle as appropriate):
Planned/ In Preparation / Submitted/ Under Revision/ Accepted [Published]

Date paper accepted for publication or anticipated date of publication:
5 May 2009

Name of journal/book:
Cytogenetic and Genome Research

Extent to which research is my own:
I conducted all the experimental and analysis under the supervision of MLD as part of the mission of the Kangaroo Genomics Centre directed by JAMG.

My contribution to writing the paper:
I wrote the draft of the manuscript and prepared figures and incorporated suggestions made by the co-authors. I corresponded with the journal editor and responded to the referee’s comments.

If paper not yet accepted, has the paper been rejected by any journals:
Not Applicable

Comments:
By using overgo probes isolated from opossum chromosome 7q and isolating specific tammar wallaby BAC clones for each overgo, I showed that the order of genes and gene blocks between opossum 7q and tammar wallaby 6q are very well conserved. Because the publication was a brief report, I will here provide background information and interpretation.
Introduction

The karyotypes of mammals show a great spread of diploid numbers and morphologies (from 2n=6 in *Muntiacus muntjak* to 2n=102 in *Tympanoctomys barrerae*), such that it was expected that the genome had been entirely scrambled by multiple rearrangements during mammalian evolution. However, modern cytological techniques and gene mapping show that this variety obscures an underlying conservatism in gene arrangement, which is much more evident in marsupials than eutherians.

Publishing the first karyotype of human in mid 1950s marked the beginning of the modern cytogenetic era (Tjio & Levan 1956, Gartler 2006). Since then there have been several major technical breakthroughs in cytology. In the 1970s the development of G-bandling techniques presented the first opportunity to expand the comparisons between the karyotypes of different species beyond the number, shape and size of the chromosomes (Sumner *et al.* 1971, Muller & Rosenkranz 1972). G-bandling chromosomes could show which part of the genomes in different lineages might be homologous. It was possible, for instance, to demonstrate considerable G-band homology between the chromosomes of primates (Dutrillaux *et al.* 1978), and the chromosomes of marsupials (Rofe & Hayman 1985).

Introducing fluorescent dyes into chromosome studies in the late 1980s was probably the most significant advance for cytogenetics (Lichter *et al.* 1988). Amongst techniques that used fluorochromes, chromosome painting spectacularly demonstrated homology across species (Wienberg *et al.* 1990, Telenius *et al.* 1992, Yang *et al.* 1999). In this technique, a single chromosome was isolated from the rest of the genome by microdissection or flow cytometry. After labelling with fluorescent dyes DNA from an isolated chromosome was hybridized to the chromosomes of another species. Labelled probes will anneal to a part or parts of the genome in target species which shares the highest sequence homology with the reference genome (Rens *et al.* 2006). This method revealed genetic homology between genomes of species that show no G-band homology (for review see Ried *et al.* 1998, Sharma & Sharma 2001). Because of the requirement for sequence homology, this technique is limited to detecting homology between eutherians, or between marsupials, although it
has been possible to demonstrate homology between the X chromosomes of marsupials and humans (Glas et al. 1999).

So far, representatives from most mammalian orders have been studied by cross-species chromosome paints, and blocks have been identified with sufficiently similar sequence. These homologous blocks are usually referred to by their location on human chromosomes (for review see Ferguson-Smith & Trifonov 2007 and references therein). Chromosome painting has been extensively applied in cytogenetic studies over the last 20 years and has made a major contribution to our understanding of genome organization and evolution in animals (Murphy et al. 2005).

Although chromosome painting has higher resolution than G-banding to reveal the homology between different genome segments in different species, it does not reveal the internal arrangements within these regions. Geneticists needed to know how single genes or blocks of genes lie on the chromosomes in different species to be able to deduce how genome organization has changed during mammal evolution.

Physical mapping of genes provided this detailed description of genome organization and reorganization. In situ hybridization using radioactive probes (genomic or cDNA) first provided a rough map of the human and other genomes (for a review see Ferguson-Smith & Trifonov 2007), but was limited in sensitivity and resolution because the use of heterologous probes could produce false signals and mis-localisation of genes. Also resolving the relative order of genes located on same chromosome was challenging, as probes for only one gene could be used at each time. Fluorescent in situ hybridization provided much better resolution, and provided the opportunity to locate two or more probes simultaneously because different dyes could be used for different probes in one experiment (Trask 1991). False-positive hybridizations could be limited by suppressing the repetitive sequences. Probes could be derived from specific locus on chromosomes, repetitive sequences of the centromeres, or a whole or part of a chromosome arm.

Of particular use were probes provided by very large genomic clones obtained by shearing the whole genome of an organism and packaging large inserts into bacterial chromosomes (known as bacterial artificial chromosomes, BACs) (Eisenbach & Giojalas 2006). This opened up this window of opportunity to locate single genes in
the genome (reviewed in Miyake & Amemiya 2004). In BAC mapping, a BAC clone that contains the sequence of interest is isolated from the genome library of the organism, and, after labelling with fluorescent dye, is hybridized to the metaphase chromosomes. Since BAC clones accept extremely large fragments of genome (100-200kb), their hybridization will result in very efficient detection of strong signal on the chromosomes. Using different fluorochromes, more than one BAC (hence more than one gene) can be localized in one experiment. This makes it possible to establish the order ORGs within genomic segments (Deakin et al. 2008).

These cytogenetic techniques are of particular value in studying unsequenced genomes, or genomes that are sequenced to a low resolution so that the assembly is incomplete and many contigs remain unassigned to chromosomes. The tammar wallaby genome is one example for which cytogenetic mapping provided substantial information from its 1.5x sequenced and fragmentally assembled genome.

The tammar wallaby karyotype was first studied in the 1960s (Graves 1967) and since then it has been included in many cytogenetic studies of cross marsupial species, including G-bandning comparisons (Rofe & Hayman 1985) and chromosome painting (Toder et al. 1997, De Leo et al. 1999, Rens et al. 2003). Like other marsupials, it has few (2n=16) and large chromosomes, making tammar wallaby very amenable to cytological study.

Many publications over decades had established that the marsupials karyotype is well conserved in both the chromosome number and G-band pattern even between distantly related species (Hayman & Martin 1974, Rofe & Hayman 1985). The diploid number of marsupial chromosomes ranges between 10 and 32, with most of the species having 2n=14 or 2n=22. Painting studies also showed that marsupial karyotypes are made up of 19 conserved segments that are present in all marsupial karyotypes and the way these segments are arranged (into chromosomes and chromosome arms) defines the difference between marsupial karyotypes. This karyotype conservation was later verified by comparative gene mappings between marsupial species. Big blocks of genes were shown to contain the same genes and in the same relative order in different marsupial genomes.
Comparative gene mapping not only helps to establish the relative gene order and gene arrangements but also benefits the genome projects. Two marsupial genomes (the gray-short-tailed South American opossum and the tammar wallaby) have been sequenced in order to inform our understanding of human genome and genome evolution. The opossum genome was sequenced at seven-fold four years ago and gene mapping played an important role in anchoring the sequences into chromosomes and also quality controlling the genome assembly. Physical mapping of 381 BAC clones helped to assign 97% of the opossum genome sequences to chromosomes (Duke et al. 2007). Gene mapping can also help to overcome the intrinsic problems of sequence assemblies such as regions of highly repetitive sequences, variations in gene copy number, or clusters of gene families that are produced by segmental duplications. Comparative gene mappings in the opossum genome helped to find some novel genes (for instance T-cell receptor; Parra et al. 2007), to locate genes that were missing from the assembly (IGF2; Lawton et al. 2007, to confirm that some genes were absent in the genome, for instance XIST; Hore et al. 2007, and to correct the location of gene families split into several clusters (MHC class I; Belov et al. 2006)

The other model marsupial (the tammar wallaby) was sequenced at a 1.5 fold level, which was not sufficiently dense to build large contigs that could be assigned to chromosomes. Therefore it was not possible to resolve the gene order in the tammar wallaby from sequence data, and it was necessary to construct a dense physical map of genes/gene blocks. By a strategy developed within the ARC Centre of Excellence for Kangaroo Genomics (KanGO), it was possible to map large regions that are conserved between the opossum and human genomes by the mapping genes at the end of these conserved blocks (Deakin et al. 2008). Comparative physical maps of all tammar wallaby chromosomes were then compared to chromosomes of opossum. The opossum genome sequence had already been obtained and assembled in contigs that had been assigned to chromosomes (Duke et al. 2007), and to the human genome.

I participated in the tammar wallaby physical mapping project, using the first part of my PhD research to examine the gene content of a region in the tammar wallaby karyotype. G-bandning had indicated that the long arm of chromosome 6 in the tammar wallaby was well conserved, and I examined the conservation of gene order
on long arm of chromosome 6 in the tammar wallaby by localising individual genes within the region.

**Results and Discussion**

The following peer reviewed publication describes the results obtained from this research. For this publication I conducted all the experiments (dry and wet lab techniques), analysed the results with help from my supervisory panel and drafted the paper.
Conservation of a chromosome arm in two distantly related marsupial species

A. Mohammadi  M.L. Delbridge  P.D. Waters  J.A. Marshall Graves

ARC Centre of Excellence for Kangaroo Genomics, Research School of Biological Sciences,
The Australian National University, Canberra, ACT (Australia)

Abstract

Marsupials, which diverged from eutherian mammals 150 million years ago (MYA), occupy a phylogenetic position that is very valuable in genome comparisons of mammal and other vertebrate species. Within the marsupials, the Australian and American clades (represented by the tammar wallaby Macropus eugenii, and the opossum Monodelphis domestica) diverged about 70 MYA. G-banding and chromosome painting suggest that tammar wallaby chromosome 6q has homology to opossum chromosome 7q. We tested this conservation by physically mapping the tammar wallaby orthologs of opossum chromosome 7q genes. We isolated 26 tammar wallaby BAC clones that contained orthologs of 16 opossum chromosome 7q genes. We used fluorescence in situ hybridization (FISH) to show that they all mapped specifically to the tammar wallaby chromosome 6q in nearly the same order as their orthologs on opossum chromosome 7q. Thus, this chromosome arm is genetically, as well as cytologically, conserved over the 55–80 million years that separate kangaroos and the opossum.

Genomic data are now available from many vertebrate species, including eutherian mammals that are relatively closely related to humans (e.g. chimpanzees, mice, and dogs) and those more distantly related (e.g. birds and fish, which diverged 310 and 450 MYA respectively) (Murphy et al., 2001; Thomas and Touchman, 2002). Sequences from different eutherian mammals are often too similar for comparisons to reveal important conserved elements, whereas the sequences from eutherian mammals and birds are often too different to align (Wakefield and Graves, 2003). Marsupials (mammalian infraclass Metatheria), which diverged from eutherian mammals 150 MYA (Kumar and Hedges, 1998), occupy a phylogenetic middle ground that is very valuable in genome comparisons of mammal and other vertebrate species.

The genome sequence of the Brazilian grey short-tailed opossum (Monodelphis domestica) (Mikkelsen et al., 2007) has proved invaluable for comparative analyses, such as the evolution of the major histocompatibility locus (Belov et al., 2006) and X chromosome inactivation (Davidow et al., 2007; Hore et al., 2007). Sequence from an Australian marsupial, distantly related to opossum, will provide data to distinguish ancestral and lineage-specific changes. However, the Australian model kangaroo, the tammar wallaby (Macropus eugenii) is the species in which most of the classic work on marsupial reproductive physiology has been carried out. This species is now being sequenced, but only to a depth of 2× in Australia (The Australian Genome Research Facility, Melbourne) and the United States (Baylor College of Medicine, Houston). Assembly of kangaroo sequence will require reference to the opossum, so detailed comparative maps are required.

The authors wish to thank The ARC Centre of Excellence for Kangaroo Genomics for funding support.
Table 1. Name and sequence of overture probes used for physical mapping

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Overture A (5'-3')</th>
<th>Overture B (5'-3')</th>
<th>Isolated AGI BACs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIGEF7</td>
<td>GAATGCCTGCTGCTGCTGCTGATG</td>
<td>TCAAAGATGTCACCAAGGTACAGCG</td>
<td>511 G21, 603 K10</td>
</tr>
<tr>
<td>MBNL2</td>
<td>AAAATAGGATTTGTTGAGAACAG</td>
<td>ACATCCGAGATCCAAATTGCTGCA</td>
<td>94 O5</td>
</tr>
<tr>
<td>S1TOR5</td>
<td>CTGATGGAGATAACATAGACAG</td>
<td>GTGATTGGCGCTGCTGCTGCTGCT</td>
<td>82 N14</td>
</tr>
<tr>
<td>KHL1</td>
<td>TTTCACAAAATGCCCCAGAGCGG</td>
<td>ACTGACTCGAGAGACCCAGCGCCCT</td>
<td>119 E18, 88 N23</td>
</tr>
<tr>
<td>SLAIN1</td>
<td>TTTTTTTGCTACCTAGTTTTCTTG</td>
<td>CAGTTCGATATCCTCACTCAAAGTAA</td>
<td>95 N1, 123 O6, 135 I6</td>
</tr>
<tr>
<td>PC12H48</td>
<td>GCCGAAGAAGTCCAGCTCTCTCCTAG</td>
<td>TATAAAAAGGCTGAGTCCTAAAGGGG</td>
<td>355 M12</td>
</tr>
<tr>
<td>SPERI</td>
<td>TGTTAGATCGGAGGCACCCATGAGC</td>
<td>ATGGGCTCGAAGGCTGCCTGAGT</td>
<td>297 F10</td>
</tr>
<tr>
<td>ARCA12</td>
<td>ATCGAGAGGGAGAGATGGGTAATG</td>
<td>TGGACCGGGCAATTCTCATTACCC</td>
<td>370 J16, 345 I22, 423 B16, 387 B23</td>
</tr>
<tr>
<td>MSH16</td>
<td>CAGGCTCTAGGTTCCTTCTCACTGAT</td>
<td>CTTAGGCTACTTCTTCTCACAGT</td>
<td>374 G23, 357 K23</td>
</tr>
<tr>
<td>DCUNID1</td>
<td>GCTGGCTGGCTCAATCCGATCAGCGG</td>
<td>CTAAAGGCTGATGCGGCGGGATTAG</td>
<td>316 I2</td>
</tr>
<tr>
<td>HPS3</td>
<td>GTTGGCTCTGTTGAGATACGCGCTTC</td>
<td>CAGACGCGAAGCCTATATGGAAATG</td>
<td>364 F12</td>
</tr>
<tr>
<td>WDR60</td>
<td>TTTATATTCTTCAGCATACATAGT</td>
<td>ATGCCTCTCAGCAAGCTGAGTTG</td>
<td>198 A5, 170 C23</td>
</tr>
<tr>
<td>PHE3</td>
<td>AGTGGCGCTTTAGAAGAAATGCGG</td>
<td>AAAGGGCTCTGCTTCCCGGATTTC</td>
<td>84 F17</td>
</tr>
<tr>
<td>NMD3</td>
<td>CTCAACAGATCGCTTGATGAAAAGT</td>
<td>TGGACCGAGACGAGACATCTTTT</td>
<td>430 I8</td>
</tr>
<tr>
<td>S1ED1</td>
<td>GGAAATAGAACGCGCATATACAGG</td>
<td>AAATGCAATTTGAGGGGACTGTTAT</td>
<td>55 P4, 70 H1</td>
</tr>
<tr>
<td>SICA1</td>
<td>GTGGCTAGACACATTGAGGAATTTG</td>
<td>TTTCGAGAGGGCATCCTTCATTCCA</td>
<td>68 H17, 84 B1</td>
</tr>
</tbody>
</table>

A great advantage for comparative genomics is the extensive conservation of marsupial karyotypes even between distantly related groups (Hayman and Martin, 1974; Role and Hayman, 1985; Rens et al., 1999, 2003). The karyotypes of the tammar wallaby (2n = 16) has been characterized (Alsop et al., 2005), and chromosome painting used to propose homologies with the karyotypes of other marsupials (Rens et al., 1999, 2003). G-banding and chromosome painting indicate that the short arm of tammar chromosome 6 corresponds to the long arm of opossum chromosome 5, whereas its long arm and pericentric region correspond to most of the long arm of the opossum chromosome 7 (Rens et al., 1999, 2003). We tested the genetic homology of tammar 6q with opossum 7q by isolating tammar BACs containing orthologs of opossum chromosome 7q genes from gene blocks that are conserved between opossum and human. We then anchored each of these blocks in tammar wallaby by localizing these genes using fluorescence in situ hybridization.

Materials and methods

BAC library screening

A male tammar wallaby genomic BAC library obtained from Arizona Genomics Institute (Tucson, AZ, USA) was screened with radioactively labeled overviews designed from the tammar wallaby trace archives (Table 1) according to Ross et al. (1999).

Chromosome preparation

Primary fibroblast cultures were established from ear biopsies of tammar wallabies (Macropus eugenii), collected under The Australian National University Animal Experimentation Ethics Committee proposals R.C.G.02.00 and R.C.G.08.03. Cultures were maintained and metaphase chromosome spreads were prepared as previously described (Koima and Graves, 2005).

Fluorescence in situ hybridization (FISH)

Extracted BAC DNAs (Promega Wizard Plus SV Miniprep DNA Purification system) were labeled in nick translation reactions with SpectrumOrange dUTP (Vysis) or SpectrumGreen dUTP (Abbott Molecular Inc., IL, USA) and hybridized to the chromosome preparations as previously described (Alsop et al., 2005). The relative order of BAC clones was determined by co-hybridizing BACs labeled with different fluorochromes to the same metaphase chromosome preparation. For each localization, approximately 80% of metaphase chromosome spreads had signals in the same position on either 3 or 4 chromosomes. The numbers and positions of signals were recorded for 8–12 metaphases for each BAC localization. After fluorescence microscopy, images were captured and enhanced using IPLab software (Scanalytics Inc., Fairfax, VA, USA).

Chromosomal location of genes in chicken, opossum, dog, and human were obtained from the latest genome assembly of each organism in the Ensembl genome browser (www.ensembl.org/).

Results and Discussion

Map construction in the kangaroo is greatly facilitated by the availability of sequence data from the opossum, which can be used to identify sequences for map-
ping orthologs to tammar wallaby chromosomes (Deakin et al., 2008). The Ensembl genome assembly (Mon-Don5) of the Brazilian grey short-tailed opossum was used to select single copy opossum genes with clear human orthologs to anchor opossum-human conserved regions. Overgo probes were designed against non-repetitive trace sequences from tammar orthologs of opossum chromosome 7q genes. Overgos were pooled and simultaneously hybridized to the BAC library filters. Dot blots were used to determine which overgo hybridized to which positive BAC clone (Deakin et al., 2008). 28 tammar wallaby BAC clones were isolated that contained tammar orthologs of the 16 opossum 7q genes (Table 1).

Fluorescence in situ hybridization was used to localize the BAC clones to tammar wallaby metaphase chromosomes. All of these BACs mapped specifically to tammar chromosome 6q. Thus, all 16 opossum 7q genes were located on the long arm of tammar wallaby chromosome 6, as predicted by the chromosome painting.

The gene order on tammar 6q was almost the same as the order of the orthologs on opossum chromosome 7q (Fig. 1). Exceptions were the reversal of gene order between KLHL1 and SLAIN1 within a single conserved

Fig. 1. The location of mapped genes (represented by colored bars) on tammar wallaby chromosome 6q compared to their orthologs in other species. Numbers represent the chromosome numbers in each species.

Chromosome arm conservation in two distantly related marsupials
opossum-human block, and reversal of order of the ABCA12 and MDH1B genes.

Our gene mapping shows that genes representing 16 blocks along opossum chromosome 7q have largely the same order along the tammar wallaby chromosome 6q, only two small interstitial inversions were detected between the gene order in tammar wallaby and opossum. Comparisons of the positions and order of these genes in other vertebrate species indicate that tammar 6q contains three conserved regions whose homologs lie on different chromosomes in other vertebrates (Fig. 1). These three blocks lie on three chromosomes in human (one of which is fragmented onto three dog chromosomes) and three in chicken, suggesting that fusion occurred in the marsupial ancestor 55–80 MYA.

The proximal regions of opossum 7q/tammar wallaby 6q, containing genes from ARHGER7 to SPERT (blue block, Fig. 1), represent a gene block that is well conserved in all species, including birds and mammals. There are only two minor internal rearrangements of gene order within this region. The central regions of opossum chromosome 7q/tammar wallaby chromosome 6q contain two mapped genes (red block, Fig. 1) that are found together in all mammals. The distal regions of opossum chromosome 7q/tammar wallaby chromosome 6q contain genes from DCUNI1D to SCHIP1 (green block, Fig. 1), which are all located together on chromosome 9 in the chicken, an arrangement that is therefore likely to be ancestral. However, they are split between human chromosomes 2 and 3, and split further in dog (which has a very fragmented karyotype). This suggests that fixation of an ancestral region occurred in the eutherian lineage.

Analysis of this small conserved region of the marsupial genome shows that comparative gene maps of the tammar wallaby and opossum will help us to reconstruct the karyotype of a mammal ancestor that lived more than 150 MYA, and chart the changes that occurred in the eutherian and marsupial lineages.

Acknowledgements

The authors wish to thank Daniel McMillan for technical help.

References

Conclusion

My studies demonstrated that the long arm of tammar chromosome 6 shares gene content with the opossum chromosome 7, and gene order differs only in two small inversions. Mapping shows that this region is represented on human chromosomes 13, and small regions of human chromosomes 2 and 3.

The comparisons between this region in the tammar and opossum are in line with the striking conservation revealed by studies of chromosome morphology (Hayman DL 1974), G-banding (Rofe & Hayman 1985) and cross-species chromosome painting (Rens et al. 2003) in the marsupial genome that indicated there have been a limited number of chromosome rearrangements in marsupials and their karyotypes are well conserved. This level of conservation is extraordinary compared to that between the genomes of eutherian mammals that diverged about the same time ago (55-80 MYA), for instance, human and mouse.

Is this tight conservation of karyotype accompanied by conservation of genes ORG families? The low rate of rearrangement might protect essential genes, keeping them safe from being translocated away from their regulatory elements. If the low rate of karyotypic change reflects the relative paucity of repetitive elements at which rearrangements often occur, it is unlikely to be correlated with a low rate of sequence change within genes and gene families.

Some gene families, however, have a very dynamic evolution. The conservation of karyotype might put some sort of constraints on the evolution of such large dynamic gene families. Olfactory receptors are one example. This gene family, the biggest in the mammalian genome, seems to be very dynamic, and different members of the family have arisen by a variety of genomic events (duplications, translocations, chromosomal fusions and fissions). This versatility may have evolved to arm the animal with a maximum number and diversity of olfactory receptors so that it may adapt more quickly to a changing environment. (Niimura & Nei 2006, Hayden et al. 2010).
To see how a conserved karyotype of marsupials had affected the evolution of such a complex and dispersed gene family, the rest of this PhD research was dedicated to analysing the olfactory receptor (OR) gene family of tammar wallaby. The ORG sequences were isolated from the first assembly of the genome, classified and were compared with the OR repertoires of another marsupial, platypus and eutherian mammals. The genomic organization of these genes were also studied by locating the major ORG clusters in tammar wallaby and a detailed comparison of one ORG cluster between different mammals and two marsupials.
CHAPTER 4: THE OLFACTORY RECEPTOR GENE

SUPERFAMILY OF THE KANGAROO MACROPUS EUGENII (THE TAMMAR WALLABY)

Authors:
Mohammadi, A., Patel, H.R., Delbridge, M.L., and Graves, J.A.M.

Current status of paper (circle as appropriate):
Planned [ ] In Preparation [ ] Submitted/ Under Revision/ Accepted/ Published

Date paper accepted for publication or anticipated date of publication:
September 2011

Name of journal/book:
---

Extent to which research is my own:
I carried out the ORG classification, phylogenetic analysis, FISH mapping and drafted parts of the manuscript. HRP extracted the ORGs from the genome assembly. MLD supervised the experiments and coordinated the manuscript. JAMG coordinated the research and edited the manuscript.

My contribution to writing the paper:
I wrote the draft of the manuscript and prepared figures and incorporated suggestions made by the co-authors.

If paper not yet accepted, has the paper been rejected by any journals:
This paper was initially sent to BMC Genomics. It was favorably reviewed by two reviewers, who offered valuable advice, but was rejected (November 2010) because the third reviewer considered that this paper contained insufficient novelty and tammar wallaby is not an interesting species for the study of evolution of ORGs. Suggested minor corrections by the other reviewers have been made and the manuscript will be submitted to a different journal in September 2011.
Comments:

Since this chapter has been prepared for submission as a major paper, it has a substantial introduction and discussion, and also a brief description of methods, and it is unnecessary to repeat them in a separated introduction and discussion.
The olfactory receptor gene superfamily of the kangaroo Macropus eugenii (the tammar wallaby)

Amir Mohammadi¹*, Hardip R. Patel², Margaret L. Delbridge¹, Jennifer A. Marshall Graves¹

¹ARC Centre of Excellence for Kangaroo Genomics, Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra 2601, Australia

²Molecular genetics division, Victor Chang Cardiac Research Institute, Darlinghurst, NSW 2010, Australia

*Corresponding author

Email addresses:

AM: amir.mohammadi@anu.edu.au

HRP: h.patel@victorchang.edu.au

MLD: margaret.delbridge@anu.edu.au

JAMG: jenny.graves@anu.edu.au
Abstract

Olfaction is an important sense for animal feeding, reproduction and survival, and many mammals have evolved large and complex families of genes to enhance this sense. The sense of smell is mediated by a group of G-protein coupled receptors known as olfactory receptors (ORs), encoded by the largest gene family in the mammalian genome. In vertebrates, the ongoing process of birth-and-death evolution has led to a huge variation in the number of ORs in each lineage. Here we describe the isolation, classification and mapping of olfactory receptor genes (ORGs) from the newly sequenced genome of the tammar wallaby. The tammar wallaby genome contains 1,753 ORG sequences (one of the highest in any mammal), including 870 pseudogenes. Based on sequence homology, these receptors are classified into two classes, 18 families, and 240 subfamilies. Major ORG clusters are found on nearly all tammar wallaby chromosomes, and they lie in conserved syntenic regions of mammalian genomes. The tammar wallaby and opossum ORGs show a high level of similarity in their sequence, total numbers and family/subfamily distributions. The tammar wallaby has a very large ORG repertoire that is very similar to that of a distantly related American marsupial, the gray short-tailed opossum. This implies that there have been relatively few lineage-specific evolutionary events (such as gene duplications or pseudogenizations) in the two marsupial lineages, so their ORG repertoires have remained relatively stable for the 80 million years since the Australian and American marsupial lineages diverged. This stability contrasts with the one-to-many relationships that often exist between the ORG repertoires of any two placental mammals with a similar divergence time.
Introduction

Olfaction is a critical sense that animals use for communicating with their surrounding environment, for finding prey and mates, as well as for avoiding dangers. The evolution of a wide array of odor receptors has been a hallmark of many placental mammals, such that odor receptors constitute the largest gene family (>2000) in mammalian genomes (Issel-Tarver & Rine 1997, Rouquier & Giorgi 2007, Niimura 2009). However, numbers and variety of olfactory receptors (ORGs) vary widely between species with different lifestyles (Freitag et al. 1998, Glusman et al. 2000a, Niimura & Nei 2007). The genomes of eutherian mammals show considerable differences in the number of ORGs they contain, even between closely related species (Gilad et al. 2003, Niimura & Nei 2005b, Quignon et al. 2005). The large size of the mammalian ORG family is thought to have been established prior to the divergence of monotreme and placental mammals (Kishida 2008), but it is hypothesized that the existing diversity was established later, during eutherian evolution. It is thus of great interest to examine the ORG family also in marsupials, the therian mammals most distantly related from placental mammals such as human and mouse.

The first step of odor perception involves specific binding of odorant molecules to olfactory receptors (ORs), which are located on the surface of the cilia of the olfactory sensory neurons in the olfactory epithelium (Buck & Axel 1991, Mombaerts 2004, Niimura & Nei 2006, Nei et al. 2008). OR proteins belong to the superfamily of seven-transmembrane domain G protein-coupled receptors. OR-encoding genes are short (≈ 1kb), intronless sequences that are organized in clusters of one to > 100 genes, scattered around the genome (Glusman et al. 2001, Godfrey et al. 2004). These clusters are the products of repeated tandem and segmental duplications and thus contain ORG sequences that are very similar (Ben-Arie et al. 1994, Sosinsky et al. 2000, Glusman et al. 2001). Almost 90% of ORG clusters have a conserved location within the same orthologous syntenic blocks, and are flanked by the same genes in many mammalian species (Aloni et al. 2006).

Availability of genomic sequences from many vertebrate species has enabled us to understand the evolutionary mechanisms shaping this large gene family. The ORG
family has evolved as a result of "genomic drift" events whereby slippage during recombination has resulted in expansion of this gene family. However, loss of certain families and subfamilies of ORGs in mammals have been attributed to adaptive evolution (Nei et al. 2008, Hayden et al. 2010). Indeed, as species have evolved to rely on other senses than olfaction, dramatic decrease in the number of ORGs have been observed (Steiger et al. 2009, Hayden et al. 2010). For example, primates have considerably lower number of ORGs than other eutherians, and it appears that the timing of this decrease in numbers coincides with their acquisition of tri-chromatic vision (Gilad et al. 2004).

These studies were, however, driven by the analysis of ORGs in eutherian mammals. ORG repertoire of the short-tailed opossum (Monodelphis domestica), a marsupial, is strikingly different from that of eutherian mammals for certain families and subfamilies (Mikkelsen et al. 2007, Niimura & Nei 2007). It was found that the opossum genome contains many specific odorant receptors not found in placental mammals. A loss of these ORGs in the eutherian lineage is consistent with the observation that the human genome has experienced relatively more segmental duplications, and hence more gains and losses in the ORGs, than opossum (Young & Trask 2002, Mikkelsen et al. 2007). Comparing the ORG repertoire of opossum with that of another marsupial could give more insights into the evolution of ORG conservation and diversity. It is shown that the ORG family has expanded twice during placental evolution, once after the divergence from monotremes, and once after the divergence of placentals from marsupials. However, for this theory to be supported all marsupials should show the evidence of this expansion and hence have OR repertoire that is similar to opossum.

Genomic sequence of Australian model marsupial, tammar wallaby (Macropus eugenii), is now available to test this hypothesis (Nilsson et al. 2004). Here we describe the ORG repertoire of the tammar wallaby, which was compiled from the first assembly of its 1.5x-sequenced genome. Comparing the sequences of tammar wallaby ORGs with opossum ORG sequences demonstrates that ORGs are conserved even between marsupials that shared a common ancestor 70 MYA. The chromosomal locations of most conserved mammalian ORG clusters were also physically mapped in the tammar wallaby, and results showed that these clusters lie within orthologous syntenic blocks shared by other mammals.
Methods

OR sequence identification (data mining) and classification

ORG sequences were identified (Figure 1) in the first draft assembly of tammar wallaby genome that was sequenced to the depth of 1.5x at the Australian Genome Research Facility (AGRF, Melbourne) and the Baylor Genome Institute (Houston, Texas). The tammar wallaby genome was searched by using the TBLASTN program (Altschul et al. 1997) and a set of 924 functional full-length OR amino acid sequences as queries (Safran et al. 2003, Niimura 2009), as previously described (Niimura & Nei 2005a). All tammar wallaby matches were then extended up to 1,000 bp on either side and optimal conceptual translation were generated by using Fasty35_t (Pearson et al. 1997). Amino acid sequences lacking start and stop codons were dynamically extended in both directions to optimum start and stop signals.

Subsequently, 250-350 amino acid tammar wallaby OR proteins that had an uninterrupted coding frame, a start and a stop codon, and had fewer than five gaps were added to the pool of vertebrate ORs and tammar wallaby genome was searched again as described above. These rounds of search and translation were repeated until no new sequence was identified.

To remove false positives from the newly identified tammar ORGs, Hidden Markov Model (HMM) search was performed by using the HMMER program (Eddy 1998). 1,221 full-length functional vertebrate OR amino acid sequences were first aligned using the MUSCLE program (Edgar 2004), and an HMM was built from this for HMM search. All the extracted and translated sequences were compared with this consensus profile, and non-OR sequences were removed from the pool with e-value greater than 0.1. Redundant sequences (100% identical), and sequences that overlapped in genomic assembly of tammar wallaby were also removed.
Classification of tammar wallaby ORGs

Tammar wallaby full-length functional ORG sequences were classified into families and subfamilies as described earlier (Glusman et al. 2000b). Full-length tammar wallaby OR amino acid sequences were aligned using the ClustalX 2.1 alignment program (Larkin et al. 2007). Multiple sequence alignments were visually inspected and gaps were manually adjusted where necessary, using the Jalview alignment editor v.2.0 (Waterhouse et al. 2009). Phylogenetic analysis was performed using standard parameters in the Geneious 4.8.2 package (Drummond AJ 2009). Neighbour-joining (NJ) trees were rooted with the opossum Beta3 adrenergic receptor sequence.

**Figure 1.** Flow chart for the isolation of OR sequences from the first draft assembly of tammar wallaby genome.
Tree representations were visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

In a neighbour-joining tree, all sequences connected by a node and ≥40% similar in amino acid sequences were classified as families. A sub-family is classified within a family for all sequences connected by a node and sharing ≥60% amino acid sequence similarity. Once the full-length functional genes were classified by this method, pseudogenes and short sequences were allocated to families and subfamilies based on their highest similarity to the full-length functional tammar wallaby ORGs.

**OR sequence analysis**

WebLogo 3 was used for generating the sequence logos (Crooks et al. 2004). To avoid sequence heterogeneity before generating the sequence logos, short parts of the sequences were removed from the N terminus (up to 28 residues upstream of the first transmembrane region, TM1) and C terminus (beyond 19 residues downstream of the seventh transmembrane region (TM7)). Also, all alignment gaps of 1-3 amino acids that were present in more than 95% of the sequences were manually deleted before the sequence logo generation. Positions of the potential transmembrane domains were predicted using TMHMM v.2.0 (Krogh et al. 2001)

**Mapping ORG clusters**

The locations of conserved ORG clusters in the opossum genome, identified as Clusters in Conservation (CLIC) by Aloni (2006) (Aloni et al. 2006), were extracted by matching the location of ORG clusters in HORDE (which is based on the MonDom1 assembly) with the latest genome assembly of opossum (MonDom5) on the Ensembl (Hubbard et al. 2009). Out of 48 conserved mammalian ORG clusters, 32 clusters were identified (including 26 previously identified (Aloni et al. 2006) and 6 new clusters that we located) in the opossum genome.

The location of ORG clusters in the tammar wallaby was predicted by linking the information from the physical and linkage maps of the tammar wallaby chromosomes (see Additional file 1). Highly conserved representatives of each CLIC were selected to determine the chromosomal location of the cluster in the tammar wallaby genome.
by fluorescence *in situ* hybridisation (FISH). Some clusters known to be large were divided into smaller sub-clusters based on their size and composition (e.g. if a cluster contained many genes from just few subfamilies, all the genes of each subfamily were considered as one sub-cluster), and a representative ORG was selected for localisation by FISH. Only ORGs that had orthologues in five species (human, mouse, rat, dog, and opossum) were selected for physical localisation to tammar chromosomes to maximise the likelihood that they would be well conserved in the tammar wallaby. Conserved representative ORGs from both ends of the clusters/sub-clusters and also from each gene subfamily present in the cluster/sub-cluster were selected (see Additional file 2). Tammar orthologues of the chosen ORGs were identified and their whole contig sequence was aligned to the opossum genome (BLAT in Ensembl genome browser) to ensure that they had the highest homology to the same opossum ORG cluster.

To isolate BAC clones for localising the ORGs, unique overgo (overlapping oligo) probes were then designed from sequences immediately flanking the chosen ORGs in their contigs. This strategy was used to minimise the possibility of non-specific hybridisation of probes to ORG coding sequence. A male tammar wallaby BAC library (Me_KBa; Arizona Genomics Institute, Tuscon, AZ, USA) was screened with radioactively labelled overgo probes (Ross MT 1999). Additional file 1 shows the orthologous ORG chosen for mapping, and the corresponding tammar wallaby genomic BAC clones that were mapped for each gene.

FISH was used to map positive BAC clones onto tammar wallaby metaphase chromosome spreads (Mohammadi *et al.* 2009). Briefly, BAC DNA was purified (Promega Wizard Plus SV Miniprep DNA Purification system) and labelled in nick translation reactions with SpectrumOrange-dUTP (Vysis) or SpectrumGreen-dUTP (Abbott Molecular Inc., IL, USA). The labelled BAC DNA was hybridized overnight to metaphase chromosome preparations at 37°C. Images were captured used a Zeiss Axioplan2 epifluorescence microscope and thermoelectronically-cooled charge-coupled device camera (RT Monochrome Spot, Diagnostic Instruments, Sterling Heights, MI, USA). The numbers and positions of signals were recorded for 8–12 metaphases for each BAC localization. After fluorescence microscopy, images were captured and enhanced using IPLab software (Scanalytics Inc., Fairfax, VA, USA).
Results

Database mining and OR identification

The 1.5x-sequenced genome of tammar wallaby (Meug_1.0) was surveyed to isolate all ORG coding regions (Figure 1). A total of 5,230 potential ORG sequences were collected. Of these, only 1,753 sequences aligned to the HMM consensus profile at the set threshold e-value of 0.1 and were considered to be tammar wallaby ORGs. Of these ORG sequences, 951 were >810 bp long containing start and stop codons and spanning all seven transmembrane regions. The other 802 were short sequences within the potential coding region that ranged in size from 174-810 bp and did not contain start or stop codon.

Of the 1,753 full-length and short ORG sequences, about half (849) contained seven transmembrane domains, no insertion or deletion mutations, and no more than 5 amino acid gaps in their entire length. These sequences were therefore considered to be functional. The other 904 sequences harboured at least one sequence change that would result in the production of a truncated protein, and so these sequences were considered to be pseudogenes. This may be an underestimation of the number of pseudogenes, as short sequences considered to be functional may contain mutations in the regions missing from the database.

Conserved sequence motifs

We found that tammar wallaby ORGs encode proteins that contain well-conserved motifs that are common to all vertebrates. We aligned 482 predicted full-length functional tammar wallaby OR protein sequences (representing most subfamilies) and displayed the conservation across the alignment using WebLogo. Like other mammalian ORG sequences, tammar wallaby ORGs are more conserved across their first, second, sixth, and seven transmembrane (TM) domains than in other TM domains (Figure 2). It has been suggested that this strong sequence conservation reflects the need for precise interactions of ORs with cell membranes (Malnic et al. 2004).
Several motifs that are common to all vertebrate OR proteins are also well conserved in the tammar wallaby OR proteins. For example, the LHTSPMY motif in the first intracellular loop (IC1), the MAYDRYVAIC motif which spans from part of the TM3 to IC2, the SY amino acids in TM5, the KAFSTC motif spanning from IC3 to TM6, and the PMLNP motif within TM7 are all well conserved. KAFSTC is often located after acidic amino acids, and can serve as a target for kinase-mediated serine/threonine phosphorylation (Bohm et al. 1997). Most G-protein coupled receptor proteins, such as the ORs, contain the PMLNP motif that is proposed to have a role in receptor internalization and desensitization (Gripentrog et al. 2000). The conservation of IC2 and IC3, unique to ORs, is proposed to be due to constraints posed by its interaction with the cytoplasmic G protein to initiate the second messenger cascade (Pilpel & Lancet 1999).

**Figure 2.** Conservation of the amino acid sequences in the OR repertoire of tammar wallaby. Sequence logo was created from an alignment of 482 full-length, intact OR sequences using the program WebLogo after manual editing. For simplicity, parts of the N- and C- termini that contained no significant sequence conservation were removed (see Material and Methods). Relative frequency of amino acids at each position is represented by the height of their letters, so that the overall height at any position shows the level of sequence conservation at that position. Predicted locations of transmembrane segments (yellow bars, TM), intracellular (IC), and extracellular (EC) domains are shown below the amino acids. Location and length of TMs have not been experimentally verified.
Not surprisingly, the most divergent part of the ORs lies within the extracellular loop 3 (EC3), which is proposed to interact with the odorant molecules (Pilpel & Lancet 1999, Irie-Kushiyama et al. 2004). This region of the tammar wallaby proteins contains only 9% sequence similarity between different protein molecules, which is within the range of the levels of similarity found in all other vertebrates (6-11%). This is consistent with a role for this region in binding very variable odor molecules.

**Tammar ORG classification**

Once the tammar wallaby ORG repertoire was identified, the functional and non-functional ORGs of the repertoire were classified into families and subfamilies, so that comparisons could be made between the size, composition and functionality of the tammar wallaby repertoire with the repertoires in other animals.

The tammar wallaby ORGs were classified into families and subfamilies (Glusman et al. 2000a). Firstly full-length functional genes were arranged into families (containing genes with 40% or greater sequence similarity) and subfamilies (60% or greater sequence similarity). Pseudogenes and short sequences were given a family and subfamily identity based on their homology to classified full-length tammar ORGs. In this way the tammar wallaby ORGs were divided into two classes, containing 18 families and 240 subfamilies (Figure 3 and Table 1). The numbers of ORGs in each family and the proportion of functional genes are broadly consistent with the classification of ORGs and ORG families in the opossum (Figure 4).

As for most other mammals, the tammar wallaby ORG repertoire contains four families of class I genes (51, 52, 55, and 56). Most of the class I genes are members of families 51 and 52, and therefore almost 15% of the entire tammar wallaby ORG repertoire lies in these two families which form a phylogenetic clade distinct from the rest of ORG families (Figure 3). In these families, a relatively large proportion (51-64%) of the genes are functional. Among tammar wallaby Class I ORG families, family 51 contains the largest number of subfamilies, suggesting that it contains ORGs that are more divergent than those in other families. In the tammar wallaby, all members of family 55 are non-functional (Table 1), suggesting that in tammar wallaby, as in human and platypus (Figure 4), family 55 is not a contributor to the functional repertoire of mammals.
Figure 3. The OR repertoire of tammar wallaby. A neighbour-joining tree of 456 full-length functional OR protein sequences rooted with opossum adrenergic beta-receptor. Genes are coded based on their family allocations and the scale corresponds to 10% sequence divergence. Only families 14, 51, and 52 form one distinct monophyletic clade, and other families intermix with each other.

Class II tammar wallaby ORGs were divided into fourteen families (1-14), of which family 5 is the biggest and family 3 the smallest. Approximately the same proportion of the ORG repertoire is found in each family in the tammar wallaby as is found in opossum, human and platypus (Figure 4). However, there are large differences in the numbers of genes classified into in different families. Among the remaining Class II families there are differences in composition between the relatively small families 7, 9 and 11, and the very large families, especially families 4 and 5. The small families
7, 9 and 11, contain 51 genes or less, of which over about 60% are functional. In addition, the genes of these families are found in only a small number of subfamilies (less than 10). In contrast, families 4 and 5 contain around 250 genes, of which only 40% are functional and which are represented by more than 20 subfamilies. This might suggest that there are different patterns of evolution acting on the different ORG families depending on their function.

Table 1. The composition of olfactory receptor gene families in tammar wallaby.

<table>
<thead>
<tr>
<th>OR Class</th>
<th>OR Family</th>
<th>Number of sequences</th>
<th>% of Functional sequences</th>
<th>Number of subfamilies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class II-1483 genes</td>
<td>1</td>
<td>110</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>174</td>
<td>61</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>268</td>
<td>39</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>246</td>
<td>37</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>101</td>
<td>51</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>27</td>
<td>59</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>135</td>
<td>40</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>21</td>
<td>80</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>138</td>
<td>47</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>51</td>
<td>62</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>123</td>
<td>50</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>86</td>
<td>37</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>UN*</td>
<td>2</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>Class I-270 genes</td>
<td>51</td>
<td>95</td>
<td>64</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>150</td>
<td>51</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>21</td>
<td>38</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>1753</td>
<td>48</td>
<td>240</td>
<td></td>
</tr>
</tbody>
</table>

*, Unclassified OR sequence
The function of the genes in the smaller ORG families 7, 9, and 11 may have constrained the evolution of these families, and adaptive evolutionary patterns responding to different environmental influences may be acting on the genes in these families. In contrast, large, divergent families, such as families 4 and 5 may be evolving by random genetic drift mechanisms, and be the families in which novel gene diversity is more frequently generated.

One interesting Class II family is family 14, which although it is moderately large (containing 86 genes), contains ORGs belonging to only a small number of subfamilies suggesting that the ORGs in this family are more homogeneous than those in other families (Table 1). Family 14 OR sequences occur also in the opossum, and in the platypus genomes, but not in eutherians. However, 63% of the genes in this
family are non-functional in tammar wallaby. Family 14 in opossum and platypus contains the lowest proportion of functional genes. This family is very large in platypus, suggesting that it has been specifically expanded in monotremes (Figure 4). Together, these results may indicate that family 14, more than other families, has been subject to a period of rapid gain and loss of genes in different lineages, as is typical of a birth and death pattern of evolution.

**Sequence homology in marsupial ORGs**

We compared the sequence homology of ORGs between two distantly related marsupials (the tammar wallaby and opossum), and with the platypus (a monotreme mammal), to determine if differential amplification of the ORG families had occurred in specific lineages. A NJ tree was constructed from functional OR protein sequences of tammar wallaby and opossum. This shows that across all ORG families (of class I and class II) there is substantial sequence homology between tammar wallaby and opossum ORGs. A roughly one-to-one relationship exists between ORGs from the two marsupials, opossum and tammar wallaby (Figure 5).

Only one opossum-specific expansion was identified in ORG family 10 (Figure 5), which arose from an expansion in only one subfamily (10AG) in opossum. All the OR10AG genes are located in one large cluster on opossum chromosome 5q (see CLIC #26 in Additional file 1), so are likely to be the products of several rounds of intra-cluster duplications of what was originally only a few genes.

**Mapping ORG clusters**

In other mammalian species, ORGs lie in clusters scattered across the genome. The location of these ORG clusters has been shown to be conserved between different mammals, including the opossum (Aloni et al. 2006). The syntenic blocks in which clusters reside show conservation of gene content and gene order. To investigate whether these gene clusters also have conserved locations in the tammar wallaby genome, we physically mapped 30 opossum gene clusters onto tammar wallaby chromosomes by FISH, and compared the locations with a predicted virtual map of the tammar wallaby conserved ORG clusters (Additional file 1). The location of the
clusters was also refined with respect to the location of flanking markers on the virtual map.

**Figure 5.** Relationship of the tammar wallaby and opossum OR repertoires. A neighbour-joining tree comparing tammar wallaby (red) and opossum (green) ORG repertoires. Numbers show the ORG families and the scale corresponds to 10% sequence divergence. Opossum adrenergic beta-receptor is used as outgroup.
A total number of 52 genomic BAC clones, representing 48 ORGs from 30 conserved mammalian ORG clusters, was localised to tammar wallaby chromosomes by FISH (Additional file 2). The locations of these ORG clusters in opossum and tammar wallaby were compared (Figure 6). The clusters are numbered according to HORDE, and syntenic blocks shared by tammar and human chromosomes are colour-coded to represent their position on human chromosomes.

The distribution of OR clusters across the tammar chromosomes is not uniform. Most of the clusters mapped to sub-telomeric or peri-centromeric regions. Tammar wallaby chromosome 2 contains the most clusters, followed by chromosome 1; no ORG cluster mapped onto chromosome 6. This distribution of clusters is similar to that observed on opossum chromosomes (Figure 6). The physical location of all but three of these ORG clusters (CLICs 9, 11, and 36) in tammar wallaby is consistent with their expected location, predicted by the location of conserved human-opossum syntenic blocks in the tammar wallaby (Additional file 1). For instance, ORG clusters 2 and 3 lie on a medial region of chromosome 2q in both marsupial species, which is homologous to part of human chromosome 1, and cluster 29 lies in a region of tammar chromosome 5q and opossum chromosome 4q that is homologous to human chromosome 11.

Clusters 11 and 36 map close to, but not in, the gene blocks conserved between the two marsupials that are expected to contain them, suggesting that FISH localisations of these clusters may be slightly misplaced, or that there have been small rearrangements. For instance, cluster 11 lies in the sub-telomeric region of tammar 4q within a region homologous to human chromosome 10, but in a pericentric region of opossum chromosome 2 homologous to human chromosome 6. However, there is a small region homologous to human chromosome 6 close to that location in tammar, and it is likely that the cluster localisation can be refined to this region. Only one ORG cluster seems to reside in a completely different genomic context in the two marsupial species. The conserved syntenic block that contains cluster 9 is located on the end of tammar chromosome 4 (homologous to human chromosome 12), but on the end of opossum chromosome 7 in a different genomic context (human chromosome 3).
Figure 6. Distribution of conserved mammalian ORG clusters. Chromosomal location of conserved mammalian ORG clusters in opossum (top) and tammar wallaby (below). Colors represent syntenic blocks shared with human chromosomes. ORG clusters are shown as black horizontal bars numbered according to CLIC numbers in HORDE. The width of these bars represents the number of genes they contain. No tammar BAC clone was detected for CLICs 1, and 25, so they are absent from the tammar map.
To determine whether the larger ORG clusters were present intact in the tammar wallaby genome, or as smaller clusters located throughout the tammar wallaby genome, two or more ORGs (based on their order in opossum counterpart clusters) were mapped for each large cluster.

We found that Cluster 23, which contains all the mammalian class I ORGs, is broken up in the tammar wallaby, and although the larger portion lies within the conserved syntenic block on 5q, some genes seem to be located on the proximal region of tammar wallaby chromosomes 2q. As these regions lie together in both opossum and human, this would seem to have resulted from a rearrangement that occurred in the tammar wallaby. Clusters 24 and 26 also seem to have undergone some rearrangement in the tammar wallaby, as some of their genes were mapped to subtelomeric 4q rather than together with their syntenic blocks on tammar wallaby chromosomes 1q and 2p, respectively.

**Preliminary analysis of ORG clusters in platypus**

Our preliminary analysis on ORGs in platypus showed that at least three of these 48 conserved clusters (CLICs 5, 26, and 27) can be traced in the platypus genome (Additional file 1). These clusters contain the same family/subfamily distribution as their counterparts in other mammalian species and reside in the same syntenic blocks.

The number and order of genes in CLIC 5 is conserved, and this cluster lies on platypus genome within the supercontig 2061 (OANA5 assembly). It is composed mostly of ORGs from family 2 in platypus, as in all mammalian species. The platypus cluster contains 10 genes, similar to opossum (12) but less than in eutherian mammals (over 50 genes). The composition of cluster in tammar wallaby is not known, because of its fragmented genome assembly. A recent burst of duplications in some ORG subfamilies of family 2 (2T, 2M, and 2L) during eutherian evolution might have shaped this difference in the number of ORGs.

Platypus CLIC 26 is flanked by genes *PTPRJ* and *P2RX3* as it is in therian mammals, and is located on Ultracontig 163 in the platypus genome (OANA5 assembly). This region is orthologous to distal 5q in opossum, which bears CLIC 26. This is a very dense cluster of ORGs (1 ORG per 10-15 Kbs) that has a very similar gene contents
in both platypus and opossum. In platypus it contains more than 80 ORGs from only 3 families; 4, 5, and 10. In opossum, however, there are more than 280 ORGs mostly from the same families, but also containing several genes transposed to the cluster from other families (8, 9, and 13). ORG families 4, 5, and 10 are also the most prevalent ORGs in CLIC 26 in human, mouse and dog.

CLIC 27, which is located 1Mb downstream from CLIC 26, is another dense cluster of more than 110 ORGs in opossum that is conserved in platypus. Despite the conservation of flanking genes, its number of ORGs in platypus is much less (12 genes). In fact, in different mammals this cluster harbors different number of ORGs but mostly belonging to ORG family 5.

Inclusion of the platypus ORG families in the comparison of ORG repertoires (Figure 4) clarifies the overall pattern of family expansion and contraction in mammals. For instance, it highlights a very large expansion of family 14 in platypus compared to therian mammals, and the very much smaller size of families 11 and 51. It also highlights expansions that appear to be features of both marsupials, compared with human and platypus; e.g. families 5, 8, 10, 13, 52.

**Discussion**

The availability of the tammar wallaby genome sequence makes it possible to discover ORGs and gene families in the major Australian branch of the marsupials, and to compare them with American marsupials, as well as with other mammals. We describe here the isolation, classification, and mapping of the major clusters of olfactory receptor sequences in tammar wallaby.

The tammar wallaby genome consists of 1,174,382 contigs covering 2,536Mb (2.2 kb per contig). Even in its incomplete state of sequencing, the current assembly of tammar wallaby genome contains 1753 ORGs, half of which are pseudogenes. The 1753 OR sequences are located on 1325 contigs (an average of 1.3 ORGs per contig). The total number of ORGs identified in tammar wallaby sequence is slightly larger than that in opossum, but the total number of tammar ORGs may have been overestimated because, even after having removed redundant sequences, two ends of a single ORG may lie on separate contigs, so counted as two separate ORGs. In the
tammar wallaby genome there are many incomplete ORG sequences located at the ends of contigs, and 45% of the ORG sequences retrieved were shorter than the smallest known vertebrate ORG (Zhang & Firestein 2002). Nevertheless, ORG families and subfamilies are very similar in size and composition between tammar wallaby and opossum (Figure 4). We therefore conclude that marsupials have a repertoire of ORs at least as large and diverse as that of placental mammals.

The tammar wallaby and opossum share a very similar ORG repertoire in terms of size, composition and sequence homology. Surprisingly, there is a one-to-one relationship between most of the tammar wallaby ORGs and their opossum ORG orthologues, as judged both by their sequence similarity and classification, and their genome context. This suggests that in both the Australian and American marsupial lineages there have been relatively few lineage-specific evolutionary events (such as gene duplications or pseudogenizations), so their ORG repertoires have remained relatively stable for the 80 million years since the Australodelphia and Ameridelphia diverged from each other (Nilsson et al. 2004).

This contrasts with the one-to-many relationships that often exist between the ORGs of any two placental mammals with a similar divergence time. For instance, one-to-many and many-to-one homology relationships are more common than one-to-one orthology between human and mouse (Godfrey et al. 2004), suggesting rapid birth-and-death evolution of ORGs in placental mammals (Nei & Rooney 2005, Niimura & Nei 2005b). During speciation any particular ORG can experience a fate different even from that of adjacent ORGs. A repeated array of duplications (mostly segmental duplications) can produce several copies of a particular ORG in one species while the same gene might remain untouched, lose activity, or accumulate mutations so fast that is no longer recognizable in another species. Duplicated genes could stay at the same location as the seeding gene, or be transposed to other locations in the genome (Sharon et al. 1999, Glusman et al. 2000b). These duplications could involve a part of, or a whole cluster of ORGs, leading to more diversification. The first outcome of this dynamic evolution is variation in the number of ORGs between species (Niimura & Nei 2007). Secondly, ORGs from two species with roughly the same size of ORG repertoire could vary extensively, as they could be the products of species-specific duplication/pseudogenization events (Gilad et al. 2003). The high level of similarity between tammar wallaby and opossum implies that such birth-and-death evolution
does not play a large role in the evolution of the ORG repertoires in marsupials, and may have been confined to particular families.

The marsupial ORG repertoires are among the largest of any sequenced mammal. For instance, humans have 802 ORGs (Malnic et al. 2004) and dogs have 1100 ORGs (Quignon et al. 2005). Rodents have similar sized OR repertoires (1400-1750) (Godfrey et al. 2004, Quignon et al. 2005), and cattle (2129) have exceptionally large repertoires (Murphy et al. 2005, Kishida et al. 2007). Among non-mammal vertebrates, the numbers of ORGs in sequenced genomes varies enormously, from 2 in elephant shark (Callorhinchus milii) to 1638 in frog (Xenopus tropicalis). There is also variation in the proportion of pseudogenes, from 12% in zebrafish to over 60% in puffer fish (for a review see Delbridge et al. 2010).

The variation in total numbers and proportion of functional ORGs between different eutherian mammals has been attributed to a combination of genomic drift and adaptive evolution to fit unique environmental niches (Nei et al. 2008, Niimura 2009, Steiger et al. 2009, Hayden et al. 2010). Unless drift plays a large part in the evolution of ORGs in marsupials, it seems likely that the large and diverse, and surprisingly stable, marsupial ORG repertoire is a result of selection for a keen sense of smell in animals that are mostly nocturnal or crepuscular, and have a two-colour vision system (Deeb et al. 2003, Croft & Eisenberg 2006). The relatively large olfactory bulb of the marsupials (Johnson 1977) is consistent with this hypothesis, and points toward a central role for the sense of smell in at least the two model marsupials, opossum and tammar wallaby.

In addition, the conservation of a large ORG repertoire in marsupials may reflect the critical importance of a keen olfactory system for the ascent of the marsupial young from the birth canal to its mother’s teat, following its birth at a very early stage of development. The well-developed olfactory epithelium of tammar wallaby newborn has been cited as evidence that the sense of smell is crucial for the young to find its mother's pouch (Hill & Hill 1955, Hughes & Hall 1984, Gemmell & Nelson 1988, Ashwell et al. 2008). The stability of the ORG repertoire in the opossum and tammar wallaby may therefore be a result of adaptation to this unique mode of reproduction of the marsupials.
In all mammals, genes coding for olfactory receptors are organized in about 100 clusters that are scattered across the genome and are located on almost all chromosomes in mammalian species (Glusman et al. 2001, Godfrey et al. 2004). These clusters vary in size from 1 to more than 150 genes in different species. Species with higher numbers of ORGs are expected to have more and bigger clusters of ORGs, as tandem duplications extend the cluster size and transpositions increase the number of clusters (Ben-Arie et al. 1994, Sosinsky et al. 2000). We identified at least 30 ORG clusters distributed around the tammar wallaby genome; nearly all located in the same conserved syntenic blocks as in other mammals. In placental mammals, most of the genes within each cluster share a high level of sequence homology, as the result of intra-cluster duplications and deletions that expand or contract the size of the ORG repertoire (Glusman et al. 2000b, Hoppe et al. 2000, Niimura & Nei 2003). It has been shown that in five mammalian species almost 90% of ORGs reside in 48 clusters that are located in conserved syntenic chromosomal blocks (Aloni et al. 2006). Our results show that of the 32 of these conserved clusters identified, only 3 clusters have been rearranged between the opossum and the tammar wallaby.

It has been proposed that most of the ORG clusters in mammals have a common ancestry, and were established before the divergence of marsupials and placental mammals at least 146 million years ago (Aloni et al. 2006). Our analysis suggests that these arrangements are far older than this. Conservation of at least three CLICs in platypus genome provides evidence that ORG amplification and formation of clusters occurred long before the divergence of monotremes from therian mammals about 166 million years ago. Future improvements in the platypus genome assembly, and localization of the entire platypus ORG repertoire could provide more evidence about the origin and potential role(s) of clustering in the evolution of ORGs.

**Conclusions**

Dynamic evolution of ORGs has resulted in substantial differences in ORG repertoires in different species of placental mammals, even those that are closely related. However, here we show that two marsupials, the tammar wallaby and opossum, share a very similar ORG repertoire, although these two species represent...
marsupial superorders that have been evolving independently for about 80 million years. Comparison with monotreme mammals shows that this number and variety is likely to have been a property of an ancestral mammal, and that there have been remarkably few changes in marsupials, perhaps because of the demands of their unique mode of reproduction.

**Authors’ Contribution**

AM carried out the ORG classification, phylogenetic analysis, FISH mapping and drafted parts of the manuscript. HRP extracted the ORGs from the genome assembly. MLD supervised the experiments and coordinated the manuscript. JAMG coordinated the research and edited the manuscript. All authors have read and approved the final manuscript.
Additional files

Additional file 1. Genomic location of CLICs in 3 species. For opossum, chromosome coordinates were extracted by matching the location of ORG clusters in HORDE (MonDom1 assembly, scaffold level) with the latest genome assembly of opossum (MonDom5, chromosome level) on the Ensemble genome browser server. Six CLICs (14, 17, 24, 25, 28, 39) were not previously located in the opossum genome. For tammar wallaby the location of CLICs were predicted on the virtual map (linkage map + physical map). The current version of platypus genome (Ornithorhynchus_anatinus 5.0) is assembled at the contig level and only 3 CLICs could be located.

<table>
<thead>
<tr>
<th>CLIC</th>
<th>Opossum</th>
<th>Tammar wallaby (virtual map)</th>
<th>Platypus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chr4: 429771811-429985979</td>
<td>Chr5: close to qtr end region: but scattered</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Chr2: 480829423-480830370</td>
<td>Chr2: 343604 (1q2 169.11)-343990 (1q2 169.30)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Chr2: 493197249-493291099</td>
<td>Chr2: 356335 (1q3 175.37)-356934 (1q3 175.67)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Chr2: 166379449-168330639</td>
<td>Chr2: 454161 (1q3 223.52)-452078 (1q3 222.49)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Chr2: 250178009-251137837</td>
<td>Chr2: 156915 (1q2 77.23)-155815 (1q2 76.69)</td>
<td>SuperConrigContig2061: 25206-78167</td>
</tr>
<tr>
<td>9</td>
<td>Chr7: 18981040-19235482</td>
<td>Chr6: 416817 (0q32 129.57)-416916 (0q32 129.6)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Chr2: 257904480-258285168</td>
<td>Chr2: 149138 (1q2 73.40)-147106 (1q2 72.40)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Chr2: 265261062-266052026</td>
<td>Chr2: 141724 (1q1 69.75)-140819 (1q1 69.31)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Chr8: 205732900-205773397</td>
<td>Chr3: 424512 (1q23 172.68)-424742 (1q23 172.78)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Chr8: 205732924-206005632</td>
<td>Chr3: 425725 (1q23 173.20)-425922 (1q23 173.01)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Chr8: 206574406-207414983</td>
<td>Chr3: 425725 (1q23 173.20)-426860 (1q23 173.68)</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Chr6: 134894285-135150300</td>
<td>Chr1: 100429 (1p1 39.03)-100105 (1p1 38.91)</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Chr6: 141329524-141968310</td>
<td>Chr1:93425 (1p1 36.31)-92294 (1p1 35.87)</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Chr1: 416181155-416681452</td>
<td>Chr1: 544713 (1q3 211.72)-545359 (1q3 211.97)</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Chr1: 16487669-1738676</td>
<td>Chr1: 360160 (1q22 139.99)-359075 (1q22 139.56)</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Chr4: 349105392-354408268</td>
<td>Chr5: 445874 (2q3 174.92)-442985 (2q2 173.78)</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Chr1: 527094269-527251006</td>
<td>Chr1:334488 (1q22 130.01)-336482 (1q22 130.78)</td>
<td></td>
</tr>
<tr>
<td>Chromosome</td>
<td>Start 1</td>
<td>End 1</td>
<td>Start 2</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>Chr5:</td>
<td>241635165-242045907</td>
<td>Chr6: 60314 (1p3 3.20)-15518 (1p3 3.94)</td>
<td>UltraContig Ultra163: 620145-1456138</td>
</tr>
<tr>
<td>Chr5:</td>
<td>285023017-288964990</td>
<td>Chr6: 59790 (1p2 18.75)-12687 (1p3 3.94)</td>
<td></td>
</tr>
<tr>
<td>Chr6:</td>
<td>16743 (1p3 3.20)-15518 (1p3 3.94)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
</tr>
<tr>
<td>Chr5:</td>
<td>285023017-288964990</td>
<td>Chr6: 59790 (1p2 18.75)-12687 (1p3 3.94)</td>
<td></td>
</tr>
<tr>
<td>Chr6:</td>
<td>16743 (1p3 3.20)-15518 (1p3 3.94)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
</tr>
<tr>
<td>Chr5:</td>
<td>286194845-288927235</td>
<td>Chr6: 15518 (1p3 3.94)-12687 (1p3 3.94)</td>
<td></td>
</tr>
<tr>
<td>Chr6:</td>
<td>15518 (1p3 3.94)-12687 (1p3 3.94)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
</tr>
</tbody>
</table>

27

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Start 1</th>
<th>End 1</th>
<th>Start 2</th>
<th>End 2</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr5:</td>
<td>286194845-288927235</td>
<td>Chr6: 15518 (1p3 3.94)-12687 (1p3 3.94)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr6:</td>
<td>15518 (1p3 3.94)-12687 (1p3 3.94)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr5:</td>
<td>288994974-291192009</td>
<td>Chr6: 12687 (1p3 3.94)-10370 (1p3 3.22)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr6:</td>
<td>12687 (1p3 3.94)-10370 (1p3 3.22)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr5:</td>
<td>292011870-293637258</td>
<td>Chr6: 9639 (1p3 3.00)-7972 (1p3 2.48)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr6:</td>
<td>9639 (1p3 3.00)-7972 (1p3 2.48)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

28

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Start 1</th>
<th>End 1</th>
<th>Start 2</th>
<th>End 2</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr4:</td>
<td>343673631-343864731</td>
<td>Chr5: in a breakpoint region between region F and G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr5:</td>
<td>240514 (1q2 14.35)-238475 (1q2 14.35)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr4:</td>
<td>240514 (1q2 14.35)-238475 (1q2 14.35)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr5:</td>
<td>294206483-294552259</td>
<td>Chr6: 12687 (1p3 3.94)-10370 (1p3 3.22)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr6:</td>
<td>12687 (1p3 3.94)-10370 (1p3 3.22)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr5:</td>
<td>294908483-295187319</td>
<td>Chr6: 7070 (1p3 2.20)-6428 (1p3 2.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr6:</td>
<td>7070 (1p3 2.20)-6428 (1p3 2.00)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

31

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Start 1</th>
<th>End 1</th>
<th>Start 2</th>
<th>End 2</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr8:</td>
<td>31469015-34003273</td>
<td>Chr3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr3:</td>
<td>358996 (1q2 14.35)-358083 (1q2 14.35)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr3:</td>
<td>358996 (1q2 14.35)-358083 (1q2 14.35)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr1:</td>
<td>166756558-168736004</td>
<td>Chr1: 166728 (1q1 64.63)-165661 (1q1 64.39)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr1:</td>
<td>166756558-168736004</td>
<td>Chr1: 166728 (1q1 64.63)-165661 (1q1 64.39)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr7:</td>
<td>206475 (1q1 64.63)-206125 (1q1 64.39)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr7:</td>
<td>206475 (1q1 64.63)-206125 (1q1 64.39)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr5:</td>
<td>270844334-271634517</td>
<td>Chr5: 358996 (1q2 14.83)-358083 (1q2 14.47)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr5:</td>
<td>358996 (1q2 14.83)-358083 (1q2 14.47)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr1:</td>
<td>171746301-171910918</td>
<td>Chr7: 201810 (1q3 108.00)-201497 (1q3 107.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr7:</td>
<td>201810 (1q3 108.00)-201497 (1q3 107.84)</td>
<td>UltraContig Ultra475: 656332-769361</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

48

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Start 1</th>
<th>End 1</th>
<th>Start 2</th>
<th>End 2</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChrX:</td>
<td>32390294-32718204</td>
<td>ChrXq</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ChrX:</td>
<td>32390294-32718204</td>
<td>ChrXq</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Additional file 2.** Orthologous ORGs in six mammalian species. These ORs represent 48 cluster/subcluster of 29 CLICs. BAC clones that were mapped for tammar wallaby representatives are also shown. Some clusters are divided into few sub-clusters and for each sub-cluster at least one BAC clone was mapped. OR names are according to HORDE.

<table>
<thead>
<tr>
<th>CLIC/Cluster</th>
<th>Human</th>
<th>Mouse</th>
<th>Rat</th>
<th>Dog</th>
<th>Opossum</th>
<th>Tammar</th>
<th>BAC for Tammar OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>OR11I1P</td>
<td>MOR122-2</td>
<td>NM_001000914</td>
<td>cOR11I3</td>
<td>OR11I3P</td>
<td>OR482</td>
<td>91D21</td>
</tr>
<tr>
<td>3</td>
<td>OR13Z2P</td>
<td>MOR217-1</td>
<td>NM_001000558</td>
<td>cOR13L2</td>
<td>OR13L4</td>
<td>OR321</td>
<td>1J13</td>
</tr>
<tr>
<td>4a</td>
<td>OR10J2P</td>
<td>MOR267-8</td>
<td>NM_001000499</td>
<td>cOR10J23</td>
<td>OR10J2</td>
<td>OR1103</td>
<td>26J8, 5711, 125M22</td>
</tr>
<tr>
<td>4b</td>
<td>OR10Z1</td>
<td>MOR267-6</td>
<td>NM_001000911</td>
<td>cOR10Z2</td>
<td>OR10Z2</td>
<td>OR1267</td>
<td>3B20, 103A10, 106D23</td>
</tr>
<tr>
<td>4c</td>
<td>OR10R2</td>
<td>MOR267-5</td>
<td>NM_001000533</td>
<td>cOR10R4</td>
<td>OR10R11</td>
<td>OR571</td>
<td>52B23, 129M13</td>
</tr>
<tr>
<td>5a</td>
<td>OR2T29</td>
<td>MOR275-1</td>
<td>NM_001000779</td>
<td>cOR2T37</td>
<td>OR2T37</td>
<td>OR1735</td>
<td>15K19, 67M20, 74G7, 88K3, 123D12</td>
</tr>
<tr>
<td>9a</td>
<td>OR5AC1P</td>
<td>MOR182-1</td>
<td>NM_001000530</td>
<td>cOR5AC3</td>
<td>OR5AC6</td>
<td>OR427</td>
<td>72A19, 103D17, 127L3</td>
</tr>
<tr>
<td>11</td>
<td>OR2V1</td>
<td>MOR276-1</td>
<td>NM_214834</td>
<td>cOR2V1</td>
<td>OR2V4</td>
<td>OR1202</td>
<td>143D19</td>
</tr>
<tr>
<td>12a</td>
<td>OR2G1P</td>
<td>MOR256-19</td>
<td>NM_001001373</td>
<td>cOR2G5</td>
<td>OR2G7</td>
<td>OR471</td>
<td>23I17, 118A21</td>
</tr>
<tr>
<td>15b</td>
<td>OR6V1</td>
<td>MOR120-1</td>
<td>NM_001001287</td>
<td>cOR6V2</td>
<td>OR6V5P</td>
<td>OR920</td>
<td>10N8, 144J23</td>
</tr>
<tr>
<td>16a</td>
<td>OR2R1P</td>
<td>MOR257-1</td>
<td>NM_001000977</td>
<td>cOR13Q4</td>
<td>OR13Q4</td>
<td>OR447</td>
<td>114I4</td>
</tr>
<tr>
<td>17a</td>
<td>OR13J1</td>
<td>MOR262-4</td>
<td>NM_001000407</td>
<td>cOR13J2</td>
<td>OR13J2</td>
<td>OR1125</td>
<td>146M7</td>
</tr>
<tr>
<td>17b</td>
<td>OR13E1P</td>
<td>MOR262-10</td>
<td>NM_001000408</td>
<td>cOR13E3</td>
<td>OR13E3</td>
<td>OR1614</td>
<td>167K7, 248H18, 249M1, 268A24</td>
</tr>
<tr>
<td>19a</td>
<td>OR13C4</td>
<td>MOR262-9</td>
<td>NM_001000398</td>
<td>cOR13C13</td>
<td>OR13D9</td>
<td>OR995</td>
<td>166F23, 210B9</td>
</tr>
<tr>
<td>21a</td>
<td>OR1K1</td>
<td>MOR138-1</td>
<td>NM_001000394</td>
<td>cOR1K2</td>
<td>OR1K2</td>
<td>OR990</td>
<td>17909, 203B8, 210E20</td>
</tr>
<tr>
<td>21b</td>
<td>OR1J2</td>
<td>MOR136-13</td>
<td>NM_001000826</td>
<td>cOR1J2</td>
<td>OR1J2</td>
<td>OR242</td>
<td>169F20, 205110, 225H18, 232G10</td>
</tr>
<tr>
<td>21c</td>
<td>OR1N1</td>
<td>MOR127-4</td>
<td>NM_001000388</td>
<td>cOR1N1</td>
<td>OR1N4</td>
<td>OR1719</td>
<td>176K7</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>----------</td>
<td>--------------</td>
<td>--------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>22b</td>
<td>OR13A1</td>
<td>MOR253-9</td>
<td>NM_001000992</td>
<td>cOR13A2</td>
<td>OR13A2</td>
<td>OR91</td>
<td>---</td>
</tr>
<tr>
<td>23a</td>
<td>OR55B1P</td>
<td>MOR42-3</td>
<td>NM_001000128</td>
<td>cOR55B4P</td>
<td>OR55B9</td>
<td>OR527</td>
<td>146p24, 148J14, 152H12, 212C22, 224B11, 252C22</td>
</tr>
<tr>
<td>23b</td>
<td>OR52N4</td>
<td>MOR34-5</td>
<td>NM_001000170</td>
<td>cOR52N8</td>
<td>OR52N4</td>
<td>OR509</td>
<td>161A23, 233H14, 239F13</td>
</tr>
<tr>
<td>23d</td>
<td>OR56A4</td>
<td>MOR40-1</td>
<td>NM_001000186</td>
<td>cOR56A4</td>
<td>OR56A4</td>
<td>OR1402</td>
<td>165O7, 208L8, 260P21</td>
</tr>
<tr>
<td>24a</td>
<td>OR6A2</td>
<td>MOR103-15</td>
<td>NM_001000553</td>
<td>cOR6A2</td>
<td>OR6A2</td>
<td>OR1320</td>
<td>29P19, 45H24, 53L1, 57J22, 102N7, 107E8, 123H7</td>
</tr>
<tr>
<td>24b</td>
<td>OR2AG1</td>
<td>MOR283-1</td>
<td>NM_001000199</td>
<td>cOR2AG7</td>
<td>OR2AG11P</td>
<td>OR653</td>
<td>26N16, 37F23, 37H22, 84P2</td>
</tr>
<tr>
<td>26-1a</td>
<td>OR4B1</td>
<td>MOR227-4</td>
<td>NM_001000616</td>
<td>cOR4B1</td>
<td>OR4B3</td>
<td>OR735</td>
<td>32C19, 50L4, 102E1</td>
</tr>
<tr>
<td>26-3a</td>
<td>OR5G1P</td>
<td>MOR175-3</td>
<td>NM_001000289</td>
<td>cOR5G9</td>
<td>OR5G1</td>
<td>OR64</td>
<td>46116, 86D11</td>
</tr>
<tr>
<td>26-3b</td>
<td>OR8K1</td>
<td>MOR194-1</td>
<td>NM_001000307</td>
<td>cOR8K1</td>
<td>OR8K1</td>
<td>OR854</td>
<td>74L3</td>
</tr>
<tr>
<td>27-1a</td>
<td>OR9I1</td>
<td>MOR211-2</td>
<td>NM_0010001286</td>
<td>cOR9I5</td>
<td>OR9I1</td>
<td>OR1058</td>
<td>62N5</td>
</tr>
<tr>
<td>27-1b</td>
<td>OR5B12</td>
<td>MOR202-10</td>
<td>NM_001000257</td>
<td>cOR5B24</td>
<td>OR5B12</td>
<td>OR1490</td>
<td>109F22, 117A15</td>
</tr>
<tr>
<td>27-2</td>
<td>OR5A2</td>
<td>MOR215-4</td>
<td>NM_001000250</td>
<td>cOR5A2</td>
<td>OR5A2</td>
<td>OR986</td>
<td>78M15, 116C7</td>
</tr>
<tr>
<td>27-3</td>
<td>OR10Q1</td>
<td>MOR266-1</td>
<td>NM_0010001289</td>
<td>cOR10Q1</td>
<td>OR10Q1</td>
<td>OR171</td>
<td>44H15</td>
</tr>
<tr>
<td>28</td>
<td>OR2AT2P</td>
<td>MOR101-2</td>
<td>NM_001000935</td>
<td>cOR2AT6</td>
<td>OR2AT6</td>
<td>OR1509</td>
<td>98L11, 106N13, 135F20, 141P9</td>
</tr>
<tr>
<td>29-2a</td>
<td>OR8G5</td>
<td>MOR171-5</td>
<td>NM_001000463</td>
<td>cOR8F2</td>
<td>OR8G2</td>
<td>OR458</td>
<td>96L1, 100E23</td>
</tr>
<tr>
<td>29-2b</td>
<td>OR8B1P</td>
<td>MOR167-3</td>
<td>NM_001000454</td>
<td>cOR8B16P</td>
<td>OR8B26</td>
<td>OR1590</td>
<td>---</td>
</tr>
<tr>
<td>31a</td>
<td>OR9K1P</td>
<td>MOR210-5</td>
<td>NM_001000062</td>
<td>cOR9K7</td>
<td>OR9K7</td>
<td>OR1738</td>
<td>14L8, 19H11, 32I17</td>
</tr>
<tr>
<td>31b</td>
<td>OR6C2</td>
<td>MOR114-1</td>
<td>NM_001000055</td>
<td>cOR6C37</td>
<td>OR6C37</td>
<td>OR356</td>
<td>180C20, 202J14, 286p13</td>
</tr>
<tr>
<td>32-1b</td>
<td>OR4L1</td>
<td>MOR247-2</td>
<td>NM_001000971</td>
<td>cOR4L1</td>
<td>OR4L1</td>
<td>OR1041</td>
<td>16 L4</td>
</tr>
<tr>
<td>32-1c</td>
<td>OR4G2P</td>
<td>MOR245-9</td>
<td>NM_001000369</td>
<td>cOR4G7</td>
<td>OR4G7</td>
<td>OR1261</td>
<td>215021, 230117, 254C19, 268F1</td>
</tr>
<tr>
<td>32-2a</td>
<td>OR11G2</td>
<td>MOR106-3</td>
<td>NM_001000839</td>
<td>cOR11G3</td>
<td>OR11G1</td>
<td>OR19</td>
<td>232013, 278M21</td>
</tr>
<tr>
<td>32-2b</td>
<td>OR11H4</td>
<td>MOR106-1</td>
<td>NM_001000835</td>
<td>cOR11H11</td>
<td>OR11H5</td>
<td>OR56</td>
<td>230117</td>
</tr>
<tr>
<td>35a</td>
<td>OR4E1P</td>
<td>MOR244-2</td>
<td>NM_001000102</td>
<td>cOR4E1</td>
<td>OR4E1</td>
<td>OR1142</td>
<td>224C8, 238N9, 247K7, 280J12</td>
</tr>
<tr>
<td>35b</td>
<td>OR10G1P</td>
<td>MOR223-6</td>
<td>NM_001000832</td>
<td>cOR10G1</td>
<td>OR10G1</td>
<td>OR248</td>
<td>224C8, 238N9, 247K7, 268G10, 280J12</td>
</tr>
<tr>
<td>36</td>
<td>OR6E1P</td>
<td>MOR254-5</td>
<td>NM_001000103</td>
<td>cOR6E1</td>
<td>OR6E4</td>
<td>OR259</td>
<td>209E19, 248K9, 283H18</td>
</tr>
<tr>
<td>38</td>
<td>OR1F1</td>
<td>MOR131-1</td>
<td>NM_214931</td>
<td>cOR1F14P</td>
<td>OR1F1</td>
<td>OR502</td>
<td>221K16</td>
</tr>
<tr>
<td>39a</td>
<td>OR1D2</td>
<td>MOR129-1</td>
<td>NM_001000714</td>
<td>cOR1D10P</td>
<td>OR1D2</td>
<td>OR1378</td>
<td>173K15, 219A5</td>
</tr>
<tr>
<td>40</td>
<td>OR4D1</td>
<td>MOR240-1</td>
<td>NM_001000713</td>
<td>cOR4D14</td>
<td>OR4D14</td>
<td>OR1437</td>
<td>256E18</td>
</tr>
<tr>
<td>45a</td>
<td>OR1I1</td>
<td>MOR12B-3</td>
<td>NM_001000418</td>
<td>cOR1I2</td>
<td>OR1I2</td>
<td>OR1091</td>
<td>154C22, 172K1, 195C15, 195C5, 234K19</td>
</tr>
<tr>
<td>46</td>
<td>OR10H1</td>
<td>MOR101-4</td>
<td>NM_001000251</td>
<td>cOR10H21</td>
<td>OR10H21</td>
<td>OR1197</td>
<td>145P11, 158B9, 192E12, 193A12, 280B8</td>
</tr>
<tr>
<td>48</td>
<td>OR13K1P</td>
<td>MOR233-2</td>
<td>NM_001000514</td>
<td>cOR13H2P</td>
<td>OR13K2</td>
<td>OR1406</td>
<td>193I4</td>
</tr>
</tbody>
</table>

**Additional file 3.** The DNA sequences of tammar wallaby olfactory receptors with their contig coordinates from the Meug_1.0 assembly. (Please find the file in attached CD-ROM)

**Additional file 4.** The protein sequences of tammar wallaby olfactory receptors with their contig coordinates from the Meug_1.0 assembly. (Please find the file in attached CD-ROM)
CHAPTER 5: THE MARSUPIAL MHC-LINKED OLFACTORY RECEPTOR GENE CLUSTER; CLUES TO THE EVOLUTION AND FUNCTION

Authors:
Mohammadi, A., Delbridge, M.L., Graves, J.A.M.

Current status of paper (circle as appropriate): Planned/ In Preparation/ Submitted/ Under Revision/ Accepted/ Published

Date paper accepted for publication or anticipated date of publication: Not Known

Name of journal/book: Mammalian Genomes

Extent to which research is my own: I conducted all the experimental and analysis work under the supervision of MLD.

My contribution to writing the paper: I wrote the draft of the manuscript and prepared figures and incorporated suggestions made by the co-authors.

If paper not yet accepted, has the paper been rejected by any journals: Not Applicable

Comments: This paper shows that the marsupial genome conservation extends to the arrangement of ORG family in one of its important clusters. Because it has a full introduction and discussion, additional background is not necessary.
The marsupial MHC-linked olfactory receptor gene cluster; clues to evolution and function.

Amir Mohammadi*, Margaret L Delbridge, Jennifer A Marshall Graves

ARC Centre of Excellence for Kangaroo Genomics, Research School of Biology, The Australian National University, Canberra 2601, Australia

*Corresponding author

Email addresses:

AM: amir.mohammadi@anu.edu.au

MLD: margaret.delbridge@anu.edu.au

JAMG: jenny.graves@anu.edu.au
Abstract

An olfactory receptor gene (ORG) cluster is located in the close vicinity of the major histocompatibility (MHC) locus in most mammalian species including marsupials and primates. Its conserved location and gene content, and proposed functional relationship with the MHC in mate choice, has attracted keen research interest. In the tammar wallaby, a model Australian marsupial, the MHC genes are, uniquely, distributed to several locations, making the tammar especially valuable for studying the relationship between a particular conserved ORG cluster and the MHC locus. Here we show that in the tammar wallaby, this ORG cluster has been split up and dispersed along with Class I MHC genes. A phylogenetic analysis and comparison of the gene content of the MHC-linked ORG cluster in seven mammalian species identifies some of these ORGs as older, more conserved, and specific to this cluster than others. Further genetic and functional investigations may demonstrate a contribution to olfaction-mediated mate choice.
Introduction

Most vertebrate species use olfaction as an important means of communication with their surrounding environments in order to acquire information for locating food, avoid predators, and enter social interactions with their kin (for a review see Knapp et al. 2006). The advantage of the sense of smell over other sensory systems is that the information can still be advertised even when the animal has left that location, and its new location can be traced.

In vertebrates, olfaction is mediated by a group of G-protein coupled receptors known as olfactory receptors (ORs) that are expressed mainly on the surface of neuroepithelial cells in the nasal cavity (Buck & Axel 1991). These OR proteins, each with seven transmembrane domains, are encoded by short, intronless olfactory receptors. The numbers of these ORGs vary extensively across vertebrates, ranging from fewer than a hundred genes in most fish to over 1500 in most mammalian species. ORGs make up the biggest gene family in mammals including marsupials (Nei & Rooney 2005, Niimura & Nei 2005a, Niimura & Nei 2007).

Like some other multigene families (e.g. MHC, and rRNA) ORGs are organized in clusters that are distributed across the genome on several chromosomes (Ben-Arie et al. 1994, Sullivan et al. 1996, Buettner et al. 1998, Strotmann et al. 1999). In mammals there are approximately 100 ORG clusters that contain from 1 to more than 100 ORGs (Glusman et al. 2001, Godfrey et al. 2004, Malnic et al. 2004, Olender et al. 2004).

ORG clusters were shaped by a series of evolutionary events, including partial or entire duplication of the original (seeding) cluster, tandem duplications of single genes, gene transpositions, and conversions (Glusman et al. 1996). Most mammalian ORG clusters are conserved across mammals (Aloni et al. 2006) and were probably established at the dawn of mammalian evolution before the divergence of monotremes (chapter 4). The chromosomal location of clusters is conserved in relation to conserved syntenic blocks, as is the gene content, gene order of the clusters, and the markers that flank them.
A major ORG cluster is located in close proximity to the major histocompatibility complex (MHC) locus in many eutherian species. This association was first discovered in human (Fan et al. 1995), and was later shown to be conserved in other mammalian species including mouse (Younger et al. 2001), dog, pig (Ando et al. 2005), and opossum (Aloni et al. 2006).

This cluster has assumed special interest in genomic studies because of its proposed role in MHC-mediated mate choice (Ziegler et al. 2002). Evidence is accumulating that in many mammalian species, individuals prefer to mate with animals having MHC haplotypes most dissimilar to their own (for a review see Penn 2002). This will increase the chance of offspring survival due to the fitness advantage of heterozygosity.

The importance of haplotypes that include particular MHC and ORG alleles was suggested by the observation that this region has the highest linkage disequilibrium in the human genome (Gyllensten & Erlich 1993). It was suggested that MHC molecules could, directly or indirectly, contribute to producing unique personal odor fingerprints. Small fragments of MHC molecules found in rodent urine have been shown to stimulate the sensory neurons within the vomeronasal organ (Leinders-Zufall et al. 2004). Peptides left over from the metabolism of MHC molecules could also produce unique odors. It is also possible that there is an indirect contribution of MHC molecules toward the uniqueness of odor by influencing the microbial flora in an animal’s body (Knapp et al. 2006).

If this cluster has a conserved function when acting together with the MHC locus, it would be expected to retain a conserved association throughout the class Mammalia. The MHC-linked ORG cluster has been studied in detail in human and mouse, revealing some conserved features (Younger et al. 2001). We now further test this hypothesis by extending the comparison to a much more distantly related mammal.

To do this, we made a BAC contig of the cluster in an Australian marsupial, the tammar wallaby, and compared it with the MHC-linked ORG clusters of eutherian mammals. The tammar wallaby was a particularly interesting model for this study, as in this species, uniquely, the MHC locus is split into clusters at several locations in the genome (Deakin et al. 2007). In most mammalian species, including the
opossum, the three classes of MHC genes are located together on a single chromosome (Belov et al. 2006). In tammar wallaby, however, the major MHC locus is located in the same syntenic block as in other mammals, but the class I MHC locus is split and distributed to several locations in the genome. We asked whether the associated ORGs were also split to the same locations.

In addition, we analysed the phylogenetic relationships between the ORGs in the cluster, and compared the overall gene content of the cluster in seven mammalian species, to search for clues to a possible functional link with MHC loci. Our comparisons identified some of the ORGs in this cluster as older, more widespread, more conserved and more specific to this cluster. These more conserved ORGs could be appropriate candidates for further genetic and behavioral studies to uncover a functional link between the ORG cluster and the MHC locus in all mammals.

**Materials and Methods**

**Probes**

Overlapping oligonucleotide (Overgo) probes were used to isolate BAC clones as previously described (Deakin et al. 2008). To screen for BACs containing unique genes that flank the ORG cluster in tammar wallaby, overgos were designed from their exonic sequences. The specific overgos to screen for ORGs were extracted from the sequence immediately adjacent to the coding sequence in the corresponding contig. If BAC end sequences contained sufficient unique sequence to produce specific probes, the overgos were derived directly from their sequences, otherwise the BAC end sequences were extended in BLAST searches against the tammar wallaby trace archives until non-repetitive matches were found, and probes were designed from these sequences. Table 1 lists the probes that were used in this study.

**BAC end sequencing**

Positive BAC clones isolated by screening were end-sequenced to provide probes for isolating the overlapping BAC clones. DNA was extracted from the BAC clones as described (Mohammadi et al. 2009). For each BAC clone, 2-4 extractions were pooled and the DNA was precipitated with ethanol (Sambrook & Russell 2001) and
Table 1. BAC clones isolated for the MHC-linked ORG cluster and flanking genes in tammar wallaby. These BACs were isolated either by using specific overgo probes from tammar wallaby sequences (overgo) or by using BAC end sequences of the flanking BACs as probes (BES).

<table>
<thead>
<tr>
<th>AGI BAC</th>
<th>Gene</th>
<th>Isolation protocol</th>
<th>AGI BAC</th>
<th>Gene</th>
<th>Isolation protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>299M18</td>
<td>WDR45L</td>
<td>overgo</td>
<td>447J2</td>
<td>OR471, OR1415</td>
<td>overgo</td>
</tr>
<tr>
<td>395E21</td>
<td>WDR45L, NARF</td>
<td>overgo</td>
<td>73I17</td>
<td>OR293</td>
<td>BES</td>
</tr>
<tr>
<td>384O16</td>
<td>NARF, HEXDC</td>
<td>overgo</td>
<td>3J5</td>
<td>OR293</td>
<td>BES</td>
</tr>
<tr>
<td>428D3</td>
<td>NARF, HEXDC</td>
<td>overgo</td>
<td>433A16</td>
<td>OR293, OR853</td>
<td>overgo</td>
</tr>
<tr>
<td>236J12</td>
<td>HEXDC, Q9H6J2</td>
<td>overgo</td>
<td>399A3</td>
<td>OR293</td>
<td>overgo</td>
</tr>
<tr>
<td>378J17</td>
<td>HEXDC, Q9H6J2</td>
<td>overgo</td>
<td>4P3</td>
<td>OR1415</td>
<td>BES</td>
</tr>
<tr>
<td>383J15</td>
<td>HEXDC, Q9H6J2</td>
<td>overgo</td>
<td>397N21</td>
<td>OR293, OR1415</td>
<td>BES</td>
</tr>
<tr>
<td>71I3</td>
<td>OR884</td>
<td>overgo</td>
<td>98C18</td>
<td>OR293, OR853</td>
<td>overgo</td>
</tr>
<tr>
<td>247B6</td>
<td>OR884</td>
<td>overgo</td>
<td>545E10</td>
<td>OR293</td>
<td>overgo</td>
</tr>
<tr>
<td>268K5</td>
<td>OR884</td>
<td>overgo</td>
<td>361A22</td>
<td>OR853, OR1147</td>
<td>overgo</td>
</tr>
<tr>
<td>33M21</td>
<td>OR884</td>
<td>overgo</td>
<td>129A23</td>
<td>OR853, OR1147</td>
<td>overgo</td>
</tr>
<tr>
<td>204N14</td>
<td>OR884</td>
<td>overgo</td>
<td>545E17</td>
<td>OR853</td>
<td>overgo</td>
</tr>
<tr>
<td>206F12</td>
<td>OR884</td>
<td>overgo</td>
<td>170I23</td>
<td>OR293, OR1147</td>
<td>overgo</td>
</tr>
<tr>
<td>248F2</td>
<td>OR884</td>
<td>overgo</td>
<td>156E3</td>
<td>OR1147</td>
<td>overgo</td>
</tr>
<tr>
<td>226O20</td>
<td>OR884</td>
<td>overgo</td>
<td>118K24</td>
<td>OR237, OR82</td>
<td>overgo</td>
</tr>
<tr>
<td>196C13</td>
<td>OR884</td>
<td>overgo</td>
<td>31M5</td>
<td>OR237, OR82</td>
<td>overgo</td>
</tr>
<tr>
<td>74M23</td>
<td>OR884, OR26</td>
<td>overgo</td>
<td>101K8</td>
<td>OR237, OR82</td>
<td>overgo</td>
</tr>
<tr>
<td>146J21</td>
<td>OR884, OR26</td>
<td>overgo</td>
<td>136I1</td>
<td>OR237, OR82</td>
<td>overgo</td>
</tr>
<tr>
<td>142F24</td>
<td>OR26</td>
<td>overgo</td>
<td>107E13</td>
<td>OR4, OR241</td>
<td>overgo</td>
</tr>
<tr>
<td>22E16</td>
<td>OR454</td>
<td>overgo</td>
<td>66N17</td>
<td>OR4, OR241</td>
<td>overgo</td>
</tr>
<tr>
<td>155G5</td>
<td>OR779, OR428</td>
<td>overgo</td>
<td>255C9</td>
<td>OR4, OR241</td>
<td>overgo</td>
</tr>
<tr>
<td>280C18</td>
<td>OR779, OR428</td>
<td>overgo</td>
<td>33B21</td>
<td>OR4, OR241</td>
<td>overgo</td>
</tr>
<tr>
<td>208P15</td>
<td>OR779, OR428</td>
<td>overgo</td>
<td>47E14</td>
<td>OR4, OR241</td>
<td>overgo</td>
</tr>
<tr>
<td>256N13</td>
<td>OR779, OR428</td>
<td>overgo</td>
<td>283O22</td>
<td>UBD, GABBR1</td>
<td>overgo</td>
</tr>
<tr>
<td>245P11</td>
<td>OR779, OR428</td>
<td>overgo</td>
<td>535F16</td>
<td>GABBR1, MOG</td>
<td>overgo</td>
</tr>
<tr>
<td>143E4</td>
<td>OR332</td>
<td>BES</td>
<td>564M21</td>
<td>GABBR1, MOG, ZNDR1</td>
<td>overgo</td>
</tr>
<tr>
<td>140O14</td>
<td>OR332</td>
<td>BES</td>
<td>451N23</td>
<td>ZNDR1</td>
<td>overgo</td>
</tr>
<tr>
<td>555C2</td>
<td>OR332</td>
<td>overgo</td>
<td>283O22</td>
<td>ZNDR1</td>
<td>overgo</td>
</tr>
<tr>
<td>432H3</td>
<td>OR332, OR471</td>
<td>BES</td>
<td>102M8</td>
<td>OR977</td>
<td>overgo</td>
</tr>
<tr>
<td>432N3</td>
<td>OR332, OR471</td>
<td>BES</td>
<td>192C17</td>
<td>OR977, OR541</td>
<td>overgo</td>
</tr>
<tr>
<td>69P14</td>
<td>OR471, OR1415</td>
<td>BES</td>
<td>193E7</td>
<td>OR977, OR541</td>
<td>overgo</td>
</tr>
<tr>
<td>427C21</td>
<td>OR471, OR1415, OR293</td>
<td>BES</td>
<td>274O11</td>
<td>OR1129, OR1507</td>
<td>overgo</td>
</tr>
</tbody>
</table>
washed with 70% ethanol. DNA pellets were resuspended in Milli-Q water to give DNA concentration greater than 120 ng/µl. Sequencing reactions were set up with BDTv3.1, primers T7 and SP6 according to (Poulsen & Johnson 2004). PCR products were ethanol precipitated, and ice-cold 70% ethanol was used for washing. Dried pellets provided DNA for sequencing at the Australian Genome Research Facility (St Lucia QLD, Australia). Raw sequences were trimmed for primer sequences in SEQUENCHER v4.8. Sequences of both strands from each end were aligned to produce a high quality sequence for either end of each BAC clone.

**BAC library screening and Fluorescence in situ hybridization (FISH)**

Each overgo was used to isolate specific BACs from a male tammar wallaby genomic BAC library obtained from Arizona Genomics Institute (Tucson, USA) using the method described by (Deakin et al. 2008). For platypus, BACs positive for these sequences were ordered from Children's Hospital Oakland Research Institute (CHORI at http://bacepac.chori.org/order_clones.php) (Table 2).

FISH was used to map positive BAC clones onto tammar wallaby chromosome spreads (Deakin et al. 2008). Briefly, BAC DNA was purified (Promega Wizard Plus SV Miniprep DNA Purification system) and labelled in nick translation reactions with SpectrumOrange-dUTP (Vysis) or SpectrumGreen-dUTP (Abbott Molecular Inc., IL, USA). The labelled BAC DNA was hybridized overnight to metaphase chromosome preparations at 37°C. Images were captured using a Zeiss Axioplan2 epifluorescence microscope and thermoelectronically cooled charge-coupled device camera (RT Monochrome Spot, Diagnostic Instruments, Sterling Heights, MI, USA). The numbers and positions of signals were recorded for 8–12 metaphases for each BAC localization. After fluorescence microscopy, images were captured and enhanced using IPLab software (Scanalytics Inc., Fairfax, VA, USA).

**BAC DNA fingerprinting**

Restriction enzyme fingerprinting of BAC clones was carried out by the Genome Science Centre (GSC) at the British Columbia Cancer Agency (Vancouver, Canada), using agar stabs of BAC clones. Briefly BAC DNAs, extracted from overnight incubations, were digested with two restriction enzymes (EcoRI and HindIII). EcoRI
and HindIII doubly digested DNA fragments were assembled into contigs using automated assembly pipeline tools at GSC. The results were viewable in Java-based Internet Contig Explorer (iCE v3.5), which is available for download from http://www.bcgsc.ca/platform/bioinfo/software/ice.

**Table 2.** BAC clones from CH236 library (CHORI) identified for probable MHC-linked ORGs in platypus. The contig address is based on OANA5 assembly

<table>
<thead>
<tr>
<th>BAC</th>
<th>Gene</th>
<th>Contig</th>
</tr>
</thead>
<tbody>
<tr>
<td>81C21</td>
<td>HEXDC, NARF, WDR45L</td>
<td>UltraContig Ultra428</td>
</tr>
<tr>
<td>45E24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21H15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>326N10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25M22</td>
<td>OR2G20p</td>
<td>SuperContig Contig11125: 6,734-22,653</td>
</tr>
<tr>
<td>360G13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>136M22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>306M2</td>
<td>OR2I2p (UBD)</td>
<td>SuperContig Contig12558: 3,188-18,677</td>
</tr>
<tr>
<td>351G21</td>
<td>OR12D11</td>
<td>SuperContig Contig32198: 45-12,957</td>
</tr>
<tr>
<td>170I20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>112J20</td>
<td>OR2G17</td>
<td>SuperContig Contig6280: 58-23,209</td>
</tr>
<tr>
<td>572K12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>391I10</td>
<td>OR2G18</td>
<td>SuperContig Contig16915: 58-15,489</td>
</tr>
<tr>
<td>142D6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75A3</td>
<td>OR2G14p</td>
<td>SuperContig Contig35889: 58-7,955</td>
</tr>
<tr>
<td>48H15</td>
<td>OR2W20p</td>
<td>SuperContig Contig21616: 48-9,974</td>
</tr>
<tr>
<td>454P3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72H22</td>
<td>OR2H14, OR2B26, OR2B22, OR2B24, OR2G22, OR2W24</td>
<td>UltraContig Ultra113: 1,071-155,796</td>
</tr>
<tr>
<td>216N13</td>
<td>Butyrophilin</td>
<td></td>
</tr>
</tbody>
</table>

**Phylogenetic analysis**

The protein sequences of ORGs in the MHC-linked ORG clusters of human, dog, platypus and opossum were obtained from Human Olfactory Repertoire Data Exploratorium (HORDE: http://genome.weizmann.ac.il/horde/). Mouse and rat OR
sequences were collected from the National Centre for Biotechnology Information (NCBI). The tammar wallaby ORG sequences were previously made available (chapter 4). OR protein sequences were aligned using the ClustalX 2.1 alignment program. Phylogenetic analysis was performed using standard parameters in the Geneious 4.8.2 package (Drummond et al. 2009). Neighbor-joining (NJ) trees were rooted with the opossum OR52B1. Tree representations were visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Results

To determine whether the content, sequence and organization of ORGS relative to MHC genes was conserved between marsupials and eutherians, consistent with a functional link, we made a detailed contig of the MHC-linked ORG cluster in the tammar wallaby and performed analysis of its content and arrangement. This was achieved by isolating BAC clones that covered the entire ORG cluster and assembling them into contigs by restriction fragment fingerprinting. The gene contig of the BAC cluster was then confirmed by PCR to detect individual ORGs, and homologies were examined between different MHC-linked ORG clusters by phylogenetic analysis.

MHC-linked ORG cluster in tammar wallaby

In order to isolate overlapping tammar wallaby BAC clones representing the MHC-linked ORG cluster in the tammar wallaby, the tammar bacterial artificial chromosome (BAC) library was screened for BACs containing the locus, and a contig was constructed from relevant BACs by restriction enzyme fingerprinting (Figure 1). The BACs were isolated from a male tammar wallaby genome library by two different approaches. Before the availability of the entire tammar wallaby OR repertoire, BAC clones covering the cluster were physically isolated stepwise from the BAC library by a BAC walking protocol (Kubat 2007). After the first genome assembly of tammar wallaby became available, the entire OR repertoire was extracted (chapter 4) and each of the ORGs that were possibly from the tammar wallaby MHC-linked ORG cluster were identified by homology search. Specific BACs for each of these ORGs were then isolated from the tammar wallaby BAC library. All BACs
identified by the two approaches were mapped to the predicted location on chromosome 2q before being used to construct the BAC contig by BAC fingerprinting.

The major MHC locus has previously been shown to lie on the pericentromeric region of chromosome 2q in the tammar wallaby. This region is part of a conserved syntenic block whose ortholog lies on the proximal region of chromosome 2q in opossum (Siddle et al. 2006). The MHC-linked ORG cluster in the opossum is located proximal to the MHC loci and is flanked by *HEXDC* and *GABBR1*. These two genes flank the MHC-linked ORG cluster in most mammalian species, so were chosen as seeders of the first round of BAC walking. Overgo probes were designed from the exons of the seeder genes, and the BAC library was screened for overlapping BACs. Their location in the genome was confirmed by FISH. If the BACs mapped to the same location as the seeder genes, then their 3’ and 5’ ends were sequenced and were used as new seeders for the next round of walking. In total 82 positive BACs were extracted by this approach.

In the second approach, all opossum MHC-linked ORGs were used as query in a BLAST search against the entire tammar wallaby genome, and the first five matches for each opossum ORG were collected. These tammar sequences were considered to be the possible orthologs for opossum ORGs. Overgo probes were designed from non-coding part of the contigs containing the ORG as described earlier (chapters 2 and 4). Relevant BACs for these probes were extracted from the library and BACs that mapped to the correct location were added to the BAC pool generated by the first approach for BAC fingerprinting, making a total of 82 BACs.

**BAC DNA fingerprinting and contig assembly**

A restriction enzyme fingerprint map of the 82 positive BACs was constructed by British Columbia Cancer Agency (Vancouver, V5Z 4E6, Canada). The restriction endonucleases *EcoRI* and *HindIII* were used to digest the BAC DNA, and DNA fragments were separated and sized by gel electrophoresis. The initial fingerprinting protocol assembled seven small contigs that included 64 BACs. In order to link these contigs together, the end sequences of BACs from both outermost ends of the contigs were used to identify the overlapping BAC contigs. This approach resulted in a single
large contig of BAC clones covering the tammar MHC-linked ORG cluster. This contig contained 17 ORGs and 8 flanking genes with homology to the MHC-linked ORG cluster in other species, as determined by gene order and sequence homologies. This ORG cluster we refer to here as the major MHC-ORG cluster (Figure 1). Thus, most of the MHC-ORGs make up one relatively large cluster close to the major MHC locus in the tammar wallaby, which is homologous to a similarly located cluster close to the MHC locus in the opossum.

Figure 1. BAC contigs of MHC-linked ORG cluster of tammar wallaby. Identity of each BAC from left to right (major cluster, minor 1 and minor 2 clusters) is represented in Table 1. The BACs length and their overlap are just rough estimations and should be treated with caution.

Two additional ORG-containing BAC contigs were identified using the restriction fragment fingerprinting protocol, and were found to map at sites other than the Major MHC cluster on chromosome 2q. One contig, containing two overlapping BACs, mapped to the telomeric region of chromosome 5q. We refer to this ORG cluster here as the Minor MHC-ORG cluster 1. Both of these BACs hybridized with overgo probes for three functional ORGs from the ORG subfamily 2G, which form part of
the MHC-linked ORG cluster in opossum, human, mouse and rat. This shows that these three ORGs are located at average intervals 20kb apart, a density similar to that of other ORGs in the same cluster in other mammalian genomes.

The second unlinked BAC contig, containing two BAC clones, was located by FISH to the distal region of tammar chromosome 1q. We refer to this cluster here as the minor MHC-ORG cluster 2. Hybridisation to specific overgo probes indicated that the minor cluster 2 contains two pseudogenes from the ORG subfamilies 2N and 2W. Their possible orthologs are located on the HEXDC side of the cluster in opossum.

The depth of the BAC coverage varies across the major MHC-ORG cluster. The region between OR332 and OR1147 has the deepest coverage and the region between OR237 and OR241 has the lowest coverage.

**MHC-linked ORG cluster in platypus**

In order to reveal the origin and understand the evolutionary path of the MHC-associated ORG cluster, we examined whether this association also exists in the genome of platypus, as we had already shown that the mammalian ORG clusters were established before the divergence of monotremes from the therian mammals (chapter 4).

The current assembly of platypus genome (OANA5) is still at the contig level, so it was impossible to track the whole of the MHC-linked ORG cluster in the platypus. Hence we used phylogenetic analysis to find any ORG that might be a platypus ortholog of MHC-associated ORGs of other mammals. All the opossum and human ORGs from the MHC-associated ORG cluster were therefore searched against the entire platypus OR repertoire for possible orthologs with more than 60% amino acid similarity. A total of 13 platypus ORGs were detected with this degree of homology. The contigs containing these ORGs were checked for any other ORG-like sequences to reveal additional ORGs in the clusters, but none were detected.

**Genomic structure of MHC-linked ORG clusters in different mammals**

In order to understand the evolutionary history of the MHC-linked ORG cluster, the genomic arrangements of the MHC locus and the MHC-linked ORG cluster was
compared in seven mammalian species. Table 3 summarizes the location and gene content of this gene block in several mammals. The same ORG cluster can be tracked in these mammalian species using its flanking gene markers. The genes UBD, GABBR1, and MOG lie on the MHC side of the ORG cluster, and the genes HEXDC, NARF, and WDR45L flank the cluster on the other side. Of these seven mammalian species, all but one have the cluster of ORGs juxtaposed with the MHC locus.

Table 3. The location and content of MHC-linked ORG cluster in mammals.

<table>
<thead>
<tr>
<th>Species</th>
<th>Location</th>
<th># of ORGs</th>
<th>Cluster starts at (bp)</th>
<th>Cluster ends at (bp)</th>
<th>MHC loci at (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opossum</td>
<td>2q</td>
<td>31</td>
<td>OR2AZ1 ENSMODG00000007082 (264,904,049)</td>
<td>OR2H2p ENSMODG0000016043 (266,227,385)</td>
<td>Upstream (1.6)</td>
</tr>
<tr>
<td>Mouse</td>
<td>17qB1</td>
<td>45</td>
<td>Olfr90 ENSMUSG000000056600 (37,222,176)</td>
<td>Olfr136 ENSMUSG00000062695 (38,473,042)</td>
<td>Downstream (1.3)</td>
</tr>
<tr>
<td></td>
<td>13qA3</td>
<td>12</td>
<td>Olfr1370 ENSMUSG00000042869 (21,164,218)</td>
<td>Olfr1359 ENSMUSG00000057727 (21,795,813)</td>
<td>--</td>
</tr>
<tr>
<td>Rat</td>
<td>20p2</td>
<td>56</td>
<td>Olr1668 ENSRN0G000000029622 (92,810)</td>
<td>Olr1750 ENSRN0G00000043481 (1,543,758)</td>
<td>Upstream (2.1)</td>
</tr>
<tr>
<td></td>
<td>17q11</td>
<td>9</td>
<td>Olr1654 ENSRN0G000000018180 (50,447,831)</td>
<td>Olr1664 ENSRN0G00000033684 (51,405,296)</td>
<td>--</td>
</tr>
<tr>
<td>Dog</td>
<td>35</td>
<td>11</td>
<td>OR2B2 ENSCAF0G00000011444 (28,097,985)</td>
<td>OR2H9p ENSCAF0G0000012227 (29,444,011)</td>
<td>Chromosome 12</td>
</tr>
<tr>
<td>Human</td>
<td>6p22.1</td>
<td>34</td>
<td>OR2B2 ENSG00000168131 (27,879,024)</td>
<td>OR2H2 ENSG00000204657 (29,556,745)</td>
<td>Upstream (0.14)</td>
</tr>
</tbody>
</table>

Among the other six species of eutherian mammals examined, the relative order of three MHC classes with respect to the ORG cluster is consistent. The MHC class I genes are always nearest to the ORG cluster, followed by MHC class III genes and then genes from MHC class II. However, the distance between the last ORG of the cluster and the first MHC gene is very inconsistent, ranging from 135kb (human) to more than 2Mb (rat). In contrast to the MHC I-III-II arrangement in eutherians, the
MHC class III genes are nearest to the ORG cluster in tammar and opossum (see figure 2), suggesting a rearrangement between MHC and ORGs in one or other lineage. Given that the arrangement of MHC genes in marsupials is the same as in frogs and birds (Belov et al. 2006), it is likely that a rearrangement occurred in the eutherian lineage.

![Figure 2](image-url)

**Figure 2.** Physical link between MHC loci and MHC-linked ORG cluster in mammals. For simplicity, only the major group of each MHC class genes are shown. Numbers denoting the chromosome number in each karyotype.

**Evolution of the subfamilies of the MHC-linked ORG cluster**

Following identification of the MHC-linked ORG cluster in different mammalian species, all of the ORGs of the cluster were compared between seven mammalian
species to examine their phylogenetic relationships (Figure 3). A phylogenetic tree was constructed after 500 rounds of bootstrapping from a multiple alignment of all functional and pseudogenes across a region between their second and seven transmembrane domains. Pseudogenes shorter than this were corrected for their missing amino acids according to their best functional match in the same species. Only a few pseudogenes that were exceptionally longer or shorter in the analysed region were removed from the analysis.

There are six ORG families making up this cluster in mammals. Comparing these with each other, and with the rest of ORGs from the same family in the genome of each mammalian species, revealed that not all of these ORGs are conserved or equally widespread. The MHC-associated ORG cluster in mammalian species can be divided into three regions (Figure 4) based largely on the gene content (ORG families and subfamilies). This provides clues to the age and importance of each region.

Region I is composed mostly of family OR2, and has a different size and gene numbers in every species. This is one of the oldest, largest and most diverse (with more subfamilies) of the mammalian ORG families. It has dozens of subfamilies, of which at least ten have representatives in this cluster. It is possible that the cluster originated with a few genes from the OR2 family, and other families and subfamilies were added later at the location. Some of the OR2 subfamilies in this cluster are very old but not specific to this cluster (2B, 2W), some are old and specific to this cluster (2I, 2H). Some are quite new (2N), particularly those that are eutherian-specific (2AD). Most of these OR2 genes are functional in every species (except for human) suggesting importance in olfaction. Some aspects of gene arrangement are also conserved in ORG Region I in all mammals; for instance, in all eutherians, the region starts with OR2B and OR2W genes, followed by one OR1F.

Subfamily 1F is present in the genome of all mammalian species including platypus and human. This region contains few ORGs (only 2 - 3 genes in any of these species). Only human, mouse, rat, and dog have an OR1F in the MHC-linked ORG cluster so it may have been added in the eutherian lineage after the divergence of marsupials. This gene is functional in most eutherians, suggesting that the addition of OR1F might have some important functional implications in eutherians. Region II seems to be rodent-specific. In human it is very small and contains two OR5 genes. However, in
**Figure 3.** Phylogenetic relationship of MHC-linked ORGs in seven mammal species. Tree was constructed from a multiple alignment of 218 ORGs (human 33, rat 57, mouse 54, dog 10, opossum 28, tammar wallaby 22, and platypus 14) after 500 rounds of bootstraps. Genes are grouped based on their subfamily allocations and bootstrap values above 80% are shown. The opossum OR52B1 genes was used as outgroup. Each branch is color-coded according to the species.

In rodents it has expanded as the result of several duplications of the OR5U and OR10AL genes. ORG5 subfamilies 5V and 5U are found only in the MHC-associated ORG cluster, suggesting that they were both located in this cluster in the common ancestor of human and rodents.

**Figure 4.** The gene content and gene order of MHC-linked ORG cluster in six mammalian species. ORG subfamilies are shown on top of the genes. All ORGs of family 1 and 12 are from subfamilies 1F and 12D respectively.
Subfamily 1AL of the rodent specific ORG Region II is present only in the mouse and rat genomes, and all the OR1AL genes in these two species are located in this cluster. The protein homology for all family 1 ORGs, especially between the 1F and 1AL protein sequences, is quite high (>50% on average), suggesting that these genes were created from a duplication of the 1F gene in this cluster in the rodent ancestor. The pattern of duplications, however, seems quite different between mouse and rat, but it is unlikely that generation of 1AL occurred independently, as the same genes in both species flank the duplicates.

It is possible that the first 10AL gene was generated in marsupials from a duplication of the 10C gene, as there is high homology between any 10AL and the only 10C gene in this cluster. There must therefore have been a subsequent duplication in the rodent lineage. The activity of OR10AL, and the low number of pseudogenes in this subfamily in mouse and rat, indicates that the OR10AL genes must be very important in rodents.

Region III is probably the oldest part of the cluster, and the most important one in terms of its association with MHC loci. Its structure is very conserved in marsupials and all eutherians except dog. This section includes ORGs from at least four families. OR12 is a small family in all mammals, ranging only from 1-12 genes in different species. It is also a very conserved family, as there are only one or two OR12 subfamilies in all mammalian species. Most of these genes belong to subfamily 12D, and all seven species (except dog) have at least one functional OR12D gene in this cluster. Most OR12D genes are located in this cluster; only mouse and rat each have another single OR12D gene elsewhere in their genomes. OR12D is probably one of the oldest ORGs in this cluster, dating back to before the divergence of monotremes from therians.

The ORG11 family appears to have been inserted into the MHC-associated ORG cluster after the divergence of the monotremes from the eutherians, as the two subfamilies 11W and 11A are not detected in the platypus genome. Most therian species have an 11A gene, except the dog, suggesting that OR11A was deleted from dog genome secondarily. OR11A seems to be the original gene, and duplication in the marsupial lineage generated the marsupial specific OR11W gene. Homology between
OR11A1 and OR11W1 in opossum is low (43%), suggesting that they diverged early in the marsupial lineage.

One highly conserved OR10C is present in the MHC-linked ORG cluster of all seven mammalian species in exactly the same orientation, and in the same location with respect to the flanking genes. It seems to be one of the oldest genes in the MHC-linked ORG cluster, dating from before the divergence of marsupials and eutherians. There are no other genes from this subfamily elsewhere in the any of these genomes.

**Discussion**

All mammalian species studied so far, possess a cluster of ORGs that is located very close to the MHC locus. In this study we show that in tammar wallaby also there is a cluster of ORGs in the vicinity of the main MHC locus on chromosome 2q. Uniquely, MHC class I genes are dispersed in tammar wallaby on several chromosomes, and as we report here so is the MHC-linked ORG cluster. However, most of the MHC genes and MHC-linked ORGs are still located together in a gene block that is conserved in mammals. We also phylogenetically compared the MHC-linked ORGs from seven different mammals and showed that some of these ORGs do not share the same conservation, and hence are less likely to play an important role in a functional correlation with MHC haplotypes.

**The MHC-linked ORG cluster of tammar wallaby**

As in other mammalian genomes, there is an ORG cluster in close proximity to the major MHC locus, which lies on the pericentromeric region of chromosome 2q in the tammar wallaby. In most mammalian species, the MHC genes (class I, II, and III) are represented as one big interrupted cluster in a conserved gene frame. In the tammar wallaby, however, the MHC class I cluster is split, and Class I genes are dispersed on almost all chromosomes.

We showed here that the some of the ORGs present in the MHC-linked ORG cluster in other mammals (including the distantly related opossum) have split from the major MHC-associated ORG cluster in tammar wallaby and have been moved, with MHC Class I genes, to two other chromosomes. For instance, some minor ORG clusters
have been mapped to roughly the same location as minor MHC clusters 1 and 2 (Siddle et al. 2009). Their proximity to the relocated MHC genes suggests that they were moved or copied together with MHC genes.

In rodents the entire MHC locus is remains intact, and is linked to the major ORG cluster. However, the MHC-linked ORG cluster has been broken up and a minor cluster has been moved to another location in the genome (on the same syntenic block of mouse and rat genomes).

Drawing any direct functional link between the relocated MHC genes and the two minor ORG clusters at this stage is impossible, as assembled sequence and a fine map of the regions containing these genes is not yet available. It is also possible that there are more ORGs in the major cluster that were not detected in this study. Again, fine mapping all the ORGs in tammar wallaby genome would provide more information to investigate the functional relationship between ORGs and the MHC loci.

**Functional relationship of ORG cluster and MHC loci in mammals**

Establishing a functional link between MHC haplotypes and associated ORGs in this cluster is not an easy task. Not all the ORGs in this cluster seem to share the same degree of conservation and activity, and therefore are likely to have different functional associations with MHC genes. MHC genes are very polymorphic; for example, more than 550 alleles have been identified in the human MHC-B locus (equivalent to MHC class I in other mammals) (Robinson et al. 2003). If different MHC haplotypes are the source of different odors, then the first requirement for any ORG to be involved in the olfaction-mediated mate choice is to be polymorphic to at least the same level as the MHC genes.

ORGs are, indeed, shown to be polymorphic (Ehlers et al. 2000, Eklund et al. 2000). However this level of polymorphism seems not to be high enough (only 10-15 alleles for some ORGs) to match MHC haplotype diversity, and contribute toward a variation in odor preferences. Apart from this, the strong linkage disequilibrium between ORGs and HLA haplotypes is in doubt. Some studies have shown that measures of linkage disequilibrium between the human ORG and HLA loci are negligible, and for some ORGs (like OR2H2), alleles are not in linkage
disequilibrium with HLA variation (Eklund et al. 2000) or do not explain findings of HLA-associated odor preference (Thompson et al. 2010).

If MHC-linked ORGs are functionally involved with MHC haplotypes in mate choice, it is expected that, compared to other ORGs of the genome, these ORGs should exhibit a distinct pattern of expression. Microarray analyses for ORG expression in human (De la Cruz et al. 2009) and mouse, however, showed that, compared to other ORGs from the same family, MHC-linked ORGs show no different patterns in their levels or locations (ectopic or conserved) of expression (Zhang et al. 2007).

Among the more than 30 different subfamilies of ORGs included in the mammalian MHC-linked ORG cluster, we have identified very old subfamilies, most of which are specific to only this cluster and are functional in almost all of the seven mammals we compared. These might be the best candidates for future studies. OR2I, OR2H, OR12D, and OR10C could be placed at the top of the priority list, as they are conserved in all mammalian species studied here. In addition, there may be medical interest in these candidates, as OR10C1 and OR5V1 have been associated with asthma (Nicolae et al. 2005) and single nucleotide polymorphisms (SNPs) within this OR12D gene are associated with asthma and the differential recognition of cigarette smoke components in Caucasian women (Santos et al. 2008)

**Conclusion**

The MHC-linked ORG cluster is one of the best-conserved olfactory gene clusters in mammals, and is thought to have a role, in association with MHC alleles, in mate choice. In tammar wallaby, the MHC locus is split, with Class I genes removed to other chromosome locations. We found that most ORGs of the tammar MHC-associated cluster remain associated with the major MHC locus in the same syntenic block as in other mammals. However, some ORGs co-migrated to other chromosome locations with the class I MHC genes. This finding does not strongly support the already controversial findings that this ORG cluster might have some functional contribution toward the olfaction-mediated mate choice. Some ORGs of the major MHC-ORG cluster, however, show signs of strong conservation and are specific to
this cluster. These could serve as good candidates for further behavioral and functional studies that might shed light on the functional involvement of ORs in mate selection based on MHC heterogeneity.
## Additional Files

**Additional file 1.** Name and sequence of overgo probes used in isolating BAC clones from tammar wallaby genomic library.

<table>
<thead>
<tr>
<th>Probe/Gene</th>
<th>Forward Sequence</th>
<th>Reverse Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WDR45L</td>
<td>TGACACTGCAAAAAAGAAACCTTACC</td>
<td>GCCAAATGCTGTGATAGGTAAGTT</td>
</tr>
<tr>
<td>NARF</td>
<td>GGCTGATGCTAAAAAATATCTTTCAG</td>
<td>ATAGGCCCAAGCTACACTGAGAAA</td>
</tr>
<tr>
<td>HEXDC</td>
<td>AACCAGATGCAATAAGTGGAGTCTTAC</td>
<td>CACTGAGTCTGACAGAACTAAC</td>
</tr>
<tr>
<td>GABBR1</td>
<td>AGGGGAAAGGGAAAAGGTAGTCAG</td>
<td>TCACTGTTTGCTTAAAACCTGACTC</td>
</tr>
<tr>
<td>MOG</td>
<td>CTAGGACCTTCTTCTTCTTCTTCTT</td>
<td>CAATCATGCTTGAATAATTGAACTT</td>
</tr>
<tr>
<td>ZNRD1</td>
<td>AGGATGCTGGAGGCGGTCACGTGTC</td>
<td>TACGGTACAGTGAGAAGAAGCTG</td>
</tr>
<tr>
<td>Meu-OR4</td>
<td>ATTCAGGGGACTCAAGTTCTTGTTC</td>
<td>CCATGTAACCCTAATTGGAACATG</td>
</tr>
<tr>
<td>Meu-OR26</td>
<td>CTTCTCTCTTTCTTTCTTTCTTCTT</td>
<td>CAAATCATGATAGCAGAAAACCA</td>
</tr>
<tr>
<td>Meu-OR82</td>
<td>CTACAGAGAAATTTCTTCTTCTTCTT</td>
<td>CAATGGAACGAGATATGAGAATG</td>
</tr>
<tr>
<td>Meu-OR237</td>
<td>CTACAGAGAAATTTCTTCTTCTTCTT</td>
<td>CAATGGAACAGATATGAGAATG</td>
</tr>
<tr>
<td>Meu-OR241</td>
<td>GATGACTGGAAGAAAAGTAACTAA</td>
<td>CCTTGAAATAAATTCCTTAGTTAC</td>
</tr>
<tr>
<td>Meu-OR293</td>
<td>CCAGGGAAGCTCAAGTTCTTCTTCTT</td>
<td>CAAATCATGCTTGAATAATTGAACTT</td>
</tr>
<tr>
<td>Meu-OR327</td>
<td>ATTCAGGGGACTCAAGTTCTTGTTC</td>
<td>CCATGTAACCCTAATTGGAACATG</td>
</tr>
<tr>
<td>Meu-OR428</td>
<td>TAGAAATACAGTTATGCTCATATGA</td>
<td>CGATCTCCTCTGACCTGACTGAC</td>
</tr>
<tr>
<td>Meu-OR471</td>
<td>CCAGGGAAGCTCAAGTTCTTCTTCTT</td>
<td>CAAATCATGCTTGAATAATTGAACTT</td>
</tr>
<tr>
<td>Meu-OR528</td>
<td>GAAGGAAACCACTACTTTCTTCTTCTT</td>
<td>TGAACGAGTCCCTATGACAAAC</td>
</tr>
<tr>
<td>Meu-OR541</td>
<td>CATTTCTCTGGGAAGCTGAGTTCG</td>
<td>GTCATCACTGACTAGCA</td>
</tr>
<tr>
<td>Meu-OR620</td>
<td>TTAATTTCTCTCACCCTTCCATCTAA</td>
<td>GGAAGCAAGGATGTAGTCTT</td>
</tr>
<tr>
<td>Meu-OR629</td>
<td>GGGAGGCAACACTTTTAATTTTGA</td>
<td>TGATCATACAGACACTCCTCAATT</td>
</tr>
<tr>
<td>Meu-OR779</td>
<td>TTTCTCTCCACAAGGATGAGATATG</td>
<td>TCCATCCTGCTTCCTCAATG</td>
</tr>
<tr>
<td>Meu-OR845</td>
<td>ACCTCAGGTTCTCATCGATTTCTCCTG</td>
<td>CACCTTTGGTCATGCAAGAAGAAG</td>
</tr>
<tr>
<td>Meu-OR884</td>
<td>GTCCACTCAGTTGAAACTGTCTTCC</td>
<td>AAAACTCATGTTAAGAAGAACAGT</td>
</tr>
<tr>
<td>Meu-OR956</td>
<td>TAAAACATAGAGGTATCTGTTTAGT</td>
<td>TGGGAGTGAGGATTAAACTAAAGA</td>
</tr>
<tr>
<td>Meu-OR977</td>
<td>AAATGATGCGATCCGCGAGCTGGT</td>
<td>CAAATCATGCTTGAATAATTGAACTT</td>
</tr>
<tr>
<td>Meu-OR1129</td>
<td>TTGGGAAGAGTTACTCATCTTCTTAGG</td>
<td>CCTTAGATATTTTGACACTAG</td>
</tr>
<tr>
<td>Meu-OR1147</td>
<td>ACCACACAGTCACAAATAAACAG</td>
<td>GAGCTGTTACAGACTCCTTATTAT</td>
</tr>
<tr>
<td>Meu-OR1415</td>
<td>TTGGTTCGGTCAGTTCTCTCTTCC</td>
<td>GCAATCGTGAATTATCAGGAAGT</td>
</tr>
<tr>
<td>Meu-OR1507</td>
<td>CTAGTTATGGGCTCCTTCTTCTGAC</td>
<td>GCAAAGGAAAAAGGCCTGACAG</td>
</tr>
<tr>
<td>Meu-OR1688</td>
<td>TCTTTAACTCTCAATCAGACTCCT</td>
<td>AAGTTGGAATGAGGCAGAGACTG</td>
</tr>
<tr>
<td>Meu-OR9</td>
<td>CATTATGCCAACCACCTCACCCTAATC</td>
<td>TGAATTATGCTGTGAGCTTGG</td>
</tr>
<tr>
<td>Meu-OR120</td>
<td>AAGGGCAAGACTTTATCTTTCTCACACT</td>
<td>GTCATAATCTACACAGGACTG</td>
</tr>
<tr>
<td>Meu-OR286</td>
<td>TCTTTTTCTCAGGATCTCCTCTTACA</td>
<td>ATGGAGAACTTGAGCAGTAAGAAG</td>
</tr>
<tr>
<td>Meu-OR364</td>
<td>TCAAGGTGTTGATGCTTAATGCTG</td>
<td>CAAACAGAAAATATCATGCAACATT</td>
</tr>
<tr>
<td>Meu-OR432</td>
<td>GTCCTCTTTCTGATTACATTTTGGC</td>
<td>CAGCAGAGAGAGGTTAGAAGAAT</td>
</tr>
<tr>
<td>Meu-OR454</td>
<td>GATTTACTGTTGCTGACACCTTCTTGA</td>
<td>GCCAAAGACGTTAGAAGTACAG</td>
</tr>
<tr>
<td>Meu-OR482</td>
<td>AAATCTGCTGTTGTTCAAGGGGTAA</td>
<td>GTTTGGAATGAGGCAGAGACTG</td>
</tr>
<tr>
<td>Meu-OR542</td>
<td>GTCAACAGATGGAACACTTCTGGA</td>
<td>CACCTTTGGTACGCTGACAG</td>
</tr>
<tr>
<td>Meu-OR576</td>
<td>ATATTAACAGACAGCCTCAACCAG</td>
<td>ATCATGTTGCGCTTTAGCTTGA</td>
</tr>
<tr>
<td>Meu-OR684</td>
<td>AACTCTTTGCATCCCAGCCCCTT</td>
<td>CAACTGGAAGGTAGGAAGGCTT</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Meu-OR692</td>
<td>GCCATAAGATAGAAAAACAGAAG</td>
<td>ATAGGCGAGATCAGAACTTCTGTT</td>
</tr>
<tr>
<td>Meu-OR756</td>
<td>GTAGAAGGCACTTACCATTACCAC</td>
<td>TGATGTCTGCAATGTTGATATG</td>
</tr>
<tr>
<td>Meu-OR853</td>
<td>AGATCCCAAGGAACTATTATTCCTCC</td>
<td>AGGGAAGATCTTAGTGGAATA</td>
</tr>
<tr>
<td>Meu-OR936</td>
<td>GGCCCAAGTTCTTGGTATCTTG</td>
<td>AAAGGTTGGAAGTCCTCACATTC</td>
</tr>
<tr>
<td>Meu-OR1053</td>
<td>TGACTGAGGAAAGGAAAGGCTTTT</td>
<td>GAGCACAAGTTGAGAGGCTT</td>
</tr>
<tr>
<td>Meu-OR1082</td>
<td>GTTTCACTATGGAAGGATCTTCC</td>
<td>TCTAGGACACAGGACTGAGAAGCT</td>
</tr>
<tr>
<td>Meu-OR1170</td>
<td>ATACAGTCTAGAGGATAGTACC</td>
<td>TCCCTATGCCTGACAGGCTACAT</td>
</tr>
<tr>
<td>Meu-OR1211</td>
<td>CCAAGGCTTTCTTCTCAATGCC</td>
<td>GTAGTGGGAGATCTGCTGATTG</td>
</tr>
<tr>
<td>Meu-OR1227</td>
<td>AGGAAGGTGAGTCAAGGAGGAA</td>
<td>TAGCCAAGCTCCTCAATTTCCTG</td>
</tr>
<tr>
<td>Meu-OR1234</td>
<td>ATCCCTTTTCTCCCTGAGCTTTCA</td>
<td>GCAAGTTTCTCTGTTGGAAGAGC</td>
</tr>
<tr>
<td>Meu-OR1309</td>
<td>ATACCTTGATATACCCAGCTCTCT</td>
<td>ATCTTCTATGCTGATAAGGAGCT</td>
</tr>
<tr>
<td>Meu-OR1381</td>
<td>GGCTCTCTCTTCTGATCTTAAACCA</td>
<td>GTAATGTTTTTAGAGGAGCTT</td>
</tr>
<tr>
<td>Meu-OR1403</td>
<td>GGAAATTGCAAAGAAAACCTAACC</td>
<td>GAGGAGAGACCTTTGTTAGT</td>
</tr>
<tr>
<td>Meu-OR1430</td>
<td>GCAATTGTTGAAAGGAGGAGGTC</td>
<td>TAGATCTCCTAGGCTCTGACTCCT</td>
</tr>
<tr>
<td>Meu-OR1439</td>
<td>TCCATTTGCTTACCTTATTTTCTC</td>
<td>CTCTCATTTTCCAGAGCCAGATA</td>
</tr>
<tr>
<td>Meu-OR1611</td>
<td>GTGCAAATGTCTACCTGTTGAAAGT</td>
<td>ACACCCTTAGTGACGTACCTCAC</td>
</tr>
<tr>
<td>Meu-OR1686</td>
<td>GTAGCCAGGTTGGTAAAGGAGAG</td>
<td>GCAACAAGGCATATTCTCTCCT</td>
</tr>
</tbody>
</table>

100
CHAPTER 6: DISCUSSION

In this project, I have studied examples of genome change in marsupials at three levels; gross chromosomal arrangement, gene family organization, and gene arrangement at the sequence level.

Marsupials have always been interesting subjects for biological and genomic comparisons. They are a group of mammals very distantly related to humans and other placental mammals. They have evolved a range of different lifestyles and morphological features to achieve successful distribution and adaptation to a wide range of ecological niches in Australia and South America.

There are three major reasons why marsupials are so valuable for comparative genetics studies. Firstly, marsupials occupy a unique position in the vertebrate phylogeny, having diverged from the eutherian lineage approximately 147 million years ago. This places them in a valuable phylogenetic position between the divergence of the reptile/bird lineage and the mammals (approximately 310 million years ago), and the rapid radiation of the eutherians approximately 105 million years ago. Secondly, marsupials have many distinctive biological characteristics, most particularly in their unique mode of reproduction, and they have adapted to some unique ecological habitats. Understanding the basics of the differences between the biology of marsupials and eutherians is a key component in revealing the evolutionary process of such characters in all mammalian species. Thirdly, small marsupials can be used as uniquely diverged model mammals for biomedical research.

The recent sequencing of two marsupial genomes (opossum and tammar wallaby), and their inclusion in vertebrate and mammalian genomic comparisons, has improved our understanding of mammalian genome organization and evolution, as well as helped to identify new genes and regulatory elements (Wakefield & Graves 2003).

In my research, I used opossum and tammar marsupial sequences to explore one of the most fascinating aspects of mammal genomes; genome stability and genome change in evolution. This thesis reports my observations on genome change at three
levels; gene arrangements on the chromosomes, the composition and genome-wide
distribution of members of the ORG family, and finally, the detailed organization and
potential functions of the MHC-linked ORG cluster in tammar wallaby. My
conclusions from experiments addressing each of these three objectives are discussed
below.

**Genome stability at the chromosomal level**

Marsupials are strikingly different from eutherians in that they have a very conserved
karyotype (Wienberg 2004). The first marsupial karyotype was published in 1961
(Sharman 1961) and since then almost all living marsupials have been karyotyped.
Marsupials have a low diploid number of chromosomes, ranging from 2n=32 in
*Aepyprymnus rufescens*, down to 2n=10 in *Wallabia bicolor* females. This extreme
conservation has its parallels in the extremely conserved reptile-bird karyotype
(Graves & Shetty 2001) thus it appears that karyotypic change was slow during most
of the evolution of tetrapods, but accelerated in eutherian mammals.

The distribution of marsupial chromosome numbers has strong modes at 2n=14 and
2n=22 (Sharman 1961); almost 65% of all extant marsupial species have one of these
chromosome numbers, and Sharman suggested that the 2n=22 karyotype was basal
for all marsupials. However, Australian marsupials seem to have a highly conserved
2n=14 karyotype, and Hayman and Martin (1965), and Rofe and Hayman (1985)
suggested that this was the ancestral marsupial karyotype (for review see Graves &
Westerman 2002b).

American and Australian marsupials are proposed to have been evolving separately
for more than 65 million years. Whether the 2n=14 karyotype common amongst
Australian marsupials, or the 2n=22 karyotype found in many South American
marsupials was ancestral has been debated for decades. Svartman and Vianna-
Morgante (1999b) showed that changes in chromosome number among South
American marsupials are largely due to Robertsonian rearrangements (centromere
fusion and fission), and the presence of interstitial telomere sequences in the large
chromosomes of some of the 2n=22 species favoured the idea that the number was
originally higher, but became lower through chromosome fusion.
The answer to the question of the ancestral marsupial karyotype has come from molecular phylogeny, that shows clearly that the South American didelphid opossums (with 2n=22) are indeed ancestral to other South American groups, as well as to the basal Australian groups with 2n=14. These data also resolve the strange case of *Dromiciops gliroides* (Microbiotheria), which, although living in South America, have 2n=14 and chromosome G-band patterns similar to that of Australian 2n=14 marsupials (Spotorno *et al.* 1997); this species falls within the Australian radiation and presumably represents a species that returned over the Gondwana land bridge to South America before its final rupture 55 MYA.

Chromosome painting comparisons of many marsupial karyotypes confirm that the conservation of karyotypes reflects homology at the DNA level. In comparative chromosome painting, the whole or part of a chromosome in one species is physically isolated (by flow sorting or microdissection) and, after labelling with fluorescent tags, is used as a probe in hybridization to the metaphase chromosomes of another species. Parts of the genome (chromosome arms or segments) in the target species that share the highest sequence homology with labelled probes will be covered with fluorescent signal and look as if it is painted (Ferguson-Smith 1997).

A series of studies, using paints from one marsupial species to detect homologous segments in others, showed that all marsupial genomes are made up of the same 19 different conserved segments, which are present in all genomes but may be rearranged (Rens *et al.* 2003). These segments are found fused together in some species, and lie separately on two chromosomes in another species. The number of rearrangements that each marsupial karyotype has experienced from the ancestral form is much lower than in eutherian mammals; for instance, there are hundreds of segments that have been rearranged between human and mouse karyotypes (for a comprehensive review see Serov *et al.* 2005)

Is it only the chromosome number and morphology that has remained conserved, or is gene content and arrangement also conserved?

Comparative gene mapping can answer this question. Comparisons of the gene arrangements (the relative order of genes in conserved syntenic blocks) between different marsupials, as well as between marsupials and other vertebrates,
complements the comparative chromosome painting and produces more detailed information on the meaning of chromosome homology (Serov et al. 2005).

The availability of a good gene map is the first requirement of such comparisons. During the last five decades, several different methods have been used to construct gene maps of marsupials, including somatic cell hybrid mapping, linkage mapping, and physical mapping (Serov et al. 2005). Rodent-marsupial somatic cell hybrids were the first attempt to construct gene maps in marsupials (for a review see Graves et al. 1989). These hybrids were very difficult to work with and very unstable, but they were useful for mapping some genes to the marsupial X chromosome (Dawson & Graves 1984, Dobrovic & Graves 1986).

Sex linkage of some markers could be established from the pattern of inheritance in kangaroos (reviewed in Cooper et al. 1977, Vandeberg et al. 1980), but developing marsupial linkage maps was difficult until the mid 1980s due to the lack of reliable, multigenerational breeding programs. The first linkage estimates for a marsupial were published on the fat-tailed dunnart (Sminthopsis crassicaudata), and since then linkage maps for several marsupials have been constructed (reviewed in Samollow 2010). Although linkage maps provide information about the order of markers within a linkage group, they usually consisted of anonymous sequence markers, greatly limiting their value in comparative genomics. It is also necessary to assign linkage groups to chromosomes using physical mapping.

Physical maps, however, can use molecular markers developed from conserved genes. Large insert clones can be localised by in situ hybridization, and are very powerful in revealing the homologies in gene order between different species at the chromosomal level (Deakin 2010). In early years, physical mapping used radioactively-labelled probes to locate the gene by autoradiography (Buongiorno-Nardelli & Amaldi 1969, John et al. 1969). Globin genes HBA, and HBB were the first genes to be physically mapped in a marsupial using marsupial cDNA clones as probes (Wainwright & Hope 1985). Radioactive in situ hybridization (RISH) using human or mouse cDNA clones as probes was very successful in establishing the chromosome location of many genes on marsupial sex chromosomes (reviewed by Graves 1996). However, the use of heterologous probes could produce false signals and mis-localisation of genes (Deakin 2010).
The introduction of fluorescence labelling overcame these obstacles and increased resolution, so that fluorescence *in situ* hybridization (FISH) revolutionized comparative mapping by producing gene localisations for many species, including those in which breeding was impracticable.

Physical mapping using FISH has now been used to construct a virtual map of the tammar wallaby, as well as the opossum genome. The availability of DNA sequence has ensured a limitless supply of markers of high reliability. The 1.5-fold genome sequencing effort of tammar wallaby produced a huge number (1,174,382) of short contigs (average length of 2.2 kb). Assigning all these short contigs to their chromosomal locations in tammar wallaby by mapping every one of them would be an impossibly time-consuming and tedious effort. As an alternative, a selected number of genes can be mapped to the chromosomes, allowing us to predict the location and order of all sequences (Deakin *et al.* 2008). This strategy is achievable since the genome of another marsupial (the opossum) is more deeply sequenced (6-fold) and is at least partially assembled at the chromosome level. By detecting blocks of genes conserved between opossum and human, we predict that these blocks are also conserved in tammar wallaby. Using BACs that contain the genes from each terminus of these blocks, and mapping them physically in the tammar wallaby, provides a framework for annotating the entire genome, and the use of the very long BACs as probes increases the signal strength and reliability.

The efficiency and accuracy of the approach used in this study for mapping genes onto tammar wallaby chromosomes had already been tested and discussed (Deakin *et al.* 2008). This efficiency is largely due to the availability of tammar wallaby genomic sequences as trace archives and a BAC library of tammar wallaby, combined with the annotated opossum genome.

In this way I prepared a physical map of tammar wallaby 6q as part of the Kangaroo Genome Project to help the assembly of its 1.5-fold sequenced genome. Previous studies, based on G-banding and on chromosome painting (Rofe & Hayman 1985, Rens *et al.* 1999, Rens *et al.* 2003) suggested that this part of the tammar genome had homology to the long arm of chromosome 7 of opossum. I used BAC mapping to see if the order of syntenic blocks, and the order of genes within this segment, was also conserved between tammar wallaby chromosome 6q and opossum chromosome 7q.
My FISH mapping results confirmed that the overall gene order of this region is well conserved between these two marsupials, with only a few exceptions. When I compared the order of mapped genes in tammar with that in birds (chicken) and two eutherians (human and dog), I found that there are more rearrangements in this region in mammals than there are in birds. This is not unexpected, since comparisons between mammals and non-mammalian species have shown that the ancestral vertebrate karyotype is generally more conserved in birds and reptiles than it is in mammals. The ancestral eutherian karyotype is highly rearranged with respect to the ancestral vertebrate karyotype. This has led to the conclusion that most of the mammalian rearrangements from the ancestral vertebrate karyotype have happened in the mammal lineage after its divergence from birds around 310 million years ago (Mikkelsen et al. 2007).

The results from my research, and similar comparative mapping of other regions of marsupial genomes, attributes most of these rearrangements to recent rearrangements in eutherian lineages, after the divergence of marsupials. Marsupial genomes seem to have suffered fewer rearrangements compared to their eutherian cousins, in reference to the bird and fish genomes.

How has the stability of marsupial karyotypes affected marsupial evolution? There is no doubt that chromosomal rearrangements can play an important role in the evolution of organisms (White 1978, King 1995, Olmo 2005). This role is manifested in the form of a direct relationship between the level of karyotypic changes and the rate of speciation (Bush et al. 1977, King 1995, Britton-Davidian 2001, Olmo 2005). However, it is still debatable whether karyotypic changes cause speciation (King 1987, King 1995) or are by-products of natural selection favoring a certain karyotype through speciation (Coyne 1984).

Placentals have been shown to have evolved faster than other vertebrates at the sequence level and this is accompanied by more chromosomal changes (Wilson et al. 1975). Eutherian mammals in general and some orders specifically (primates, rodents) have evolved much faster than other mammalian lineages with more karyotypic changes (Bush et al. 1977). This contrasts with metatherians, in which a limited number of species (despite a long history of evolution) is linked with a very stable karyotype. The karyotypic evolution for marsupials is estimated around 0.124
changes in the number of chromosome arms per million years and the speciation rate at around 1.0 (Wilson et al. 1975). This is markedly different from the karyotypic evolution for primates, lagomorphs, and rodents at around 0.413, 0.403, and 0.253, with a speciation rate of 2.6, 1.6, and 1.6 per million years, respectively (Wilson et al. 1975).

Is chromosome change linked to species diversity? Today the number of living primate species is more than 650 (Rylands & Mittermeier 2009) and there are more than 2270 rodent species (Wilson & Reeder 2005): these two groups (40% of all mammalian species) have dominated every corner of the globe. In contrast, the 334 species of marsupials occupy only two continents. Major unanswered questions are how marsupials ended up with such a stable karyotype and whether it has limited their diversity? I have tried to partly address these fundamental questions in the following sections.

**Genome change in the olfactory receptor gene family**

Is the unusual genome stability and strong conservation of the marsupial karyotype reflected in greater stability of the numbers and arrangements of genes in large gene families? To test this possibility, I looked at one of the largest gene families in mammalian genomes, the olfactory receptor gene (ORG) family, and compared the ORG families between marsupial species, and between marsupials and eutherians.

Vertebrates rely on a family of receptors for acquiring the chemical information from their environment. The importance of this task is reflected in the huge size of the ORG family. ORGs are short sequences (around 1kb) comprising a single exon. They belong to the superfamily of G protein-coupled receptors. In mammals they are the biggest gene family in the genome, ranging in number from few hundred genes to over 2,000 in some species (reviewed comprehensively in Delbridge et al. 2010). This huge diversity is the result of their ‘birth-and-death’ pattern of evolution, which is triggered by random duplication events and provides more and different gene copies for expression in different species (Chapter 4).

Genes of the ORG repertoire of the tammar wallaby were extracted from the first tammar assembly by homology searches using a series of vertebrate ORG sequences.
as the query. The tammar wallaby ORG sequences were cross-checked and purified from other related sequences as described (Chapter 4).

The tammar wallaby genome was found to contain as many diverse ORGs as other mammalian species (opossum and eutherians). The ORG repertoire of the tammar wallaby consists of over 1,700 ORG sequences, half of which are pseudogenes. These sequences can be classified into 18 families and 240 subfamilies. This diversity is not unexpected, as tammar wallaby, being a nocturnal marsupial, relies heavily on its sense of smell for food and social and environmental interactions (for a review see Delbridge et al. 2010).

The number of ORGs found in the tammar wallaby is similar to that found in the other sequenced marsupial genome, the opossum, and ORGs are distributed over the same number of families and almost the same number of subfamilies in the two species. The similarities between these two marsupial ORG superfamilies, however, extend beyond the total numbers of genes and the family distribution.

We established a one-to-one relationship between most of the tammar wallaby and opossum ORGs, suggesting that these two marsupials have retained a very conserved ORG repertoire since their divergence about 70 MYA. The organization and genomic distribution of ORGs also seems to be very conserved, as the major ORG clusters (accounting for around 90% of the entire ORG repertoire) are located in the same syntenic blocks in the two marsupial genomes (Aloni et al. 2006).

These two marsupials have been evolving independently for about 70 million years, so it is surprising that they have such a very similar ORG repertoire. Eutherian mammals of the same divergence time have ORG families that have diverged wildly. Like tammar wallaby and the opossum, human and mouse, or human and dog, shared the same ancestor about 70-80 MYA but have very different ORG repertoires. Human and mouse each have 802 and 1,391 ORGs respectively, but of these, 372 are unique for human and 350 are unique for mouse, and only 430 ORGs are conserved between the two species (Niimura & Nei 2005b).

Retaining the same genes, and having suffered fewer evolutionary events that alter gene number (such as gene duplications, and pseudogenization of the duplicated
genes) can be attributed either to a lower frequency of duplication, or to convergent evolution in response to a greater selection pressure on the ORGs of marsupials compared to eutherian mammals. This latter possibility is plausible because their unique mode of reproduction might necessitate that marsupials retain a conserved set of ORGs that are vital for the newborn in vital climb up to the pouch from the birth canal (Veitch et al. 2000). Also, most marsupials (and therefore their common ancestor) are dichromatic, nocturnal (or crepuscular) and semi-social animals that use olfactory cues as an important mean of absorbing the environmental information (e.g. food, danger, presence of mates).

However, a more plausible explanation is that this high level of conservation in ORG family reflects the lower rate of segmental duplications in marsupials compared to other amniotes. This does seem to be the case. Only 1.7% of the opossum genome sequences are the result of segmental duplications, whereas the proportions of the same sequences in mouse, human and chicken are 5.2%, 5.3%, and 10.4%, respectively (Mikkelsen et al. 2007). Opossum segmental duplications are also shorter, more closely spaced within chromosomes, and less likely to be dispersed among chromosomes than those of humans. The ORG family relies mostly on segmental duplications (intra-cluster duplications) for its expansion and diversity. Marsupial genomes (opossum and tammar wallaby, for example) are shown to include fewer macro-rearrangements than eutherian genomes. Fewer genome-wide opportunities for duplications and dispersal would lead to a more stable ORG family.

ORGs are organized in clusters throughout the genomes, and comparing the gene content and gene order of an orthologous ORG cluster in two marsupials could test this theory. The difference between ORG repertoires of different species is reflected in the number of genes in their ORG clusters. If two marsupials share a conserved ORG family (due to an exceptionally stable genome), they should also share very similar ORG clusters in reference to other mammals (namely eutherians).

**Genome changes in the MHC-linked ORG cluster**

ORGs, like genes in many multigene families, are organized in clusters of two to more than 100 genes distributed across mammalian genomes. I showed that the
location of most of the tammar wallaby ORG clusters is conserved within the same syntenic blocks of genes, which are shared between opossum and the tammar wallaby (Chapter 4). One of these conserved ORG clusters in mammalian genomes is physically linked to the MHC locus, giving rise to speculation that members of this cluster might be involved in olfaction-mediated mate choice (Penn 2002).

There are some conserved features of the ORG cluster adjacent to the MHC locus shared by all mammals, but there are major differences between different eutherian mammals. For example, human, and rodents have two subfamilies (5V, and 5U) in this cluster that are absent from the dog MHC-associated cluster, and the duplication events for these two subfamilies are, again, different between human and rodents (Chapter 5).

In order to investigate the gene content and order of the MHC-associated cluster in tammar wallaby, I constructed a BAC contig over this cluster and mapped the ORGs in the cluster. To explore the possibility of any functional significance of the physical juxtaposition of ORGs and MHC (e.g. particularly favorable haplotypes), the ORGs of this cluster were detailed in the tammar wallaby and phylogenetically compared with those in seven mammalian species.

The MHC locus in tammar wallaby has a unique structure. MHC class I genes in the tammar wallaby, unlike other mammals (even the opossum), are distributed in clusters at several locations across the genome. However, all class II, class III and a few class I genes remain in the same conserved location as in other mammals (Deakin et al. 2007, Siddle et al. 2009). Relocation of class I genes away from the major cluster, and away from class II (antigen-presenting) and class III genes has been suggested to be due to the proximity of retroviral (KERV) elements, and may be driven by selection for a new role for Class I genes (Siddle et al. 2009). It is therefore of great interest to ask whether the ORGs are found only at a single (original) location, or whether they, too, have been duplicated and dispersed.

I found that tammar wallaby, like other mammals, has a cluster of ORGs in the same syntenic block as the major MHC locus located on tammar wallaby chromosome 2q. This major sub-cluster (containing at least 17 ORGs) is conserved in close proximity to the major MHC locus on chromosome 2q and contains the same ORG content.
(subfamily and gene number per subfamily), as does the MHC-linked ORG cluster in the opossum. The major ORG cluster in tammar shares many conserved features with its ortholog in opossum, and the differences between the tammar wallaby and opossum clusters are very minor compared to the differences between any two eutherian mammals. Some minor rearrangements, unique to tammar wallaby, are observable in this ORG cluster.

However, my results show that MHC-linked ORGs in tammar wallaby are also distributed into at least three sub-clusters located in three different regions of the genome. One of the minor ORG clusters containing two genes is close to a cluster of MHC class I genes on chromosome 5q. This suggests that ORGs have duplicated and co-migrated to new sites along with MHC genes.

Although analysis of the MHC-linked ORG cluster in tammar wallaby is not complete, it is evident that the major MHC-linked ORG cluster is more conserved between marsupial species than it is between eutherian species. Again, this reflects a pressure that marsupial genomes are experiencing to remain remarkably similar even at the level of segmental duplication and cluster organization across the entire infraclass and through time.

Thus marsupial genome stability is reflected even at the most detailed level of organization. Firstly I showed that the gene arrangements within the conserved syntenic blocks are highly conserved between two marsupials. Secondly the ORG repertoires of these two marsupials show a high level of conservation in terms of gene birth and death. And finally a stable ORG repertoire is reflected in its organization in clusters.

**Conclusions**

This PhD thesis reports research aimed to understand how the genome of marsupials has changed in evolution at three different levels of observation; at the gross chromosome level by comparing the gene arrangements, at the level of gene families by examining the numbers and organization of members of the ORG family, and at the level of detailed analysis for one interesting ORG cluster between two sequenced marsupial genomes.
For a long time it was known that marsupials have exceptionally conserved and less rearranged karyotypes than any other mammalian groups. Here I add more supporting results that this conservation extends beyond the karyotype (chromosome numbers and morphology) to gene content and arrangement, to the number and organization of the ORG superfamly (the number of genes, its subfamily composition, relationship between orthologs, and distribution of ORG clusters over the genome), and also to the detailed arrangement within the MHC-linked ORG cluster (location of the major cluster, gene content, subfamily composition). Minor rearrangements were detected in the tammar wallaby genome at all these three levels compared to the opossum genome, but the extent of change seems to be much less than for other mammals at all these levels.

The next immediate steps in testing the genome stability of marsupials at the most detailed level could involve mapping all the ORGs in tammar wallaby and some other marsupials. Sequencing a few BACs covering the MHC-linked ORG cluster in tammar wallaby will greatly help to detail the stability of the ORG repertoire distribution across the marsupial genomes. Localization of MHC-linked ORGs in platypus also will help to understand the evolution and importance of this gene cluster and will provide a basis for further functional studies.

Marsupials are not alone in having a stable karyotype and hence a stable genome. Other vertebrate groups like birds (Griffin et al. 2007), turtle (Matsuda et al. 2005), and squamates (snakes, lizards) (Olmo 2005) also have very stable genomes. Bird karyotypes are remarkably uniform with around 80 chromosomes consisting of 7 pairs of large chromosomes (macrochromosomes) and a many very small chromosomes (microchromosomes) (Belterman & Boer 1990). The turtles and snakes also have a karyotype very well conserved and similar to that of birds (Matsuda et al. 2005) although their ancestors diverged from the birds’ ancestor 200 and 210 MYA, respectively (Iwabe et al. 2005, Muller & Reisz 2005). Chromosome painting confirms chromosome homology over very large genetic distances (e.g. turtles and birds) (Graves & Shetty 2000). However the genome stability in class Reptilia seems to be different from what is observed in Metatheria and is more ancestral.

Bird karyotypes are very similar even between ratites and carinates (Shetty et al. 1999), and reflect the karyotype of a proto-karyotype of the vertebrate ancestor.
(Nakatani *et al.* 2007), especially in that most of their microchromosomes are one-to-one counterparts of each other (Burt 2002). Only few vertebrate-wide fusion and some recent, avian-specific fission events mark the difference between the karyotypes of the bird and vertebrate ancestor (Nakatani *et al.* 2007). The karyotype of the early ancestors of marsupials changed extensively and experienced a lot of fusions following the divergence from the rest of the therians (Nakatani *et al.* 2007).

One more difference between the stable genomes of marsupials and birds arises from the speciation events. Having a stable genome may have restricted the marsupials to very slow speciation, while birds, despite their stable genome, have diverged much more quickly and today there are around 10,000 species of birds (Gill 2006). It has been speculated that although evolution through genome and karyotype change is a constant observation in mammals, birds do not need to reshuffle their genomes for evolution. A large number of chromosomes could facilitate the speciation by random segregation and crossing over (Griffin *et al.* 2007). Likewise there are at least 60 species of dasyurid marsupials that have virtually identical G-banded karyotypes (Young *et al.* 1982). On the other hand, other marsupial groups, such as the rock wallabies of genus Petrogale, have undergone rapid chromosome evolution in small groups on rocky outcrops or islands that enforce small independently evolving subspecies (Eldridge & Johnston 1993)

After the second round of whole genome duplication in the ancestor of Gnathostomata, which duplicated the number of chromosomes to $2n \approx 80$, the number of chromosomes started to decrease gradually due to smooth fusion events in the ancestor of all extant vertebrates (except for cartilaginous fish like sharks) (Nakatani *et al.* 2007). During the past 200 million years however, some lineages such as the metatherians experienced a sudden and deep reduction in the number of chromosomes. The ancestral mammalian karyotype was $2n=46$ (Kemkemer *et al.* 2009) and it reduced to $2n=22$ in an ancestral metatherian (Svartman & Vianna-Morgante 1999a), then $2n=14$ in an Australian metatherian ancestor (Graves & Westerman 2002b) and remained stable at this chromosomal number. In contrast the eutherians evolved a wide range of karyotypes from this ancestral karyotype. How marsupial karyotypes ended up being so different from other mammals and yet very uniform across the whole infraclass is an interesting question.
Chromosomal changes and karyotype evolutions have been hypothesized to facilitate speciation by either of two processes (Olmo 2005). One hypothesis is that mutations that increase the fitness or even some slightly deleterious mutations and chromosomal changes will be fixed very rapidly in small populations that have been isolated from the mother populations for some generations (King 1987, King 1995). These fixed rearrangements could serve as reproduction barriers and hence promote the emergence of a new species. Others put the stress on achieving a karyotype optimal for an adaptive zone (Bickham & Baker 1979). When a new species emerges, its karyotype will keep changing until it reaches a point where more changes will be deleterious and maladaptive for that environment.

The isolation necessary to fix rearrangements on the first theory could be achieved through a geographical barrier or some social behaviors favoring the formation of small demes of animals (Wilson et al. 1975, Bush et al. 1977, King 1995). Most mammalian species are thought to have evolved through this process of social isolation. Behaviors like polygamy, mother-offspring bonding (increasing the inbreeding), high mobility (migration to far distances), strong territoriality, and other factors (like food resources, predators pressure, patchiness of the environment), could potentially drive animals to form social structures with a small effective population size. This therefore creates a bigger opportunity for the rapid fixation of genome changes (Wilson et al. 1975).

A slower rate of chromosomal change in marsupials however seems not to be due to the lack of these behaviors and social structuring. There is a strong mother-offspring bonding in marsupials (Renfree 1995) and they are shown to be polygamous rather than monogamous (Medellin 2004). Many marsupials also have a good mobility and can and do travel around and live in small groups almost isolated from other groups. Some marsupials are also territorial animals (Tyndale-Biscoe 2005).

Drawing the exact picture of what happened 147 MYA when marsupials started diverging from the rest of Theria is almost impossible. But probably some environmental and/or physiological changes (like reproduction mode) might have facilitated or dictated a rapid karyotypic changes (intensive fusions) to form an ancestral karyotype of 2n=22 which was the optimum genome organization for the environment and biology of the first marsupials.
A more likely scenario is that some lineage-specific (therefore recently evolved) sequence features of eutherian chromosomes (e.g. repetitive sequences) mediate frequent changes. The absence of these sequences from chromosomes in marsupials, monotremes, birds and reptiles may be what is enforcing chromosome stability. Repetitive sequences can promote chromosome change by facilitating recombination events (Ferguson-Smith & Trifonov 2007). This is consistent with a lower repetitive sequence content of the opossum genome.

Finally, I conclude that the genome organization (at three levels of karyotype, gene family, and gene arrangements) is very old and heavily conserved for two marsupials (representing two major families of American and Australian marsupials) that diverged 70 MYA. This genome stability is an adaptive response in order to gain the optimum fitness level for the environment in which the first marsupials emerged and some unique biological features of marsupials were established.

Future directions

Marsupials possess a very conserved genome, but they are not an exception in this regard. Other vertebrate groups (like birds, fish, snakes, turtle, etc.) also have a very stable genome. In contrast, eutherian mammals have been evolving fast with a very dynamic genome. Is genome stability the general rule and only eutherian mammals are the exception here? Why do eutherians reshuffle their genome so quickly and deeply, while birds and fish have retained karyotypes similar to those of their ancestral genomes and yet are evolving at the same rate (if not faster than) as eutherians. Are eutherians evolving toward a more stable genome? Are marsupials evolving a eutherian way of evolution? Will these groups stick to their own stable/unstable genomes? These are fascinating questions and can only be answered by comparing more mammalian genomes and expanding the comparisons to more distant vertebrate relatives.

So far only two marsupial genomes have been sequenced, and only a few comparisons have been made between them. My results obtained from comparing tammar wallaby and opossum genomes with each other and with other mammalian groups show that marsupial genomes are more stable and have evolved more slowly
than members of other mammalian groups with the same divergence time. It would be particularly interesting to sequence groups of marsupials that have undergone rapid chromosomal change (e.g. the rapidly differentiating subspecies of rock wallabies, genus Petrogale) or the dasyurids that have undergone rapid speciation without gross chromosome change.

Obviously it would be a great advantage to sequence the genomes of more marsupials, including species that represent all the families, as well as improving the sequence depth of tammar wallaby sequencing. This will provide more data to test the hypothesis that the marsupial genome is unusually stable. Is this genome stability marsupial-wide or restricted to some families? How has it been affected by differences in the lifestyle of marsupials? For instance, do the Tasmanian devil (as the largest extant carnivore marsupial) and wombat (a vegetarian that spends most of its life underground) have different requirements for their sense of smell in their lifestyles? Are adaptations reflected in their genomes by having varied (possibly smaller) ORG repertoires?

Sequencing more marsupials will also have some conservation applications by finding the genomic basis of marsupial specific diseases (like the Tasmanian Devil’s facial tumour). Improving the annotation of the tammar wallaby genome and that of platypus (the only sequenced monotreme) will increase the quality and the depth of mammal-wide comparisons for genome evolution studies. The Tasmanian devil genome is now undergoing sequencing, and isolating its ORG repertoire with location of the genes and clusters will provide more opportunity to test the stable genome hypothesis.

Finally, if genome stability was an adaptive answer to the environment in which the first marsupials evolved, what are those ecological elements that shaped a genome that would stay unchanged during the next 166 million years? How did marsupials manage to keep such a stable genome through evolutionary bottlenecks? Reconstructing the ancestral genome of all marsupials and of each family would greatly help us in our attempt to answer these fundamental questions and this is achieved only by comparing more marsupial genomes.
REFERENCES


Kemkemer, C., M. Kohr, D. Cooper, L. Froenicke, J. Hogel, H. Hameister & H. Kehrer-Sawatzki (2009). Gene synten comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. *BMC Evolutionary Biology*, 9, 84.


odorant receptor allele telomeric to the human major histocompatibility complex. *Genet Test*, 12, 481-6.


APPENDIX 1: THE OLFACTORY RECEPTOR GENE FAMILY OF MARSUPIALS

Authors:

Delbridge, M.L., Mohammadi, A., and Graves, J.A.M.

Current status of paper (circle as appropriate):

Planned/ In Preparation / Submitted/ Under Revision/ Accepted [Published]

Date paper accepted for publication or anticipated date of publication:

July 2010

Name of journal/book:


Extent to which research is my own:

I prepared figures 1 and 3 and table 1 and I contributed toward reviewing the literature.

My contribution to writing the paper:

I wrote the first draft for the review, which the first author referred to in compiling the final manuscript.

If paper not yet accepted, has the paper been rejected by any journals:

Not Applicable

Comments:

This manuscript was an invited contribution to a book on ‘Marsupial Genetics and Genomics’.
Chapter 21
The Olfactory Receptor Gene Family of Marsupials

Margaret L. Delbridge, Amir Mohammadi, and Jennifer A. Marshall Graves

Abstract Olfaction in vertebrates is mediated mainly by a large family of olfactory receptors in the olfactory epithelium that belong to the superfamily of G protein-coupled receptors. Olfactory systems are well conserved among vertebrates, including marsupials, but there is a large variation in the numbers of olfactory genes in different animals. Most marsupials are nocturnal so depend on their sense of smell to locate food, avoid predators and identify potential mates in similar ways to other mammals. The olfactory bulbs are quite large in adult marsupials, suggesting that the sense of smell is very important in these animals. In addition, very undeveloped newborn marsupials have the special challenge of locating the pouch unassisted. It is likely that these newborns utilise their sense of smell for this unique pouch-finding task. The olfactory system is one of the few systems that is sufficiently developed in newborn marsupials to accomplish the task of finding the pouch. The opossum OR repertoire of one marsupial, the American opossum, is one of the largest characterised in mammals so far, containing over 1,500 genes. Interestingly comparisons of the opossum OR repertoire with the repertoire in an Australian marsupial, the tammar wallaby, suggests that a large conserved OR repertoire may be a feature of marsupials. The OR repertoires of the two marsupials show a high degree of similarity in total gene numbers and range of genes. This is unlike placental mammals, where the OR repertoires show a greater range. Results from these comparisons provide evidence for both the major forces (adaptation and genomic drift) behind the “birth-and-death” theory for the evolution of OR genes.

Keywords Olfactory receptor genes · Marsupials · Olfaction

M.L. Delbridge (✉)
ARC Centre of Excellence for Kangaroo Genomics, Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia
e-mail: margaret.delbridge@anu.edu.au

21.1 Introduction

21.1.1 The Importance of Olfaction

The sense of smell is so sensitive that it can detect and discriminate thousands of odours at very low concentration. Most animals rely on their sense of smell to find food and water, and avoid environmental dangers such as fire and predators. Animals use body odours to convey information about their identity, sex, health, and reproductive status to other individuals, and signal their home territories to other animals of the same and different species. Mammals also utilize their sense of smell to identify individuals and the mating readiness of other animals (Knapp et al., 2006). Olfaction therefore, is one of the most important senses animals utilize to learn about their environment.

21.1.2 Structure of the Olfactory Epithelium

Olfaction is mediated by the olfactory receptors that are located in the cilia of olfactory sensory neurons in the olfactory epithelium of mammals. Olfactory sensory neurons are bipolar neurons that span the olfactory epithelium and terminate externally in a number of fine cilia. Like other epithelial cells, olfactory sensory neurons have only a very short life and are constantly being regenerated from the basal cells of the olfactory epithelium approximately every 30 days (Moulton, 1974).

Odorant molecules become immersed in a layer of mucus, then bind to olfactory receptor proteins on the surface of the cilia. Binding of an odour molecule to the olfactory receptor activates the coupled G protein inside the olfactory sensory neuron. This event activates the conversion of ATP to the second messenger, cyclic AMP (cAMP), in the cytosol. cAMP opens up ligand-gated sodium (Na+) channels, reduces the potential across the plasma membrane, and produces an action potential to be sent to the olfactory bulb in forebrain (reviewed by Menini et al., 2004).

21.1.3 The Organization of the Olfactory System

One of the features of the olfactory system in all animals from insects to mammals is the specificity of the olfactory receptor cells. It is a widely held view that most olfactory neurons express only one olfactory receptor gene, whereas all the other olfactory receptor genes in that neuron (over 1,000 in some mammalian species) are silent. Initial mapping of receptors in the olfactory epithelium suggested that olfactory neurons expressing a single type of receptor were distributed in one of four anterior-posterior bands or expression “zones” in the olfactory epithelium (Vassar et al., 1993). More recently it has been shown that single types of olfactory receptors are not confined to these four zones, but are found in multiple overlapping zones in the olfactory epithelium. Therefore a given olfactory receptor type is generally
confined to a zone of expression, but many different olfactory receptor types are expressed in overlapping areas within a zone (Iwema et al., 2004; Miyamichi et al., 2005) (Fig. 21.1).

The olfactory bulb is organised into discrete regions called glomeruli, of which there are approximately 2,000 in mice (Menini et al., 2004). The position of each glomerulus in the olfactory bulb is roughly conserved between individuals, and each glomerulus represents the gathering together of the axons of all olfactory sensory neurons that express the same kind of olfactory receptor (Fig. 21.1) (Menini et al., 2004). The olfactory bulb is symmetrical around the midline, and so each olfactory sensory neuron converges to one or two glomeruli on each side of the olfactory bulb (Mombaerts, 2006). Therefore the olfactory receptor types are represented spatially by the organization of the glomeruli in the olfactory bulb. Mitral cells deliver olfactory signals from one glomerulus to multiple clusters of cortical neurons in several areas of the olfactory cortex, so that the spatial organization of the olfactory epithelium and olfactory bulb is not reflected in the organization of the olfactory cortex. The olfactory receptor can detect multiple odour molecules with different affinities,
and a single odour is detected by multiple olfactory receptors. The combination of olfactory receptors reacting to a single odour and the transmission of signals from multiple glomeruli to similar areas in the olfactory cortex results in the processing and recognition of complex odour signals (Malnic et al., 1999; Menini et al., 2004).

The olfactory receptor expressed by the neuron has been shown to have a role in the specification of the neurons for particular glomeruli. Rather than a particular type of olfactory receptor-containing neuron “targeting” a specific glomerulus, one model proposes that during development it is the coalescence of the axons of olfactory sensory neurons in a self-sorting mechanism that results in the formation of glomeruli. This self-sorting mechanism is perhaps mediated by interactions between olfactory receptor protein fragments located within the axon growth cones. This model is not inconsistent with a mechanism that would allow constant renewal of the olfactory neurons in the olfactory epithelium. The processes by which the complex connections between the olfactory neurons, glomeruli and regions of the olfactory cortex are laid down during development and throughout adult life are still not well understood (reviewed by Mombaerts, 2006).

In addition the evidence in support of the “one receptor-one neuron” hypothesis is not rigorously conclusive (Mombaerts, 2004), but there are only a limited number of exceptions that have been reported (Rawson et al., 2000; Sato et al., 2007; Tian and Ma, 2008). A variation on this hypothesis suggests that each neuron expresses just a few olfactory receptors, perhaps zero, one or two (Mombaerts, 2004). Neurons that express zero receptors would be negatively selected, neurons expressing one or more olfactory receptors would be positively selected. Neurons that express two or more olfactory receptors may persist only if there is no conflict between the expressed receptors; otherwise they will be selected against. This hypothesis proposes that immature olfactory sensory neurons would be likely to express more than one olfactory receptor with differing levels of expression, and that down-regulation of all other olfactory receptors would occur following coalescence of an axon into a glomerulus. This hypothesis also proposes that an olfactory sensory neuron can switch between the olfactory receptors that it expresses, so that it expresses different olfactory receptors sequentially during its lifespan (reviewed in Mombaerts, 2004). The choice of olfactory receptor expression in the olfactory sensory neurons is therefore proposed to be governed by selective forces similar to those proposed for the choice of alleles expressed by T and B lymphocytes (reviewed in Cedar and Bergman, 2008).

21.2 Olfactory Receptor Genes

The olfactory epithelium of vertebrates has a well conserved structure (Stoddart, 1980; Ache and Young, 2005). Olfactory receptors are a subfamily of the G protein-coupled receptor (GPCR) family. Olfactory receptors were first identified by their expression in the olfactory epithelium of the rat (Rattus norvegicus) (Buck and Axel,
Like other members of the GPCR family, they are transmembrane proteins, with seven α-helical transmembrane domains composed of 20–28 hydrophobic amino acids each (Feingold et al., 1999). The third, fourth and fifth transmembrane domains, and the third extracellular loop are the most variable regions of the olfactory receptor protein, and are suggested to be the regions that play a role in binding to external odour molecules (Ngai et al., 1993; Pilpel and Lancet, 1999; Irie-Kushiyama et al., 2004). The first, second, sixth and seventh transmembrane domains, and the second intracellular loop of the olfactory receptors contain highly conserved motifs, that are proposed to be involved in protein-membrane interactions, and may initiate the intracellular signaling pathways to the brain (Malnic et al., 2004).

Olfactory receptor proteins are encoded by genes of the olfactory receptor (OR) gene family. This gene family is the largest in vertebrates, having between 700 and 1,500 members in different mammals, and accounting for 3–5% of all genes in the genome (Ben-Arie et al., 1993; Glusman et al., 1996). The vertebrate OR gene family has no homology to the olfactory genes of invertebrates, despite the similarity in the nervous pathways to the brain and similarities in the patterns of evolution of each gene superfamily (reviewed in Sanchez-Gracia et al., 2009). Vertebrate OR genes are short, intronless sequences, approximately 1 kb in length, that are organized in clusters and are present on almost all chromosomes. There is a large variation in the number of OR genes and the fraction of pseudogenes identified within a family in different species, and up to 50% of OR genes are non-functional in some species (Table 21.1).

### 21.2.1 Regulation of OR Gene Expression

The expression of olfactory receptors in the olfactory epithelium is not restricted to functional genes. There is widespread expression of human olfactory receptor pseudogenes in the olfactory epithelium (Zhang et al., 2007). It has been suggested that this indicates that the choice of olfactory receptor to be expressed in a given neuron is largely stochastic, whereby different OR genes are expressed and the expression is stabilised only once a functional gene is expressed (reviewed in Fuss and Ray, 2009).

It has also been demonstrated that OR genes are mono-allelically expressed. The expression of the paternally and maternally inherited allele is random. As for the mammalian X chromosome, which also undergoes random epigenetic silencing, active and inactive alleles replicate asynchronously (Singh et al., 2003). The mammalian OR gene loci are enriched in long interspersed nuclear elements (LINEs) to a level similar to that of the LINE-rich X chromosome, and it has been suggested that the LINE elements at the OR loci contribute to allelic exclusion, in the same way they are proposed to act as way-stations for spreading inactivation along the X chromosome (Kambere and Lane, 2009). However, unlike X inactivation or genomic imprinting (Goldmit and Bergman, 2004) silencing of OR alleles does not
Table 21.1 Numbers of OR genes in 20 vertebrate species

<table>
<thead>
<tr>
<th>Common name</th>
<th>Species name</th>
<th>T&lt;sup&gt;a&lt;/sup&gt;</th>
<th>F&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P/T&lt;sup&gt;c&lt;/sup&gt;</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amphioxus</td>
<td>Branchiostoma floridae</td>
<td>43</td>
<td>31</td>
<td>12</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>Danio rerio</td>
<td>176</td>
<td>154</td>
<td>22</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Pufferfish</td>
<td>Tetraodon nigroviridis</td>
<td>34</td>
<td>11</td>
<td>23</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Elephant shark</td>
<td>Callorhinchus milii</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Medaka</td>
<td>Oryzias latipes</td>
<td>98</td>
<td>68</td>
<td>30</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Stickleback</td>
<td>Gasterosteus aculeatus</td>
<td>159</td>
<td>102</td>
<td>57</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Fugu</td>
<td>Takifugu rubripes</td>
<td>125</td>
<td>47</td>
<td>78</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Frog</td>
<td>Xenopus tropicalis</td>
<td>1,638</td>
<td>824</td>
<td>814</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Lizard</td>
<td>Anolis carolinensis</td>
<td>146</td>
<td>112</td>
<td>34</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Chicken</td>
<td>Gallus gallus</td>
<td>433</td>
<td>211</td>
<td>222</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Platypus</td>
<td>Ornithorhynchus anatinus</td>
<td>718</td>
<td>265</td>
<td>453</td>
<td>Niimura and Nei (2007)</td>
</tr>
<tr>
<td>Opossum</td>
<td>Monodelphis domestica</td>
<td>1,492</td>
<td>1,188</td>
<td>619</td>
<td>Niimura and Nei (2007)</td>
</tr>
<tr>
<td>Tammar wallaby</td>
<td>Macroopus eugenii</td>
<td>1,753</td>
<td>601</td>
<td>1,152</td>
<td>Moharramadi et al. (in preparation)</td>
</tr>
<tr>
<td>Cow</td>
<td>Bos taurus</td>
<td>2,129</td>
<td>970</td>
<td>1,159</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Dog</td>
<td>Canis lupus familiaris</td>
<td>1,100</td>
<td>811</td>
<td>289</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Mouse</td>
<td>Mus musculus</td>
<td>1,391</td>
<td>1,037</td>
<td>354</td>
<td>Niimura and Nei (2005)</td>
</tr>
<tr>
<td>Rat</td>
<td>Rattus norvegicus</td>
<td>1,765</td>
<td>1,205</td>
<td>560</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Macaque</td>
<td>Macaca mulatta</td>
<td>606</td>
<td>309</td>
<td>297</td>
<td>Niimura and Nei (2007)</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>Pan troglodytes</td>
<td>813</td>
<td>380</td>
<td>433</td>
<td>Niimura (2009)</td>
</tr>
<tr>
<td>Human</td>
<td>Homo sapiens</td>
<td>802</td>
<td>387</td>
<td>415</td>
<td>Niimura and Nei (2007)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Total number of genes  
<sup>b</sup>Number of functional genes  
<sup>c</sup>Number of pseudogenes and truncated genes

appear to be stable, for if the expressed allele is non-functional, then the other allele is subsequently expressed (Feinstein et al., 2004).

Local regulation of the MOR28 OR gene cluster on mouse chromosome 14 by the cis-acting H element has also been investigated as a trans-acting mechanism for the widespread regulation of OR gene expression. The H element has been proposed to interact with only one MOR28 OR gene at a time, through a physics
association of the H element with an OR gene promoter (Serizawa et al., 2003).
This association then regulates the expression of the MOR28 gene cluster. This
mechanism does not appear to be transferable as the H element on chromosome
14 does not regulate any other OR genes, either on chromosome 14 outside the
MOR28 cluster or other chromosomes (Fuss et al., 2007; Nishizumi et al., 2007;
reviewed by Fuss and Ray, 2009).

The most likely model for the determination of which OR gene will be expressed
by any particular neuron is probably stochastic. As all neurons express a functional
OR gene, there is the possibility that immature olfactory neurons switch between
expression of OR genes before the neuron has established a stable connection to
the brain. A negative feedback signal has been proposed to prevent continuation of
the olfactory receptor choice process as soon as a functional olfactory receptor is
expressed. This permits a second choice if, for example, a non-functional pseudo-
gene was expressed. Interestingly it also appears that the second choice of receptor
for expression is not random, but is restricted to the class of gene from which the
first OR gene was expressed. The negative feedback of an expressed functional OR
gene prevents the further expression of other OR genes and ensures that only a sin-
gle gene will be expressed from each olfactory sensory neuron (Shykoff et al., 2004,
reviewed in Fuss and Ray, 2009).

It is not known if the cis-regulation of the MOR28 cluster by a locus control
region (LCR) such as the H element, will be generally applicable to the regulation of
other OR gene clusters, or if this type of regulation is a consequence of the proximity
of this cluster to the T-cell receptor α locus. Multiple LCRs may be present in a
zebrafish OR gene cluster (Nishizumi et al., 2007), but the H element is the first
LCR identified so far in mammals. It remains to be seen whether multiple LCRs
and this type of cis-acting regulation also exist in other OR gene clusters. It is not
yet known whether mechanisms of silencing such as methylation and/or histone
modification have a role in the regulation of olfactory receptor choice, but it may be
that modifications such as these are involved in the feedback mechanism to stabilise
gene expression rather than be involved in OR gene switching (reviewed in Fuss and

21.2.2 Vertebrate Olfactory Receptor Gene Families

There are four major families of GPCR genes that encode olfactory receptor pro-
teins. The neurons of the olfactory epithelium in the nose express two of the gene
families; the OR and trace amine-associated receptor (TAAR) gene families. The
vomeronasal organ is a specialised region of the nasal epithelium in mammals that
is thought to be largely responsible for the detection of pheromones, and it expresses
the other two gene families; the vomeronasal receptor type 1 and 2 (V1R and V2R)
gene families. The four families of olfactory receptors are expressed in functionally
different types of neurons, and use different signal transduction pathways (Nei et al.,
2008). The TAAR receptors share sequence similarity with the OR receptors, and
also are expressed in different zones within the olfactory epithelium, and so have
been classified as a second class of olfactory receptors (Liberles and Buck, 2006). TAARs may have a role in detecting social cues and may be a separate group of receptors in the nasal epithelium responsible for detecting pheromones (Stephen, 2009).

Olfactory gene repertoires are characterised by species-specific and lineage-specific expansion and reduction of specific gene families. The mammalian vomeronasal organ seems to be specialised for the detection of pheromones, and analysis of the morphology of the vomeronasal organ and the functional repertoire of the V1R and V2R gene families suggests that the function of this organ has been greatly enhanced in the rodents. V1R genes are intronless like OR genes, but the V2R genes contain introns and their proteins have a long extracellular N-terminus (Nei et al., 2008). The mouse expresses over 150 functional V1R genes, whilst the rat genome contains over 100 V1R genes. In contrast, humans have only five V1R genes, chimpanzees have none, and the dog and cow have only 8 and 32 V1R genes respectively. Like the rodent V1R repertoire, the V2R repertoire also seems to be enhanced particularly in the rodents, consisting of over 100 intact genes in both mouse and rat (Hasin-Brumshtein et al., 2009).

In fish, which have no vomeronasal organ, there is a single V1R-like gene that is expressed in the main olfactory epithelium (Pfister and Rodriguez, 2005), but there has been a lineage-specific expansion of V2R-like genes (Hashiguchi and Nishida, 2006). This suggests that the V1R and V2R gene repertoires have undergone specific expansions in the rodent lineage, along with the development of the vomeronasal organ. Interestingly, the South American grey, short-tailed opossum (Monodelphis domestica) also has well developed V1R and V2R repertoires, consistent with having a well-developed vomeronasal organ in this marsupial mammal (Grus et al., 2005; Young and Trask, 2007). Another mammal in which there has been a lineage specific expansion of the V1R gene repertoire is the platypus. This semi-aquatic monotreme has undergone the largest lineage-specific V1R gene expansion of any mammal so far investigated, and it also has a large V2R repertoire. In contrast, the main olfactory system of the platypus seems to be somewhat reduced. This suggests that the vomeronasal system is a more important sensory organ than the main olfactory system of the nose in the platypus, which forages underwater with the eyes, ears and nostrils closed (Grus et al., 2007).

Olfactory receptors in the main nasal epithelium are encoded by the OR gene family that is found in all vertebrates. In eutherian mammals, the OR gene family can contain between 700 and 1,500 genes in different species. In contrast, the OR gene family is much smaller in fish, ranging from about 40–140 genes in different fish species. The fish OR gene family is more divergent than the mammalian OR gene repertoire, suggesting that fish might be able to detect a wider range of odorants, but that the mammalian OR repertoire might be capable of discriminating between more structurally similar odorants (Niimura and Nei, 2005). It is unlikely that the number of OR genes is directly related to the sensitivity of the sense of smell. Comparisons of olfactory abilities must also take into account the size of the olfactory epithelium, the fraction of the brain devoted to processing olfactory information, and the ability to memorize the signals from new odours.
The OR gene family is classified into two major classes: Class I and Class II. In mammals, Class I olfactory receptors represent only 10% of the total olfactory receptor repertoire, whereas Class II contains approximately 90% of the total repertoire. Class I receptors have been proposed to be responsible for detecting water-soluble odorants, whereas Class II receptors are thought to bind air-borne odorants. In support of this hypothesis is evidence that in goldfish, the only functional OR genes have homology to Class I genes, and all Class II genes are non-functional pseudogenes (Freitag et al., 1998). In amphibians olfaction occurs in two anatomically separate nasal cavities, and both Class I and Class II olfactory receptors are expressed. Class I OR genes are expressed in a water-filled nasal cavity and Class II OR genes are expressed in the air-filled nasal cavity (Freitag et al., 1995). Further evidence for the distinction between the functions of the two classes is provided by studies of the dolphin, in which a large array of Class II OR genes were found to be nonfunctional pseudogenes. It was suggested that the Class II OR genes were lost in these aquatic mammals when they evolved from land mammals and lost the ability to detect air-borne odorants (Freitag et al., 1998). It is unlikely however that there is a functional distinction between Class I and Class II OR genes. High numbers of functional Class I OR genes were found in the human and mouse genomes (Glusman et al., 2001; Zhang and Firestein, 2002), suggesting that Class I OR genes are likely to respond to additional ligands in these species, as well as water-soluble odour molecules in aquatic animals.

The blurring of the distinction between Class I and Class II OR genes has prompted a reclassification of the OR gene family in vertebrates, although the well-established Class I and Class II nomenclature is still in common usage. The availability of genome sequence data and its use for phylogenetic analysis of vertebrate OR gene families led to classification into nine groups (α, β, γ, δ, ε, ζ, η, θ), each of which had a single or small number of founding genes prior to the divergence of fish and tetrapods. The α and γ groups are the only groups that have been retained by, and greatly expanded in, the tetrapods, and these two groups represent the Class I and Class II genes respectively in mammals (Niimura and Nei, 2005; Niimura and Nei, 2006; Nei et al., 2008; Niimura, 2009).

### 21.2.3 Ectopic Expression of OR Genes

The discovery of olfactory receptor genes was based partly on the observation that they were specifically expressed in the olfactory epithelium (Buck and Axel, 1991). They are also expressed in the olfactory bulb, where they are likely to play a role in the coalescence of olfactory sensory neurons into glomeruli (Mombaerts, 2006). Several studies have shown that OR genes can be expressed in tissues other than the olfactory epithelium, which may suggest other roles for these receptors. Expression of various OR genes has been reported in the testis and germ cells, as well as a wide range of other tissues such as the tongue, erythroid cells, prostate, placenta, brain, and peripheral nervous system (see Feldmesser et al., 2006). Expression of OR genes in the testis has been proposed to play a role in sperm chemotaxis or
olfaction-driven mate choice, whereas expression of OR genes in a wide range of tissues has been suggested to indicate a role for the OR genes in cell-cell recognition (Ziegler et al., 2002; Dreyer, 1998; Fukuda et al., 2004).

Database mining and microarray experiments (Feldmesser et al., 2006; Zhang et al., 2007) have now shown that widespread expression of OR genes in non-olfactory tissues may be a consequence of transcription from leaky promoters during the decondensation of chromatin that is required for transcriptional regulation. Leaky expression of genes may be evolutionarily important by generating potential diversity (Rodriguez-Trelles et al., 2005; Feldmesser et al., 2006). However, microarray analysis of OR gene expression shows that ectopic patterns of expression of orthologous OR genes are conserved between chimpanzee and human, suggesting that they evolved more slowly because they serve a selectable function (Branscomb et al., 2000; De la Cruz et al., 2009). This is circumstantial evidence that at least some OR genes may have gained new functions in non-olfactory tissues.

### 21.2.4 Evolution of OR Genes

The OR gene family in vertebrates is characterized by lineage-specific expansions and contractions in the numbers of OR genes, described as a “birth-and-death” pattern of evolution (reviewed in Nei and Rooney, 2005). This model predicts that the composition of a gene family may be unique in different lineages. For example, in one species, particular genes may be duplicated one or more times, resulting in expansion of the family, whereas in another species the same genes may be lost or become pseudogenes, resulting in contraction of the family. For example, the most recent common ancestor of human and mouse is estimated to have had 754 functional OR genes, but there has been a lineage-specific contraction in the OR repertoire to 388 functional genes in humans and a lineage-specific expansion of the repertoire to 1,037 functional genes in the mouse. Of the 754 ancestral OR genes, 193 and 267 genes have become new non-functional pseudogenes in human and mouse, respectively, whilst humans have lost 249 of the ancestral pseudogenes that have been proposed to be part of the OR repertoire of their most recent common ancestor with mice (Nei and Rooney, 2005).

Class I and Class II OR genes are further subdivided based on protein sequence similarity. OR gene families encode proteins with more than 40% protein similarity, and OR gene subfamilies encode proteins with more than 60% protein similarity. Currently there are 18 gene families and 300 subfamilies that have been characterised in humans (Olender et al., 2008), but the publication of whole genome sequences has made it possible to characterise OR gene families bioinformatically in a number of vertebrate species. This has shown that there are large numbers of OR genes that vary widely between different species. Another hallmark of the OR gene families in vertebrates is that in addition to functional gene members, they all contain significant numbers of disrupted genes, the numbers of which also
vary widely between species. The disrupted genes have interruptions to their coding region, suggesting that they are non-functional pseudogenes.

Another feature of the OR gene family is that the genes are located in a number of clusters on nearly all chromosomes in the genome. Fish have one of the smallest OR repertoires, and it has been proposed that one of the major events in the evolution of the OR gene families was the adaptation of the olfactory system to a non-aquatic environment (Glusman et al., 2001). Early duplications in the land mammals gave rise to the major families of Class II OR genes found in all mammals. Subsequently there have been local expansions of different OR families and subfamilies to form large gene clusters (Glusman et al., 2001). This pattern of duplication means that OR genes are found at many chromosomal locations, and related OR genes are often present in the same clusters (Sullivan et al., 1996; Rouquier et al., 1998; Zhang and Firestein, 2002), many of which are conserved across species (Aloni et al., 2006).

Segmental duplications between chromosomes have also contributed to the wide distribution of OR gene clusters, resulting in the location of members of different subfamilies on different chromosomes. Other mechanisms that are likely to have contributed to the evolution of this gene family are tandem duplication, which has generated long tandem arrays of closely related genes, gene conversion and recombination which may allow conversion of non-functional to functional genes or vice versa, and retrotransposition (reviewed in Kambere and Lane, 2007). Some OR gene clusters are broken up by many interspersed non-OR genes (Olender et al., 2008), whereas other clusters contain OR genes exclusively (Glusman et al., 2001; Lane et al., 2002). The presence of related OR genes in single clusters suggests that different clusters may largely represent different subfamilies of OR genes and therefore might be responsible for detecting different types of odorant molecules. It may be that gene clusters are preserved so that all the genes within a cluster can be transcriptionally regulated by common control elements (Krumlauf, 1992).

Variation in the total numbers of functional and non-functional OR genes in the repertoires of different animals has been suggested to reflect an adaptive response to the environment. It has been suggested that there have been three major events in the evolution of the vertebrate olfactory system; first, the expansion of the olfactory system as a response to the adaptation to a terrestrial environment, secondly, the reduction in the olfactory system in the primate lineage along with the development of trichromatic vision, and thirdly the development of the pheromone system in the rodent lineage (Kambere and Lane, 2007; Olender et al., 2008). This theory of adaptive evolution is supported by the relatively few genes in the OR repertoire of all fish species (approximately 100 OR genes) compared to a 8–10 times larger number of genes in the OR repertoires of terrestrial mammals (between approximately 800 and 1,500), suggesting an expansion of the OR repertoire has occurred during the adaptation of vertebrates to a terrestrial existence (Nimura and Nei, 2006; Saraiva and Korsching, 2007; Nei et al., 2008). In addition, there are many fewer OR genes in aquatic mammals (dolphins, whales), consistent with loss of OR genes following the readaptation of these mammalian species to the water (Freitag et al., 1998). This does not explain, though, why a smaller number of fish OR genes can detect the
same range of odorants as the larger mammalian OR repertoire (Niimura and Nei, 2005).

In primates, lower numbers of functional OR genes and higher numbers of pseudogenes seems to have resulted from increased pseudogenisation of OR genes. This may have occurred when the olfactory system became relatively less important with the increasing reliance of primates on the development of trichromatic colour vision, or the development of higher brain mechanisms such as memory. However, this does not explain why cattle with dichromatic colour vision have a high proportion of pseudogenisation in their OR repertoires, or why the numbers of functional OR genes varies so much among terrestrial mammals (Nei et al., 2008).

Diversity of the OR repertoire in different animals was clearly generated by gene duplication and deletion events, then adaptation of species to different environmental niches with different requirements for olfaction. Such gene duplication and deletion events occur more or less randomly, so genomic drift may play an important role in the generation of diverse OR gene repertoires, and may wholly account for instances in which species-specific expansions and contractions are difficult to explain by adaptation (Nei et al., 2008). Random gene duplication can increase the number of genes which are then fixed by positive selection or chance (Nei et al., 2008).

If genetic drift has a significant effect on the OR gene repertoire of mammals, it would be expected that copy number variation should exist among individuals of the same species. Indeed, up to 30% of functional genes in the human and mouse OR repertoires are polymorphic for copy number (Young et al., 2008; Nozawa et al., 2007). There is also a high degree of allelic variation in human and mouse OR genes, which may account for the variation observed in sensitivity to odours observed between different individuals. Selection for a diverse repertoire can explain why the exceptionally large numbers of functional OR genes and pseudogenes are tolerated in vertebrate genomes. Gene conversion between different functional and non-functional OR copies could rapidly generate new functional genes (reviewed in Keller and Vosshall, 2008; Hasin-Brumshtein et al., 2009).

The size of OR gene families, as well as the proportions of intact OR genes and pseudogenes, varies widely among vertebrate species (Table 21.1). Different subsets of pseudogenes are found even between different human populations, or different dog breeds (Tacher et al., 2005; Rouquier and Giorgi, 2007), implying very rapid evolution.

### 21.3 The Olfactory System in Marsupials

Most marsupials are nocturnal, and forage alone or in small groups at night, although some marsupials, such as the kangaroos, may continue to forage at either end of the day. During the night therefore, the auditory and olfactory senses are more important than vision for these animals. The relative importance of the sense of smell for the marsupials is confirmed by the observation that all marsupials have
a relatively large olfactory bulb. Like most other vertebrates, marsupials use their sense of smell to find food, avoid predators, and to identify animals and marked territories of their own and other species. This is particularly important for species that live a solitary existence and meet only to mate (reviewed by Croft and Eisenberg, 2006).

A unique feature of marsupial reproduction is the birth of the young at a very early developmental stage. The tiny newborn marsupial young manages to move unaided from the urogenital opening to the pouch some distance away, where it attaches to a teat (reviewed in Pask and Renfree, Chapter 14). At birth, the eyes and ears of the pouch young are very undeveloped, but the olfactory system is relatively well developed, suggesting that both gravity and the sense of smell guide the newborn marsupials to the pouch (Gemmell and Rose, 1989; Hughes et al., 1989).

21.4 Olfactory Apparatus of Marsupials

Exactly which signals guide the newborn marsupial to the pouch is still unclear, and the relative importance of different stimuli may be different for different marsupial species. At birth, the olfactory epithelium of a number of marsupials, including the brush-tailed possum (Trichosurus vulpecula), rat-kangaroo (Potorous tridactylus) and quoll (Dasyurus hallucatus), is well developed. Olfactory receptor neurons with cilia are observed in the olfactory epithelium of the newborn possum, and may be capable of responding to odours. However, the apparent maturity of the olfactory epithelium contrasts with the morphological immaturity of the olfactory bulb in these species (Gemmell and Rose, 1989; Gemmell and Nelson, 2004). Indeed in both the opossum (M. domestica) and the tammar wallaby (Macropus eugenii), the olfactory bulb appears to be relatively undeveloped at birth, and there is no evidence for the presence of glomeruli (Ashwell et al., 2008; Schneider et al., 2009) until approximately 15 days after birth in the opossum and 25 days after birth in the tammar wallaby. It is not until approximately 54 days after birth that the olfactory bulb seems to be fully developed in the tammar wallaby (Ashwell et al., 2008). In the tammar wallaby, the vomeronasal organ, which is responsible for the detection of pheromones, is very immature in appearance at birth and does not have neurons extending to the olfactory bulb until day 5 after birth (Ashwell et al., 2008).

The olfactory system of most marsupials undergoes further development and differentiation during the first 25–40 days of pouch life. Although the olfactory epithelium of marsupials appears to be functional at birth, it has been suggested that the olfactory bulb and signaling pathway to the brain is too immature at birth to allow the newborn to detect odour and use this sense to locate the pouch (Gemmell and Rose, 1989; Ashwell et al., 2008). However, behavioural studies have shown that newborn tammar wallabies are attracted toward the smell of the mother’s pouch, indicating that these newborns have a functional olfactory system (Schneider et al., 2009). This suggests that although the olfactory system is not fully developed at birth, it may be sufficiently developed to contribute to finding the pouch.
21.5 Marsupial OR Repertoire

The koala (*Phascolarctos cinereus*) was the first marsupial from which OR genes were isolated (Glusman et al., 2000). Twenty-five OR genes were identified, all of which belonged to the class II OR genes, and five of which were pseudogenes. From these limited numbers, it was estimated that marsupials must have a very small OR repertoire of up to 50 genes (Glusman et al., 2000).

The availability of genome sequence of the opossum (*M. domestica*) provided the first opportunity to isolate almost the whole OR repertoire from a marsupial. Surveying the first genome assembly of the opossum resulted in the detection of 1,518 OR genes, more than 60% of which are functional (Aloni et al., 2006). Somewhat different estimates have since been reported for the opossum OR gene repertoire, ranging from 1,492 genes of which 80% are functional (Niimura and Nei, 2007), to 1,548 genes of which 77% are functional (Kishida, 2008). However, these differences are small and mostly attributable to different methodologies and/or different versions of the assembled genome being used.

The availability of genome data from the opossum and the platypus enables a more complete account of the evolution of OR gene repertoire. Examination of the OR repertoires of several mammalian species reveals several major duplications during vertebrate evolution that led to differences in the total number of OR genes in different lineages. In the monotremes, as well as amphibians, birds and fish, there are fewer than 1,000 OR genes, suggesting that the OR repertoire was relatively small prior to the mammalian radiation. The large OR repertoire in the opossum implies that there was a sudden burst of OR gene duplication and diversification between the divergence of the monotremes from Theria (marsupials and eutherians), and the divergence of marsupials from eutherians. During this short period (between 166 and 148 million years ago, Bininda-Emonds et al., 2007), there was a gain of approximately 350 genes to increase the total number of OR genes in the most recent common ancestor of the therian mammals from 152 to 492. A further 350 genes were gained in the eutherian lineage between 145 and 105 million years ago, between the divergence of the marsupials from the Eutheria and prior to the radiation of eutherian mammals. There was a remarkable gain of approximately 750 genes in the marsupial lineage, and further significant increases (up to 1,500) in the number of genes then occurred in the OR repertoires of the rodent lineages, whilst significant decreases (down to about 300) in the number of genes occurred in the primate lineage (Niimura and Nei, 2007) (Fig. 21.2).

Most of the expansion in the marsupial lineage can be accounted for by expansion in a small number (four or five) of gene lineages. The majority (80–90%) of the mammalian OR repertoire is made up of Class II OR genes, leaving only 10–20% represented by the Class I OR genes. Niimura and Nei (2007) divided the Class II genes into 34 subgroups by constructing phylogenetic trees, so that each clade contained genes that had the greatest sequence similarity both within and between species. They found that the representation of the Class II subgroups can be quite different between species (Niimura and Nei, 2007). Some Class II subgroups appeared to be quite stable throughout mammalian evolution,
Fig. 21.2 Evolutionary changes in functional gene numbers in mammalian OR gene repertoires. The numbers of OR genes in each species is given in each box. “+” represents a gain of approximately 150 genes and “−” represents a loss of approximately 150 genes. Divergence times are estimated in millions of years ago (mya) (Bininda-Emonds et al., 2007). There has been significant expansion in OR gene numbers in the most recent common ancestor of the Theria and the Eutheria, as well as lineage specific expansion in the marsupials and rodents. Generally, there have been fluctuations in the numbers of OR genes in separate lineages following the eutherian radiation, including significant amounts of both gain and loss of genes from the OR gene repertoires. Adapted from Niimura and Nei (2007).

containing approximately the same numbers of genes in different species. However, other Class II subgroups contained vastly different numbers of genes in different species. The expansion in the opossum lineage can be accounted for mostly by expansion in four OR Class II groups, presumably because these OR genes encode receptors that detect odours that are important for opossums. Interestingly, some of the same subgroups are expanded also in the rodent lineages. This analysis provides valuable insight into the gains and losses of OR genes that have occurred during mammalian evolution, but the function of each of the subgroups remains unclear. Further investigation is needed to ascertain whether there is a relationship between these subgroups and their function and/or genomic location. Additionally there is limited information about which ligands elicit a response from which OR gene families, so it cannot be determined, for example, whether the same OR gene family in marsupials and rodents is responding to the same ligand (Niimura and Nei, 2007).

Information from marsupials also makes it possible to interpret the evolution of the TAAR family of OR receptors in the main olfactory epithelium, which detect amines in the urine that function as sex pheromones in the mouse (Liberles and Buck, 2006). In all tetrapods the TAAR family is located in a conserved cluster on
human chromosome 6, mouse chromosome 10 and opossum chromosome 2. Only one or two TAAR genes are found in the frog and the chicken, suggesting that these genes encode less important odour receptors in amphibians and birds. A similar pattern of lineage-specific gain and loss of genes is observed in this gene family. Subfamily I of the TAAR family is expanded in the opossum as well as the mouse (Hashiguchi and Nishida, 2007), which is consistent with the detection of amines in the urine being important for the identification of oestrus in female opossums (Zuri et al., 2003).

The repertoire of opossum OR genes also shows that this pattern of lineage-specific gain and loss of genes has occurred to a much greater degree in the other pheromone receptor genes, the V1R and V2R genes of the vomeronasal epithelium. The numbers of V1R and V2R genes in opossum is similar to those of the rodents, whereas there has been a loss of these genes in primates, dog and cow (Grus et al., 2007; Young and Trask, 2007). This is consistent with the well-developed vomeronasal system in the opossum, which has been suggested to relate to the reliance of marsupials, as well as rodents, on the detection of pheromones, and the loss of this ability in other mammals. The platypus has the largest V1R and V2R repertoires of any mammal, which may have been selected for because of its reliance on these receptors underwater, and the orientation of its vomeronasal organ to open into the oral cavity rather than the nasal cavity as is usual for most other mammals (Grus et al., 2007).

Another remarkable feature of the V1R and V2R repertoires is their rapid and extreme expansions within mammal lineages. This is evident from phylogenies of V1R and V2R genes that form lineage-specific clades, in which genes within a species are most closely related and there are very few genes with one-to-one orthology between different species. This means that the evolution of these genes has been much more rapid than that of the OR genes of the main olfactory epithelium (Grus et al., 2007; Grus and Zhang, 2008), in which orthologous clusters of OR genes can still be recognised in different species (Aloni et al., 2006).

Data from the opossum and tammar genomes also shows that as well as an increase in the total number of genes in the marsupial lineage, there has been remarkable conservation of the marsupial OR repertoire. In almost all terrestrial marsupials and eutherians whose genomes have been sequenced, the Class I genes consistently represent between 12 and 18% of the total OR repertoire, and most of them are functional. This implies that the Class I genes have been functional since before the divergence of the marsupials from the eutherians 148 million years ago. The size and the composition of the OR repertoire is well conserved between the tammar wallaby and opossum, indicating that the marsupial OR repertoire has been conserved for more than 80 million years (Fig. 21.3a). This extent of conservation is unusual; it is not found between any two eutherian mammals and might be the result of some functional constraints in marsupials resulting from similar biological needs.

The OR51 family is an example of a Class I family that appears to have undergone lineage-specific expansion in the therians. Within family OR51 there is almost a one-to-one relationship between the tammar and opossum OR genes of this family (Fig. 21.3b). This suggests that the expansion of this family occurred in the
Fig. 21.3 (a) A comparison of the composition of the total OR gene repertoires in mammals. Numbers next to the sections represent the family number. Class I families are represented in the excised portion of the pie charts. Comparisons between species suggest there has been a loss of family OR14 genes in eutherians, expansion of class I genes (especially family OR51) prior to the divergence of the therians, and a high level of similarity between two marsupial species. (b) The phylogenetic relationship between 125 genes of the OR51 family from the opossum and tammar wallaby. A class II OR gene is used as outgroup. Bootstrap values are shown for important nodes. In most cases there is a one-to-one relationship between the opossum and tammar wallaby genes. Absence of species-specific expansions and/or contractions in this family indicates an important function for this family in marsupials. Similar sized human and marsupial OR51 families suggest that there was an expansion of this family between the divergence of the monotremes 166 mya, and the divergence of the marsupials from the eutherians 148 mya (Mohammadi et al., in preparation).
marsupial lineage prior to the divergence of these two marsupials approximately 80 million years ago (Mohammadi et al., in preparation). Similar numbers of OR51 genes are found in humans and marsupials, whereas low numbers of family OR51 are found in platypus, suggesting that the expansion of the family occurred in the therian lineage following its divergence from the monotremes, between 148 and 166 million years ago. Comparisons with an outgroup, such as the counterpart of the OR51 family in another vertebrate, would be necessary to confirm the expansion of this lineage in therians. The analysis of therian Class I OR genes such as this may be an opportunity to examine the role of olfactory receptors in detecting airborne molecules and/or gaining new functions in land-living mammals.

Data from the opossum and tammar genomes also shows that as well as an increase in the total number of genes, the diversity of the OR genes has increased since the divergence of Theria from the monotremes. There are only 130 subfamilies in the platypus genome, compared to 240 OR gene subfamilies in the opossum and 286 in the tammar wallaby. These duplications have affected mostly the gene families OR8, OR11, OR13 and OR51, which have expanded up to 20 times compared to their counterparts in platypus. Further analysis including a vertebrate outgroup will confirm whether these expansions are specific to therians, or whether the contrast between platypus and therian OR family sizes are the result of lineage specific loss of genes in the platypus (Mohammadi et al., in preparation).

In contrast, the Class II OR14 gene family may have undergone a significant loss of genes during therian evolution; the platypus genome contains more than 250 members of the OR14 gene family, but only one third of them have been retained in marsupials, and this family has been completely lost from the eutherian genome. As all the OR14 genes of platypus belong to only six subfamilies, this implies coincidental tandem duplications of a few ancestral genes. Further analysis using a vertebrate outgroup will confirm whether this is the case, or if there has been an expansion of these Class II genes in platypus (Mohammadi et al., in preparation) (Fig. 21.3a).

21.6 Conclusion

The sense of smell is an important means by which vertebrates receive information from the environment and communicate with each other. The olfactory system responds to airborne odour molecules and pheromones. One important area of research in this field examines how the connections between the olfactory epithelium, olfactory bulb and olfactory cortex of the brain are established and regulated. Olfactory neurons contain positional information within in the epithelial layer, they express one or a few olfactory receptors, and their axons can coalesce with other neurons of the same type to form glomeruli in the olfactory bulb. Genomic imprinting mechanisms may play a role in the regulation of olfactory gene expression influencing the construction of the olfactory system as well as detection of odour molecules, so the investigation of epigenetic regulation in marsupials is likely to
provide valuable insight into the regulation of olfactory genes as it has other gene systems in mammals (reviewed in Hore et al., Chapter 12). Marsupials are ideal organisms in which to study the development of the olfactory connections through the olfactory bulb to the olfactory cortex, as they are born at an undeveloped stage and a significant proportion of the complexity of olfactory bulb development takes place in the pouch.

The marsupial genome harbours one of the largest mammalian OR repertoires, containing up to 1,500 olfactory receptor genes, that respond in combinatorial codes to odour molecules to allow marsupials to process and recognize an almost unlimited range of odours. Comparisons between the repertoires of different animals has suggested that different families and subfamilies of OR genes have been expanded in response to the adaptation of animals to different environments. In marsupials there has been expansion of both the OR repertoire of the main olfactory epithelium and the V1R and V2R repertoires of the vomeronasal organ, similar to that observed in rodents. This suggests that the sense of smell and detection of pheromone signals are at least as important to marsupials as they are to rodents, perhaps due to marsupials nocturnal habits and unique reproductive biology. These unique evolutionary adaptations of the marsupials make the characterisation of the marsupial OR repertoire essential to any comprehensive study of the olfactory system.

References


APPENDIX 2: PRESENTATIONS

Oral Presentations:


Poster Presentations:

