RATTINI (RODENTIA, MURINAE) SPECIES RELATIONSHIPS AND INVOLVEMENT AS RESERVOIRS FOR SCRUB TYPHUS: A COMPARATIVE MOLECULAR CYTOGENETIC AND GENE EXPRESSION APPROACH

By
Daleen Badenhorst

Dissertation presented for the Degree of Doctor of Philosophy at the University of Stellenbosch

Supervisor: Prof Terence J. Robinson
Co-Supervisor: Dr Gauthier Dobigny
Faculty of Science
Department of Botany and Zoology

December 2011

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2011

ABSTRACT

The *Rattus sensu lato* complex, which harbours the important biomedical model species *Rattus norvegicus*, represents a group of rodents that are of clinical, agricultural and epidemiological importance. This study presents a comprehensive comparative molecular cytogenetic investigation of this complex. Karyotypes of 11 *Rattus s. l.* species and *Hapalomys delacouri* were analysed using conventional cytogenetic techniques. These data suggest that, contrary to previous claims, the usefulness of karyotypes for identifying these Asian murid species is limited, as few species-specific chromosomal characters could be identified. In order to understand the events that shaped the morphology and composition of the extant karyotypes, nine out of the 11 *Rattus s. l.* species were analysed by chromosome painting. This allowed the detection of a rare convergent sex-autosome translocation present in *H. delacouri* and *B. savilei*, whose establishment is hypothesised to have been favoured by the presence of interstitial heterochromatic blocks (IHBs) at the boundaries of the translocated segments. These results indicate that *Rattus s. l.* is characterised by slow to moderate rates of chromosome evolution in contrast to the extensive chromosome restructuring identified in most other murid rodents. Based on these data the first comprehensive putative Rattini ancestral karyotype was constructed. Their integration with published comparative maps enabled a revision of the previously postulated ancestral murid karyotype. BAC-mapping unequivocally demonstrated that the widely reported polymorphisms affecting chromosome pairs 1, 12 and 13 in the *Rattus* karyotypes are due to pericentric inversions. The analysis of genomic features, such as telomeres, Ag-NORs and satellite DNA suggest a constrained pattern of chromosome evolution. The investigated rat satellite I DNA family appears to be taxon (*Rattus*) specific, and of recent origin (consistent with a feedback model of satellite evolution).

The comparative nature of the study led to the further analysis of the species within *Rattus s. l.* complex and their possible involvement as reservoirs of scrub typhus using a qPCR gene expression approach based on real-time PCR. The structure and transcription of syndecan-4, which had previously been linked to *Orientia tsutsugamushi* (causative agent of scrub typhus) infection, was compared among Rattini (typhus-positive) and Murini (typhus-negative) rodents. Although, it was not possible to conclusively link the structural variation observed between Rattini and Murini with carrier status, a link was identified between underexpression of syndecan-4 in Murini and seropositive Rattini rodents, compared to those that were

seronegative. This suggests that the reduced levels of syndecan-4 transcription in Murini and Rattini is linked to the poor carrier status of Murini, and to increased longevity of seropositive Rattini (i.e., predominant host of *Orientia*), highlighting aspects that need further investigation.

OPSOMMING

Die *Rattus sensu lato* kompleks, wat die belangrike biomediese model spesie *Rattus norvegicus* huisves, bevat 'n groep knaagdiere wat van kliniese, landbou, en epidiomologiese belang is. Hierdie studie bevat 'n omvattende sitogenetiese molekulêre vergelyking binne hierdie kompleks. Die kariotipes van 11 *Rattus s. l.* spesies en *Hapalomys delacouri* is ondersoek met die gebruik van konvensionele sitogenetiese tegnieke. Die data dui daarop, in kontras met vorige bevindinge, dat die bruikbaarheid van kariotipes om Asiese knaagdiere te identifiseer beperk is, aangesien min spesies-spesifieke chromosoom merkers geidentifiseer kon word. Om die gebeure wat tot die morfologie en komposisie van die huidige kariotipes gelei het, te verstaan, is nege van die 11 *Rattus s. l.* spesies met behulp van chromosoom fluoressente hibridisasie ondersoek. Dit het die ontdekking van 'n rare konvergente geslagschromosoom-outosoom translokasie in *H. delacouri* en *B. savilei* tot gevolg gehad. Die vaslegging hiervan is heel moontlik bevoordeel deur die teenwoordigheid van interkalerende heterchromatien blokke (IHBs) by die grens van die translokeerde segmente. Hierdie resultate dui daarop dat *Rattus s. l.* deur 'n stadige tempo van chromosoom verandering gekenmerk word, wat in skrille kontras staan met die hoë aantal chromosoom herrangskikkings wat in meeste ander murid knaagdiere geidentifiseer is. Vanuit hierdie data kon die eerste oerouer kariotipe van die Rattini bepaal word. Die gebruik van hierdie data in kombinasie met beskikbare vergelykende kaarte het dit moontlik gemaak om die vorige hipotetiese oerouer murid kariotipe te hersien. BAC-kartering het dit moontlik gemaak om sonder twyfel vas te stel dat die polimorfisme wat chromosoom pare 1, 12 en 13 in die *Rattus* kariotipe affekteer, die resultaat van perisentriese omsetting is. Die analise van genomiese eienskappe, soos telomere, Ag-NORs en satelliet DNA dui op 'n beperkte patroon van chromosoom evolusie. Die bestudeerde rot satelliet DNA familie blyk takson (*Rattus*) spesifiek te wees met 'n onlangse oorsprong, wat ooreenstem met die terugvoer model van satellite-evolusie.

Die vergelykende aard van hierdie studie het gelei tot die verdere analise van die spesies in die *Rattus s. l.* kompleks in terme van hul moontlike rol as draers van "scrub typhus" deur gebruik te maak van qPCR geen uitdrukking, wat gebasseer is op "real-time" PCR. Die struktuur en transkripsie van syndecan-4, wat in die verlede aan *Orientia tsutsugamushi* (veroorsakende agent van scrub typhus) infeksie gekoppel is, is tussen Rattini (typhus-positief) en Murini (typhus-negatief) knaagdiere vergelyk. Ten spyte daarvan dat dit nie

moontlik was om die strukturele variasie tussen Rattini en Murini met draer status te koppel nie, is daar wel 'n skakel tussen die verlaagde uitdrukking van syndecan-4 in Murini en sero-positiewe Rattini knaagdiere, in vergelyking met die wat sero-negatief was, gevind. Dit stel voor dat die verlaagde vlakke van syndecan-4 transkripsie in Murini en Rattini aan die swak draer status van Murini, asook die verhoogde langslewendheid van sero-positiewe Rattini (i.e., oorheersende gasheer van *Orientia*), gekoppel is. Hierdie bevindinge beklemtoon sekere belangrike aspekte vir verdere navorsing.

ACKNOWLEDGEMENTS

There are a number of persons and institutions to which I am greatly indebted. Firstly, I wish to thank my principal supervisor, Prof. Terry Robinson for his faithful encouragement, guidance, and patience during the past three years, as well as with the preparation of this manuscript. I have benefited greatly under his tutelage and believe this will hold me in good stead in future endeavours. I am sincerely grateful to Terry for affording me with the opportunity to thrice travel abroad (twice to France and once to Portugal) where I was able to work on certain aspects of this study and benefit from working with individuals who are experts in their respective fields. This was made possible through a joint research grant awarded under a France/South Africa scientific cooperation agreement to T.J Robinson and G. Dobigny. I am also very obliged to Terry for sponsoring my participation at the Zoological Society of Southern Africa's (ZSSA) meeting in 2009, as well as the International Conference on Rodent Biology and Management (ICRBM) that was held in Bloemfontein in April 2010 where I had the opportunity to present my results in the form of oral presentations.

I would also like to warmly thank my co-supervisor, Dr Gauthier Dobigny who was always available to answer any questions and queries, as well as give advice at any time, even though we were separated by 1,000s of kms. In addition, I would like to thank him for hosting my two visits to his laboratory in Montpellier, France (Centre de Biologie pour la Gestion des Populations), and for making my French experience such an enjoyable one. I am also thankful to Gauthier for visiting South Africa in September 2010 and being willing to spend an intensive week with me while assisting with certain aspects of this dissertation. His insightful comments and understanding has been more helpful than he may think.

I am also very obliged to Drs Raquel Chaves and Filomena Adega who hosted my visit to their laboratory (Institute for Biotechnology and Bioengineering, Centre of Genetics and Biotechnology of the University of Trás-os-Montes) in Vila Real, Portugal. Their boundless enthusiasm was infectious and I am truly grateful for their guidance, patience, hospitality and for making me feel so welcome.

I would also like to thank my mentor-friends-colleagues Clement Gilbert, Caroline Tatard, Marie Pagès, Paul Waters, Adriaan Engelbrecht, Leigh Richards and Victor Rambau who offered their time, knowledge and enthusiasm. I am much indebted to Clement who

introduced me to the wonderful world of cytogenetics, cell culturing, G- and C-banding, and FISH. I would like to thank Caroline and Marie for their assistance, warmth and hospitality during my time in France. I am also very obliged to Caroline for her help with the gene expression aspects and for allowing me to benefit from her technical expertise. Thanks must go to Paul for his help with the BAC clone aspects of this study before, during and after a two week visit to South Afria in 2009. I am truly grateful to Adriaan and Victor for all their help over the years, as well as their companionship. They, together with Leigh, helped make all the hours spent in the lab or in the microscope room very pleasant ones.

I would also like to thank all my lab mates and colleagues from the EGG (Evolutionary Genomics Group) lab, who helped make these past three years truly enjoyable. There are too many to list, but I would like to especially thank Nina du Toit (who also kindly helped translate the summary into Afrikaans), Ann Mclachlan, Rina Groenewald, Minette Karsten, Petra Ros, Savel Daniels, Sophie Von der Heyden, Prince Kaleme, BettineVan Vuuren and Nico Solomon.

Finally I would like to thank all my family and friends who have supported me unwaiveringly over the years and have always shown a great interest in what I was doing. Special thanks must go to my parents, Pieter and Mareva, for their loving support and encouragement throughout my career. Thank you for always believing in me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENERAL INTRODUCTION

RODENTS AS RESERVOIR HOSTS OF DISEASE

The number of new "human" diseases linked to rodents that act as reservoirs has increased significantly in the past few years leading to a revival of interest in reservoir ecology (Mills and Childs, 1998; Vorou *et al.*, 2007), and to the phylogenetic relationships of these rodents (Jansa and Weksler, 2004). The importance of understanding host and vector interactions and their applications in human health was recognized as early as the 1900s (Mills and Childs, 1998). Rodents play an important role in the natural life cycle of several zoonoses as they can harbour a disease without the disease causing illness in the animal (Arntzen *et al.*, 1991; Mills and Childs, 1998; Carrara *et al*., 2007; Prentice and Rahalison, 2007). Consequently, the disease can be spread to other rodents, and also to humans, via urine, saliva and faeces, or through vectors such as fleas and ticks. Fleas and ticks feed off infected rodents thus becoming infected themselves (i.e., they digest blood containing pathogenic bacteria), and transfer the disease to humans through their bite. Fleas and ticks generally move to human hosts when their animal host (i.e., rodents) die (Keeling and Gilligan, 2000).

Given the high population densities of rodents and the commensal nature of several of these species (Wolfe *et al.,* 2007) it is not surprising that some are important zoonotic reservoirs of human pathogens. For example, *Rattus* was originally restricted to Asia before human-mediated dispersal allowed the spread of *Rattus* species to regions outside of Asia. For example, *Rattus rattus*, *R. norvegicus* and *R. exulans* colonized along the routes of human

1

migration throughout the Indo-Pacific region, a pattern that has been elegantly illustrated for *R. exulans* which displays a close association with human settlements (Matisoo-Smith and Robins, 2004; Robins *et al.,* 2008). Many of these species are highly invasive in new areas of colonization – for example *R. tanezumi* in South Africa (Taylor *et al.,* 2008; Bastos *et al.,* 2011), *R. rattus* in New Zealand (Chappell, 2004), Madagascar (Ganzhorn, 2003; Tollenaere *et al.,* 2010) and Northwest Uganda (Borchert *et al.,* 2007).

Rodents are important zoonotic reservoirs for a number of agents of major infectious diseases (see Meerburg *et al.,* 2009 for recent review) that have been transferred to humans in the past, for example rodent-borne hemorrhagic fevers (hantavirus and arenavirus) (Lee *et al*., 1981; González Ittig and Gardenal, 2004), as well as plague (Keeling and Gilligan, 2000; Prentice and Rahalison, 2007) and typhus (Wolfe *et al*., 2007). Other examples include Lyme disease (Richter *et al*., 1999), hepatitis E virus (He *et al*., 2002), the tick borne encephalitis virus (Charrel *et al*., 2004), and tularaemia (Grunow and Finke, 2002). Black rats (*R. rattus*) are one of the most important reservoirs of human rodent-borne pathogens and this species has had a profound impact on human history, for example through bubonic plague which devastated European populations from the Middle Ages to the end of the 1700s. This continued until *R. norvegicus* replaced *R. rattus* as the predominant vector due, in large part, to the former's tolerance of colder climates (Audoin-Rouzeau, 1999). However, the black rat is still the most common vector of virulent diseases in Central Africa, India and other countries where dense populations of the species occur (Audoin-Rouzeau, 1999; Cavagna *et al*., 2002); it has also been linked to the re-emergence of plague in Madagascar (Bosier *et al*., 2002). In addition, *Rattus sensu lato* species, which includes *R. norvegicus* and *R. rattus*, are known to carry many of the clinically significant zoonoses including scrub typhus, the focus of this investigation.

An accurate systematic framework is pivotal in conducting epidemiological surveys (i.e., the study of factors that contribute to illness and health within populations; Carrara *et al.*, 2007). Chromosomes have been very important for elucidating the systematics of problematic species in these surveys (Lecompte *et al.*, 2003; Venturi *et al.*, 2004). Clarifying the mechanisms of how pathogens affect their hosts, and the reservoir's affect on the pathogen (i.e., replication and fitness of the pathogen), are important for enabling and improving predictions of extinction, emergence or re-emergence of zoonotic pathogens, and their response to changes in their environment. For example, pathogen transmission from a reservoir host through natural or anthropogenic causes (Carrara *et al.*, 2007) may be affected by changes in the reservoir host's habitat.

Scrub typhus in humans

Among the rickettsial diseases, scrub typhus is of medical significance and is often difficult to diagnose as the symptoms are generally non-specific (Kelly *et al.,* 2002; Parola *et al.,* 2005a, b; Lerdthusnee *et al.,* 2008). Clinical symptoms of scrub typhus, an acute febrile zoonotic disease, range from a mild fever (including, for example, headaches and nausea) to more severe cases where multiple organ failure may occur that could result in death (for example Wisseman, 1991; Lerdthusnee *et al.,* 2008; reviewed in Seong *et al.,* 2001). Its severity is dependent on the infective strain and treatment is important given 1-30% mortality associated with untreated scrub typhus (Baxter, 1996; Demma *et al*., 2006). Scrub typhus occurs throughout Australia, Asia and from the Far East to the Middle East (Rosenberg, 1997; Parola *et al.,* 2005a, b). Outbreaks of scrub typhus continue to be reported. For example, several thousand cases including more than 30 deaths have recently occurred in the Maldives (2001-2003) (Lewis *et al.,* 2003). In Thailand approximately 3000 cases of scrub typhus infections were reported each year for the period 1999-2006; these included several

deaths (Lerdthusnee *et al.,* 2008). Worldwide more than a billion people are at risk for infection (Rosenberg, 1997; Coleman *et al.,* 2003). Scrub typhus was believed to be a disease of the past, however, this may largely be due to the lack of sensitive diagnostic clinical tools compounded by the fact that the symptoms of scrub typhus are non-specific (i.e., fever and headaches) (Ong *et al*., 2001; Demma *et al*., 2006). For example, Ong *et al*. (2001) noted that since the introduction of specific serological tests for *rickettsiales* in 1998, endemic typhus has been increasingly noted in Singapore. Importantly, many scrub typhus cases are most likely undetected and hence under-reported in developing countries (i.e., Africa and Asia) where the quality of public health care may be low (i.e., a lack of proper laboratory facilities) (Ong *et al*., 2001; Parola *et al.,* 2005b; Demma *et al*., 2006).

Orientia tsutsugamushi (Hayashi, 1920), the causative agent of scrub typhus, is an obligate intracellular bacterium (reviewed in Seong *et al.,* 2001) originally referred to as *Rickettsia tsutsugamushi*. It was recently elevated to its own genus (*Orientia*) as it is phylogenetically distinct from other *Rickettsia* (Parola *et al.,* 2005b). Scrub typhus is transmitted by numerous species of *Leptotrombidium* chigger mites (larval trombiculid mites) to an animal host (Frances *et al.,* 1999). The disease is spread to humans when they are bitten by a chigger mite infected with *O. tsutsugamushi* (Lerdthusnee *et al.,* 2008). Rodents act as hosts for chiggers and as rodents have an extensive geographic distribution, their ecological plasticity often extends into human settlements and thus humans have an increased likelihood of coming into contact with infected chigger mites (Coleman *et al.,* 2003). This is exacerbated by habitat destruction, climate change and the effects of alien transport, all of which can lead to shifts in animal distributions and increased rodent/human contacts that can result in the transmission of new and re-emerging animal-borne pathogens (Mills and Childs, 1998; Wolfe *et al.,* 2007).

***RATTUS SENSU LATO* COMPLEX**

Phylogenetic review

The Family Muridae (Order Rodentia) comprises in excess of 1300 species that have a cosmopolitan distribution which occur in a variety of habitats from arid deserts to tropical forests (Jansa and Weksler, 2004). The subfamily Murinae (rats and mice) alone is thought to encompass more than 500 recognised species (Musser and Carleton, 2005). The earliest cladogenic events in Murinae are estimated to have occurred approximately 12.3 million years ago (mya). Their natural distribution encompasses Eurasia, Australia, Southeast Asia and Africa (Lecompte *et al.,* 2008). Species diversity is further elevated by the relatively high number of endemic taxa restricted to islands and archipelagos (Michaux *et al.*, 2007).

The *Rattus s. l.* complex, with a hypothesised origin in the tropics (Gadi and Sharma, 1983), is species-rich and, not unexpectedly, shows a great degree of variability with regards to the geographic distribution, habitat utilisation, morphology and life history patterns of its constituent species (Wilson and Reeder, 2005); these characteristics are thought to reflect successive radiations (Michaux *et al.*, 2007). *Rattus sensu lato* (tribe Rattini) is nested within the Murinae (Figure 1.1; Chaimee and Jaeger, 2000; Musser and Carleton, 1993, 2005). *Maxomys* is the most distant sister group to *Rattus*, followed by *Leopoldamys* and *Niviventer* both of which are closely related to each other (Usdin *et al.*, 1995; Verneau *et al.*, 1997; Lecompte *et al.,* 2008; Pagès *et al.,* 2010). The close morphological relationship between *Niviventer* and *Leopoldamys* initially suggested by Musser and Newcomb (1983) was supported by Watts and Baverstock (1994) using immunological techniques (microcomplement fixation of albumin). They also found a distant relationship of these genera to *Maxomys*. The next assemblage to diverge, based on sequence data, was *Berylmys*

followed by *Sundamys* (i.e., Verneau *et al.,* 1998; Lecompte *et al.,* 2008). DNA/DNA hybridizations demonstrated a close relationship between *Berylmys, Bandicota* and *Sundayms* with *Rattus* taxa (Ruedas and Kirsch, 1997). The *Rattus* assemblage (*Rattus sensu stricto*) was the last group to diverge (Usdin *et al.*, 1995; Verneau *et al.*, 1997; Lecompte *et al.,* 2008; Pagès *et al.,* 2010). Some geneticists believe that *Bandicota* should be included in *Rattus s. s.* – i.e., a paraphyletic *Bandicota*/*Rattus* clade (Usdin *et al.*, 1995; Verneau *et al.*, 1997, 1998).
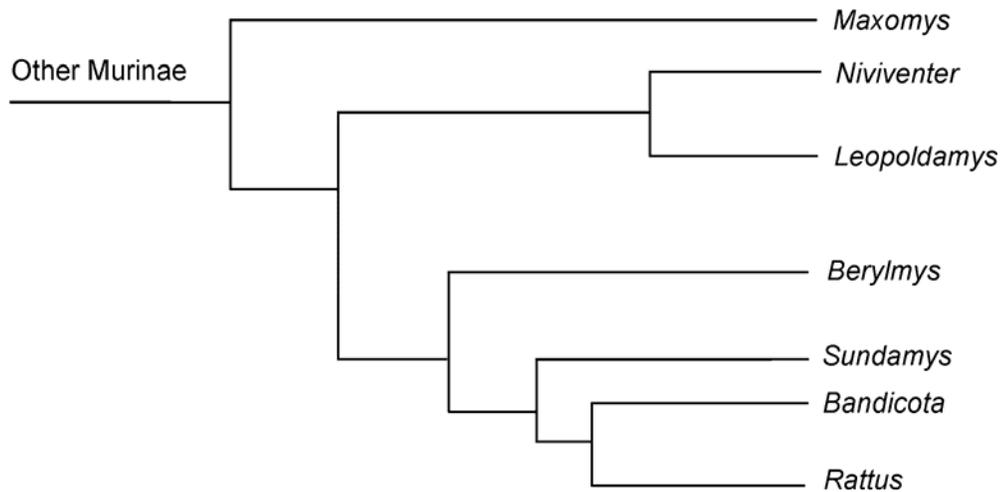
Figure 1.1. Simplified molecular tree depicting the relationships between Southeast Asia rodents of the *Rattus s. l.* complex. Modified from Verneau *et al.* (1997, 1998) and Lecompte *et al.* (2008).

Robins *et al.* (2008) placed the deepest divergence within *Rattus* at approximately 3.5 mya, with *R. norvegicus* diverging at the earliest ~2.9 mya. Their estimates (based on mtDNA sequence) suggest that the sister species *R. rattus* and *R. tanezumi* diverged from *R. exulans* ~2.2 mya, while *R. rattus* and *R. tanezumi* diverged as recently as ~0.4 mya. These results place the diversification of the *Rattus* assemblage somewhat earlier than those of Verneau *et al.* (1998) calculated on LINE-1 (L1) sequence variation, and those of Lecompte *et al.* (2008) using nuclear (growth hormone receptor, GHR, exon 10 and interphotoreceptor retinoid binding protein , IRBP, exon 1) and mitochondrial markers (cytochrome b).

Lecompte *et al.* (2008) calculated the divergence of other Murinae from Rattini to be 11.3 mya, somewhat later than the 9.7 mya calculated by Rowe *et al.* (2008). Estimates place the *Micromys* - Rattini divergence at 9.7 mya followed by *Rattus s. l.* at 8.4 mya, and *Rattus s. s.* at 5.4 mya (Lecompte *et al.,* 2008; *Rattus s. s.* group of Lecompte includes *Berylmys* and *Bandicota*), dates that are largely consistent with earlier reports by Verneau *et al.* (1998).

Despite the recent advances in the taxonomy of *Rattus* (i.e., Usdin *et al.*, 1995; Verneau *et al.,* 1997, 1998; Jansa and Weksler, 2004; Lecompte *et al.,* 2008; Rowe *et al.,* 2008; Pagès *et al.,* 2010) that resulted in the removal of species which were originally included within the *Rattus* clade (such as those of *Niviventer*), the boundaries of *Rattus* and its taxonomy are still unclear (Usdin *et al.*, 1995; Jansa and Weksler, 2004). Due to a lack of morphological data and a well resolved taxonomy, black rats are often confused with other Asian *Rattus* species. For example, *R. rattus* is often confused with *R. tanezumi* (Aplin *et al.*, 2003a).

Cytogenetic data and taxonomy

Maxomys with 17 species (originally considered a subgenus of *Rattus* by Ellerman, 1961 but elevated to a separate genus by Misonne, 1969) is karyotypically diverse. Chromosome number varies from 2n = 34-52 (i.e., *M. hellwaldii* 2n = 34, *M. rajah* and *M. whitehead* 2n = 36, *M. inas* 2n = 42, *M. surifer* and *M. moi* 2n = 52) (see Rickart and Musser, 1993 and references therein). All 18 species within *Niviventer* that diverged following *Maxomys*, are reported to have similar karyotypes (2n = 46) (Ray-Chaudhuri and Pathak, 1970; Yong 1970; Musser, 1981; Gadi and Sharma, 1983; Wang *et al.,* 2003). Marshall (1977) classified *Niviventer* as a distinct subgenus, whereas it was previously allied with *M. bartelsii*. Musser (1981) subsequently elevated it to separate generic rank and divided it into two separate "divisions", namely the *N. andersoni* division and *N. niviventer* division.

Leopoldamys (six species) is the sister clade to *Niviventer* and was first described by

Ellerman (1961). It is reported to have a chromosome complement of 2n = 42 (Rickart and

Musser, 1993). However, there are exceptions: *L. edwardsi* (2n = 42 or 44; Yong, 1970;

O'Brien *et al.,* 2006) and *L. neilli* (2n = 44; Marshall, 1977; Tsuchiya *et al.,* 1979).

Berylmys (four species) is regarded as the sister genus to *Rattus* (Ellerman, 1961;

Gadi and Sharma, 1983). Cytogenetic data are limited for this genus, although a

chromosome complement of 2n = 40 has been reported for *B. bowersi* and *B. berdmorei*

(Yong, 1968, 1973; Markvong *et al*., 1973; Duncan *et al*., 1970; Marshall, 1977; Yosida,

1980). *Sundamys* (three species), the next clade to diverge following *Berylmys*, is reported to

have a chromosomal complement of 42 (*S. muelleri*; Yong, 1968; Yosida, 1973) and was

considered to be distinct from *Rattus* by Musser and Newcomb (1983), and Musser and

Heaney (1992).

Bandicoot rats were originally grouped with *Nesokia* in a single genus by Blanford

(1891) but later recognised as distinct taxa - *Bandicota* and *Nesokia* (Ellerman, 1961).

Bandicota comprises three species, *B. bengalensis* (lesser bandicoot-rats), *B. indica* (great

bandicoot-rats) and *B. savilei* (Savile's bandicoot rat) (Gadi and Sharma, 1983; Wilson and

Reeder, 2005). The chromosome complements of these rodents vary markedly from each

other (Gadi and Sharma, 1983). The *B. indica* karyotype varies from 2n = 42-46 (Markvong

et al., 1973; Hsu and Benirschke, 1974). The chromosome complement of *B. i. indica* (2n =

42) observed by Gadi and Sharma (1983) is similar to that recorded by Avirachan *et al*.

(1971), whereas *B. savilei* has 2n = 44 (Markvong *et al*., 1973).

The last group to diverge based on the published topology (i.e., Usdin *et al.*, 1995;

Verneau *et al.,* 1997, 1998; Jansa and Weksler, 2004; Lecompte *et al.,* 2008; Pagès *et al.,*

2010) was the *Rattus* assemblage (65 species). The majority of *Rattus* species are reported to

have a diploid number of 42. There are, however, exceptions that show 2n = 38 including *R. rattus, R. fuscipes* and *R. frugivorous* (Raman and Sharma, 1977; Yosida, 1980; Gadi and Sharma, 1983; Cavagna *et al.,* 2002). *Rattus norvegicus* standardised karyotype was published by the Committee for a Standard Karyotype in 1973; G- and C- banding karyotypes have been well characterised (Yosida and Sagai, 1972; Levan, 1974; Gadi and Sharma, 1983; Guilly *et al.*, 1999; Hamta *et al.*, 2006; O'Brien *et al.*, 2006). The available cytogenetic data (i.e., G- and C-bands) for species investigated in this study (*Rattus s. l.*) are summarised in Table 1.1. Species represent six of the seven *Rattus s. l.* genera and thus allow a comprehensive comparative molecular cytogenetic investigation of this complex.

Table 1.1. Citation summary for cytogenetic data available for representatives of the *Rattus sensu lato* complex investigated in this study. Dashes indicate absence of data.

Species	2n (standard karyotype)	G-band karyotype	C-band karyotype
Bandicota			
B. indica	42-46	Gadi *et al.* (1982) and Gadi and Sharma (1983)	Gadi *et al.* (1982)
B. savilei	44	-	-
Berylmys			
B. bowersi	40	-	-
B. berdmorei	40	-	-
Leopoldamys			
L. edwardsi	42 or 44	O'Brien *et al.* (2n = 44; 2006)	-
L. neilli	44	-	-
Maxomys			
M. surifer	52	-	-
Niviventer			
N. fulvescens	46	Jiang (1995)	Baskevich and Kuznetsov (2000)
Rattus			
R. exulans	42	Yosida and Sagai (1972, 1973) and Yosida (1975)	
R. losea	42	-	Yosida (1980)
R. norvegicus	42 (Committee for a Standard Karyotype in 1973)	Yosida and Sagai (1972, 1973); Levan (1974); Gadi and Sharma (1983); Guilly *et al.* (1999); Hamta *et al.* (2006) and O'Brien *et al.* (2006)	
R. tanezumi	42	Yosida and Sagai (1975) and Yosida (1980)	

9

Two hypothetically interrelated ancestral karyotypes have been suggested to explain the evolutionary relationships within *Rattus s. l.* (Raman and Sharma 1977; Gadi and Sharma 1983). The first posits an ancestral diploid number of 2n = 46 from which the *Niviventer* clade (2n = 46) is thought to be derived. The second suggests an ancestral karyotype of 2n = 42 that gave rise to the *Rattus, Berylmys* and *Leopoldamys* genera. This ancestral karyotype differs from the former through a series of fusions and inversions. The Australian and Asian species of *Rattus* are hypothesized to have evolved independently from the same ancestor (the 2n = 42 ancestral karyotype) based on differences in molar morphology (Misonne, 1969; Taylor and Horner, 1973; Seddon and Baverstock, 2000). The bandicoot rats most likely evolved from a taxon that became further diversified due to the development of large sex chromosomes (Gadi and Sharma, 1983).

THE RAT GENOME AND ITS ONLINE RESOURCES

It is imperative that rat genomic resources match the species importance as a model for biomedical research (Twigger *et al.,* 2008). As part of this, the Rat Genome Database (RGD, http://www.rgd.mcw.edu) was established in 1999 in conjunction with the University of California, Santa Cruz (UCSC), Ensembl, US National Center for Biotechnology Information (NCBI) and Mouse Genome Informatics (MGI) (Aitman *et al.,* 2008). The first high-quality draft sequence of the rat genome (*R. norvegicus*, strain BN/NHsdMcwi) was published in 2004 and was used comparatively against mouse and human (Gibbs *et al.,* 2004). This has subsequently been updated with completed sequences obtained from targeted BACs (~ 55 Mb; Worley *et al.,* 2008).

The RGD still continues to obtain and incorporate new genomic datasets into its database (i.e., quantitative trait loci (QTL), gene models, strain-specific length polymorphisms (SSLPs) for diagnostic PCR amplification products, sequence and map data).

These datasets have undergone significant growth and advancement since 2005 (see Table 1.2 for an example of the most recent update of genomic data provided by the RGD and Laulederkind *et al.,* 2011). Genetic and core genomic datasets (i.e., QTLs and strains) are obtained through a variety of sources via local bioinformatic analyses and the literature (Dwinell *et al.,* 2009; Laulederkind *et al.,* 2011). The availability and increasing collection of molecular markers (i.e., microsatellites, single nucleotide polymorphisms, SNPs, for QTLs and disease gene mapping) has enabled the development of a high-density rat genetic map which is freely available and is invaluable in genetic studies of disease phenotype and physiology (Aitman *et al.,* 2008).

Table 1.2. Figures showing increase in the number of data elements found in the Rat Genome Database since 2005 (RGD, http:// www. rgd.mcw.edu; Laulederkind *et al.,* 2011).

	2005	2009/2011
Rat genes	8442	35,427/40,000
Known Genes		27,782
Pseudogenes		7492
Human genes	3432	26,703
Mouse genes	4101	35,351
Rat QTLs	808	1552/1771
Strains	738	1876/2209
Rat SSLPs	9730	12,813
Number of genes with GO annotations	3850	17,167
Number of genes with experimentally supported GO	2740	13,392
Number of GO annotations	39,861	167,767

Conventional cytogenetic analyses are also of great importance in rat genetics and the updated cytogenetic map constructed from G-banded *R. norvegicus* prometaphase chromosomes comprising 535 individual bands has greatly improved the previously available ideogram (see the Rat Genome Browser in Ensembl,

http:www.ensembl.org/Rattus_norvegicus) enabling the interpretation of biological data (Hamta *et al.,* 2006).

In summary, the RGD database together with NCBI Ensembl and UCSC resources (see the available online documentation about the NCBI and UCSC resources, i.e., Map Viewer and BLAST; http://www.ncbi.nlm.nih.gov and http://www.genome.ucsc.edu) are important tools that promote rat research. These tools enable easy access to available rat data through various formats that include pathway diagrams, linear genome viewers and biological ontologies. The continued support of public databases and bioinformatics ensures the future of the rat as a genetic model for complex diseases, physiology, pharmacological studies, as well as the development of new diagnostic and treatment approaches (Dwinell *et al.,* 2009).

COMPARATIVE CYTOGENETICS

Chromosome painting and genomic comparative maps

Chromosome rearrangements serve as powerful evolutionary characters and phylogenetic landmarks (Murphy *et al.,* 2004; Wienberg, 2004). Information on karyotype evolution and the mechanisms involved in repatterning of chromosomes are critical in determining their significance in speciation. For example, rearrangements may influence gene expression due to genetic duplications and deletions, often affecting the phenotype (Ferguson-Smith and Trifonov, 2007). Chromosome painting, also known as Zoo-fluorescence *in situ* hybridization (Zoo-FISH), is a method frequently used for the identification of orthologies and rearrangements spanning mammalian orders (Murphy *et al.,* 2004) and this has lead to the rapid and economical mapping of syntenies and the charting of genomes (Wienberg and Stanyon, 1995; Stanyon *et al*., 1999). This aids the exploration of poorly studied or unknown species (Cavagna *et al.,* 2002), as Zoo-FISH data can be used to develop comparative maps that result from alignments of chromosomes of related species

12

(Ferguson-Smith and Trifonov, 2007). However, technical difficulty increases over large evolutionary distances. Consequently it is not feasible to conduct Zoo-FISH analyses between placental mammals (eutherians) and marsupials (except for a small portion of the X chromosome – see Glas *et al.,* 1999), monotremes, as well as reptiles, fish, amphibians and other vertebrate classes of birds (Ferguson-Smith and Trifonov, 2007; Martins *et al.,* 2010). Additionally, Zoo-FISH is unable to resolve segments ≤4 Mb in size (Murphy *et al*., 2004). FISH using whole chromosome probes results only in a rather gross estimate of the degree of genomic rearrangement, especially in the absence of reciprocal painting (see below) since the precise origin (and borders) of chromosomal segments resulting from multiple translocations cannot be determined with accuracy (Murphy *et al*., 2004; Ferguson-Smith and Trifonov, 2007). Moreover, it cannot detect intrachromosomal rearrangements. In these instances probes (such as bacterial artificial chromosome, BACs, clones) that hybridize to subregions of chromosomes are able to more precisely elucidate intrachromosomal rearrangements (Ried *et al*., 1993; Wienberg and Stanyon, 1995). The utility of BAC clones is underscored by their specificity - they can serve as positional markers at points of interest, and for the identification of intrachromosomal rearrangements such as inversions. In brief, they permit a finer scale analysis than do whole chromosome or even microdissected painting probes (Ried *et al*., 1993; Wienberg and Stanyon, 1995; Ferguson-Smith and Trifonov, 2007).

As indicated above, reciprocal chromosome painting is often applied to corroborate the chromosomal syntenies in two independent tests, and more accurately define the breakpoint boundaries. Reciprocal chromosome painting is required when the probe representing one chromosome of the donor species hybridizes to more than one chromosome pair in the target species. The chromosomes of the target species are flow-sorted and following degenerate oligonucleotide primed PCR (DOP-PCR) are characterised by hybridizing flow-sorts back to the chromosomes of the original donor species using FISH.

The chromosome paint (probe) contains a mixture of DNA that will hybridize to the entire chromosome of origin and can be visualized by fluorescent microscopy. Probes are generated using DOP-PCR which relies on partially degenerated oligonucleotide primers to amplify an assortment of related target DNA (Telenius *et al.*, 1992).

Chromosome painting, first introduced in 1988 to human cytogenetics (Pinkel *et al.,* 1988), has become common place in other mammals as well as fishes, birds and some invertebrates (see Martins *et al.,* 2010 for recent review). Wienberg and colleagues (1990) were among the first to pioneer chromosome painting in primates where they applied this tool for the first comparative cytogenetic analysis between Japanese macaque and human. Since then numerous studies have been conducted that compare the genomes of eutherian mammals, as well as birds (see Griffin *et al.,* 2007; Ferguson-Smith and Trifonov, 2007, Martins *et al.,* 2010 for review). The development of DOP-PCR in 1992 (Telenius *et al.,* 1992, see above) made it feasible to generate probes for any given vertebrate species, whereas studies prior to this were limited by the availability of painting probes. The increasing accessibility of probes from more than 100 mammalian species has helped revolutionize the field of comparative cytogenetics (Martins *et al.,* 2010). However, human chromosome-specific painting probes have been the mainstay in the field and most comparative studies have been conducted using the human genome as a reference. For example, human probes have been applied to the genomes of approximately 70 mammalian species and as a consequence comparative chromosome maps have been established between humans and representatives of nearly all 18 extant eutherian orders (Ferguson-Smith and Trifonov, 2007; Graphodatsky *et al.,* 2007).

Chromosome painting has also been important for the development of putative ancestral karyotypes for several primate groups (i.e., the ancestral anthropoids karyotype and

ancestral hominoid karyotype), rodents, carnivores, Cetartiodactyls, Perrissodactyls; including the establishment of the ancestral eutherian karyotype (AEK) (see Ferguson-Smith and Trifonov, 2007; Stanyon *et al.,* 2008). The AEK has since been revised and refined through further cytogenetic studies and bioinformatic analyses (see Robinson and Ruiz-Herrera, 2008).

Additionally, chromosome painting has provided insights concerning the rate of chromosome evolution occurring within different families. For example, Ursidae, Candidae, Mephitidae families and the muroid superfamily appear to exhibit rapid rates of karyotype evolution (see Martins *et al.,* 2010 and references therein). The latter group is believed to display the fastest rate of chromosome evolution among mammal species (i.e., Graphodatsky, 2007), as indicated by reciprocal painting between rats and mice which found that the rate of divergence between these two species is tenfold greater than that observed between human and cat (Stanyon *et al.,* 1999). Another example of differing rates of karyotypic evolution has been identified between mammals and birds, where chromosome painting has indicated that birds are karyotypically more stable than mammals (Wienberg, 2004; Griffin *et al.,* 2007). For instance, birds are separated into two different groups, namely Neognathae and Paleognathae. Zoo-FISH studies using chicken chromosome (Neognathaea) probes onto emu *Dromaius novaehollandis* (Paleognathae) and *Rhea americana* (Paleognathae) chromosomes revealed a high degree of karyotype conservation between these two avian groups (Shetty *et al.,* 1999; Guttenbach *et al.,* 2003).

Phylogenetics using chromosomal characters

Cross-species chromosome painting has paved the way for the growth of phylogenomics, a field where evolutionary relationships may be inferred through the use of chromosomal syntenies as phylogenetic characters. Given the premise that chromosome

rearrangements are infrequent and have low rates of convergence these phylogenetically informative markers are often considered ideal candidates for use in the reconstruction of the evolutionary history of lineages (Rokas and Holland, 2000). They are sometimes more informative than gene sequences (Veyrunes *et al.*, 2006). The employment of Zoo-FISH in this context has contributed towards the unravelling of the eutherian tree, allowed for proposals of ancestral karyotypes (see above), and the placing of these within a timescale for diversification (Rokas and Holland, 2000; Wienberg, 2004; Ferguson-Smith and Trifonov, 2007; Graphodatsky *et al.,* 2008).

Heterochromatin and its role in genome evolution

Constitutive heterochromatin demonstrates considerable variability in amount and distribution in rodents (Mascarello and Hsu, 1976; Baker and Barnett, 1981; Hamilton *et al.,* 1990 among others). The evolutionary histories of many eukaryotic genomes have, to a large extent, been influenced by heterochromatin as shown for numerous mammalian taxa (reviewed in Adega *et al.,* 2009). Heterochromatin heterozygosity in *M. terricolor* has been reported to reduce meiotic fitness between hybrids, and possibly act as a reproductive barrier in *M. terricolor* speciation (Sharma *et al.,* 2003). Chromosomal variability linked with differences in heterochromatin may, at times, be correlated with species' status. For example, species within the rodent genus *Onychomys* and *Xerus* can be differentiated based on the number of heterochromatic short arms or blocks (Baker and Barnett, 1981; Robinson *et al.*, 1986; Hamilton *et al.,* 1990) indicating an absence of gene flow between taxa. In addition, somewhat controversially, Thelma *et al*. (1988) reported that the heterochromatic (C-blocks) differences may on their own promote speciation in *Nesokia indica* without leading to other structural changes such as fusions. They designed crossing experiments between *Nesokia* specimens with X and Y chromosomes displaying different C-banding

16

patterns. Their results suggest that the heterochromatin was important for reproductive success, and that the heterochromatin may promote differentiation and potentially speciation.

Higher eukaryotic genomes often contain a significant proportion of repetitive sequences (Chaves *et al*., 2000b). Repetitive DNA can be organised into two main categories. One comprises DNA sequences, mainly transposable elements that are intermingled with other sequences and interspersed throughout the genome. The other consists of highly repeated families of repeats called satellite DNA consisting of arrays of similar or identical tandemly arranged units (reviewed by Slamovits and Rossi, 2002). In general, tandem repetitive sequences are located within constitutive heterochromatin usually located close to the centromeres and telomeres of chromosomes. However, repetitive sequences have also been detected interstitially (Charlesworth *et al.,* 1994; Dean and Schmidt, 1995), as well as in X-autosome translocations where interstitial heterochromatic blocks occurring between the autosome and sex chromosome are proposed to act as barriers to the spread of X-inactivation to the translocated autosomal segment (Dobigny *et al.,* 2004b).

In mammals the pericentromeric heterochromatic regions generally consists of satellite DNA (i.e., Garagna *et al*., 2001). Recent studies have attempted to elucidate the function of satellite DNA and its organization in the heterochromatic regions of the chromosomes of mammals. In mouse and human satellite DNA is thought to play a role in centromere function (Garagna *et al*., 2001; Hartman and Scherthan, 2003; Chaves *et al*., 2004). Both molecular and *in situ* studies of *M. domesticus* satellite DNA sequences indicate that certain features in the composition and organisation of the satellite DNA are correlated with the propensity of the telocentric chromosomes to become fused at the centromeric regions i.e., Robertsonian translocations (Redi *et al*., 1990). In Robertsonian translocations,

17

two chromosomes (telocentric or acrocentric) fuse together at or near to their centromeres generating a bi-armed metacentric chromosome with an internal centromere (Garagna *et al.*, 2001).

Satellite DNA found in these pericentromeric regions are important evolutionary markers for genome divergence and speciation (Modi *et al.*, 1996; Chaves *et al.,* 2005) as sequence analysis of these repeats indicate that they are evolving rapidly (Chaves *et al.*, 2000a, b). The satellite I sequence has been studied in a number of species including sheep, goats, cattle and pigs (Chaves *et al.*, 2000b; Chaves *et al.*, 2003a,b; Chaves *et al.*, 2005; Adega *et al.*, 2005, 2006), but to date, these have not been extensively studied in rodents. Karyotype analysis, together with satellite DNA sequence analysis, aids the clarification of important aspects of repetitive sequence evolution as well as genome evolution (Chaves *et al.*, 2000b). Satellite DNA is, however, not free of homoplasy as these regions may predispose to breakpoint reuse (Slamovits *et al.*, 2001).

GENE EXPRESSION APPROACH

Real-time PCR (RT-PCR) is one of the methods of choice for quantifying mRNA in biological samples derived from various sources. Real-time PCR, also known as quantitative PCR (qPCR), is commonly used where low levels of gene expression are analysed. Furthermore, it is an accurate high-throughput technique capable of quantifying small changes in gene expression (Pfaffl, 2001; Bustin, 2002; Ginzinger, 2002; Huggett *et al.,* 2005). In addition, it avoids post-PCR handling and the use of radioactivity, and only small quantities of RNA are needed (Radonić *et al.,* 2004).

Real-time PCR amplifies cDNA (complementary DNA) produced by the reverse transcription of mRNA that serves as a link between DNA and proteins (Pfaffl, 2001; Radonić *et al.,* 2004). A number of detection systems can be used for qPCR analysis such as

18

molecular beacons (Tyagi and Kramer, 1996), hybridization probes (Bernard and Wittwer, 2000) and scorpions (Whitcombe *et al.,* 1999; Thelwell *et al.,* 2000). The simplest detection technique for qPCR, and the method followed in the present investigation, involves the use of double strand DNA (dsDNA) binding dyes, i.e., SYBR Green I fluorescence dye (Giulietti *et al.,* 2001; Pfaffl, 2001).

During qPCR, the accumulation of PCR product during the exponential phase is measured by the increase of fluorescent signal of the SYBR Green I dye. SYBR Green I is undetectable when in its free form (unbound), however, once it binds to dsDNA it begins to emit a fluorescent signal (Giulietti *et al.,* 2001). Despite this advantage, SYBR green not only detects the target but also primer-dimers and non-specific PCR products as it binds to all dsDNA. This disadvantage can be countered by analysing melting curves of each sample in the reaction on the LightCycler, and through optimization of the PCR reaction and the careful design of primers (Kellogg *et al.,* 1994; Giulietti *et al.,* 2001). SYBR green has been successfully used in numerous studies together with the LightCycler, i.e., viral load identification (Schröter *et al.*, 2001), and cytokine quantifications (Giulietti *et al.,* 2001).

GENERAL AIMS OF THE STUDY

The paucity of data in the *Rattus s. l.* complex highlights a critical need for more extensive systematic research on this medically and agriculturally important group of rodents. It is intended that this dissertation will contribute to resolving some of the problematic relationships within the *Rattus s. l.* complex and provide a molecular comparative cytogenetic perspective for this group of rodents. Prior to this investigation comparative molecular cytogenetic studies within *Rattus s. l.* were largely limited to *R. norvegicus* and *R. rattus* (see for example Cavagna *et al.,* 2002). The *Rattus s. l.* complex was selected as it is known to have an association with several pathogens, specifically scrub typhus, in addition to which a

completely sequenced genome and densely annotated genetic map of *R. norvegicus* (biomedical model) is available (see for example Aitman *et al.,* 2008 and Twigger *et al.,* 2008).

In broad terms, the investigation analysed the karyotypic diversity in Rattini using conventional (banding) cytogenetic techniques. Moreover it comprised a comprehensive comparative cytogenetic investigation that relied on chromosome painting probes previously isolated from *R. norvegicus* (see Stanyon *et al.,* 1999), as well as *M. surifer* and *R. rattus* that were provided by the Cambridge Resource Centre for Comparative Genomics (UK); these were subsequently characterised in Stellenbosch. The aims of the molecular cytogenetic aspect of my study were in the first instance to develop comparative genomic maps within the *Rattus s. l.* complex by reciprocal chromosome painting using Zoo-FISH. These data (i.e., regions of synteny conservation and breakpoints junctions) enabled the construction of a chromosomal phylogeny for these rodents and permitted the examination of the utility of these data for deciphering phylogenetic relationships within *Rattus s. l*. Secondly, I set out to describe the mode (i.e., the type of rearrangements) and tempo (the rate of accumulation) of genome repatterning within the *Rattus s. l.* complex, and to construct a putative ancestral karyotype for these rodents. Thirdly, genomic features such as the distribution of the satellite I DNA family was investigated as these features may have facilitated (if not promoted) chromosome evolution in *Rattus s. l*. The analysis of the satellite I DNA repeat family was completed during a short visit to the Institute for Biotechnology and Bioengineering, Centre of Genetics and Biotechnology of the University of Trás-os-Montes, Vila Real, Portugal in the laboratory of Drs Raquel Chaves and Filomena Adega.

The second focus of my study dealt with the analysis of the *Rattus s. l.* complex and its role as a reservoir for the human pathogen, scrub typhus. This relied on a qPCR gene

expression approach and real-time PCR. In broad terms the aim here was to compare the sequence and expression levels of syndecan-4 (considered a good candidate for this comparative study as it has been linked to *O. tsutsugamushi* infection, see for example Kim *et al.,* 2004) between Rattini species that are known to be natural reservoirs for the typhus agents, and Murini species that are not. The gene expression aspects were completed during two three month visits to the Centre de Biologie pour la Gestion des Populations, UMR 022 INRA-IRD-CIRAD-SupAgro, Campus de Baillarguet, CS30016, 34988 Montferrier-sur-Lez, France, in Dr Gauthier Dobigny's laboratory.

CHAPTER 2

NEW KARYOTYPIC DATA FOR ASIAN RODENTS (RODENTIA, MURIDAE) [*]

INTRODUCTION

The non-ambiguous identification of species is critical for fundamental as well as applied research as it is the departure point for many aspects of comparative biology including systematics (see for example Sites and Marshall, 2003 and references therein), biodiversity assessments, conservation (Crandall *et al.*, 2000; Moritz, 2002), biogeography (i.e., Avise, 2009), community ecology and pests control (i.e., Hoffman *et al.,* 2002). In particular, the lack of accurate species-specific identification of hosts is an important consideration for current epidemiological surveys since (i) numerous pathogens are carried by a restricted range of hosts, and (ii) each rodent species possesses its own life traits. These two aspects are pivotal since they may lead to important differences in the role a host species can play in circulation and transmission of pathogens. Additionally, some species-specific ecological characteristics (such as commensality and hence the propensity to transport with humans) and evolutionary properties (population structure and connectivity) are also central to epidemiological processes (reviewed in Mills and Childs, 1998). In such a context, karyotypic data has been shown to be valuable, especially for identifying rodents that belong to sibling species complexes.

---

[*] The work described in this chapter has been published as part of the following manuscript: D. Badenhorst, V. Herbreteau, Y. Chaval, M. Pagès, T. J. Robinson, W. Rerkamnuaychoke, S. Morand, J.-P. Hugot, G. Dobigny. (2009). New karyotypic data for Asian rodents (Rodentia, Muridae) with the first report of B-chromosomes in the genus *Mus*. Journal of Zoology (London), 44-56. Print ISSN 0952-8369.

The *Rattus s. l.* complex (i.e., *Rattus s. s.* and allied genera) is a biologically important group of species given its role as pests and hosts of important human pathogens (see Introduction – Chapter 1). This complex includes the biologically important index species, *R. norvegicus* (Aplin *et al.*, 2003a) for which a vast amount of data, both genomic and immunogenetic, has been generated (see for example Aitman *et al.,* 2008 and Twigger *et al.,* 2008). The other species of *Rattus s. l.* have, however, attracted less attention and surveys (particularly of wild populations of these species which are of epidemiological importance) are largely lacking. In addition, somewhat surprisingly given their epidemiological importance, karyotypic data on species in this complex are limited (see Table 1.1 - Chapter 1). Numerous systematic modifications (reviewed in Wilson and Reeder, 2005) have been proposed since the pioneering cytotaxonomic work published in the late 1960s to the 1970s (see Marshall, 1977 and references therein), further complicating interpretations of correspondence between karyotype and host species. For example, many (if not most) of the taxa that were originally considered as different *R. rattus* subspecies (e.g., *tanezumi*, *argentiventer*, *rattus*) are now regarded as synonyms or have been elevated to distinct species. In the same manner, many formerly congeneric species have been split into different genera (e.g., *Berylmys*, *Maxomys*, *Leopoldamys*, and *Niviventer*) (see Table 2.1). Consequently, species-specific karyotypes have been difficult to delimit making a cytotaxonomic approach to species identification problematic.

In an attempt to address these shortcomings, conventional cytogenetic analysis of twelve rodent species was undertaken. These included *H. delacouri* (selected as outgroup taxon, see Chapter 3), as well as representatives of most lineages within the Rattini tribe (*sensu* Lecompte *et al.*, 2008), i.e., *R. exulans*, *R. losea*, *R. tanezumi*, *L. edwardsi, L. neilli*, *M. surifer*, *N. fulvescens*, *B. berdmorei*, *B. bowersi, B. indica* and *B. savilei*. These rodent species were found in different localities in Thailand (see Table 2.1 for distribution patterns

of these rodents) and formed part of an epidemiological survey in this region (ANR program 00121-05, from the French Ministry of Research). In particular, the collection focused on murids that are suspected of being important reservoirs for several human pathogens, specifically scrub typhus (see Chapter 5). The first goal was to develop identification "tools" for use at the species-level. That this approach has application is illustrated by the Asian Hantaviruses (responsible for lethal haemorrhagic fevers)/rodents association found in Cambodian *Rattus* specimens that could not be discriminated at the species level (Reynes *et al.*, 2003). The overarching aim of this aspect of the study was to re-assess the reliability of karyotype-based diagnoses of Asian rodents in an updated taxonomic framework. It is important to note that the use of *B. indica* and *L. neilli* samples, as well as several of the additional specimens of each species of *Rattus s. s*. (see Table 2.2) was precluded in other aspects of the thesis (Chapter 3 and 4) due to poor quality of the chromosome preparations and limited quantities of cell suspension.

Table 2.1. Taxonomic classification of *Rattus s. l.* and *H. delacouri* and simplified geographic range maps for each genus (Wilson and Reeder, 2005; Global Biodiversity Information Facility, GBIF, http://www.gbif.org).

Species	Reference	Distribution	Simplified geographic distribution maps
Bandicota indica	(Bechstein, 1800)	Throughout Southeast Asia, from India to Indonesia. Recent immigration into Thailand.	
Bandicota savilei	Thomas, 1916	From Burma to Vietnam, Southeast Asia and occurs throughout Thailand.	
Berylmys bowersi	(Anderson, 1879)	Southern China, Burma, Northern Laos, Northern Vietnam, Malaysia, Indonesia, Thailand, (present in northern upland areas and southern peninsula), and Northeast India	
Berylmys berdmorei	(Blyth, 1851)	Southern China, Cambodia, Laos, Southern Vietnam, Burma, Thailand, Laos, Champassak Province, Cambodia, Mondulkiri Province, and Vietnam,	

Leopoldamys edwardsi	(Thomas, 1882)	Northwest of India, North of Burma, South and Central China, North of Laos, Vietnam, Northern Thailand (i.e., Loei). Phrae Province, and Cambodia (Cambodian Mondulkiri Province)	
Leopoldamys neilli	(Marshall, Jr., 1976)	Saraburi Province, North and Southwest Thailand	
Maxomys surifer	(Miller, 1900)	South of Burma, Laos, Vietnam, Cambodia, Thailand, Malaya, Indonesia (Sumatra, Java, Borneo), and South of China	
Niviventer fulvescens	(Gray, 1847)	From South Himalayas through Bangladesh, South China, Thailand, Cambodia, Laos, Cambodia, Malay Peninsula, Indonesia (Sumatra, Java and Bali),	
Rattus exulans	(Peale, 1848)	Cosmopolitan distribution. Distribution in Asia: Bangladesh, Centre and South Burma, Thailand, Cambodia, South Laos (absent from Luang Prabang Province), Centre and South Vietnam, China (south west of Hainan), Malaysia, Indonesia, and Philippines	
Rattus losea	(Swinhoe, 1871)	Cosmopolitan distribution. Distribution in Asia: Taiwan, Southeast China, Hainan, Vietnam, Laos, Cambodia, and Thailand	
Rattus norvegicus	(Berkenhout, 1769)	Cosmopolitan distribution, reported throughout the world in all main ports and major cities. Distribution in Asia: native from Northern China recorded mainly within the central plain and the South of Thailand.	
Rattus tanezumi	Temminck, 1844	Unclear systematics, often mistaken for *Rattus rattus* in publications. Distribution in Asia: Cambodia, Lao PDR and Thailand.	
Hapalomys delacouri	Thomas, 1927	Southern China (Hainan Isl and S Guangxi on the mainland), Northern laos, Central Vietnam, (limits unknown), Laos, and Thailand (Loei)	
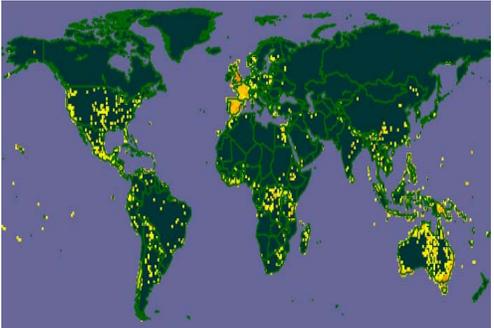
MATERIAL AND METHODS

Tissue samples, cell culture and chromosome preparation

Material was collected in 2006 and 2007 in Thailand (ANR-00121-05 'Biodiversity of rodent-borne hantaviruses in Southeast Asia'; coord. J.P. Hugot, 2006-2008) in the surrounding area of Phrae (18°09'N, 100°08'E), Loei (17°29'N, 101°43'E) and Kalasin (16 °49'N, 103 °53'E) using locally made cage traps (Table 2.2). Voucher specimens have been deposited in the mammal collections of the Kasetsart University (Bangkok, Thailand) and Centre de Biologie et de Gestion des Population (CBGP; Montferrier sur Lez, France) and are available upon request. Except where otherwise detailed, taxonomy follows Musser and Carleton (2005).

Table 2.2. Specimens investigated in the present study. "+" indicates specimens identified using DNA sequencing (Pagès *et al.,* 2010). Individual numbers refer to the CBGP Asian rodent collection (Montpellier, France). Sex, origin, diploid (2n) and autosomal fundamental (NFa) numbers, type of biological material (bm: bone marrow; cc: cell culture), as well as number of B chromosomes (Bs) are provided for each specimen. "?" indicates unavailable data.

Species	Id.	Individuals			Cytogenetics			
	DNA	individual	Sex	Origin	material	2N	NFa	Bs
Rattus losea	+	R4203	F	Phrae	bm	42	58	
	+	R4229	M	Loei	bm	42	58	
	+	R4230	M	Loei	bm	42	57	
	+	R4260	F	Loei	bm	42	58	
	+	R4724	F	Loei	cc	42	56	
Rattus tanezumi	+	R4113	M	Loei	bm	42	54	
	+	R4481	F	Phrae	bm	42	55	
	+	R4003	F	Kalasin	bm	42	57	
	-	R4008	M	Kalasin	bm	42	55	
	+	R4016	F	Phrae	bm	42	52	
	-	R4096	F	Loei	bm	42	52	
	-	R4145	F	Phrae	bm	42	54	
	+	R4182	F	Phrae	bm	42	56	
	+	R4436	M	Phrae	bm	43	59	1
	-	R4447	M	Phrae	bm	42	55	
	-	R4450	M	Phrae	bm	42	56	
	-	R4542	F	Loei	bm	42	56	
	+	R4075	M	Loei	bm	42	57	

26

	-	R4033	F	Phrae	bm	42	58	
	-	R4035	F	Phrae	bm	42	59	
	-	R4100	F	Loei	bm	42	59	
Rattus exulans	+	R4103	M	Loei	bm	42	58	
	-	R4104	M	Loei	bm	42	?	
	-	R4112	M	Loei	bm	42	54	
	+	R4140	F	Phrae	bm	42	?	
	-	R4216	M	Loei	bm	42	58	
Bandicota indica	+	R4000	M	Kalasin	bm	45	?	
	+	R4265	F	Loei	bm	44	?	
	+	R4141	M	Phrae	bm	43	?	
	+	R4142	M	Phrae	bm	43	?	
Bandicota savilei	+	R4143	M	Phrae	bm	45	?	
	+	R4017	F	Phrae	bm	43	?	
	+	R4408	F	Loei	cc	43	60	1
	+	R4021	M	Phrae	bm	43	?	
	+	R5239	M	Loei	cc	42	54	
	+	R5240	M	Loei	cc	42	54	
Leopoldamys edwardsi	+	R4070	M	Loei	bm	42	54	
	+	R4098	M	Loei	bm	42	54	
	+	R4222	F	Loei	bm	42	54	
Leopoldamys neilli	+	R4477	F	Phrae	bm	44	50	
	+	R4099	M	Loei	bm	52	64	
	+	R4107	M	Loei	bm	52	64	
Maxomys surifer	+	R4108	M	Loei	bm	52	64	
	+	R4223	M	Loei	bm	52	64	
	+	R4259	F	Loei	bm	52	64	
	+	R4404	M	Loei	cc	52	64	
	+	R4071	M	Loei	bm	46	52	
Niviventer fulvescens	+	R4409	M	Loei	cc	44 (rearr.)		
	+	R4519	F	Loei	bm	46	52	
Berylmys berdmorei	+	R4266	F	Loei	bm	40	62	
	+	R4406	F	Loei	cc	41	64	1
Berylmys bowersi	+	R4102	F	Loei	bm	40	62	
	+	R4400	M	Loei	cc	40	62	
Hapalomys delacouri	+	R5237	M	Loei	cc	48	92	

Metaphase spreads were obtained from cell lines established from tail, ribs or ear clips using standard protocols. Fibroblasts were cultured using DMEM (supplemented with 15% foetal calf serum) and/or Amniomax (Gibco) culture medium; incubation was at 37°C with 5% CO_2. Chromosome preparations for several species (refer to Table 2.2) were obtained from bone marrow harvests prepared in the field using the air-drying method on femoral bone marrow following yeast stimulation (Lee and Elder, 1980). Species

identifications were based on correspondence in both morphological (refer to Marshall, 1977; Corbet and Hill, 1992; Aplin *et al.*, 2003a, b) and molecular data (see Pagès *et al.*, 2010 for DNA-based analyses).

The application of G- and C-banding followed conventional protocols (Seabright, 1971; Sumner, 1972). For G-band analysis, air-dried chromosome spreads were aged overnight at 37°C followed by incubation in a drying oven at 65°C for 3 hrs to further "harden" the chromosomes for trypsin treatment. C-banding entailed ageing air-dried slides overnight at 37°C, their placement in 0.2 M hydrochloric acid for 30 min and a 5% barium hydroxide treatment for 20 s – 5 min, depending on the species analysed. This was followed by exposure to a 2x SSC solution (0.3 mol/NaCl, 0.03 mol/l sodium citrate) at 55°C for 30 min. The preparations were stained with a 5% giemsa buffered solution (pH 7.0). Images were captured using Genus 3.7 software (Applied Imaging). A minimum of ten complete metaphases were analysed for each specimen. Diploid (2n), fundamental (NF; i.e., total numbers of chromosomal arms in one cell) and autosomal fundamental (NFa; i.e., NF excluding sex chromosomes) numbers were determined, and karyotypes were ordered in decreasing size or according to the *R. norvegicus* format.

RESULTS

All specimens were analyzed using conventional staining thus allowing an overall description of the species-specific karyotypes (Table 2.2). In addition, G-banded karyotypes were constructed where feasible. However, in several instances slight variation and/or ambiguities in NF/NFa values were detected through conventional staining (Giemsa only) reflecting the difficulty in categorizing acrocentric/subtelocentric chromosomes where chromosomal contraction can be a confounding factor. Although not uncommon in mammalian cytogenetics, subtle differences in chromosome morphology (such as telocentric

28

vs. subtelocentric) have been considered important in murine cytotaxonomy where this variability has been used to underpin the species identification of several taxa (e.g., Marshall, 1977, p. 407-408). In an attempt to constrain ambiguities introduced by terminology, all chromosomes that have no visible short arm were viewed as telocentric and those that possess minute arms as acrocentric elements (subtelocentric chromosomes in some schemes) for NF/NFa assessment. If the short arms were more pronounced and clearly and repeatedly observed, chromosomes were considered to be bi-armed (submetacentric and metacentric).

Rattus exulans (Peale, 1848), *R. losea* (Swinhoe, 1871) and *R. tanezumi* Temmink, 1844

All three *Rattus* species have an identical diploid number of 2n = 42 and largely invariant karyotypes. In both *R. exulans* (NFa = 56) and *R. losea* (NFa = 56) the karyotype comprises one large submetacentric chromosomal pair and seven small metacentrics; the remaining autosomes are all acrocentric while the X is a medium sized acrocentric (Figure 2.1). Though it was not possible to unequivocally identify the Y chromosome due to the poor quality of the preparations, the *Rattus* Y chromosome has been previously described as a small acrocentric element in the literature (see for example Gadi and Sharma, 1982 and Hamta *et al.,* 2006). In the case of *R. tanezumi* (NFa = 54), the karyotype is similar to *R. exulans* and *R. losea* except the large submetacentric chromosome pair is absent and presents rather as a large acrocentric pair (pair 1, Figure 2.1). An additional 23 specimens of *Rattus s. s.* were karyotyped (Table 2.2; data not shown). These included *R. losea* (n = 4), *R. tanezumi* (n = 12) and *R. exulans* (n = 7). All but one specimen were determined to have a diploid number of 42 with NF values ranging from 56-61, 58-60 and 54-59 in *R. exulans*, *R. losea* and *R. tanezumi*, respectively. The 2n = 43 *R. tanezumi* specimen was characterised by the presence of a G-negative and C-positive B chromosome (Figure 2.1d), as has previously been documented in this taxon (Yosida, 1976, 1977b; Gadi and Sharma, 1983). The largely

overlapping NF values are attributable to variation in the short arms (p arms) of several pairs

resulting in their classification as acrocentric and/or submetacentric homologs (see Table

2.3). The polymorphisms appear to correspond to variation in pairs 1, 9 and 13 of *R. rattus*

(*sensu* Yosida and colleagues 1971a, b, 1976 and 1977a) (see also Chapter 3).

Table 2.3. Chromosomal polymorphism observed for pairs 1, 9 and 13 (*sensu* Yosida and colleagues; see text for details) in the *Rattus* specimens investigated herein. "M" and "A" correspond to meta-/submetacentric and acrocentric morphs, respectively. As for Table 2.2, individual numbers refer to the CBGP Asian rodent collection (Montpellier, France). "?" indicates unavailable data.

Species	Individuals	Polymorphism in *Rattus* species		
		pair 1	pair 9	pair 13
Rattus losea	R4203	MM	MM	AA
	R4229	MM	MM	AA
	R4230	MM	MM	AA
	R4260	MM	MM	AA
	R4724	MM	AA	AA
Rattus tanezumi	R4113	AA	AA	AA
	R4481	AA	AM	AA
	R4003	AM	AM	AM
	R4008	AA	AM	AA
	R4016	AA	AA	AA
	R4096	AA	AA	AA
	R4145	AA	AA	AA
	R4182	AA	MM	AA
	R4436	AA	?	?
	R4447	AM	AA	AA
	R4450	AM	AM	AA
	R4542	AA	MM	AA
	R4075	AM	MM	AA
Rattus exulans	R4033	MM	MM	AA
	R4035	MM	MM	AA
	R4100	MM	MM	AM
	R4103	MM	MM	AA
	R4104	?	?	?
	R4112	MM	AA	AA
	R4140	?	?	?
	R4216	MM	AA	MM

The C-bands are largely centromeric; however *R. exulans* and *R. losea* show shared

interstitial bands, especially with regards to pair 1 (Figure 2.1). These differ from *R.*

30

tanezumi where C-positive material is largely restricted to the centromeric regions of all chromosomes. In addition, the centromeric C-bands stain darker and are more defined in *R. tanezumi* than either *R. exulans* or *R. losea.*
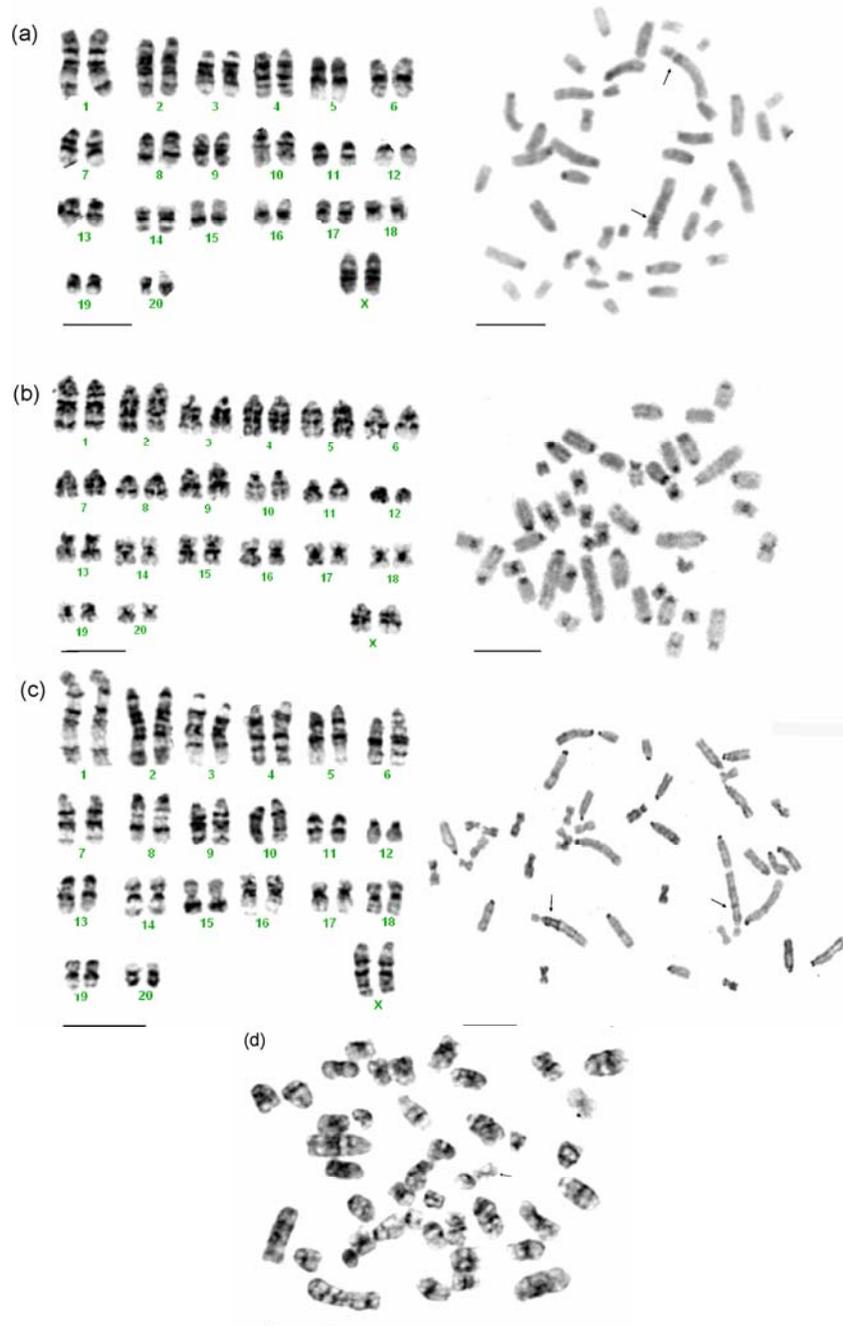
Figure 2.1. G-banded karyotypes and C-banded metaphases of **(a)** *R. exulans,* **(b)** *R. tanezumi,* and **(c)** *R. losea.* **(d)** G-banded metaphase of male *R. tanezumi* R4436 (2n = 43).

31

Arrows show interstitial C-positive bands in pair 1 (panels a and c) and the B chromosome found in *R. tanezumi* (panel d). Scale bar = 10 μm.

Berylmys berdmorei (Anderson, 1879) and *B. bowersi* (Blyth, 1851)

Two *B. berdmorei* female specimens (2n = 40 and 41), and one male and female *B. bowersi* specimen (2n = 40) were karyotyped (Figure 2.2). Both species' karyotypes comprised two large (pairs 1 and 2), three medium-sized (pairs 4, 5, and 6) and one small submetacentric pair of chromosomes (pair 11). Additionally there are one large (pair 3) and seven medium-sized metacentric chromosomal pairs (pairs 12-18), as well as four medium-sized (pairs 7-10) and one small acrocentric pair of chromosomes (pair 19) (NF = 66/NFa = 64 in *B. berdmorei* and NF = 64/NFa = 62 in *B. bowersi*). The X is a medium-size acrocentric chromosome and the *B. bowersi* Y is a small acrocentric element that is entirely C-positive (see Figure 2.2b). The discrepancy between diploid number of the *B. berdmorei* (2n = 41) specimen and the other *Berylmys* specimens (2n = 40) is believed to be due to the presence of an extra C-negative submetacentric element that did not match any of the other elements of the karyotype using G-banding comparison (Figure 2.2) suggesting that this may be a B chromosome. The heterochromatin appears to be reduced in *B. berdmorei* in comparison to *B. bowersi* (Figure 2.2b and d).
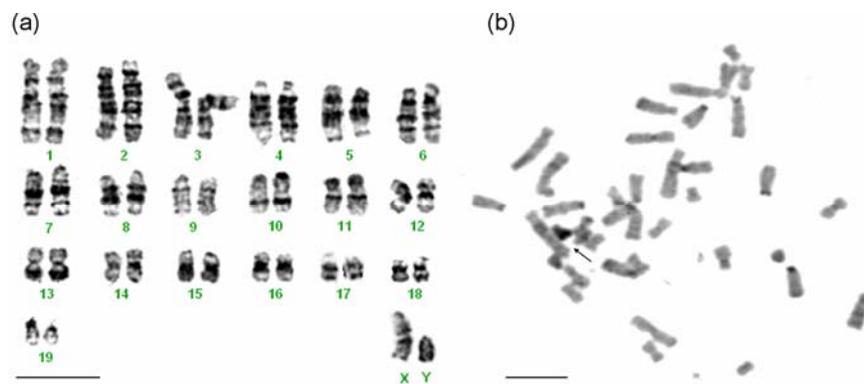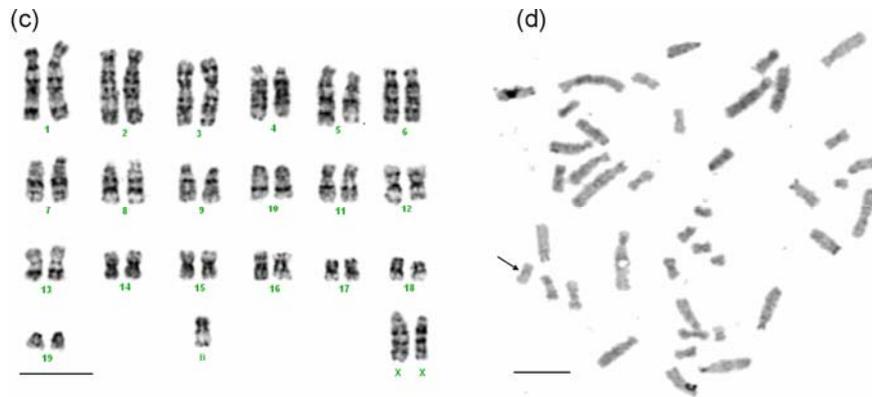
Figure 2.2. **(a)** G-banded karyotype and **(b)** C-banded metaphase of a male *B. bowersi*; as well as the **(c)** G-banded karyotype and **(d)** C-banded metaphase spread of a female *B. berdmorei*. The arrows indicate the C-positive Y chromosome of *B. bowersi* and C-negative B chromosome found in *B. berdmorei*. Scale bar = 10 μm.

Bandicota indica (Bechtein, 1800) and *B. savilei* Thomas, 1916

Eight bandicoot-rats, including two *B. indica* (one male, one female) and six *B. savilei* (four males, two females) specimens were karyotyped. The *B. indica* specimens were characterised by 2n = 44 and 45, respectively. Unfortunately, well resolved G- and C-banded chromosomes could not be produced due to the poor quality of the bone marrow preparations. That said, the 2n = 45 male clearly possessed 26 bi-armed chromosomes (NF = 71; Figure 2.3) including a moderately large submetacentric X. The rest of the karyotype was composed of acrocentric elements, most probably including the Y (which would result in NFa = 68) although its unambiguous identification was not possible from the preparations. The 2n = 44 female was characterised by the same karyotypic morphology, except that it possessed one more acrocentric and one less bi-armed chromosome.

Figure 2.3. G-banded metaphase of a male *B. indica*. Scale bar = 10 µm.

Three males and two females of *B. savilei* were characterised by an identical diploid number of 43. The diploid number of an additional male specimen was determined to be 2n = 45. Unfortunately, as with *B. indica*, the preparations were of very poor quality precluding detailed analysis. Consequently the fundamental number could not be determined confidently from these five specimens. Nevertheless, it proved possible to establish a cell line for one of the female specimens allowing for a more detailed investigation of the *B. savilei* karyotype. This specimen was characterised by 2n = 43 (Figure 2.4a) and has 23 meta- and submetacentric elements, and the remaining chromosomes are acrocentric (NF = 64, NFa = 60). Among the bi-armed chromosomes, two correspond to the large metacentric X pair (as confirmed by Zoo-FISH using a *R. norvegicus* X paint; see Chapter 3), and one submetacentric element was a B chromosome – the first report of such an occurrence in this species. C-banding showed that the X-long arms as well as several acrocentric and metacentric centromeric regions were C-positive (Figure 2.4b). More surprisingly, the B chromosome did not appear strongly stained following C-band analysis (Figure 2.4). Although the Y chromosome could not be unambiguously identified, it has been previously described as a small acrocentric element by Markvong *et al.* (1973).

34
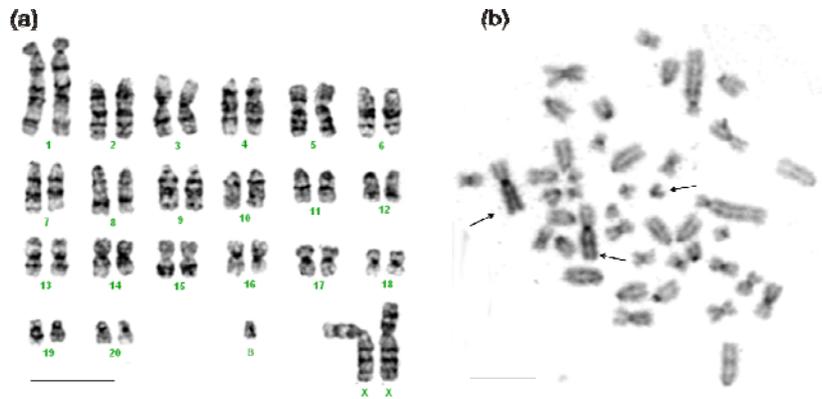
Figure 2.4. **(a)** G-banded karyotype and **(b)** C-banded metaphase of a female *B. savilei*. Arrows indicate the X chromosomes and C-negative B chromosome. Scale bar = 10 μm.

Leopoldamys edwardsi (Thomas, 1882) and *L. neilli* (Marshall Jr, 1976)

The five *L. edwardsi* specimens investigated in this study all display identical karyotypes with 2n = 42 chromosomes. Most of the autosomes are acrocentric, except for four of the five largest autosomal pairs (pairs 1, 2, 3 and 4 in Fig. 2.5a) which possess short arms, as well as the three smallest ones which are metacentric (pairs 18, 19 and 20). The X is a medium size acrocentric chromosome, whereas the Y is C-positive and dot-like (and most likely acrocentric) (NF = 56/NFa = 54). Clearly observable blocks of C-positive heterochromatin were detected in several pericentromeric and telomeric regions, as well as in an interstitial position on one of the largest pairs (Figure 2.5b; pair 2 on Figure 2.5a).
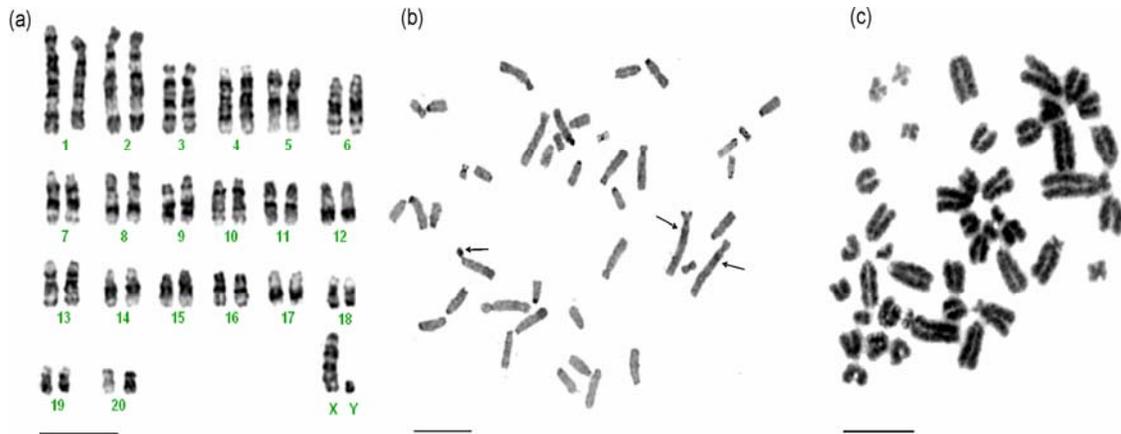
Figure 2.5. (a) G-banded karyotype, **(b)** C-banded metaphase of a male *L. edwardsi* and **(c)** conventionally stained metaphase of a female *L. neilli*. Arrows indicate C-positive interstitial bands and minute Y. Scale bar = 10 µm.

A single *Leopoldamys* female trapped in the vicinity of limestone karsts was referred to as *L. neilli* on the basis of morphology as well as unambiguous molecular results (Pagès *et al.*, 2010). Even though the preparation was of poor quality, ruling out G-banding, the specimen unambiguously displayed 2n = 44 with two large submetacentric and two small metacentric pairs of chromosomes (NF = 52, Figure 2.5c).

Niviventer fulvescens **(Gray, 1847)**

Of the three specimens analysed, a single male specimen was used to establish a fibroblast culture but the resulting lines appeared to have undergone *de novo* rearrangement in culture (2n = 44). Nonetheless, the rearranged karyotype (Figure 2.6) corresponds relatively well to the bone marrow karyotype of the other female specimen which was characterised by 2n = 46 (NF = 54/NFa = 52; Figure 2.6). All chromosomes are acrocentric except for the submetacentric pair 2, as well as the three of the shortest autosomal pairs which are metacentric. The X is a large acrocentric chromosome, and the Y is a tiny acrocentric element (Figure 2.6).
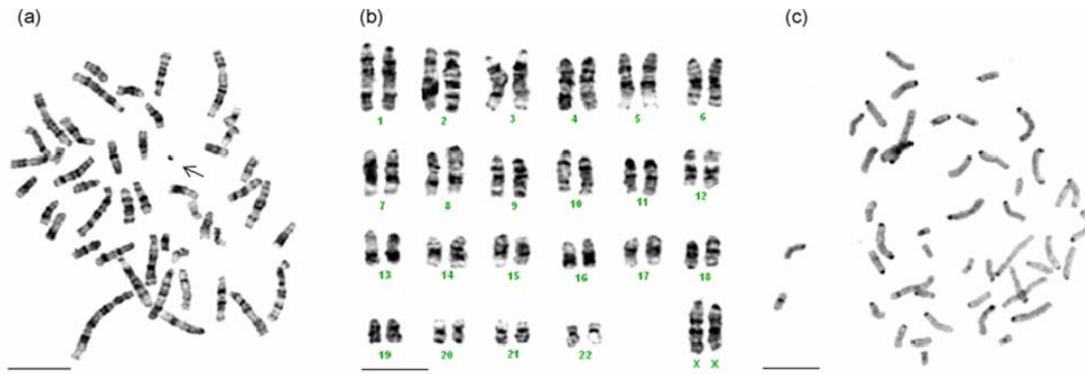
Figure 2.6. **(a)** G-banded metaphase of a male *N. fulvescens* from a *de novo* rearranged cell line. **(b)** G-banded karyotype and **(c)** C-banded metaphase of a female *N. fulvescens*. Arrow indicates minute Y. Scale bar = 10 µm.

Maxomys surifer **(Miller, 1900)**

Five males and one female of *M. surifer* were analysed and the G-banded karyotype of *M. surifer* is presented for the first time (Figure 2.7a). The species has a diploid number of 2n = 52 that comprises three metacentric pairs (pairs 4, 5 and 14 in Figure 2.7) and four submetacentric (pairs 10, 12, 18 and 24) chromosomal pairs (NF = 66/NFa = 64). The X chromosome is acrocentric, while the Y chromosome, one of the smallest elements, is entirely heterochromatic (Figure 2.7b), and appears acrocentric in morphology.
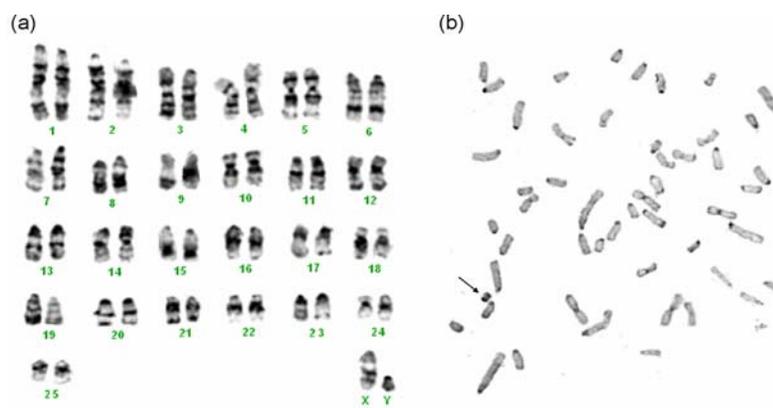
Figure 2.7. **(a)** G-banded karyotype and **(b)** C-banded metaphase of a male *M. surifer*. The arrow indicates the C-positive Y chromosome. Scale bar = 10 µm.

37

Hapalomys delacouri **Thomas, 1927**

A fibroblast cell culture was established from a male *H. delacouri* specimen and its karyotype was characterised by 2n = 48 chromosomes and an NFa = 92 (NF = 95; Figure 2.8a). All the autosomes were bi-armed (metacentric or submetacentric). The metacentric X and the acrocentric Y were easily recognizable since they were, respectively, the largest and one of the smallest unpaired elements in the karyotype. Interestingly, two classes of C-positive material were identified on staining intensity suggesting a different repeat composition for this heterochromatic material (Figure 2.8b). The genome appears to be particularly rich in heterochromatin with some chromosomes displaying entirely heterochromatic long arms. Zoo-FISH using a *R. norvegicus* X paint (see Chapter 3) showed that the metacentric X was the result of an X-autosome translocation (i.e., it has a XY1Y2 chromosomal composition). Similarly, the extra C-negative metacentric element that did not match any of the other elements of the karyotype using G-banding was, following FISH (Chapter 3), shown to be a B chromosome (Figure 2.8a).
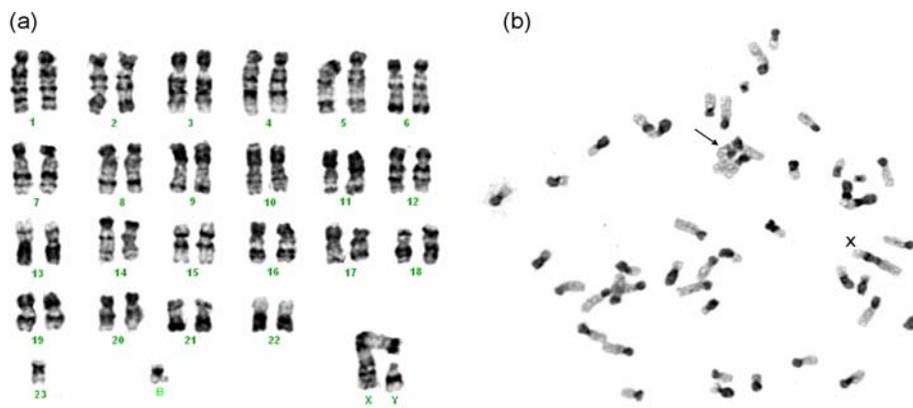
Figure 2.8. **(a)** G-banded karyotype and **(b)** C-banded metaphase of a male *H. delacouri*. The arrow indicates the C-negative Y chromosome. Scale bar = 10 μm.

DISCUSSION

***Rattus exulans* (Peale, 1848), *R. losea* (Swinhoe, 1871) and *R. tanezumi* Temmink, 1844**

The karyotypes of *R. exulans, R. losea* and *R. tanezumi* correspond well to the conventionally stained and G-banded karyotypes previously reported in the literature (Yosida *et al.,* 1971; Yosida and Sagai, 1972; Markvong *et al.,* 1973; Yosida, 1973; Yosida and Sagai, 1973, Levan, 1974; Marshall, 1977; Tsuchiya *et al.,* 1979; Gadi and Sharma, 1983; Baskevich and Kuznetsov, 1998). Although all three *Rattus* species investigated herein are characterized by 2n = 42, several geographical variants have been reported. For example, the *R. rattus* 2n = 38 arrangement that conforms to the Oceanian variant, described by Yosida *et al.* (1974) and the Ceylon type (2n = 40) that is identical to the 2n = 38 arrangement except for an additional chromosome pair that arose from a centric fission event (see Bastos *et al.,* 2011).

The largely overlapping NFa values identified between *R. exulans* (NFa = 54-59)*, R. losea* (NFa = 56-58) and *R. tanezumi* (NFa = 54-59) karyotypes are attributable to variation in the acrocentric and submetacentric morphology of pairs 1, 9 and 13 of Yosida and colleagues' karyotypic descriptions (1971a, b, 1976 and 1977a; see Table 2.3)*.* This variation in centromere position has been noted in several *Rattus* species and subspecies (Yosida *et al.,* 1971a, 1971b, 1974; Yosida, 1973, 1977a; Yosida and Sagai, 1973, 1975; Gadi and Sharma, 1983; Motokawa *et al.*, 2001) and is thought to result from pericentric inversion (Yosida *et al.*, 1971a, 1971b; Yosida, 1976, 1977a) and/or heterochromatin amplification/deletion (Yosida and Sagai, 1975; Baskevich and Kuznetsov, 1998). C-band analysis shows that the discrepancy seen between *R. tanezumi* to *R. losea* and *R. exulans* with respect to the acrocentric vs. submetacentric morphology of pair 1 is not due to heterochromatin variability (Figure 2.1). The chromosomes and/or the segments involved are

39

small and thus difficult to interpret, even when using G-banding, however BAC-mapping analysis conducted in the present investigation (see Chapter 3) demonstrated that the variation is attributable to pericentric inversions.

Importantly, Marshall (1977) proposed a number of apparently fixed (diagnostic) features among *R. exulans*, *R. losea* and *R. tanezumi* (which is synonymous with *R. r. thai sensu* Marshall; see Wilson and Reeder, 2005) in Thailand based on the morphology of these three chromosome pairs. On the basis of conventional staining, Marshall, following the chromosome nomenclature used by Yosida and colleagues, argues (pp. 407-408) that: (i) pair 1 is acrocentric in *R. tanezumi* and submetacentric in *R. exulans* and *R. losea*, (ii) pair 9 is submetacentric in *R. losea* but acrocentric in *R. exulans* and *R. tanezumi*, and (iii) pair 11 (most probably corresponding to pair 13 *sensu* Yosida and colleagues) is submetacentric in *R. exulans*, whereas it is acrocentric in *R. losea* and *R. tanezumi*. However, taking other studies into account (see above) and our own data (Table 2.3), none of these chromosomal morphs (1, 9 or 13) can be considered as species-specific. Moreover, heterozygotes for each of these three pairs have been reported in *R. exulans* (Table 2.3) and *R. tanezumi* (Table 2.3 and Chapter 3 – this study) suggesting that these chromosomal polymorphisms were present in their last common ancestor and are therefore ancestral to the *Rattus s. s.* clade (Yosida *et al.*, 1971a, b; Yosida and Sagai, 1975, Yosida, 1977a), and date back ~2.2 mya (Robins *et al.*, 2008).

Berylmys berdmorei (Anderson, 1879) and *B. bowersi* (Blyth, 1851)

Data are limited for these species as karyotypes are known only from conventionally stained chromosomal preparations (Yong, 1968; Markvong *et al.*, 1973; Duncan *et al.*, 1970; Marshall, 1977; Yosida, 1980). The findings reported in this study are in good agreement with previous reports (reviewed in Musser and Newcomb, 1983)

indicating a 2n = 40 and NFa = 64 for both species. However, an additional female *B. berdmorei* specimen was found to have 2n = 41 due to the presence of a B chromosome. B chromosomes vary from species to species (Vujošević and Blagojević, 2004) and the staining intensity identified here most probably reflects underlying compositional differences of the satellite repeat families present in the two types of B chromosomes (C-negative and C-positive), as well as differing chromatin compaction. It should be pointed out, however, that B chromosomes are very poorly documented, especially in mammals. In particular, fine-scale analysis using recent molecular methods (i.e., gene contents, epigenetics) are lacking, thus making any discussion about the origin of the various C-band staining intensities speculative.

The results shown here differ in some respects from those of Yong (1968) and Yosida (1973) who identified six pairs of acrocentric, eight pairs of metacentric and five pairs of submetacentric autosomes. They also show a medium sized acrocentric X chromosome and a small acrocentric Y in specimens of *B. bowersi*. This mirrors other conflicting reports regarding the *B. berdmorei* karyotype. Marshall (1977) described six telocentric, six subtelocentric and seven metacentric pairs of autosomes in *B. bowersi* from Thailand, whereas Yong (1968) recorded eight pairs of metacentrics, nine pairs of submetacentric, two pairs of acrocentric autosomes and a medium sized submetacentric X in specimens from Malaysia. In contrast, Markvong *et al.* (1973) and Duncan *et al.* (1970) reported five pairs of submetacentrics, eight pairs of metacentrics and six pairs of acrocentrics in this species from Thailand and from Vietnam. The X and Y chromosomes of *B. bowersi* described by Yong (1968) and Yosida (1973) are comparable to those of the present study. Conversely the *B. berdmorei* X has previously been reported as submetacentric (Yong, 1968), while our specimens clearly possessed an acrocentric X which is in agreement with both Duncan *et al.* (1970) and Markvong *et al.* (1973). Furthermore, the data from the present study does not

support the species-specific differences that have been proposed to differentiate the karyotypes, i.e., a submetacentric vs. acrocentric pair 6, and an acrocentric vs. submetacentric pair 11 in *B. berdmorei* and *B. bowersi*, respectively (see Marshall, 1977). Given that the C-banding data show the presence of several C-positive short arms (Figure 2.2b and c) which may be responsible for variations in chromosome morphology (i.e., the interpretation of acrocentric vs. submetacentric), it seems prudent to conclude that the two *Berylmys* species cannot be distinguished cytogenetically with any confidence.

Bandicota indica (Bechtein, 1800) and *B. savilei* Thomas, 1916

The difference between the 2n = 44 and 2n = 45 karyotypes of *B. indica* specimens analysed here suggest that a Robertsonian polymorphism may be present in *B. indica*. Based on the limited data, however, it is not possible to reach a definitive conclusion. Consequently, other explanations (e.g., X-autosome translocations, B chromosomes) cannot be excluded at this stage. For example, *B. indica* has been previously reported to contain 1-3 B chromosomes (Gadi *et al.,* 1982; Gadi and Sharma, 1983). Importantly, however, our results are slightly different to those previously published for this species collected from Thailand (2n = 46; Markvong *et al.,* 1973; Marshall, 1977), but are consistent with data on specimens from India (2n = 42 for *B. indica indica*, and 2n = 44 for *B. indica nemorivaga*; Ranjini, 1967; Avirachan *et al*., 1971; Gadi and Raman, 1977; Gadi and Sharma, 1983). When taken together, the variation in 2n suggests a more complex systematic situation than currently thought within *B. indica*, and/or that there is a chromosomal polymorphism that remains to be described.

The *B. savilei* karyotype described herein differs from previous reports on specimens collected in Thailand (2n = 44, NFa = 58; Markvong *et al*., 1973). Markvong *et al*. (1973) described a karyotype consisting of six pairs of small metacentrics and two pairs of

submetacentrics and thirteen pairs of acrocentrics, in comparison to the karyotype presented here which has 23 meta- and submetacentric elements and only ten pairs of acrocentrics. Furthermore, the X of *B. savilei* was previously identified as a medium-sized acrocentric which differentiates it from other lesser bandicoot rats (Markvong *et al*., 1973; Gadi and Sharma, 1983). However, the X was identified unambiguously in the present study by Zoo-FISH analysis (see Chapter 3) as a large submetacentric chromosome. Therefore, males of *B. savilei* should have a 2n = 43+1B karyotype (i.e., XY1Y2 males), which was confirmed by Zoo-FISH analysis (Chapter 3).

Leopoldamys edwardsi (Thomas, 1882) and *L. neilli* (Marshall Jr. 1976)

The *L. edwardsi* specimens investigated in this study had 2n = 42, in agreement with other reports (Yong, 1970; Rickart and Musser, 1993 and references therein) although *L. edwardsi* (from an unknown locality) was recently proposed to have 2n = 44 chromosomes (Sablina *et al.*, unpubl., quoted in the "Atlas of Mammalian Chromosomes", ed. S.J. O'Brien *et al.*, 2006). In addition, there is a discrepancy in the morphology of the Y chromosome: the Y identified by Yong (1970) is an average-sized acrocentric, while here the Y clearly appears as a dot like element (Figure 2.5), suggesting there may be more fundamental differences underlying these observations. Geographically these may underpin the separation of Malaysian populations of *L. edwardsi* (Yong, 1970) from Asian populations of *L. edwardsi* (this study). The findings for *L. neilli* in the present investigation are in good agreement with previous reports (NF = 52 in Marshall, 1977), but differ from the NF = 50 noted by Rickart and Musser (1993).

It would appear, therefore, that the two *Leopoldamys* species can be distinguished by karyotypes: *L. edwardsi* with 2n = 42/NF = 56 and *L. neilli* with 2n = 44/NF = 52. A third species, *L. sabanus*, which is morphologically similar to *L. edwardsi*, has also been described

from Thailand (Wilson and Reeder, 2005). The *L. sabanus* karyotype differs slightly from that of *L. edwardsi* since Marshall (1977) recorded a 2n = 42/NF = 54 complement for this species. Unfortunately, no material was available to examine the delimitation of the *L. edwardsi*/*L. sabanus* species on karyotypic data. However, it can be anticipated that karyotypic differences between these two taxa are subtle, therefore unreliable for diagnostic purposes.

Niviventer fulvescens (Gray, 1847)

The taxonomy within *Niviventer* is in need of revision (Wilson and Reeder, 2005). The results of the preliminary molecular (Pagès *et al.*, 2010) and morphological (Chaval *et al.* unpubl.) investigations are concordant and indicate that the specimens analysed here form a natural grouping referable to *N. fulvescens* (Pagès *et al.*, 2010). There are only a few published karyotypic reports on subspecies/species of *Niviventer* and these are limited to conventionally stained preparations. The only exception to this is *N. confucianus* (Wang *et al.*, 2003). Except for tiny differences in centromeric placement, *N. rapit*, *N. bukit*, *N. hinpoon* (Yong, 1969; Markvong *et al.*, 1973; Marshall, 1977), *N. langbianis* (Baskevich and Kuznetsov, 2000), *N. fulvescens* (Duncan *et al.*, 1970, 1974) and *N. cremoriventer* (Yong, 1969) are all invariant and closely resemble the cytotype that is presented here for *N. fulvescens*. It is important to note that *N. rapit* and *N. bukit* from Thailand have since then been collapsed into *N. fulvescens* (see Musser, 1981; Abe, 1983). The cytogenetic criteria proposed to discriminate all these species are based on subtle shape differences of some chromosome pairs (see Table 2 in Wang *et al.*, 2003) which are most probably due to variation in heterochromatin, as observed between *N. tenaster*, *N. fulvescens* and *N. langbianis* (all from Vietnam) following C-banding protocols (Baskevich and Kuznetsov, 2000). Although it cannot be stated with any certainty, it is highly probable that karyotypic

44

variation within the genus will be subtle making reliable cytotaxonomic identification problematic.

Maxomys surifer (Miller, 1900)

The chromosome complement identified herein is in agreement with available data for *M. surifer* from Southeast Asia (see Rickart and Musser, 1993 and references therein) including Thailand (Marshall, 1977), all of which were determined by conventional staining. However, some differences exist between these studies and the findings reported here, especially with respect to the identification of the X chromosome. The X was described as a metacentric in specimens collected from localities in Thailand (Marshall, 1977), Vietnam and Malaysia (see Rickart and Musser, 1993 and references therein), resulting in 2n = 52/NFa = 62. In sharp contrast, the present study shows the X chromosome to be one of the largest chromosomal elements and is clearly acrocentric in morphology. Zoo-FISH analysis using a *R. norvegicus* X painting probe (Chapter 3) confirmed this observation. The *M. surifer* specimen is consequently characterised by NFa = 64, and not by NFa = 62. When one considers the morphology of the X chromosome, the difference in NFa (NFa = 62 vs. 64) may be due to (i) a misidentification of the X in the past, (ii) a polymorphism that would involve the X chromosome, or (iii) the co-existence of different but cryptic *Maxomys* species that differ in the morphology of the X chromosome. Further sampling and karyotyping is required to address this point.

Hapalomys delacouri Thomas, 1927

There is very little karyotypic data on *Hapalomys* with a single report available for *H. longicaudatus* (Yong *et al.*, 1982). The present findings differ significantly from those on *H. longicaudatus* in that Yong *et al.* (1982) describe specimens from Malaysia as having 2n = 50, with the X and Y chromosomes being of similar morphology, and all the autosomes as

acrocentric. This suggests that cytotaxonomy may be a valuable diagnostic tool within the genus. These results constitute the first karyotypic data for this very poorly documented species.

CONCLUSIONS

Despite a central role as reservoirs and/or vectors of several human diseases, reliable systematic data necessary to underpin accurate species-specific geographical, ecological and genetic information on Asian rodents are limited and many taxonomic associations remain unclear. Cytotaxonomy has been of great help in similar epidemiological investigations on other continents i.e., Africa, frequently providing unambiguous identification of species including those that are morphologically cryptic (e.g., Duplantier *et al.*, 1990; Volobouev *et al.*, 2002; Dobigny *et al.*, 2003a; Granjon and Dobigny, 2003; Rambau *et al.*, 2003; Veyrunes *et al.*, 2004). However, apart from *Hapalomys*, the preliminary results presented here suggest that this does not hold for the Asian species analysed in this study. Although most genera can routinely be distinguished on karyotype, few species-specific chromosomal characters could be identified. For instance, although *L. neilli* and *L. edwardsi* can be delimited cytogenetically, the two *Berylmys* and the three *Rattus* species analysed are indistinguishable on karyotype. Although it remains to be confirmed, a similar situation may exist within *Niviventer*, *Maxomys*, *Bandicota*, as well as between *L. edwardsi* and *L. sabanus*. These findings suggest that new diagnostic features should be sought as a matter of urgency for use in these epidemiologically important rodent groups. DNA-based methods such as primer-specific analysis that have been used successfully in Africa (e.g., Lecompte *et al.*, 2005) and Europe (e.g., Michaux *et al.*, 2001) are compelling candidates that should be prioritised for testing in Asian murids.

CHAPTER 3

CHROMOSOMAL PHYLOGENY, EVOLUTION AND BAC CLONE ANALYSIS OF ASIAN RODENTS (RODENTIA, MURIDAE)

INTRODUCTION

Chromosomal rearrangements are believed to play a pivotal role in speciation within Rodentia (Rumpler and Dutrillaux, 1990; King, 1993; Searle, 1998; Richard *et al.*, 2003a), the most species-rich order of mammals (Wilson and Reeder, 2005; see Introduction – Chapter 1). The *Rattus s. l.* complex falls within the suborder Myomorpha which contains a significant number of species that are characterised by extraordinary chromosomal variation and high rates of chromosomal evolution (e.g., Graphodatsky, 1989; Murphy *et al.,* 2001; Romanenko *et al.,* 2006, among others).

One of the most powerful methods available in modern comparative cytogenetic studies involves the use of cross-species chromosome painting which is used to determine the extent of chromosomal homologies among taxa. This has, however, been successfully applied to only a few species within Myomorpha, and almost all of these fall within Muridae. Examples include instances of cross-species chromosome painting using mouse specific painting probes against *Cricetulus griseus* (Yang *et al.,* 2000; Romanenko *et al.,* 2007), *Mesocricetus auratus* (Romanenko *et al.,* 2006), *R. norvegicus* (Guilly *et al.,* 1999; Stanyon *et al.,* 1999), *R. rattus* and *R. frugivorous* (Cavagna *et al.,* 2002), *Rhabdomys pumilio* (Rambau and Robinson, 2003), *Otomys irroratus* (Engelbrecht *et al.,* 2006), seven *Apodemus* species (Matsubara *et al.,* 2004) and, within *Mus*, *Nannomys, Coelomys* and *Pyromys* (Veyrunes *et al.,* 2006). The relationship between mouse and rat, in particular, has been

47

extensively investigated (Grutzner *et al.,* 1999; Guilly *et al.,* 1999; Stanyon *et al.,* 1999; Helou *et al.,* 2001; Nilsson *et al.,* 2001; Cavagna *et al.,* 2002). However, somewhat surprisingly, no comparisons have been conducted within the *Rattus s. l.* complex using chromosome painting. The only exception to this is the study by Cavagna *et al.* (2002) who reported the complete orthology of 16 autosomal chromosomes of *R. rattus* (2n = 38) and *R. frugivorous* (2n = 38) with *R. norvegicus* (2n = 42).

In these types of studies it is often useful to use painting probes developed from index species such as the laboratory mouse (*M. musculus*) or rat (*R. norvegicus*) since this provides access to annotated sequence assemblies (Serikawa, 1992; Yamada *et al*., 1994; Jacob *et al*., 1995; Toyota *et al*., 1996; Steen *et al*., 1999; Aitman *et al*., 2008; i.e., Rat Genome Database, http://rgd.mcw.edu/, and Mouse Genome Informatics, http://www.informatics.jax.org/) that are generally absent for most taxa. Nonetheless, their selection as donor species for the production of chromosome paints is not without complications. For example, the use of mouse painting probes can be problematic due to the rearranged nature of the mouse karyotype (for example Bourque *et al.,* 2004; Murphy *et al.,* 2005) resulting in difficulties when interpreting hybridization patterns (Romanenko *et al*., 2007). In an attempt to avoid this, and given the high degree of synteny reported between *R. norvegicus* and *R. rattus* (Cavagna *et al*., 2002), painting probes isolated from the Norway rat (*R. norvegicus*) were utilised to identify conserved syntenies and chromosomal characters among representatives of the *Rattus s. l.* complex and *H. delacouri. Hapalomys delacouri* was selected for inclusion in this study as outgroup taxon based on its apparent basal position in Murinae as determined by molecular analyses (see Materials and Methods).

In addition to the commercially available *R. norvegicus* painting probes, *M. surifer* and *R. rattus* chromosomes were flow-sorted, characterised and used for reciprocal

chromosome painting schemes allowing for the unambiguous identification of synteny junctions between these taxa. Additionally *M. surifer* and *R. rattus* paints were applied to a broader set of taxa in order to scrutinize synteny junctions that were identified by non-reciprocal Zoo-FISH analysis with whole *R. norvegicus* chromosome paints. These data (shared syntenic junction breakpoints) can be scored as cytogenetic markers for the retrieval of phylogenetic relationships within Rattini following the principles outlined in Dobigny *et al.* (2004a). This approach allowed (i) an assessment of both the mode and tempo of karyotypic evolution of *Rattus s. l.*; (ii) the reconstruction of the Rattini ancestral karyotype and (iii) provides additional insights into the ancestral chromosomal elements that possibly comprised the Murinae ancestral karyotype.

An additional focus of this investigation involved the use of BAC clones to examine the intrachromosomal morphological differences identified among pairs 1, 12, and 13 of the *Rattus* karyotypes (orthologs of pairs 1, 9 and 13 of Yosida's karyotypes, see Chapter 2). Opinion is divided as to whether the morphological differences detected among different species are the result of heterochromatin amplification/deletion or pericentric inversions (Chapter 2 and references therein). Since a significant shortcoming of chromosome painting using whole chromosome paints is the inability to identify intrachromosomal rearrangements, the BAC-based analysis contributed significantly to clarifying the underlying cause of the variable chromosome morphology of pairs 1, 12 and 13 detected between *Rattus* taxa. This was considered important in the present investigation since pericentric inversions may be useful binary characters whereas heterochromatin can be highly variable and often homoplasic in nature (Dobigny *et al.,* 2004a). Moreover, heterochromatin addition/deletion and inversions may have different evolutionary consequences, i.e., silencing of genes located close to heterochromatic blocks (the so-called Position Effect Variegation; Grewal and Moazed, 2003) while inversions may modify the linear order of chromatin and potentially

impact on gene function (Patton and Sherwood, 1983; Brown and O'Neill, 2010). In addition, in some instances inversions may cause a local adaptive advantage by suppressing recombination and creating a partial reproductive barrier (i.e., the suppressed recombination speciation model, Brown and O'Neill, 2010).

MATERIAL AND METHODS

Tissue samples and cytogenetics

Material was collected in 2006 and 2007 in Thailand (ANR program 00121-05, from the French Ministry of Research, see Chapter 2) (Table 3.1). Establishment of fibroblast cell lines, chromosome harvests and slide preparation followed the same protocols as those outlined in Chapter 2. Chromosomes were counterstained with 4',6-diamidino-2-phenylindole (DAPI, Vector Laboratories) to facilitate identification. Chromosomes of *Rattus s. l.* and *H. delacouri* were ordered according to Badenhorst *et al.* (2009; Chapter 2).

Table 3.1. List of species and specimens included in the present investigation; sampling localities and grid references are provided for each locality.

Species	Specimen	Sex	Origin	Grid reference	Material	2n
Rattus losea	R4724	F	Loei	$17^{o}29$'N, $101^{o}43$'E	cc	42
Rattus tanezumi	R4003	F	Kalasin	$16^{o}49$'N, $103^{o}53$'E	bm	42
	R4008	M	Kalasin	$16^{o}49$'N, $103^{o}53$'E	bm	42
	R4182	F	Phrae	$18^{o}09$'N, $100^{o}08$'E	bm	42
	R4075	M	Loei	$17^{o}29$'N, $101^{o}43$'E	bm	42
Rattus exulans	R4033	F	Phrae	$18^{o}09$'N, $100^{o}08$'E	bm	42
	R 4035	F	Phrae	$18^{o}09$'N, $100^{o}08$'E	bm	42
Bandicota savilei	R4143	M	Phrae	$18^{o}09$'N, $100^{o}08$'E	bm	45
	R4408	F	Loei	$17^{o}29$'N, $101^{o}43$'E	cc	43
Berylmys berdmorei	R4406	F	Loei	$17^{o}29$'N, $101^{o}43$'E	cc	41
Berylmys bowersi	R4400	M	Loei	$17^{o}29$'N, $101^{o}43$'E	cc	40

Leopoldamys edwardsi	R5239	M	Loei	17°29'N, 101°43'E	cc	42
	R5240	M	Loei	17°29'N, 101°43'E	cc	42
Niviventer fulvescens	R4409	M	Loei	17°29'N, 101°43'E	cc	#44 (rearr.)
	R4519	F	Loei	17°29'N, 101°43'E	bm	46
* *Maxomys surifer*	R4404	M	Loei	17°29'N, 101°43'E	cc	52
Hapalomys delacouri	R5237	M	Loei	17°29'N, 101°43'E	cc	48
* *Rattus rattus*	JPQ3554	F	Clapiers, France	43°39'47"N, 03°53'34"E	cc	38

Specimen numbers refer to the CBGP Asian rodent collection (Montpellier, France). Diploid (2n) number and type of biological material used (bm, bone marrow; cc, cell culture) are indicated for each specimen. # refer to Chapter 2. * The *R. rattus* and *M. surifer* specimens used for the production and characterisation of flow-sorted chromosome paints.

Flow-sorting and probe labelling

Chromosome-specific painting probes were prepared from flow-sorted *R. norvegicus, M. surifer* and *R. rattus* fibroblasts and provided by the Cambridge Resource Centre for Comparative Genomics, UK. Chromosomes were separated on AT:GC ratio and size (Stanyon *et al.,* 1999). Flow-sorted chromosomes were amplified using 6MW primers and DOP-PCR (Telenius *et al.,* 1992). Fluorescent labelling was done using biotin and/or digoxigen-dUTP (Roche).

Flow-sorted peaks were assigned by hybridizing each of the fluorescently labelled flow-sort to DAPI banded metaphase chromosome spreads of the species of origin, i.e., *R. rattus* and *M. surifer*. Double-colour hybridizations were used to resolve ambiguities where chromosomes were also present in another peak (either alone, or associated with other chromosomes). The flow-karyotype and characterised peaks of *M. surifer* and *R. rattus* are shown in Figure 3.1. The *R. norvegicus* flow-sorts were previously characterised by Stanyon *et al.* (1999).
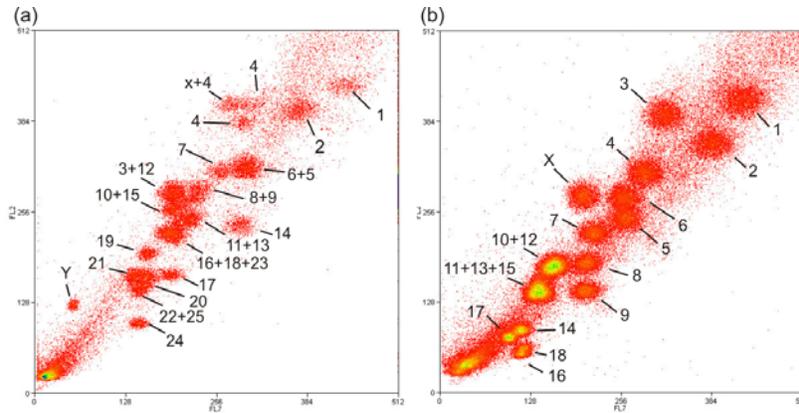
Figure 3.1. Flow-sorted karyotypes of **(a)** *M. surifer* (MSU, 2n = 52, XY) and **(b)** *R. rattus* (RRA, 2n = 38, XX) showing the flow-peaks and their correspondence with the respective chromosomes (see text for details).

Chromosome painting

Cross-species chromosome painting (reviewed in Rens *et al.,* 2006) followed Deuve *et al.* (2006) and Gilbert *et al.* (2006) with slight modifications. A total of 100-150 ng of biotin- or digoxigen-dUTP labelled probe was precipitated together with 50 ng of salmon sperm DNA in 1/10 volume of Na-Acetate and four volumes of 100% ethanol (-70°C for minimum 2 hrs, preferably left overnight). After 15 min centrifugation at 13,000 rpm the pellet was washed in ice-cold 70% ethanol, dried for 30 min at 37°C and resuspended in 15 μl hybridization buffer (50% deionised formamide, 10% dextran sulphate, 2x SSC, 0.5 M phosphate buffer, pH 7.3).

Metaphase chromosomes were cleared of the extraneous cytoplasm through a pepsin pre-treatment that improved probe accessibility. Pepsin pre-treatment entailed incubating slides in the enzyme at 37°C followed by two washes in 2x SSC for 5 min each at room temperature following which the slides were passed through an ethanol series (70, 80, 90 and 100%). Slides were heat-aged at 65°C for 2 hrs and denatured by incubation in 70% formamide/0.6x SSC solution at 65°C for 30 s-1 min depending on the probe/target species used. The slides were subsequently dehydrated through an ethanol series. Preannealed

probes (denatured at 65°C for 10 min and then preannealed by incubation at 37°C for 60 min) were applied to metaphase preparations, cover-slipped and sealed with rubber cement. Hybridization took place in humid chamber at 37°C for two to three (for problematic probes) nights. Two post-hybridization washes in formamide 50%/1x SSC (5 min each) were carried out following hybridization. These were followed by two 5 min rinses in 2x SSC and a final 4XT (4x SSC, 0.05% Tween 20) rinse for 10 min. All post-hybridization washes were conducted at 41-43°C (variation dependent on painting probe used and target species). Fluorochromes were used to visualize labelled paints: Cy3-avidin (Amersham Biosciences) in the case of biotin-dUTP and FITC (Roche) for digoxigen-dUTP. Fluorochrome detection solution comprised 4XT/relevant antibody in a 250 µl final volume at 37°C for 20-45 min. Slides were subsequently washed thrice in 4XT at 37°C, counterstained with DAPI (6 µl DAPI 2 mg/ml in 50 ml 2x SSC), and mounted using an antifade solution (Vectashield).

The images were captured with a CCD camera coupled to an Olympus BX60 fluorescence microscope and analysed using Genus 3.7 Software (Applied Imaging). Signals were assigned to specific chromosomes according to morphology, size and DAPI banding. When DAPI bands were insufficient for chromosome identification, FISH was conducted on previously G-banded metaphase preparations. In these instances G-banded images were captured and the slides subsequently destained in a 100% methanol and ethanol series for 10 min. The times used in the chromosome painting protocol denaturation steps outlined above were accordingly halved (i.e., 30 s instead of 1 min) for previously G-banded metaphase preparations.

Reciprocal chromosome painting between *R. norvegicus* and *R. rattus*, as well as between *M. surifer* and *R. norvegicus*, was applied to corroborate the chromosomal syntenies on one hand, and to allow the more precise identification of some of the breakpoints and their

53

borders on the other. *Maxomys surifer* and *R. rattus* paints were not used in instances where the karyotypes were conserved with respect to *R. norvegicus*. Thus the *M. surifer* and *R. rattus* paints (following reciprocal painting with *R. norvegicus* chromosomes) were used to clarify breakpoint junctions identified using *R. norvegicus* chromosome paints.

Phylogenetic analyses

Comparative chromosome maps generated in this study allowed the identification of segments that were rearranged within the *Rattus s. l.* complex. Parsimony analyses of chromosomal rearrangements were conducted by scoring syntenic disruptions and/or segmental associations (i.e., the characters); their presence/absence as the character state following the principles outlined in Dobigny *et al.* (2004a). In other words, the 'breakpoints' are considered as characters (Robinson and Seiffert, 2004; Trifonov *et al.,* 2008). Zoo-FISH analysis using *R. norvegicus, R. rattus* and *M. surifer* chromosome paints (unidirectional FISH and at times including reciprocal FISH) was conducted in order to assess orthology and define the boundaries of the various syntenies.

Variations in chromosome morphology and centromere position (intrachromosomal rearrangements) were also included in the parsimony analysis and their presence/absence scored as character states. In addition to Robertsonian fusions, gonosome autosome fusions and fissions, several instances involving more subtle differences in chromosomal morphology among orthologs (i.e., no visible short arms vs. prominent short arms) were noted. Four types of rearrangements could theoretically explain these morphological differences, (i) pericentric inversions, (ii) transpositions (iii) centromeric shifts and (iv) neocentromere formation. Little attention has, however, been paid to scoring these characters in phylogenetic studies although the distinction between pericentric inversions and transpositions was made by Robinson *et al.* (1998). Pericentric inversions are two-break rearrangements that results in a reversal of the

54

centromere containing segment, and in an altered gene order. A transposition is a rare type of structural change; it requires three breakpoints followed by a shift of a chromosomal segment to another region of the same chromosome while maintaining the original orientation of the translocated segment (Therman and Susman, 1993). Centromeric shifts, i.e., the strict repositioning of only the centromere with no change in DNA marker order (i.e., O'Neill *et al.,* 2004), would similarly necessitate three breaks. Finally, neocentromere formation (or neocentromerization as it is occasionally referred to), is the newest of the possible explanations for a shift in chromosomal centromere position, and is just beginning to be understood. The seeding of a new centromere requires no breaks as a new centromere simply emerges at some point along the chromosome after the inactivation of the old centromere and, as with a strict centromeric shift, there is consequently no disruption to gene order (see Marshall and Choo, 2009 and Brown and O'Neill, 2010 for review). In the absence of BAC-mapping and interpretable comparisons of G-banded patterns it is impossible to state unequivocally which of the competing hypotheses applies to the changed chromosomal morphology detected among the Rattini although parsimony (least number of breaks) allows for the ready distinction between mechanisms (i)-(iii). Using this criterion pericentric inversions would be considered the most parsimonious solution. BAC mapping on the three inversion polymorphisms detected (RNO 1, 12 and 13 orthologs – see below) within *Rattus* supports this contention since the ordering of the BACs is inverted in each instance. This in turn raises the principle of homologous change, where canalized rearrangements, referred to as karyotypic orthoselection by White (1973), has driven karyotype evolution in Rattini. The mechanisms underpinning neocentromerization on the other hand remain ill-defined (Marshall *et al.,* 2008; Marshall and Choo, 2009; Rocchi *et al.,* 2009; Brown and O'Neill, 2010) and difficult to demonstrate empirically (requiring high resolution BAC-mapping, e.g., at 5 centimorgan intervals, the mapping of the centromere protein CENP-A by chromatin

immunoprecipitation (ChIP)-on-chip-experiments, and/or the use of centromeric α satellite DNA probes; see Amor *et al.,*2004; Lomiento *et al.,*2008; Stanyon *et al.,* 2008). Moreover, there are no reports of neocentromere detection within Rattini suggesting, on the bulk of evidence, that multiple neocentromere seeding is less likely an explanation for the observed differences than pericentric karyotypic orthoselection.

In order to interpret the data with respect to other Murinae species, and to aid polarisation of characters within *Rattus s. l.*, the results were compared to published comparative chromosomal maps available for representative species of Murinae. The representative species were: *A. sylvaticus* (Matsubara *et al.,* 2004), *M. platythrix* (Matsubara *et al.,* 2003), *R. pumilio* (Rambau and Robinson*,* 2003), *N. mattheyi* and *C. pahari* (Veyrunes *et al.,* 2006). As these taxa were all studied using mouse (*M. musculus*) paints, the rat characters identified in the present study (*R. norvegicus*, RNO, chromosomes were used as reference to define the characters) were converted to their respective *M. musculus* (MMU) syntenic associations based on the mouse and rat painting studies of Helou *et al.* (2001), Stanyon *et al.* (1999) and, where applicable, the Ensembl genome browser (http://www.ensembl.org). The latter enabled a reappraisal of the rat and mouse syntenic associations reported in Stanyon *et al.* (1999), i.e., the incorrectly assigned association of RNO 20 with MMU 17/11 rather than MMU 17/10 and permitted the correct identification of the segments of MMU 17 and 10 that are involved in this syntenic association (see Figure 3.4). *Hapalomys delacouri* was selected as the primary outgroup taxon as recent sequence based phylogenies place the species basal to Murinae (Pagès *et al.* upubl.; present study – Chapter 5) making it the most appropriate outgroup taxon for this study. The corresponding schematic representation of the painting experiments implemented in the present investigation, together with the published data, is shown in Figure 3.2.
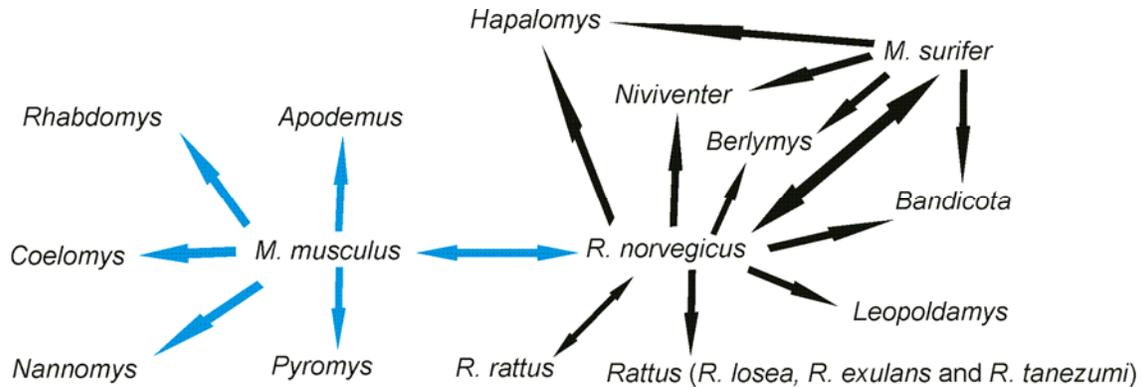
Figure 3.2. A schematic representation of the procedure followed in the construction of the cladistic framework from combined datasets, one from the chromosome painting strategies followed in this study (Rattini), and the second from published data (i.e., permitting the inclusion of additional species of Murinae). Black arrows indicate cross-species hybridization experiments conducted in this study from which the binary characters were identified and subsequently coded with respect to *M. musculus* based on published data using mouse whole chromosome paints (blue arrows). Arrows indicate the target species. ←→ = reciprocal FISH analysis. See text for details.

To ensure a conservative and unambiguous approach, polymorphic chromosome pairs were not included in the cladistic analysis (i.e., the RNO 1, 12 and 13 orthologs investigated in the BAC clone analysis; see below). Moreover, to facilitate comparisons of banding patterns only intrachromosomal rearrangements (pericentric inversions, see above) involving segments that were conserved as whole elements/chromosomes in Rattini were included in the cladistic analysis (except for RNO 15, 16 and 20 that were involved in a complex segmental association occurring only in the outgroup taxon *H. delacouri*). The assignment of character states based solely on banding patterns that involved interchromosomal rearrangements (i.e., conserved as whole element in Rattini but disrupted in additional Murinae taxa) were difficult to determine unambiguously. In addition, no *a priori* polarization (i.e., synteny breakage resulting from fission vs. segmental association resulting from fusion) was assumed, and character states were all strictly inferred through outgroup (*H. delacouri*) comparisons. This allowed the *a posteriori* reconstruction of chromosome evolution in *Rattus s. l.*

57

The most parsimonious phylogenetic tree was obtained by maximum parsimony analyses using PAUP* 4.0b10 (heuristic searches, branch swapping and Tree-Bissection-Reconnection options; Swofford, 2002). The support for each node was examined by bootstrap analysis of 1,000 replicates.

Mapping the chromosome rearrangements onto a consensus molecular tree

As an alternative to the PAUP analysis, chromosomal characters identified by cross-species chromosome painting using *R. norvegicus, R. rattus* and *M. surifer* paints were mapped to a sequence-based consensus tree derived from molecular data that include L1 amplification sites (Verneau *et al.,* 1997, 1998), cytochrome b and IRBP gene sequences (Lecompte *et al.,* 2008), and the most recently published analysis of the cytochrome b, Cytochrome Oxydase I (COI) and IRBP gene sequences (Pagès *et al.,* 2010). The syndecan-4 sequences used were based on the present study (Chapter 5). This approach allowed the determination of the polarity of karyotype evolution among the various rodent families and the identification of synapomorphic chromosomal rearrangements associated with karyotypic evolution of Rattini.

FISH with BACs

Informative BAC clones were selected spanning the regions of interest (in the p arms and flanking the centromere on the q arm of target chromosomes) from the Wellcome Trust Sanger Institute Ensembl contigs (http://www.ensembl.org) and were obtained from the Children's Hospital Oakland-BACPAC Resources, Oakland, California, USA (http://www.bacpac.chori.org). Their selection was based on their spanning potential inversion breakpoint sites identified through G-band comparisons. These occur on orthologs that correspond to *R. norvegicus* chromosomes 1, 12 and 13 as identified by Yosida and colleagues (Yosida *et al.,* 1971, 1974; Yosida, 1973, 1977a; Yosida and Sagai, 1973, 1975;

see Table 3.2 and Figure 3.3) and present as both acrocentric and submetacentrics (see

Chapter 2; Badenhorst *et al.,* 2009 for a review) within *Rattus* species. The *R. tanezumi*

specimen analysed in the present study was polymorphic (i.e., heterozygous) for the

acrocentric and submetacentric morphs of the *R. norvegicus* orthologs RNO 1, 12 and 13.

This allowed the rapid comparison of the two chromosomal morphs and whether the

orientation of the BAC clones was modified (i.e., due to an inversion) or retained (i.e., the

chromosomal variation is the result of heterochromatic variation or centromeric shifts).

BAC clones were received as bacterial LB agar stab cultures and were handled

according to the supplier's instructions (http://www.bacpac.chori.org/vectorsdet.htm). The

bacteria were streaked out to single colonies on a LB agar plate containing an antibiotic (12.5

µg/ml chloramphenicol) by dipping a sterilised inoculation wire loop in the bacteria growing

within the punctured area of the stab culture, and running this wire loop lightly over the

surface of the prepared LB agar plate. The colonies were allowed to grow overnight at 37°C

following which, single colonies were picked for isolation. DNA (~1 µg) was extracted from

each BAC clone using Wizard Plus SV Miniprep DNA purification system (Promega) and

labelled by standard nick translation with either biotin- or digoxigen-dUTP (Roche). In

addition, 15% glycerol stocks (250 µl 60% glycerol stock + 750 µl overnight culture) were

prepared for long-term storage of the clones at -80°C.

Table 3.2. BAC clones used in the study with their positions on the rat ideogram, clone
names, and relevant accession numbers. The map positions were confirmed by both NCBI
and UCSC genome browsers.

	No.	Chromosome position	Clone ID	Accession #	End accession #	
					SP6	T7
RNO 1	1	1p13	CH230-347K17	AC135531	BZ154483	BZ154484
	2	1p11	CH230-99D20	AC109433	BH314535	BH314564
	3	1q11	CH230-365B17	/	BZ116707	BZ116708
	4	1q12	CH230-447J17	/	BZ248996	BZ248997
RNO 12	5	12p12	CH230-30M1	/	BH293553	BH293555
	6	12p11	CH230-70E3	/	BH273680	BH273682

	7	12q11	CH230-IJ13	AC095368	BH297088	BH297089
	8	12q12	CH230-204G18	AC123562	BH366346	BH366347
RNO 13	9	13p13	CH230-52N12	AC103093	BH318049	BH318051
	10	13p11	CH230-231G7	/	BZ113582	BZ113584
	11	13q11	CH230-221F10	/	BH353297	BZ094368
	12	13q12	CH230-128D5	AC106070	BH279605	BH279609

/ These BAC clones have not yet been fully sequenced, however, all BAC clones have had both ends sequenced (SP6 and T7)
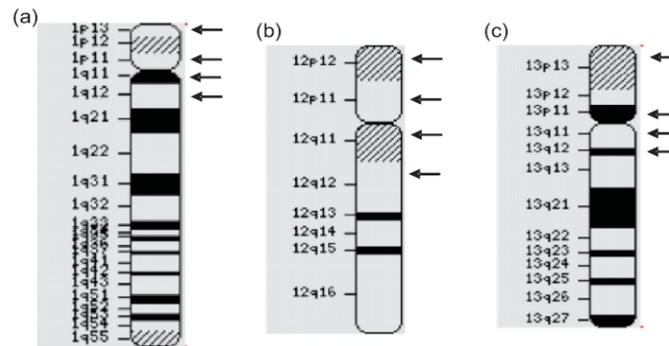
Figure 3.3. Chromosome ideograms of polymorphic *R. norvegicus* chromosomes **(a)** 1, **(b)** 12, and **(c)** 13 with arrow heads pointing to designated regions in which BAC clones were selected for BAC analysis (ideograms redrawn from Hamta *et al.,* 2006).

Dual-colour FISH reactions were set up to determine orientation of BAC clones in the different target regions. Labelled probes were hybridized to chromosome preparations, cover-slipped and sealed with rubber cement. Sealed chromosome preparations with probe were denatured together on a hot plate at 65°C for 3 min. Hybridization took place in a humid chamber at 37°C overnight. Post-hybridization washes consisted of a first wash in 0.4x SSC/0.3% Tween 20 for 5 min at 60°C, followed by second wash in 2x SSC/0.1% Tween 20 for 1 min at room temperature. The remainder of the detection protocol was carried out as previously described (see above - chromosome painting). Images were captured with a CCD camera coupled to an Olympus BX60 fluorescence microscope and analysed using Genus 3.7 Software (Applied Imaging).

RESULTS

Chromosome painting using flow-sorted chromosomes of *R. norvegicus* allowed the identification of orthologous chromosomes and chromosomal segments among the different species analysed. In addition *M. surifer* and *R. rattus* paints were successfully used to confirm the homologies identified by the *R. norvegicus* paints, and to precisely define the breakpoint boundaries of identified rearrangements. A refinement of rat and mouse syntenic associations and the positions of breakpoint junctions identified in this study are presented in Figure 3.4. Diploid numbers varied from 40 – 52 (Chapter 2) with four species (*R. exulans, R. tanezumi, R. losea* and *L. edwardsi*) characterised by an identical 2n = 42 mirroring the situation in *R. norvegicus.*
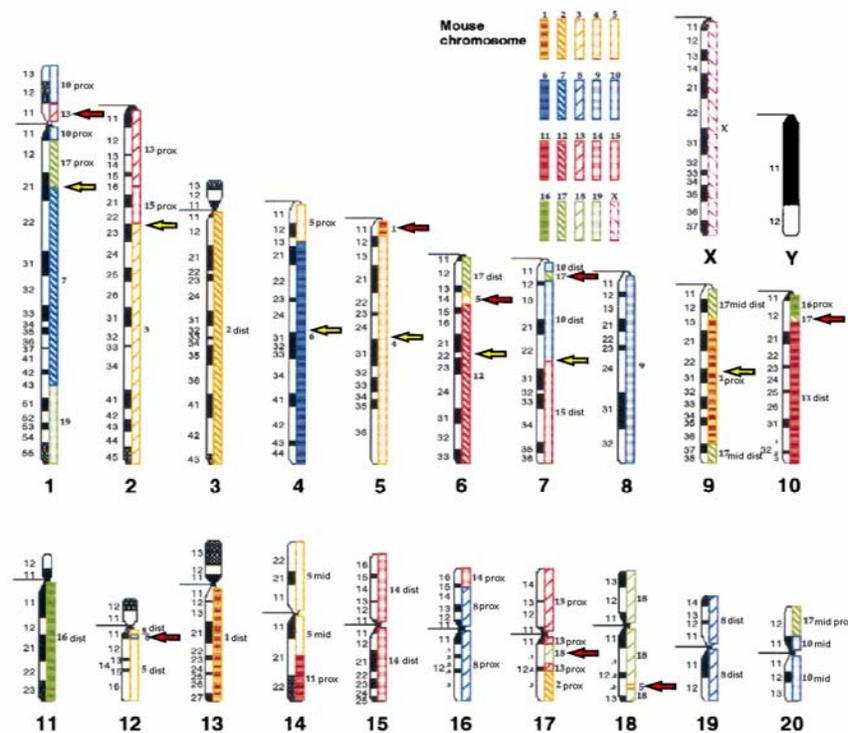
Figure 3.4. Schematic representation of hybridization results using mouse chromosome paints on rat chromosomes (shown against rat ideogram, redrawn from Helou *et al.,* 2001). The rat chromosomes are numbered below and homology to mouse chromosomes is indicated on the right. "dist", "mid", "prox" refer respectively to the distal, middle and proximal segments of the *M. musculus* chromosomes. Red arrows indicate small conserved *M. musculus* segments/bands in the rat genome that were not previously reported by Stanyon *et*

61

al. (1999; the only available reciprocal painting study between rat and mouse). Yellow areas denote the breakpoint junctions of the interchromosomal rearrangements identified in the present study.

Characterisation of *Rattus rattus* (RRA) and *Maxomys surifer* (MSU) flow karyotypes

Figure 3.1 shows the flow-karyotypes of *M. surifer* (male) and *R. rattus* (female) with the assignments of the flow-sorts presented in Table 3.3. The 38 chromosomes of *R. rattus* were resolved into 16 peaks (Table 3.3). Fourteen peaks contained a single chromosome pair (RRA 1-9, 14, 9-11, 16-18, X) and two contained more than one chromosome; one peak contained a mix of two different chromosomes (RRA 10+12), whereas the other comprised three different chromosomes (RRA 11, 13+15).

The 52 chromosomes of *M. surifer* were resolved into 20 peaks. Twelve contained a single chromosome pair (MSU 1, 2, 4, 7, 14, 17, 19-21, 24, Y), with MSU 4 identified in two separate peaks (peaks D and T; refer to Table 3.3). This most likely reflects differing amounts of heterochromatin between the two homologs although this was not readily apparent on C-band analysis (data not shown). Seven peaks contained a mix of two different chromosomes (MSU X+4, 5+6, 8+9, 11+13, 10+15, 22+25, 3+12), and one peak contained a mix of three chromosomes (MSU 16+18+23). The assignment of one of the seven peaks containing two different chromosomes by FISH (peak N, refer to Table 3.3) was unsuccessful and it is assumed that MSU 3 and 12, which were not present in any of the other flow-sorts, were contained in this peak.

As detailed by Stanyon *et al.* (1999), the *R. norvegicus* (2n = 42) flow-karyotype comprised 21 individual peaks (Table 3.3) of which two contained more than one chromosome. One comprised RNO 11+15, and the other RNO 13+14+15. Importantly, RNO 15 is common to both peaks and consequently 2-colour FISH was useful for differentiating between non-pure sorts involving RNO 15 from RNO 11, 13 and 14. Thus 19

chromosomes of the *R. norvegicus* genome are fully resolved with the identification of only

two chromosomes being ambiguous (RNO 13 and 14; see Stanyon *et al.,* 1999). The

identification of RNO 13 and 14 could not be unambiguously clarified using *M. surifer* and

R. rattus chromosome paints, nor with dual-colour FISH, since their orthologs (RNO 13:

MSU 11 and RRA 10, and RNO 14: MSU 8 and RRA 11; see Table 3.3) are contained in

non-pure flow-sorts. Fortunately, however, both chromosomes are easily distinguished on G-

band patterns.

Table 3.3. Flow-sorted peak assignments (A-V) for *M. surifer* (MSU) and *R. rattus* (RRA).
Those for *R. norvegicus* (RNO) are taken from Stanyon *et al.* (1999).

	Peak																					
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Corresponding MSU chromosomes	1	2	X, 4	4	Y	24	14	17	19	6, 5	22, 25	7	8, 9	3, 12	10, 15	11, 13	16, 18, 29	21	20	4	/	/
Corresponding RRA chromosomes	1	3	2	4	X	6	5	7	8	9	10, 12	11, 13, 15	14	17	16	18	/	/	/	/	/	/
Corresponding RNO chromosomes	1	2	10	*	X	20	12	19	Y	6	18	17	16	13, 14, 15	11, 15	7	9	8	4	5	3	/

Grey shading indicates peaks for which 2-colour FISH was useful for differentiating non-pure
sorts involving MSU 4 and X, and RNO 11 and 15. MSU chromosome numbers correspond
to the published G-band karyotype (Badenhorst *et al.,* 2009-Chapter 2), and RRA was
arranged according to Cavagna *et al.* (2002).
* Result of a spontaneous translocation between RN0 4 and 8 in the cell line used for the
characterisation of the *R. norvegicus* flow-karyotype reported by Stanyon *et al.* (1999).

Reciprocal chromosome painting between *Rattus rattus* (RRA), *Maxomys surifer* (MSU)

and *R. norvegicus* (RNO)

Bi-directional FISH between *R. rattus* and *R. norvegicus* confirmed the findings of

Cavagna *et al.* (2002) and showed the largely conserved nature of the genomes of these two

species. As expected, the only exceptions to this are the fused chromosomes RRA 1 and 4

(corresponding to Robertsonian fusion (Rb) 5.7 and Rb 9.11 of the unfused *R. norvegicus*

orthologs; Figure 3.5a). As the genomes of these two species are largely invariant, *R. rattus*

was not considered useful for reciprocal analysis to delimit the breakpoint junctions identified

within the *Rattus s. l.* complex using *R. norvegicus* chromosome paints.

The results of the cross-species chromosome painting of *R. norvegicus* chromosome

paints onto *M. surifer* chromosomes showed that sixteen *R. norvegicus* chromosomes (RNO

3, 7-20 and X) are conserved *in toto* in the *M. surifer* karyotype (see MSU 2, 5-9, 11, 14-16,

18, 22-25 and X in Figure 3.5). Five *R. norvegicus* chromosomes produced two signals,

respectively, on *M. surifer* chromosomes (corresponding to: RNO 1 = MSU 21 and 1, RNO 2

= MSU 3 and 4, RNO 4 = MSU 10 and 13, RNO 5 = MSU 12 and 17, and RNO 6 = MSU 19

and 20). Comparisons of the G-banding patterns show that there are a number of

intrachromosomal differences, attributed to pericentric inversions (see above), between the

two karyotypes that have altered the morphology of MSU 4, 5, 8-10, 12, 14, 16, 18, 22, 23

and 25 in comparison to their *R. norvegicus* orthologs.

The results of the reciprocal analysis of *M. surifer* whole chromosome paints to *R.

norvegicus* chromosomes are shown in Figure 3.5a. Reciprocal chromosome painting

between *M. surifer* and *R. norvegicus* confirmed the unidirectional assignments based on

painting *R. norvegicus* probes to *M. surifer* (see Figure 3.5), and permitted the more precise

delimitation of sub-chromosomal homologies in *M. surifer* (i.e., RNO1p+1q prox

homologous to MSU 21 and RNO1q dist homologous to MSU 1). Importantly, and contrary

to the outcome based on *R. rattus* paints, the *M. surifer* chromosome paints proved useful for

delimiting breakpoints identified by the *R. norvegicus* Zoo-FISH analysis of Rattini. This

applies specifically to the segmental associations in *B. berdmorei, B. bowersi, B. savilei* and

H. delacouri (i.e., syntenic association of RNO 9 and 11) and the disruptions of *R. norvegicus*

chromosomes in *N. fulvescens, B. savilei* and *H. delacouri* (i.e., RNO 1, 7 and 9) (see below).

The only exception was the disruption of RNO 2. Although G-banding provides convincing evidence that MSU 3 and 4 are homologous to RNO 2, (Figure 3.5) this could not be confirmed by FISH as MSU 3 is presumed to be present in peak N which failed to hybridize when conducting the flow-sort assignments (see above). Thus, cross-hybridization to delimit breakpoint junctions in RNO 2 was only possible using the *M. surifer* probe corresponding to MSU 4 (MSU 4 = peak C, D and T; Table 3.3.) but it seems reasonable to assume that the breakpoints are identical in all three genera.
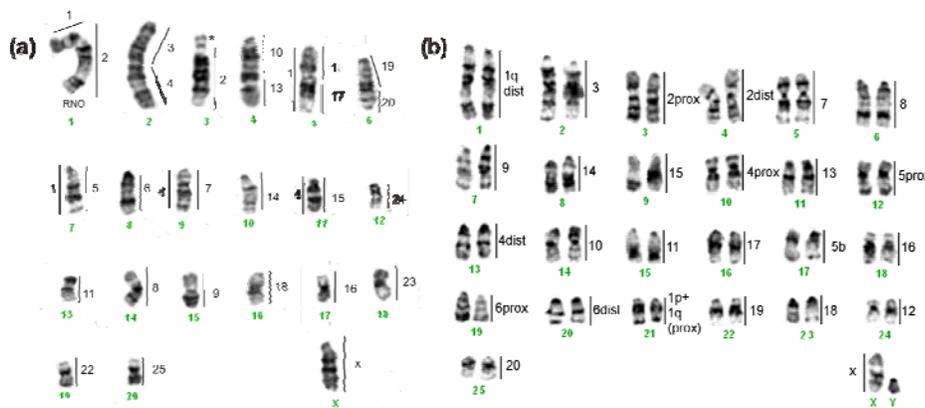
Figure 3.5. **(a)** Regions of orthology between *R. norvegicus* (2n = 42), *M. surifer* (2n = 52) and *R. rattus* (2n = 38) chromosomes based on reciprocal painting between the two species and mapped to the *R. norvegicus* G-banded half karyotype (2n = 42). Segments orthologous to *M. surifer* are shown on the right of each chromosome and to *R. rattus* Rb fusion chromosomes (RRA 1 and 4) on the left. **(b)** G-banded karyotype of male *M. surifer* (2n = 52) with regions of orthology to *R. norvegicus* (numbered on the right, except for X) as determined by cross-species chromosome painting. * Indicates blocks that were not hybridized by any of the chromosome paints and which correspond to heterochromatic regions identified through C-banding (refer to Chapter 2).

Cross-species chromosome painting

The half-karyotype comparisons of G-banded chromosomes of the ten species under investigation in this study compared to that of *R. norvegicus* is presented in Figure 3.6. This provides an effective summary of the Zoo-FISH results. Details of each Zoo-FISH experiment are provided below. Examples of cross-species chromosome painting among the different species using *R. norvegicus* chromosomes are presented in Figure 3.7.

65
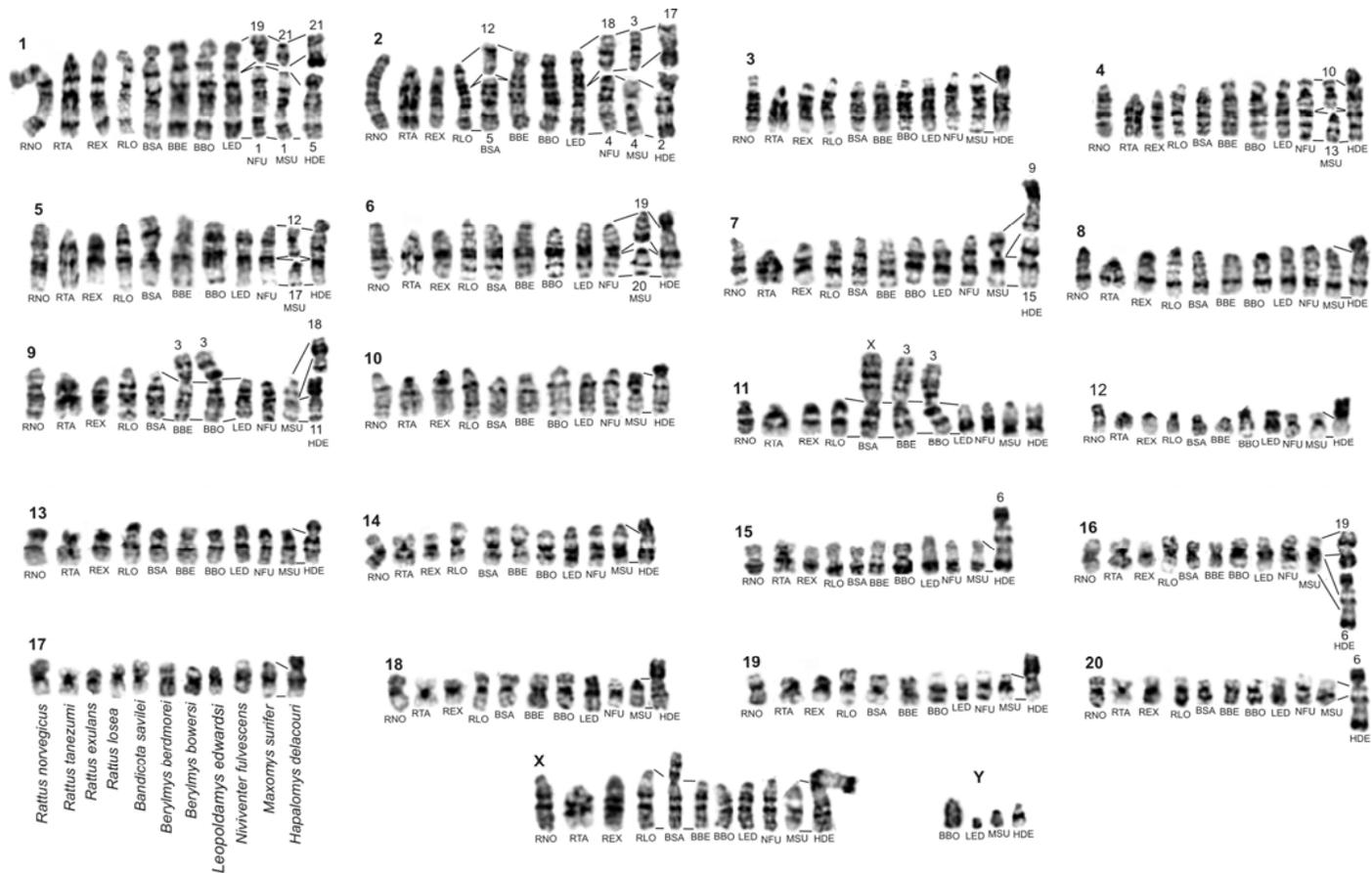
Figure 3.6. G-banded half-karyotype comparison between *R. norvegicus* and the ten species analysed in this study showing genome-wide chromosomal correspondence defined by painting (using *R. norvegicus, R. rattus* and *M. surifer* chromosome paints) and banding homologies. Karyotypes were arranged according to *R. norvegicus* standard karyotype (Committee for a standardized karyotype of *R. norvegicus*, 1973).
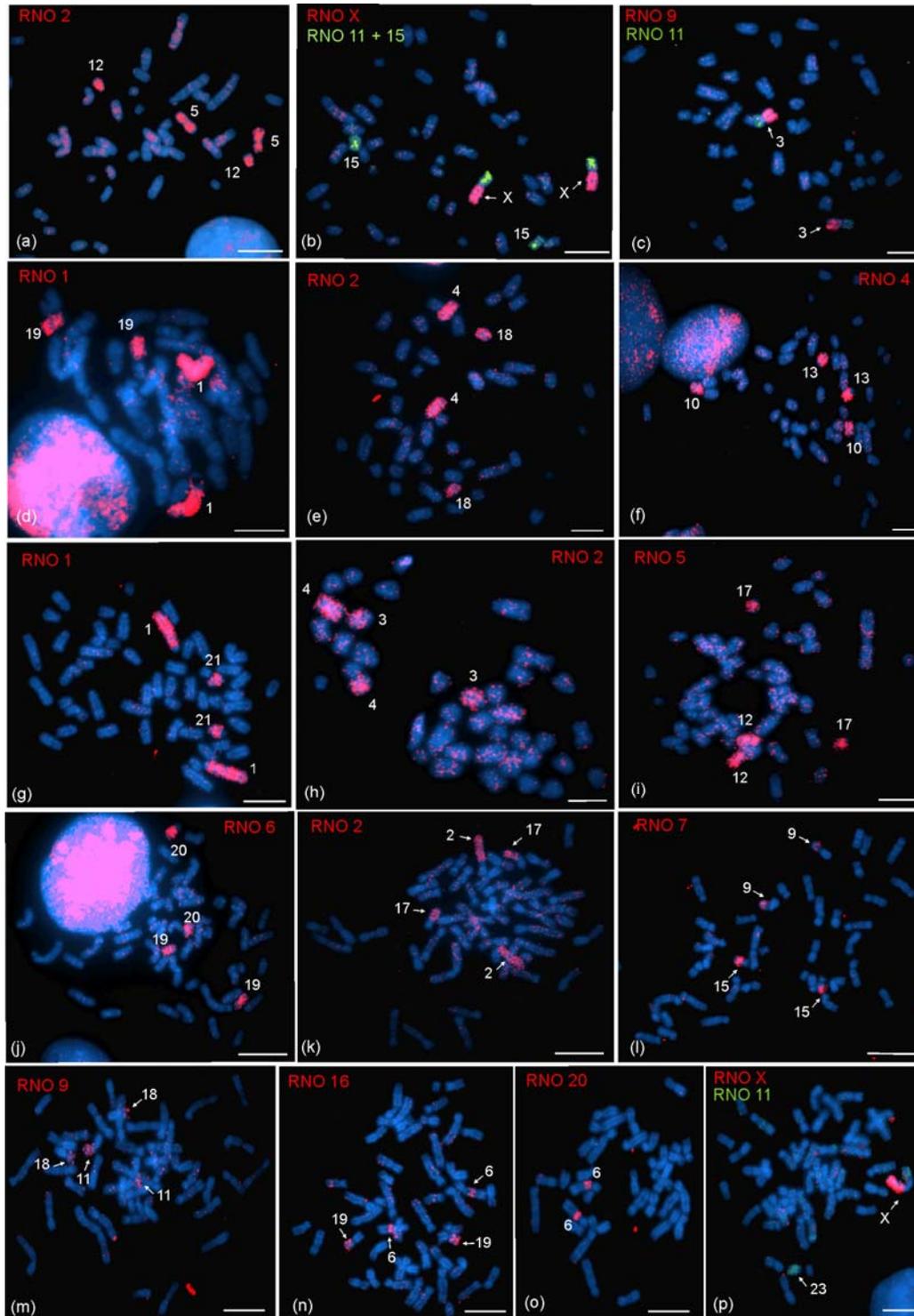
Figure 3.7. Examples of FISH and double-colour FISH experiments using various *R. norvegicus* (RNO) chromosome specific painting probes on species of Asian rodents analysed in the present investigation. Chromosome numbers of the target species are indicated in white and the RNO probes used in red (Cy3-labelled) or green (FITC-labelled). Panels **(a)** and **(b)** present FISH of RNO 2 and RNO X and 11 on metaphase chromosomes of *B. savilei* showing the disruption of RNO 2 as well as X-autosome translocation (RNO X and

11). Panel **(c)** shows FISH of RNO 9 and 11 on metaphase chromosomes of *B. berdmorei* showing their syntenic associations in BBE 3. Panels **(d)** and **(e)** present the hybridization of RNO 1 and 2, respectively on metaphase chromosomes of *N. fulvescens* illustrating the disruption of RNO 1 and 2. Metaphase chromosomes of *M. surifer* hybridized with **(f)** RNO 4, **(g)** RNO 1 **(h)** RNO 2, **(i)** RNO 5, and **(j)** RNO 6 chromosome paints, showing the disruption of RNO 4, 1, 2, 5 and 6 respectively. Hybridization of **(k)** RNO 2, **(l)** RNO 7, **(m)** RNO 9, **(n)** RNO 16, **(o)** RNO 20 and **(p)** RNO X and 11 on metaphase chromosomes of *H. delacouri*. Panels (k), (l) and (m) show the disruption of RNO 2, 7 and 9 chromosome pairs, respectively, while panel (n) demonstrates the disruption of RNO 16 with the syntenic association of a segment of RNO 16 within HDE 6. Panel (o) illustrates the conservation of RNO 20 in HDE 6, while panel (p) shows the same X-autosome translocation as detected in *B. savilei* (refer to panel b). Note the lack of hybridization signal on the heterochromatic arms of *H. delacouri* submetacentric autosomes (i.e., panels k, l, m and n). Scale bar = 10 µm.

(i) Hybridization of Rattus norvegicus *(RNO) whole chromosome painting probes to chromosomes of* R. exulans *(REX),* R. tanezumi *(RTA),* R. losea *(RLO) and* Leopoldamys edwardsi *(LED).*

The karyotypes of the *Rattus s. s.* representatives (*R. losea*, *R. exulans* and *R. tanezumi*) and *L. edwardsi* are strongly conserved based on both the Zoo-FISH and G-band data (Figure 3.8). Consequently *M. surifer* chromosome paints were not hybridized to the metaphase spreads of these species. All have 2n = 42 chromosomes, and there is no detectable intrachromosomal repatterning evident in *R. exulans, R. losea* and *R. tanezumi* when compared to *R. norvegicus* (Figure 3.8a, note that polymorphic rearrangements involving RNO 1, 12 and 13 orthologs are not considered when comparing G-band patterns; see BAC-mapping below). Comparison of G-banding patterns show that there are a number of intrachromosomal differences that have altered the morphology of LED 1, 3, 9, 12, 14, 15, 16 and 17 in comparison to their *R. norvegicus* orthologs. These intrachromosomal rearrangements are considered to be the result of pericentric inversions (see above).
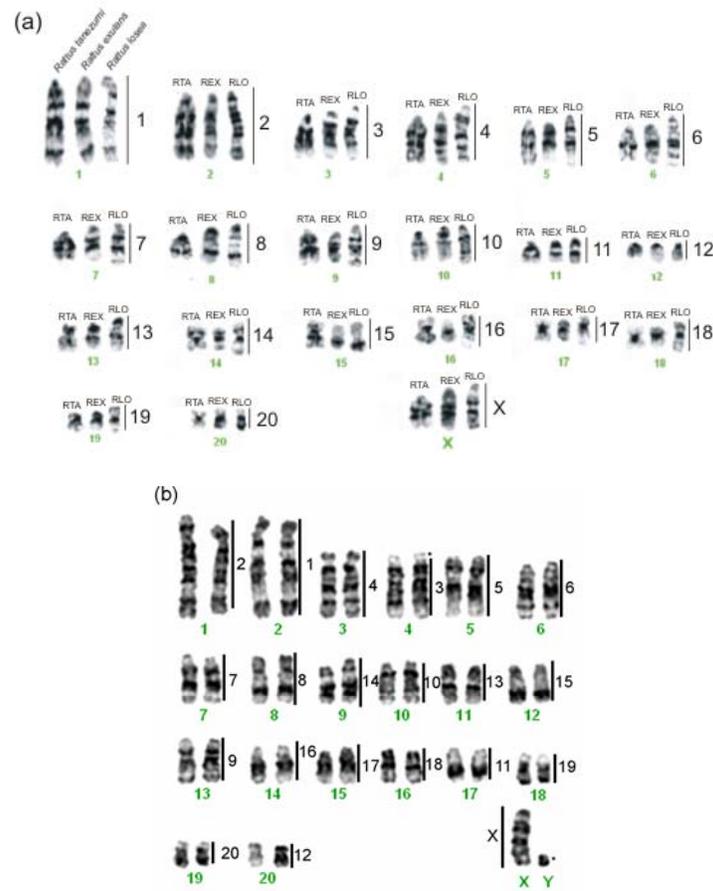
Figure 3.8. G-banded karyotypes of **(a)** *Rattus s. s*. (2n = 42), arranged according to *R. norvegicus* standard karyotype (Committee for a standardized karyotype of *Rattus norvegicus*, 1973) and **(b)** male *L. edwardsi* (2n = 42) with the assignments of regions of homology to *R. norvegicus* (numbered on the right, except for the X in *L. edwardsi*) as determined by cross-species chromosome painting. Note that *Rattus s. s*. comprises a comparison of G-banded half karyotypes of *R. tanezumi, R. exulans* and *R. losea*. * Indicates blocks that were not hybridized by any of the chromosome paints and which correspond to heterochromatic regions identified by C-banding analysis (Chapter 2).

(ii) Hybridization of Rattus norvegicus *(RNO) and* Maxomys surifer *(MSU) whole chromosome painting probes to chromosomes of* Bandicota savilei *(BSA).*

Twenty *R. norvegicus* (RNO 1, 3-20 and X) chromosomes are conserved *in toto* in the *B. savilei* genome (BSA 1-4, 6-11, 13-20 and X; Figure 3.9a). Two interchromosomal rearrangements and the presence of a B chromosome differentiate *B. savilei* from *R. norvegicus*—a disruption of RNO 2 corresponding to BSA 5 and 12, and a sex-autosome

69

association involving RNO X and 11 that corresponds to the large metacentric BSA X chromosome (i.e., there is a XY1Y2 sex chromosome system in the species). The XY1Y2 system was confirmed by the analysis of a *B. savilei* male specimen (Figure 3.9b) where Y1 represents the original Y and Y2 represents the unfused autosome (homolog of RNO 11); the other autosomal RNO 11 homolog is fused with the original X. Furthermore, the sex-autosome association was confirmed through the hybridization of *M. surifer* paints, MSU X and 15 (orthologs of RNO X and 11) respectively. The disruption of RNO 2 could only be confirmed using MSU 4 (as previously addressed; see above), which hybridized to BSA 5, corresponding to the distal segment of RNO 2. As expected, no hybridization was detected on the B chromosome (Chapter 2). The G-band patterns between *R. norvegicus* and *B. savilei* showed good correspondence indicating an absence of detectable intrachromosomal rearrangements. The only exceptions to this included BSA 3, 5 and 9 whose morphology differs, respectively, from RNO 5, 2 and 9, as a result of pericentric inversions (see above).
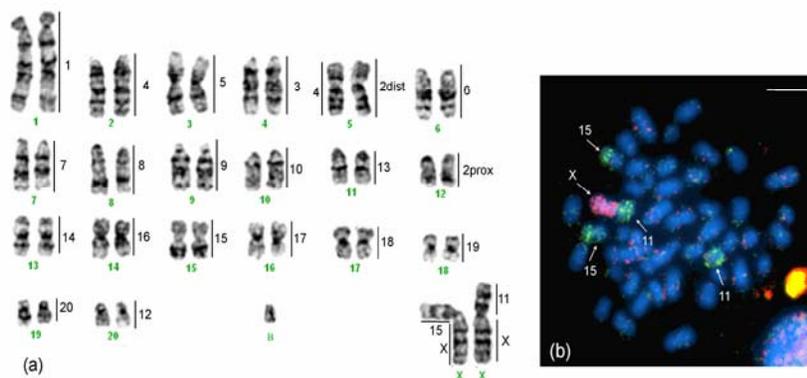
Figure 3.9. G-banded karyotype of **(a)** female *B. savilei* (2n = 42 + 1B chromosome), with the assignments of regions of orthology to *R. norvegicus* (numbered on the right) and *M. surifer* where applicable (numbered on the left) as determined by cross-species chromosome painting. **(b)** FISH hybridization of RNO X, 11 and 15 (both RNO 11 and 15 are in the same probe) on metaphase chromosomes of male *B. savilei* specimen showing the X-autosome fusion of RNO X and 11 and the unfused homolog of RNO 11 (Y2). Scale bar = 10 μm.

(iii) Hybridization of Rattus norvegicus *(RNO) and* Maxomys surifer *(MSU) whole chromosome painting probes to chromosomes of* Berylmys berdmorei *(BBE) and* Berylmys bowersi *(BBO).*

All *R. norvegicus* chromosomes show almost complete conservation in both *B. berdmorei* (2n = 40 + 1B) and *B. bowersi* (2n = 40). The only difference between *R. norvegicus* and these two species is the presence of a single syntenic association (corresponding to RNO 9 and 11) in each (BBE 3 and BBO 3) (Figure 3.10, panels a and b). The RNO 9+11 syntenic association was confirmed using *M. surifer* paints MSU 15 and MSU 7 (corresponding to RNO 9 and 11, respectively). The *B. berdmorei* specimen analysed here was also noteworthy for the presence of a single B chromosome (Chapter 2). Reasonable banding homology is observed between both *Berylmys* species (*B. berdmorei* and *B. bowersi*) and *R. norvegicus* suggesting the absence of intrachromosomal rearrangements at the level of resolution obtained in this study. The exceptions to this are differences in the morphology of chromosomes BBE and BBO 2, 5 and 6 in comparison with RNO 2, 5 and 4 that are due to pericentric inversions (see above). Lastly, one homolog of BBE 5 differs from the other through heterochromatic amplification of the terminal end of the q arm (i.e., there is a heteromorphism in this specimen). The heterochromatic nature of this difference was confirmed by the absence of hybridization to this region using *R. norvegicus* painting probes (Figure 3.10, panel c), as well as through C-band analysis (C-positive material at terminal end of BBE 5; Chapter 2).
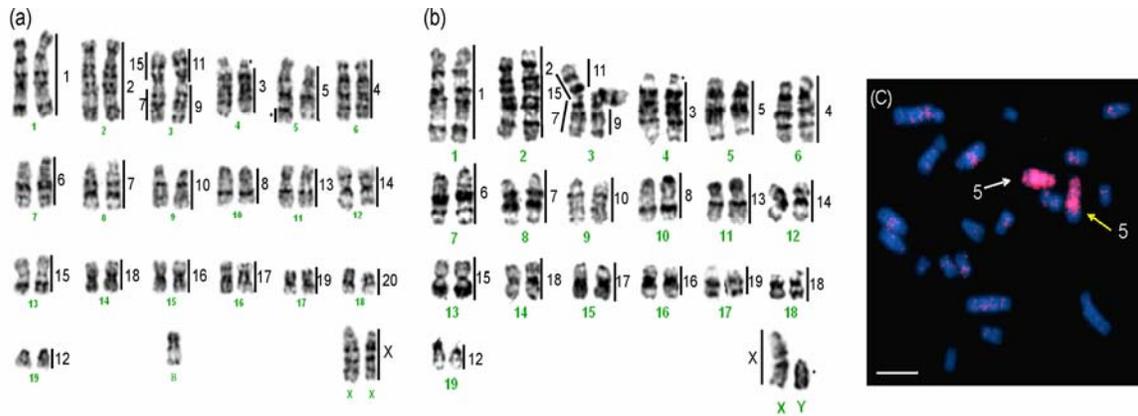
Figure 3.10. G-banded karyotype of **(a)** female *B. berdmorei* (2n = 40 + 1B), and **(b)** male *B. bowersi* (2n = 40) with regions of orthology to *R. norvegicus* (shown to the right, except for the X in *B. bowersi*) and *M. surifer* where applicable (to the left) as determined by cross-species chromosome painting. * Indicate blocks that were not hybridized by any of the chromosome paints and correspond to heterochromatic regions identified by C-banding (refer to Chapter 2). **(c)** Hybridization of RNO 5 on metaphase chromosomes of *B. berdmorei* depicting lack of hybridization at terminal end of the q arm of the RNO 5 homolog (yellow arrow) in *B. berdmorei*. Scale bar = 10 μm.

(iv) Hybridization of Rattus norvegicus *(RNO) and* Maxomys surifer *(MSU) whole*

chromosome painting probes to chromosomes of Niviventer fulvescens *(NFU).*

Nineteen *R. norvegicus* (RNO 3-20 and X) chromosomes are conserved *in toto* in

the *N. fulvescens* genome (see NFU 2, 3, 5-17, 20-22 and X in Figure 3.11). Two *R.*

norvegicus chromosome paints (RNO 1 and 2) each produced two signals (i.e., on different

chromosomes) in the *N. fulvescens* karyotype: RNO 1 = NFU 1 and 19, RNO 2 = NFU 4 and

18 (see Figure 3.11). Good banding homology is observed between *N. fulvescens* and *R.*

norvegicus, and this showed that there are several intrachromosomal differences between the

two karyotypes, recognized as pericentric inversions (see above) that have altered the

morphology of NFU 2, and 12-17 in comparison to their *R. norvegicus* orthologs (Figure

3.11). In addition, the use of *M. surifer* painting probes confirmed the disruption of RNO 1

where MSU 1 and 2 hybridized respectively to NFU 19 and 1 (corresponding to RNO 1p+1 q

proximal and RNO 1q distal). The disruption of RNO 2 could only be confirmed using MSU 4 which hybridized to NFU 4 (corresponding to distal RNO 2, see above).
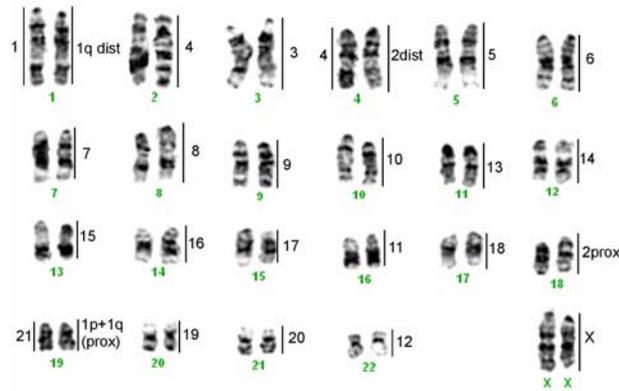
Figure 3.11. G-banded karyotype of female *N. fulvescens* (2n = 46) with regions of orthology to *R. norvegicus* (illustrated to the right) and *M. surifer* where applicable (illustrated to the left) as determined by cross-species chromosome painting.

(v) Hybridization of Rattus norvegicus *(RNO) and* Maxomys surifer *(MSU) whole chromosome painting probes to chromosomes of* Hapalomys delacouri *(HDE).*

Sixteen *R. norvegicus* chromosomes (RNO 3-6, 8, 10-15, 17-20 and X) were retained in their entirety either as a conserved block, or as a single chromosome in the *H. delacouri* genome (see HDE 1, 3, 4, 6-8, 10, 12-14, 16, 20, 22, 23 and X in Figure 3.12a). Five *R. norvegicus* chromosomes produced two signals in the *H. delacouri* karyotype (RNO 1 = HDE 5 and 21, RNO 2 = HDE 2 and 17, RNO 7 = HDE 9 and 15, RNO 9 = HDE 11 and 18, and RNO 16 = HDE 6 (one segment only) and 9). The disruptions of *R. norvegicus* orthologs in the *H. delacouri* karyotype were confirmed using *M. surifer* chromosome paints (for example MSU 4, the homolog of distal RNO 2, corresponds to HDE 2; Figure 3.12a). A complex autosomal segmental association was identified in HDE 6 comprising a segmental association of RNO 20, 15, and a segment of 16 (Figure 3.12, panel b). Moreover, as with *B. savilei*, an X-autosome association was detected that involves RNO X and 11. This sex-

73

autosome translocation was confirmed through the hybridization of MSU X and 15 (homologs of RNO X and 11, respectively). As to be expected, no hybridization signal was detected on the C-negative B chromosome identified in the *H. delacouri* karyotype, nor on the C-positive heterochromatic arms of each of the autosomes (Chapter 2).

Good correspondence was noted in the banding patterns of the syntenic segments that are conserved between *R. norvegicus* and *H. delacouri* (Figure 3.12). The only exceptions were some intrachromosomal rearrangements, namely pericentric inversions (see above) that differentiate HDE 6, 10, 13, 16, 20, and 23 and their *R. norvegicus* orthologs (i.e. RNO 20 and 15).
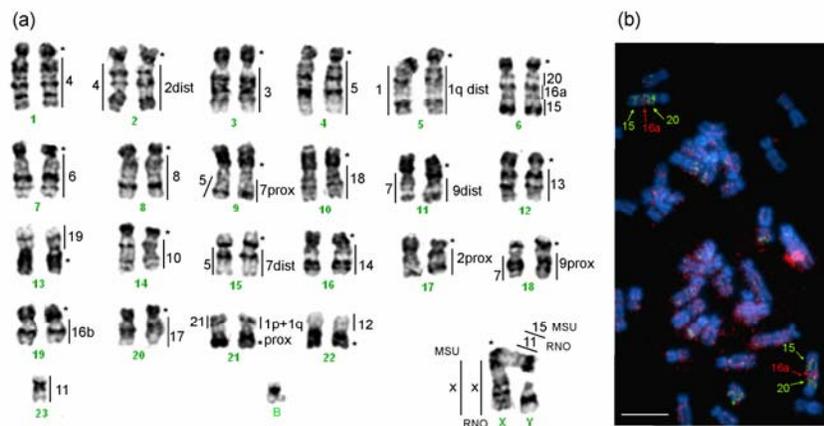
Figure 3.12. G-banded karyotype of **(a)** male *H. delacouri* (2n = 48) with regions of orthology to *R. norvegicus* (numbered to the right) and *M. surifer* where applicable (numbered to the left) as determined by cross-species chromosome painting. * Indicates blocks that were not hybridized by any of the chromosome paints and which correspond to heterochromatic regions identified by C-banding (Chapter 2). **(b)** Double-colour FISH hybridization to *H. delacouri* metaphase chromosomes using *R. norvegicus* chromosome paints RNO 20 and 15 were detected with Cy3 (in red), and RNO 16 was detected with FITC (in green) showing the segmental association involved in the formation of HDE 6. Note that the disruption of RNO 16 is considered an ambiguous character as it is not possible to precisely discern the position of the breakpoint (reciprocal analysis was not useful in this regard). Consequently the two segments corresponding to the disrupted RNO 16 are denoted as "a" and "b" respectively. Scale bar = 10 μm.

The chromosomal orthologies identified between the nine species of the *Rattus s. l.* complex and *H. delacouri* investigated in this study using *R. norvegicus, R. rattus* and *M. surifer* painting probes are summarised in Table 3.4.

Table 3.4. Conserved chromosomal regions detected in Rattini species and *H. delacouri* (HDE) using *R. norvegicus* (RNO) painting probes. Intrachromosomal rearrangements identified following comparisons with their respective *R. norvegicus* orthologs are denoted by +.

R. *norvegicus*	*R.* *losea*	*R.* *tanezumi*	*R.* *exulans*	*B.* *bowersi*	*B.* *berdmorei*	*B.* *savilei*	*L.* *edwardsi*	*N.* *fulvescens*	*M.* *surifer*	*H.* *delacouri*
1	1	1	1	1	1	1	2	19, 1	21, 1	21, 5
2	2	2	2	2+	2+	5+, 12	1+	18, 4	3, 4+	17, 2
3	3	3	3	4	4	4	4	3	2	3
4	4	4	4	6+	6+	2	3+	2+	10+, 13	1
5	5	5	5	5+	5+	3+	5	5	12+. 17	4
6	6	6	6	7	7	6	6	6	19, 20	7
7	7	7	7	8	8	7	7	7	5+	9, 15
8	8	8	8	10	10	8	8	8	6	8
9	9	9	9	3	3	9+	13	9	7	18, 11
10	10	10	10	9	9	10	10	10	14+	14
11	11	11	11	3	3	X	17+	16+	15+	23+, X
12	12	12	12	19	19	20	20	22	12	22
13	13	13	13	11	11	11	11	11	11	12
14	14	14	14	12	12	13	9+	12+	14+	16+
15	15	15	15	13	13	15	12+	13+	15+	6+
16	16	16	16	16	15	14	14+	14+	16+	19, 6
17	17	17	17	15	16	16	15+	15+	17+	20+
18	18	18	18	14	14	17	16+	17+	18+	10+
19	19	19	19	17	17	18	18	20	22+	13+
20	20	20	20	18	18	19	19	21	20+	6+
X	XX	XX	XX	X	X	X	X	XX	X	X
Y				Y			Y		Y	Y

Phylogenetic analysis

Phylogenetic analysis based on chromosome rearrangements

A total of 23 binary characters were included in the data matrix (Table 3.5). Among these, 13 correspond to segmental associations/disruptions of *R. norvegicus* segments that most probably arose as a result of either fusion/fission events, while the remaining 10 were interpreted as inversions (denoted as "inv" in Table 3.5).

Table 3.5. Matrix of chromosomal characters corresponding to presence (1) or absence (0) of segmental associations of *R. norvegicus* (RNO) chromosomal fragments and inversions (inv). In addition the rat characters were converted to orthologous *M. musculus* (MMU) chromosomes to infer presence/absence of segmental associations and inversions in additional Murinae taxa from published cross-species painting studies that utilised *M. musculus* painting probes. These included *R. pumilio* (RPU, Rambau and Robinson, 2003), *N. mattheyi* (NMA, Veyrunes *et al.,* 2006), *M. musculus* (MMU, Stanyon *et al.,* 1999; Helou *et al.,* 2001), *C. pahari* (CPA, Veyrunes *et al.,* 2006), *M. platythrix* (MPL, Matsubara *et al.,* 2003) and *A. sylvaticus* (ASY, Matsubara *et al.,* 2004). Remaining abbreviations represent *H. delacouri* (HDE) and representatives of the *Rattus s. l.* complex: *R. rattus* (RRA), *R. tanezumi* (RTA)*, R. exulans* (REX), *R. losea* (RLO)*, M. surifer* (MSU), *N. fulvescens* (NFU), *L. edwardsi* (LED), *B. savilei* (BSA), *B. berdmorei* (BBE), and *B. bowersi* (BBO).

Characters based on RNO chromosomes	MMU association	Chromosome segments flanking breakpoints/characters		RNO	RRA	REX	RTA	RLO	BSA	BBO	BBE	LED	NFU	MSU	HDE	RPU	NMA	MMU	CPA	MPL	ASY
		RNO	MMU																		
1 (1p/1q prox)+1q dist	(10/13/10/17)+(7/19)	1q prox+1q dist	17 prox+7	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
2 2 prox+2 dist	(13/15)+3	2 prox+2 dist	15 prox+3	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0
3 4 prox+4 dist	(5/6 prox)+6 dist	4 prox+4 dist	6 prox+6 dist	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	?	1
4 5 prox+5 dist	(1/4 prox)+4 dist	5 prox+5 dist	4 prox+4 dist	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
5 5+7	(1/4)+(10/17/10/15)	5+7	4+10 dist	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 6 prox+6 dist	(17/5/12 prox)+12 dist	6 prox+6 dist	12 prox+12 dist	1	1	1	1	1	1	1	1	1	1	0	1	?	1	1	1	1	1
7 7 prox+7 dist	(10/17/10)+15	7 prox+7 dist	10 dist+15	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
8 9 prox+9 dist	(17/1 prox)+(1 dist/17)	9 prox+9 dist	1 prox+1 dist	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
9 9+11	(17/1/17)+16	9+11	17 mid dist+16 dist	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	?
10 X+11	X+16	X+11	X+16 dist	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
11# 16a+16b	(14/8 prox)+8 prox	16a+16b	(14prox/8prox)a+(8 prox)b	1	1	1	1	1	1	1	1	1	1	1	0	?	?	?	?	?	?

12 #	15+16a+20	14+(14/8 prox)+(17/10)	15+16a	14 dist+(14 prox/8 prox)a or 14 dist+(8 prox/14 prox)b	0	0	0	0	0	0	0	0	0	0	0	1	?	?	0	?	?	?
13 #	15+16a+20	14+(14/8 prox)+(17/10)	16a+20	(14 prox/8 prox)a+(10/17 mid prox) or (8 prox/14 prox)b+(10/17 mid prox)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	?	0
14	inv (10)	inv (16/17/11)	inv (10)	inv (11 dist)	1	1	1	1	1	1	1	1	1	1	0	1	?	1	?	1	?	?
15	inv (11)	inv (16)	inv (11)	inv (16)	0	0	0	0	0	0	0	0	1	1	1	1	?	1	?	1	1	?
16	inv (14)	inv (5/11)	inv (14)	inv (5 mid)	0	0	0	0	0	0	0	0	1	1	1	1	?	?	?	?	?	?
17	inv (15)	inv (14)	inv (15)	inv (14)	0	0	0	0	0	0	0	0	1	1	1	1	1	?	?	?	?	1
18 *	inv (16)-1	inv (14/8)	inv (16)-1	inv (8 prox)-1	0	0	0	0	0	0	0	0	1	1	0	1	?	?	?	?	?	?
19	inv (17)	inv (13/18/13/2)	inv (17)	inv (13 prox)	0	0	0	0	0	0	0	0	1	1	1	1	1	?	?	?	?	?
20	inv (18)	inv (18)	inv (18)	inv (18)	0	0	0	0	0	0	0	0	1	1	1	1	1	?	?	1	1	?
21	inv (19)	inv (8)	inv (19)	inv (8 dist)	0	0	0	0	0	0	0	0	0	0	1	1	1	?	?	?	?	?
22	inv (20)	inv (17/10)	inv (20)	inv (10 mid)	0	0	0	0	0	0	0	0	0	0	1	1	?	?	?	?	?	?
23 *	inv (16)-2	inv (14/8)	inv (16)-2	inv (8 prox)-2	0	0	0	0	0	0	0	0	0	0	1	0	?	?	?	?	?	?

"+" denotes position of the breakpoint. "?" represents unknown or ambiguous character states resulting from chromosomal disruption and/or the lack of reciprocal painting data. The abbreviations "dist", "mid", "prox" refer to the distal, middle and proximal segments of chromosomes respectively. "a" and "b" refer to unidentifiable chromosomal segments resulting from the uncertain location of the breakpoint involved in the disruption of RNO 16 in *H. delacouri* (see #). In addition, segment "a" is involved in the complex segmental association identified in *H. delacouri* where the chromosome end (i.e., MMU 8 prox or 14 prox of RNO 16 ortholog) involved in the segmental association is not known (see #). * indicates two different breakpoints/characters identified involving inversion of RNO 16 ortholog.

The cladistic analysis retrieved a single most parsimonious tree with L = 25 steps (Figure 3.13). Bootstrap support was moderate to high (BP = 55-90) and global statistics were high (CI = 0.92 and RI = 0.963). Note that moderate bootstrap support is not surprising given the low number of informative characters in the analysis, a finding that is inherent with cytogenomic analyses (Dobigny *et al.,* 2004a). The *a posteriori* polarized chromosome changes (using *H. delacouri* as outgroup) were inferred directly from the tree topology obtained herein. The *Rattus s. l.* complex was retrieved as a monophyletic group in our analysis, with the topology of *Rattus s. l.* showing a basal position for *M. surifer*, followed by *N. fulvescens*, *L. edwardsi* and, lastly, a node that is largely unresolved (i.e., a number of unresolved polytomies comprising *B. berdmorei* and *B. bowersi*, all the *Rattus* taxa, and *B. savilei*).

Importantly, the *a posteriori* polarization of chromosome changes (including additional taxa) shows 14 synapomorphies, six autapomorphies and one homoplasic character (Figure 3.13). Note that character 10 does not represent an autapomorphy in *B. savilei*, as this character is also present in the outgroup taxon, *H. delacouri* (see below). Two synapomorphies were retrieved that unite Rattini, namely the segmental association of RNO 7prox+7dist (RNO 7) and RNO 16prox+16dist (RNO 16), characters 7 and 11, respectively (Table 3.5). Inversions of RNO 19 and 20 (characters 21 and 22) unite *Rattus s. l.* to the exclusion of *M. surifer* (basal to *Rattus s. l.*). Two synapomorphies underpins the remainder of *Rattus s. l.*, excluding *N. fulvescens* and *M. surifer* (segmental association that gave rise to RNO 1 and RNO2, character 1 and 2). Six synapomorphies were retrieved supporting the grouping of *Rattus s. s., B. savilei, B. bowersi* and *B. berdmorei* (inversions of RNO 11 and 14 to 18; characters 15 to 20), the most robust node in the analysis (BP = 90). A single syntenic association was shared between *R. rattus* and the *Berylmys* genera (*B. bowersi* and *B. berdmorei*), namely Rb 9.11 (character 9) that supported their grouping in the karyotypic

tree. Five autapomorphies were identified in *M. surifer* (fissions of RNO 4 to 6, inversions of RNO 10 and 16; characters 3, 4, 6, 14 and 23, respectively) and a single autapomorphy was present in *R. rattus,* Rb 5.7 (character 5). Rattini was united with Murini, Arvicanthini and Apodemini on the basis of a single character—the syntenic association of the whole RNO 9 segment (character 8). An example of convergent evolution in this dataset is the fission of RNO 2 (character 2) in *B. savilei*, *N. fulvescens, M. surifer*, the additional taxa (Murini, Arvicanthini and Apodemini) and *H. delacouri* (outgroup).
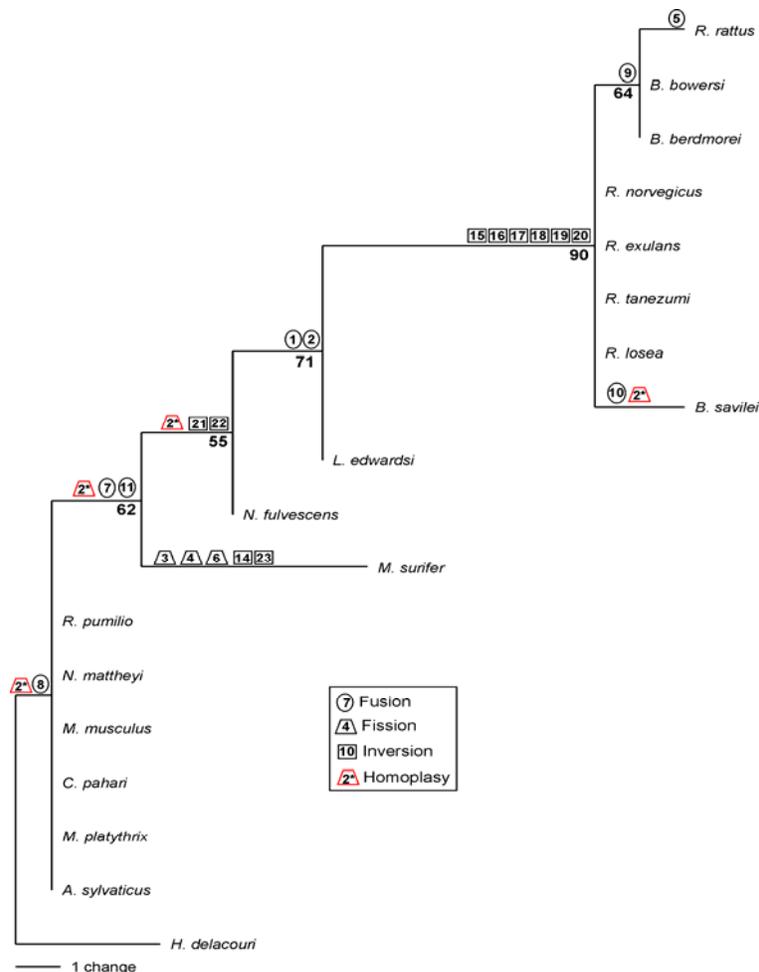
Figure 3.13. Chromosomal phylogeny retrieved by PAUP. All chromosomal changes were mapped *a posteriori* to the tree. Numbers on branches refer to characters described in Table 3.5. Numbers in bold at each node indicate bootstrap support. Note that character 10 also occurs in the outgroup taxon, *H. delacouri* (see below).

Mapping the chromosomal rearrangements onto a Rattini consensus molecular tree

In total, 10 synapomorphies, six autapomorphies and four homoplasic characters were identified following the mapping of the chromosomal characters to a Rattini sequence based phylogenetic tree (Verneau *et al.,* 1997; Lecompte *et al.,* 2008; Pagès *et al.,* 2010; this study - Chapter 5) (Figure 3.14). These findings were largely consistent with the previous cladistic analysis except for a disruption of three associations that were considered synapomorphic signatures in the PAUP analysis, but are shown herein to be homoplasic characters. One of the disrupted associations corresponded to the character uniting *Berylmys* with *R. rattus* (Rb 9.11, Figure 3.13) in the PAUP analysis, but which appears as a convergent event when mapped to the more robust molecular tree's topology (see Figure 3.14). The other important discrepancy between the PAUP-based cladistic analysis and the mapping of chromosome changes to the sequence-based tree involved the two synapomorphies that unite *Rattus s. l.* (fusions that gave rise to RNO 1 and 2) to the exclusion of *N. fulvescens* and *M. surifer*. Thus the association of *L. edwardsi* and *N. fulvescens* as sister taxa, as seen by the robustly supported sequence data, was lost in the PAUP analysis. As a consequence, the segmental associations that gave rise to RNO 1 and 2 represent homoplasic characters (i.e., either reversal or convergent events, see below). An additional convergent character that was identified between the outgroup taxon, *H. delacouri*, and *B. savilei* in both methods utilised deserves special mention. This convergent event involved a sex-autosome translocation (character 10) comprising *R. norvegicus* chromosome orthologs RNO X and 11 in both species (discussed below). The remaining synapomorphies, autapomorphies and homoplasies are all consistent with the PAUP analysis and the sequence-based tree's topology.
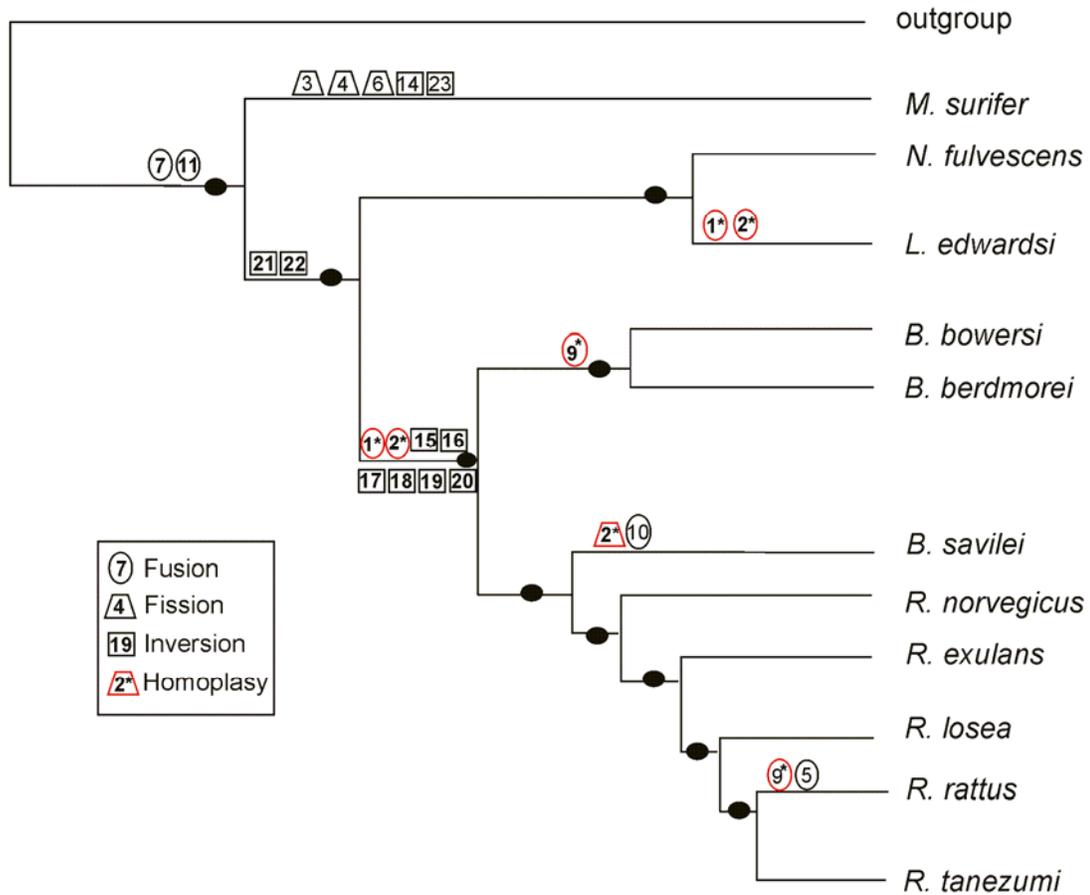
Figure 3.14. Phylogeny generated by mapping the chromosome changes to the *Rattus s. l.* molecular phylogenetic tree (modified from Verneau *et al.,* 1997; Lecompte *et al.,* 2008; Pagès *et al.,* 2010; this study - Chapter 5). Numbers on branches refer to characters described in Table 3.5. Black ovals indicate strong nodal support (BI>0.95; BP>95). Note that character 10 also occurs in the outgroup taxon, *H. delacouri*.

FISH with BACs

The order of BAC clones located on the q arm is unchanged in both the acrocentric and submetacentric morphs of *R. tanezumi* chromosomes RTA 1, 12 and 13 (that correspond to the submetacentric orthologs of RNO 1, 12 and 13, respectively), thus indicating that these BAC clones are not included in the breakpoint region (data not shown). Importantly, however, the order of the BAC clones selected for their positions on the p arms of chromosome pairs RNO 1, 12 and 13 (all submetacentric in *R. norvegicus*) are inverted in the

81

acrocentric form of the heteromorphic pairs RTA 1, 12 and 13 of *R. tanezumi*. For example,

the order of the BAC clones on the p arm (from the terminal segment to the proximal

segment of the submetacentric RNO 12) is: BAC clone 5 (12p12-Cy3) followed by BAC

clone 6 (12p11-FITC). The order is inverted in the *R. tanezumi* acrocentric morph (Figure

3.15b). This confirms that the changes in morphology of these chromosomes is due to

pericentric inversions in the three pairs and are therefore not simply a reflection of

centromeric shifts or heterochromatic arm variability in *R. tanezumi* (Figure 3.15).
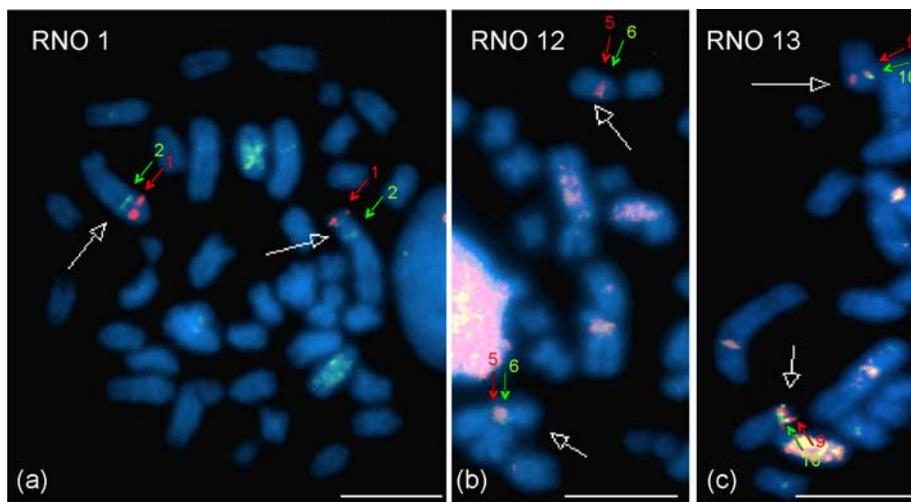
Figure 3.15. Representative metaphase spreads from a *R. tanezumi* specimen that is heterozygous for the acrocentric and submetacentric morphs of chromosome orthologous to RNO 1, 12, 13 (white arrows). Green and red arrows indicate the localization of BAC clones that map to **(a)** *R. norvegicus* (RNO) 1p (BAC clone 1 red, BAC clone 2 green); **(b)** RNO 12p (BAC clone 5 red, BAC clone 6 green); **(c)** RNO 13p (BAC clone 9 red, BAC clone 10 green). Scale bar = 10 μm. See text for details.

In addition, these data also identified an error in the Ensembl database. This

resource shows *R. norvegicus* BAC clone 2 located at band RNO 1p11 (Table 3.2.). FISH

mapping of this clone showed, however, that it hybridized to RNO 1q11 (Figure 3.16) – in

other words within the long arm (and not the short arm) of *R. norvegicus* chromosome 1, and

similarly so in *R. tanezumi*.
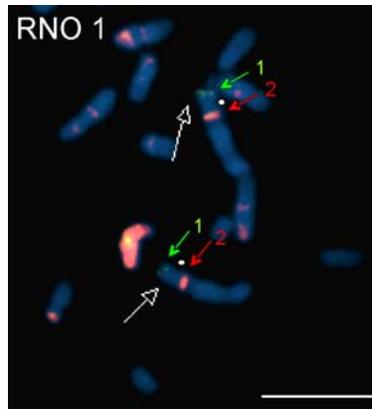
Figure 3.16. Representative metaphase spread of *R. norvegicus* (RNO) showing localization of BAC clones (green and red arrows) on chromosome pair 1 (white arrow), centromere position indicated by white oval. The BAC clone selected for 1p13 (BAC clone 1, green) confirms the assignment given in Ensembl database while BAC clone 2, reportedly mapping to RNO1p11 by Ensembl, is located on the q arm (approximately 1q11) in both *R. norvegicus* and *R. tanezumi*. Scale bar = 10 µm.

DISCUSSION

Phylogenetic analysis and chromosome evolution within Rattini

Cross-species painting using *R. norvegicus* whole chromosome paints as well as reciprocal painting between *R. norvegicus*, *M. surifer* and *R. rattus* facilitated the generation of genome-wide comparative maps among nine Rattini species and *H. delacouri*. In total seven different murine genera were investigated in this study - *Bandicota*, *Leopoldamys*, *Rattus*, *Berylmys*, *Niviventer*, *Maxomys* and *Hapalomys*. These comparative maps, together with published data (from representatives of Murini, Arvicanthini and Apodemini), allowed the tracking of chromosomal rearrangements that have occurred during the evolution of *Rattus s. l.*, as well as permitting an assessment of their usefulness in discriminating phylogenetic relationships among species.

The cladistic analysis (Figure 3.13) and mapping of chromosomal characters onto the molecular tree (Figure 3.14) demonstrated that pericentric inversions, which were

83

confirmed in three pairs of rat chromosomes (RTA 1, 12 and 13) by BAC-mapping analysis (see above), appear to predominate in the chromosome evolution of *Rattus s. l.* Importantly, both methods retrieved several cytogenetic signatures that support some of the natural groupings at higher systematic levels (i.e., the basal position of *M. surifer* within Rattini). In addition, both methods identified two cytogenetic signatures uniting the Rattini (syntenic associations that gave rise to RNO 7 and RNO 16, character 7 and 11 respectively). However, character 11 is not considered a good cytogenetic signature since it requires further confirmation (i.e., reciprocal painting) in the Murinae taxa included in this study (*R. pumilio, A. sylvaticus, M. musculus, N. mattheyi*, and *M. platythrix*).

The monophyly of *B. bowersi, B. berdmorei, B. savilei* and the *Rattus* assemblage is well supported by six intrachromosomal rearrangements in the form of pericentric inversions (characters 15 to 20), thus representing new derived synapomorphies for this group. This is consistent with prior reports that suggested pericentric inversions play a key role in the karyotype evolution of *Rattus* and *Bandicota* genera (Gadi and Sharma, 1983). It has been suggested that these rearrangements may promote rapid genetic diversification among populations by causing changes to gene order, and repressing recombination which prevents admixture of the newly evolved combinations of alleles (Noor *et al.,* 2001; Hoffman *et al.,* 2004; Ayala and Coluzzi, 2005). This allows different combinations of alleles to evolve independently enabling important adaptive opportunities and the protection of favourable combinations of adaptive alleles. It has been postulated that these processes may explain the association of balanced inversion clines along adaptive gradients occurring along geographical clines (Ayala and Coluzzi, 2005; Kirkpatrick and Barton, 2006) (for further discussion of these aspects see FISH with BACs below where the presence of pericentric inversions were unambiguously confirmed).

The unresolved polytomy of the *Rattus* lineage with *B. savilei* (Figure 3.13) does not refute the suggestion that *Bandicota* should be included in *Rattus s. s* based on L1 data (Usdin *et al*., 1995; Verneau *et al*., 1997, 1998; Chaimanee and Jaeger, 2001) (Chapter 5). However, the suggestion that the large metacentric chromosome pair in *B. bowersi* and *B. berdmorei* arose as a result of an Rb fusion of two *R. rattus* chromosomes (Yosida, 1973) is refuted. In fact, this metacentric pair arose due to an Rb fusion of two *R. norvegicus* chromosomes (Rb 9.11), a character that is shared with *R. rattus* (see Figure 3.13), and which corresponds to RRA 4 (Cavagna *et al.,* 2002). This most likely represents a convergent evolutionary event given that the basal position of *Berylmys* to the *Bandicota* and *Rattus* assemblage is strongly supported by numerous DNA-based studies (i.e., Usdin *et al*., 1995; Verneau *et al*., 1997, 1998; Chaimanee and Jaeger, 2001; Lecompte *et al.,* 2008; Pagès *et al.,* 2010; Chapter 5 – this study).

The chromosomal tree does not support the close association of *L. edwardsi* and *N. fulvescens* detected in molecular analyses (i.e., Verneau *et al.,* 1997; Lecompte *et al.,* 2008; Pagès *et al.,* 2010; Chapter 5 - this study), since *L. edwardsi* is basal to *Rattus s. s.* (which includes *B. savilei* and *Berylmys* represented by *B. bowersi* and *B. berdmorei*) in the PAUP analysis (see Figure 3.13). Two cytogenetic characters (specifically two fusions that gave rise to RNO 1 and 2) unite *L. edwardsi* with the *Rattus* assemblage, *Berylmys*, and *B. savilei* to the exclusion of *N. fulvescens* (Figure 3.13). It is suggested that these two fusions represent homoplasic characters given the robust and congruent topologies resulting from various DNA-based methods (L1 amplification sites, mitochondrial and nuclear gene sequences). There are two possible interpretations for the occurrence of these homoplasic characters: (i) their absence in *N. fulvescens* represents a reversal, or (ii) they arose separately in *L. edwardsi* and *Rattus s. s.* (including *B. berdmorei, B. bowersi* and *B. savilei*) and consequently represent convergent events (Figure 3.14). Comparisons of G-band patterns

with other *Niviventer* species were not useful in clarifying which of the two mechanisms was most likely. This was exacerbated by the fact that the few available studies are limited to conventional staining (Baskevich and Kuznetsov, 2000; Li *et al.,* 2008), the only exception being *N. confucianus* (Wang *et al.,* 2003). However, homologs of *N. fulvescens* in *N. confucianus* (corresponding to RNO 1 or 2 orthologs) could not be unambiguously identified. Nonetheless, these findings are in agreement with suggestions that *L. edwardsi* is chromosomally more closely related to *Berylmys, Bandicota* and *Rattus* than to *Niviventer* (Gadi and Sharma, 1983) thus supporting hypothesis (ii). On the other hand, the disruption of RNO 2 (representing a convergent event between *B. savilei* and *M. surifer*, *N. fulvescens*, and the additional Murinae taxa included in the study, Figure 3.13) may actually represent a site of breakpoint reuse. For example, the disruption of RNO 2 appears to correspond to a breakpoint previously identified in *Mus*, *Peromyscus* and *Cricetulus* (Mlynarski *et al.,* 2010) thus supporting the "Fragile Breakage Model" of chromosome evolution proposed by Pevzner and Tesler (2003). These authors suggest that breakpoint reuse is frequent and that regions between syntenic blocks have a propensity to undergo rearrangement events.

A convergent sex chromosome-autosome translocation was identified in both *B. savilei* and *H. delacouri* involving a fusion of RNO X and 11. This is notable as X-autosome fusions are considered to be highly deleterious chromosomal rearrangements (Ashley, 2002; reviewed in Dobigny *et al.,* 2004b) and, consequently, are not expected to be susceptible to convergence. The deleterious nature of autosome-gonosome translocations are thought to reflect the effects of X-inactivation on the autosome and the differing replication times (autosomes and sex chromosomes replicate at different times in cell division; Ashley, 2002; Dobigny *et al.,* 2004b). The process of X-inactivation is required to compensate for the unequal X-gene dosage between different sexes in mammals (one X in males and two Xs in females) through transcriptional silencing of most genes on one X in females (Lyon, 1961;

Brockdorff, 1998; Boumil and Lee, 2000; Avner and Heard, 2001). Nevertheless, despite these limitations, X-autosome translocations are not that uncommon in mammals (in rodents, i.e., Gerbillinae: Ratomponitina *et al.,* 1986; Dobigny *et al.,* 2002a, 2004b; Murinae: Veyrunes *et al.,* 2004; Bathyergidae: Deuve *et al.,* 2006; in bats: Tucker, 1986; Richards *et al.,* 2010; in shrews: Pack *et al.,* 1993; in carnivores: Fredga, 1972; in bovids: Rubes *et al.,* 2008).

It has been postulated that an interstitial heterochromatic block (IHB) that occurs between the ancestral X chromosome and the autosomal component acts as a barrier preventing the spread of X-inactivation to the autosomal component (Ratomponitina *et al.,* 1986; Dobigny *et al.,* 2004b). Interstitial heterochromatic blocks are believed to consist of mostly repetitive sequences, i.e., rDNA and telomeres (Parish *et al.,* 2002; Dobigny *et al.,* 2003b, 2004b; Veyrunes *et al.,* 2004), and these blocks are thought to represent a "regulatory super structure" for differential replication timing (Dobigny *et al.,* 2004b). This epigenomic hypothesis applies to both *H. delacouri* and *B. savilei* in this study, as pericentromeric IHBs are present between the X chromosome and the autosomal component in these two unrelated species (Chapter 2). Furthermore, the X-autosome translocation was identified in an additional male *B. savilei* specimen (see Results), showing that this mutation is fixed within *B. savilei* and any ill-effects of the spread of X-inactivation have been countered. These results imply that although the X and autosomal component are structurally linked, they remain functionally separate from each other.

Tempo of chromosomal change in the *Rattus sensu lato* complex

The rate of chromosomal evolution was determined using chromosomal changes identified in this study (see Figure 3.14) and nodal dates estimated from a published molecular analysis (Lecompte *et al.,* 2008, which included representatives of each murine

genus investigated in this study). This makes it possible to calculate the rate of chromosomal

reorganization that characterises each branch of the Rattini tree (Figure 3.17) using the ratio

between the number of chromosome rearrangements and the time of divergence.
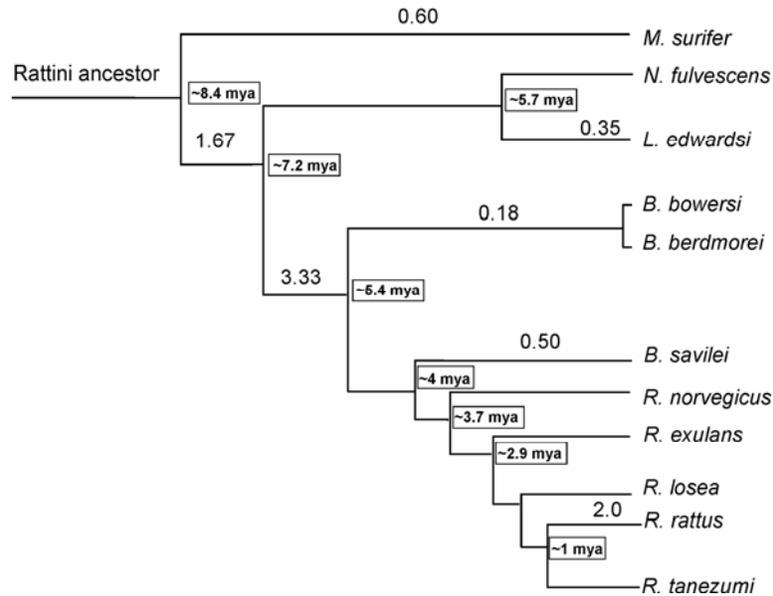
Figure 3.17. Rates of chromosome evolution against divergence times in Rattini. Numbers in squares indicate molecular divergence estimates in million years (mya) as inferred by Lecompte *et al.* (2008). Numbers above the branches represent the average rates of rearrangement per million years.

Karyotypes may differ by many chromosome rearrangements within the same

family (i.e., Muridae), and even at the species level. For example, 19 mouse autosomes

correspond to 49 orthologous regions in the rat genome (*Mus*/*Rattus* split ~12 mya; Guilly *et al.,* 1999; Stanyon *et al.,* 1999; Helou *et al.,* 2001; Nilsson *et al.,* 2001). However, although

at least 19 chromosome rearrangements (fusions, fissions and pericentric inversions) have

occurred within *Rattus s. l.* (Figure 3.14), the variation is somewhat muted when one takes

each species/lineage separately into account (see Figure 3.17). Evolutionary rates of

chromosomal reorganization are extremely low in nearly all Rattini branches (≤0.6

rearrangements per million years, represented hereafter as R/myr) with pericentric inversions

predominating among the chromosomal rearrangements identified. Slightly accelerated rates of karyotypic evolution occurred during a short phase at the nodes uniting *N. fulvescens* and *L. edwardsi* with the rest of *Rattus s. l.* and the node uniting *Berylmys* with *B. savilei* and the *Rattus* assemblage. These involved two (1.67 R/myr) and six pericentric inversions (3.33 R/myr) respectively that were fixed within a short time period (1.2 and 1.8 myr, respectively). This was followed by a period of stasis, with the only exceptions being a single fusion in the karyotypes of *B. bowersi* and *B. berdmorei*, a fission and fusion in *B. savilei* karyotype, and two fusions which restructured the *R. rattus* karyotype. As a consequence, the *Rattus s. l.* species analysed herein exhibit a high degree of genome conservation and, if inversions were to be excluded, *L. edwardsi* and *Rattus s. s.* have retained largely invariant karyotypes since their last common ancestor ~7 mya.

The rate of chromosome evolution estimated in the present study is remarkably constrained in comparison to the extensive chromosome restructuring identified in most other murid rodents, such as those in *Mus* (i.e., Britton-Davidian *et al.,* 2000; Veyrunes *et al.,* 2004; Pialek *et al.,* 2005). For example, the four subgenera of *Mus* which diverged almost instantaneously within 1 myr, are represented by a rate of 13 mutations per million years (Veyrunes *et al.,* 2006). Moreover, 16 Rb fusions became fixed in the *M. minutoides* clade in less than 1 million years (Veyrunes *et al.,* 2010). Similarly, the recently diverged West African gerbilline *Taterillus* species (~0.4 mya; Dobigny *et al.,* 2002a) have undergone extremely rapid rates of chromosome evolution (e.g., *T. tranieri, T. pygargus, Taterillus* sp1 and *Taterillus* sp2) and they are characterised by an rate of 45 changes per myr (Dobigny *et al.,* 2005). Other examples of rapid chromosome evolution exist in non-rodent mammals such as arctic and red foxes that diverged 2 mya with their karyotypes already differing by over 30 chromosome rearrangements (Nash *et al.,* 2001). However, the most remarkable example of rapid chromosome evolution in non-rodent mammals is provided by muntjac

deer. The Chinese (*Muntiacus muntjak reevesi*) and Indian muntjac (*M. vaginalis*), which last shared a common ancestor only 0.9-3.7 mya, already have extremely diverse karyotypes comprising 46 and six/seven chromosomes, respectively (Yang *et al.,* 1997a, b; Wang and Lan, 2000; Tsipouri *et al.,* 2008).

In contrast, constrained chromosome evolution as was observed in Rattini has also been reported in other rodents such as squirrels (Sciuridae; Richard *et al.,* 2003b; Stanyon *et al.,* 2003; Li *et al.,* 2004, 2006). The latter studies demonstrated that squirrels (i.e., *Menetes berdmorei*, *Sciurus carolinensis* and *Callosciurus erythraeus*), in marked contrast to the mouse, have undergone slow to moderate genome evolution with their karyotypes showing extensive conservation with the human one. American and Australian marsupials are another example of chromosomal stasis and even though these two species groups diverged ~70 mya they still display a high degree of genome conservation (i.e., Rens *et al.,* 1999, 2001, 2003).

In summary therefore, it is apparent that chromosome evolution can vary between lineages (O'Brien *et al.,* 1999; Murphy *et al.,* 2005), and that not all rodents undergo the rapid and extensive genome rearrangement originally hypothesised for murids. This study highlights the fact that rampant chromosomal rearrangements in murids may, in fact, be limited to Murini (i.e. *Mus*: Veyrunes *et al.,* 2006, 2010), while Rattini are constrained. Similar patterns of contrasting rates of chromosome evolution (conserved vs. highly rearranged) have been reported in other mammalian lineages, for example equids are characterised by very rapid rates (2.92 to 22.2 R/Myr), while tapirs and rhinoceros display much slower rates (≤0.77 R/Myr) (Trifonov *et al.,* 2008).

Reconstructing the Rattini and Murinae ancestral karyotypes

The mapping of polarized chromosome changes along the karyotypic phylogeny allows an objective assessment of ancestral states for the reconstruction of the putative

Rattini ancestral karyotype (RAK; Figure 3.18). Moreover, comparisons between the putative *Mus* ancestral karyotype (Veyrunes *et al.,* 2006), *R. pumilio* (Arvicanthini; Rambau and Robinson, 2003), *A. sylvaticus* (Apodemini; Matsubara *et al.,* 2004) and putative ancestral Rattini karyotype (determined here) with *H. delacouri* (basal to Murinae, Pagès *et al.* unpubl.; Chapter 5) allowed a critical re-assessment of ancestral syntenies for the reconstruction of the basal Murinae ancestral karyotype (Table 3.6 and Figure 3.18). In short, the conserved combination of chromosomal segments occurring in different tribes may signify ancestral syntenies but do not necessarily define (as synapomorphies) specific groups (see Robinson and Ruiz-Herrera, 2008). Syntenies in *R. pumilio* and *A. sylvaticus* were assigned using mouse paints, and although several extremely small conserved segments of *M. musculus* have been identified in *R. norvegicus* (see Figure 3.4; Helou *et al.,* 2001; Nilsson *et al.,* 2001), most were absent in these studies. These small conserved fragments of *M. musculus* chromosomes may have been undetected in *R. pumilio* and *A. sylvaticus* at the resolution of FISH provided in the latter studies.
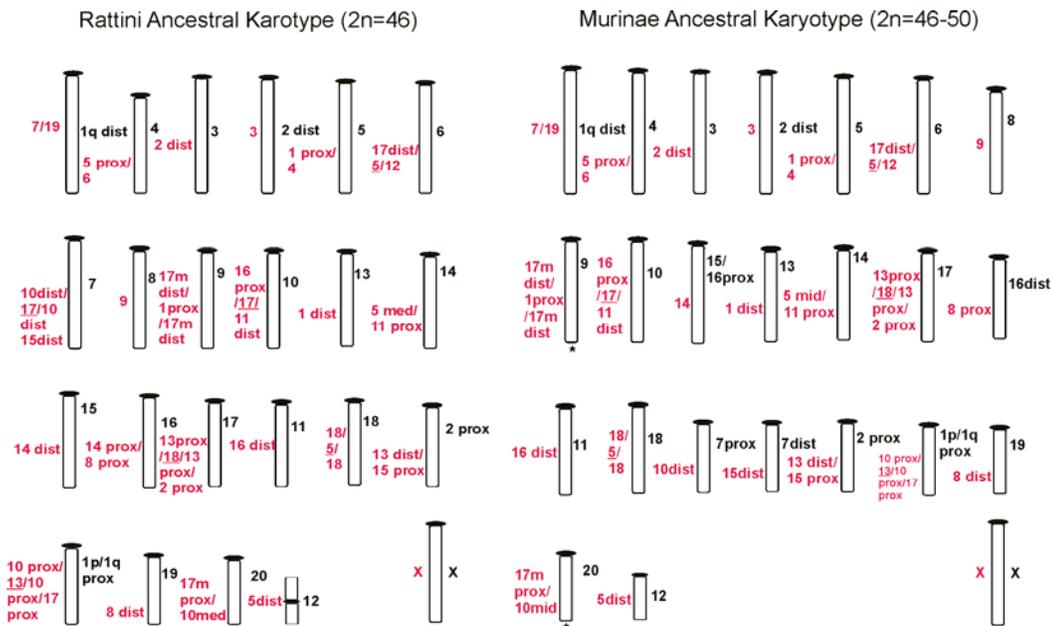
Figure 3.18. Putative ancestral karyotype of Rattini (left) and Murinae (right) with *R. norvegicus* and *M. musculus* homologies numbered on the right (black) and left (red)

respectively. * indicates ambiguous ancestral states. Underlined *M. musculus* segments represent small, generally undetected conserved segments between mouse and rat (see text for details). "dist", "mid", "prox", "m prox" refer to the distal, middle, proximal and mid proximal segments of the chromosomes respectively.

The analyses presented herein suggest that the RAK had 2n = 46 and that comprised, with one exception (a single small metacentric pair), only acrocentric chromosomes. It shares 16 autosomal pairs that were conserved *in toto* (entire block or syntenic segment: MMU 7/19, 10prox/17prox, 13dist/15prox, 3, 2dist, 5prox/6, 1prox/4, 9, 16dist, 5dist, 1dist, 5med/11prox, 14dist, 13prox/2prox, 18, and 8 dist) with the 2n = 46 acrocentric ancestral *Mus* karyotype proposed by Veyrunes *et al.* (2006).

Critical comparisons between the RAK, *Mus* ancestral karyotype, representatives of Apodemini and Arvicanthini with *H. delacouri* (Table 3.6) suggests that a 2n = 46-50 acrocentric chromosome complement is characteristic of the base of Murinae (Figure 3.18b). The variable diploid number is due to uncertainty regarding the character states of RNO 9 and 20 (see Table 3.6) where the latter (RNO 20) may potentially be fused with another chromosome (fused with RNO 16 in *H. delacouri*) in the ancestral state. As a consequence the Murinae ancestral karyotype would be 2n = 46. It is, however, not possible to unambiguously identify the RNO 20 ortholog in *Mus* ancestral and *R. pumilio* karyotypes through G-band comparisons since there are two conserved syntenic associations detected in the *R. norvegicus* karyotype corresponding to MMU 17/10 (see Figure 3.4), of which one corresponds to RNO 1p/1q prox, and the other to RNO 20. The RNO 9 ancestral state may be disrupted based on inferences from *H. delacouri* and *Mus* ancestral karyotypes and as a result, the Murinae ancestral karyotype would be 2n = 50. Unfortunately conservation of the RNO 9 breaksite location cannot be confirmed based on G-band comparisons (there is no MSU 1/17 association detected in *Mus* ancestral karyotype). Reciprocal painting data is

required to confirm the true homology of many of the mouse segments, particularly the ones involved in the characters of interest detailed above. Future studies may benefit from the inclusion of *R. norvegicus* painting probes corresponding to these rearrangements to ensure the unambiguous identification of their chromosome orthologs since these may be biased by the highly rearranged nature of the mouse genome (Bourque *et al.,* 2004; Murphy *et al.,* 2005; Romanenko *et al*., 2007). Moreover the inclusion of other Murinae taxa and families such as Cricetidae would assist in further defining the composition of rodent ancestral karyotypes.

Table 3.6. Presence and/or absence of putative Rattini ancestral syntenic associations in Murinae ancestral karyotype. Chromosomal correspondence between the Rattini ancestral and *M. musculus* is shown and the presence and absence of Rattini ancestral segment is indicated by "+" and "-", respectively.

Rattini ancestral	*M. musculus*	*Mus* ancestral	Arvicanthini (*R. pumilio*)	Apodemini (*A. sylvaticus*)	*H. delacouri* (basal)
1q dist	7/19	+	+	+	+
1p/1q prox	10prox/13/17prox	+	+	+	+
2 prox	13dist/15prox	+	-	+	+
2dist	3	+	+	+	+
3	2dist	+	+	+	+
4	5prox/6	+	-	+	+
5	1prox/4	+	-	-	+
6	17dist/5/12	-	-	+	+
#7	10dist/17/15dist	-	-	-	-
8	9	+	-	+	+
9	17dist/1prox/17dist	-	+	+	-
10	16prox/17/11dist	-	+	+	+
11	16dist	+	-	+	+
12	5dist/8	+	-	+	+
13	1dist	+	+	+	+
14	5mid/11prox	+	-	+	+
*15	14dist	+	+	+	+
*16	14prox/8prox	-	-	-	-
17	13prox/18/2prox	+	-	-	+
18	18/5	+	+	+	+
19	8 dist	+	+	+	+
20	17mid prox/10 mid	-	+	+	+
X	conserved				

# indicates disruption of Rattini 7 in Murinae ancestral karyotype corresponding to MMU 10 and 15 elements. * refers to the presence of the whole MMU 14 element in Murinae ancestral karyotype (not disrupted as seen in Rattini ancestral karyotype). Regions shaded in grey denote ambiguous characters resulting in ambiguity of Murinae ancestral karyotype diploid number (2n = 46-50, see text for details). "dist", "mid", "prox" refer to the distal, middle and proximal segments of the chromosomes respectively.

Importantly, these findings provide the first comprehensive hypothesis of the putative RAK and revise the previously postulated ancestral murid karyotypes of Stanyon *et al.* (2004) and Engelbrecht *et al.* (2006). Stanyon *et al.* (2004) postulated that the murid ancestral karyotype had a diploid number of 2n = 54 based on the conservation of proposed ancestral syntenic associations from alignments of representatives of Murini (*M. platythrix* and *M. musculus*), Rattini (*R. rattus* and *R. norvegicus*), Arvicanthini (*R. pumilio*), and Apodemini (*A. sylvaticus*). Many of the ancestral syntenic associations were conserved in the reappraisal of the murid ancestral karyotype by Engelbrecht *et al.* (2006; 2n = 52) which included a representative of Otomyini (*O. irroratus*). It was possible to retrieve all the syntenic associations suggested by Stanyon *et al.* (2004) in the putative Murinae ancestral karyotype hypothesised herein except for four cases where syntenic associations were lost. These were MMU 6a, 6b, 11a and 12a, where MMU 6a and 6b correspond to one ancestral state (fused) in the present study (MMU 5prox/6); elements corresponding to MMU 11a or 12a could not be identified (Table 3.6). Furthermore, the segmental association suggested by Engelbrecht *et al.* (2006; MMU 10b+17b or MMU 10c+17c fused with MMU13b+15a) that led the authors to propose a 2n = 52 putative murid ancestral karyotype was not retrieved in the MAK proposed here. The two ancestral karyotypes presented here (the putative RAK and the revised MAK) may be useful in large-scale investigations of genome organization rather than the widely used and highly rearranged mouse genome.

FISH with BAC

Chromosome painting using BAC clones together with a representative of *Rattus s. s.* (*R. tanezumi*) has unequivocally shown that the polymorphisms affecting chromosome pairs 1, 12 and 13 (*R. norvegicus* orthologs) are due to pericentric inversions. To date, polymorphisms for these three pairs have been reported only in *Rattus* species (see

94

Badenhorst *et al.,* 2009, Chapter 2). The pericentric inversions detected in these rodents appear as floating polymorphisms (Chapter 2 and see Yosida *et al.,* 1971a, b, 1974; Yosida, 1973, 1976; 1977a; Yosida and Sagai, 1973, 1975; Gadi and Sharma, 1983; Motokawa *et al.,* 2001) that may be segregating independently in these rodents, or at the very least within the three *Rattus* species (*R. exulans, R. losea* and *R. tanezumi*) analysed in Chapter 2 (see Table 2.3), since they last shared a common ancestor ~3.7 mya (Figure 3.17). Widely distributed inversion polymorphisms have also been reported in other rodents including *Oligoryzomys nigripes* (Bonvicino *et al.,* 2001; Paresque *et al.,* 2007), *M. erythroleucus* (Dobigny *et al.,* 2010a) and *O. irroratus* (Engelbrecht *et al.,* 2011).

It has been postulated that the presence of inversions may aid speciation by preventing hybridization (see for example in *Drosophila pseudoobscura* and *D. persimilis*: Noor *et al.,* 2001 and *Anopheles gambiae*: Coluzzi *et al.,* 2002). In the latter studies it was also suggested that the chromosome inversions were linked with adaptation to different geographical areas implying a potentially adaptive value to these chromosome rearrangements; similar findings have been reported for humans (Stefansson *et al.,* 2005) and *O. nigripes* (Bonvicino *et al.,* 2001; Paresque *et al.,* 2007). Similarly Yosida (1980) postulated that a pericentric inversion in the acrocentric pair 1 of an insular Asian black rat, most likely *R. tanezumi* (2n = 42), population in Southeast Asia conferred a selective advantage to survival in warmer climates. However, the latter finding should be treated with caution, and deserves further investigation due to the use of predominantly conventionally stained, non-banded karyotypes. In addition, *R. tanezumi* represents a complex where species-specific boundaries and hence taxonomy is unresolved (Usdin *et al.*, 1995; Jansa and Weksler, 2004) with extensive taxonomic revisions still being undertaken in these geographic areas (e.g., Pagès *et al.,* 2010). Moreover, note that several *Rattus* species (i.e., *R. exulans*) were and still are transported by humans (see for example Yosida, 1980; Matisoo-Smith and

Robins, 2004) thus rendering interpretations of the geographic patterns and possible associated adaptive value of inversion polymorphisms in these rodents problematic.

Importantly, the utility of BAC clones for deciphering the nature of intrachromosomal rearrangements is supported in this study. BAC clones should be further utilised to clarify the nature of the other intrachromosomal rearrangements identified herein. These include, for example, the acrocentric and submetacentric morphology of RNO 14 - 20 that are thought to have been rearranged due to pericentric inversions (see Materials and Methods) in the *Rattus s. l.* complex. The *Rattus s. s.* complex (*sensu* Lecompte *et al.,* 2008) is characterised by metacentric morphology of RNO 14 to 20 orthologs, while species occurring outside this group are largely characterised by acrocentric morphs of these *R. norvegicus* orthologs.

In conclusion, taking into account chromosomal orthologies identified among the species studied by FISH and the good correspondence of the G-banding patterns of their chromosomes, this study strongly suggests that the karyotypic conservation reflected in Rattini rodents may extend to the conservation of gene order. Consequently, it seems likely that the comparative maps will permit the transfer of gene mapping data from model (i.e., *R. norvegicus*) to non-model species (i.e., *R. exulans, R. losea, R. tanezumi, L. edwardsi, N. fulvescens, M. surifer* and *H. delacouri*). Together with the increasing availability of complete genome sequences for many species (see for example the Genome 10K project, G10K, http://www.genome10k.org/), the present study will aid the future exploration of genetic factors that differentiate between species that are important zoonotic reservoirs, and those that are not (Chapter 5).

CHAPTER 4

DISTRIBUTION OF REPETITIVE ELEMENTS AND THEIR ROLE IN THE CHROMOSOMAL EVOLUTION OF ASIAN RODENTS (RODENTIA, MURIDAE)

INTRODUCTION

Repetitive DNAs are ubiquitous components of eukaryotic genomes that offer important insights when examining the molecular basis of chromosomal evolution (Garagna *et al.,* 1995, 2001; Shan *et al.,* 2003; Kalitsis *et al.,* 2006; Gauthier *et al.,* 2010; Koo *et al.,* 2010). They occur as tandem repeats, such as satellite DNAs (satDNA), mini- and microsatellites, and in some multigenic families including the genes for ribosomal RNAs (found in the nucleolar organizer regions, or NORs, of chromosomes). Repetitive DNAs are also present as dispersed repeats that show high variability regarding their organization and location for example, mobile elements such as transposons and retrotransposons (Charlesworth *et al.*, 1994). Satellite DNAs are normally found in centromeric/telomeric heterochromatic regions, and often show high variability with regard to nucleotide sequence, reiteration frequency and distribution in the genome. Sequence analysis has confirmed their rapid evolution (e.g., Chaves *et al.,* 2000a, b; 2005) that is thought to be facilitated by concerted evolution - a process that can lead to the development of species-specific satDNAs (Ugarković and Plohl, 2002). In concerted evolution, mutations of monomer units of satDNA are fixed and homogenized within the genome of a species through molecular drive, a genetic event that operates at both the population and molecular level (Dover, 1986). In spite of the effects of concerted evolution, however, satellite repeat sequences may still be quite variable

97

with some repeats being species-specific, while others are broadly conserved across distantly related species (Ellingsen *et al.,* 2007; Acosta *et al.,* 2007).

Another theoretical consideration, in relation to satDNA evolution, concerns copy number variation. Different satellite repeats vary with respect to copy number, and it has been hypothesised that these may have existed in the form of a 'library' in a common ancestor of closely related species that have amplified differentially during cladogenesis (Fry and Salser, 1977). Consequently species-specific profiles may occur as a result of disparities in copy number within this library (Meštrović *et al.* 1998; Ugarković and Plohl, 2002; Palomeque and Lorite, 2008). Molecular analyses of satellite repeats have confirmed changes in copy number among closely related species (Ugarković and Plohl, 2002) including rodents (Slamovits *et al.,* 2001).

Telomeres, on the other hand, consist of tandemly repeated TTAGGG hexamers that are most frequently located at terminal positions where they play an important role in the stability of the chromosome, as well as in the maintenance of chromosomal integrity (Cech, 2004). However, telomeric repeats have been reported at internal sites of the chromosomes in the genomes of a number of species, and these are referred to as intrachromosomal (or interstitial) telomeric sequences (ITSs, Ruiz-Herrera *et al.,* 2008). Heterochromatic ITSs (Het-ITSs) have been identified in rodents (see for example, Castiglia *et al.,* 2006; Ventura *et al.,* 2006) and comprise large blocks of telomeric-like repeats located at centromeric or pericentromeric regions. Het-ITSs are believed to be remnants of ancestral chromosomal rearrangements that occurred during the evolution of different taxa (Finato *et al.,* 2000; Go *et al.,* 2001). Telomeric-like repeats at internal sites (generally at subterminal locations and not at pericentromeric or centromeric regions), are referred to as short-ITSs (s-ITSs). Studies indicate that a correlation exists between these sites and evolutionary chromosomal

reorganisation (Ruiz-Herrera *et al.,* 2008). In addition, it has been suggested that these telomeric repeats may themselves be involved in Robertsonian translocations as telomere shortening (which occurs with cellular aging in mammalian somatic cells) has been linked to increased end-to-end chromosome fusions (Counter *et al.,* 1992; Slijepcevic *et al.,* 1998).

It has also been proposed that rDNA clusters may contain telomeric repeats, and that their occurrence at, or close to break sites, provides the functional telomeres necessary to ensure the viability of the fissioned products (Imai, 1991; Hall and Parker, 1995). The hypothesis that rDNA clusters may have a role in healing/stabilization of naked chromosome ends is supported by studies in insects (*Neopridion abietis*, Rousselet *et al.,* 1998, 2000) and plants (*Hypochoeris radicata*, Hall and Parker, 1995). It is further supported by the regular co-localization of rDNA genes with telomeric repeats in other organisms including lemmings, *Myopus schisticolor* (Liu and Fredga, 1999), and gerbils, *Taterillus* spp. (Dobigny *et al.,* 2003b), as well as the eels *Anguilla anguilla* and *A. rostrata* (Salvadori *et al.,* 1995).

The investigation of repeats (satDNA, telomeric-like repeats and rDNA clusters) in *Rattus s. l.* and *H. delacouri* in the present study is of interest given the potential that these motifs have for facilitating chromosomal rearrangement and observations of (i) a number of rearrangements and breakpoints within *Rattus s. l.* and *H. delacouri* (Chapter 3) and (ii), two X-autosome translocations which display interstitial heterochromatic blocks (IHBs; Chapters 2 and 3). Additionally, since the chromosomal distribution of NORs has been used in a phylogenetic context in species as diverse as bovids (Gallagher *et al.,* 1999; Nguyen *et al.,* 2008), frogs (Wang *et al.,* 2009), fish (Gold and Li, 1994), the common vole (Mazurok *et al.,* 2001) and in species of the genus *Apodemus* (Matsubara *et al.,* 2004), the phylogenetic utility of NORs based on location and numbers in *Rattus s. l.* and *H. delacouri* was also investigated.

MATERIALS AND METHODS

Chromosome preparation and DNA extraction

Protocols detailing the establishment of fibroblast cell lines, chromosome harvests and preparations are presented in Chapter 2. Genomic DNA (gDNA) was extracted from tissue samples and/or pelleted fibroblasts of ten species (Table 4.1) using the QIAamp DNA Micro kit (Qiagen) following manufacturers' instructions.

Table 4.1. List of species and specimens included in the study. Coordinates for each sampling locality are provided.

Species	Field number	Sex	Origin	Grid reference	2N
Rattus losea	R 4724	F	Loei	17°29'N, 101°43'E	42
Rattus tanezumi	R 4003	F	Kalasin	16°49'N, 103°53'E	42
Rattus exulans	R 4033	F	Phrae	18°09'N, 100°08'E	42
Bandicota savilei	R 4408	F	Loei	17°29'N, 101°43'E	43
Berylmys berdmorei	R 4406	F	Loei	17°29'N, 101°43'E	41
Berylmys bowersi	R 4400	M	Loei	17°29'N, 101°43'E	40
Leopoldamys edwardsi	R 5239	M	Loei	17°29'N, 101°43'E	42
Niviventer fulvescens	R 4519	F	Loei	17°29'N, 101°43'E	46
Maxomys surifer	R 4404	M	Loei	17°29'N, 101°43'E	52
Hapalomys delacouri	R 5237	M	Loei	17°29'N, 101°43'E	48
* Rattus norvegicus	/	F	Stellenbosch, South Africa	33°56'S, 18°51'E	42

Diploid (2n) numbers are indicated for each specimen. Field numbers refer to the CBGP Asian rodent collection (Montpellier, France). * The *R. norvegicus* specimen was obtained through the pet trade and as a consequence its precise origin is unknown. / = not applicable.

Satellite probe isolation and sequencing

PCR primers specific to the satellite I DNA family (sat I DNA) were designed on the basis of sequence data available for *R. norvegicus* sat I DNA in the NCBI database (370bp, acc. no: V01570 J00784). These are referred to as primer pair A in this text (see Table 4.2). An additional primer set (primer pair B) was designed from satellite sequence isolated from the template generated from primer pair A (Table 4.2). PCR amplification of

gDNA was performed in a 50 µl final volume using standard methods. Cycling parameters entailed an initial denaturing step of 94^oC for 3 min followed by 25 cycles of: 94^oC for 45 s, 55/62^oC (see Table 4.2) for 45 s and 72^oC for 1 min 30 s. A final extension of 72^oC for 10 min completed the programme. The 50 µl reaction mixture contained 100-300 ng gDNA, 5 µl buffer (10X), 3 µl MgCl$_2$ (25 mM), 2 µl dNTPs (2.5 mM), 5 µl of each forward and reverse primer (10 µM), 1 µl Taq (5 U). The remaining volume was made up with dH$_2$O. The mixture used for labelling with either digoxigen-11-dUTP (Roche) or biotin-16-dUTP (Roche) contained 5 µl buffer (10X), 2 µl MgCl$_2$ (25 mM), 5 µl dACG (2.5 mM), 1.5 µl dUTP-biotin or DIG (1 mM), 5 µl of each forward and reverse primer (10 µM), 1 µl Taq (5 U), 5 µl DNA (PCR product of the first amplification). This was made up to a final volume of 50 µl with dH$_2$O. The PCR programme selected was the same as for the first round of amplification.

Table 4.2. Primer sequences and the optimal primer annealing temperatures used to isolate satellite sequences from genomic DNA of species within the *Rattus s. l.* complex and *H. delacouri*.

Template		Primer sequence	Annealing Tm
(A) Satellite I	Forward	5'-ATTCAGGCTGCCTCTTGTGT-3'	62^oC
	Reverse	5'-GGATTCTCAAGTTGCGACG -3'	
(B) Satellite I	Forward	5'-TCCCAGTAGCCTGCTCTTGT-3'	55^oC
	Reverse	5'-TCAGTTCGTTAAAACGTTGCTC-3'	

Amplified products were electrophoresised in 1% agarose gels, excised and purified using the Wizard SV Gel and PCR Clean-up System (Promega) following manufacturer's recommendations. In instances where multiple bands were present on the agarose gel (i.e., *N. fulvescens* in Figure 4.1, see Results), which may indicate multiple monomers of the same repeat, fragments were excised separately and tested in autologous painting experiments. Since these resulted in identical hybridization patterns only a single purified fragment was selected for sequencing.

101

All cycle-sequencing reactions were performed using BigDye Chemistry and products were analysed on an automated sequencer (AB 3100, Applied Biosystems). Nucleotide sequences were edited and aligned using ClustalW Multiple Alignment in Bioedit v. 7.0.9.0 (Hall, 1999). DNA sequences were analysed using BLASTN searches to investigate possible sequence identity in EMBL and Genbank. Information on the size and number of tandem repeats present in the isolated nucleotide sequences was done using Tandem Repeat Finder (TRF) v 4.00 (Benson, 1999). In addition, Dot-matrix homology analysis was performed using Dotter v 3.0 (Sonnhammer and Durbin, 1995) in instances where TRF analysis was unsuccessful (no tandem repeats identified).

Telomeric probe isolation

A telomeric probe containing the repeat motif $(TTAGGG)_n$ was constructed and biotin-labeled by PCR as described by Ijdo *et al.* (1991) with minor modifications. PCR was performed in the absence of DNA template (primers serve as both template and primer) using primers $(CCCTAA)_5$ and $(TTAGGG)_5$ at 94°C for 1 min; 55°C for 30 s; 72°C for 1 min (10 cycles); 94°C for 1 min; 60°C for 30 s; 72°C for 1 min 30 s (30 cycles), followed by final extension of $72^{\circ}C$ for 5 min. The 50 μl reaction mixture contained 5 μl buffer (10X), 3 μl $MgCl_2$ (25 mM), 4 μl dNTPs (2.5 mM), 1 μl of each primer (0.1 μM), 1 μl Taq (5 U), 35 μl dH_2O. The mixture used for labelling with either biotin-16-dUTP (Roche) or digoxigen-11-dUTP (Roche) contained 5 μl buffer (10X), 3 μl $MgCl_2$ (25 mM), 5 μl dACGs (2.5 mM), 1.5 μl dUTP-DIG or biotin (1 mM), 1 μl of each primer (0.1 μM), 1 μl Taq (5 U), 5 μl DNA (PCR product of the first amplification), 27.5 μl dH_2O. The specifications of the programme used for the second amplification were: 94°C for 4 min; followed by 30 cycles of $94^{\circ}C$ for 1 min, 60°C for 45 s; 72°C for 30 s; and a final extension of $72^{\circ}C$ for 10 min.

Detection of Nucleolar Organiser Regions (NORs)

The distribution of NORs was investigated using silver staining following Goodpasture and Bloom (1975) and Barch (1997). To facilitate the identification of chromosomes bearing NORs, silver staining was carried out on previously DAPI-banded slides.

Fluorescence *in situ* hybridization (FISH) for satellite and telomere detection

A total of 100-150 ng of biotin- or digoxigen-dUTP labelled probe was precipitated together with 50 ng of salmon sperm DNA in three volumes of 100% ethanol (-80°C for minimum 1 hr or at -20°C overnight). After 25 min centrifugation at 13000 rpm, supernatant was removed and the pellet was dried for 30 min at 37°C and resuspended in 15 µl hybridization buffer (50% deionised formamide, 10% dextran sulphate, 2x SSC, 0.5 M phosphate buffer, pH 7.3).

Slides were heat-aged overnight at 65°C. Dry slides were placed in a 1x PBS solution (2 x 5 min) before fixation in 1% Formaldehyde solution in 1x PBS for 20 min, and washed in 1x PBS (2 x 5 min). Slides were dehydrated through ethanol series (70, 90 and 100%, 5 min each), denatured by incubation in 70% formamide/30% 2x SSC solution at 65°C for 30 s, and dehydrated through ethanol series (70, 90 and 100%, 5 min each). Hereafter, preannealed probes (denatured at 80°C for 10 min and immediately placed on ice) were applied to metaphase preparations, cover-slipped and sealed with rubber cement. Hybridization took place in humid chamber at 37°C for two nights. Following the hybridization period, post-hybridization washes were made up of two washes in formamide 50%/2x SSC for 5 min each, and two 5 min washes in 2x SSC followed by a final rinse in 4XT (4x SSC, 0.05% Tween 20) for 5 min. All post-hybridization washes were conducted at 42°C except for 4XT (room temperature). Fluorochromes were used to visualize labelled

103

paints. These were FITC-conjugated avidin (Vector) for biotin-dUTP, and anti-digoxigenin-rhodamine (Roche) for digoxigen-dUTP. The fluorochrome detection solution comprised 4x SSC/relevant antibody in a 200 µl final volume at 37°C for 45 min-1 hr. Slides were subsequently washed thrice in 4XT at room temperature, counterstained and mounted using an antifade solution containing DAPI (Vectashield). The images were captured with a CCD camera coupled to an Olympus BX60 fluorescence microscope and analysed using Genus Imaging Software (Applied Imaging). Signals were assigned to specific chromosomes according to morphology, size and DAPI banding.

RESULTS

Satellite sequence analysis

Isolation of satellite I DNA sequence and analysis

Satellite I DNA primers (Table 4.2) were successfully employed in the PCR amplification of the sat I DNA sequence from the gDNA of the species under investigation (examples of isolated PCR products visualized on an agarose gel is shown in Figure 4.1).
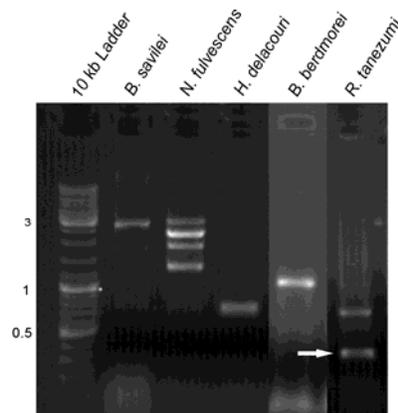
Figure 4.1. Examples of the PCR amplification of genomic DNA using primers designed for *R. norvegicus* sat I DNA. Arrow indicates the expected fragment size of rat satellite I DNA (370 bp) in *R. tanezumi*.

Sequences were obtained from the amplified PCR products for all species of *Rattus s. s.* (i.e., *R. exulans*, *R. losea* and *R. tanezumi*) and for *N. fulvescens*, *B. savilei*, *B. berdmorei* and *H. delacouri*. Unfortunately, sequencing of the PCR products from *M. surifer*, *L. edwardsi* and *B. bowersi* proved problematic and these species had to be excluded from further sequence analysis. No *R. norvegicus* sat I DNA sequence homology was found with *B. savilei*, *H. delacouri*, *N. fulvescens* and *B. berdmorei* following BLASTN analysis, nor in pair-wise comparisons among these non-*Rattus s. s.* species (Figure 4.2a). There was, however, significant homology between the *R. norvegicus* sat I DNA sequence (acc. no: V01570 J00784) available on Genbank and those of *R. exulans* (97%), *R. losea* (91%) and *R. tanezumi* (90%) indicating that the satDNA is significantly conserved within *Rattus* (Figure 4.2b).

(a)

```
              10        20        30        40        50        60        70        80        90        100
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
N. fulvescens  ------------------------GAGGGCATCGGATCCCATTACAGATGGTTGTGAGCCAAAATGTGGTTGCTGGGAATTGAACTCAGGACCTCTGGAA
B. savilei     ----------------------------------------------------------------------------------------------------
H. delacouri   AGAGAGGGGANTGCCACTCCCCTCTGCAGGNCTGCATGACATACCTNTCCTCTATGGGCTGTTATACTCTACAGATTCCCCTGCCCTTTATCTTCTTCAG
B. berdmorei   TAGACTTGTGCCAAGACCAGGGCAGGTCAATGTCTAAAAGTTCATGGCCGACCACACCCTGGCCCGTATACAATGTNAGCTTGNNCCTGGNACACTGCAC

              110       120       130       140       150       160       170       180       190       200
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
N. fulvescens  GAATAGTCAGTGTTCGTCTGAGCCATCTCTCCAGACCATAT--TCCCTTTCTTGATGTGGACTTTTTAAAACAGTATCAACTTTAGCCCTGACCCCCTGT
B. savilei     ----------------------------------------------------------------------------------------------------
H. delacouri   GCATACCAGGACACCGATGACTCAGGATGGTTTTTACCCGT--GNAGCNGGCTGACACTGAAGACTGACATGCATGTCTGCCCTGTGCCCTCTTCCTCAC
B. berdmorei   CTGTCCCCAGNCNACCATGATCCANAACAAACCATCCCAGCTGGGGAGTGGCNGAGNTTGTGCTTTCAGATGTTGATCTGCAAGAGAAAAGAGCTGAGGN

              210       220       230       240       250       260       270       280       290       300
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
N. fulvescens  TCCATACCCGGGATCTGCAGTGGGSCCGTCCCTCACCTTG---TSTGTGTACAGATCTKTGK---CAACTCGGKACACATTGTCCCTCCCCTCCTGGCGGK-
B. savilei     GACAGAAGCAGATTCW-TGATGATGYCGGGGCTCAACAGC---TCTACG--CTGTCCAGCCT---CCGCTGCAGGCGAGCCAGCTCCTGCCTC-------AG-
H. delacouri   AGAAACCACAAGAAAGGCCCTTGTGCCCATGCCCCTCATCCCTCTGCTTTCTGACTAACCTGGTTCTTCCATGTGCTGGGCCTCCTGATTTATGAGCAGA
B. berdmorei   -CTNCNGAGGTGATAGCAAACTCTGCCTAAACTCCTAGTGT-TGGACGCTACAACTGAGGGGTCTACCTYGATTGGAGAGCTAGNAAGCTAAAAGGCAAA

              310       320       330       340       350       360       370       380       390       400
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
N. fulvescens  -----SAGSATCAGGCTGAGCTCCTGCCTGACCTCCTCT---AGAAGCCATG-GGCCATNCAGATGTCCCTGAAGATCGGTGTACGC----CACACTCTGAGT
B. savilei     ----CTGCTGCCCGCTGAACCGCTCCCGTGCCT--TCC---GGATATCCTG-ACGAATG---GTCTCCCGGAAG------TAGCCGC----TGATGCCCAAGA
H. delacouri   CTTCCCACAACACTCGTCTGCGTTTGCACACTCAATTAAAGAAAATTTTA-GGCATAGATGAAAACTCAAGNTATAAACAGCTACATATGGAATTTAAAA
B. berdmorei   --TACCCATGCACACCT-TCTGGGGGTATCCTTGATCNGATGNTCTNANATGGGNAAAGTTAGTGCTAAGGAGAATGGCCTNTTCNGTGGCGGCNCGGAG

              410       420       430       440       450
         ....|....|....|....|....|....|....|....|....|....|
N. fulvescens  TTCTCACCTGCAGGGGGCAGGTGAG-G------AGAGGTGAGAAGGGAGAAGTGGAG
B. savilei     CCAGGAGAGAGAGAGAGAGAGAGAGAG------AGAGAGAGAGAGAGAGAGAGAGAG
H. delacouri   TTAAAAGGCAGAATGACCACAAGGCTG------TGACATAAACACACTGGAATAAGT
B. berdmorei   AGAGGAGATGGTGGGATGGGTGGTGGGCCCCGTTGGCCGTNACTNGCNGGGGGAGAG
```

(b)

```
                 10        20        30        40        50        60        70        80
R. norvegicus  ATTTGGTGAGCCGAATCTGAGTAACATGGTGAATTCGGTTAGTTCCCAGTGGGCCGCAGTAGTAATAAGTGGACTGTTCG
R. losea       ..G.C...T....T....C.A....T.....C.........A.A.....................T.......G.
R. tanezumi    ....C..............N......T.............A.A.........................G.
R. exulans     ...........................................A.A..........................

                 90       100       110       120       130       140       150       160
R. norvegicus  GGATATAGAAAAATTTTCACGAATGGAATCAAAATTTCTAGGCGAATTCAGTGAGTTCCCAATGGGACATGCTTGTATGA
R. losea       ............................................................A.............A....
R. tanezumi    ...............................................................C....
R. exulans     ..................................................................

                170       180       190       200       210       220       230
R. norvegicus  GTTTGCTTTTCAATATGGAGCAACATTTTGAGGAACGGAATCTAAGTATCATGGTGAATTCAGTAGGTTCCCAATAGG
R. losea       ............C..T...........A.C........C....A...................G....
R. tanezumi    ........C....C..AT......T...AA.C......T......T...G.......TC.T.......G....
R. exulans     ...................N..........NN.........................N.......
```

(c)

```
                 10        20        30        40        50
R. norvegicus  T GGT GAAT T CAGT TAGT T CCCAGT GGGACA T GCT T GT AAT AAGT T TGCT T
R. exulans     . . . . . . . . . . A . . . . . . . A . A . . . . . CA . . A . . . . . . . . . GGA . . G
R. losea       . . . . . . . . . . A . . . . . . . A . A . . C . . CAG . A . . . . . . . G . A . . G
R. tanezumi    . . . . . . . . . G . . . . . . . A . A . . C . GCAG . A . . . . . . . GGA . . G

                 60        70        80        90
R. norvegicus  T T CAAGAT AT AGCAACA T T T T GAGGAACGGAAT CT AAGT T T C
R. exulans     . . . . . . . . G . . A . . A . . . . . C . C . . . . . . . . . . . . . A . .
R. losea       . . . . . C . . . . A . . A . . . . . A . C . . . . . C . A . . AAA . .
R. tanezumi    . . GGG . . . . . . . A . . A . . . . . C . C . . . T . . . . . . A . . A . . . .
```

Figure 4.2. **(a)** Sequences from *B. savilei*, *H. delacouri*, *N. fulvescens* and *B. berdmorei* showing the lack of significant homology in species outside of *Rattus s. s*. and **(b)** multiple alignment of four *Rattus* sat I DNA sequences (238 bp) displaying significant homology by ClustalW. **(c)** Multiple alignment of tandem repeat sequences identified in the four *Rattus* taxa (92 bp) by TRF v 4.00. Dots indicate identity with the *R. norvegicus* sat I DNA sequence in (b) and (c).

No tandem repeats were identified by the TRF v 4.00 analysis in sequences of the non-*Rattus s. s.* species (*B. savilei*, *H. delacouri*, *N. fulvescens* and *B. berdmorei*). In contrast, a largely conserved 92 bp reiterated subunit was identified in *Rattus s. s*. species. This was repeated twice in *R. exulans*, *R. losea*, and *R. tanezumi* sat I DNA and four times in *R. norvegicus* (determined by TRF v 4.00, see Discussion). Pair-wise comparisons between *R. norvegicus* 92 bp repeat subunit and *Rattus s. s*. 92 bp subunits of this study indicated 82% nucleotide similarity with *R. exulans*, 76% similarity with *R. losea* and 75% sequence similarity with *R. tanezumi* (see Figure 4.2c).

Given the apparent lack of sequence homology between *B. savilei*, *H. delacouri*, *N. fulvescens* and *B. berdmorei* and the *R. norvegicus* sat I DNA sequence using BLASTN

analysis, a dot-matrix approach was used to determine the presence of repetitive nucleotides within these sequences. Dot-matrix approaches are frequently used to rapidly identify and visualize repeats within a single sequence. This involves comparisons of the sequence against itself, represented by the main diagonal in the output, with regions of similarity (presence of repeats) within the sequence shown by matches (diagonal lines appearing on the plot) away from the main diagonal (Mount, 2007). The dot-matrix analysis indicates little or no internal substructure in these sequences (Figure 4.3). Some complementary repetitive nucleotides/strands are, however, evident in *B. savilei* and to a lesser extent *H. delacouri*.
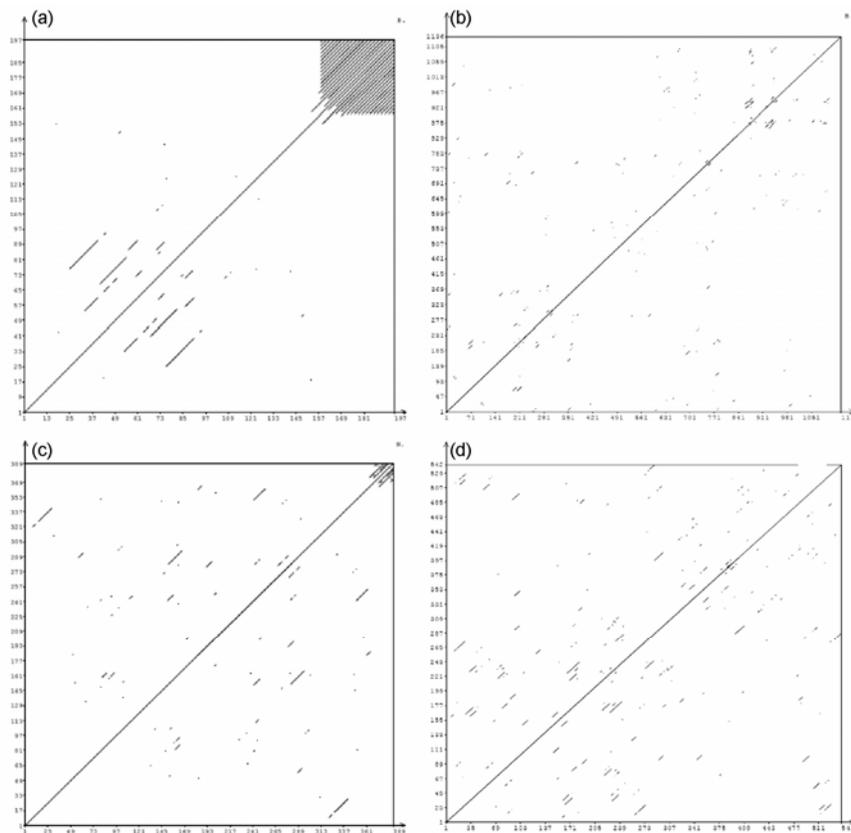
Figure 4.3. Dot-matrix analysis of the sequences isolated in **(a)** *B. savilei,* **(b)** *B. berdmorei,* **(c)** *H. delacouri,* **(d)** and *N. fulvescens* with each sequence plotted against itself. A window size of 21 with a stringency of 50 was employed. Lines appearing on the plot indicate regions of similarity.

Fluorescent in situ *hybridization of satellite I DNA*

Physical mapping of the labelled PCR products generated by the sat I DNA primers A and B was conducted using FISH, examples of which are shown in Figure 4.4. The labelled PCR products were hybridized to species of origin (autologous painting) as well as to the chromosomes of the other species (heterologous painting). Hybridization signals obtained in *Rattus s. s.* and non-*Rattus s. s.* species displayed very different outcomes. Hybridization patterns in *N. fulvescens*, *B. savilei*, *H. delacouri* and *B. berdmorei* (non-*Rattus s. s.* species) showed interspersed patterns with both autologous as well heterologous painting experiments (Figure 4.4 panel e to h). No centromeric signal was detected in non-*Rattus s. s.* while the heterochromatic Y-chromosomes of *B. bowersi*, *L. edwardsi*, *M. surifer* and *H. delacouri* were negative for signal following both autologous and heterologous painting. The same dispersed hybridization patterns were detected when the isolated sequences from these non-*Rattus s. s.* species was hybridized in heterologous painting experiments to the chromosomes of species of *Rattus s. s.* (*R. exulans, R. losea, R. tanezumi* and *R. norvegicus,* Figure 4.4 panel i).

In contrast, hybridization patterns of sequences isolated from *R. exulans, R. losea* and *R. tanezumi* were invariably centromeric following autologous painting experiments (i.e., Figure 4.4 panel a). Similar patterns (centromeric) were identified with heterologous painting experiments using *R. losea* isolated satellite sequence as representative of the *Rattus s. s.* group against chromosomes of *Rattus s. s.* (i.e., Figure 4.4 panel b to d). No hybridization signal was detected in heterologous painting experiments using *R. losea* satellite against the chromosomes of the non-*Rattus s. s.* species (*B. savilei, B. bowersi, B. berdmorei, N. fulvescens, L. edwardsi, M. surifer* and *H. delacouri*). In summary, therefore, the hybridization patterns detected on the chromosomes of the various species are (1)

108

centromeric when using probes from species within *Rattus s. s.* and (2), interspersed in instances when using non-*Rattus s. s.* species as the probe source. Signal intensities varied between the two hybridization patterns detected but the intensity was generally higher at centromeres than at the interspersed sites of hybridization (Figure 4.4).
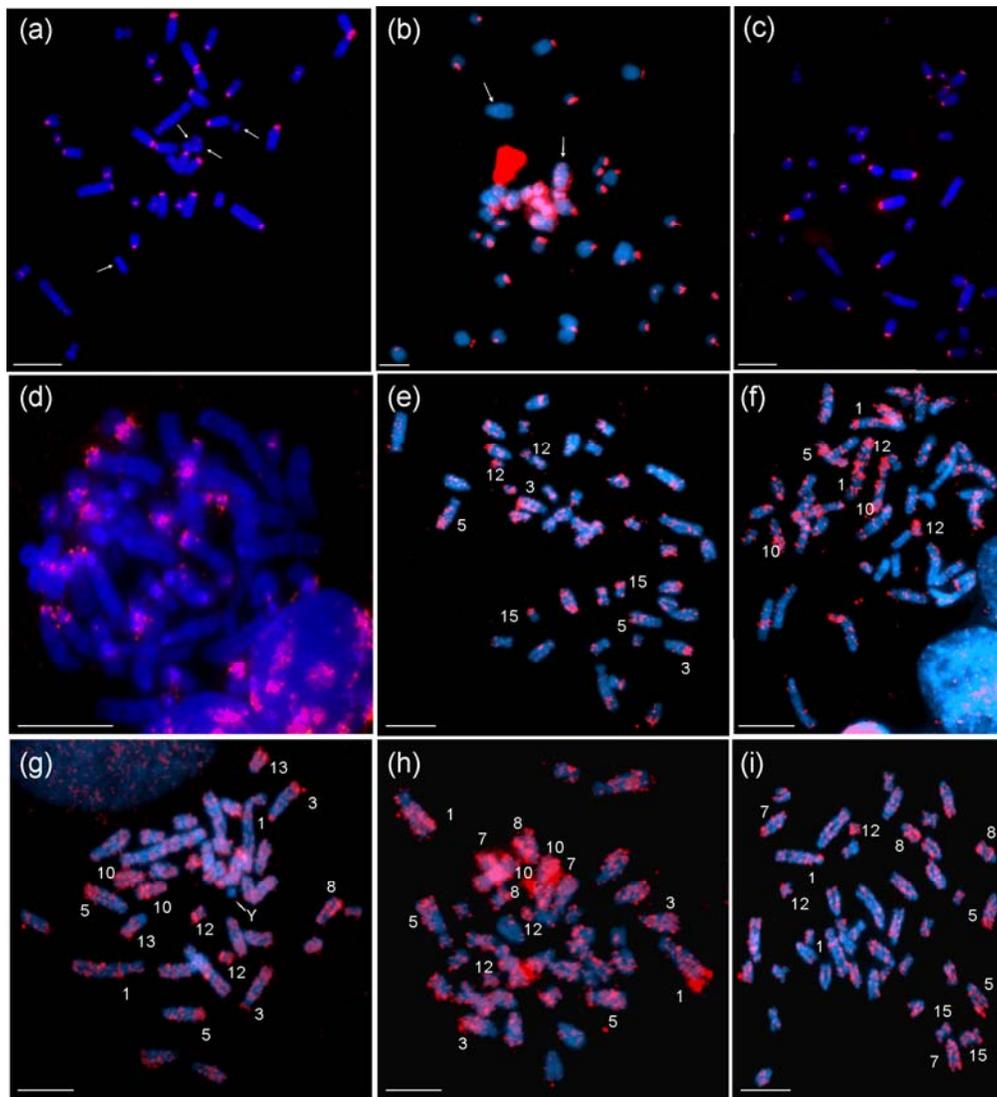
Figure 4.4. Examples of FISH experiments showing the two types of hybridization patterns detected using isolated satellite sequence probes. Centromeric hybridization patterns were obtained from: **(a)** *R. losea* satellite painted to a *R. losea* metaphase spread, **(b)** *R. losea* to *R. exulans* **(c)**, *R. losea* to *R. norvegicus*, and **(d)** *R. losea* to *R. tanezumi*. Examples of interspersed hybridization signals were detected when **(e)** *N. fulvescens* amplicon was painted to *B. berdmorei*, **(f)** *N. fulvescens* to *N. fulvescens*, **(g)** *H. delacouri* to *L. edwardsi*, **(h)** *B. berdmorei* to *B. berdmorei*, and **(i)** *B. berdmorei* to *R. losea*. Chromosome numbers correspond to those of *R. norvegicus* (RNO) thus facilitating comparisons of the patterns

detected in the various rodent genomes. Arrows on panels (a) and (b) indicate an absence of signal on the orthologs of RNO 12 and 13 (panel a) and RNO 1 (panel b) in *R. losea* and *R. exulans* respectively. Note that poor quality of bone marrow preparations of *R. tanezumi* (panel d) precluded the identification of *R. norvegicus* orthologs in this species. All metaphases in panel (c) display strong to weak centromeric signal. Scale bar = 10 μm.

Telomeric sequence distribution

As expected, the telomeric $(TTAGGG)_n$ probe hybridized to the termini of all the chromosomes in the species investigated. Examples of telomeric FISH results are shown in Figure 4.5. Other than the strongly fluorescent pericentromeric regions of HDE 3, 10, 11, 12, 13, 16, 19 and 22 in *H. delacouri*, and MSU 4 in *M. surifer*, no consistent interstitial signals were detected. Moreover no interstitial telomeric signals were evident at the junction (IHB) of the X-autosome translocations identified in *H. delacouri* and *B. savilei* (see panels a and e of Figure 4.5), or at the sites of the Robertsonian fusions in *B. bowersi* and *B. berdmorei* which result from the head-to-head fusion of acrocentrics chromosomes (see panel b in Figure 4.5).
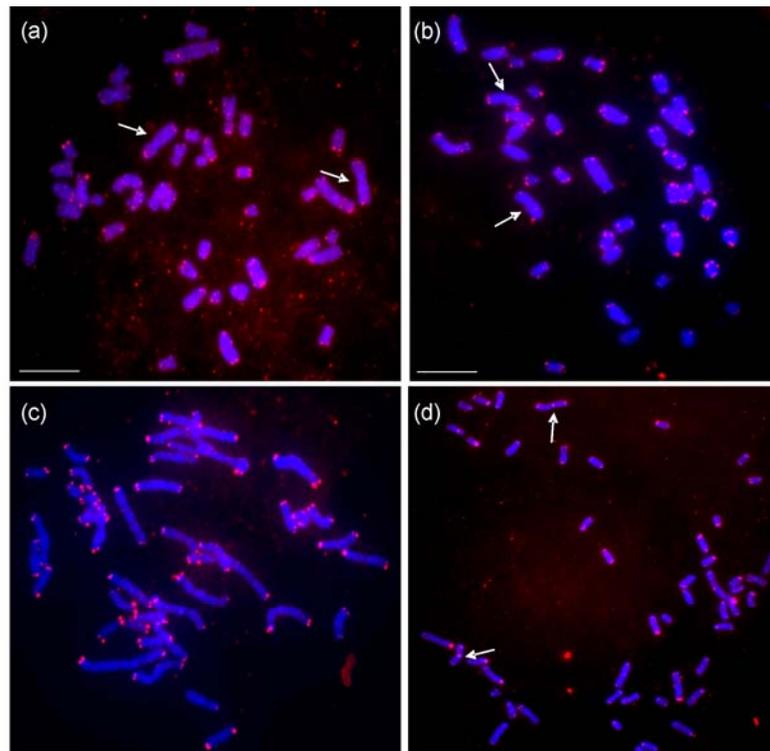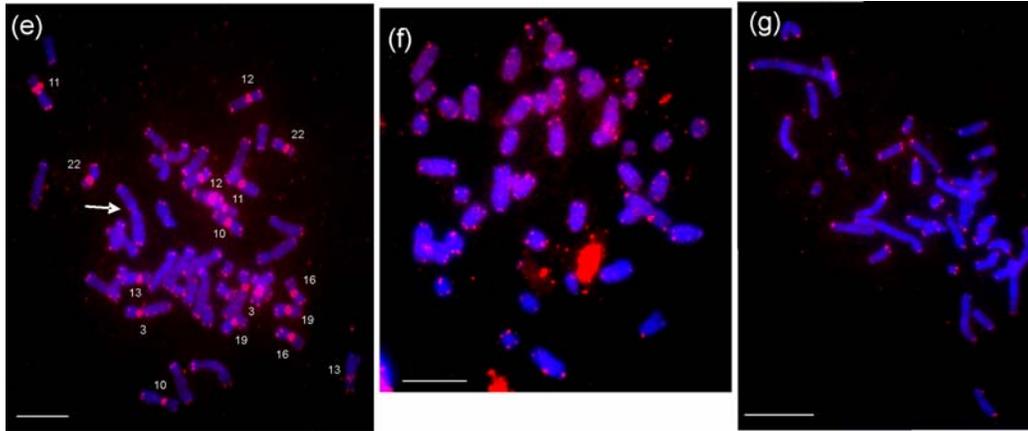
Figure 4.5. Distribution of telomeric repeats $(TTAGGG)_n$ on metaphase spreads of **(a)** *B. savilei*, **(b)** *B. bowersi* (representative of *Berylmys*), **(c)** *R. losea* (representative of *Rattus s. s.*), **(d)** *M. surifer*, **(e)** *H. delacouri*, **(f)** *N. fulvescens* and **(g)** *L. edwardsi*. The numbering in (e) corresponds to the chromosomes of *H. delacouri* (Chapter 2). Note the absence of interstitial telomeric signal at the breakpoint of the fusion between RNO X and 11 in *B. savilei* and *H. delacouri*, and between RNO 9 and 11 in *B. bowersi* (panels a, b and e, white arrow) as well as the presence of telomere-like sequences in the pericentromeric regions of *H. delacouri* (panel e) and *M. surifer* (panel d, white arrow). Scale bar = 10 µm.

Nucleolar Organiser Regions (NORs)

Detection of NORs by Ag-staining

Silver staining detected a variable number of NOR-bearing autosomal chromosome pairs (Table 4.3 and Figure 4.6); two NOR pairs were identified in *B. savilei*, seven in *H. delacouri* and eight in *M. surifer*. *Niviventer fulvescens* and *L. edwardsi* both show five NORs bearing chromosome pairs, and both *B. bowersi* and *B. berdmorei* are characterised by four pairs. *Rattus exulans*, *R. losea* and *R. tanezumi* all contained three NOR bearing chromosome pairs, a finding that is consistent with data on *R. norvegicus* (Szabo *et al.,* 1978; Kodama *et al.,* 1980; Sasaki *et al.,* 1986; Cavagna *et al.,* 2002, see Table 4.3).

111

Table 4.3. Number of chromosome pairs showing nucleolar organizer regions (NORs) in the ten species analysed in the present study (numbering according to their respective G-banded karyotypes; see Chapter 2), including *R. norvegicus* and *R. rattus* from published data (see #, based on silver staining, Cavagna *et al.,* 2002). * = Correspondence to *R. norvegicus* orthologous chromosomes identified by cross-species chromosome painting using *R. norvegicus* paints (Chapter 3) and / = not applicable.

	No. NORs pairs	Chromosome no.	Respective *R. norvegicus* ortholog *
R. exulans (2n = 42, XX)	3	REX 3, 11 and 12	3, 11 and 12
R. losea (2n = 42, XX)	3	RLO 3, 11 and 12	3, 11 and 12
R. tanezumi (2n = 42, XX)	3	RTA 3, 11 and 12	3, 11 and 12
B. savilei (2n = 43, XX)	2	BSA 10 and 20	10 and 12
B. bowersi (2n = 40, XY)	4	BBO 4, 7, 10 and 9	3, 6, 8 and 10
B. berdmorei (2n = 41, XX)	4	BBE 4, 7, 10 and 9	3, 6, 8 and 10
L. edwardsi (2n = 42, XY)	5	LED 2, 3, 4, 10 and 20	1, 4, 3, 10 and 12
N. fulvescens (2n = 46, XX)	5	NFU 1, 6, 8, 9 and 10	1q dist, 6, 8, 9 and 10
M. surifer (2n = 52, XY)	8	MSU 1, 2, 4, 6, 7, 14, 17 and 19	1q dist, 3, 2 dist, 8, 9, 10, 5 dist and 6 prox
H. delacouri (2n = 48, XY)	7	HDE 1, 3, 6, 7, 8, 10 and 13	4, 3, (20+16a+15), 6, 8, 18 and 19
R. norvegicus (2n = 42)#	3	RNO 3, 11 and 12	/
R. rattus (2n = 38)#	2	RRA 5, 8 and 16	RNO 3, 8 and 12

There is, however, a strong conservation of NOR location among species. For example, the NOR on RNO 3 orthologs is present in all species except *B. savilei* and *N. fulvescens*, and the NOR on RNO 10 is conserved in all Rattini except for those in *Rattus s. s.* (*R. exulans, R. losea, R. tanezumi, R. norvegicus* and *R. rattus*). The remaining NORs loci are present in a minimum of two species, while those on RNO 5 represent an autapomorphy in *M. surifer*, and as do those on RNO 18 and 19 orthologs in *H. delacouri*. NOR locus on RNO 11 appears to be characteristic of the *Rattus* group but, importantly, is absent in *R. rattus*. NOR sites were not detected at the interstitial heterochromatic blocks of the X-autosome translocations identified in *H. delacouri* and *B. savilei* (see panels b and d of Figure 4.6). However, two NOR loci in *M. surifer* (MSU 17 and 19) correspond to fissioned chromosome products of RNO 5 and RNO 6 (Chapter 3).
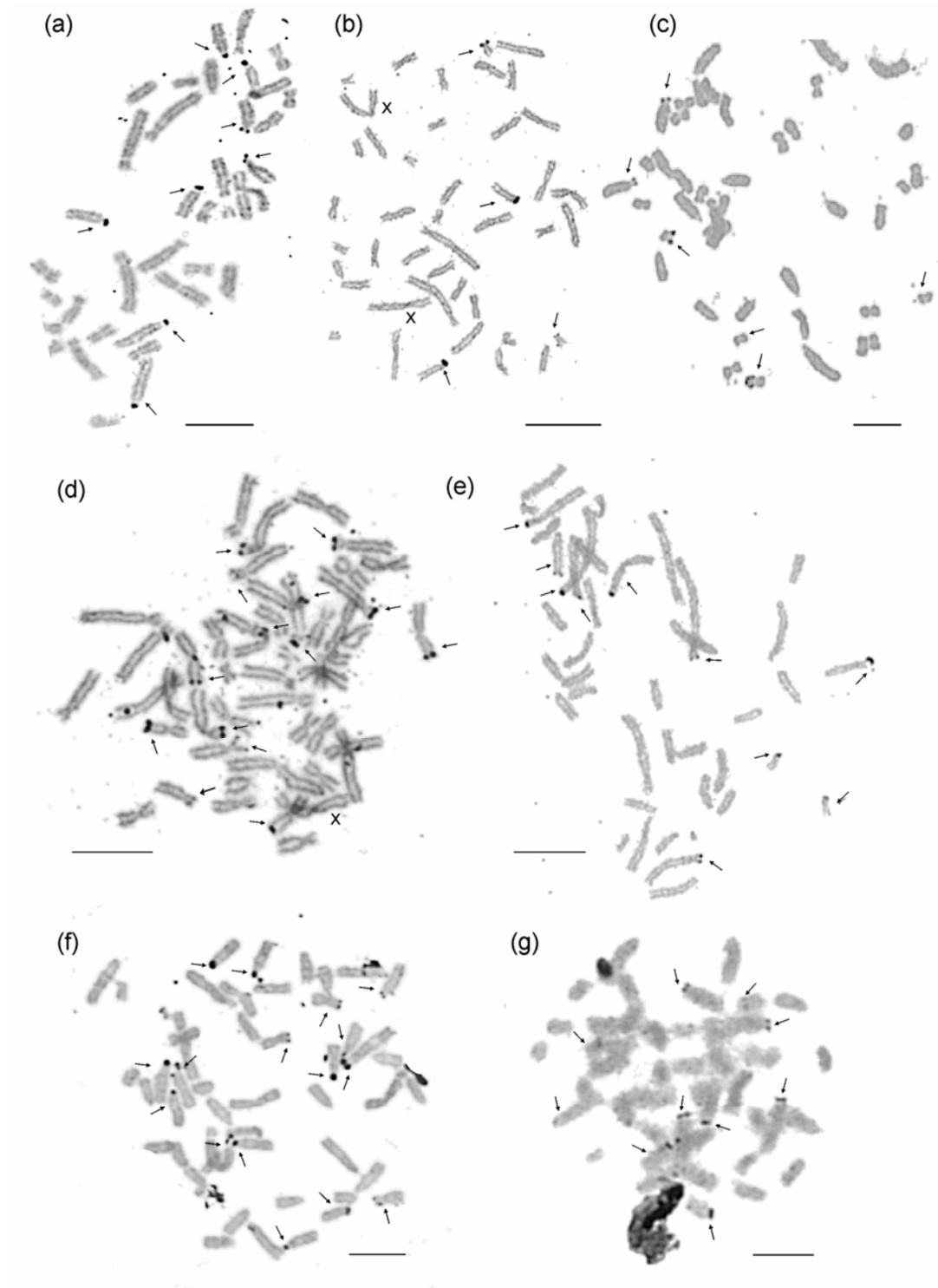
Figure 4.6. Nucleolar organizer regions (NORs) in **(a)** *B. berdmorei,* n = 8; **(b)** *B. savilei*, n = 4; **(c)** *R. losea*, n = 6; **(d)** *H. delacouri,* n = 14; **(e)** *L. edwardsi*, n = 10; **(f)** *M. surifer*, n = 16; **(g)** *N. fulvescens,* n = 10. The X chromosomes in *B. savilei* and *H. delacouri* (panels b and d) that are involved in X-autosome translocations are indicated.

DISCUSSION

Satellite I DNA repeat analysis and distribution

Satellite I DNA family variants were isolated from gDNA of six species of *Rattus s. l.* and *H. delacouri* using two pairs of primers designed from *R. norvegicus* sat I DNA family consensus sequence and examined from two different perspectives—(i) their molecular composition, and (ii) their physical localization.

Molecular composition

Satellite I DNA sequences were successfully isolated from seven species (sequencing was unsuccessful for *B. bowersi, L. edwardsi* and *M. surifer*). Analysis of the sequences obtained for the non-*Rattus s. s.* species (*H. delacouri*, *N. fulvescens*, *B. savilei*, and *B. berdmorei*) showed no sequence homology to *R. norvegicus* sat I DNA suggesting possible mispriming. There was also a lack of homology among the *H. delacouri, N. fulvescens, B. savilei,* and *B. berdmorei* sequences (see Figure 4.2a) in spite of the fact that strong, single band PCR products were obtained for all species (see Figure 4.1). This finding held for both primer sets used in the investigation. BLASTN searches of the *H. delacouri, N. fulvescens, B. savilei,* and *B. berdmorei* sequences failed to returned matches with Genbank satDNA sequences. This suggests that they may in fact be new species-specific satellite families, as it is unlikely they are from the same family considering that they display no homology to each other. This explanation would be consistent with the process of concerted evolution (amplification of new variants during speciation) that may lead to the generation of species-specific satDNA (Hamilton *et al.,* 1992; Jobse *et al.,* 1995; Nijman and Lenstra, 2001; Adega *et al.,* 2008; Louzada *et al.,* 2008). However, there was only some indication (although not compelling) from the dot-matrix of short repeat elements within *B. savilei* and to lesser extent *H. delacouri* and these are different from those of the sat I DNA. None were

114

identified within *N. fulvescens* and *B. berdmorei* which is inconsistent with the presence of satellite DNA (repetitions of short DNA sequences). As a consequence it cannot be excluded these findings may be a technical artefact of the investigation or, alternatively, a reflection of some other repeat family such as long interspersed nuclear elements (LINE-1; see Waters *et al.*, 2004 and below).

In contrast to these findings, there was a high degree of *R. norvegicus* sat I DNA sequence conservation (>90%) within *Rattus s. s.* (*R. exulans*, *R. losea* and *R. tanezumi*). This was corroborated by TRF analysis that showed the presence of a tandemly repeated 92 bp subunit in *R. exulans, R. tanezumi* and *R. losea* (Figure 4.2c). This corresponds to the previously described 92 bp repeat subunit for *R. norvegicus* sat I DNA by Pech *et al.* (1979) (>75 % pair-wise sequence similarity), however, the number of copies of this subunit differed (*R. norvegicus* has four copies and *R. exulans, R. tanezumi* and *R. losea*, two copies each). As the primers used in the present investigation were not designed from regions flanking the *R. norvegicus* sat I DNA, but rather from sequence internal to it (thus ensuring amplification of the correct satellite sequence), the difference in copy number may be related to this (final alignments of *Rattus* satDNA sequences included 238 nucleotides).

Importantly, although *B. savilei* is considered to be congeneric with *Rattus* in some molecular phylogenies (Usdin *et al.*, 1995; Verneau *et al.*, 1997, 1998), the *B. savilei* sequence showed no meaningful homology with rat sat I DNA (also supported by FISH analysis, see below). These findings, together with the conservation of satellite sequence within *Rattus s. s.*, suggest that the rat sat I DNA family is *Rattus*-specific. However, it should be noted that this satellite family may occur in *Cricetus cricetus* (Adega *et al.* unpubl.), suggesting that broader taxonomic sampling is required. Should the taxon-specificity of the rat sat I DNA family be correct, it provides an opportunity to date the

appearance of this repeat family in this group of rodents. Molecular clock estimates place the divergence of *Rattus s. l.* at ~8.4 mya, the *Rattus s. s., Berylmys* and *Bandicota* divergence at ~5.4 mya (Lecompte *et al.,* 2008), and the deepest divergence within *Rattus* at ~3.5 mya (Robins *et al.,* 2007, 2008). *Rattus norvegicus* is thought to have diverged ~2.9 mya (Robins *et al.,* 2007, 2008). It seems likely, therefore, that the rat sat I DNA sequence based on these results has a recent origin, and its presence in the common ancestor to *Rattus s. s*. at ~3.5 mya may reflect the minimum age of this satellite family.

Physical mapping of satellite I DNA

Cross-species *in situ* hybridization data support the findings of the molecular analysis, namely that the sat I DNA repeat is absent in non-*Rattus s. s.* species, but conserved within *Rattus s. s.*

(i) Physical mapping within non-*Rattus s. s*. species

The isolated non-*Rattus s. s.* sequences resulted in an interspersed FISH pattern (similar to the LINE-1 patterns detected in Waters *et al*., 2004) in both heterologous and autologous painting experiments. There was some indication from the dot-matrix of short repeat elements in *B. savilei* and *H. delacouri* PCR amplicons (Figure 4.3) that may be responsible for the interspersed nature of the FISH signal seen here. These results, however, must be treated with caution as dot-plot analysis simply gives a gross assessment of similarity within a sequence and low stringency settings were used (see Results). Additionally, FISH using *N. fulvescens* and *B. berdmorei* probes showed similar interspersed patterns and no significant repeat elements were identified in these two sequences in the dot-matrix (Figure 4.3 and 4.4). In conclusion, although there are reports indicating interspersed signals co-localized with constitutive heterochromatic regions that are considered hotspots for chromosome evolution (e.g., co-localization of CCR4/10sat DNA with C-positive interstitial

bands in *Peromyscus*; Louzada *et al.,* 2008), this did not hold in the present study and the dispersed nature of the FISH patterns observed herein are most likely non-specific, and possibly artifacts of the methods used.

(ii) Physical mapping within *Rattus s. s.*

The rat sat I DNA family is reported to concentrate in centromeric DNA (Sternes and Vig, 1995) although there appears to be some variability in its location in *Rattus s. s.* For example, all *R. norvegicus* chromosomes carry centromeric sat I DNA, while only eight pairs of chromosomes in *R. tanezumi* show this. In addition, within all four *Rattus* karyotypes there were several autosomes with reduced amounts of centromeric sat I DNA that manifested in the form of faint hybridization signals. This observation in itself is not surprising given that satDNA generally varies with regards to its abundance, sequence, and chromosomal distribution, at times even evident between closely related species (Charlesworth *et al.,* 1994; Adega *et al.,* 2008; Louzada *et al.,* 2008) as was observed here. Moreover, comparable findings have been reported in *Microtus* and *Arvicola* rodent karyotypes where several autosomes displayed reduced amounts of pericentromeric satDNA (Msat-160) (i.e., *M. savii, M. brachycercus, A. terrestris* and *A. sapidus*; Acosta *et al.,* 2010). Amplifications and contractions, as well as loss of satellite arrays in chromosomes of the karyotype, can alter the structure and distribution of a repeat sequence in pericentromeric heterochromatin (Acosta *et al.,* 2010). Possibly leading to the diminution even complete deletion of rat sat I DNA in the pericentromeric heterochromatin of *Rattus s. s* (i.e., reduction of chromosomes carrying centromeric sat I DNA in *R. tanezumi*).

The rat sat I DNA appears to have a relatively recent origin having persisted minimally for ~3.5 mya. If this holds, its localization at the centromeres in species of *Rattus s. s.* and its apparent absence in species outside of this assemblage may be consistent with the

feedback model of satellite evolution (Nijman and Lenstra, 2001). This evolutionary model proposes that three phases are present in satDNA evolution. The first is characterised by rapid contractions and expansions that leads to fluctuations in copy number. Mutational decay leads to the appearance of new sequence variants as well as preventing recombination of some repetitive units. This initiates the second phase where different sequence variants are deleted and/or amplified independently, this ongoing loss of heterogeneity leads to the final phase where the interactions between the repeat sequences are completed and the repeats become inactive (e.g., no further changes in copy number). Each phase is characterised by different levels of sequence identity where the initial phase is favoured by sequence homogeneity (Nijman and Lenstra, 2001; Slamovits and Rossi, 2002 for review). According to this model, the new rat sat I DNA family may be undergoing the initial phase of its life history, as the three *Rattus* satDNA sequences (*R. exulans, R. losea* and *R. tanezumi*) currently displays a high degree of overall sequence conservation (>90% similarity/sequence homogeneity) with *R. norvegicus*.

Distribution of telomeric sequences

As expected, telomeric signals were detected at the ends of the chromosomes in all nine *Rattus s. l.* species and *H. delacouri* (Figure 4.5), an observation that is consistent with reports that these structures are crucial for maintaining the stability of chromosomes (Bolzan and Bianchi, 2006 among others). Although, interstitial telomeric sequences (ITS) have been previously detected at sites of chromosomal rearrangements in a number of published studies (reviewed by Ruiz-Herrera *et al.,* 2008), this is not a universal observation (Garagna *et al.,* 1995; Fagundes *et al.,* 1997; Fagundes and Yonenaga-Yassuda, 1998; Garagna *et al.,* 2001; Andrades-Miranda *et al.,* 2002; Viera *et al.,* 2004; Gauthier *et al.,* 2010). The present study falls within the latter category since no het-ITS nor s-ITS were identified at the evolutionary

118

breakpoint sites of the fusions identified herein (X-autosome translocation in *H. delacouri* and *B. savilei*, and the rearrangements that gave rise to *B. berdmorei* and *B. bowersi* pair 3).

The absence of interstitial telomeric signals may result from one of the following: (1) telomeric sequences were lost during the fusion process, suggesting that the telomere sequences were located distally to the breakpoints, (2) the amount of telomeric sequence remaining at the fusion sites is insufficient to be detected at the resolution provided by FISH or (3), that the telomeric sequences have eroded following the fusion event (see for example Slijepcevic *et al.,* 1998; Dobigny *et al.,* 2003b; Ropiquet *et al.,* 2010). The absence of het-ITS in *Rattus s. l.* is most likely due to the removal of the telomere sequence during the fusion event rather than through decay as suggested by inference from data on other species. For example, het-ITS were not detected at the centromeric or pericentromeric regions of the most recent chromosome rearrangements to occur in *R. rattus*—the Rb fusion chromosomes RRA 1 and 4 (Cavagna *et al.,* 2002)—a species which is estimated to have diverged very recently from its sister, *R. tanezumi* (~0.4 mya; Robins *et al.,* 2008). Het-ITS and s-ITS have, on the other hand, been identified at the junctions of Robertsonian and tandem fusion translocations in *Taterillus* species of similar age (*T. tranieri* and *T. pygarus* are thought to have shared a common ancestor 0.4 mya, Dobigny *et al.,* 2003b). Similarly, and again in contrast to other species (Parish *et al.,* 2002; Dobigny *et al.,* 2003b, 2004b), there was no evidence of telomere co-localization with the C-positive IHB identified in the X-autosome translocations of *H. delacouri* and *B. savilei*. In summary therefore, it would seem probable that the loss of signal at rearrangement junctions in *Rattus s. l.* is a reflection of telomere sequences that were located distally to the breakpoints (see above) rather than the erosion of interstitial telomeric sequences.

Interestingly, and in contrast to the findings relating to rearrangement breakpoints, pericentromeric telomeric signals were noted in some *H. delacouri* chromosomes (HDE 3, 10, 11, 12, 13, 16, 19 and 22) as well as in one chromosome pair of *M. surifer* (MSU 4) (Figure 4.5). Since these sites do not correlate with evolutionary breakpoints, the presence of pericentromeric signal here may reflect differences in the sequence composition of the pericentromeric heterochromatin. Similar patterns have been observed in other vertebrate species (for example Moyzis *et al.,* 1988; Garagna *et al.,* 1997; Faravelli *et al.,* 1998; Metcalfe *et al.,* 2004; Santos *et al.,* 2004; Gilbert *et al.,* 2008) and molecular analyses of these pericentromeric ITSs have substantiated these reports (reviewed by Ruiz-Herrera *et al.,* 2008). In these instances ITS were shown to represent arrays of telomeric-like repeats (see Faravelli *et al.,* 2002). Moreover, the presence of strong fluorescent signals of pericentromeric repeats detected in *H. delacouri* probably simply reflects the greater quantity of C-positive material in HDE 3, 10, 11, 12, 13, 16, 19 and 22 compared to that of *M. surifer* (MSU 4) (see Chapter 2) and hence the increased copy number of telomeric-like repeats in these C-positive regions.

Nucleolar Organizer Regions

Reports documenting the species-specific chromosomal distribution of NORs have led researchers to use these chromosomal landmarks as cytogenetic markers for species discrimination (Gold and Li, 1994; Gallagher *et al.,* 1999; Mazurok *et al.,* 2001; Matsubara *et al.,* 2004; Nguyen *et al.,* 2008; Wang *et al.,* 2009). However, there are several important considerations (Dobigny *et al.,* 2004a) when using these data. First, intraspecific NOR variation is well known within mammals and small sample size can result in spurious conclusions. Examples of intraspecific variation extend to species that form the focus of the present study. For example, Yosida (1979), Sasaki *et al.* (1986) and Wang *et al.* (2003) have

documented intraspecific variation in NOR number and position in *Rattus* and *Niviventer*, and even in single strains of *R. norvegicus* and *N. confucianus* (Sasaki *et al.,* 1986; Wang *et al.,* 2003). Secondly, silver staining, the method followed in this study, may reveal non-ribosomal loci (i.e., in hedgehogs, Sanchez *et al.,* 1995 and gerbils, Dobigny *et al.,* 2002b) and of course will only detect active NORs. Analysis by FISH, which would circumvent the lack of specificity of silver staining (Sanchez *et al.,* 1995), was attempted in this study (5S rRNA and 28S rRNA genes isolated from gDNA), but was unsuccessful. Nonetheless, even with the use of single specimens and silver staining, the results obtained in the present investigation are useful since published NOR data is currently limited to few species within *Rattus s. l.* (i.e., Szabo *et al.,* 1978; Kodama *et al.,* 1980; Sasaki *et al.,* 1986; Cavagna *et al.,* 2002). Consequently, the present study extends what is known of Rattini NORs by including data on *M. surifer, L. edwardsi, N. fulvescens, B. bowersi, B. berdmorei, B. savilei, R. exulans, R. losea* and *R. tanezumi,* as well as an outgroup species *H. delacouri*.

The chromosomal location of several NORs are conserved across some species (i.e., orthologs of *R. norvegicus* pair 3, 4, 6, 8, 9, 10 and 12, see Table 4.3) limiting their use in a phylogenetic context. Even the NOR loci on *R. norvegicus* pair 11 that appeared, in the present study, to unite *R. exulans, R. losea, R. tanezumi* and *R. norvegicus* to the exclusion of the other species analysed herein (Table 4.3) proved unsatisfactory in a broader context. Following the inclusion of published data on *R. rattus* it was found that this character is not shared by the whole *Rattus s. s.* complex and its absence on chromosome 11 in *R. rattus* is most likely a consequence of the rearrangement that gave rise to RRA 4 (Cavagna *et al.,* 2002).

The data showed no co-localization of NORs with the IHB detected in the X-autosome translocations of *H. delacouri* and *B. savilei*, although this is equivocal given the

121

reliance only on silver staining (see Parish *et al.,* 2002; Dobigny *et al.,* 2003b, 2004b). Importantly, two of the NOR loci identified in *M. surifer* (MSU 17 and 19) correspond to chromosomes that have arisen through a fission event (disruptions of RNO 5 and 6, see Chapter 3) supporting the hypothesis that rDNA clusters may co-localize at, or close to, chromosomal breakpoints and provide the necessary functional telomeres needed to stabilize fission products (i.e., Hall and Parker, 1995; Rousselet *et al.,* 1998, 2000). Further evidence is provided by reports that rDNA can cap broken ends of human chromosomes (Zankl and Huwer, 1978) and chromosomal fissions were associated with rDNA in *N. abietis* (Rousselet *et al.,* 2000).

In conclusion, the rat sat I DNA family based on this study appears to have a recent origin and is specific to the *Rattus s. s.* group of species, a suggestion supported by both the molecular and physical mapping analysis conducted in this study. Although the great majority of satDNA sequences are known to evolve rapidly (Slamovits and Rossi, 2002; Ugarković and Plohl, 2002; Chaves *et al.,* 2005; Adega *et al.,* 2009; Acosta *et al.,* 2010) there are, however, several studies where slow to moderate satDNA evolution has been reported (e.g., Volobouev *et al.,* 1995; Mravinac *et al.,* 2002; Cafasso *et al.,* 2003; Li and Kirby, 2003; Pons and Gillespie, 2003; Robles *et al.,* 2004; Adega *et al.,* 2008). Bovine satDNA represents an example of rapid satDNA evolution, where new bovine sequence variants arose within 0.5-1 myr (cattle species diverged ~0.2 to 5 mya; Nijman and Lenstra, 2001). On the other hand, species within the insect genus *Pimelia* (Coleoptera) that are separated by ~5 or 6 myr contain nearly the same quantities of PIM357 satDNA (Pons *et al.,* 1997). Similarly, satDNA in the beetle *Palorus ratzeburhii* is indistinguishable from its counterparts in congeneric species which diverged ~7 mya (Meštrović *et al.,* 2000) and 170 bp HindIII satDNA is conserved in six species of the sturgeon family, even though these species are believed to have diverged ~80 mya (de la Herran *et al.,* 2001). The present study falls within

the latter category (slow to moderate satDNA evolution), as the rat sat I DNA family is conserved within all the *Rattus* species at the molecular level (>90% sequence similarity) and in location (centromeric). This pattern of constrained change in repetitive sequence evolution is also demonstrated by Ag-NORs analysis (chromosome location largely conserved between species) and the telomeres are located, as expected, at the ends of chromosomes (except for MSU 4 of *M. surifer* within the *Rattus s. l.* complex). Importantly, these findings extend those based on Zoo-FISH analysis using whole *R. norvegicus, M. surifer* and *R. rattus* chromosome paints (Chapter 3), which similarly showed the karyotypes of the *Rattus s. s.* group (*R. losea*, *R. exulans, R. tanezumi* and *R. norvegicus*) were highly conserved, implying a close phylogenetic relationship among species.

CHAPTER 5

RODENT INTERACTIONS WITH *ORIENTIA TSUTSUGAMUSHI*: A GENE EXPRESSION AND STRUCTURAL INVESTIGATION OF SYNDECAN-4

INTRODUCTION

The importance of zoonotic (animal-borne) diseases that are transmissible to humans is underscored by the recent emergence of infectious diseases resulting in outbreaks in various parts of the world (Lerdthusnee *et al.,* 2008). Several of these new "human" diseases are linked to rodents of the *Rattus s. l.* complex that act as reservoirs (Mills and Childs, 1998; Vorou *et al*., 2007; Meerburg *et al.,* 2009), of which many are highly invasive and easily exported to other parts of the world (e.g., *R. norvegicus, R. exulans* and *R. losea*) (Coleman *et al.,* 2003). A number of these outbreaks occurred in Asian countries and the epidemiology of rodent borne diseases, and the threats posed to humans in Asian countries have not been well defined (Kelly *et al.,* 2002).

In Thailand the primary rodent reservoir hosts of scrub typhus, the focus of this study, are *R. rattus, R. norvegicus, R. exulans, R. losea* and *B. indica* (Family Muridae, Order Rodentia; Coleman *et al.,* 2003; Lerdthusnee *et al.,* 2003, 2008). A number of other Rattini species have been linked to *O. tsutsugamushi* (causative agent of scrub typhus), for example *B. berdmorei, B. savilei,* and *R. argentiventer* (Coleman *et al.,* 2003). *Rattus norvegicus* has been reported to be the primary carrier of *O. tsutsugamushi* in areas of human habitation (Liu *et al.,* 2003). In addition, the preliminary French ANR Thailand data (S. Morand, pers. comm. and see Community Ecology of Rodents and their Pathogens in South-East Asia,

CERoPATH, website: http://www.ceropath.org) suggests that all species within the *Rattus s. l.* complex are carriers of typhus. Although Muridae also comprises the Murini lineage, no survey has identified any *Mus* species as the primary reservoir of the virus in Thailand. For example, a recent survey identified *M. caroli* and *M. cervicolor* rodents infested with chigger mites (infected chigger mites transmit *O. tsutsugamushi* to the animal host), however, none of these animals tested positive for *O. tsutsugamushi* infection while, conversely, Rattini species from the same localities tested positive (Coleman *et al.,* 2003). A single *M. caroli* specimen from the Chantaburi province, Thailand, tested positive for *O. tsutsugamushi* (Wangroongsarb *et al.,* 2008) representing the only reported case, to the author's knowledge, of this species' possible involvement with the disease in this country. Studies in laboratory settings have shown that mice can successfully be experimentally infected with *O. tsutsugamushi*, however, they become significantly moribund by day 10 post infection (Ewing *et al.,* 1978; Koh *et al.,* 2004) suggesting that Murini rodents may succumb rapidly upon infection. This limits the likelihood of detecting infected mice in the wild, and prevents mice acting as effective reservoirs for *O. tsutsugamushi*, whereas *Rattus* species appear to be able to maintain *O. tsutsugamushi* infections for long periods (months or longer; Walker *et al.,* 1973; Traub *et al.,* 1975; Van Peenen *et al.,* 1977; Frances *et al.,* 2000).

An essential step in the *O. tsutsugamushi* infectious process is its attachment to and the uptake by a host cell (Kim *et al.,* 2004). In a study by Ihn *et al.* (2000) the treatment of a host cell with heparin sulphate (HS) was found to prevent *O. tsutsugamushi* infection in a dose dependant manner. Consequently it was hypothesized that a common surface feature, heparin sulphate proteoglycans (HSPGs), is required for the attachment of *O. tsutsugamushi*. HSPG act as cellular receptors for numerous viruses (e.g., human immunodeficiency virus, foot-and-mouth, and herpes viruses) and bacteria, (e.g., *Helicobacter pylori, Chlamydia trachomatis, Streptococcus pyogenes* and *S. mutans* (Rostand and Esko, 1997; Wadstrom and

Ljungh, 1999). One of the major HSPGs on mammalian cell surfaces are the syndecans (Carey, 1997) which play a central role in the cell adhesion and uptake processes of various microbial pathogens (Saphire *et al.,* 2001). The only syndecan component located on most mammalian cell surfaces is syndecan-4 which comprises 5 exons and has a variable transcript length for example, 2488 bp in rat (Ensembl acc. no: ENSRNOT00000019386), 2458 bp in mouse (Ensembl acc. no: ENSMUST00000017153), and 2613 bp in human (Ensembl acc. no: ENST00000372733). Syndecan-4, together with integrins, is found to localize into focal adhesions (Woods and Couchman, 1994). Longely *et al*. (1999) suggested that syndecan-4 regulates stress fiber formation, focal adhesion, and exerts an influence on both cell migration and morphology. Kim *et al.* (2004) investigated whether there was a correlation between *O. tsutsugamushi* infection and HSPGs, specifically syndecan-4. Their findings indicated that *Orientia* infection was dependant on expression levels of syndecan-4 on the cell surface of rat embryo fibroblasts. They concluded that syndecan-4 plays a key role in *O. tsutsugamushi* infection in a dosage dependant manner.

At the outset it was hypothesized that syndecan-4 may occur in different chromosomal environments (and/or chromosomal fragments) that have been reshuffled within Rattini in comparison to Murini, resulting in variable rodent host/syndecan-4 interactions (typhus-positive and negative). However, it was subsequently concluded that the variability is not due to position effect (i.e., that changes in structural genomic environment/chromosome location alter gene function) given the conserved physical location of syndecan-4 in reservoir and non-reservoir Rattini and Murini rodents. Specifically, the FISH data (Chapter 3) show that the physical location of syndecan-4 on distal RNO 3, as identified using the Ensembl genome browser, is conserved between all the Rattini species (i.e., RNO 3 is not disrupted within the tribe), and distal RNO 3 is orthologous to MMU 2

126

(corresponding to the location of syndecan-4 in mouse; Spring *et al.,* 1994; Helou *et al.,* 2001; Nilsson *et al.,* 2001).

Taking the above data into consideration, it seems reasonable to hypothesize that the variable rodent host/syndecan-4 interactions between Rattini and Asian Murini may be due to intrinsic (genetic) differences. Thus it seems likely that the reservoir (Rattini) and apparent non-reservoir (Asian Murini rodents) status is a result of structural (at the sequence/protein level) differences, and/or the differential expression (mRNA) of syndecan-4 protein in these two lineages. This hypothesis is tested here using a novel candidate gene approach that focuses on the mRNA sequence and expression levels of syndecan-4 in these murine rodents. In so doing, selected syndecan-4 mRNA sequences will be scrutinised in a phylogenetic framework permitting the identification of functionally shared mutations (amino acids) that underpin the reservoir and non-reservoir status of rodents in these lineages. Additionally, 3D structural modifications were examined to identify functional differences among them. Studies of gene expression entailed comparisons of syndecan-4 transcripts between Rattini and Murini, as well as between *O. tsutsugamushi* seropositive and seronegative Rattini specimens.

MATERIALS AND METHODS

Samples

All rodents (Appendix 1) were live-caught in the wild by researchers from the CBGP (courtesy of the programme ANR-00121-05 "Biodiversity of rodent-borne hantaviruses in Southeast Asia"; coord. J.P. Hugot, 2006-2008). Rodents were selected to represent various murid lineages (see Lecompte *et al.,* 2008) with a particular emphasis on the Rattini and Murini rodents. African Murini, Praomyini and Arvicanthini were included for comparative purposes (African Murini occur outside the scrub typhus

infection/distribution, see Rosenberg, 1997; Parola *et al.,* 2005a, b) allowing the polarisation of intrinsic factors that may characterise *Orientia* infection in Asian Murini and Rattini rodents.

To avoid misidentification due to the ambiguous nature of some morphological characters, and frequent sibling species co-occurrence in the wild, all rodents were unambiguously identified through cytogenetic and/or molecular techniques (data not shown, but see Badenhorst *et al.,* 2009 and Pagès *et al.,* 2010 for Asian rodents; Dobigny *et al.,* 2008, 2010b for African rodent representatives, and Appendix 1 for summary). Epidemiological surveys for scrub typhus infection were conducted through serological screening and, where feasible, Rattini specimens of the same species were selected in order to include both seronegative and seropositive specimens for scrub typhus infection (refer to Appendix 1; unpublished data from ANR Biodiversity ANR 07 BDIV 012, project CERoPATH, "Community Ecology of Rodents and their Pathogens in a changing environment", coord. S. Morand, 2008-2011; CERoPATH website: http://www.ceropath.org).

Structural and sequence analysis of syndecan-4

Amplification and sequencing of syndecan-4

Although syndecan-4 is ubiquitously expressed in cells and tissues (Kim *et al.,* 1994; Rioux *et al.,* 2002) in the present study the spleen was specifically selected since it is considered particularly suitable for *Orientia* detection (Wangroongsarb *et al.,* 2008). Tissues collected in the field were preserved in RNA Later (Sigma) and mRNA transcripts of syndecan-4 cDNAs were amplified and subsequently sequenced.

cDNA (1:10) was used in syndecan-4 PCR reactions (PCR and qPCR) and was sequenced for phylogenetic analyses. In order to isolate the whole mRNA transcript of

syndecan-4 (including all 5 exons), primers were designed from the conserved mRNA regions of mouse *M. musculus* (Ensembl acc. no: ENSMUSG00000017009) and Norway rat *R. norvegicus* (Ensembl acc. no: ENSRNOG00000014297) using Primer 3 software (Untergasser *et al.,* 2007).

The syndecan-4 primers used were 5' GCCTGYCTGCCTGYTTGCG 3' (forward) and 5' CAAYGAGTTCTACGCATGAAGC 3' (reverse). PCR was performed in 25 µl volumes containing 2 µl cDNA (1:10), 1X Buffer (containing 15 mM of $MgCl_2$), 0.1 mM dNTP, 1 µM of both primers, 0.5 U Taq polymerase (Qiagen). PCR reactions were carried out at $94^{o}C$ for 3 min, followed by 40 cycles of $94^{o}C$ for 30 s, $55^{o}C$ for 1 min, $72^{o}C$ for 1.30 min, and a final elongation step of $72^{o}C$ for 10min.

All PCR products were visualized on a 1.5% agarose gel and samples that amplified satisfactorily were sent for sequencing (Macrogen, Korea).

Phylogenetic reconstruction

Nucleotide sequences were edited and aligned using ClustalW Multiple Alignment and visually checked (for accuracy) in Bioedit v. 7.0.9.0 (Hall, 1999). I was unable to obtain syndecan-4 sequences from several of the taxa initially selected as outgroups for the phylogenetic analysis (i.e., representatives of the Gerbillinae: *Gerbillus nigeriae* and *Gerbilliscus kempi* and Arvicolinae: *Microtus arvalis* and *Myodes glareolus*). Consequently use was made of *Acomys johannis* (Muridae, Deomyinae) (from Northern Cameroon; Dobigny *et al.,* 2008.), *Spermophilus tridecemlineatus* (Sciuridae; Ensembl acc. no: ENSTOG00000009310), *Dipodomys ordii* (Heteromyidae; Ensembl acc. no: ENSFM00500000273193**)** and *Homo sapiens* (Primates: Hominidae; Ensembl acc. no: ENSG00000124145). Bayesian inference (BI), maximum likelihood (ML) and maximum parsimony (MP) approaches were employed to retrieve phylogenetic relationships. The

model of evolution best suited to the data was selected using MODELTEST v 3.7 (Posada and Crandall, 1998). Both the model parameters and model of evolution were determined using the Akaike Information Criterion (AIC). This programme examines the fit of 56 models that allow for either a gamma distribution of substitution rate variation among sites (G), the proportion of invariable sites (I), or a combination of both (I + G). In the case of the syndecan-4 dataset (see below), the best-fit substitution model was GTR+G which was used to run each of the maximum likelihood and Bayesian analyses.

Bayesian analysis was conducted using MrBayes v. 3.1 (Huelsenbeck and Ronquist, 2001). We ran five Metropolis-coupled chains in a run of 1,000,000 generations, and trees were sampled every 100 generations to check that each run converged on a stable log-likelihood value against generation time for each. The first 10,000 generations were discarded as burn-in. The sampled trees were used to generate an optimal tree showing all compatible partitions and the support for the nodes was given by posterior probabilities. Maximum likelihood analysis was conducted using PhyML v. 3.0 (Guidon and Gascuel, 2003). Gamma shape parameters and probabilities of substitution were estimated from the data. Maximum parsimony analyses were performed using PAUP* 4.0b10 (heuristic searches, branch swapping and Tree-Bissection-Reconnection options, Swofford, 2002). For maximum parsimony (in PAUP) and maximum likelihood (in PhyML), bootstrap values (BP) were calculated after 1,000 replicates under the closest stepwise addition option. FigTree v. 1.2.2 and InkScape were used in phylogram construction and editing.

Mapping of amino acid changes and bioinformatic analysis of amino acid sequences.

Mapping of amino acid changes on gene tree

The amino acid changes were reported *a posteriori* on a consensus molecular tree from available published data (Lecompte *et al.,* 2008). To do so, nucleotide sequence data

was translated into amino acid sequence data in Bioedit v. 7.0.9.0 (Hall, 1999). Those amino acid changes (characters) important for Murini and Rattini evolution were identified using PAUP (the heuristic search option, Tree-Bisection-Reconnection) and mapped onto an independently obtained consensus tree topology (Lecompte *et al.* 2008). Although this is fully congruent with the syndecan-4 gene tree topologies obtained from the BI, MP and ML analyses used in this study, it was preferentially used since it is more robust than the syndecan-4 gene tree topologies (see below). This permitted the unambiguous tracking of changes in primary structure within the syndecan-4 protein that have been fixed during the evolution of Murinae.

Bioinformatic inference of the tertiary protein structure from amino acid sequences

As any lineage specific amino acid changes identified here may have important impacts on the 3D tertiary structure of the protein and on the functionality of the protein product, bioinformatic analyses were conducted using representative amino acid chains from lineages of interest (Rattini and Murini). These analyses were carried out to predict tertiary protein structures, as well as to identify conserved protein domains, ligand-binding sites and DNA-binding sites. The focus was on differences between Murini and Rattini syndecan-4 protein sequences which are of particular interest given possible differences in rodent host cell – *Orientia* interactions among species of these two clades. The amino acid sequences of two species of Murini (*M. musculus* and *N. mattheyi*) and Rattini (*R. norvegicus* and *B. savilei*) were subjected to InterProScan annotation (http://www.ebi.ac.uk/Tools/InterProScan/) that utilises the most frequently used protein databases (i.e., PRINTS, PROSITE, Pfam, ProDom), as well as the PDB resource (Protein Data Bank, contains information about experimentally determined protein structures and nucleic acids; http://www.pdb.org/pdb/home/home.do). This permits the identification of

sequence motifs through comparison with a database of protein domains, functional sites, and protein families (Apweiler *et al.,* 2001; Zdobnov and Apweiler, 2001; Quevillon *et al.,* 2005).

In order to accurately determine a protein's structure, experimental methods of X-ray crystallography and nuclear magnetic resonance (NMR) are needed (i.e., Wüthrich, 1990; Zhang, 2008, 2009). Unfortunately, both of these methods are time consuming, expensive and require specialised equipment (Zhang, 2008, 2009). As an alternative, protein prediction servers have been developed that allow for rapid predictions of possible protein structures. Tertiary structural predictions (i.e., the 3D fold of the polypeptide structure in space) were carried out using the I-TASSER server (http://zhang.bioinformatics.ku.edu/I-TASSER/) that employs a protein structure predictive algorithm which combines multiple thread alignments, structural refinements, *in silico* modelling, and iterative simulations (Zhang, 2008, 2009; Ambrish *et al.,* 2010). It is regarded as the most effective prediction technique in the Critical Assessment of protein Structure Prediction (CASP) experiments (Zhang, 2008).

The accuracy of the predicted 3D protein models are indicated by C- and TM-scores. The C-score is a confidence score and is based on the significance of the threading template alignments as well as the convergence parameters of the structure assembly simulations. Its value typically ranges from -5 to 2 and the higher the C-score, the greater the confidence in the predicted model (Zhang, 2008, 2009; Ambrish *et al.,* 2010). TM-scores are produced following an algorithm that determines the degree of similarity between the topology of two protein structures; in other words, it assesses the quality of the predicted protein structure in relation to a native and experimentally determined known protein structures. TM-scores between 0 and 1 were used to indicate the reliability of the structural predictions, with a TM score <0.17 indicating a random model selection from the PDB library, while a value >0.5 indicates a correct topology (Zhang and Skolnick, 2004). In

summary, reconstructions of predicted tertiary structure and scores (TM and C-score) allow comparisons of the predicted 3D model's topology against the experimentally determined (known) topology in the PDB thus indicating whether a predicted model is reliable or not.

The predicted tertiary (3D) structures were submitted to ProFunc (a server which executes a number of structure and sequence based analyses) to determine if any additional information on the functional parameters of the 3D structures could be retrieved (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc; Laskowski *et al.,* 2005). The ProFunc analysis also screens for nest sites (structural motifs frequently found in functionally important regions of protein structures; Watson *et al.,* 2002). The functional significance of a nest site is indicated by a nest score that is based on: (i) the conservation scores of its principal residues, (ii) the number of NH-atoms that are accessible to the solvent, (iii) and whether the nest is associated with one of the larger clefts (specific surface sites where molecules bind) that occur on the protein surface. A nest score >2 indicates a functionally significant nest (Laskowski *et al.,* 2005).

Molecular visualising software, RASMOL v 2.7.5 (Sayle and Milner-Whilte, 1995) was used for viewing the predicted 3D protein structures, and for determining some of the characteristic components of the 3D structure. A schematic outline of the structural and functional strategies followed in this study is shown in Figure 5.1.
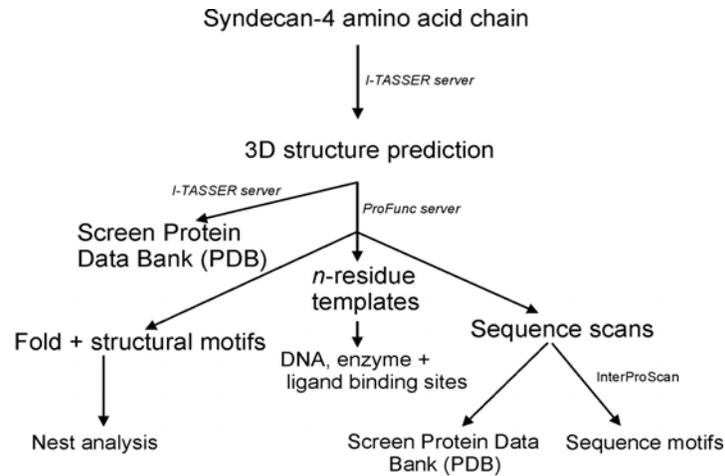
Figure 5.1. Schematic representation of the structural and/or functional analyses followed for syndecan-4 amino acid sequences (*R. norvegicus, B. savilei, N. mattheyi* and *M. musculus*). Arrows indicate the outputs; servers utilised for the analyses are designated alongside the arrow line.

Gene expression

RNA extraction

Total RNA was extracted from spleen tissue preserved in RNA Later® (Sigma) using the RNeasy mini Kit (Qiagen) according to the manufacturers' specifications with slight modifications for animal tissues, as well as the implementation of an additional DNase (Qiagen) step. Briefly, spleen tissue was ground to small pieces in liquid nitrogen using sterilized pestle and mortar. The RNA fraction was absorbed onto a silica-based matrix, DNase-treated, and eluted with 50 µl RNase-free water by centrifugation. The quality of the RNA fraction was determined using direct visualization on a 1% agarose gel, and quantified using a Nanodrop spectrophotometer (Thermoscientific). All samples were adjusted to a final concentration of 150 ng/µl using RNase-free water, and stored at -80°C for cDNA synthesis.

Reverse transcription PCR

cDNA was produced using ImProm-II[TM] Reverse transcription system (Promega) according to the manufacturers instructions for oligo(dT)$_{15}$ primed cDNA-synthesis. Briefly, cDNA synthesis was performed using 600 ng (150 ng/µl) of template RNA. The reaction involved initial denaturation for 5 min at 70°C, annealing for 5 min at 25°C, synthesis for 60 min at 42°C, and a final step at 70°C for 15 min to inactivate the reverse transcriptase. The cDNA obtained was diluted 1:10 and stored at -20°C prior to its use in PCR reactions for sequencing analysis, as well as for the design of qPCR-specific primers (see below, Table 5.1).

LightCycler real-time PCR

Selection of reference gene

A reference gene should ideally be one of the housekeeping genes, i.e., a gene expressed at a constant level in all tissues that is not influenced by experimental procedures (Radonić *et al.,* 2004). β-actin (which encodes cytoskeleton protein; Giulietti *et al.,* 2001) was selected here given its common usage in expression studies, and its expression in most cell types (including the spleen). In addition, β-actin copy numbers are generally constant under different types of treatment, and in different samples (Giulietti *et al.,* 2001).

Amplification and sequencing of β-actin

β-actin was sequenced in order to design primers in conserved regions, a prerequisite for qPCR analysis (see below). β-actin has an expected transcript length of 1263 bp and comprises 6 exons (Ensembl acc. no: ENSRNOG00000034254). The "murid" β-actin primers were obtained from Tayeh and colleagues (see Tayeh *et al.,* 2010) and cDNA (1:10) was used in PCR reactions. The PCR reaction was performed in a 25 µl volume containing 2

μl cDNA (1:10), 1X Buffer (containing 15 mM of $MgCl_2$), 2 mM $MgCl_2$, 0.2 mM dNTP, 1 μM of both forward (5' AGTTCGCCATGGATGACG 3') and reverse primer (5' CTCATCGTACTCCTGCTTGC 3'), 0.5 U Taq polymerase. The PCR programme used for amplification entailed an initial denaturing step of $95^{o}C$ for 3 min followed by 10 cycles (touchdown cycle) of $94^{o}C$ for 30 s, $65^{o}C$ for 55 s, $72^{o}C$ for 1 min, and another 30 cycles of $94^{o}C$ for 30 s, $55^{o}C$ for 55 s, $72^{o}C$ for 1 min and a final elongation step of $72^{o}C$ for 10 min (Tayeh *et al.,* 2010).

All PCR products were visualized on a 1.5% agarose gel and those that had amplified satisfactorily were sent for direct sequencing (Macrogen, Korea).

Primer design for qPCR analyses

In order to increase the chances of successful qPCR amplification (refer to Appendix 1 for samples/sequences used to design syndecan-4 and β-actin primer sequences), two sets of primers were designed for syndecan-4 and the reference gene, β-actin (Table 5.1) using LightCycler Probe Design Software 2.0 (Roche). Importantly, primers were designed to fall in conserved regions of the gene thus avoiding the use of degenerate nucleotides as far as possible. Both the forward and reverse primers for each gene were designed (a) to fall in separate exons (Table 5.1) further restricting gDNA amplification, (b) to generate an amplicon size of ~150 bp (short amplicons increase the efficiency of PCR amplifications and may increase the sensitivity of the detection; Yin *et al.,* 2001), and (c) to work optimally at a melting temperature (Tm) of approximately $60^{o}C$.

Table 5.1. Primer sequences for qPCR analyses of syndecan-4 and β-actin.

Primer pair name		Sequence (5'-3')	Exon
Syndecan-4 A	Forward	GTRTCCATGTCCAGCAC	4
	Reverse	CCTTCTTCTTCATGCGGT	5
Syndecan-4 B	Forward	CCCAAGGAACTGGAAGAGAAT	4
	Reverse	CGCCGCCCACYATCAGA	5

136

β-actin A	Forward	ACAATGAGCTGCGTGTG	3
	Reverse	ACAGCCTGGATGGCTAC	4
β-actin B	Forward	TCCGTAAAGACCTCTATGCCAA	3
	Reverse	CAGAGTACTTGCGCTCAG	4

qPCR analyses

The qPCR analysis was designed to consist of three experimental stages: qPCR 1 determined the optimal specificity of the different qPCR primer pairs designed for both the target and reference genes; qPCR 2 allowed the determination of optimal cDNA and primer concentrations that allow for repeatable, quantifiable results; and qPCR 3 quantified the expression levels of the target (syndecan-4) gene relative to those of β-actin. All qPCR reactions were performed in a LightCycler 480 (Roche) using 384-well microtitre plates. Importantly, each qPCR was performed within the same 384-well plate in order to avoid between-run differences. Once samples and the PCR mix were aliquoted into the plates, sealed, centrifuged and placed in the LightCycler rotor, the following protocol was used in all qPCR reactions: an initial denaturing step at 95°C for 10 min, followed by 45 cycles of 95°C for 5 s, 62°C for 15 s and 72°C for 10 s, and lastly a final cooling step to 40°C. A negative control was included in all reactions.

(i) qPCR 1

Both sets of syndecan-4 and β-actin primers (Table 5.1) were tested on all the samples to determine which primer pair (i.e., either primer pair A or B of syndecan-4 or β-actin) was optimal for amplification and quantification of the target and reference genes by qPCR 2 and qPCR 3 analysis. The 7 μl reaction for the LightCycler reaction contained 1 μl cDNA (1:20), 2.14 μl RNase and DNase-free water, 0.18 μl forward primer (0.3 μM), and 0.18 μl reverse primer (0.3 μM), and 3.5 μl SYBR Green I master mix (Roche, containing:

137

fast start Taq polymerase, reaction buffer, dNTPs, $MgCl_2$, and SYBR Green I dye). First the

PCR mix was aliquoted into the 384-well plate, followed by cDNA dilutions. Melting (Tm)

curves for all samples were analysed using the LightCycler software 1.5 in order to determine

whether the target and reference gene were amplifying optimally and that there are no

contaminating PCR products. The PCR products were also visualized on a 2.5% agarose gel

in order to determine the quality of amplification.

(ii) qPCR 2

In this reaction, the selected primer pairs (as determined in qPCR 1) and cDNA

concentrations were tested to identify optimal concentrations for qPCR 3. The efficiency of

each PCR reaction (for the different primer concentrations) was calculated using LightCycler

software 1.5. Ideally, the efficiency should fall in the range of 1.8-2 with 2 indicating

maximal efficiency. The efficiency of the PCR reaction is represented by the standard curve

that describes the overall quality of the PCR amplification (i.e., Giulietti *et al.,* 2001). The

slope of a standard curve (concentrations of standard samples, i.e., serial dilution are plotted

against crossing points of the samples) obtained from a perfect amplification reaction must

approximate 3 with an ideal value of 3.3 (see for example Giulietti *et al.,* 2001; Yuan *et al.,*

2006). Quantitative analyses of mRNA (transcripts, hence expression levels) rely on the

determination of the crossing point (CP) which is defined as the point at which the

fluorescence rises appreciably above the background fluorescence (i.e., Pfaffl, 2001).

To do so, cDNA of all samples was pooled and diluted to obtain a dilution series of

1:10, 1:20, 1:40, 1:80 and 1:160. Different primer concentrations (0.2, 0.3, 0.4, and 0.5 µM)

were tested at the same time. The PCR mixes utilised for the different primer concentrations

(each containing 1 µl cDNA from the dilution series) with a final volume of 7 µl were: (a)

2.26 µl RNase and DNase-free water, 0.12 µl forward primer (0.2 µM), and 0.12 µl reverse

primer (0.2 μM), and 3.5 µl SYBR Green I master mix; (b) 2.14 µl RNase and DNase-free water, 0.18 µl forward primer (0.3 μM), and 0.18 µl reverse primer (0.3 μM), and 3.5 µl SYBR Green I master mix; (c) 2.02 µl RNase and DNase-free water, 0.24 µl forward primer (0.4 μM), and 0.24 µl reverse primer (0.4 μM), and 3.5 µl SYBR Green I master mix, and (d) 1.9 µl RNase and DNase-free water, 0.3 µl forward primer (0.5 μM), and 0.3 µl reverse primer (0.5 μM), and 3.5 µl SYBR Green I master mix. All samples were analysed in triplicate in the same 384-well plate.

(iii) qPCR 3

Optimal target and reference gene primer pairs (identified in qPCR 1) and the optimal primer and cDNA concentrations (identified in qPCR 2) were utilised to amplify both the target and reference gene in all the samples. All samples were analysed in triplicate (within the same plate) from which the mean CP values were used for further comparative analysis of relative gene expression (see below). PCR reactions (final volume of 7 µl for both the target and reference gene) contained 1 µl cDNA, 1.9 µl RNase and DNase-free water, 0.3 µl forward primer (0.5 μM), and 0.3 µl reverse primer (0.5 μM), and 3.5 µl SYBR Green I master mix.

(iv) Calculation of relative expression levels of the target (syndecan-4) gene

The fluorescent signal from SYBR green dye I was measured in real time during the PCR programmes. Signal increases as PCR product in reaction increased, and this plateaus once the PCR reagents are depleted (i.e., Giulietti *et al.,* 2001; Vandesompele *et al.,* 2002). 'Fit Point Method' (CP measured at constant fluorescence level; Rasmussen, 2001) was performed in the LightCycler software 1.5 (Roche) following the manufacturer's instructions.

Pairwise relative quantification of syndecan-4 was calculated using the Pfaffl (2001) method which takes cDNA load variation, RT-PCR efficiency, and RNA integrity into account. In essence, one pool of samples was considered as the "control" and another as the "sample" group, with syndecan-4 the target gene and β-actin as the reference gene (see Equation one by Pfaffl, 2001). A number of calculations were conducted to investigate the expression levels of syndecan-4 within or between Murini, Asian Murini, Rattini, *Orientia*-seropositive and *Orientia*-seronegative rodents.

The following relative syndecan-4 expression levels were compared: (i) African Murini vs. Asian Murini, (ii) Rattini samples vs. the remainder of samples, (iii) Rattini samples vs. Murini (iv) Rattini samples vs. Asian Murini, (v) Rattini rodents seropositive for *Orientia* infection vs. all the seronegative rodents, (vi) Rattini rodents seropositive for *Orientia* infection vs. seronegative Rattini rodents, and, when feasible, (vii) congeneric Rattini rodents seropositive for *Orientia* vs. seronegative rodents of the same genus (e.g. within the *Bandicota* genera), (viii) Murini vs. seronegative Rattini, (ix) Asian Murini vs. seronegative Rattini, (x) Murini vs. seropositive Rattini, and (xi) Asian Murini vs. seropositive Rattini. This allowed the determination of R which corresponds to the relative expression ratio that, in turn, reflects the expression of the target gene in the "control" group vs. that of the "sample" group. For instance, if two species, A and B, are considered as the control and sample group respectively, values of R>1 (or R<1) reflect an overexpression (or underexpression) of the target gene in the control species A, relative to sample species B. R values ranging between 0.8-1.2 were considered to be similar and thus potentially ambiguous.

RESULTS

Structural and sequence analysis of syndecan-4

Phylogenetic analysis

Syndecan-4 transcripts of expected length and sequence (when compared to the available mouse and rat syndecan-4 sequences, i.e., Ensembl acc. no: ENSMUSG00000017009 and Ensembl acc. no: ENSRNOG00000014297, respectively) were obtained following PCR amplification. None of the sequences displayed stop codons following translation into amino-acids. The final alignments of syndecan-4 sequences included 586 nucleotides derived from 55 specimens representing 22 taxa (refer to Appendix 1).

The maximum likelihood (ML) and Bayesian analyses (BI) phylogenetic trees produced using the best-fitting model (GTR+G) are shown in Figure 5.2 and 5.3, respectively. The maximum parsimony analysis (MP), following a heuristic search, retrieved 20 equally parsimonious trees and the strict consensus (L = 544) is shown in Figure 5.4. The topologies obtained from all three methodologies are largely similar except for the tribal relationships which are either unresolved (MP analysis, Figure 5.4), or poorly supported (BI and ML analyses, Figure 5.2 and 5.3).

The monophyly of all murine tribes represented here are well supported in the three topologies: Praomyini (*Praomys nov. sp.*, *Mastomys natalensis, M. erythroleucus* and *M. kollmannspergeri*; 1.00 PP in BI, 100% BP in both ML and MP), Arvicanthini (*Arvicanthis nilocitus* and *Lemniscomys zebra*, (1.00 PP, 100%, BP in both ML and MP), Murini (*N. mattheyi, M. musculus, M. caroli, M. cooki* and *M. cervicolor*; 1.00 PP, 100% BP in both ML and MP) and Rattini (*M. surifer, N. fulvescens, L. edwardsi, B. bowersi, B. berdmorei, B.*

141

indica, B. savilei, R. exulans, R. losea, R. tanezumi and *R. norvegicus*; 1.00 PP, 95% BP in

both ML and MP).

Figure 5.2. Maximum likelihood tree obtained from the syndecan-4 marker dataset. Statistical support for the nodes is indicated when BP>70%. Arrow indicates the basal placement of *H. delacouri* (HDE). Specimen numbers and abbreviations correspond to those provided in Appendix 1.
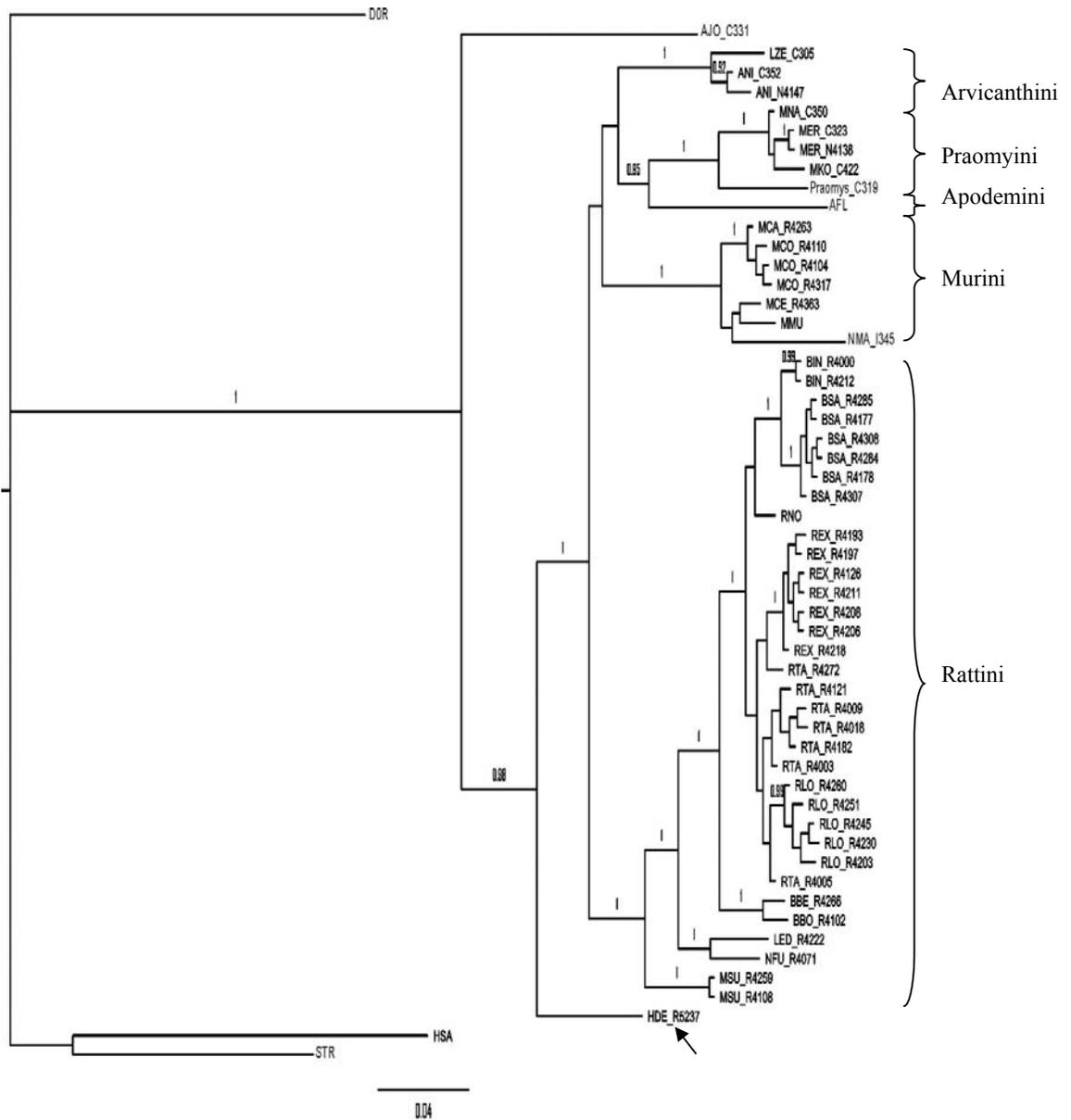
Figure 5.3. Tree obtained using Bayesian inference analysis of the syndecan-4 dataset under the GTR+G model. Statistical support for the nodes is indicated when PP>0.92. Arrow indicates the basal placement of *H. delacouri* (HDE). Specimen numbers and abbreviations correspond to those provided in Appendix 1.

143
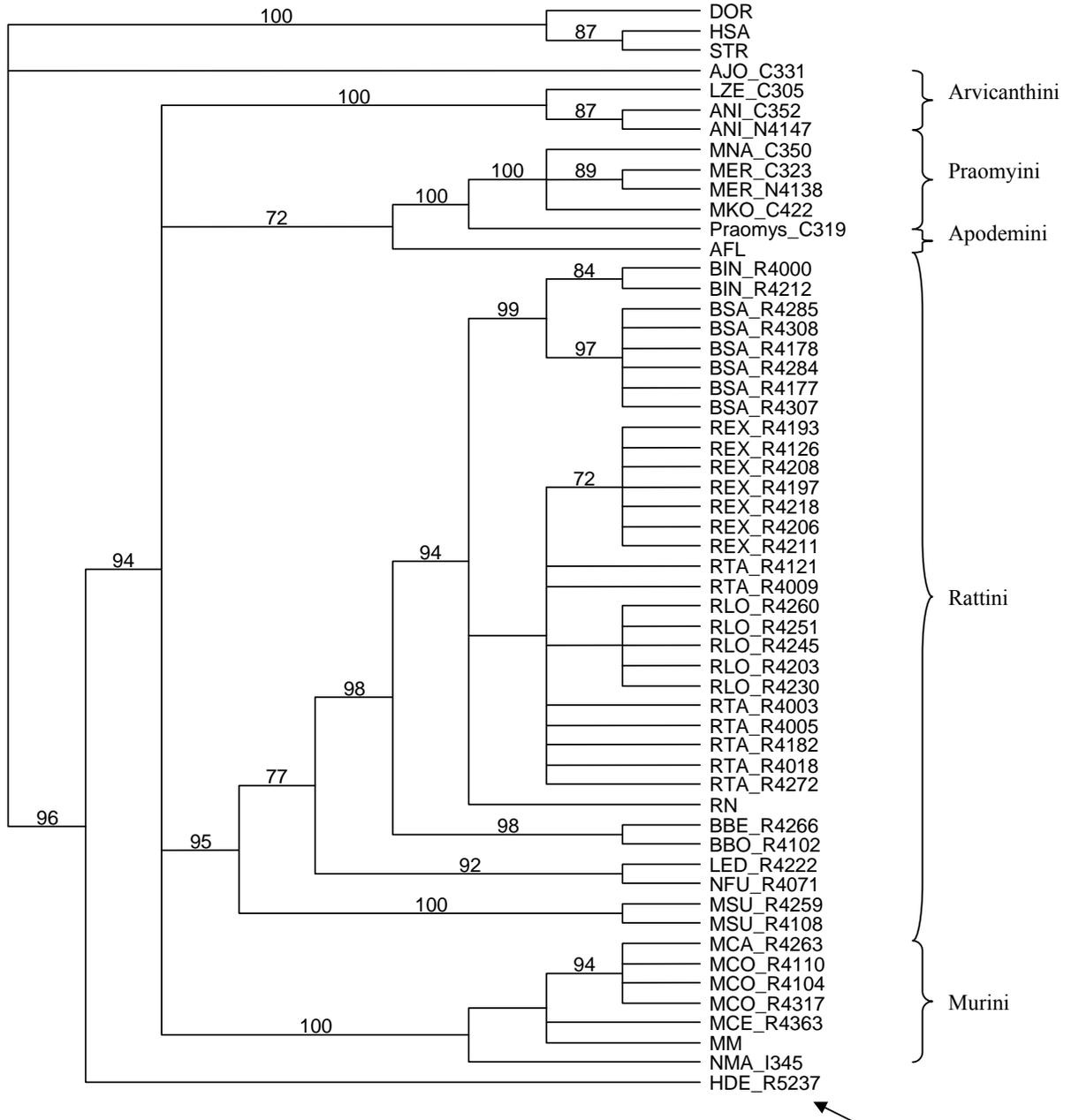
Figure 5.4. Maximum parsimony consensus tree from the combined syndecan-4 dataset following an exhaustive heuristic search in PAUP. Statistical support for the nodes is indicated when BP>70%. Tree scores: CI = 0.676, RI = 0.825, and HI = 0.324. Arrow indicates the basal placement of *H. delacouri* (HDE). Specimen numbers and abbreviations correspond to those provided in Appendix 1.

Hapalomys delacouri was the most basal lineage in all three topologies (0.98 PP,

100 % BP, 96% MP), a finding consistent with Pagès *et al.* (unpublished results using nuclear

144

GRH, IRBP and mitochondrial cytochrome b genes). Praomyini and Apodemini were sister lineages in all three types of analysis with good (0.95 in BI) to moderate support (72% and 73% in MP and ML, respectively), supporting recent phylogenetic conclusions based on cytochrome b and IRBP sequences (Lecompte *et al.,* 2008). The placement of Apodemini and Praomyini as sister lineage to Arvicanthini was retrieved by ML and BI analysis (unresolved in MP) which is similarly reflected in the Lecompte *et al.* (2008) study, however these relationships are poorly supported. Other inter-tribal relationships were unresolved, irrespective of the methodology used.

The Murini formed a monophyletic clade that included the African *N. mattheyi*. The internal Murini relationships were, however, largely unresolved in all three topologies. The internal relationships within the Rattini (*Rattus s. l.* complex) were quite well resolved and consistent with recent molecular analyses (i.e., Verneau *et al.,* 1997; Lecompte *et al.* 2008; Robins *et al.,* 2008; Pagès *et al.,* 2010). Only the relationships within *Rattus* (*R. exulans, R. losea, R. tanezumi* and *R. losea*), the most derived of the lineages examined, are unresolved, as was the placement of *B. savilei* and *B. indica.* These findings parallel those of Pagès *et al.* (2010) where the monophyly of *Rattus* was moderately to poorly supported.

Mapping of amino acid changes and bioinformatic analyses of amino acid sequences

Mapping of amino acid changes on gene tree

The examination of amino acid differences between Murini and Rattini showed that few amino acid changes could be linked to the typhus reservoir lineages (Rattini) and non-reservoir lineage (Murini) (Figure 5.5). Position numbers of the different amino acids (Figure 5.5) correspond to the syndecan-4 amino acid sequences alignment obtained using ClustalW (presented in Appendix 2).

145

Comparisons of amino acid residues (i.e., those that were invariant within a clade, but differed between clades) revealed three Rattini defining characters (Figure 5.5), namely a synapomorphy of serine (S) 134, a deletion of residue 55 (D), and an insertion of an aspartic acid (D) 122 residue. Three clade-defining characters were identified for Murini: a proline (P) 72 residue, a glutamic acid (E) 78 residue and the deletion of four amino acid residues 49-52. Two synapomorphies were identified for Asian Murini (i.e., excluding the African *N. mattheyi*) - alanine (A) 114 and methionine (M) 124. Two homoplasic characters were also identified for Rattini and Apodemini, specifically the phenylalanine (F) 35 and phenylalanine (F) 49 amino acid residues.
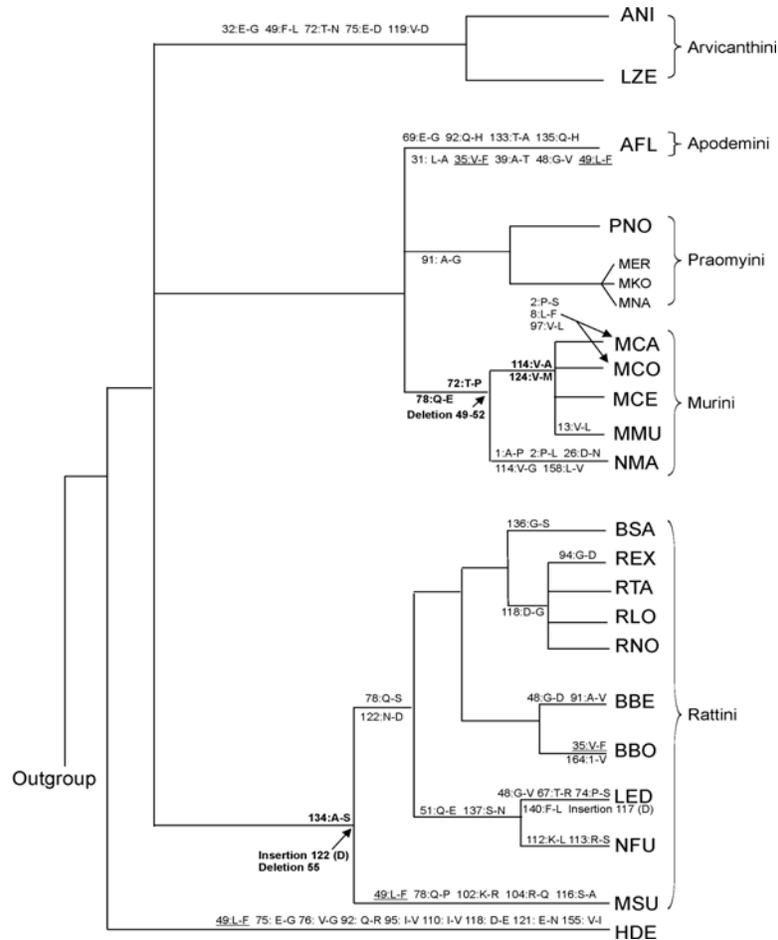
Figure 5.5. Syndecan-4 amino-acids mapped *a posteriori* to the branches of the murine (Rattini, Murini, Arvicanthini, Apodemini and Praomyini) species tree redrawn from

146

Lecompte *et al.* (2008). Rattini and Murini (including Asian Murini) defining amino acid characters are indicated in bold; homoplasic amino acid residues are underlined. Specimen numbers and abbreviations correspond to those provided in Appendix 1. Amino acid abbreviations are: alanine (A); aspartic acid (D), glutamic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L), methionine (M), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S), threonine (T), and valine (V).

<u>Bioinformatic inference of tertiary protein structure from amino acid sequences</u>

The InterProScan search using the syndecan-4 *M. musculus*, *N. mattheyi*, *B. savilei* and *R. norvegicus* protein sequences yielded seven matching, conserved sequence motifs that were identical in all analysed protein sequences. The reference codes of the seven sequence motifs (names in brackets) were: tmhmm (transmembrane regions), SM00294 (4.1m), PS00964 (syndecan), PTHR10915:SF3 (syndecan-4), PTHR10915 (syndecan), PF01034 (syndecan) and motif SignalP (signal-peptide).

The *R. norvegicus*, *B. savilei*, *N. mattheyi* and *M. musculus* syndecan-4 protein 3D models were obtained from I-TASSER (Figures 5.6, 5.7a and 5.8a). The main structural features - α-helices, β-strands, H-bonds and turns (protein structures must contain turns which allow the peptide backbone to fold) - of the predicted protein models were highly variable among the four 3D structures obtained (Table 5.2), as were the topologies (Figures 5.6, 5.7a and 5.8a). All four models show TM scores <0.5 indicating only moderately reliable models (incorrect topology); these are most likely responsible for the differences in topology observed between the four structures (see Table 5.4 for TM and C-scores). Thus any inferences made from these findings need to be treated with caution.

Table 5.2. The main structural features and TM- and C-scores of the predicted syndecan-4 protein models obtained from the I-TASSER server.

Predicted syndecan-4 3D model	α-helices	β-strands	H-bonds	turns	TM-score	C-score
R. norvegicus	5	0	93	32	0.36±0.12	-3.19

B. savilei	3	7	91	30	0.30±0.10	-3.80
M. musculus	14	0	124	22	0.38±0.13	-2.93
N. mattheyi	7	0	107	24	0.36±0.12	-3.11

The predicted 3D protein structures (*M. musculus, N. mattheyi, B. savilei* and *R. norvegicus*) were subjected to ProFunc analysis and, as expected, searches using FASTA against existing PDB entries identified significant homology (>96%) with the *H. sapiens* syndecan-4 protein cytoplasmic domain (PDB acc. code: 1ejp(A)). No enzyme active sites, ligand or DNA binding templates were identified in any of the models. However, several nest sites (structural motifs found in functionally important regions of protein structures) were identified in all four predicted proteins. Importantly, although the amino acid sequences between *R. norvegicus* and *B. savilei*, and between *M. musculus* and *N. mattheyi,* are similar, different nest sites were identified between the four species-specific 3D models. This may be due to the variability of the moderately supported (Tm<0.5) predicted 3D structures (see Figure 5.6, 5.7 and 5.8), as the nest sites are determined through structure-based analyses of the 3D structures (in ProFunc). Note that the positions of the different amino acids in the nest sites that are listed below correspond to the syndecan-4 amino acid sequences alignment obtained using ClustalW (Appendix 2). One functionally important nest site (>2.0) was identified in *N. mattheyi* (the residue range of glutamic acid 102 – glutamic acid 104, nest score = 3.77), whereas two nest sites were detected in the *B. savilei* protein model (aspartic acid 116 – glycine 118, nest score = 4.00, and proline 87 – asparagine 89, nest score = 3.16). The 3D *R. norvegicus* syndecan-4 protein model only contained two functionally significant nests (serine 59 – serine 61, nest score = 4.0, and aspartic acid 45 – leucine 49, nest score = 3.50), while three nests that were functionally significant were identified in the *M. musculus* protein model (serine 178 – aspartic acid 180, nest score = 3.97, glycine 9 – phenylalanine 11, nest score = 3.89, and glycine 48 – glutamic acid 56, nest score = 2.55). Note that in the

148

M. musculus syndecan-4 protein the latter nest corresponds to glycine 48-glutamic acid 53

(see the ClustalW alignment in Appendix 2 where residues 49-52 are deleted in *M.*

musculus). Importantly, one of the nest sites from *R. norvegicus* (aspartic acid 45 – leucine

49) overlaps with amino acid residues 49-52 which are deleted in Murini (Figure 5.7b and c),

and one of the Murini (*M. musculus*) nest sites (glycine 48 – glutamic acid 56) overlaps with

amino acid residue aspartic acid-55 which is deleted in Rattini (Figure 5.8b and c).
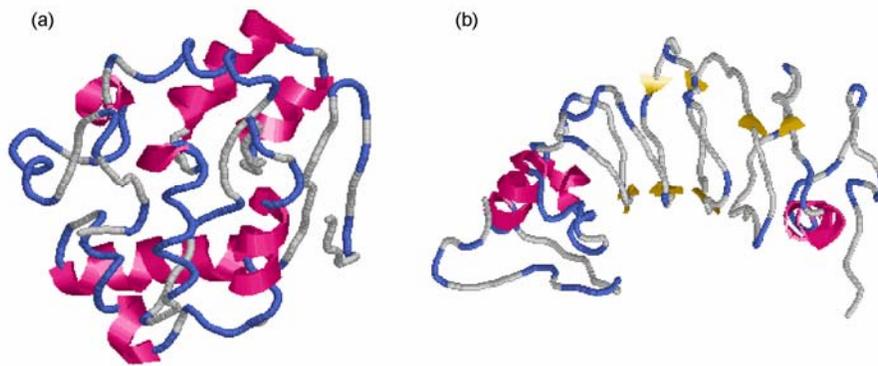
Figure 5.6. **(a)** Predicted 3D protein model of *M. musculus* syndecan-4. **(b)** Predicted 3D protein model of *B. savilei* syndecan-4. The main structural features are depicted: α-helices in pink, β-strands in yellow (secondary structures), beta turns in blue and other parts of the strand in grey.
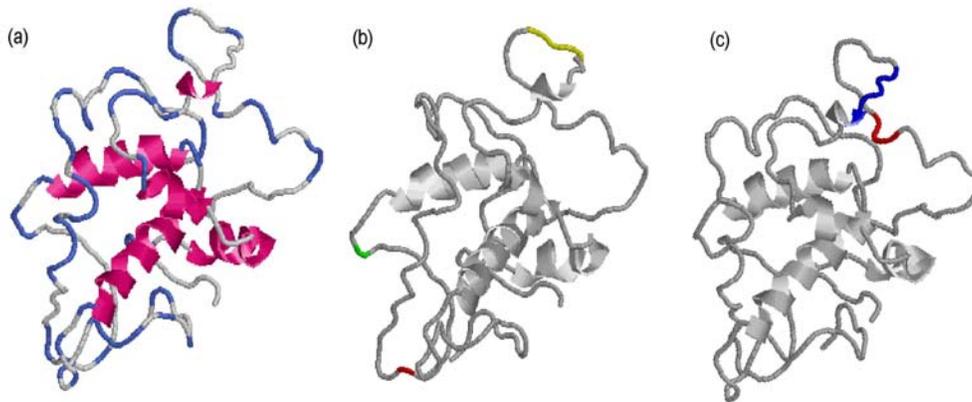
Figure 5.7. **(a)** Predicted 3D protein model of *R. norvegicus* syndecan-4 with α-helices depicted in pink, beta turns in blue, and other parts of the strand in grey. **(b)** Image illustrating the location of the amino acid changes identified between the Rattini and Murini clade. Relevant amino acid residues present in *R. norvegicus* syndecan-4 3D protein structure are: leucine 49 – aspartic acid 52 (yellow, deletion in Murini); insertion aspartic acid 122 (red) and serine 134 (green). **(c)** Illustration of the functionally significant nest sites identified through ProFunc analysis of the 3D model; serine 59 – serine 61 (red) and aspartic

acid 45 – leucine 49 (blue). The angle of the 3D models in (c) and (d) were adjusted in RASMOL to ensure that all highlighted amino acid residues and nest sites are visible.
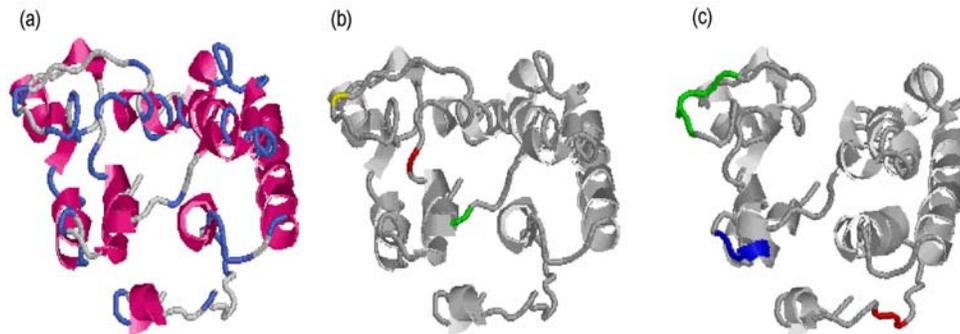
Figure 5.8. (a) Predicted 3D protein model of *M. musculus* syndecan-4 with α-helices depicted in pink, beta turns in blue and other parts of the strand in grey. **(b)** Image illustrating relevant amino acid residues present in *M. musculus* syndecan-4 3D protein structure: aspartic acid 55 (yellow, deletion in Rattini); proline 72 (red) and glutamic acid 78 (green). **(c)** Illustration of the functionally significant nest sites identified through ProFunc analysis of the 3D model: serine 178 – aspartic acid 180 (red); glycine 9 – phenyalanine 11 (blue) and glycine 48 – glutamic acid 56 (green). The angle of the 3D models in (c) and (d) were adjusted in RASMOL to ensure that all highlighted amino acid residues and nest sites are visible.

Gene expression analysis

RNA extraction

It is expected that intact, total RNA visualized on a agarose gel should at least show two distinctive bands representing 28S and 18S rRNA (ribosomal RNA), whereas a completely denatured RNA fragment will appear only as a smear (Dheda *et al.,* 2004; Huggett *et al.,* 2005). All RNA samples that yielded low RNA concentrations (<150 ng/µl) and did not display the 28S and 18S rRNA bands clearly on the agarose gel (Figure 5.9), were re-extracted until clear 18S and 28S bands were obtained. When concentrations were >150 ng/µl they were adjusted to 150 ng/µl for reverse transcription PCR.
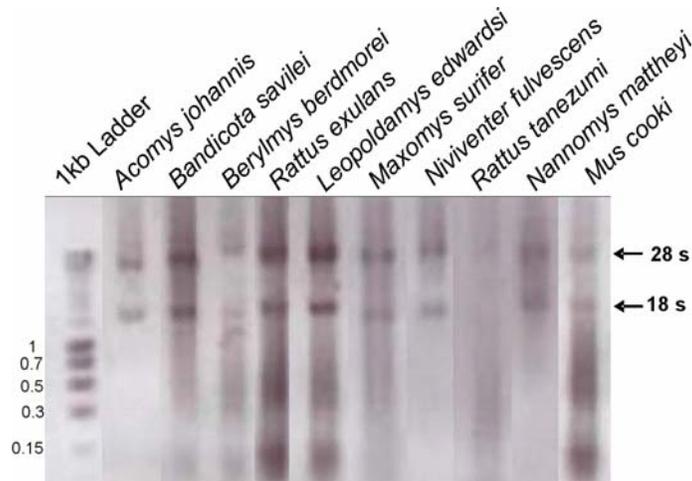
Figure 5.9. RNA extracts from several samples of this study visualized on a 1.5% agarose gels. Examples of superior quality RNA extracts are seen in *B. savilei, R. exulans*, and *L. edwardsi* lanes. Samples were re-extracted where the 28S and 18S rRNA bands are very faint, i.e., *B. berdmorei*, or where the RNA is denatured to the extent where only a smear can be seen, i.e., *R. tanezumi*.

Reverse transcription PCR

The reverse transcription using syndecan-4 and β-actin specific primers was successful for all species and amplicons showed fragments of expected length on electrophoresis (data not shown).

LightCycler real-time PCR

(i) qPCR 1

Products generated by syndecan-4 primer pair A and β-actin pair B (Table 5.1) yielded superior Tm curves and, as a consequence, subsequent qPCR reactions (namely qPCR 2 and 3, see below) were performed using these primer pairs. Variation in Tm curves generated by syndecan-4 corresponded to the different rodent genera used in the study (Figure 5.10). The selection of syndecan-4 primer pair A was further corroborated by visualization on agarose gel where only single PCR products (bands) were observed (data not shown).

151
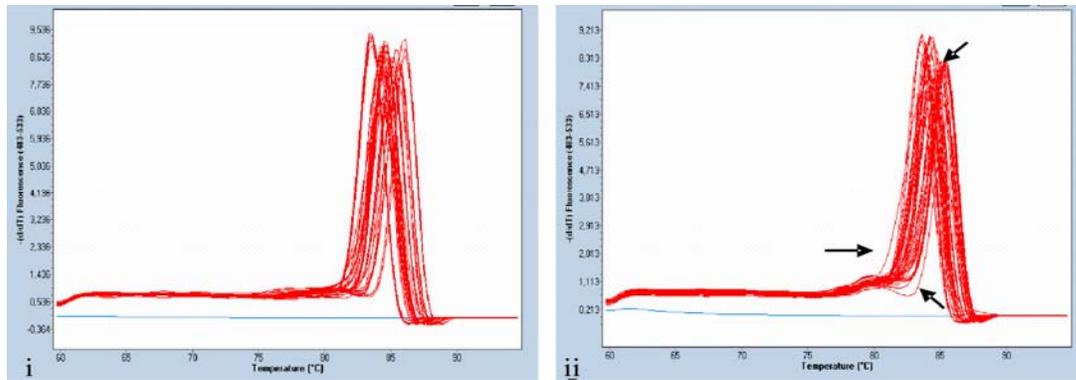
Figure 5.10. (i) Melting curves of samples of each representative of species of Rattini and Murini amplified using primer pair A and **(ii)** primer pair B of syndecan-4. If the PCR reaction generated more than one amplicon due to primer-dimers or contamination, the melting curve analysis will display more than one melting peak (i.e., more than one melting curve per sample). Arrows indicate substandard melting curves in primer pair B.

The β-actin primer pair A (see Table 5.1) displayed an unsatisfactory melting curve, while that of primer pair B was optimal (Figure 5.11). Visualization on agarose gels further corroborated the β-actin primer pair B as superior since no additional bands were seen in any of the samples, while primer pair A led to the amplification of additional bands in several samples (data not shown).
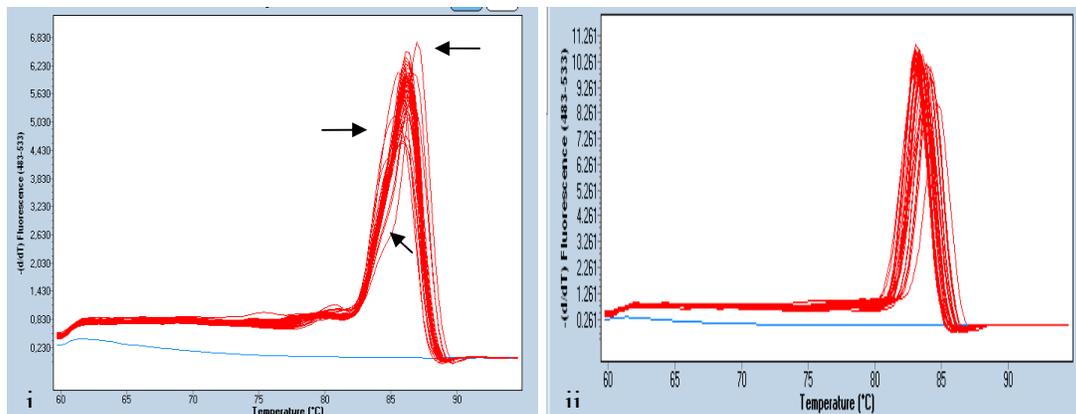
Figure 5.11. (i) Melting curves of samples (representatives of species of Rattini and Murini) amplified using primer pair A and **(ii)** primer pair B of β-actin. Arrows indicate unsatisfactory melting curves.

152

(ii) qPCR 2

The optimal primer concentration for both syndecan-4 and β-actin genes (Figure 5.12 and 5.13) was determined to be 0.5 µM. A 1:20 cDNA dilution, which was tested at the same time as the primer concentration, yielded the best results for both markers in the dilution series tested. An efficiency of 1.934 (Figure 5.12; standard curve indicates the efficiency of the PCR reactions) was obtained for syndecan-4 using a 0.5 µM primer concentration, while the lowest was 1.696 for a primer concentration of 0.2 µM (data not shown). Consequently qPCR 3 reactions were performed using a 0.5 µM primer pair A concentration for syndecan-4 together with a 1:20 cDNA sample dilution.
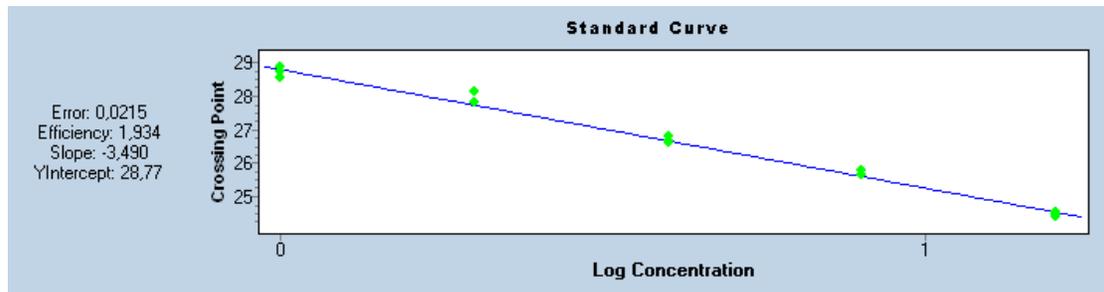
Figure 5.12. Optimal standard curve obtained for syndecan-4 using a 0.5 µM primer concentration with highest efficiency of 1.934 and a slope of -3.4. Crossing point (CP) values are plotted against input cDNA copy number. The green dots represent the concentrations of the cDNA dilution series. A perfect amplification reaction should generate a standard curve with an efficiency of '2' and a slope of -3, in reality, however, all reactions generally exhibit efficiency <2.

The greatest PCR efficiency for β-actin (Figure 5.13) was 1.916 based on a 0.5 µM primer concentration (similar to syndecan-4 primer pair A), while the lowest was 1.710 with a primer concentration of 0.2 µM (data not shown). The qPCR 3 reactions were conducted using a 0.5 µM primer pair B concentration for β-actin and a 1:20 cDNA sample dilution.
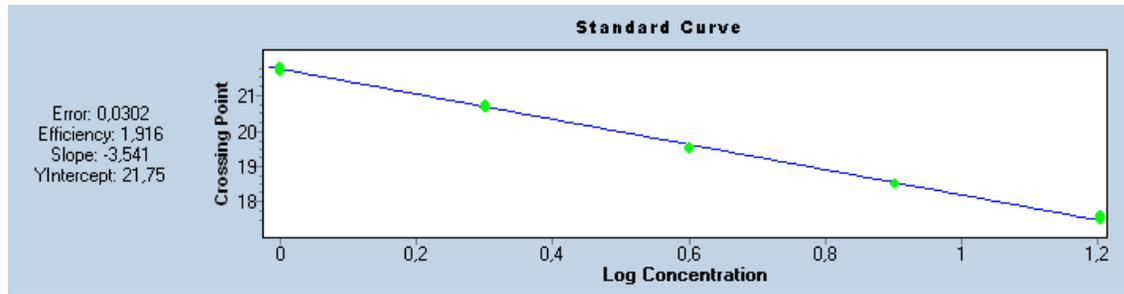
Figure 5.13. Optimal standard curve obtained for β-actin using a 0.5 μM primer concentration with highest efficiency of 1.916 and a slope of -3.5. Crossing point (CP) values are plotted against input cDNA copy number. The green dots represent the concentrations of the cDNA dilution series.

(iii) qPCR 3

Following the identification of optimal cDNA and primer concentrations, and based on the standard curve calculated from the cDNA dilution series included in the 384-well plates, PCR efficiencies were 1.976 and 1.942 for the amplification of syndecan-4 and β-actin in qPCR 3 respectively (Figure 5.14). The standard deviations (STD) of the CP of the triplicates for each sample were not significant (STD<0.5, see Table 5.3). A STD>0.5 would have indicated a problem with the experimental design or technique (e.g., significant differences between cDNA or primer load; i.e., Karlen *et al.,* 2007).
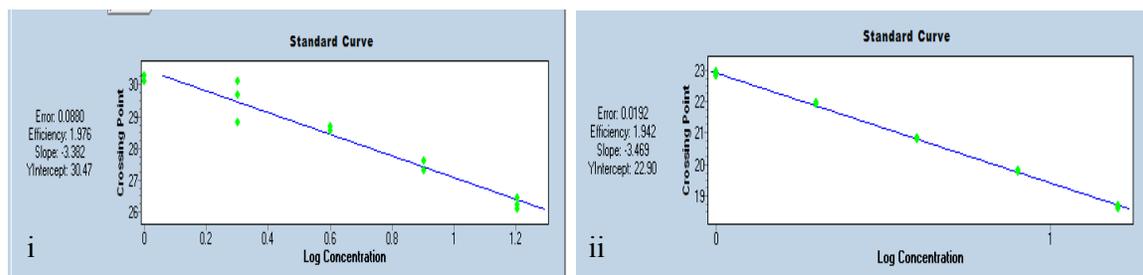
Figure 5.14. (i) Standard curve of a dilution series included for the amplification of the target gene, syndecan-4, and **(ii)** reference gene, β-actin to determine the PCR efficiency of the qPCR 3 experiments.

Table 5.3. Mean crossing point (CP) and standard deviation (STD) values for all samples of Rattini (blue), Murini (yellow) and the additional taxa (green) that have been included for comparative purposes in the study (see Materials and Methods), including the qPCR 3 efficiencies of the target and reference gene.

PCR efficiency		1.976	1.942				
				Syndecan-4	β-actin	Syndecan-4	β-actin
	Sample Name	Target	Reference	STD CP	STD CP	Mean CP	Mean CP
Acomys johannis	C331	Syndecan-4A	β-actin	0.03	0.04	25.22	17.00
Arvicanthis niloticus	C352	Syndecan-4A	β -actin	0.17	0.09	25.83	17.59
Mastomys erythroleucus	C323	Syndecan-4A	β–actin	0.15	0.06	27.97	21.54
Bandicota indica	R4212	Syndecan-4A	β–actin	0.13	0.03	27.45	18.10
Bandicota indica	R4000	Syndecan-4A	β-actin	0.10	0.05	30.49	21.57
Bandicota savilei	R4285	Syndecan-4A	β-actin	0.08	0.03	25.15	18.57
Bandicota savilei	R4308	Syndecan-4A	β-actin	0.08	0.05	27.69	19.99
Bandicota savilei	R4178	Syndecan-4A	β-actin	0.02	0.02	28.67	21.42
Bandicota savilei	R4284	Syndecan-4A	β–actin	0.12	0.13	27.06	18.88
Bandicota savilei	R4177	Syndecan-4A	β–actin	0.16	0.02	25.10	18.03
Bandicota savilei	R4307	Syndecan-4A	β–actin	0.06	0.04	26.74	17.53
Berylmys bowersi	R4102	Syndecan-4A	β–actin	0.04	0.05	29.17	18.08
Berylmys berdmorei	R4266	Syndecan-4A	β–actin	0.12	0.01	29.16	18.47
Leopoldamys edwardsi	R4222	Syndecan-4A	β–actin	0.08	0.07	26.16	17.68
Leopoldamys edwardsi	R4370	Syndecan-4A	β–actin	0.11	0.04	24.96	16.85
Niviventer fulvescens	R4071	Syndecan-4A	β–actin	0.38	0.08	24.43	17.22
Maxomys surifer	R4259	Syndecan-4A	β–actin	0.04	0.07	23.78	18.33
Maxomys surifer	R4108	Syndecan-4A	β–actin	0.19	0.05	28.00	21.81
Rattus exulans	R4208	Syndecan-4A	β–actin	0.12	0.07	26.69	18.40
Rattus exulans	R4197	Syndecan-4A	β–actin	0.04	0.08	26.77	19.00
Rattus exulans	R4218	Syndecan-4A	β–actin	0.04	0.02	25.18	17.17
Rattus exulans	R4206	Syndecan-4A	β–actin	0.06	0.06	24.89	17.08
Rattus exulans	R4211	Syndecan-4A	β–actin	0.07	0.09	25.98	18.11
Rattus exulans	R4193	Syndecan-4A	β–actin	0.12	0.03	26.10	17.25
Rattus exulans	R4126	Syndecan-4A	β–actin	0.13	0.04	25.00	17.60
Rattus losea	R4260	Syndecan-4A	β–actin	0.22	0.04	24.76	16.85
Rattus losea	R4251	Syndecan-4A	β–actin	0.11	0.05	24.90	17.72
Rattus losea	R4245	Syndecan-4A	β-actin	0.15	0.02	26.91	18.78
Rattus losea	R4230	Syndecan-4A	β–actin	0.02	0.05	25.00	16.63
Rattus losea	R4203	Syndecan-4A	β–actin	0.06	0.05	25.27	18.51
Rattus tanezumi	R4009	Syndecan-4A	β–actin	0.14	0.03	27.02	20.64
Rattus tanezumi	R4272	Syndecan-4A	β–actin	0.18	0.09	26.43	17.61
Rattus tanezumi	R4121	Syndecan-4A	β–actin	0.05	0.07	26.31	19.01
Rattus tanezumi	R4003	Syndecan-4A	β–actin	0.06	0.05	29.02	21.57
Rattus tanezumi	R4005	Syndecan-4A	β–actin	0.11	0.07	26.65	20.89
Rattus tanezumi	R4182	Syndecan-4A	β–actin	0.19	0.04	23.42	16.89
Rattus tanezumi	R4018	Syndecan-4A	β-actin	0.17	0.06	31.43	21.30
Mus caroli	R4263	Syndecan-4A	β–actin	0.13	0.07	27.20	18.55
Mus cervicolor	R4363	Syndecan-4A	β–actin	0.07	0.03	32.05	22.12
Mus cooki	R4101	Syndecan-4A	β–actin	0.06	0.03	28.46	22.28
Mus cooki	R4317	Syndecan-4A	β–actin	0.16	0.01	26.49	19.25
Mus cooki	R4110	Syndecan-4A	β–actin	0.06	0.07	28.07	21.31
Nannomys mattheyi	I345	Syndecan-4A	β–actin	0.22	0.02	26.78	19.01

155

(iv) Calculation of relative expression levels of the target (syndecan-4) gene

The relative spleen-specific expression ratio R was calculated for syndecan-4 and for each of the possible combinations of pooled samples (see Materials and Methods). R values are shown in Table 5.4, several of which deviate significantly from R = 1 (i.e., R<0.8 and R>1.2). In particular, they show that rodents that are seropositive for *O. tsutsugamushi* express syndecan-4 significantly less than rodents that are seronegative for *O. tsutsugamushi*. Both (African and Asian) Murini and Asian Murini were found to express syndecan-4 at significantly lower levels than seronegative Rattini rodents. In contrast, no significant differences in syndecan-4 expression levels were identified between Rattini (reservoir lineage) and Murini (non-reservoir lineage) lineages when all samples (both seronegative and seropositive) were considered (Rattini ~ Murini). Not unexpectedly, Murini rodents displayed slightly elevated syndecan-4 transcript levels compared to seropositive Rattini rodents. Murini and Asian Murini rodents generally express syndecan-4 at levels that are lower than in seronegative Rattini rodents, while most of the seropositive Rattini rodents express syndecan-4 at less than twice the levels of their negative counterparts (i.e., seropositive *Bandicota* vs. seronegative *Bandicota*). For example, seropositive *R. tanezumi* express syndecan-4 at levels that are three times less than those of seronegative *R. tanezumi*.

Table 5.4. Relative expression ratios (R) for syndecan-4 following Pfaffl (2001). Values in bold deviate markedly from 1 (0.8<R<1.2). Expression levels of syndecan-4 in the control pool relative to the sample pool are indicated. + = rodents that are seropositive for the causative agent of scrub typhus, *O. tsutsugamushi*; - = rodents that are seronegative. See text for details.

Control pool/sample pool	R	Relative expression levels of syndecan-4
Asian rodents vs. additional taxa	1.15	Asian rodents ~ additional taxa
Rattini vs. remainder	1.10	Rattini ~ remainder
Rattini vs. Murini	1.06	Rattini ~ Murini
Rattini vs. Asian Murini	1.06	Rattini ~ Asian Murini
+ rodents vs. – rodents	**0.56**	+ rodents << - rodents
+ Rattini vs. – Rattini	**0.52**	+ Rattini << - Rattini
+ Bandicota vs. *- Bandicota*	**0.59**	*+ Bandicota << - Bandicota*

+ *B. indica* vs. - *B. indica*	0.80	+ *B. indica* ~ - *B. indica*
+ *B. savilei* vs. - *B. savilei*	**0.53**	+ *B. savilei* << - *B. savilei*
+ *Rattus* vs. – *Rattus*	**0.58**	+ *Rattus* << - *Rattus*
+ *R. exulans* vs. - *R. exulans*	1.18	+ *R. exulans* ~ - *R. exulans*
+ *R. tanezumi* vs. - *R. tanezumi*	**0.22**	+ *R. tanezumi* <<< - *R. tanezumi*
+ *R. losea* vs. - *R. losea*	**0.69**	+ *R. losea* < - *R. losea*
Murini vs. – Rattini	**0.70**	Murini < - Rattini
Asian Murini vs -Rattini	**0.71**	Asian Murini < - Rattini
Murini vs. +Rattini	**1.36**	Murini > + Rattini
Asian Murini vs. +Rattini	**1.37**	Asian Murini > + Rattini

DISCUSSION

Examples where closely related species within the same genus react differently to pathogenic infections have previously been reported. A case in point is work by Isaäcson *et al.* (1981) and Arntzen *et al.* (1991) who demonstrated a difference in susceptibility between sibling species of the *Mastomys* species complex (*M. natalensis* and *M. coucha*) to experimental infections with the bacteria *Yersinia pestis* (the cause of bubonic plague). Similar findings have been reported for *M. erythroleucus* and *M. natalensis* where the latter is a natural reservoir for Lassa fever, while the other is apparently not (Lecompte *et al.,* 2006). In similar fashion, murine recombinant β_3 integrins have been reported to be species-specific. For example, recombinant integrin prevents hantavirus infection and this may explain why *M. musculus* is not a suitable host for hantavirus, while other Murinae lineages (i.e., *B. indica, R. rattus, R. norvegicus, A. flavicollis* and *A. agrarius*; Mackow and Gavrilovskaya, 2009) are. These examples suggest that closely related species sometimes possess different intrinsic (genetic) properties that make them effective reservoir or non-reservoir hosts (Arntzen *et al.*, 1991; González Ittig and Gardenal, 2004; Tayeh *et al.,* 2010). These findings reinforce the suggestion that intrinsic factors may be responsible for the different epidemiological patterns (and relationships with the causative agent of scrub typhus, *O. tsutsugamushi*) between the Murini and Rattini.

Structural and sequence analysis of syndecan-4

Phylogenetic analysis

The monophyly of the tribes (Rattini, Murini, Arvicanthini, Praomyini and Apodemini) is well supported by the syndecan-4 sequence analysis, a finding consistent with Lecompte *et al.* (2008). However, the inter-tribal relationships are unresolved in all three methodologies utilised in the present study. This too is reflected by the Lecompte *et al.* (2008) study, although with greater nodal support. The relationships of the *Mus* subgenera within Murini were unresolved in the present study as with Lecompte *et al.* (2008); this is due to inconsistencies in the branching order of *N. mattheyi* and *M. musculus* between the three methods used herein. However, the MP analysis (100% BP) supports the monophyly of *Mus* and *N. mattheyi*—a finding that is consistent with Veyrunes *et al.* (2005).

Rattini as identified in this gene phylogeny corresponds to the *Rattus s. l.* group of species based on the most recent molecular studies using L1 amplification events, and mtDNA and nuclear gene sequences (Verneau *et al.*, 1997; Jansa *et al.,* 2006; Lecompte *et al.,* 2008; Pagès *et al.,* 2010). The placement of *M. surifer, N. fulvescens, L. edwardsi* and *B. berdmorei* and *B. bowersi* in the syndecan-4 analysis is completely consistent with Verneau *et al.* (1997), Lecompte *et al.* (2008) and Pagès *et al.* (2010). Some authors have proposed that *Bandicota* be included within the *Rattus s. s.* complex based on biochemical, immunological (Niethammer, 1977; Gemmeke and Niethammer, 1984) and L1 amplification events (Usdin *et al*., 1995; Verneau *et al*., 1997). The syndecan-4 sequence-based study weakly supports - but does not reject - the inclusion of *Bandicota* within *Rattus*. *Bandicota indica* and *B. savilei* formed a monophyletic clade in all three analyses with good nodal support (1.0 PP, 98% BP, 99% MP). *Rattus norvegicus* was identified as a sister species to the *Bandicota* representatives in the ML and BI analyses though with low nodal support (0.61

158

PP and 46% ML), while the MP analysis could not resolve the *Bandicota*/*Rattus*

relationships. Consequently, the phylogenetic analysis of syndecan-4 in this study using

three different methodological approaches does not contradict the inclusion of *Bandicota*

within *Rattus*.

Hapalomys delacouri was originally placed within *Micromys* by Musser and

Carleton (2005) on the basis of morphology. Lecompte *et al.* (2008) identified *Micromys* as

the sister lineage of the *Rattus s. l.* group on mitochondrial cytochrome b and nuclear IRBP

gene sequences, a finding echoed by Rowe *et al.* (2008) and Michaux *et al.* (2007). The

present study is the second phylogenetic investigation (the other was Pagès *et al.,* unpubl.) to

include this poorly documented Asian species, and its placement is consistent with the

outcome of the Pagès *et al.* (unpubl.) investigation. The latter was based on several

mitochondrial and nuclear gene sequences that indicate that *H. delacouri* is most likely not

related to *Micromys,* nor is it the furthest sister lineage to *Rattus s. l.* complex, but rather that

it represents a basal lineage within Murinae. However, future molecular investigations that

include all known and/or suspected tribal lineages of Murinae are required to more accurately

define the phylogenetic placement of *Hapalomys* within Murinae. Importantly, the present

investigation indicates that syndecan-4 is a robust phylogenetic marker for the study of

Murinae evolution, as the syndecan-4 gene phylogeny is highly congruent with the suspected

species phylogeny reported in the literature (i.e., Lecompte *et al.,* 2008; Pagès *et al.,* 2010).

*Host/*Orientia *interactions and syndecan-4 amino acid sequence/structure evolution*

The phylogenetic analysis of syndecan-4 sequences show that Murini constitutes a

monophyletic clade. This is important as *N. mattheyi* is included in the Murini, but the

species occurs outside of the distribution of scrub typhus and is consequently not subjected to

an *Orientia*-mediated selection regime. This appears to refute the possibility that only Asian

Murini rodents have evolved a protective syndecan-4 genotype following strong *Orientia*-driven selection that would have provided a mechanism against typhus infection. If this had been the case, one would have expected *N. mattheyi* – believed to have never been exposed to scrub typhus – not to carry the mutation, and consequently group outside the Murini in the syndecan-4-based phylogeny. As Rattini forms a monophyletic clade, it is also feasible that Rattini rodents (and not those in Murini) may have evolved an adaptive mutation in the syndecan-4 receptor that may confer a degree of protection against the detrimental effects of *Orientia*. Since all Rattini rodents represent effective reservoirs of scrub typhus infection, one would expect all Rattini representatives to possess this adaptive mutation in their syndecan-4 gene sequences, as was observed here.

Although the good agreement between species phylogeny and the syndecan-4 gene topology does not allow the unambiguous identification of a strong *Orientia*-mediated selective regime acting on the syndecan-4 genotype of murine rodents (i.e., differentiating between *Orientia*-mediated selection and regular phylogenetic divergence), it does not refute it. It is, however, possible that selection may be acting on the primary structure of the syndecan-4 protein, as a number of amino acid characters were identified at the primary sequence level that clearly differentiate the Murini and Rattini lineages in this study. Three amino acid characters were identified that are specific to Rattini (refer to Figure 5.5); similarly, three Murini specific amino acid changes were identified, of which two differentiate *Mus* from *N. mattheyi*. The two amino acid changes that are Asian *Mus* specific (excluding *N. mattheyi*; see Figure 5.5), alanine-114 and methionine-124 are important, as it is possible that they may confer an adaptive advantage to *Mus* that would allow Asian mice to escape *Orientia* infection in nature. This would account for an *Orientia*-driven selection regime in *Mus* exposed to scrub typhus and their absence in closely related species, such as *N. mattheyi*, that are not subjected to selection.

160

In addition, several of the amino acid characters correspond to indels (insertions/deletions) in Rattini (deletion of aspartic acid-55 and insertion of aspartic acid-122) and Murini (deletion of amino acid residues 49-52). Indels were not identified in any of the other taxa included in the analysis (i.e., Arvicanthini, Apodemini and Praomyini). Consequently it seems likely that these may have a functional impact on the Rattini vs. Murini syndecan-4 proteins. In support of this, one of the nest sites identified in Murini (in the *M. musculus* 3D protein model), that extends from glycine-48 to glutamic acid-53 (glycine-48 to glutamic acid-56 in the ClustalW syndecan-4 amino acid sequence alignment, Appendix 2), overlaps with amino acid residue 55 (aspartic acid) which is deleted in Rattini. Moreover, one of the *R. norvegicus* nest sites (aspartic acid 45 to leucine 49 in *R. norvegicus*), overlaps with amino acid residue 49 that corresponds to the deletion of amino acid residues 49-52 in Murini. Murini must have evolved this deletion before *Nannomys* diverged from extant *Mus* (~6.6 mya, Lecompte *et al.*, 2008), which explains why *N. mattheyi* - which occurs outside the distribution of scrub typhus - shares this deletion with Asian Murini. However, one would anticipate that were this the case, the *Orientia*-mediated selection would have waned once *Nannomys* diverged from extant *Mus* ~6.6 mya, as it no longer co-occurs with scrub typhus. As African *N. mattheyi* has nearly the same syndecan-4 genotype as Murini, and shares almost all the amino acid changes that differentiate Murini from Rattini (except for two, see above). It is probable that the indels identified in the syndecan-4 protein sequences are not undergoing strong *Orientia*-mediated selection (or at the very least this is not the only selective pressure the indels are experiencing).

That the indels may be undergoing other selective pressures not linked to *Orientia* is supported by the different nest sites identified in *B. savilei* and *R. norvegicus* (both effective reservoirs of *Orientia*). Syndecan-4 also acts as a receptor for a variety of other viruses, i.e., foot-and-mouth disease and cytomegalovirus (see Kim *et al.*, 2004) and their ranges

encompass both Asia and Africa. For example, foot-and-mouth disease is carried by rats (i.e., Epoke and Coker, 1991), while cytomegalovirus is inherently species-specific with both mouse (*M. musculus*) and rat (*R. norvegicus*) cytomegalovirus strains. It is believed that cytomegaloviruses have evolved with their hosts, and this protracted co-evolution has shaped their genetic content (Teterina *et al.,* 2009). It's possible therefore that viruses such as that which cause foot-and-mouth disease and the cytomegalovirus may, rather than *O. tsutsugamushi* (or, alternatively, along with *Orientia*), be influencing syndecan-4 structural evolution.

In addition, no amino acid differences could be observed between the seropositive (n = 15, Appendix 1) and seronegative (n = 19, Appendices 1 and 2) Rattini specimens. For example, the *Bandicota* species, *B. indica* and *B. savilei*, have identical amino-acid chains, as do members of *Rattus,* namely *R. losea, R. exulans, R. tanezumi* and *R. norvegicus.* Thus primary amino acid sequences do not differentiate between seropositive and seronegative individuals and cannot, at present, be linked with increased susceptibility to *Orientia* infection in seropositive specimens. This implies that seronegative animals may have the same probability of becoming infected as those that are seropositive but simply have not come into contact with *O. tsutsugamushi.* Alternatively, these seronegative rodents may have even been previously infected with *Orientia*, as only the current serological status of the rodents at the point of collection is known.

Gene expression

Candidate gene approaches are powerful tools for examining the genetic architecture of complex traits (Kwon and Goate, 2000; Zhu and Zhao, 2007; Lee *et al.,* 2009 among others). However, these are often limited by the requirement of *a priori* knowledge of the biochemical, physiological or functional aspects of possible candidate genes (Zhu and

Zhao, 2007 for review). Fortunately earlier work by Longley *et al.* (1999) and Kim *et al.* (2004) provided the framework for the selection of syndecan-4 as a candidate gene for scrub typhus infectivity.

Importantly, Murini and seropositive Rattini rodents both showed significant reductions of syndecan-4 expression levels relative to seronegative Rattini rodents (refer to Table 5.4; seropositive Rattini<(all or Asian) Murini<seronegative Rattini). No significant differences in syndecan-4 expression were observed when comparing representatives of the Rattini and Murini clades to each other (Rattini~Murini). As a consequence it seems possible that the underexpression of syndecan-4 in comparison to seronegative Rattini may protect Murini against *Orientia*-infection (non-reservoir), whereas in seropositive Rattini species (reservoir), it may be a result of the host-immune response to infection (see below).

The underexpression of syndecan-4 in Murini relative to seronegative Rattini may be a good explanation for the poor host-pathogen affinity observed between Murini (apparent non-reservoir) and scrub typhus. It has been suggested that *O. tsutsugamushi* requires recognition by a cellular receptor, syndecan-4, for its cellular attachment (Kim *et al.,* 2004). The attachment of *O. tsutsugamushi* to host cell plasma membranes is a crucial step in the infection and uptake into the cell (Ihn *et al.,* 2000; Seong *et al.,* 2001; Kim *et al.,* 2004) and significantly reduced levels of syndecan-4 transcripts at the cell surface may reduce/inhibit the uptake of *O. tsutsugamushi* into the host cell, and/or diffusion between cells. This is supported by findings that rat embryo fibroblast cells (REF-syn4 cell line), which overexpress syndecan-4 prior to infection, have increased *O. tsutsugamushi* infectivity (Kim *et al.,* 2004; syndecan-4 overexpression has also been linked with the increased infectivity of host cells by *Neisseria gonorrhoeae,* Freissler *et al.,* 2000). Conversely, those expressing reduced levels of syndecan-4 (REF-syn4AS cell line) exhibited diminished infectivity

following infection with *O. tsutsugamushi* (Kim *et al.,* 2004; Fears and Woods, 2006 for review). Therefore *O. tsutsugamushi* infectivity appears to be dependent on syndecan-4 and its' expression level on the cell surface. As a consequence, Murini are rarely seropositive for scrub typhus in the wild as the reduced syndecan-4 transcripts may limit entry of the bacterium into the host cells of wild Asian Murini undergoing *Orientia*-mediated selection. To date, only a single *M. caroli* seropositive specimen (single specimen was tested) has been detected during epidemiological surveys of Thailand (Wangroongsarb *et al.,* 2008) and only four seropositive *M. musculus* specimens (out of 37) have been reported in China in two separate surveys (Liu *et al.,* 2003; Yang *et al.,* 2008). No information was provided on whether the wild seropositive mice in the latter studies displayed gross symptoms of infection (i.e., ruffled fur and reduced activity). Since inbred strains of mice experimentally infected with *Orientia* are reported to become significantly moribund by day 10 following infection, and begin to die by day 16 (i.e., Van Peenen *et al.,* 1977; Ewing *et al.,* 1978; Frances *et al.,* 2000; Koh *et al.,* 2004). Consequently *Orientia* infection in wild seropositive mice may also be lethal (i.e., mice have reduced infectivity as a result of decreased syndecan-4 expression, however, if infection occurs it is highly deleterious).

The different expression patterns between seropositive and seronegative Rattini may be explained by two hypotheses: (i) the variable expression levels may make the one group (seropositive, underexpression) more susceptible to *Orientia* infection than the other (seronegative, overexpression), thereby making syndecan-4 underexpression a factor favouring infection in Rattini, or (ii) the reduced syndecan-4 expression levels in seropositive animals are a consequence of *Orientia*-infection in the host. As underexpression of syndecan-4 prior to infection was shown to reduce infectivity (Kim *et al.,* 2004, see above), it seems unlikely that underexpression in Rattini (hypothesis i) would predispose towards *Orientia* infection (i.e., seropositive status). Moreover, as seropositive and seronegative

164

rodents are indistinguishable based on amino acid chains (they have similar structure and function, see above), and seronegative rodents may display the same trend of reduced syndecan-4 transcription following infection, it is likely that Rattini may have an immune response that modulates the expression of syndecan-4 upon infection (hypothesis ii). In this regard it is noteworthy that seropositive Rattini rodents are able to persist in the wild for long periods of time (Walker *et al.,* 1973; Traub *et al.,* 1975; Van Peenen *et al.,* 1977; Frances *et al.,* 2000) and display no ill-effects to *Orientia* infection (no significant difference between uninfected and infected animals; Van Peenen *et al.,* 1977; Takahashi *et al.,* 1990; Strickman *et al.,* 1994; Frances *et al.,* 2000). Consequently, the reduced levels of syndecan-4 may be a host-defence mechanism that influences the intensity of infection and the degree of debilitation inflicted on the host by the *O. tsutsugamushi* bacterium. Seropositive Rattini animals that, for example, overexpress syndecan-4 are possibly not detected in the wild as the *O. tsutsugamushi* infection is lethal in these instances.

Several studies have suggested that syndecan-4 may play an important role in host-defence mechanisms (Ishiguro *et al.,* 2002; Smith *et al.,* 2006a, b). For example, it has been reported that syndecan-4 expression is up-regulated in response to *H. pylori* infection (human and non-human primates are the principal hosts, Whary and Fox, 2004; Smith *et al.,* 2006a, b), and that the overexpression of syndecan-4 leads to accelerated cell surface shedding of syndecan-4 ectodomains thereby serving as a possible protective mechanism (Schmidt *et al.,* 2005; Smith *et al.,* 2006a, b). As the syndecan-4 shed from epithelial cell surfaces may act as "decoy receptors" for bacteria, resulting in decreased colonisation of the bacteria in the host (Smith *et al.,* 2006a, b). Syndecan-4 plays an important role in both cell migration and invasion (Charni *et al.,* 2009) and *O. tsutsugamushi* can invade a broad range of host cells through intracellular invasion and migration process (i.e., epithelial cells, fibroblasts and endothelial cells; Ihn *et al.,* 2000; Kim *et al.,* 2002, 2004). Thus the reduction of cell

receptors (underexpression) rather than increased "decoy receptors" (overexpression) in seropositive Rattini reduces the number of receptors available for bacterial attachment. This prevents extensive cell to cell spreading (or minimally it limits the degree/rate of infection) and, as a consequence, Rattini rodents do not display deleterious symptoms of scrub typhus infection. In similar fashion, the down-regulation of syndecan-4 has been reported to play a role in heptoma (hepatocellular carcinoma) pathogenesis by reducing heptoma cell invasion and migration (Charni *et al.,* 2009).

Although the detection of antibodies (i.e., seropositive Rattini herein) is associated with infection, it does not necessarily indicate active infection (Wangroongsarb *et al.,* 2008). For example, active *O. tsutsugamushi* has been isolated up to eight weeks after infection in *R. rattus*, while antibodies have been detected up to 10 and 11 months after infection in *R. rattus* and *R. mindanesi* (Van Peenen *et al.,* 1977; Frances *et al.,* 1999). It has been reported that antibodies to *O. tsutsugamushi* neutralize infection by hampering the entry mechanisms of the bacterium and increase the uptake of *Orientia* by phagocytes (i.e., polymorphonuclear leukocytes) and macrophages (see Seong *et al.,* 2001 for review). Thus it is possible that the down-regulation of syndecan-4 may persist together with the prolonged presence of *Orientia* antibodies (i.e., 10 to 11 months after infection) following active infection, consequently limiting or preventing reinfection with scrub typhus.

In conclusion, although the structural analyses identified lineage-specific differences of syndecan-4 primary structures between Rattini and Murini none of these could be linked unambiguously to reservoir or non-reservoir scrub typhus status. More convincingly, however, the gene expression analysis showed an important link between underexpression of syndecan-4 in Murini and seropositive Rattini rodents compared to seronegative Rattini. Although speculative, it seems possible that Murini are ineffective reservoirs for scrub typhus

as the reduced syndecan-4 transcripts limit bacterium access to the host's cells. On the contrary, reduced syndecan-4 transcription following initial *Orientia*-infection in Rattini may contribute to a lower infectivity (non-lethality) and hence to the reservoir status of Rattini—findings that underscore the importance of syndecan-4 in the *O. tsutsugamushi* infection process.

CHAPTER 6

CONCLUDING REMARKS

CHROMOSOME EVOUTION IN *RATTUS SENSU LATO*

Cytotaxonomy

Information on chromosomes of the *Rattus s. l.* complex has been largely restricted to unbanded karyotypes (see Table 1.1; Chapter 1). For example, the last comprehensive cytogenetic investigation of this complex was that of Gadi and Sharma (1983) in which the cytogenetic relationships within *Rattus, Nesokia* and *Bandicota* genera were assessed using conventional techniques. Therefore the present investigation is the most comprehensive study to update the available karyotypic data and to provide G- and/or C-band karyotypes for several species for which data are lacking, namely *B. savilei, L. edwardsi, M. surifer, B. berdmorei, B. bowersi, N. fulvescens* and *H. delacouri*.

Although the chromosomal phylogeny generated in this study is broadly consistent with that based on sequence data (i.e., Usdin *et al.*, 1995; Verneau *et al.*, 1997, 1998; Chaimanee and Jaeger, 2001; Lecompte *et al.,* 2008; Pagès *et al.,* 2010, syndecan-4 gene tree – this study, Chapter 5), there were two problematic relationships which could not be resolved by the present study: the paraphyly of *Rattus s. s.* and the placement of *H. delacouri* in Rattini. Usdin *et al*. (1995) and Verneau *et al.* (1997, 1998) had proposed that *Bandicota* (represented here by *B. savilei*) should be placed within the *Rattus* assemblage. However, analysis of repetitive elements (rat sat I DNA family and NORs, see Chapter 4) did not support the paraphyly of *Rattus s. s*., while these relationships were uncertain in the Zoo-FISH analysis (i.e., due to an unresolved polytomy between *B. savilei* and *Rattus s. s.*, see

168

Chapter 3). These findings are reflected in the recent comprehensive taxonomic revision of this complex (by Pagès *et al.* 2010), where the hypothesised branching order of *Bandicota* within the *Rattus* assemblage could not be rejected. Further sampling to include representatives of all the *Rattus* species groups, as defined by Musser and Carleton (2005), is required before any final conclusions can be drawn on the paraphyly of *Rattus*.

A second point of interest concerns *H. delacouri* which was placed within the *Micromys* division by Musser and Carleton (2005) on the basis of morphological characteristics. Lecompte *et al.* (2008) show that *Micromys* is the furthest sister lineage of the Rattini (no representative of *Hapalomys* was included in their study) leading these authors to suggest that *Hapalomys* should be considered Murinae *incertae sedis*. Recent sequence-based phylogenies have placed *H. delacouri* basal to Murinae (Pagès *et al.* upubl.; present study – see Chapter 5) contrary to the suggested placement of *Hapalomys* within the *Micromys* division, and basal to *Rattus s. l.* Surprisingly, the chromosomal data indicated a close association/similarity between *H. delacouri* and *B. savilei* as both species share the same X-autosome translocation (see Chapter 3), a chromosome rearrangement that should not be prone to convergence (Rokas and Holland, 2000). However, neither sequence, morphological nor the remaining chromosomal data (i.e., entirely heterochromatic short arms in *H. delacouri*, see Chapters 2 and 3) supports a close association between the species suggesting that the X-autosome translocation represents a convergent event. Even if Lecompte *et al.* (2008) and Musser and Carleton (2005) findings were accurate, and *H. delacouri* was basal to *Rattus s. l.*, the X-autosome translocation is not conserved across successive nodes (*M. surifer, L. edwardsi, N. fulvescens, B. berdmorei* and *B. bowersi*) to *B. savilei* as would be expected. Further sampling to include representatives of all tribal lineages of Murinae is required before any final conclusions can be drawn about the phylogenetic placement of *Hapalomys* within Murinae.

Mode and tempo of chromosome evolution

Comparative chromosome painting experiments (using *R. norvegicus, M. surifer* and *R. rattus* chromosome paints) identified fusion and fission events that shaped the karyotype evolution of the taxa under investigation. This finding is not unexpected as these two types of rearrangements are common in mammals (Kolnicki, 2000; Ferguson-Smith and Trifonov, 2007). However, as evidenced by the present study, changes to the short arms of many autosomes (thought to be pericentric inversions) predominated in the karyotype evolution of the *Rattus s. l.* complex. The presence of pericentric inversions was only unambiguously confirmed in three pairs of rat chromosomes (RNO 1, 12 and 13 orthologs) through BAC-mapping (Chapter 3). Based on these findings, the unlikelihood of neocentromere formation within Rattini, and the principle of parsimony (a shift and transposition which do not involve a change in gene order, would require three breaks, a pericentric inversion only two), the other changes that involved short arm modification in the various species were similarly considered to result from pericentric inversions (see Chapter 3 - Materials and Methods, pp 48-50). Importantly, however, the presence of numerous intrachromosomal rearrangements within *Rattus s. l.* is consistent with reports that the *Rattus* genome has undergone considerably more intrachromosomal rearrangements than *Mus* (Zhao *et al.,* 2004; Mlynarski *et al.,* 2010).

Arguments that chromosome rearrangements may affect reproductive success and thus play a role in the emergence of new species (speciation) have increasingly met with criticism (reviewed in Brown and O'Neill, 2010) with newer models suggesting that chromosome rearrangements are not reducing fitness, but may in fact be reducing gene flow by curbing recombination in these regions, i.e., the so called suppressed-recombination models. In this context, and as mentioned in Chapter 3, inversions are believed to play an

important role in conferring an adaptive advantage, given that these mutations may act as a genetic filter and ultimately lead to reproductive isolation (e.g., Noor *et al.,* 2001; Hoffman *et al.,* 2004; Ayala and Coluzzi, 2005). However, the suppressed-recombination models, as with most theories of chromosome speciation, have also met with criticism. For example, rearranged chromosomes between human and chimpanzees showed no significant differences in protein or gene divergence (Zhang *et al.*, 2004). In addition, it has been reported that sterile hybrids were produced even when the parental species displayed similar chromosomal complements (Coyne and Orr, 1998; Brown and O'Neill, 2010).

The present study does not to support the chromosomal theory of speciation since the mode of chromosome evolution within *Rattus s. l.* was determined to be one of mainly constrained change (i.e., very few interchromosomal rearrangements have occurred during the karyotypic evolution of this complex, Chapter 3). Consequently it seems that chromosome rearrangements were not as vital in the speciation process of these rodents as they have been in other murids (i.e., *Nannomys, Mus, Coelomys* and *Pyromys* subgenera of the genus *Mus*; Veyrunes *et al.,* 2006). Similar findings have been reported where mammalian orders may contain taxa that display slower rates of chromosome evolution (i.e., Sciuridae family among muroid rodents, Richard *et al.,* 2003b; Stanyon *et al.,* 2003; Li *et al.,* 2004, 2006). There are no persuasive explanations for the variation in rates of chromosome evolution between various mammalian orders and taxa, although contributory factors such as differing mutation rates, environmental effects, population size, mobile elements, retroviruses as well as staggered double strand breaks have been mooted (Casals and Navarro, 2007; Ferguson-Smith and Trifonov, 2007; Ranz *et al.,* 2007; Brown and O'Neill, 2010).

In addition, comparisons between partially or fully sequenced mammalian genomes have shown that evolutionary breakpoints are not randomly distributed in the genome, but

171

rather cluster in hotspots augmented with tandem repeats and these areas often associate with fragile sites (i.e., Slamovits and Rossi, 2002; Murphy *et al.,* 2005; Robinson *et al.,* 2006; Ruiz-Herrera *et al.,* 2006; Ruiz-Herrera and Robinson, 2007, Ruiz-Herrera *et al.,* 2008; Adega *et al.,* 2009). As mentioned elsewhere (see Chapters 1 and 4), repetitive elements have been linked to chromosome rearrangements and are considered important markers for genome evolution. The pattern of slow chromosomal evolution in *Rattus s. l.* is also reflected in the analysis of repetitive elements (rat sat I DNA, telomeres and NORs) which were found to be largely conserved with respect to distribution and location (Chapter 4). This underscores the usefulness of the rat as index species in comparative genomic studies since the mouse (*M. musculus*) has an extremely rearranged karyotype (i.e., Romanenko *et al.,* 2006, 2007) and is highly divergent even when compared to other closely related Murini rodents (Matsubara *et al.,* 2003; Veyrunes *et al.,* 2006). The rat genome is largely conserved (intrachromosomal rearrangements predominate among species) and many rat fragments are similar to human prompting observations that the rat genome is structurally closer to the human genome than the mouse (Zhao *et al.,* 2004).

RODENT INTERACTIONS WITH SCRUB TYPHUS

Sequence and expression levels of syndecan-4 (see Chapter 5) were investigated to determine the mechanisms behind the host interactions with scrub typhus. Although the findings of the structural analysis proved inconclusive, a correlation was nonetheless identified regarding underexpression of syndecan-4 in Murini as well as seropositive Rattini specimens in comparison to seronegative Rattini. The pattern of syndecan-4 underexpression is thought to underpin the divergent responses of the Rattini and Murini lineages to scrub typhus infection (Chapter 5). While clearly tentative, these findings should be tested using laboratory investigations (controlled infections of the apparent reservoir and non-reservoir

taxa) and longitudinal field studies in order to establish the precise effect of infection on populations. Care will need to be taken to account for the fact that laboratory studies may not reflect the transmission of scrub typhus and infection of hosts in the wild. An interesting line of further investigation concerns observations (Chapter 5) that the deleterious effects of *O. tsutsugamushi* on the Rattini hosts may be moderated following reductions of syndecan-4 transcripts in host cells (seropositive rodents survive for long periods in the wild; Walker *et al.,* 1973; Traub *et al.,* 1975; Van Peenen *et al.,* 1977; Frances *et al.,* 2000). Monitoring of syndecan-4 expression levels during scrub typhus infections in humans to determine whether similar or divergent expression patterns of syndecan-4 are obtained, may provide additional insights into human immune responses to *Orientia* infection.

Interestingly, rats are generally more frequent reservoirs of disease than mice since pathogens commonly have a broad spectrum of hosts in rats whereas in mice they are often very species-specific (see Table 2 in Mills, 2006). For example, leptospirosis and toxoplasmosis can be carried by several *Rattus* species, while only *M. musculus* (of all Murini tested) is reported to act as a host (Kijlstra *et al.,* 2008; Taylor *et al.,* 2008). Moreover, *M. musculus* rodents are not considered natural reservoirs for hantaviruses even though they exhibit the hallmarks of hantavirus infection when infection is induced experimentally (Araki *et al.,* 2003). These findings are supported by serological studies where wild *M. musculus* rodents were seronegative even when co-occurring with other wild rodents that demonstrated high levels of endemic infection (Kantakamalakul *et al.,* 2003; Zuo *et al.,* 2004). The latter finding (hantavirus infection and apparent non-reservoir status of *M. musculus* in the wild) would appear to mimic the interaction of Murini with scrub typhus shown in the present study. *Mus* interactions with hantaviruses may consequently also be a result of intrinsic factors which is supported by reports that have linked recombinant β_3 integrins to the poor hantavirus reservoir status of *M. musculus* (reviewed in Mackow and Gavrilovskaya, 2009).

β_3 integrins mediate the cellular entry of hantavirus (Gavrilovskaya *et al.,* 1998) and, interestingly, syndecan-4 is associated with focal adhesions that comprise either β_1 or β_3 integrin subunits (Woods and Couchman, 1994). Recent studies have identified a link between the signalling networks that are facilitated by syndecan and integrin adhesion receptor families (Filla *et al.,* 2006; Morgan *et al.,* 2007). In addition, both stress fiber formation and focal adhesions require the interaction of integrins and syndecan-4, since neither is individually sufficient (van der Flier and Sonnenberg, 2001 and Morgan *et al.,* 2007). It seems plausible, therefore, that the underexpression of syndecan-4 together with recombinant β_3 integrin may be linked with the poor host-pathogen affinity observed between Murini and scrub typhus, as well as hantaviruses. As tantalizing as this may appear, further experimental confirmation is required since the precise role of both β_3 integrin and syndecan-4 in infection of host cells by bacteria (i.e., *Orientia*) and viruses (i.e., hantavirus) has not yet been determined.

The increased longevity of seropositive Rattini rodents is beneficial to both the host, and the invading *O. tsutsugamushi* bacterium, as a shorter-host life span would not aid the transmission of the pathogen (Mills and Childs, 1998; Smith *et al.,* 2006a, b; Carrara *et al*., 2007; Prentice and Rahalison, 2007). Rattini rodents, which serve as effective reservoirs for numerous pathogens and generally show no sign of illness (i.e., Van Peenen *et al.,* 1977; Takahashi *et al.,* 1990; Strickman *et al.,* 1994; Mills, 2006), may have coevolved with these pathogens. In future, investigations of how scrub typhus infection and underexpression of syndecan-4 affects receptor function, endothelial cell responses, signalling pathways and immune response may be fundamental to understanding the pathogenesis of *O. tsutsugamushi* in their hosts, and may aid identification of suitable therapeutic targets for scrub typhus infection.

REFERENCES

Abe, H. 1983. Variation and taxonomy of *Niviventer fulvescens* and notes on *Niviventer* group of rats in Thailand. *J Mamm Soc Jpn* **9**: 151-161.

Acosta, M.J., Marchal, J.A., Martínez, S., Puerma, E., Bullejos, M., Díaz la de Guardia, R., Sánchez, A. 2007. Characterisation of the satellite DNA Msat-160 from the species *Chionomys nivalis* (Rodentia, Arvicolinae). *Genetica* **130**: 43-51.

Acosta, M.J., Marchal, J.A., Fernández-Espartero, C., Romero-Fernández, I., Rovatsos, M.T., Giagia-Athanasopoulou, E.B., Gornung, E., Castiglia, R., Sánchez, A. 2010. Characterisation of the satellite DNA Msat-160 from species of *Terricola* (*Microtus*) and *Arvicola* (Rodentia, Arvicolinae). *Genetica* **138**: 1085-1098.

Adega, F., Chaves, R., Guedes-Pinto, H. 2005. Chromosome restrictive enzyme digestion in domestic pig (*Sus scrofa*) constitutive heterochromatin arrangement. *Genes Genet Syst* **80**: 49-56.

Adega, F., Chaves, R., Kofler, A., Krausman, P.R., Masabanda, J., Wienberg, J., Guedes-Pinto, H. 2006. High-resolution comparative chromosome painting in the Arizona collared peccary (*Pecari tajacu*, Tayassuidae): a comparison with the karyotype of pig and sheep. *Chromosome Res* **14**: 243-251.

Adega, F., Chaves, R., Guedes-Pinto, H. 2008. Suiformes orthologous satellite DNAs as a hallmark of collared and white-lipped peccaries (Tayassuidae) evolutionary rearrangements. *Micron* **39**: 1281-1287.

Adega, F., Guedes-Pinto, H., Chaves, R. 2009. Satellite DNA in the karyotype evolution of domestic animals—clinical considerations. *Cytogenet Genome Res* **126**: 12-20.

Aitman, T.J., Critser, J.K., Cuppen, E., *et al*. 2008. Progress and prospects in rat genetics: a community view. *Nat Genet* **40**: 516-522.

Ambrish, R., Kucukural, A., Zhang, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**: 725-738.

Amor, D.J., Bentley, K., Ryan, J., Perry, J., Wong, L., Slater, H., Choo, K.H. 2004. Human centromere repositioning "in progress". *Proc Natl Acad Sci USA* **101**: 6542-6547.

Andrades-Miranda, J., Zanchin, N.I., Oliveira, L.F., Langguth, A.R., Mattevil, M.S. 2002. (T2AG3)n telomeric sequence hybridization indicating centric fusion rearrangements in the karyotype of the rodent *Oryzomys subflavus*. *Genetica* **114**: 11-16.

Aplin, K.P., Chesser, T., Have, J.T. 2003a. Evolutionary biology of the genus *Rattus*: profile of an archetypal rodent pest. In: Grant R. Singleton, Lyn A. Hinds, Charles J. Krebs and Dave M. Spratt, 2003. *Rats, mice and people: rodent biology and management*. ACIAR Monograph No. 96. Australian Centre for International Agricultural Research, pp 487-498.

Aplin, K.P., Frost, A., Tuan, N.P., Lan, L.P., Hung, N.M. 2003b. Identification of rodents of the genus *Bandicota* in Vietnam and Cambodia. In: Grant R. Singleton, Lyn A. Hinds, Charles J.

Krebs and Dave M. Spratt, 2003. *Rats, mice and people: rodent biology and management*. ACIAR Monograph No. 96. Australian Centre for International Agricultural Research, pp 531-540.

Apweiler, R. Attwood, T.K., Bairoch, A., *et al.* 2001. InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* **29**: 37-40.

Araki, K., Yoshimatsu, K., Lee, B-H., Kariwa, H., Takashima, I., Arikawa, J. 2003. Hantavirus-specific CD8-T-cell responses in newborn mice persistently infected with Hantaan virus. *J Virol* **77**: 8408-8417.

Arntzen, L., Wadee, A.A., Isaäcson, M. 1991. Immune response of two *Mastomys* sibling species to *Yersinia pestis*. *Infect Immun* **59**: 1966-1971.

Ashley, T. 2002. X-Autosome translocations, meiotic synapsis, chromosome evolution and speciation. *Cytogenet Genome Res* **96**: 33-39.

Audoin-Rouzeau, F. 1999. The black rat (*Rattus rattus*) and the plague in ancient and medieval Western Europe. *Bull Soc Pathol Exot* **92**: 422-426.

Avirachan, T.T., Mehta, H.J., Sugandhi, M.R. 1971. Chromosomes of the genus *Bandicota*. *Mamm Chrom Newsl* **12**: 62-67.

Avner, D., Heard, E. 2001. X-chromosome inactivation: counting, choice and initiation. *Nat Genet* **2**:59-67.

Ayala, F.J., Coluzzi, M. 2005. Chromosome speciation: humans, Drosophila, mosquitoes. *Proc Natl Acad Sci USA* **102**: 6535-6542.

Badenhorst, D., Herbreteau, V., Chaval, Y., Pagès, M., Robinson, T.J., Rerkamnuaychoke, W., Morand, S., Hugot, J-P., Dobigny, G. 2009. New karyotypic data for Asian rodents (Rodentia, Muridae) with the first report of B-chromosomes in the genus *Mus. J Zool (Lond)* **279**: 44-56.

Baker, R.J., Barnett, R.K. 1981. Karyotypic Orthoselection for Additions of Heterochromatic Short Arms in Grasshopper Mice (*Onychomys*: Cricetidae). *Southwest Naturalist* **26**: 125-131.

Barch, M.J. 1997. *The AGT Cytogenetics Laboratory Manual*, 3rd edition. Philadelphia: Lippincott-Raven Publishers, pp 301.

Baskevich, M.I., Kuznetsov, G.V. 1998. Cytogenetic differentiation of species from the genus *Rattus* (Rodentia, Muridae) from the Dalat Plateau in southern Vietnam. *Zool Zh* **77**: 1321-1328.

Baskevich, M.I., Kuznetsov, G.V. 2000. Preliminary chromosomal results of *Niviventer* Marshall, 1976 (Mammalia, Rodentia, Muridae) from the Dalat plateau in Southern Vietnam. In: *Isolated vertebrate communities in the tropics. Proceedings of the 4th International Symposium Bonn Germany, Bonner Zoologische Monographien* **46**: 351-356. Rheinwald, G. (Eds).

Bastos, A.D.S., Nair, D., Taylor, P.J. *et al.* 2011. Genetic monitoring detects an overlooked cryptic species and reveals the diversity and distribution of three invasive *Rattus* congeners in South Africa. *BMC Genetics* **12**: 26.Baxter, J. 1996. The Typhus group. *Clin Dermatol* **14**: 271-278.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**: 573-580.

Bernard, P.S., Wittwer, C.T. 2000. Homogeneous amplification and variant detection by fluorescent hybridization probes. *Clin Chem* **46**: 147-148.

Blanford, W.T. 1891. *The fauna of British India: Mammalia*, Delhi Publication Division, pp 1-617.

Bolzan, A.D., Bianchi, M.S. 2006. Telomeres, interstitial telomeric repeat sequences, and chromosomal aberrations. *Mut Res* **612**: 189-214.

Bonvicino, C., D'Andrea, P., Borodin, P. 2001. Pericentric inversion in natural populations of *Oligoryzomys nigripes* (Rodentia, Sigmodontinae). *Genome* **44**: 791-796.

Borchert, J.N., Mach, J.J., Linder, T.J., Ogen-Odoi, A., Angualia, S. 2007. Invasive rats and bubonic plague in Northwest Uganda. In: *Managing Vertebrate Invasive Species: Proceedings of an International Symposium*. Fort Collins, CO.

Bosier, P., Rahalison, L., Rasolomaharo, M., Ratsitorahina, M., Mahafaly, M., Razafimahefa, M., Duplantier, J., Ratsifasoamanana, L., Chanteau, S. 2002. Epidemiologic features of four successive annual outbreaks of bubonic plague in Mahajanga, Madagascar. *Emerg Infect Dis* **8**: 311-316.

Boumil, R.M., Lee, J.T. 2000. Forty years of decoding the silence in X chromosome inactivation. *Hum Mol Genet* **10**: 2225-2232.

Bourque, G., Pevzner, P.A., Tesler, G. 2004. Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome Res* **14**: 507-516.

Britton-Davidian, J., Catalan, J., da Graça Ramalhinho, M., Ganem, G., Auffray, J.C., Capela, R., Biscoito, M., Searle, J.B., da Luz Mathias, M. 2000. Rapid chromosomal evolution in island mice. *Nature* **403**: 158.

Brockdorff, N. 1998. The role of Xist in X-inactivation. *Curr Opin Genet Dev* **8**: 328-333.

Brown, J.D., O'Neill, R.J. 2010. Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annu Rev Genomics Hum Genet* **11**: 291-316.

Bustin, S.A. 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* **29**: 23-39.

Cafasso, D., Cozzolino, S., De, L.P., Chinali, G. 2003. An unusual satellite DNA from *Zamia paucijuga* (Cycadales) characterised by two different organizations of the repetitive unit in the plant genome. *Gene* **311**: 71-79.

Carey, D.J. 1997. Syndecans: multifunctional cell-surface co-receptors. *Biochem J* **327**: 1-16.

Carrara, A., Coffey, L.L., Aguilar, P.V., Moncayo, A.C., Travassos Da Rosa, A.P.A., Nunes, R.T., Tesh, R.B., Weaver, S.C. 2007. Venezuelan equine encephalitis virus infection of cotton rats. *Emerg Infect Dis* **13**: 1158-1165.

Casals, F., Navarro, A. 2007. Inversions: the chicken or the egg? *Heredity* **99**: 479-480.

Castiglia, R., Garagna, S., Merico, V., Oguge, N., Corti, M. 2006. Cytogenetics of a new cytotype of African *Mus* (subgenus *Nannomys*) *minutoides* (Rodentia, Muridae) from Kenya: C- and G-banding and distribution of (TTAGGG) n telomeric sequences. *Chromosome Res* **14**: 587-594.

Cavagna, P., Stone, G., Stanyon, R. 2002. Black rat (*Rattus rattus*) genomic variability characterised by chromosome painting. *Mamm Genome* **13**: 157-163.

Cech, T.R. 2004. Beginning to understand the end of the chromosome. *Cell* **116**: 273-279.

Chaimanee, Y., Jaeger, J. 2001. Evolution of *Rattus* (Mammalia, Rodentia) during the Plio-Pleistocene in Thailand. *Historical Biol* **15**: 181-191.

Chappell, R. 2004. *Motutapere Island rodent invasion 2004*. Internal report, file ISL 0013. Department of Conservation, Coromandel. 3pp.

Charlesworth, B., Sniegowski, P., Stephan, W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215-220.

Charni, F., Friand, V., Haddad, O., *et al.* 2009. Syndecan-1 and syndecan-4 are involved in RANTES/CCL5-induced migration and invasion of human hepatoma cells. *Biochim Biophys Acta* **1790**: 1314-1326.

Charrel, R.N., Attoui, H., Butenko, A.M., *et al*. 2004. Tick-borne virus diseases of human interest in Europe. *Clin Microbiol Infect* **10**: 1040-1055.

Chaves, R., Guedes-Pinto, H., Heslop-Harrison, J.S., Schwarzacher, T. 2000b. The species and chromosomal distribution of the centromeric α-satellite I sequence from sheep in the tribe Caprini and other Bovidae. *Cytogenet Cell Genet* **91**: 62-66.

Chaves, R., Heslop-Harrison, J.S., Guedes-Pinto, H. 2000a. Centromeric heterochromatin in the cattle rob (1;29) translocation: α-satellite I sequences, *in-situ Msp*I digestion patterns, chromomycin staining and C-bands. *Chromosome Res* **8**: 621-626.

Chaves, R., Adega, F., Santos, S., Guedes-Pinto, H., Heslop-Harrison, J.S. 2002. In situ hybridization and chromosome banding in mammalian species. *Cytogenet Genome Res* **96**: 113-116.

Chaves, R., Adega, F., Heslop-Harrison, J.S., Guedes-Pinto, H., Wienberg, J. 2003a. Complex satellite DNA reshuffling in the polymorphic t (1;29) Robertsonian translocation and evolutionary derived chromosomes in cattle. *Chromosome Res* **11**: 641-648.

Chaves, R., Adega, F., Wienberg, J., Guedes-Pinto, H., Heslop-Harrison, J.S. 2003b. Molecular cytogenetic analysis and centromeric satellite organization of a novel 8;11 translocation in sheep: a possible intermediate in biarmed chromosome evolution. *Mamm Genome* **14**: 706-710.

Chaves, R., Santos, S., Guedes-Pinto, H. 2004. Comparative analysis (Hippotragini versus Caprini, Bovidae) of X-chromosome's constitutive heterochromatin by in situ restriction endonuclease digestion: X-chromosome constitutive heterochromatin evolution. *Genetica* **121**: 315-325.

Chaves, R., Guedes-Pinto, H., Heslop-Harrison, J.S. 2005. Phylogenetic relationships and the primitive X chromosome inferred from chromosomal and satellite DNA analysis in Bovidae. *Proc R Soc B* **272**: 2009-2016.

Coleman, R.E., Monkanna, T., Linthicum, K.J., *et al*. 2003. Occurrence of *Orientia tsutsugamushi* in small mammals from Thailand. *Am J Trop Med Hyg* **69**: 519-524.

Coluzzi, M., Sabatini, A., Della Torre, A., Di Deco, M., Tetrarca, V. 2002. A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**: 1415-1418.

Corbet, G.B., Hill, J.E. 1992. *The mammals of the Indomalayan region: A systematic review*. Natural History Museum Publications and Oxford University Press, London and Oxford, U.K., 488p.

Counter, C.M., Avilion, A.A., LeFeuvre, C.E., Stewart, N.G., Greider, C.W., Harley, C.B., Bacchetti, S. 1992. Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J* **11**: 1921-1929.

Coyne, J.A., Orr, H.A. 1998. The evolutionary genetics of speciation. *Phil Trans R Soc Lond B* **353**: 287-305.

Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M., Wayne, R.K. 2000. Considering evolutionary processes in conservation biology. *Tr Ecol Evol* **115**: 290-295.

Culhane, A.C., Howlin, J. 2007. Molecular profiling of breast cancer: transcriptomic studies and beyond. *Cell Mol Life Sci* **64**: 3185-3200.

de la Herran, R., Fontana, F., Lanfredi, M., Congiu, L., Leis, M., Rossi, R., Ruiz Rejon, C., Ruiz Rejon, M., Garrido-Ramos, M.A. 2001. Slow rates of evolution and sequence homogenisation in an ancient satellite DNA family of sturgeons. *Mol Biol Evol* **18**: 432-436.

Dean, C., Schmidt, R. 1995. Plant genomes: a current molecular description. *Annu Rev Plant Physiol Plant Mol Biol* **46**: 395-418.

Demma, L.J., McQuiston, J.H., Nicholson, W.L., Murphy, S.M., Marumoto, P., Sengebau-Kingzio, J.M., Kuartei, S., Durand, A.M., Swerdlow, D.L. 2006. Scrub typhus, republic of Palau. *Emerg Infect Dis* **12**: 290-295.

Deuve, J.L., Bennett, N.C., O'Brien, P.C.M., Ferguson-Smith, M.A., Faulkes, C.G., Britton-Davidian, J., Robinson, T.J. 2006. Complex evolution of X and Y autosomal translocations in the giant mole-rat, *Cryptomys mechowi* (Bathyergidae). *Chromosome Res* **14**: 681-691.

Dheda, K., Huggett, J.F., Bustin, S.A., Johnson, M.A., Rook, G., Zumla, A. 2004. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* **37**: 112-119.

Dobigny, G., Aniskin, V., Volobouev, V. 2002a. Explosive chromosome evolution and speciation in the gerbil genus *Taterillus* (Rodentia, Gerbillinae): a case of two new cryptic species. *Cytogenet Genome Res* **96**: 117-124.

Dobigny, G., Ozouf-Costaz, C., Bonillo, C., Volobouev, V. 2002b. "Ag-NORs" are not always true NORs. new evidence in mammals. *Cytogenet Genome Res* **98**: 75-77.

Dobigny, G., Granjon, L., Aniskin, V., Bâ, K., Volobouev, V. 2003a. A new sibling species of *Taterillus* (Rodentia, Gerbillinae) from West Africa. *Mamm Biol* **68**: 299-316.

Dobigny, G., Ozouf-Costaz, C., Bonillo, C., Volobouev, V. 2003b. Evolution of rRNA genes clusters and telomeric evolution during explosive genome repatterning in *Taterillus* (Rodentia, Gerbillinae). *Cytogenet Genome Res* **103**: 94-103.

Dobigny, G., Ducroz, J., Robinson, T.J., Volobouev, V. 2004a. Cytogenetics and cladistics. *Syst Biol* **53**: 470-484.

Dobigny, G., Ozouf-Costaz, C., Bonillo, C. 2004b. Viability of X autosome translocations in mammals: an epigenomic hypothesis from a rodent case-study. *Chromosoma* **113**: 34-41.

Dobigny, G., Aniskin, V., Granjon, L., Cornette, R., Volobouev, V. 2005. Recent radiation in West African *Taterillus* (Rodentia, Gerbillinae): the concerted role of chromosome and climatic changes. *Heredity* **95**: 358-368.

Dobigny, G., Lecompte, E., Tatard, C., Gauthier, P., Bâ, K., Denys, C., Duplantier, J.-M., Granjon, L. 2008. An update on the taxonomy and geographic distribution of the cryptic species *Mastomys kollmannspergeri* (Muridae, Murinae) using combined cytogenetic and molecular data. *J Zool* **276**: 368-374.

Dobigny, G., Catalan, J., Gauthier, P., *et al.* 2010a. Geographic patterns of inversion polymorphisms in a wild African rodent, *Mastomys erythroleucus*. *Heredity* **104**: 378-386.

Dobigny, G., Tatard, C., Kane, M., Gauthier, P., Brouat, P., Ba, K., Duplantier, J.M. 2010b. A cytotaxonomic and DNA-based survey of rodents from Northern Cameroon and Western Chad. *Mammal Biol.* DOI:10.1016/j.mambio.2010.10.002.

Dover, G.A. 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet* **2**: 159-165.

Duncan, J.F., van Peenen, P.F.D., Ryan, P.F. 1970. Somatic chromosomes of eight mammals from Con son Island, South Vietnam. *Caryologia* **23**: 173-181.

Duncan, J.F., Irsiana, R., Chang, A. 1974. Karyotypes of five taxa of *Rattus* (Rodentia, Muridae) from Indonesia. *Cytologia* **39**: 295-302.

Duplantier, J.M., Britton-Davidian, J., Granjon, L. 1990. Chromosomal characterisation of three species of the genus *Mastomys* in Senegal. *Zool Syst Evolut Forsch* **28**: 289-298.

Dwinell, M.R., Worthey, E.A., Shimoyama, M. *et al.* 2009. The Rat Genome Database 2009: variation, ontologies and pathways. *Nucleic Acids Res* **37**: D744-D749.

Ellerman, J.R. 1961. *Rodentia Pt. II. Fauna of India - Mammalia 3* (ed. M.L. Roonwal), Delhi Publication-Division, pp 483-873.

Ellingsen, A., Slamovits, C.H., Rossi, M.S. 2007. Sequence evolution of the major satellite DNA of the genus *Ctenomys* (Octodontidae, Rodentia). *Gene* **392**: 283-290.

Elliott, R.M. 2008. Transcriptomics and micronutrient research. *Br J Nutr* **99**: S59-65.

Engelbrecht, A., Dobigny, G., Robinson, T.J. 2006. Further insights into the ancestral murine karyotype: the contribution of the *Otomys-Mus* comparison using chromosome painting. *Cytogenet Genome Res* **112**: 126-130.

Engelbrecht, A., Taylor, P.J., Daniels, S.R., Rambau, R.V. 2011. Chromosomal polymorphisms in the African vlei rats, *Otomys irroratus* (Muridae: Otomyini) detected using banding techniques and chromosome painting: inversions, centromeric shifts, and diploid number variation. *Cytogenet Genome Res.* DOI: 10.1159/000323416.

Epoke, J., Coker, A.O. 1991. Intestinal colonization of rats following experimental infection with *Campylobacter jejuni. East Afr Med J* **68**: 348-351.

Ewing, E.P., Takeuchi, A., Shirai, A., Osterman, J.V**.** 1978. Experimental infection of mouse peritoneal mesothelium with scrub typhus rickettsiae: an ultrastructural study. *Infect Immun* **19:** 1068-1075.

Fagundes, V., Vianna-Morgante, A.M., Yonenaga-Yassuda, Y. 1997. Telomeric sequences localization and G-banding patterns in the identification of a polymorphic chromosomal rearrangement in the rodent Akodon cursor (2n: 14, 15 and 16). *Chromosome Res* **5**: 228-232.

Fagundes, V., Yonenaga-Yassuda, Y. 1998. Evolutionary conservation of whole homologous chromosome arms in the Akodont rodents *Bolomys* and *Akodon* (Muridae, Sigmodontinae): maintenance of interstitial telomeric segments (ITBs) in recent event of centric fusion. *Chromosome Res* **6**: 643-648.

Faravelli, M., Moralli, D., Bertoni, L., Attolini, C., Chernova, O., Raimondi, E., Giulotto, E. 1998. Two extended arrays of a satellite DNA sequence at the centromere and at the short-arm telomere of Chinese hamster chromosome 5. *Cytogenet Cell Genet* **83**: 281-286.

Faravelli, M., Azzalin, C.M., Bertoni, L., Chernova, O., Attolini, C., Mondello, C., Giulotto E. 2002. Molecular organization of internal telomeric sequences in Chinese hamster chromosomes. *Gene* **283**: 11-16.

Fears, C. Y., Woods, A. 2006. The role of syndecans in disease and wound healing. *Matrix Biol* **25:** 443-456.

Ferguson-Smith, M.A., Trifonov, V. 2007. Mammalian karyotype evolution. *Nature* **8**: 950-962.

Filla, M.S., Woods, A., Kaufman, P.L., Peters, D.M. 2006. Beta1 and beta3 integrins cooperate to induce syndecan-4-containing cross-linked actin networks in human trabecular meshwork cells. *Invest Ophthalmol Vis Sci* **47**: 1956-1967.

Finato, A.O., Varella-Garcia, M., Tajara, E.H., Taddei, V.A., Morielle-Versute, E. 2000. Intrachromosomal distribution of telomeric repeats in *Eumops glaucinus* and *Euntops perotis* (Molossidae, Chiroptera). *Chromosome Res* **8**: 563-569.

Frances, S.P., Watcharapichat, P., Phulsuksombati, D., Tanskul, P., Linthicum, K.J. 1999. Seasonal occurrence of *Leptotrombidium deliense* (Acari: Trombiculidae) attached to sentinel rodents in an orchard near Bangkok, Thailand. *J Med Entomol* **36**: 869-874.

Frances, S.P., Watcharapichat, P., Phulsuksombati, D. 2000. Development and persistence of antibodies to *Orientia tsutsugamushi* in the roof rat, *Rattus rattus* and laboratory mice following attachment of naturally infected *Leptotrombidium deliense. Acta Trop* **77**: 279-285.

Fredga, K. 1972. Comparative chromosome studies in mongooses (Carnivora, Voverridae). I. Idiograms of 12 species and karyotype evolution in Herpestinae. *Hereditas* **71**: 1-74.

Freissler, E., Meyer auf der Heyde, A., David, G., Meyer, T.F., Dehio, C. 2000. Syndecan-1 and syndecan-4 can mediate the invasion of OpaHSPG-expressing *Neisseria gonorrhoeae* into epithelial cells. *Cell Microbiol* **2**: 69-82.

Fry, K., Salser, W. 1977. Nucleotide sequences of HS-a satellite DNA from kangaroo rat *Dipodomys ordii* and characterisation of similar sequences in other rodents. *Cell* **12**: 1069-1084.

Gadi, I.K., Sharma, T., Raman, R. 1982. Supernumerary chromosomes in *Bandicota indica nemorivaga* and a female individual with XX/X0 mosaicism. *Genetica* **58**: 103-108.

Gadi, I.K., Sharma, T. 1983. Cytogenetic relationships in *Rattus, Cremnomys, Millardia, Nesokia* and *Bandicota* (Rodentia: Muridae). *Genetica* **61**: 21-40.

Gallagher, D.S., Davis, S.K., De Donato, M., Burzlaff, J.D., Womack, J.E., Taylor, J.F., Kumamoto, A.T. 1999. A molecular cytogenetic analysis of the tribe Bovini (Artiodactyla: Bovidae: Bovinae) with an emphasis on sex chromosome morphology and NOR distribution. *Chromosome Res* **7**: 481-492.

Ganzhorn, J.U. 2003. Effects of introduced *Rattus rattus* on endemic small mammals in dry deciduous forest fragments of western Madagascar. *Anim Conserv* **6**: 147-157.

Garagna, S., Broccoli, D., Redi, C.A., Searle, J.B., Cooke, H.J., Capanna, E. 1995. Robertsonian metacentrics of the house mouse lose telomeric sequences but retain some minor satellite DNA in the pericentromeric area. *Chromosoma* **103**: 685-692.

Garagna, S., Ronchetti, E., Mascheretti, S., Crovella, S., Formenti, D., Rumpler, Y., Romanini, M.G. 1997. Non-telomeric chromosome localization of (TTAGGG)n repeats in the genus *Eulemur. Chromosome Res* **5**: 487-491.

Garagna, S., Marziliano, N., Zuccotti, M., Searle, J.B., Capanna, E., Redi, C.A. 2001. Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. *Proc Natl Acad Sci* **98**: 171-175.

Gauthier, P., Hima, K., Dobigny, G. 2010. Robertsonian fusions, pericentromeric repeat organization and evolution: a case study within a highly polymorphic rodent species, *Gerbillus nigeriae. Chromosome Res* **18**: 473-486.

Gavrilovskaya, I.N., Shepley, M., Shaw, R., Ginsberg, M.H., Mackow, E.R. 1998. β3 Integrins mediate the cellular entry of hantaviruses that cause respiratory failure. *Proc Natl Acad Sci USA* **95**: 7074-7079.

Gemmeke, H., Niethammer, J. 1984. Zur Taxonomie der Gattung *Rattus* (Rodentia, Muridae). *Z Saugetierkunde* **49**: 101-116.

Gibbs, R.A., Weinstock, G.M., Metzker, M.L. *et al*. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution**.** *Nature* **428:** 493-521.

Gilbert, C, O'Brien, P.C.M., Bronner, G., Yang, F., Hassanin, A., Ferguson-Smith, M.A., Robinson, T.J. 2006. Chromosome painting and molecular dating indicate a low rate of chromosomal evolution in golden moles (Mammalia, Chrysochloridae). *Chromosome Res* **14**: 793-803.

Gilbert, C., Maree, S., Robinson, T.J. 2008. Chromosomal evolution and distribution of telomeric repeats in golden moles (Chrysochloridae, Mammalia). *Cytogenet Genome Res* **121**: 110-119.

Ginzinger, D.G. 2002. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Haematol* **30**: 503-512.

Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R., Mathieu, C. 2001. An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. *Methods* **25**: 386-401.

Glas, R., Graves, J.A.M., Toder, R., Ferguson-Smith, M., O'Brien, P.C. 1999. Cross-species chromosome painting between human and marsupial directly demonstrates the ancient region of the mammalian X. *Mamm Genome* **10**: 1115-1116.

Go, Y., Rakotoarisoa, G., Kawamoto, Y., Randrianjafy, A., Koyama, N., Hirai, H. 2000. PRINS analysis of the telomeric sequence in seven lemurs. *Chromosome Res* **8**: 57-65.

Gold, J.R., Li, Y. 1994. Chromosomal NOR karyotypes and genome size variation among squawfishes of the genus *Ptychocheilus* (Teleostei: Cyprinidae). *Copeia* **1994**: 60-65.

González Ittig, R.E., Gardenal, C.N. 2004. Recent range expansion and low levels of contemporary gene flow in *Calomys musculinus*: its relationship with the emergence and spread of Argentine haemorrhagic fever. *Heredity* **93**: 535-541.

Goodpasture, C., Bloom, S. E. 1975. Visualization of nucleolar organizer regions in mammalian chromosomes using silver staining. *Chromosoma* **53**: 37-50.

Granjon, L., Dobigny, G. 2003. Chromosomally characterised murid rodents from the edges of Lake Chad: evidence for an African biogeographical crossroads, or a centre of endemism? *Mammal Rev* **33**: 77-91.

Graphodatsky, A.S. 1989. Conserved and variable elements of mammalian chromosomes. In: Halnan CRE, editor. *Cytogenetics of Animals*. Oxon: CAB International. p 95-123. AB International, Oxon, UK.

Graphodatsky, A. S. 2007. Comparative Chromosomics. *Mol Biol* **41**: 361-375.

Graphodatsky, A.S., Yang, F., Dobigny, G., *et al*. 2008. Tracking genome organization in rodents by zoo-FISH. *Chromosome Res* **16**: 261-274.

Grewal, S.I., Moazed, D. 2003. Heterochromatin and epigenetic control of gene expression. *Science* **301**: 798-802.

Griffin, D.K., Robertson, L.B.W., Tempest, H.G., Skinner, B.M. 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet. Genome Res* **117**: 64-77.

Grunow, R., Finke, E. 2002. A procedure for differentiating between the intentional release of biological warfare agents and natural outbreaks of disease: its use in analyzing the tularemia outbreak in Kosovo in 1999 and 2000. *Clin Microbiol Infect* **8**: 510-521.

Grutzner, F., Himmelbauer, H., Paulsen, M., Ropers, H.H., Haaf, T. 1999. Comparative mapping of mouse and rat chromosomes by fluorescence in situ hybridization. *Genomics* **55**: 306-313.

Guidon, S., Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696-704.

Guilly, M., Fouchet, P., de Chamisso, P., Schmitz, A., Dutrillaux, B. 1999. Comparative karyotype of rat and mouse using bidirectional chromosome painting. *Chromosome Res* **7**: 213-221.

Guttenbach, M., Nanda, I., Feichtinger, W., Masabanda, J.S., Griffin, D. K., Schmid, M. 2003. Comparative chromosome painting of chicken autosomal paints 1–9 in nine different bird species. *Cytogenet Genome Res* **103**: 173-184.

Hall, K.J., Parker, J.S. 1995. Stable chromosome fission associated with rDNA mobility. *Chromosome Res* **3**: 417-422.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95-98.

Hamilton, M. J., Honeycutt, R. L., Baker, R. J. 1990. Intragenomic movement, sequence amplification and concerted evolution in satellite DNA in harvest mice (*Reithrodontomys*): evidence from in situ hybridization. *Chromosoma* **99**: 321-329.

Hamilton, M.J., Hong, G., Wichman, H.A. 1992. Intragenomic movement and concerted evolution of satellite DNA in *Peromyscus*: evidence from in situ hybridization. *Cytogenet Cell Genet* **60**: 40-44.

Hamta, A., Adamovic, T., Samuelson, E., Helou, K., Behboudi, A., Levan, G. 2006. Chromosome ideograms of the laboratory rat (*Rattus norvegicus*) based on high- resolution banding, and anchoring of the cytogenetic map to the DNA sequence by FISH in sample chromosomes. *Cytogenet Genome Res* **115**: 158-168.

Hartman, N., Scherthan, H. 2003. Characterisation of ancestral chromosome fusion points in the Indian muntjac deer. *Chromosoma* **112**: 213-220.

He, J., Innis, B.L., Shrestha, M.P. 2002. Evidence that rodents are a reservoir of hepatitis E virus for humans in Nepal. *J Clin Microbiol* **40**: 4493-4498.

Helou, K., Walentinsson, A., Levan, G., Stahl, F. 2001. Between rat and mouse zoo-FISH reveals 49 chromosomal segments that have been conserved in evolution. *Mamm Genome* **12**: 765-771.

Hoffman, S.L., Subramanian, G.M., Collins, F.H., Venter, J.C. 2002. *Plasmodium* , human and *Anopheles* genomics and malaria. *Nature* **415**: 702-709.

Hoffman, A.A., Sgro, C.M., Weeks, A.R. 2004. Chromosomal inversion polymorphisms and adaptation. *Trends Ecol Evol* **19**: 483-488.

Hsu, T.C., Benirschek, K. 1974. *An atlas of mammalian chromosomes.* Vol 8. Berlin, Heidelberg, New York. Spring-Verlag.

Huelsenbeck, J. P., Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754-755.

Huggett, J., Dheda, K., Bustin, S.A., Zumla, A. 2005. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun* **6**: 279-284.

Ihn, K.S., Han, S.H., Kim, H.R., Huh, M.S., Seong, S.Y., Kang, J.S., Han, T.H., Kim, I.S., Choi, M.S. 2000. Cellular invasion of *Orientia tsutsugamushi* requires initial interaction with cell surface heparan sulfate. *Microb Pathog* **28**: 227-233.

Ijdo, J.W., Wells, R.A., Baldini, A., Reeders, S.T. 1991. Improved telomere detection using a telomere repeat probe (TTAGGG)n generated by PCR. *Nucleic Acids Res* **19**: 4780.

Imai, H. T. 1991. Mutability of constitutive heterochromatin (C bands) during eukaryotic chromosomal evolution and their cytological meaning. *Jpn J Genet* **66**: 635-661.

Isaäcson, M., Arntzen, L., Taylor, P. 1981. Susceptibility of members of the *Mastomys natalensis* species complex to experimental infection with *Yersinia pestis*. *J Infect Dis* **144**: 80.

Ishiguro, K., Kadomatsu, K., Kojima, T., *et al*. 2001. Syndecan-4 deficiency leads to high mortality of lipopolysaccharide-injected mice. *J Biol Chem* **276**: 47483–47488.

Jacob, H.J., Brown, D.M., Bunker, R.K., *et al*. 1995. A genetic linkage map of the laboratory rat, *Rattus norvegicus*. *Nat Genet* **9**: 63-69.

Jansa, S.A., Weksler, M. 2004. Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. *Mol Phylogenet Evol* **31**: 256-276.

Jansa, S.A., Barker, F.K., Heaney, L.R. 2006. The pattern and timing of diversification of Philippine endemic rodents: evidence from mitochondrial and nuclear gene sequences. *Syst Biol* **55**: 73-88.

Jiang, Q. 1995. Study on chromosomal classification of *Rattus confucianus*, *R. confucianus lotipes* and *R. fulvescens*. *J Sun Yatsen Univ* **1**: 98-103.

Jobse, C., Buntjer, J.B., Haagsma, N., Breukelman, H.J., Beintema, J.J., Lenstra, J.A. 1995. Evolution and recombination of bovine DNA repeats. *J Mol Evol* **41**: 277-283.

Kalitsis, P., Griffiths, B., Choo, K.H.A. 2006. Mouse telocentric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. *Proc Natl Acad Sci* **103**: 8786-8791.

Kantakamalakul, W., Sirtantikorn, S., Thongcharoen, P., Singchai, C., Puthavathana, P. 2003. Prevalence of rabies virus and Hantaan virus infections in commensal rodents and shrews trapped in Bangkok. *J Med Assoc Thai* **86**: 1008-1014.

Karlen, Y., McNair, A., Perseguers, S., Mazza, C., Mermod, N. 2007. Statistical significance of quantitative PCR. *BMC Bioinformatics* **8**: 131.

Keeling, M.J., Gilligan, C.A. 2000. Metapopulation dynamics of bubonic plague. *Nature* **407**: 903-906.

Kellogg, D.E., Rybalkin, I., Chen, S., Mukhamedova, N., Vlasik, T., Siebert, P.D., Chenchik, A. 1994. TaqStart Antibody™ "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. *BioTechniques* **16**: 1134-1137.

Kelly, D.J., Richards, A.L., Temenak, J., Strickman, D., Dasch, G.A. 2002. The past and present threat of rickettsial diseases to military medicine and international public health. *Clin Infect Dis* **34**: 145-169.

Kijlstra, A., Meerburg, B., Cornelissen, J., De Craeye, S., Vereijken, P,. Jongert, E. 2008. The role of rodents and shrews in the transmission of *Toxoplasma gondii* to pigs. *Vet Parasitol* **156**: 183-190.

Kim, C.W., Goldberger, O.A., Gallo, R.L., Bernfield, M. 1994. Members of the syndecan family of heparan sulfate proteoglycans are expressed in distinct cell-, tissue-, and development-specific patterns. *Mol Biol Cell* **5**: 797-805.

Kim, M.K., Seong, S.Y., Seoh, J.Y., Han, T.H., Song, H.J., Lee, J.E., Shin, J.H., Lim, B.U., Kang, J.S. 2002. *Orientia tsutsugamushi* inhibits apoptosis of macrophages by retarding intracellular calcium release. *Infect Immun* **70**: 4692-4696.

Kim, H.R., Choi, M.S., Kim, I.S., 2004. Role of Syndecan-4 in the cellular invasion of *Orientia tsutsugamushi*. *Microb Pathog* **36**: 219-225.

King, M. 1993. *Species Evolution: The Role of Chromosomal Change*. Cambridge University Press, Cambridge.

Kirkpatrick, M., Barton, N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419-434.

Kodama, Y., Yoshida, M., Sasaki, M. 1980: An improved silver staining technique for nucleolus organizer regions by using nylon cloth. *Jap J Human Genet* **25**: 229-233.

Koh, Y.S., Yun, J.H., Seong, S.Y., Choi, M.S., Kim, I.S. 2004. Chemokine and cytokine production during *Orientia tsutsugamushi* infection in mice. *Microb Pathog* **36:** 51-57.

Kolnicki, R.L. 2000. Kinetochore reproduction in animal evolution: cell biological explanation of karyotypic fission theory. *Proc Natl Acad Sci USA* **97**: 9493-9497.

Koo, D-H., Nam, Y-W., Choi, D., Bang, J-W., de Jong, H., Hur, Y. 2010. Molecular cytogenetic mapping of *Cucumis sativus* and *C. melo* using highly repetitive DNA sequences. *Chromosome Res* **18**: 325-336.

Kwon, J.M., Goate, A.M. 2000. The candidate gene approach. *Alcohol Res Health* **24**: 164-168.

Laulederkind, S.J.F., Shimoyama, M., Hayman, G.T. *et al.* 2011. The Rat Genome Database curation tool suite: a set of optimized software tools enabling efficient acquisition, organization, and presentation of biological data. *Database* **2011**: bar002.

Laskowski, R.A., Watson, J.D., Thornton, J.M. 2005. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* **33** Web Server issue: W89-W93.

Lecompte, E., Denys, C., Granjon, L., Volobouev, V. 2003. Integrative systematics: contributions to Mastomys phylogeny and evolution. In: Grant R. Singleton, Lyn A. Hinds, Charles J. Krebs and Dave M. Spratt, 2003. *Rats, mice and people: rodent biology and management*. ACIAR Monograph No. 96. Australian Centre for International Agricultural Research, pp 536-540.

Lecompte, E., Brouat, C., Duplantier, J.M., Galan, G., Granjon, L., Loiseau, A., Mouline, K., Cosson, J.F. 2005. Molecular identification of four cryptic species of *Mastomys* (Rodentia, Murinae). *Biochem Syst Ecol* **33**: 681-689.

Lecompte, E., Fichet-Calvet, E., Daffis, S., *et al*. 2006. *Mastomys natalensis* and Lassa fever, West Africa. *Emerg Infect Dis* **12**: 1971-1974.

Lecompte, E., Aplin, K., Denys, C., Catzeflis, F., Chades, M., Chevret, P. 2008. Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. *BMC Evol Biol* **8**: 199.

Lee, M.R., Elder, F.F.B. 1980. Yeast stimulation of bone marrow mitosis for cytogenetic investigations. *Cytogenet Cell Genet* **26**: 36-40.

Lee, H.W., Lee, P.W., Baek, L.J., Song, C.K., Seong, I.W. 1981. Intraspecific transmission of Hantaan virus, etiologic agent of Korean hemorrhagic fever, in the rodent *Apodemus agrarius*. *Am J Trop Med Hyg* **30**: 1106-1112.

Lee, J.H., Park, I.H., Gao, Y., Li, J.B., Li, Z., Daley, G.Q., Zhang, K., Church, G.M. 2009. A robust approach to identifying tissue specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS Genet* **5**: e1000718.

Lerdthusnee, K., Khuntirat, B., Leepitakrat, W., *et al*. 2003. Scrub typhus: vector competence of *Leptotrombidium chiangraiensis* chiggers and transmission efficacy and isolation of *Orientia tsutsugamushi*. *Ann N Y Acad Sci* **990**: 25-35.

Lerdthusnee, K., Nigro, J., Monkanna, T., *et al*. 2008. Surveys of rodent-borne disease in Thailand with a focus on scrub typhus assessment. *Integr Zool* **3**: 267-273.

Levan, G. 1974. Nomenclature for G-bands in rat chromosomes. *Hereditas* **77**: 37-52.

Lewis, M.D., Yousuf, A.A., Lerdthusnee, K., Razee, A., Chandranoi, K., Jones, J.W. 2003. Scrub typhus re-emergence in the Maldives. *Emerg Infect Dis* **9**: 1638-1641.

Li, Y.X., Kirby, M.L. 2003. Coordinated and conserved expression of alphoid repeat and alphoid repeat-tagged coding sequences. *Dev Dyn* **228**: 72-81.

Li, T., O'Brien, P.C.M., Biltueva, L., Fu, B., Wang, J., Nie, W., Ferguson-Smith, M.A., Graphodatsky, A.S., Yang, F. 2004. Evolution of genome organizations of squirrels (Sciuridae) revealed by cross-species chromosome painting. *Chromosome Res* **12**: 317-335.

Li, T., Wang, J., Su, W., Nie, W., Yang, F. 2006. Karyotypic evolution of the family Sciuridae: inferences from the genome organizations of ground squirrels. *Cytogenet Genome Res* **112**: 270-276.

Li, Y., Wu, Y., Harada, M., Lin, L-K., Motokawa, M. 2008. Karyotypes of three rat species (Mammalia: Rodentia: Muridae) from Hainan Island, China, and the valid specific status of *Niviventer lotipes. Zoolog Sci* **25**: 686-692.

Liu, W.S., Fredga, K. 1999. Telomeric (TTAGGG)n sequences are associated with nucleolus organizer regions (NORs) in the wood lemming. *Chromosome Res* **7**: 235-240.

Liu, Y.X., Zhao, Z.T., Yang, Z.Q., Zhang, J.L., Xu, J.J., Wu, Q.Y., Peng, Z.L., Miao, Z.S. 2003. Epidemiological studies on host animals of scrub typhus of the autumn-winter type in Shandong Province, China. *Southeast Asian J Trop Med Public Health* **34**: 826-830.

Lomiento, M., Jiang, Z., D'Addabbo, P., Eichler, E.E., Rocchi, M. 2008. Evolutionary-new centromeres preferentially emerge within gene deserts. *Genome Biol* **9**: R173.

Longley, R.L., Woods, A., Fleetwood, A., Cowling, G.J., Gallagher, J.T., Couchman, J.R. 1999. Control of morphology, cytoskeleton and migration by syndecan-4. *J Cell Sci* **112**: 3421-3431.

Louzada, S., Paco, A., Kubickova, S., Adega, F., Guedes-Pinto, H., Rubes, J., Chaves, R. 2008. Different evolutionary trails in the related genomes *Cricetus cricetus* and *Peromyscus eremicus* (Rodentia, Cricetidae) uncovered by orthologous satellite DNA repositioning. *Micron* **39**: 1149-1155.

Lyon, M.F. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**: 372-373.

Mackow, E.R., Gavrilovskaya, I.N. 2009. Hantavirus regulation of endothelial cell functions. *Thromb Haemost* **102**: 1030-1041.

Markvong, A., Marshall, J.T., Gropp, A. 1973. Chromosomes of rats and mice of Thailand. *Nat Hist Bull Siam Soc* **25**: 23-32.

Marshall, J.R. 1977. Family Muridae. In: Lekagul, B., Mcneely, J.A.: *Mammals of Thailand*. Bangkok: Ass. Conservation of Wildlife, pp 1-97, 396-487.

Marshall, O.J., Chueh, A.C., Wong, L.H., Choo, K.H. 2008. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet* **82**: 261-82.

Marshall, O.J., Choo, K.H. 2009. Neocentromeres come of age. *PLos Genet* **5**: e1000370.

Martins, C., Cabral-de-Mello, D.C., Valente, G.T., Mazzuchelli, J., de Oliveira, S.G. 2010. Cytogenetic Mapping and Contribution to the Knowledge of Animal Genomes. In: *Advances in Genetics Research Volume 4 (ISBN 978-1-61728-764-0)*. Edited by:Nova Science Publishers, Inc., pp 1-81.

Mascarello, J. T., Hsu, T.C. 1976. Chromosome evolution in woodrats, genus *Neotoma* (Rodentia: Cricetidae). *Evolution* **30**: 152-169.

Matisoo-Smith, E., Robins, J.H. 2004. Origins and dispersals of Pacific peoples: Evidence from mtDNA phylogenies of the Pacific rat. *Proc Natl Acad Sci* **101**: 9167-9172.

Matsubara, K., Nishida-Umehara, C., Kuroiwa, A., Tsuchiya, K., Matsuda, Y. 2003. Identification of chromosome rearrangements between the laboratory mouse (*Mus musculus*) and the Indian spiny mouse (*Mus platythrix*) by comparative FISH analysis. *Chromosome Res* **11**: 57-64.

Matsubara, K., Nishida-Umehara, C., Tsuchiya, K., Nukaya, D., Matsuda, Y. 2004. Karyotypic evolution of *Apodemus* (Muridae, Rodentia) inferred from comparative FISH analyses. *Chromosome Res* **12**: 383-395.

Mazurok, N.A., Rubtsova, N.V., Isaenko, A.A., Pavlova, M.E., Slobodyanyuk, S.Y., Nesterova, T.B., Zakian, S.M. 2001. Comparative chromosome and mitochondrial DNA analyses and phylogenetic relationships with common voles (*Microtus*, Arvicolidae). *Chromosome Res* **9**: 107-120.

Meerburg, B., Singleton, G., Kijlstra, A. 2009. Rodent-borne diseases and their risks for public health. *Crit Rev Microbiol* **35**: 221-270.

Meštrović, N., Mravinac, B., Juan, C., Ugarkovic, D., Plohl, M. 2000. Comparative study of satellite sequences and phylogeny of five species from the genus *Palorus* (Insecta, Coleoptera). *Genome* **43**: 776-785.

Metcalfe, C.J., Eldridge, M.D.B., Johnson, P.G. 2004. Mapping the distribution of the telomeric sequence (T2AG3)n in the 2n = 14 ancestral marsupial complement and in the macropodines (Marsupialia: Macropodidae) by fluorescence in situ hybridization. *Chromosome Res* **12**: 405-414.

Michaux, J. R., Kinet, S., Filipucci, M.G., Libois, R., Besnard, A., Catzeflis, F. 2001. Molecular identification of three sympatric species of wood mice (*Apodemus sylvaticus*, *A. flavicollis*, *A. alpicola*) in Western Europe (Muridae, Rodentia). *Mol Ecol Notes* **1**: 260-263.

Michaux, J., Chevret, P., Renaud, S. 2007. Morphological diversity of Old World rats and mice (Rodentia, Muridae) mandible in relation with phylogeny and adaptation. *J Zool Sys Evol Res* **45**: 263-279.

Mills, J.N., Childs, J.E. 1998. Ecologic studies of rodent reservoirs: their relevance for human health. *Emerg Infect Dis* **4**: 529-537.

Mills, J.N. 2006. Biodiversity loss and emerging infectious disease: An example from the rodent-borne hemorrhagic fevers. *Biodiversity* **7**: 9-13.

Misonne, X. 1969. African and Indo-Australian Muridae, Evolutionary trends. *Annals Mus r afr cent ser* **84-182**: 1-129.

Mlynarski, E.E., Obergfell, C.J., O'Neill, M.J., O'Neill, R.J. 2010. Divergent patterns of breakpoint reuse in Muroid rodents. *Mamm Genome* **21**: 77-87.

Modi, W. S., Gallagher, D. S., Womack, J. E. 1996. Evolutionary histories of highly repeated DNA families among the Artiodactyla (Mammalia). *J Mol Evol* **42**: 337-349.

Morgan, M.R., Humphries, M.J., Bass, M.D. 2007. Synergistic control of cell adhesion by integrins and syndecans. *Nat Rev Mol Cell Biol* **8**: 957-969.

Moritz, C. 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. *Syst Biol* **51**: 238-254.

Motokawa, M., Lu, K.H., Harada, M., Lin, L.K. 2001. New records of the Polynesian rat *Rattus exulans* (Mammalia: Rodentia) from Taiwan and the Ryukyus. *Zool Stud* **40**: 299-304.

Mount, D.W. 2007. Dot matrix pairwise sequence comparison. *Cold Spring Harb Protoc.* Doi:10.1101/pdb.top31.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., Wu, J.R. 1988. A highly repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci USA* **85**: 6622-6626.

Mravinac, B., Plohl, M., Mestrovic, N., Ugarkovic, D. 2002. Sequence of PRAT satellite DNA ''frozen'' in some *Coleopteran* species. *J Mol Evol* **54**: 774-783.

Murphy, W.J., Stanyon, R., O'Brien, S.J. 2001. Evolution of mammalian genome organization from comparative gene mapping. *Genome Biol* **2**: 1-8.

Murphy, W.J., Pevzner, P.A., O'Brien, J. 2004. Mammalian phylogenomics comes of age. *Trends Genet* **20**: 631-639.

Murphy, W.J., Larkin, D.M., Everts-van der Wind, A., *et al*. 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**: 613-617.

Musser, G.G. 1981. Results of the Archbold expedition no. 105. Notes on systematics of Indo-Malayan Rodents and description of new genera and species from Ceylon, Sulawesi and the Philippines. *Bull Am Mus Nat Hist* **168**: 229-334.

Musser, G. G., Newcomb, C. 1983. Malaysian murids and the giant rat of Sumatra. *Bull Am Mus Nat Hist* **174**: 327-598.

Musser, G.G., Heaney, L.R. 1992. Philippine rodents: Definitions of *Tarsomys* and *Limnomys* plus a preliminary assessment of phylogenetic patterns among native Philippine murines (Murinae, Muridae). *Bull Am Mus Nat Hist* **211**: 1-138.

Musser, G.G., Carleton, M.D. 1993. Family Muridae. In: *Mammal Species of the World a Taxonomic and Geographic Reference*. Edited by: Wilson, D.E., Reeder, D.M. Smithsonian Institution Press, Washington D.C., pp 501-755.

Musser, G.G., Carleton, M.D. 2005. Superfamily Muroidea. In: *Mammal species of the world a taxonomic and geographic reference Volume 2*. 3rd edition. Edited by: Wilson, D.E., Reeder, D.M. Baltimore, Johns Hopkins University., pp 894-1531.

Nash, W.G., Menninger, J.C., Wienberg, J., Padilla-Nash, H.M., O'Brien, S.J. 2001. The pattern of phylogenomic evolution of the Canidae. *Cytogenet Cell Genet* **95**: 210-224.

Nguyen, T.T., Aniskin, V.M., Gerbault-Seureau, M., Planton, H., Renard, J.P., Nguyen, B.X., Hassanin, A., Volobouev, V.T. 2008. Phylogenetic position of the saola (*Pseudoryx nghetinhensis*) inferred from cytogenetic analysis of eleven species of Bovidae. *Cytogenet Genome Res* **122**: 41-54.

Niethammer, J. 1977. Versuch der Rekonstruktion der phylogenetischen Beziehungen zwischen einigen zentralasiatischen Muriden. *Bonn Zool Beitr* **28**: 236-247.

Nijman, I.J., Lenstra, J.A. 2001. Mutation and recombination in cattle satellite DNA: a feedback model for the evolution of satellite DNA repeats. *J Mol Evol* **52**: 361-371.

Nilsson, S., Helou, K., Walentinsson, A., Szpirer, C., Nerman, C., Stahl, F. 2001. Rat-mouse and rat-human comparative maps based on gene homology and high-resolution zoo-FISH. *Genomics* **74**: 287-298.

Noor, M.A.F., Grams, K.L., Bertucci, L.A., Reiland, J. 2001. Chromosomal inversion and the reproductive isolation of species. *Proc Natl Acad Sci USA* **98**: 12084-12088.

O'Brien, S.J., Menotti-Raymond, M., Murphy, W.J., *et al.* 1999. The promise of comparative genomics in mammals. *Science* **286**: 458-481.

O'Brien, S.J., Meminger, J.C., Nash, W.G. 2006. *Atlas of Mammalian chromosomes*. John Wiley and Sons, Inc. Hoboken, Canada, pp 281, 282, 287.

O'Neill, R.J., Eldridge, M.D., Metcalfe, C.J. 2004. Centromere dynamics and chromosome evolution in marsupials. *J Hered* **95**: 375-381.

Ong, A.K.Y., Tambyah, P.A., Ooi, S., Kumarasinghe, G., Chow, C. 2001. Endemic typhus in Singapore - A re-emerging infectious disease? *Singapore Med J* **42**: 549-552.

Pack, S.D., Borodin, P.M., Serov, O.L., Searle, J.B. 1993. The X-autosome translocation in the common shrew (*Sorex araneus* L.): late replication in female somatic cells and pairing in male meiosis. *Chromosoma* **102**: 355-360.

Pagès, M., Chaval, Y., Herbreteau, V., Waengsothorn, S., Cosson, J.F., Hugot, J.P., Morand, S., Michaud, J. 2010. Revisiting the taxonomy of the Rattini tribe: a phylogeny-based delimitation of species boundaries. *BMC Evol Biol* **10**: 184.

Palomeque, T., Lorite, P. 2008. Satellite DNA in insects: a review. *Heredity* **100**: 564-573.

Paresque, R., de Jesus Silva, M., Yonenaga-Yassuda, Y., Fagundes, V. 2007. Karyological geographic variation of *Oligoryzomys nigripes* Olfers 1818 (Rodentia, Cricetidae) from Brazil. *Genet Mol Biol* **30**: 43-53.

Parish, D.A., Vise, P., Wichman, H.A., Bull, J.J., Baker, R.J. 2002. Distribution of LINEs and other repetitive elements in the karyotype of the bat *Carollia*: implications for X-chromosome inactivation. *Cytogenet Genome Res* **96**: 191-197.

Parola, P., Davoust, B., Raoult, D. 2005b. Tick- and flea-borne rickettsial emerging zoonoses. *Vet Res* **36**: 469-492.

Parola, P., Paddock, C.D., Raoult, D. 2005a. Tick-borne Rickettsioses around the World: Emerging diseases challenging old concepts. *Clin Microbiol Rev* **18**: 719-756.

Patton, J. L., Sherwood, S.T. 1983. Chromosome evolution and speciation in rodents. *Annu Rev Ecol Syst* **14**: 139-158.

Pech, M., Igo-Kemenes, T., Zachau, H.G. 1979. Nucleotide sequence of a highly repetitive component of rat DNA. *Nucleic Acids Res* **7**: 417-432.

Pevzner, P., Tesler, G. 2003. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proc Natl Acad Sci USA* **100**: 7672-7677.

Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**: E45.

Pialek, J., Hauffe, H.C., Searle, J.B. 2005. Chromosomal variation in the house mouse. *Biol J Linn Soc* **84**: 535-563.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., Gray, J. 1988. Fluorescence in situ hybridization with human chromosome-specific libraries detection of trisomy-21 and translocations of chromosome-4. *Proc Nat Acad Sci USA* **85**: 9138-9142.

Pons, J., Bruvo, B., Juan, C., Petitpierre, E., Plohl, M., Ugarkovic, D. 1997. Conservation of satellite DNA in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). *Gene* **205**: 183-190.

Pons, J., Gillespie, R.G. 2003. Common origin of the satellite DNAs of the Hawaiian spiders of the genus *Tetragnatha*: possible selective constraints on the length and nucleotide composition of the repeats. *Gene* **313**: 169-177.

Posada, D., Crandall, K.A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Prentice, M.B., Rahalison, L. 2007. Plague. *Lancet* **369**: 1196-1207.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res* **33**: W116-W120.

Radonić, A., Thulke, S., Mackay, I.M., Landt, O., Siegert, W., Nitsche, A. 2004. Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun* **313**: 856-862.

Raman, R., Sharma, T. 1977. DNA replication, G- and C-bands and meiotic behaviour of supernumerary chromosomes of *Rattus rattus* (Linn.). *Chromosoma* **45**: 111-119.

Rambau, R.V., Robinson, T.J. 2003. Chromosome painting in the African four-striped mouse *Rhabdomys pumilio:* Detection of possible murid species contiguous segment combinations. *Chromosome Res* **11**: 91-98.

Rambau, R.V., Stanyon, R., Robinson, T.J. 2003. Molecular genetics of *Rhabdomys pumilio* subspecies boundaries: mtDNA phylogeography and karyotypic analysis by fluorescence *in situ* hybridization (FISH). *Mol Phylogenet Evol* **28**: 564-575.

Ranjini, P.V. 1967. *Studies on mammalian chromosomes.* Ph.D. thesis, Banaras Hindu University.

Ranz, J.M., Maurin, D., Chan, Y.S., von Grotthuss, M., Hillier, L.W., Roote, J., Ashburner, M., Bergman, C.M. 2007. Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biol* **5**: e152.

Rasmussen, R. 2001. Quantification on the LightCycler. In: Meuer, S, Wittwer, C. and Nakagawara, K, eds. *Rapid Cycle Real-time PCR, Methods and Applications*. Springer Press, Heidelberg.

Ratomponirina, C., Viegas-Péquignot, E., Dutrillaux, B., Petter, F., Rumpler, Y. 1986. Synaptonemal complexes in Gerbillidae: probable role of intercaled heterochromatin in gonosome–autosome translocations. *Cytogenet Cell Genet* **43**: 161-167.

Ray-Chaudhuri, S.P., Pathak, S. 1970. Chromosome analysis in the genus *Rattus*. *Mamm Chrom Newsl* **11**: 135-136.

Redi, C.A., Garagna, S., Zuccotti, M. 1990. Robertsonian chromosome formation and fixation: the genomic scenario. *Biol J Linn Soc* **41**: 235-255.

Rens, W., O'Brien, P.C., Yang, F., Graves, .J.A, Ferguson-Smith, M.A. 1999. Karyotype relationships between four distantly related marsupials revealed by reciprocal chromosome painting. *Chromosome Res* **7**: 461-474.

Rens, W., O'Brien, P.C., Yang, F., Solanky, N., Perelman, P., Graphodatsky, A.S., Ferguson, M.W., Svartman, M., De Leo, A.A., Graves, J.A., Ferguson-Smith, M.A. 2001. Karyotype relationships between distantly related marsupials from South America and Australia. *Chromosome Res* **9**: 301-308.

Rens, W., O'Brien, P.C., Fairclough, H., Harman, L., Graves, J.A., Ferguson-Smith, M.A. 2003. Reversal and convergence in marsupial chromosome evolution. *Cytogenet Genome Res* **102**: 282-290.

Rens, W., Fu, B., O'Brien, P.C.M., Ferguson-Smith, M. 2006. Cross-species chromosome painting. *Nat Protoc* **1**: 783-790.

Reynes, J.M., Soares, J.L., Hüe, T., Bouloy, M., Sun, S., Kruy, S.L., Flye Sainte Marie, F., Zeller, H. 2003. Evidence of the presence of Seoul virus in Cambodia. *Microbes Infect* **5**: 769-773.

Richard, F., Lombard, M., Dutrillaux, B. 2003a. Reconstruction of the ancestral karyotype of eutherian mammals. *Chromosome Res* **11**: 605-618.

Richard, F., Messaoudi, C., Bonnet-Garnier, A., Lombard, M., Dutrillaux, B. 2003b. Highly conserved chromosomes in an Asian squirrel (*Menetes berdmorei*, Rodentia: Sciuridae) as demonstrated by ZOO-FISH with human probes. *Chromosome Res* **11**: 597-603.

Richards, L.R., Rambau, R.V., Lamb, J.M., Taylor, P.J., Yang, F., Schoeman, M.C., Goodman, S.M. 2010. Cross-species chromosome painting in bats from Madagascar: the contribution of Myzopodidae to revealing ancestral syntenies in Chiroptera. *Chromosome Res* **18**: 635-53.

Richter, D., Endepols, S., Ohlenbusch, A., Eiffert, H., Spielman, A., Matuschka, F. 1999. Genospecies diversity of Lyme disease spirochetes in rodent reservoirs. *Emerg Infect Dis* **5**: 291-296.

Rickart, E. A., Musser, G.G. 1993. Philippine rodents: chromosomal characteristics and their significance for phylogenetic inference among 13 species (Rodentia, Muridae, Murinae). *Am Mus Novit* **3064**: 1-34.

Ried, T., Arnold, N., Ward, D.C., Wienberg, J. 1993. Comparative high resolution maping of human and primate chromosomes by fluorescence *in situ* hybridization. *Genomics* **18**: 381-386.

Rioux, R., Landry, R.Y., Bensadoun, A. 2002. Sandwich immunoassay for the measurement of murine syndecan-4. *Lipid Res* **43**: 167-173.

Robins, H., Hingston, M., Matisoo-Smith, E., Ross, H.A. 2007. Identifying *Rattus* species using mitochondrial DNA. *Mol Ecol Notes* **7**: 717-729.

Robins, J.H., McLenachan, P.A., Phillips, M.J., Craig, L., Ross, H.A., Matisoo-Smith, E. 2008. Dating of divergences within the *Rattus* genus phylogeny using whole mitochondrial genomes. *Mol Phylogenet Evol* **49**: 460-466.

Robinson, T.J., Skinner, J.D., Haim, A.S. 1986. Close chromosomal congruence in two species of ground squirrel: *Xerus inauris* and *X. princeps* (Rodentia: Sciuridae). *SA J Zool* **21**: 100-105.

Robinson, T.J., Harrison, W.R., Ponce de Leon, F.A., Davis, S.K., Elder, F.F.B. 1998. A molecular cytogenetic analysis of the X chromosome repatterning in the Bovidae: transpositions, inversions and phylogenetic inference. *Cytogenet Cell Genet* **80**: 179-184.

Robinson, T.J., Seiffert, E.R. 2004. Afrotherian origin and interrelationships: new views and future prospects. *Curr Top Dev Biol* **63**: 37-59.

Robinson, T.J., Ruiz-Herrera, A., Frönicke, L. 2006. Dissecting the mammalian genome - new insights into chromosomal evolution. *Trends Genet* **22**: 297-301.

Robinson, T.J., Ruiz-Herrera, A. 2008. Defining the ancestral eutherian karyotype: a cladistic interpretation of chromosome painting and genome sequence assembly data. *Chromosome Res* **16**: 1133-1141.

Robles, F., de la, H.R., Ludwig, A., Ruiz, R.C., Ruiz, R.M., Garrido-Ramos, M.A. 2004. Evolution of ancient satellite DNAs in sturgeon genomes. *Gene* **338**: 133-214.

Rocchi, M., Stanyon, R., Archidiacono, N. 2009. Evolutionary new centromeres in primates. *Prog Mol Subcell Biol* **48**: 103-152.

Rokas, A., Holland, P.W.H. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol* **15**: 454-459.

Romanenko, S.A., Perelman, P.L., Serdukova, N.A., *et al.* 2006. Reciprocal chromosome painting between three laboratory rodent species. *Mamm Genome* **17**: 1183-1192.

Romanenko, S.A., Volobouev, V.T., Perelman, P.L., *et al.* 2007. Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroidea, Rodentia) inferred from chromosomal painting and banding comparison. *Chromosome Res* **15**: 283-297.

Ropiquet, A., Hassanin, A., Pagacova, E., Gerbault-Seureau, M., Cernohorska, H., Kubickova, S., Bonillo, C., Rubes, J., Robinson, T.J. 2010. A paradox revealed: karyotype evolution in the four-horned antelope occurs by tandem fusion (Mammalia, Bovidae, *Tetracerus quadricornis*). *Chromosome Res* **18**: 277-286.

Rosenberg, R. 1997. Drug-resistant scrub typhus: paradigm and paradox. *Parasitol Today* **13***: 131-132.

Rostand, K.S., Esko, J.D. 1997. Microbial adherence to and invasion through proteoglycans. *Infect Immun* **65**: 1-8.

Rousselet, J., Géri, C., Hewitt, G. M., Lemeunier, F. 1998. The chromosomes of *Diprion pini* and *D. similis* (Hym.: Diprionidae): implications for karyotype evolution. *Heredity* **81**: 573-578.

Rousselet, J., Monti, L., Auger-Rozenberg, M.A., Parker, J.S., Lemeunier, F. 2000. Chromosome fission associated with growth of ribosomal DNA in *Neopridion abietis* (Hymenoptera, Diprionidae). *Proc R Soc Lond B* **267**: 1819-1823.

Rowe, K.C., Reno, M.L., Richmond, D.M., Adkins, R.M., Steppan, S.J. 2008. Pliocene colonization and adaptive radiations in Australia and New Guinea, Sahul: Multilocus systematics of the old endemic rodents, Muroidea: Murinae. *Mol Phylogenet EvoL* **47**: 84-101.

Rubes, J., Kubickova, S., Pagacova, E. Cernohorska, H., Di Berardino, D., Antoninova, M., Vahala, J., Robinson, T.J. 2008. Phylogenomic study of spiral-horned antelope by cross-species chromosome painting. *Chromosome Res* **16**: 935-947.

Ruedas, L.A., Kirsch, J.A.W. 1997. Systematics of Maxomys Sody, 1936 (Rodentia: Muridae: Murinae): DNA/DNA hybridization studies of some Borneo-Javan species and allied Sundaic and Australo-Papuan genera. *Biol J Linn Soc London* **61:** 385-408.

Ruiz-Herrera, A., Castresana, J., Robinson, T.J. 2006. Is mammalian chromosomal evolution driven by regions of genome fragility? *Genome Biol* **7**: R115.

Ruiz-Herrera, A., Robinson, T.J. 2007. Chromosomal instability in Afrotheria: Fragile sites, evolutionary breakpoints and phylogenetic inference from genome sequence assemblies. *BMC Evol Biol* **7**: 199.

Ruiz-Herrera, A., Nergadze, S.G., Santagostino, M., Giulotto, E. 2008. Telomeric repeats far from the ends: mechanisms of origin and role in evolution. *Cytogenet Genome Res* **122**: 219-228.

Rumpler, Y., Dutrillaux, B. 1990. *Chromosomal Evolution and Speciation in Primates*. R.B.C., University of the Basque Country (ed.), Vol. 23, Springer-Verlag, Berlin, Heidelberg, New York.

Salvadori, S., Deiana, A.M., Elisabetta, C., Floridia, G., Rossi, E., Zu¡ardi, O. 1995. Colocalization of (TTAGGG)*n* telomeric sequences and ribosomal genes in Atlantic eels. *Chromosome Res* **3**: 54-58.

Sanchez, A., Jimenez, R., Burgos, M., Stitou, S., Zurita, F., Diaz de la Guardia, R. 1995. Cytogenetic peculiarities in the Algerian hedgehog: silver stains not only NORs but also heterochromatic blocks. *Heredity* **75**: 10-16.

Santos, J.H., Meyer, J.N., Skorvaga, M., Annab, L.A., Van Houten, B. 2004. Mitochondrial hTERT exacerbates free-radical-mediated mtDNA damage. *Aging Cell* **3**: 399-411.

Saphire, A.C., Bobardt, M.D., Zhang, Z., David, G., Gallay, P.A. 2001. Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages. *J Virol* **75**: 9187-9200.

Sasaki, M., Nishida, C., Kodama, Y. 1986. Characterisation of silver-stained nucleolus organizer regions (Ag-NORs) in 16 inbred strains of the Norway rat, *Rattus norvegicus. Cytogenet Cell Genet* **41**: 83-88.

Sayle, R.A., Milner-White, E.J. 1995. RASMOL: biomolecular graphics for all. *Trends Biochem Sci* **20**: 374.

Schmidt, A., Echtermeyer, F., Alozie, A., Brands, K., Buddecke, E. 2005. Plasmin- and thrombin-accelerated shedding of syndecan-4 ectodomain generates cleavage sites at Lys(114)-Arg(115) and Lys(129)-Val(130) bonds. *J Biol Chem* **280**: 34441-34446.

Schröter, M., Zöllner, B., Schäfer, P., Laufs, R., Feucht, H. 2001. Quantitative detection of hepatitis C virus RNA by light cycler PCR and comparison with two different PCR assays. *J Clin Microbiol* **39**: 765-768.

Seabright, M. 1971. A rapid banding technique for human chromosome. *Lancet* **2**: 971-972.

Searle, J.B. 1998. Speciation, chromosomes, and genomes. *Genome Res* **8**: 1-3.

Seddon, J.M., Baverstock, P.R. 2000. Evolutionary lineages of *RT1.Ba* in the Australian *Rattus*. *Mol Biol Evol* **17**: 768-772.

Seong, S.Y., Choi, M.S., Kim, I.S. 2001. *Orientia tsutsugamushi* infection: overview and immune responses. *Microbes Infect* **3**: 11-21.

Serikawa, T. 1992. Rat genetic map and a rat/mouse/human comparative gene map. *Transplantation Proceedings* **31**: 1537-1540.

Shan, F., Yan, G., Plummer, A. 2003. Cyto-evolution of *Boronia* genomes revealed by fluorescence in situ hybridization with rDNA probes. *Genome* **46**: 507-513.

Sharma, T., Bardhan, A., Bahadur, M. 2003. Reduced meiotic fitness in hybrids with heterozygosity for heterochromatin in the speciating *Mus terricolor* complex. *J Biosci* **28**: 189-198.

Shetty, S., Griffin, D.K., Graves, J.A.M. 1999. Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Res* **7**: 289-295.

Sites, J.W. Jr., Marshall, J.C. 2003. Delimiting species: a Renaissance issue in systematic biology. *Tr Ecol Evol* **18**: 462-470.

Slamovits, C.H., Cook, J.A., Lessa, E.P., Rossi, M.S. 2001. Recurrent amplifications and deletions of satellite DNA accompanied chromosomal diversification in South American tuco-tucos (genus *Ctenomys*, Rodentia: Octodontidae): a phylogenetic approach. *Mol Biol Evol* **18**: 1708-1719.

Slamovits, C.H., Rossi, M. 2002. Satellite DNA: agent of chromosomal evolution in mammals. A review. *J Neotrop Mammal* **9**: 297-308.

Slijepcevic, P. 1998. Telomeres and mechanisms of Robertsonian fusion. *Chromosoma* **107**: 136-140.

Smith, M.F., Novotony, J., Carl, V.S., Comeau, L.D. 2006a. *Helicobacter pylori* and toll-like receptor agonists induce syndecan-4 expression in an NF-kB-dependent manner. *Glycobiology* **16**: 221-229.

Smith, M.F., Gautam, J.K., Black, S.G., Ernst, P.B. 2006b. Microbial-induced regulation of syndecan expression: important host defense mechanism or an opportunity for pathogens? *ScientificWorldJournal* **4**: 442-445.

Sonnhammer, E.L.L., Durbin, R. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**:GC1-10.

Spring, J., Goldberger, O.A., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., Bernfield, M. 1994. Mapping of the syndecan genes in the mouse: linkage with members of the myc gene family. *Genomics* **21**: 597-601.

Standard karyotype of the Norway rat, *Rattus norvegicus*. Committee for a standardized karyotype of *Rattus norvegicus*. 1973. *Cytogenet Cell Genet* **12**: 199-205.

Stanyon, R., Yang, F., Cavagna, P., O'Brien, P.C.M., Bagga, M., Ferguson-Smith, M.A., Wienberg, J. 1999. Reciprocal chromosome painting shows that genomic rearrangement between rat and mouse proceeds ten times faster than between humans and cats. *Cytogenet Cell Genet* **84**: 150-155.

Stanyon, R., Stone, G., Garcia, M., Frönicke, L. 2003. Reciprocal chromosome painting shows that squirrels, unlike murid rodents, have a highly conserved genome organization. *Genomics* **82**: 245-249.

Stanyon, R., Yang, F., Morescalchi, A.M., Galleni, L. 2004. Chromosome painting in the long-tailed field mouse provides insights into the ancestral murid karyotype. *Cytogenet Genome Res* **105**: 406-411.

Stanyon, R., Rocchi, M., Capozzi, O., Roberto, R., Misceo, D., Ventura, M., Cardone, M., Bigoni, F., Archidiacono, N. 2008. Primate chromosome evolution: ancestral karyotypes, marker order and neocentromeres. *Chromosome Res* **16**: 17-39.

Steen, R.G., Kwitek-Black, A.E., Glenn, C., *et al*. 1999. A High-Density Integrated Genetic Linkage and Radiation Hybrid map of the laboratory rat. *Genome Res* **9**: 1-8.

Stefansson, H., Helgason, A., Thorleifsson, G., *et al.* 2005. A common inversion under selection in Europeans. *Nat Rev Genet* **37**: 12-137.

Sternes, K.L., Vig, B.K. 1995. Satellite I DNA in transformed rat cells. *Cancer Genet Cytogenet* **79**: 64-69.

Strickman, D., Smith, C.D., Corcoran, K.D., Ngampochjana, M., Watcharapichat, P., Phulsuksombati, D., Tanskul, P., Dasch, G.A., Kelly, D.J. 1994. Pathology of *Rickettsia tsutsugamushi* infection in *Bandicota savilei*, a natural host in Thailand. *Am J Trop Med Hyg* **51**: 416-423.

Sumner, A.T. 1972. A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res* **75**: 304-306.

Swofford, D.L. 2002. *PAUP* 4.0. Phylogenetic Analysis Using Parsimony (*and Other methods).* Sinauer Associates, Sunderland, Massachusetts.

Szabo, P., Lee, M., Elder, F., Prensky, W. 1978. Localization of 5S RNA and rRNA genes in the Norway rat. *Chromosoma* **65**: 161-172.

Takahashi, M., Murata, M., Hori, E., Tanaka, H., Kawamura, A. Jr. 1990. Transmission of *Rickettsia tsutsugamushi* from *Apodemus speciosus*, a wild rodent, to larval trombiculid mites during the feeding process. *Jpn J Exp Med* **60**: 203-208.

Tayeh, A., Tatard, C., Kako-Ouraga, S., Duplantier, J-M., Dobigny, G. 2010. Rodent host cell/Lassa virus interactions: Evolution and expression of α-Dystroglycan, LARGE-1 and LARGE-2 genes, with special emphasis on the *Mastomys* genus. *Infect Genet Evol* **10**: 1262-1270.

Taylor, J.M., Horner, B.E. 1973. Systematics of native Australian Rattus (Rodentia: Muridae). Results of Archbold Expeditiomn 98. *Bull Am Mus Nat Hist* **150**: 1-130.

Taylor, P.J., Arntzen, L., Hayter, M., Iles, M., Frean, J., Belmain, S.R. 2008. Understanding and managing sanitary risks due to rodent zoonoses in an African city: beyond the Boston Model. *Integr Zool* **3**: 38-50.

Telenius, H., Pelmear, A., Tunnacliffe, A., Carter, N.P., Behmel, A., Ferguson-Smith, M.A., Nordenskold, M., Pfagner, R., Ponder, B. 1992. Cytogenetic analysis by chromosome painting using DOP-PCR amplified sorted chromosomes. *Gene Chromosome Canc* **4**: 257-264.

Teterina, A., Richter, D., Matuschka, F.R., Ehlers, B., Voigt, S. 2009. Identification of a novel betaherpesvirus in *Mus musculus*. *Virol J* **6**: 225

Thelma, B.K., Juyal, R.C., Tewari, R., Rao, S.R.V. 1988. Does heterochromatin variation potentiate speciation? A study in *Nesokia. Cytogenet Cell Genet* **47**: 204-208.

Thelwell, N., Millington, S., Solinas, A., Booth, J., Brown, T. 2000. Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Res* **28**: 3752-3761.

Therman, E., Susman, M. 1993. *Human chromosomes: structure, behaviour, and defects*. New York: Springer, 3rd edn, pp 275-276.

Tollenaere, C., Brouat, C., Duplantier, J-M. *et al*. 2010. Phylogeography of the introduced species *Rattus rattus* in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. *J Biogeogr* **37**: 398-410.

Toyota, M., Canzian, F., Ushijima, T., Hosoya, Y., Kuramoto, T., Serikawa, T., Imai, K., Sugimura, T., Nagao, M. 1996. A rat genetic map constructed by representational difference analysis markers with suitability for large-scale-typing. *Proc Natl Acad Sci* **93**: 3914-3919.

Traub, R., Wisseman, C.L., Jones, M.R., O'Keefe, J.J. 1975. The acquisition of *Rickettsia tsutsugamushi* by chiggers (trombiculid mites) during the feeding process. *Ann NY Acad Sci* **266**: 91-114.

Trifonov, V.A., Stanyon, R., Nesterenko, A.I.*, et al.* 2008. Multidirectional cross-species painting illuminates the history of karyotypic evolution in Perissodactyla. *Chromosome Res* **16**:89-107.

Tsipouri, V., Schueler, M.G., Hu, S. NISC Comparative Sequencing Program, Dutra, A., Pak, E., Riethman, H., Green, E.D. 2008. Comparative sequence analyses reveal sites of ancestral chromosomal fusions in the Indian muntjac genome. *Genome Biol* **9**: R155.

Tsuchiya, K., Yosida, T.H., Moriwaki, K., Ohtani, S., Kulta-Uthai, S., Sudto, P. 1979. Karyotypes of twelve species of small mammals from Thailand. *Report Hokkaido Inst Public Health* **29**: 26-29.

Tucker, PK. 1986. Sex chromosomeYautosome translocations in the leaf-nosed bats, family Phyllostomidae. I. Mitotic analyses of the subfamilies Stenodermatinae and Phyllostominae. *Cytogenet Cell Genet* **43**: 19-27.

Twigger, S.N., Pruitt, K.D., Fernández-Suárez, X.M., *et al.* 2008. What everybody should know about the rat genome and its online resources. *Nat Genet* **40**: 523-527.

Tyagi, S., Kramer, F.R. 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* **14**: 303-308.

Ugarković, D., Plohl, M. 2002. Variation in satellite DNA profiles—causes and effects. *EMBO J* **21**:5955-5959.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**: W71-W74.

Usdin, K., Chevret, P., Catzeflis, F.M., Verona, R., Furano, A.V. 1995. L1 (LINE-1) retrotransposable elements provide a "fossil" record of the phylogenetic history of murid rodents. *Mol Biol Evol* **12**: 73-82.

Van Peenen, P.F.D., Ho, C.M., Bourgeois, A.L. 1977. Indirect immunofluorescence antibodies in natural and acquired *Rickettsia tsutsugamushi* infections of Phillippine rodents. *Infect Immun* **15**: 813-816.

van der Flier, A., Sonnenberg, A. 2001. Function and interactions of integrins. *Cell Tissue Res* **305**: 285-298.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**: 34.

Ventura, K., Silva, M.J., Fagundes, V., Christoff, A.U., Yonenaga-Yassuda, Y. 2006. Non-telomeric sites as evidence of chromosomal rearrangement and repetitive (TTAGGG) n arrays in heterochromatic and euchromatic regions in four species of *Akodon* (Rodentia, Muridae). *Cytogenet Genome Res* **115**: 169-175.

Venturi, F.P., Chimimba, C.T., van Aarde, R.J., Fairall, N. 2004. The distribution of two medically and agriculturally important cryptic rodent species, *Mastomys natalensis* and *M. coucha* (Rodentia: Muridae) in South Africa. *Afr Zool* **39**: 235-245.

Verneau, O., Catzeflis, F., Furano, A.V. 1997. Determination of the evolutionary relationships in *Rattus sensu lato* (Rodentia:Muridae) using L1 (LINE-1 amplification events). *J Mol Evol* **45**: 424-436.

Verneau, O., Catzeflis, F., Furano, A.V. 1998. Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. *Proc Natl Acad Sci USA* **95**: 11284-11289.

Veyrunes, F., Catalan, J., Sicard, B., Robinson, T.J., Duplantier, J.M., Granjon, L., Dobigny, G., Britton-Davidian, J. 2004. Autosome and sex chromosome diversity among the African pygmy mice, subgenus *Nannomys* (Murinae; *Mus*). *Chromosome Res* **12**: 369-382.

Veyrunes, F., Britton-Davidian, J., Robinson, T.J., Calvet, E., Denys, C., Chevret, P. 2005. Molecular phylogeny of the African pygmy mice, subgenus *Nannomys* (Rodentia, Murinae, *Mus*): implications for chromosomal evolution. *Mol Phylogenet Evol* **36**: 358-369.

Veyrunes, F., Dobigny, G., Yang, F., O'Brien, P.C.M., Catalan, J., Robinson, T.J., Britton-Davidian, J. 2006. Phylogenomics of the genus *Mus* (Rodentia; Muridae): extensive genome repatterning is not restricted to the house mouse. *Proceedings Proc R Soc B* **273**: 2925-2934.

Veyrunes, F., Catalan, J., Tatard, C., Cellier-Holzem, E., Watson, J., Chevret, P., Robinson, T.J., Britton-Davidian, J. 2010. Mitochondrial and chromosomal insights into karyotypic evolution of the pygmy mouse, *Mus minutoides*, in South Africa. *Chromosome Res* **18**: 563-574.

Viera, A., Ortiz, M.I., Pinna-Senn, E., Dalmasso, G., Bella, J.L., Lisanti, J.A. 2004. Chromosomal localization of telomeric sequences in three species of *Akodon* (Rodentia, Sigmodontinae). *Cytogenet Genome Res* **107**: 99-102.

Volobouev, V., Vogt, N., Viegas-Péquignot, E., Malfoy, B., Dutrillaux, B. 1995. Characterisation and chromosomal location of two repeated DNAs in three *Gerbillus* species. *Chromosoma* **104**: 252-259.

Volobouev, V., Ducroz, J.F., Aniskin, V., *et al*. 2002. Chromosomal characterisation of *Arvicanthis* species (Rodentia: Murinae) from Western and Central Africa: implications on taxonomy. *Cytogenet Genome Res* **96**: 250-260.

Vorou, R.M., Papavassiliou, V.G., Tsiodras, S. 2007. Emerging zoonoses and vestor-borne infections affecting humans in Europe. *Epidemiol Infect* **135**: 1231-1247.

Vujošević, M., Blagojević, J. 2004. B chromosomes in populations of mammals. *Cytogenet Genome Res* **106**: 247-256.

Wadstrom, T., Ljungh, A. 1999. Glycosaminoglycan-binding microbial proteins in tissue adhesion and invasion: key events in microbial pathogenicity. *J Med Microbiol* **48**: 223-233.

Walker, J.S., Gan, E., Chye, C.T., Muul, I. 1973. Involvement of small mammals in the transmission of scrub typhus in Malaysia: isolation and serological evidence. *Trans R Soc Trop Med Hyg* **67**: 838-845.

Wang, W., Lan, H. 2000. Rapid and parallel chromosomal number reductions in muntjac deer inferred from mitochondrial DNA phylogeny. *Mol Biol Evol* **17**: 1326-1333.

Wang, J., Zhao, X., Koh, H.S., Deng, Y., Qi, H. 2003. Chromosomal polymorphisms due to heterochromatin growth and pericentric inversions in white-bellied rat, *Niviventer confucianus*, from China. *Hereditas* **138**: 59-64.

Wang, B., Jiang, J., Xie, F., Chen, X., Dubois, A., Liang, G., Wagner, S. 2009. Molecular Phylogeny and Genetic Identification of Populations of Two Species of *Feirana* Frogs (Amphibia: Anura, Ranidae, Dicroglossinae, Paini) Endemic to China. *Zoolog Sci* **26**: 500-509.

Wangroongsarb, P., Saengsongkong, W., Petkanjanapong, W., Mimgratok, M., Panjai, D., Wootta, W., Hagiwara, T. 2008. An Application of Duplex PCR for Detection of *Leptospira* spp. and *Orientia tsutsugamushi* from Wild Rodents. *Jpn J Infect Dis* **61**: 407-409.

Waters, P.D., Dobigny, G., Pardini, A.T., Robinson, T.J. 2004. LINE-1 distribution in Afrotheria and Xenarthra: implications for understanding the evolution of LINE-1 in eutherian genomes. *Chromosoma* **113**: 137-144.

Watson, J.D., Milner-White, E.J. 2002. The conformations of polypeptide chains where the main-chain parts of successive residues are enantiomeric. Their occurence in cation and anion-binding regions of proteins. *J Mol Biol* **315:**183-191.

Watts, C.H.S., Baverstock, P.R. 1994. Evolution in some South-east Asian Murinae (Rodentia), as assessed by microcomplement fixation of albumin, and their relationship to Australian Murines. *Aust J Zool* **42**: 711-722.

Whary, M.T., Fox, J.G. 2004. Natural and experimental *Helicobacter* infections. *Comp Med* **54:** 128-158.

Whitcombe, D., Theaker, J., Guy, S.P., Brown, T., Little, S. 1999. Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotechnol* **17**: 804-807.

White, M.J.D. 1973. *Animal cytology and evolution*, 3rded. Cambridge University Press, London.

Wienberg, J., Jauch, A., Stanyon, R., Cremer, T. 1990. Molecular cytotaxonomy of primates by chromosomal in situ suppression hybridization. *Genomics* **8**: 347-350.

Wienberg, J., Stanyon, R. 1995. Chromosome painting in mammals as an approach to comparative genomics. *Curr Opin Genetics Dev* **5**: 792-797.

Wienberg, J., Stanyon, R. 1997. Comparative painting of mammalian chromosomes. *Curr Opin Genetics Dev* **7**: 784-791.

Wienberg, J. 2004. The evolution of eutherian chromosomes. *Curr Opin Genetics Dev* **14**: 657-666.

Wilson, D.E., Reeder, D.A.M. 2005. Mammal species of the World: a taxonomic and geographic reference. Baltimore, USA: The John Hopkins University Press.

Wisseman, C. L. 1991. *Rickettsial infections*. G. T. Strickland Hunter's tropical medicine 7th ed. Saunders Philadelphia, PA, pp 256-286.

Wolfe, N.D., Dunavan, C.P., Diamond, J. 2007. Origins of major human infectious diseases. *Nature* **447**: 279-283.

Woods, A., Couchman, J.R. 1994. Syndecan 4 heparan sulfate proteoglycan is a selectively enriched and widespread focal adhesion component. *Mol Biol Cell* **5**: 183-92.

Worley, K.C., Weinstock, G.M., Gibbs, R.A. 2008. Rats in the genomic era. *Physiol Genomics* **32**: 273-282.

Wüthrich, K. 1990. Protein structure determination in solution by NMR spectroscopy. *J Biol Chem* **265**: 22059-22062.

Yamada, J., Kuramoto, T., Serikawa, T. 1994. A rat genetic linkage map and comparative maps for mouse or human homologous rat genes. *Mamm Genome* **5**: 63-83.

Yang, F., O'Brien, P.C.M., Wienberg, J., Ferguson-Smith, M.A. 1997a. A reappraisal of the tandem fusion theory of karyotype evolution in the Indian muntjac using chromosome painting. *Chromosome Res* **5**: 109-117.

Yang, F., O'Brien, P.C.M., Wienberg, J., Neitzel, H., Lin, C.C., Ferguson-Smith, M.A. 1997b. Chromosomal evolution of the Chinese muntjac (*Muntiacus reevesi*). *Chromosoma* **106**: 37-43.

Yang, F., O'Brien, P.C., Ferguson-Smith, M.A. 2000. Comparative chromosome map of the laboratory mouse and Chinese hamster defined by reciprocal chromosome painting. *Chromosome Res* **8**: 219-227.

Yang, L.P., Zhao, Z.T., Li, Z., Wang, X.J., Liu, Y.X., Bi, P. 2008. Comparative analysis of nucleotide sequences of *Orientia tsutsugamushi* in different epidemic areas of scrub typhus in Shandong, China. *Am J Trop Med Hyg* **78**: 968-972.

Yin, J.L., Shackel, N.A., Zekry, A., McGuinness, P.H., Richards, C., Putten, K.V., McCaughan, G.W., Eris, J.M., Bishop, G.A. 2001. Real-time reverse-transcriptase polymerase chain reaction (RT-PCR) for measurement of cytokine and growth factor mRNA expression with fluorogenic probes or SYBR Green I. *Immunol Cell Biol* **79**: 213-221.

Yong, H.S. 1968. Karyotypes of four Malayan rats (Muridae, genius Rattus fisher). *Cytologia* **33**: 174-180.

Yong, H.S. 1969. Karyotypes of Malayan rats (Rodentia, Muridae, genus *Rattus* Fischer). *Chromosoma* **27**: 245-267.

Yong, H.S. 1970. A Malayan view of *Rattus edwardsii* and *R. sabanus* (Rodentia: Muridae). *Zool J Linn Soc* **49**: 359-370.

Yong, H.S. 1973. Chromosomes of the pencil-tailed treemouse, *Chiropodomys gliroides* (Rodentia, Muridae). *Malays Nat J* **26**: 159-162.

Yong, H.S., Dhaliwal, S.S., Lim, L. 1982. Karyotypes of *Hapalomys* and *Pithecheir* (Rodentia, Muridae) from Peninsular Malaysia. *Cytologia* **47**: 535-538.

Yosida, T.H., Tsuchiya, K., Moriwaki, K. 1971a. Frequency of chromosome polymorphism in *Rattus rattus* collected in Japan. *Chromosoma* **33**: 33-40.

Yosida, T.H., Tsuchiya, K., Moriwaki, K. 1971b. Karyotypic differences of Black rats, *Rattus rattus*, collected in various localities of East and Southeast Asia and Oceania. *Chromosoma* **33**: 252-267.

Yosida, T.H., Sagai, T. 1972. Banding pattern analysis of polymorphic karyotypes in the black rats by a new differential staining technique. *Chromosoma* **37**: 387-394.

Yosida, T.H. 1973. Evolution of karyotypes and differentiation in 13 *Rattus* species. *Chromosoma* **40**: 285-297.

Yosida, T.H., Sagai, T. 1973. Similarity of Giemsa banding patterns of chromosomes in several species of the genus *Rattus*. *Chromosoma* **41**: 93-101.

Yosida, T.H., Kato, H., Tsuchiya, K., Sagai, T., Moriwaki, K. 1974. Cytogenetical survey of black rats, *Rattus rattus*, in Southwest and central Asia, with special regard between three geographical types. *Chromosoma* **45**: 99-109.

Yosida, T. H. 1975. Diminition of heterochromatic C-bands in relation to the differentiation of *Rattus* species. *Proc Jpn Acad* **51**: 659-663

Yosida, T.H., Sagai, T. 1975. Variation of C-bands in the chromosomes of several subspecies of *Rattus rattus*. *Chromosoma* **50**: 283-300.

Yosida, T.H. 1976. Frequencies of chromosome polymorphism (pairs no. 1, 9 and 13) in the black rat, *Rattus rattus diardii* (Rodentia, Muridae). *Chromosoma* **36**: 256-262.

Yosida, T.H. 1977a. Frequencies of chromosome polymorphisms in pairs no. 1, 9, and 13 in three geographical variants of Black rats, *Rattus rattus*. *Chromosoma* **60**: 391-398.

Yosida, T.H. 1977b. Supernumerary chromosomes in the black rat (*Rattus rattus*) and their distribution in three geographical variants. *Cytogenet Cell Genet* **18**: 149-159.

Yosida, T.H. 1979. A comparative study on nucleolus organizer regions (NORs) in 7 *Rattus* species with special emphasis on the organizer differentiation and species evolution. *Proc Jap Acad* **55**: 481-486.

Yosida, T.H. 1980. *Cytogenetics of the black rat: Karyotype evolution and species differentiation*. University of Tokyo Press, Tokyo, pp 1-255.

Yuan, J.S., Reed, A., Chen, F., Stewart, C.N. 2006. Statistical analysis of real-time PCR data. *BMC Bioinformatics* **7**: 85.

Zankl, H., Huwer, H. 1978. Are NORs easily translocated to deleted chromosomes? *Hum Genet* **42**: 137-142.

Zdobnov, E.M., Apweiler, R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **7**: 847-848.

Zhang, Y., Skolnick, J. 2004. Automated structure prediction of weakly homologous proteins on a genomic scale. *Proc Natl Acad Sci USA* **101**: 7594-7599.

Zhang, J., Wang, X., Podlaha, O. 2004. Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Res* **14**: 845-51.

Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40.

Zhang, Y. 2009. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins* **9**: 100-113.

Zhao, S., Shetty, J., Hou, L., *et al*. 2004. Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. *Genome Res* **14**: 1851-1860.

Zhu, M., Zhao, S. 2007. Candidate gene identification approach: progress and challenges. *Int J Biol Sci* **3**: 420-427.

Zuo, S., Wu, X., Sun, P. *et al*., 2004. Study on the molecular epidemiology of hantavirus carried by hosts in northern suburb of Beijing. *Chin J Epidemiol* **25**: 421-424.

APPENDICES

Appendix 1. Rodents investigated in this study, with field numbers, geographic (or Ensembl database) origin indicated. The status (typhus-positive or typhus-negative) of each rodent is reported. Additionally, the table shows whether the sample was utilised for gene expression and/or sequence analysis. The qPCR primer design is also specified. * indicates that the sample was used only for syndecan-4 qPCR primer design. # marks the four samples used in the bioinformatic analysis of predicted 3D protein structure. "/" not applicable and dashes indicate absent data.

Species	Abbreviation	Field number	origin	Grid reference	*Orientia*	Gene expression	Sequencing	qPCR primer design	Reference for taxonomic identification
Homo sapiens	HSA	/	ENSG00000124145	-	/		X		/
Spermophilus tridecemlineatus	STR	/	ENSTOG00000009310	-	/		X		/
Dipodomys ordii	DOR	/	ENSFM00500000273193	-	/		X		/
Gerbillus nigeriae	GNI	N4172	Niger, Gangara	-	/		Unsuccessful		Dobigny *et al.* (unpubl.)
Gerbilliscus kempi	GKE	C333	Cameroon, Gamnaga	-	/		Unsuccessful		Dobigny *et al.* (2010b)
Microtus arvalis	MAR	Ma0504Y27	France, Jura	-	/		Unsuccessful		CBGP collections, Montpellier, France
Myodes arvalis	MAR	LS05040542	France, Jura	-	/		Unsuccessful		CBGP collections, Montpellier, France
Myodes glareolus	MGL	MgCz017	Tcheck republic, Studenec	-	/		Unsuccessful		CBGP collections, Montpellier, France
Myodes glareolus	MGL	MgSk010	Slovaquie, Rhozanovce	-	/		Unsuccessful		CBGP collections, Montpellier, France
Lemniscomys zebra	LZE	C305	Cameroon, Gamnaga	-	/		X	X	Dobigny *et al.* (2010b)
Arvicanthis niloticus	ANI	C352	Cameroon, Maga	10º50'N, 14 º56' E	/	X	X	X	Dobigny *et al.* (2010b)
Arvicanthis niloticus	ANI	N4147	Niger, Gaya	-	/		X		Dobigny *et al.* (unpubl.)
Mastomys natalensis	MNA	C350	Cameroon, Maga	10º50'N, 14º56' E	/		X	X	Dobigny *et al.* (2010b)
Mastomys erythroleucus	MER	C323	Cameroon, Baldama	10º58'N, 14º04'E	/	X	X	X*	Dobigny *et al.* (2010b)

Species	Code	ID	Location	Coordinates					Reference
Mastomys erythroleucus	MER	N4138	Niger, Kojimairi	-	/		X		Dobigny *et al.* (unpubl.)
Mastomys kollmannspergeri	MKO	C422	Cameroon, Kongola	10°37'N, 14°25'E	/		X	X	Dobigny *et al.* (2010b)
Praomys nov. sp.	Praomys	C319	Cameroon, Gamnaga	-	/		X	X	Dobigny *et al.* (2010b)
Apodemus flavicollis	AFL	AP0507Y01R	France, Jura	-	/		X		CBGP collections, Montpellier, France
Acomys johannis	AJO	C331	Cameroon, Gamnaga	-	/	X	X		Dobigny *et al.* (2010b)
Hapalomys delacouri	HDE	R5287	Thailand, Loei	-	/		X		Pagès *et al.* (unpubl.)
Bandicota indica	BIN	R4000	Thailand, Kalasin	16°49'N, 103°52'E	+	X	X		ANR-00121-05
Bandicota indica	BIN	R4212	Thailand, Loei	17°48'N, 101°67'E	-	X	X	X	ANR-00121-05
Bandicota savilei	BSA	R4285	Thailand, Phrae	18°23'N, 100°29'E	+	X	X	X*	ANR-00121-05
Bandicota savilei	BSA	R4308	Thailand, Phrae	18°23'N, 100°32'E	+	X	X#	X	ANR-00121-05
Bandicota savilei	BSA	R4178	Thailand, Phrae	18°23'N, 100°32'E	+	X	X		ANR-00121-05
Bandicota savilei	BSA	R4284	Thailand, Phrae	18°23'N, 100°29'E	-	X	X	X*	ANR-00121-05
Bandicota savilei	BSA	R4177	Thailand, Phrae	18°23'N, 100°32'E	-	X	X		ANR-00121-05
Bandicota savilei	BSA	R4307	Thailand, Phrae	18°23'N, 100°32'E	-	X	X		ANR-00121-05
Berylmys berdmorei	BBE	R4266	Thailand, Loei	17°49'N, 101°68'E	-	X	X	X	Pagès *et al.* (2010)
Berylmys bowersi	BBO	R4102	Thailand, Loei	17°48'N, 101°67'E	-	X	X	X*	ANR-00121-05
Leopoldamys edwardsi	LED	R4370	Thailand, Phrae	18°23'N, 100°35'E	-	X	X		Pagès *et al.* (2010)
Leopoldamys edwardsi	LED	R4222	Thailand, Loei	17°49'N, 101°68'E	+	X			Pagès *et al.* (2010)
Maxomys surifer	MSU	R4259	Thailand, Loei	17°49'N, 101°68'E	+	X	X	X	ANR-00121-05
Maxomys surifer	MSU	R4108	Thailand, Loei	17°49'N, 101°68'E	-	X	X		ANR-00121-05
Niviventer fulvescens	NFU	R4071	Thailand, Loei	17°49'N, 101°68'E	-	X	X	X*	ANR-00121-05
Rattus exulans	REX	R4193	Thailand, Loei	18°22'N, 100°27'E	+	X	X		ANR-00121-05
Rattus exulans	REX	R4126	Thailand, Loei	18°22'N, 100°27'E	+	X	X		ANR-00121-05
Rattus exulans	REX	R4211	Thailand, Loei	17°47'N, 101°66'E	+	X	X		ANR-00121-05
Rattus exulans	REX	R4208	Thailand, Loei	17°47'N, 101°66'E	+	X	X		ANR-00121-05

Rattus exulans	REX	R4197	Thailand, Phrae	18º22'N, 100º29'E	-	X	X	X*	ANR-00121-05
Rattus exulans	REX	R4218	Thailand, Phrae	17º49'N, 101º68'E	-	X	X	X	ANR-00121-05
Rattus exulans	REX	R4206	Thailand, Phrae	17º47'N, 101º66'E	-	X	X		ANR-00121-05
Rattus "losea"	RLO	R4203	Thailand, Phrae	18º22'N, 100º29'E	+	X	X	X	Pagès *et al.* (2010)
Rattus "losea"	RLO	R4230	Thailand, Loei	17º46'N, 101º69'E	-	X	X		Pagès *et al.* (2010)
Rattus "losea"	RLO	R4260	Thailand, Loei	17º48'N, 101º68'E	+	X	X		ANR-00121-05
Rattus "losea"	RLO	R4251	Thailand, Loei	17º48'N, 101º68'E	-	X	X		ANR-00121-05
Rattus "losea"	RLO	R4245	Thailand, Loei	17º48'N, 101º68'E	-	X	X		ANR-00121-05
Rattus tanezumi	RTA	R4003	Thailand, Kalasin	16º49'N, 103º52'E	-	X	X	X*	Pagès *et al.* (2010)
Rattus tanezumi	RTA	R4009	Thailand, Kalasin	16º49'N, 103º52'E	+	X	X	X	ANR-00121-05
Rattus tanezumi	RTA	R4005	Thailand, Kalasin	16º49'N, 103º52'E	+	X	X		ANR-00121-05
Rattus tanezumi	RTA	R4182	Thailand, Phrae	18º23'N, 100º35'E	+	X	X	X*	ANR-00121-05
Rattus tanezumi	RTA	R4018	Thailand, Phrae	18º23'N, 100º35'E	-	X	X	X*	ANR-00121-05
Rattus tanezumi	RTA	R4272	Thailand, Phrae	18º23'N, 100º32'E	-	X	X		ANR-00121-05
Rattus tanezumi	RTA	R4121	Thailand, Phrae	18º23'N, 100º35'E	-	X	X		ANR-00121-05
Rattus norvegicus	RNO	/	ENSRNOT00000019386	-	/		X#	X	/
Mus caroli	MCA	R4263	Thailand, Loei	17º48'N, 101º66'E	/	X	X		ANR-00121-05
Mus cervicolor	MCE	R4363	Thailand, Phrae	18º23'N, 100º29'E	/	X	X		ANR-00121-05
Mus cooki	MCO	R4101	Thailand, Loei	17º48'N, 101º68'E	/	X	X	X	ANR-00121-05
Mus cooki	MCO	R4110	Thailand, Loei	17º48'N, 101º67'E	/	X	X	X	ANR-00121-05
Mus cooki	MCO	R4317	Thailand, Phrae	17º48'N, 101º68'E	/	X	X		ANR-00121-05
Nannomys mattheyi	NMA	I345	South Africa	-	/	X	X#		CBGP collections, Montpellier, France
Mus musculus	MMU	/	ENSMUST00000017153	-	/		X#	X	/

Appendix 2. Multiple alignments of syndecan-4 protein sequences used in this study. Specimen numbers and abbreviations follow those in Appendix 1. Dots indicate identity with the *R. norvegicus* (RNO) syndecan-4 protein sequence, while dashes signify deletions (i.e., indels).

```
                    10        20        30        40        50        60        70
          ····|····|····|····|····|····|····|····|····|····|····|····|····|····|
RNO       APLLLLLLGGFPVAPGESIRETEVIDPQDLLEGRYFSGALPDDEDAGGLEQDSD-FELSGSGDLDDTEEP
BIN_R4000 ....................................................-................
BIN_R4212 .............X......................................-................
BSA_R4285 ....................................................-................
BSA_R4308 ....................................................-................
BSA_R4178 ....................................................-................
BSA_R4284 .................X..................................-................
BSA_R4177 ....................................................-................
BSA_R4307 ....................................................-................
BBE_R4266 .................................................D..-................
BBO_R4102 .............................F......................-................
LED_R4222 ...........V.............................V.E...-................R...
MSU_R4259 ...........V..................X................F...-................
MSU_R4108 ...........V..................X................F...-................
NFU_R4071 ...........V..............................X...E...-................
REX_R4193 ....................................................-.....X........
REX_R4126 ....................................................-................
REX_R4208 ..X.................................................-................
REX_R4197 ....................................X...............-.....X........
RTA_R4121 ...................................X.X..............-................
REX_R4218 ..X.................................................-................
REX_R4206 ....................................................-................
REX_R4211 ....................................................-................
RTA_R4009 ..................................................X.-................
RLO_R4260 ....................................................-................
RLO_R4251 ..............X...................................X.-................
RLO_R4245 ....................................................-................
RTA_R4003 ..................................................X.-................
RLO_R4203 ..............X...................................X.-................
RLO_R4230 ....................................................-................
RTA_R4005 .................................................XX.-................
RTA_R4182 .......................................X..........X.-................
RTA_R4018 ....................................................-................
RTA_R4272 .........................D...............X..X.....X.-................
MCA_R4263 .S.....X.....V..........................-----.D....................
MCE_R4363 ...........V............................-----.D....................
MCO_R4110 .S....F....V.............X..............-----.D.........X.....
MCO_R4104 .S....F....V.....................X......-----.D.........X.....
MCO_R4317 .S....F....V............................-----.D.........G.....
MMU       ............LV..........................-----.D....................
NMA       PL.........V...........X.N..........X...-----.D.........X....
HDE_R5237 ........................................D....F....D................
AJO_C331  .............A..A.......A............D......SPG.EY.D...........A.
LZE_C305  ...........V...............G.........................D................
ANI_C352  ...........V...............G.........................D................
ANI_N4147 .......XX....V.............G.........................D................
MNA_C350  X...F......AV.......................................D................
MER_C323  ....F......AV......................................D...........G..
MER_N4138 ....F......AV......................................D..XX.......G..
MKO_C422  ....F......AV......................................D...........A..
PRO C319  ...........VXX.................G...........X.......D.........G.Q...
```

208

```
                    80          90         100         110         120         130        140
                ----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
RNO             R T F P E V I S P L V P L D N H I P E N A Q P G I R V P S E P K E L E E N E V I P K R V P S - D V G D D D V S N K V S M S S T S Q G S N I F
BIN_R4000       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . S . . . .
BIN_R4212       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . S . . . .
BSA_R4285       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . - . . . . . . . . . . . . . . . S . . . .
BSA_R4308       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . S . . . .
BSA_R4178       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . S . . . .
BSA_R4284       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . S . . . .
BSA_R4177       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . - . . . . . . . . . . . . . . . S . . . .
BSA_R4307       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . - . . . . . . . . . . . . . . . S . . . .
BBE_R4266       . . . . . . . . . . . . . . . . . . . . V . . . . . . . . . . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . .
BBO_R4102       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . .
LED_R4222       . . . S . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . D . . . . . . . . . . . . . . . . . N . . L
MSU_R4259       . . . . . . P . . . . . . . . . . . . . . . . . . . . . . . . . . R . Q . . . . . . . . A - . . . X . . . . . . . . . . . . . . . . . . . .
MSU_R4108       . . . . . . P . . . . . . . . . . . . . . . . . . . . . . . . . . R . Q . . . . . . . . A - . . . X . . . . . . . . . . . . . . . . . . . .
NFU_R4071       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . I S . . . - . . . . . . . . . . . . . . . . . N . . .
REX_R4193       . . . . . . . . . . . . . . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . - G . . . . X . . . . . . . . . . . . . . . . . .
REX_R4126       . . . . . . . . . . . . . . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
REX_R4208       . . . . . . . . . . . . . . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
REX_R4197       . . . . . . . . . . . . . . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . - G . . . . X . . . . . . . . . . . . . . . . . .
RTA_R4121       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
REX_R4218       . . . . . . . . . . . . . . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . - G . . . . X . . . . . . . . . . . . . . . . . .
REX_R4206       . . . . . . . . . . . . . . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
REX_R4211       . . . . . . . . . . . . . . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
RTA_R4009       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X - G . . . . . . . . . . . . . . . . . . . . . . .
RLO_R4260       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
RLO_R4251       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
RLO_R4245       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
RTA_R4003       . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - X . . . . . . . . . . . . . . . . . . . . . . .
RLO_R4203       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
RLO_R4230       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
RTA_R4005       . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
RTA_R4182       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - G . . . . . . . . . . . . . . . . . . . . . . .
RTA_R4018       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . - G . . . . X . . . . . . . . . . . . . . . . . .
RTA_R4272       . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X - G . . X . X . . . . . . . . . . . . . . . . . .
MCA_R4263       . P . . . . . E . . . . . . . . . . . . . . . . . . L . . . . . . . . . . . . . . . A . - . . . - . M . . . . . . . . A . . . . .
MCE_R4363       . P . . . . . E . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . A . - . . . - . M . . . . . . . . A . . . . .
MCO_R4110       . P . . . . . E . . . . . . . . . . . . . . . . . . L . . . . . . . . . . . . . . . A . - X . . - . M . . . . . . . . A . . . . .
MCO_R4104       . P . . . . . E . . . . . . . . . . . X . . . . . . L . . . . . . . . . . . . . . . A . - X . . - . M . . . . . . . . A . . . . .
MCO_R4317       . P . . . . . E . . . . . . . . . . . . . . . . . . L . . . . . . . . . . . . . . . A . - X . . - . M . . . . . . . . A . . . . .
MMU             . P . . . . . E . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . A . - . . . - . M . . . . . . . . A . . . . .
NMA             . P . . . . . E . . . . . . X X X . . . . . . . . . . . . . . . . . . . . . . . . G . . . . . - . . . . . . . . . . . . A . . . . .
HDE_R5237       . . . . G G . O . . . . . N . . . X . . H . V . . . . . . . . . . . . . . . . . V . . . . . - E . . N - . I . . . . A . . . . .
AJO C331        . . . . . . . Q . . . . N . . L . . K P . . . . . . . . . . . . . . . . . . . . . . . - - G T E N . I . . . . . A . D . S . .
LZE_C305        . N . . D . . Q . . . . . . . . . . . . . . . . . . . . . . . . . . . I . . . - . D . - . . . . . . . . . A . . . . .
ANI_C352        . N . . D . . Q . . . . . . . . . . . . . . . . . . . . . . . . . . . L . . . - . D . - X . . . . . . . . . A . . . . .
ANI_N4147       . N . . D . . Q . . . . . . . . . . . . . . . . . . . . . . . . . . . L . . . - . D . - . . . . . . . . . A . . . . .
MNA_C350        K . . . . . . Q . . . . . . . . . G . . . S . . . . . . . . . . . . . . . . . . . . - . . . . . . . . . . . . A . . . . .
MER_C323        K . . . . . . Q . . . . . . . . . G . . . S . . . . . . . . . . . . . . . . . . . . - . . . . - . . . . . . . . A . . . . .
MER_N4138       K . . . . . . Q . . . . . . . . . G . . . S . . . . . . . . . . . . . . . . . . . . - . . . . - . . . . . . . . A . . . . .
MKO_C422        K . . . . . . Q . . . . . . . . . G . H . S . . . . . . . . . . . . . . . . . . . . - . . . . - . . . . . . . . A . . . . .
PRO C319        G . . . . . . Q . . . . . . . . . G . G . . . X . . . . S . . . . . . . . . . . . . - G . . . - . . . . . . . . . A . . . . .
```

209

```
                    150         160         170         180         190
             - - - | - - - - | - - - - | - - - - | - - - - | - - - - | - - - - | - - - - | - - - - | - -
RNO          E R T E V L A A L I V G G V V G I L F A V F L I L L L V Y R M K K K D E G S Y D L G K K P I Y K K A P T K E F X
BIN_R4000    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
BIN_R4212    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
BSA_R4285    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X .
BSA_R4308    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
BSA_R4178    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
BSA_R4284    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
BSA_R4177    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
BSA_R4307    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
BBE_R4266    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
BBO_R4102    . . . . . . . . . . . . . . . . . . . V . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
LED_R4222    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
MSU_R4259    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
MSU_R4108    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
NFU_R4071    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
REX_R4193    . . . . . . . . . . . . . . . . . . . V . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
REX_R4126    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
REX_R4208    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
REX_R4197    . . . . . . . . . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
RTA_R4121    . . . . . . . . . . . X . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
REX_R4218    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
REX_R4206    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
REX_R4211    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
RTA_R4009    . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
RLO_R4260    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
RLO_R4251    . . . . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
RLO_R4245    . . . . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . X . . .
RTA_R4003    . . . . . . . . . . . X . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
RLO_R4203    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . L . . . . . . . X . . .
RLO_R4230    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
RTA_R4005    . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
RTA_R4182    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
RTA_R4018    . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
RTA_R4272    . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
MCA_R4263    . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . X . . .
MCE_R4363    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
MCO_R4110    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
MCO_R4104    . . . . . . . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
MCO_R4317    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
MMU          . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . N . . .
NMA          . . . . . . . . . . . . . . . . V . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
HDE_R5237    . . . . . . . . . . . . . I . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
AJO C331     . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
LZE_C305     . . . . . . . . . . . . . . . . I . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . X .
ANI_C352     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X .
ANI_N4147    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X .
MNA_C350     . . . . . . . . . . . X . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
MER_C323     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
MER_N4138    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . . .
MKO_C422     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X .
PRO C319     . K . . . . . . . . . M . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . X . X -
```

210