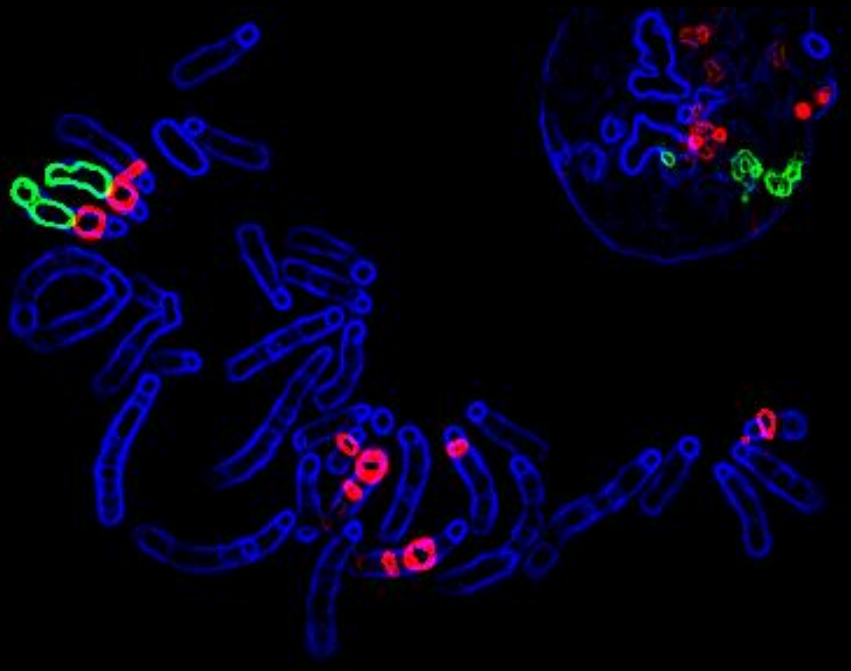
Universidade de Trás-os-Montes e Alto Douro

KARYOTYPE RESTRUCTURING IN RODENTIA:

FROM EVOLUTION TO CANCER



Sandra Louzada Gomes Pereira

2013

UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO

KARYOTYPE RESTRUCTURING IN RODENTIA: FROM EVOLUTION TO CANCER

SANDRA LOUZADA GOMES PEREIRA

2013

Este trabalho foi expressamente elaborado com vista à obtenção de grau de Doutor em Genética Molecular, Comparativa e Tecnológica.

Tese de Doutoramento financiada pela Fundação para a Ciência e Tecnologia, "Programa Operacinal Potencial Humano – POPH", co-financiado pelo Fundo Social Europeu (POPH/FSE) e por fundos nacionais (POPH-QREN).

Orientadora: Professora Doutora Raquel Chaves, Centro de Genómica e Biotecnologia, Universidade de Trás-os-Montes e Alto Douro, CGB/IBB/UTAD

Co-Orientadora: Doutora Filomena Adega, Centro de Genómica e Biotecnologia, Universidade de Trás-os-Montes e Alto Douro, CGB/IBB/UTAD

AGRADECIMENTOS

A presente tese de doutoramento não teria sido possível sem a colaboração de muitas pessoas e instituições às quais gostaria de exprimir os meus agradecimentos.

À Universidade de Trás-os-Montes e Alto Douro (UTAD), na pessoa do seu Magnífico Reitor, Professor Doutor Carlos Alberto Sequeira e também ao anterior Reitor, Professor Doutor Armando Mascarenhas Ferreira pela disponibilização dos meios para a realização desta tese.

À Escola de Ciências da Vida e do Ambiente da UTAD, na pessoa do seu Presidente Professor Doutor António Fontainhas Fernandes, pela disponibilização dos meios necessários para este trabalho.

Às Comissões Directivas do Curso Doutoral em Genética Molecular Comparativa e Tecnológica, na pessoa do actual Director Professor Doutor Valdemar Carnide, e do antigo Director Professor Doutor Henrique Guedes-Pinto pelas condições que foram fornecidas e que permitiram a realização deste trabalho.

Ao Centro de Genómica e Biotecnologia do Instituto de Bioengenharia e Biotecnologia, na pessoa do Professor Doutor Henrique Guedes-Pinto, pelas condições facilitadas que permitiram a realização deste trabalho.

À Fundação para a Ciência e a Tecnologia, pela confiança depositada na concessão da bolsa de doutoramento (SFRH/BD/25813/2005), sem a qual teria sido impossível a concretização deste trabalho. Quero ainda referir que parte do trabalho se integrou no projecto com a referência POCI/BIA-BCM/58541/2004, financiado pela FCT e COMPETE/QREN/UE.

À Professora Doutora Raquel Chaves, minha orientadora, por me ter aceitado como orientanda e pela confiança que depositou em mim. Agradeço-lhe todos os ensinamentos (quer nas aulas, quer no laboratório), a disponibilidade constante e a preocupação. Foi para mim um privilégio poder trabalhar com uma pessoa que tanto me inspirou pela sua inteligência, sabedoria e capacidade de trabalho e que ao mesmo tempo sempre demonstrou a simplicidade de uma amiga com quem podíamos contar. Obrigada por me guiar durante esta grande aventura.

À Doutora Filomena Adega, minha co-orientadora, por me ter acolhido como orientanda, por tudo o que me ensinou quando cheguei ao laboratório, pela motivação constante e pelos conselhos. Agradeço ter sido um valioso exemplo de investigadora e por ter contribuído para o meu sucesso durante esta aventura. Vou para sempre recordar, o carinho, as conversas, as cantorias e sobretudo a amizade que transcendeu a parte profissional e que permanece.

Ao Doutor Jiri Rubes por me ter recebido no seu laboratório e a todos os membros da sua equipe com quem tive o prazer de trabalhar no período de tempo que passei lá, por me terem ajudado, mostrado novas técnicas e pelos bons momentos que passamos juntos.

Aos meus queridos colegas de laboratório, Ana Borges (AnaLu), Ana Vieira da Silva (Netinha), Ana do Paço (Paçito), Jorge Pereira (Jójó), Sara Santos (Sarita), Susana Meles (Susy) e João Coutinho (Jonhy), com os quais formei uma família que me acompanhou ao longo destes últimos anos. Obrigada por tornarem os meus dias tão agradáveis com as conversas, a risota, as "parvoíces", mas sobretudo pelo apoio, por toda a ajuda, companheirismo e pelo ombro amigo. Todos vocês marcaram a minha vida de diferentes maneiras, e por serem excelentes pessoas (mesmo do melhor que há!), NUNCA vos irei esquecer. Começamos como colegas de doutoramento e hoje a nossa amizade vai para além das bancadas do laboratório. Vou sempre recordar os cafezinhos e as longas conversas com a Ana Lu, as viagens que partilhei com a minha companheira Netinha, os comentários hilariantes e eficiência da Paçito, as piadas do Jorge (mas ele terá um agradecimento especial!), a disponibilidade e apoio constante da Sarita, a boa disposição e as histórias da Susy e o "relax" do Johny. Gostaria de fazer um agradecimento especial ainda á Ana Neta pela rápida e pormenorizada revisão da minha tese.

Aos restantes colegas de curso de Pós-graduação em especial ao Rui Abreu, Cláudia Baptista, Fernando Pimentel e Daniela Sá pela amizade.

À Karine com quem partilhei bancada, pela amizade e apoio.

Aos meus primos, Lídia, Pedro e ao meu querido Tomás, pelos belos e divertidos momentos que passei com vocês nos "intervalos" do trabalho e que muito me ajudaram. Pelo vosso apoio constante e disponibilidade absoluta. Um agradecimento especial ao pequeno Tomás pelas brincadeiras.

Aos meus restantes familiares, aos meus primos (em especial à Marlene que sempre me deu força e incentivo), tios e avós, por todo o apoio e pelos bons momentos. E também á minha "nova" família que me acolheu de braços abertos, em especial aos meus sogros, Maria e Gomes, e à minha cunhadinha Patrícia pela preocupação constante e carinho. Foi na família que encontrei o suporte e o carinho que foram essenciais durante todo este percurso, e devo dizer que me sinto uma sortuda por ter uma família assim.

À minha irmã Joana, umas das pessoas mais importantes da minha vida, por me ter ouvido tantas vezes e pela contenção das minhas angústias, pelos conselhos, por me mostrar que as coisas podem ser vistas de uma outra perspectiva e por acreditar em mim. O teu apoio e amizade foram muito importantes.

Aos meus pais, que eu amo muito, agradeço tudo o que fizeram por mim ao longo da vida, pois é graças a eles que cheguei a esta etapa. Por me terem mostrado o que é importante, por tudo que me ensinaram, pelo apoio incondicional, por terem sempre acreditado no meu valor, e por fazerem tudo para que fosse feliz. Muito muito obrigada.

Ao meu querido esposo Jorge, o meu príncipe e amor da minha vida, por me mostrar o que é a verdadeira felicidade e também por me ter "aturado" nos meus momentos menos bons.

A todas as outras pessoas que directa ou indirectamente me ajudaram, o meu muito obrigada!

"I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale".

Marie Curie (1867 - 1934)

ABSTRACT

The order Rodentia represents the most abundant and diversified order of living mammals. Record high rates of karyotype evolution are found in the rodent's superfamily Muroidea (the most evolutionary successful mammalian species) making them the perfect organisms for studying chromosome evolution and powerful tools also in the study of chromosome rearrangements and their consequences in cancer. The major goal of the present thesis was the analysis of the karyotype restructuring dynamics, both during species evolution and cancer.

Chromosomal evolutionary events are disclosed by the comparative analysis of different species karyotype using cytogenetic techniques, allowing the fast generation of large scale comparative maps in diverse groups including Rodentia.

The present thesis describes the construction of high resolution chromosome maps of three Rodentia species, namely one Muridae species, *Praomys tullbergi*, and two Cricetidae *Cricetus cricetus* and *Peromyscus eremicus*. One important outcome presented here is the delineation of the Ancestral Muroidea Karyotype (AMK), based in the analysis of the *Mus musculus* syntenic associations outlined by several works and supported by the results obtained. The assembling of these comparative maps permitted the disclosure of genome architecture, as well as the delineation of the chromosome evolutionary history since the common Muroidea ancestor for these species. *Peromyscus eremicus* reveled to possess a highly conserved genome sharing most of the identified syntenic association with the AMK. *Cricetus cricetus* and *Praomys tullbergii*, on the other hand, showed to have more derivative genomes accounting a large number of large-scale rearrangements occurred since the AMK, mostly fusion events. The construction of the comparative maps allowed also the identification of the evolutionary breakpoint regions. The presence of repetitive sequences at evolutionary breakpoints has been shown by whole genome alignment studies, and constitutive heterochromatin (CH) has been considered as hotspot for structural chromosome rearrangements. Here we have found a high co-localization of CH with the identified evolutionary breakpoints for the species in analysis (*P. tullbergi*, *C. cricetus* and *Peromyscus eremicus*), clearly indicating its involvement in the structural chromosome rearrangements. Besides, its constituents, such as the satellite DNA, are most likely the responsible for promoting the genomic plasticity and consequently the higher rates of chromosome rearrangements observed.

Satellite DNAs and thus thought to be implicated in karyotype restructuring, both in species evolution and cancer. Satellite DNAs are highly repeated sequences, characterized by a a dynamic behavior and the major constituents of functional centromeres; however being also found in telomeres and interstitial positions. The potential functional importance of satellite DNAs and

the existence of a whole range of satellite sequences either conserved or divergent, even between closely related species, highlight the importance of studying satellite DNA. In the presented study two repetitive sequences (CCR4/10sat and PMSat) were isolated *de novo* using laser microdissection, physically mapped and molecularly characterized. Both sequences revealed to be shared by different rodent species enlightening a dynamic behavior and possible implication in karyotype architecture in rodents (Cricetidae). While CCR4/10sat evolution seems to be related with intragenomic movements, the evolutionary pathway of PMSat occurred through copy number variations, culminating in different profiles.

Cancer chromosomes are known to exhibit high levels of complexity and the ability to constantly evolve. Understanding the genetic etiology of the cancer genome is important to comprehend the mechanisms for cancer initiation and progression. The last part of this thesis was dedicated to the genetic/cytogenetic characterization of two DMBA-induced rat mammary tumor cell lines: HH-16 cl.2/1 and HH-16.cl.4. The cytogenetic analysis of both cell lines revealed significant changes in their karyotypes, suggesting the presence of chromosomal instability (CIN) and chromosome structure instability (CSI). It has been demonstrated that CSI can influence tumorigenesis by deregulating expression of specific target genes or by promoting gene fusion. In the present case it was clear the implication of chromosome rearrangements and karyotype restructuring in tumor progression, specifically by causing changes in two oncogenes copy number (*Mycn* and *Erbb2*). Both cell lines showed different expression profiles regarding the intensely studied *Erbb2*. Besides, the expression of *Erbb2* in the HH-16.cl.4 rat cell line appears to be affected by global genome demethylation (after 5-Aza-2'-deoxicitidine), suggesting the action of negative regulators of *Erbb2* expression. The different outcomes for both tumor cell lines, regarding cytogenetic characterization, gene expression and methylation analysis, suggests different mechanisms involved in tumor progression. This study highlights HH-16 cl.2/1 and HH-16.cl.4 potential as models for studying *Erbb2* associated mechanisms and as experimental tools to assist in the generation of new biotherapies.

The present thesis resulted in the elaboration of five articles that were submitted/published in scientific journals.

SUMÁRIO

A ordem Rodentia representa a mais abundante e diversificada ordem de mamíferos. A análise dos cariótipos/genomas de roedores da superfamília Muroidea tem revelado a ocorrência de elevadas taxas de evolução para estas espécies, o que as torna bons modelos para o estudo da evolução de cromossomas, dos rearranjos cromossómicos e das consequências dos mesmos durante o processo tumoral. O principal objectivo desta tese consistiu na análise da dinâmica envolvida na reestruturação dos cariótipos durante a evolução de algumas espécies de roedores e durante o processo de cancro num modelo celular da espécie *Rattus norvegicus*.

Os eventos que ocorrem durante a evolução dos cromossomas têm sido desvendados através da análise comparativa dos cariótipos de espécies diferentes usando técnicas de citogenética ("Comparative chromosomics") que permitem a elaboração de mapas comparativos entre os mais diversos grupos, incluindo os Rodentia. Na presente tese foram construidos mapas cromossómicos de elevada resolução de três espécies de roedores, *Praomys tulbergi* (PTU, Muridae), *Cricetus cricetus* (CCR) e *Peromyscus eremicus* (PER), ambas Cricetidae. Um resultado importante foi o delineamento do cariótipo ancestral putativo dos Muroidea (AMK), baseado nas associações sinténicas do genoma modelo *Mus musculus* determinadas neste trabalho e em trabalhos anteriores. A análise destes mapas comparativos permitiu desvendar a arquitectura, assim como delinear a história evolutiva dos cromossomas desde o ancestral Muroidea até aos cariótipos das espécies em análise. A espécie PER revelou possuir um genoma extremamente conservado, partilhando muitas associações sinténicas com o AMK. As restantes espécies, CCR e PTU possuem genomas mais derivativos, demonstrando a ocorrência de um considerável número de rearranjos desde o AMK. A construção dos referidos mapas permitiu ainda a identificação das regiões de "breakpoint", regiões estas que apresentam uma grande instabilidade. A presença de sequências repetitivas (constituintes da heterocromatina constitutiva -HC) nesses "breakpoints" evolutivos tem sido demonstrada em vários estudos, tendo sido a HC considerada como uma região propícia ("hotspot") à ocorrência de rearranjos cromossómicos. Nos genomas analisados (PTU, CCR e PER) foi encontrada uma elevada co-localização da HC com os "breakpoints" evolutivos, indicando o envolvimento da HC na ocorrência dos rearranjos cromossómicos observados, sendo o seu maior constituinte - o DNA satélite - o melhor candidadato a promover esta plasticidade. O DNA satélite é composto por sequências altamente repetidas e dinâmicas e é o principal constituinte dos centrómeros funcionais, tendo também sido descrito em regiões teloméricas e intersticiais.

O cariz enigmático das sequências de DNA satélite , conjugado com a existência de uma grande variedade de famílias diferentes, conservadas ou divergentes entre espécies e possível função nos genomas, ilustram a importância do estudo destas sequências. No presente trabalho é descrito o isolamento *de novo* de duas sequências repetitivas (CCR4/10sat e PMSat) por microdissecção a laser, e a sua caracterização molecular. Ambas as sequências revelaram estar presentes no genoma de diferentes espécies de roedores, apresentando, no entanto diferentes perfis, facto que evidencia a sua importância e dinamismo na reestruturação destes genomas. Enquanto que a evolução do CCR4/10sat parece dever-se a movimentos intragenómicos, as características demonstradas pelo PMSat indicam que o seu percurso evolutivo está relacionado com variações no número de cópias.

A constante evolução e o elevado grau de dinamismo são também característicos dos genomas tumorais. A parte final do presente trabalho descreve a caracterização genética/citogenética de duas linhas celulares comerciais de tumor de mama de ratazana: HH-16 cl.2/1 e HH-16.cl.4. A análise citogenética revelou a presença de alterações consideráveis nos seus cariótipos, sugerindo a ocorrência de instabilidade cromossómica ("chromosomal instability"- CIN) e instabilidade estrutural dos cromossomas ("chromosome structure instability"- CSI). Vários estudos revelaram que a CSI pode influenciar a tumorigénese através da desregulação de genes específicos ou mediante fusão de genes. O trabalho permitiu determinar os rearranjos cromossómicos que das linhas celulares em análise, bem como de dois oncogenes - *Mycn* e *Erbb2*, afectados por esses mesmos rearranjos. Ambas as linhas celulares revelaram diferentes níveis de expressão do gene *Erbb2*. Para além disso, na linha HH-16.cl.4 a expressão deste gene parece ser afectada pela desmetilação global do genoma (tratamento com 5-aza-2'-deoxicitidina), sugerindo a acção de reguladores negativos da expressão do *Erbb2*. Tendo por base os resultados de caracterização citogenética, expressão de genes e análise de metilação, sugeriu-se o envolvimento de mecanismos diferentes na progressão tumoral das duas linhas celulares, evidenciando o potencial das linhas HH-16 cl.2/1 e HH-16.cl.4 como modelo celular para o estudo dos mecanismos epigenéticos associados ao *Erbb2*, bem como potenciais ferramentas experimentais para o desenvolvimento de novas bioterapias.

A presente tese de doutoramento resultou na elaboração de cinco artigos científicos que foram submetidos/publicados em jornais científicos.

LIST OF PUBLICATIONS

This thesis is based on the collection of the following papers throughout the PhD period:

Paper I: Chaves R, Louzada S, Meles S, Wienberg J, Adega M (2012) *Praomys tullbergi* (Muridae, Rodentia) genome architecture decoded by comparative chromosome painting with Mus and Rattus. *Chromosome Res.* 20(6):673-683.

Paper II: Vieira-da-Silva A*, Louzada S*, Adega F, Chaves R (2012) A high-resolution comparative chromosome map of *Cricetus cricetus* and *Peromyscus eremicus* reveals the involvement of constitutive heterochromatin in breakpoint regions (Submitted)
* Vieira-da-Silva A and Louzada S contributed equally to the present work (2012)

Paper III: Louzada S, Paço 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(8): 1149-1155.

Paper IV: Louzada S, Vieira-da-Silva A, Kubickova S, Adega F, Rubes J, Chaves R (2012) An ancient satellite DNA in *Peromyscus* genome that evolves by copy number fluctuation: does the sequence matters? (Submitted)

Paper V: Louzada S, Adega F, Chaves R (2012) Defining the Sister Rat Mammary Tumor Cell Lines HH-16 cl.2/1 and HH-16.cl.4 as an In Vitro Cell Model for Erbb2. *PLoS One* 7(1):e29923.

From the work described resulted the following communications published in refereed proceedings of conferences:

- Louzada S, Adega F, Chaves R (2010) Molecular cytogenetic characterization of a cell clone from a rat fibrosarcoma cell line. *Chromosome Res* 18:718.

- Vieira-da-Silva A, Louzada S, Adega F, Guedes-Pinto H, Chaves R (2010) An orthologous satellite DNA family between Muridae and Cricetidae (Rodentia). *Chromosome Research* 18:7

- Louzada S, Vieira-da-Silva A, Adega F, Guedes-Pinto H, Chaves R (2009) Mouse and rat uncover the chromosome restructuring in *Peromyscus eremicus* (Cricetidae, Rodentia) – Chromosomes 1 and 5 in focus. *Chromosome Research* 17(1): S168.

- Vieira-da-Silva A, Louzada S, Adega F, Guedes-Pinto H, Chaves R (2009) Comparative analysis of two *Cricetus cricetus* chromosomes with *Mus musculus* and *Rattus norvegicus* using chromosome painting. *Chromosome Research* 17(1):S167.

- Adega F, Louzada S, Vieira-da-Silva A, Guedes-Pinto H, Kofler A, Wienberg J, Chaves R (2008) The freewill of evolution in the structuring of Rodentia genomes. *Chromosome Research* 0: 19.

- Louzada S, Vieira-da-Silva A, Paço A, Kubickova S, Adega F, Guedes-Pinto H, Rubes J and Chaves R (2008). Evolutionary chromosome repositioning of orthologous satellite DNA in the related genomes *C. cricetus* and *P. eremicus* (Rodentia,Cricetidae). *Chromosome Research*16:1046.

- Adega F, Louzada S, Vieira-da-Silva A, Guedes-Pinto H, Kofler A, Wienberg J and Chaves R (2008) The freewill of evolution in the restructuring of Rodentia genomes. *Chromosome Research* 16: 1046.

- Louzada S, Vieira-da-Silsa A, Kubickova S, Adega F, Guedes-Pinto H, Chaves R, Jíri J (2007) *Cricetus cricetus* and *Peromyscus eremicus* (Rodentia, Cricetidae) share common repetitive DNA sequences. *Chromosome Research* 15(2):43.

ABBREVIATIONS

A, C, T, G	Adenine, cytosine, thymine, guanine
ACdK	Ancestral Cricetidae karyotype
ACnK	Ancestral Cricetinae karyotype
AEK	Ancestral *Ellobius* karyotype
AMdK	Ancestral Muridae karyotype
AMiK	Ancestral *Microtus* karyotype
AMK	Ancestral Muroidea karyotype
AMnK	Ancestral Murinae karyotype
ASdK	Ancestral Sciuridae karyotype.
BAC	Bacterial Artificial Chromosomes
BER	Basepair-excision repair
bp	Base pair
CCR	*Cricetus cricetus*
CENP-A	Centromere protein A
CENP-B	Centromere protein B
CH	Constitutive heterochromatin
CIN	Chromosomal instability
CSI	Chromosome structure instability
DMBA	7,12-dimethylbenz[a]anthrazene
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferases
DSBs	Double stranded breaks
FISH	Fluorescent in situ hybridization
HOR	Higher-order repeat
HP1	Heterochromatin protein 1
HSR	Homogeneously staining region
Kb	Kilo bases
LINE1	Long interspersed elements 1
Ma	Million years ago
MAR	*Microtus arvalis*
Masat	Mouse major satellite
Mb	Mega bases
MFISH	Multiplex FISH
Misat	Minor satellite
MMR	Mismatch-repair
MMU	*Mus musculus*
MS3	Mouse satellite 3
MS4	Mouse satellite 4
NER	Nucleotide-excision repair
PER	*Peromyscus eremicus*
PRAT	Major satellite from *Palorus ratzeburgii*
PSU	*Phodopus sungorus*
PTU	*Praomys tullbergi*
RAK	Ancestral Rodentia karyotype
RNA	Ribonucleic acid
RNO	*Rattus norvegicus*
RPCS	Repetitive PuvII *Ctenomys* sequence

rRNA	Ribosomal RNA
satDNA	Satellite DNA
siRNA	Small interfering RNAs
SKY	Spectral karyotyping

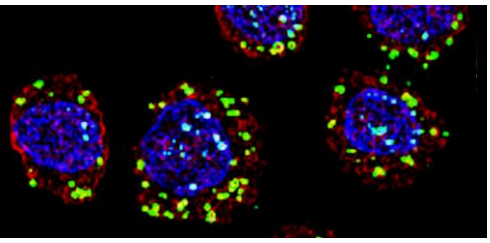
INDEX

FIGURES AND TABLES INDEX

CHAPTER I – INTRODUCTION

CHAPTER III – GENERAL DISCUSSION AND FUTURE PERSPECTIVES

CHAPTER I |



INTRODUCTION

The human genome (entire hereditary information of an organism that is encoded in their DNA) provides the underlying code for human biology (ENCODE Project Consortium 2012). Cells within an organism contain a complete copy of these instructions, written in the four-letter language of DNA (A, C, T and G). These nucleic acids are arranged into units called genes, which are organized in the **chromosomes** (Reece 2004). This organization ensures not only proper gene function but also an accurate distribution of genes to daughter cells during cell division. Chromosomes are thus the ultimate determinants of the organization of all living organisms (Sumner 2003). The entire chromosome set of a species is known as a **karyotype**, which can be thought of as a global map of the nuclear genome.

Genomes have the ability to evolve throughout local changes in nucleotide sequences or by changes in the karyotypes, by means of chromosome rearrangements that can result in dramatic phenotypic consequences and are assumed to play an important role in the evolution of species and in cancer. Reproductive isolation and tumorigenic karyotypic transformation can be initiated through the same structural rearrangements, therefore karyotype restructuring can drive both speciation and carcinogenesis (Ye *et al.* 2009).

The major goal of the present thesis was the analysis of the karyotype restructuring dynamics, both during evolution and cancer, in Rodentia species. Specifically, three objectives may be outlined: i) the comparative analysis of the genome of various Rodentia species by means of comparative chromosome painting in an evolutionary perspective; ii) the molecular and cytogenetic characterization of satellite DNA families, and the analysis of their dynamic behavior in the light of the species' karyotype evolution; iii) cytogenetic and molecular characterization of two rat tumor cell lines, highlighting the chromosome rearrangements effect in gene expression, and validating the use of these cell lines as cellular models for breast cancer research, namely in the elucidation of the epigenetic events involved in the regulation of Erbb2 expression.

This thesis is divided in three major parts: **chromosomal evolution** (Section 1; Papers I and II), **satellite DNA dynamics and evolution** (Section 2; Papers III and IV) and **chromosomes and cancer** (Section 3; Paper V). For each of these parts a literature review was

performed; the Results chapter will be presented as individual papers being some of them already published, and others submitted. A general discussion is made at the end in order to integrate and correlate all the data achieved.

1. CHROMOSOMES AND EVOLUTION

In the 1900s a series of experiments by Theodor Boveri gave the definitive demonstration that chromosomes are the vectors of heredity. Eukaryotic chromosomes present differences in morphology (shape and size) within the karyotype and karyotypes vary in terms of number and organization even between closely related species (Sumner 2003).

Going back to 1859, Darwin introduced the biologists to the concept that allied species are descended from a common ancestor and that species change gradually over long periods of time. The idea of evolution as the principle for the origin of biodiversity can be applied to chromosomes, being the chromosomal diversity found the result of the action of different mechanisms during the process of chromosome evolution, elucidating the high plasticity of the genomes at the chromosomal level. Moreover, the wide diversity of karyotypes found, combined with evidence that chromosomal rearrangements might reduce the fertility of heterozygous hybrids (King 1993), has led some researchers to argue for a causative role of chromosomal change in speciation.

1.1 GENOME CONSERVATION AND KARYOTYPE RESTRUCTURING

The first approach in the study of chromosomal evolutionary events was the comparative analysis of several species karyotype. Initial attempts to identify chromosome homologies were based on chromosome banding patterns (Dutrillaux *et al.* 1980, Nash and O'Brien 1982). Comparative studies were facilitated when molecular techniques were incorporated into cytogenetics, allowing DNA level comparison even between phylogenetically distant or highly rearranged species. With the advent of advanced molecular approaches, a new term emerged - **Comparative Chromosomics** - used to define the field of cytogenetics using methodologies which allow further resolution of comparative maps (Graphodatsky 2007). Comparative chromosome painting and Zoo-FISH revealed to be a powerful tool in comparative chromosome studies, allowing the construction of large-scale comparative maps mostly in mammalian groups. This technique is based in cross-species fluorescent *in situ* hybridization (FISH) using chromosome-specific DNA sequences as painting probes allowing the definition of chromosomal homologue segments between species (Figure 1.1) (Wienberg *et al.* 1990, Scherthan *et al.* 1994).

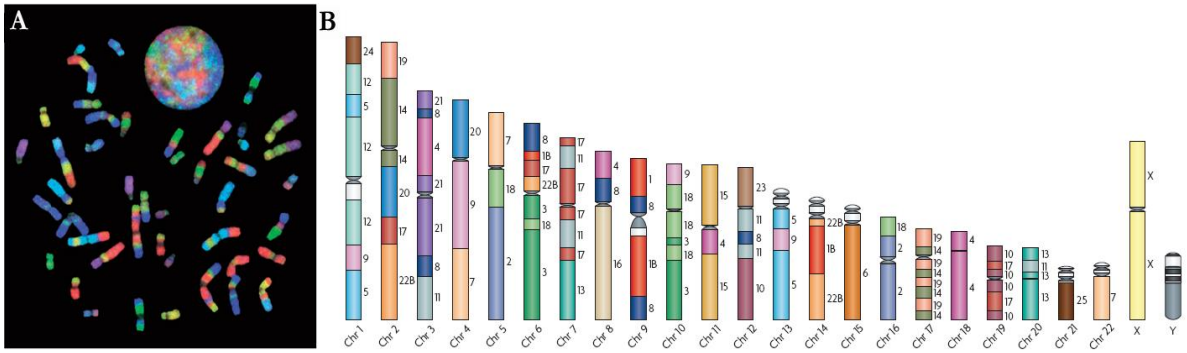


Figure 1.1| Example of the use of Zoo-FISH for the construction of comparative maps. A) Human metaphase and interphase nucleus after hybridization with a chromosome-specific paint probe set derived from gibbon chromosomes. B) The analysis of the painting experiments enabled the construction of a homology map of gibbon chromosomal segments on human chromosomes where syntenic associations can be observed, e.g. syntenic association 7/9 in chromosome 4 (adapted from Ferguson-Smith and Trifonov 2007).

Comparative chromosome painting permitted the disclosure of **syntenic segments**, defined as large blocks of DNA often extending to whole chromosomes or chromosome arms, which are shared by different species. Contiguous syntenic segments which are homologous to regions belonging to different chromosomes in another species are designated as **syntenic associations** (exemplified in Figure 1.1B) (Froenicke 2005). Comparative studies showed remarkable interspecies chromosome segments conservation, but also demonstrated that between species the syntenic blocks are assembled in different combinations, resulting in distinct chromosome number and chromosome morphology, reflected in the karyotype variability (Ferguson-Smith and Trifonov 2007). The reassembling of those segments and consequent karyotype restructuring is promoted by chromosome rearrangements, being the most common the translocations, inversions (paracenric and pericentric), fusions and fissions (Pevzner and Tesler 2003a). Duplications, deletions and heterochromatin additions/eliminations were also responsible for changes in chromosomes during evolution (Bailey *et al.* 2004, Adega *et al.* 2009). Conserved segments that are fused together in one species can be separated on different chromosomes in another. Chromosome numbers can increase or decrease by fission or fusion events, and segments within blocks can be inverted and centromeres repositioned. The analysis of the most parsimonious scenarios is the dominant approach in genome rearrangement study uncovering the evolutionary history (Ferguson-Smith and Trifonov 2007).

Comparative chromosome painting allowed the fast generation of large-scale comparative maps in Eutheria (placental mammals), however, its resolution presents some limitations. This methodology fails to determine the orientation of each conserved block within a chromosome, it does not allow the identification of intrachromosomal rearrangements such as inverted segments, and it is not efficient in revealing syntenies between distantly relates species

(Murphy *et al.* 2005). Cytogenetic analysis can be in some cases complemented with the use of BAC (Bacterial Artificial Chromosomes with cloned DNA fragments), allowing the detection of more detailed homologies (Goureau *et al.* 2001). With the advent of large-scale genome sequencing of eukaryotic genomes and the use of powerful algorithms to promote their alignment and comparative analysis, an exquisite molecular resolution at the level of single-base pair differences, as well as identification of gene order and changes in synteny was accomplished (Froenicke *et al.* 2006). Complete sequencing of genomes has confirmed the extensive levels of conserved synteny originally found by cytogenetic comparative mapping, but the high density of markers afforded by complete sequence also results in a more complex view of chromosomal evolution, with remarkable levels of intrachromosomal rearrangement (Eichler and Sankoff 2003). Nevertheless, the drawbacks pointed to comparative chromosome painting do not invalidate this methodology, in fact, several species were analysed using this technique and many more will be, as this is clearly easier and faster than sequencing a species genome. In this way, for the species whose genome has already been sequenced, the use of painting probes derived from different species combined with comparative sequencing projects is definitely the more efficient approach in comparative studies.

1.2 RESOLVING PHYLOGENIES IN MAMMALIA

Traditionally, the analysis of mammalian phylogenies was restricted to fossil records and morphological characters. In the past years, data from a range of research disciplines, such as molecular systematics, genome sequencing and comparative cytogenetics, have disclosed the evolutionary relationships between humans and their mammalian relatives. Mammalian phylogeny and evolution is now the driving force behind comparative genomic analysis, investigating the details of mammalian genomes and how they evolved (e.g. Engelbrecht *et al.* 2006, Robinson and Ruiz-Herrera 2008). Together, these tools are now converging on a well-established phylogeny and timescale of mammalian species.

1.2.1 FROM HOMOLOGY MAPS TO THE ANCESTRAL KARYOTYPE

The search for the ancestral mammalian karyotype has a long tradition in cytogenetics. The first comparative chromosome maps drawn outlined the segments with conserved homology between human and other species. Chromosome homology maps of higher resolution were also prepared from chromosome-specific paints from other animals, such as the domestic dog (Graphodatsky *et al.* 2000), gibbon (Müller *et al.* 2003), the house mouse (Romanenko *et al.* 2006), several rodent species (reviewed in Romanenko *et al.* 2012), among others.

ANCESTRAL KARYOTYPE DELINEATION | The cladistic analysis of Zoo-FISH data was used in the construction of ancestral karyotypes. This method relies in the identification of primitive/ancestral chromosome traits (sympleisiomorphies) and shared derived chromosome traits (synapomorphies), assisted by parsimony analyses of the chromosome evolutionary rearrangements direction (Chowdhary *et al.* 1998, Wienberg *et al.* 2000). In order to define the conserved and derived syntenic associations, it is important the comparison with an outgroup, a distantly related taxon known to be phylogenetically outside the group of species under study (Wienberg 2004). If the syntenic association is present in the outgroup, then, according with the parsimony principle, it is considered as ancestral, while others are classed as common derived characters.

The first reconstruction of the ancestral eutherian karyotype was based in cladistic analysis of Zoo-FISH data of seven non-primate species, representing three orders, and was performed by Chowdhary and colleagues (1998). Several further analysis of the ancestral eutherian karyotype have been made since then, each providing additional insights into the organization of this ancestral karyotype (e.g. Wienberg 2004, Froenicke 2005, Ferguson-Smith and Trifonov 2007). Using a similar strategy it was possible to reconstruct ancestral karyotypes of different mammalian groups, such as primates, carnivores and rodents (reviewed in Graphodatsky *et al.* 2011). Regarding rodents, the delineation of the ancestral Muroidea karyotype (AMK) has been proposed by several authors (e.g. Stanyon *et al.* 2004, Engelbrecht *et al.* 2006, Romanenko *et al.* 2007). Recently, data from all comparative studies in Muroid rodents was compiled and besides the suggestion of the AMK, were also presented putative ancestral karyotypes for Cricetidae (ACdK) and Muridae (AMdK) (Figure 1.2), all based in the analysis of the syntenic associations defined using mouse (*Mus musculus*) paints (reviewed in Romanenko *et al.* 2012). Also Chaves *et al.* (2012) proposed an high precision Muroidea ancestral karyotype (Muridae/Cricetidae and Murine) based in a broad species analysis combining previous reported comparative maps together with newly presented data.

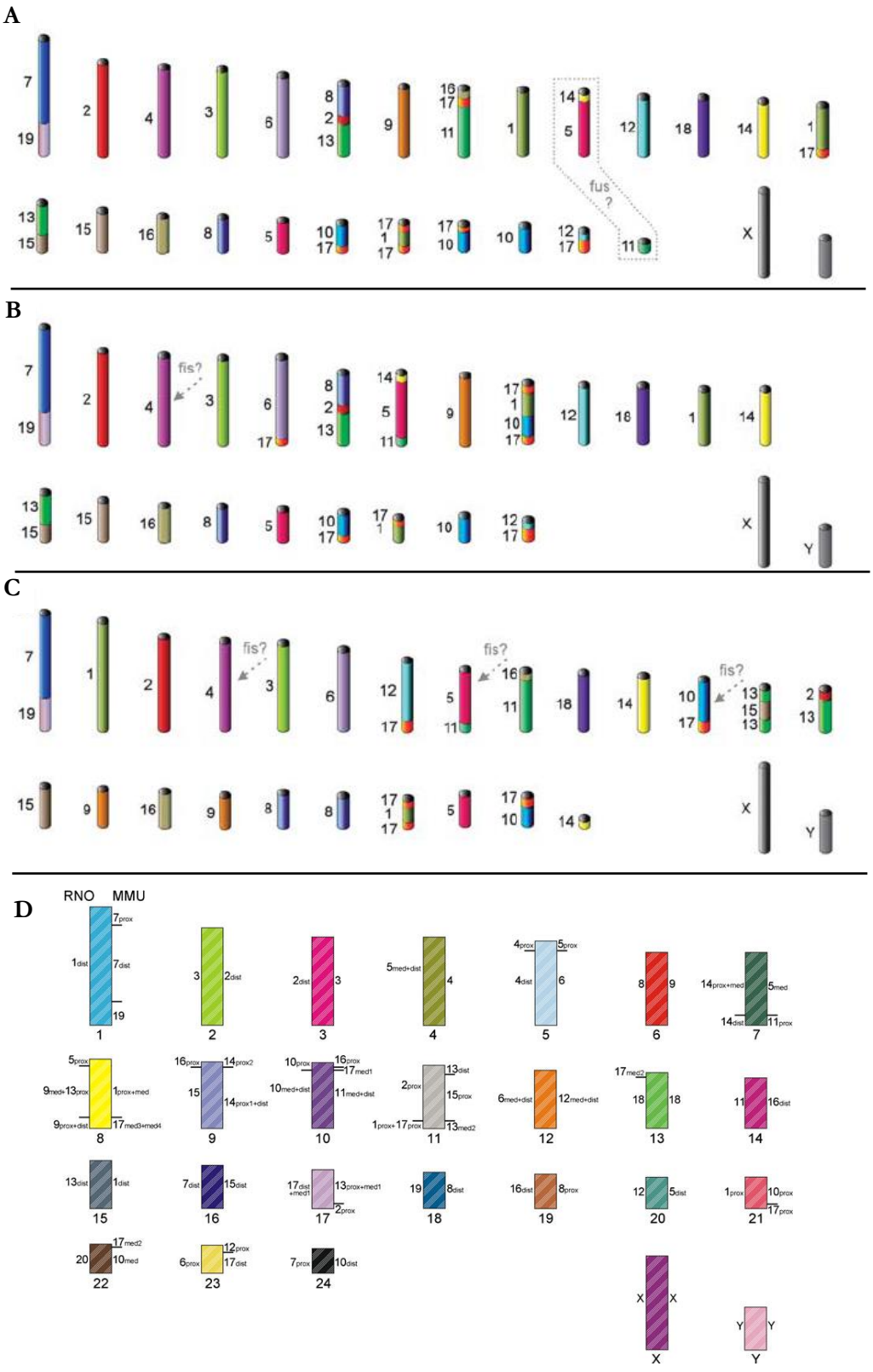


Figure 1.2|Putative ancestral karyotypes for Rodentia proposed by different authors. A) Ancestral Muroidea karyotype (AMK). B) Ancestral Cricetidae karyotype (ACdK). C) Ancestral Muridae Karyotype (AMdK). Different colors correspond to different mouse chromosomes. Some elements state is still ambiguously determined in the ancestral karyotypes, being represented by dashed grey frame and arrows (from Romanenko *et al.* 2012). D) Putative AMK with *Mus musculus* (MMU) and *Rattus norvegicus* (RNO) homologies (from Chaves *et al.* 2012).

CONSTRUCTING COMPARATIVE MAPS IN RODENTIA | Within Rodentia, comparative chromosome studies have been particularly productive in the analysis of non-muroid families, such as the Sciuridae (squirrels), whose karyotypes are highly conserved and retain many ancestral conditions (Stanyon *et al.* 2003, Li *et al.* 2006). On the contrary, the superfamily of muroid rodents, including the important laboratory animals, have highly rearranged karyotypes in comparison with humans, and for that reason cross-species chromosome painting with human probes was sometimes difficult to interpret (Ferguson-Smith *et al.* 1998). This problem was overcome by the use of painting probes from different rodents, such as *Mus musculus*, in cross-species experiments, being afterwards the human homologies inferred from human-mouse comparative maps based in the genome sequencing data (e.g. Romanenko *et al.* 2006). *Mus musculus* paint probes have been the most commonly used in rodent comparative studies (Cricetidae and Muridae families), but recent works describe the use of chromosome paints from rodent species belonging to other subfamilies, namely Cricetinae, Arvicolinae and Sigmodontinae (reviewed in Romanenko *et al.* 2012).

PHYLOGENETIC TREES | The number of syntenic segments per haploid set provides a measure of the relationship between species. When compared with the human genome, most eutherians have 30 to 40 separated segments of homology (O'Brien *et al.* 1999). Some species are exceptional, such as dogs and gibbons, and have about twice as many conserved segments (Wienberg *et al.* 1990, Yang *et al.* 1999), the rat shares about 100 segments with human (O'Brien *et al.* 1999), while the mouse is unique in having almost 200 blocks (Nilsson *et al.* 2001). By comparing different species chromosomes, a phylogenic tree can be constructed based on the minimum number of rearrangements occurred since the ancestral or the maximum number of shared syntenic segments. Molecular cytogenetic data and the increasing availability of partially or fully sequenced genomes from a variety of vertebrate species have fueled advances in phylogenomics, the phylogenetic reconstructions using genomic data (Robinson and Yang 2012). Because chromosomal rearrangements are such unique events, the probability that they occurred twice in different lineages (convergence) is low (Wienberg 2004). Therefore, the chromosome rearrangements identified by comparative studies revealed to be reliable evolutionary signatures or landmarks and have been used in the construction of phylogenetic trees elucidating phylogenetic relationships between species (e.g. Rokas and Holland 2000, Romanenko *et al.* 2007, Nie *et al.* 2012). During these last years, several comparative studies performed attempted to resolve some of the Muroids complex phylogenies, and phylogenetic trees were constructed for different muroid families, as it is exemplified in figure 1.3 (reviewed in Romanenko *et al.* 2012).

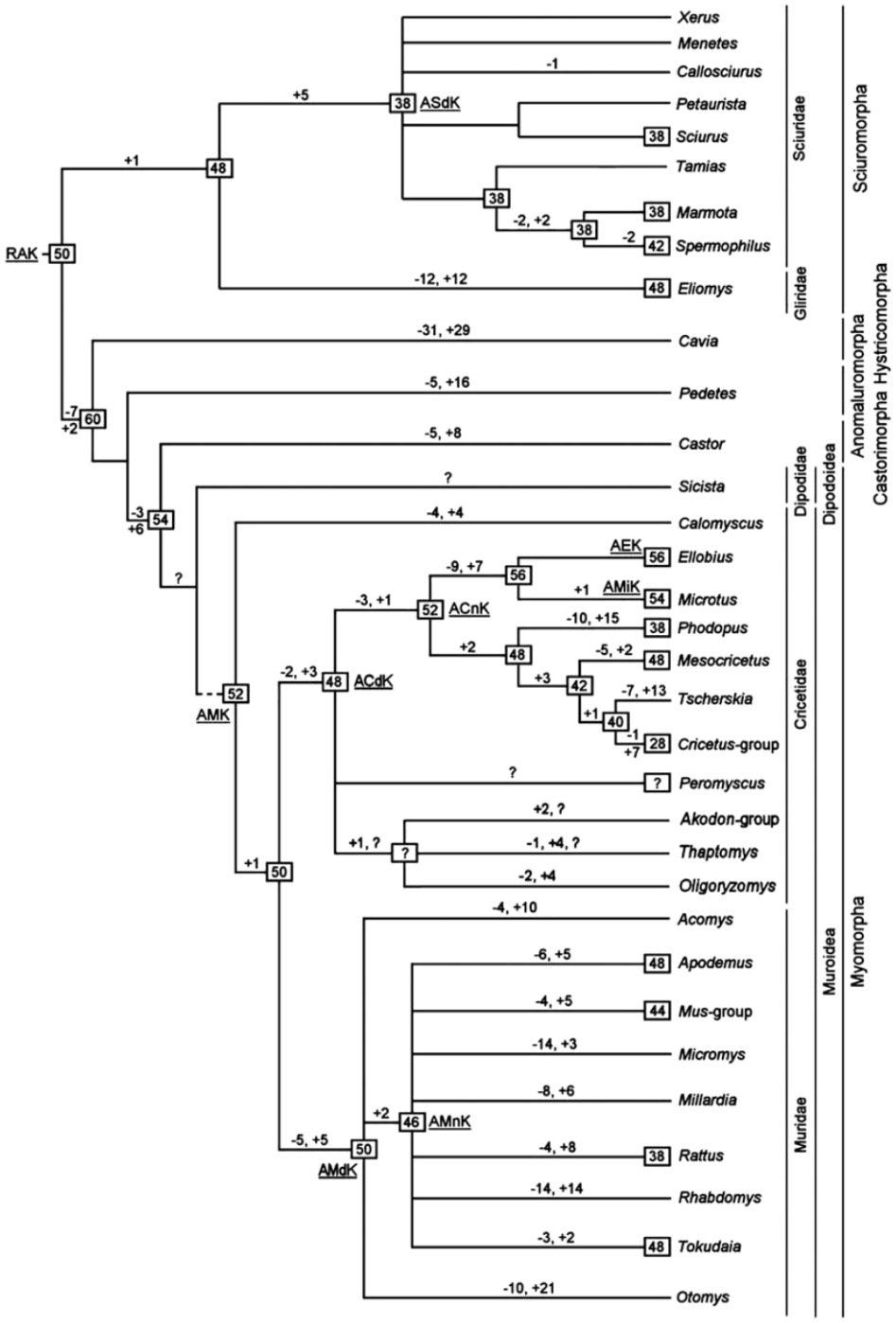


Figure 1.3| Putative rodent evolutionary tree. This tree is based in comparative data from several Muroid rodents, showing chromosome evolution to the genus level. RAK—ancestral Rodentia karyotype; ACdK—ancestral Cricetidae karyotype; ACnK—ancestral Cricetinae karyotype; AEK—ancestral *Ellobius* karyotype; AMdK—ancestral Muridae karyotype; AMK—ancestral Muroidea karyotype; AMiK—ancestral *Microtus* karyotype; AMnK—ancestral Murinae karyotype; ASdK— ancestral Sciuridae karyotype. Presumable ancestral diploid number characters for node are shown in black frames. Minus sign indicates chromosome fissions, plus sign indicates chromosome fusions, and question mark indicates unresolved positions (from Romanenko *et al.* 2012).

RATES OF CHROMOSOMAL REARRANGEMENT | The chromosome painting data now available for many species belonging to different eutherian orders, as well as data from alignment of genome sequencing and radiation hybrid maps, helped to estimate the average rate of evolutionary rearrangements (Murphy *et al.* 2005). The rates of chromosomal rearrangement vary radically not only among different lineages but also between sex chromosomes and autosomes (Eichler and Sankoff 2003). In most eutherian orders, there are species presenting a slow rate of chromosome evolution considered as the "default" frequency (Wienberg 2004). The default rate of eutherian chromosome evolution was calculated as approximately one rearrangement within 10 million years. Detailed investigation within groups has suggested that at different times, the rate of evolution, as well as the prevailing type of rearrangement, can change greatly (Murphy *et al.* 2005). For example, in the lineage that extends from the eutherian ancestor to the primate ancestor, during 50 million years, only three rearrangements took place (Froenicke *et al.* 2006). In contrast, a sudden karyotype diversification occurred in the gibbon lineage, with 24 rearrangements leading to the common gibbon ancestor and then multiple rearrangements subsequently leading to the karyotypes of the extant species (Müller *et al.* 2003, Wienberg 2005,). During the same period, karyotype evolution within the great apes group was extremely slow.

Record high rates of karyotype evolution are found in muroid rodents (Romanenko *et al.* 2006, 2007, Sitnikova *et al.* 2007), canids (Yang *et al.* 1999, Graphodatsky *et al.* 2000) and gibbons (Müller *et al.* 2003). The evolutionary rate between mouse and rat appears to be ten times greater than that found between humans and cat, or between humans and chimpanzees (Stanyon *et al.* 1999). Nevertheless, each of these mammalian orders contains groups with slower rates of chromosomal rearrangement, namely Sciuridae family among rodents (Stanyon *et al.* 2003, Li *et al.* 2004, 2006), felidae among carnivore (Perelman *et al.* 2005) and apes among primates (Froenicke 2005). The variation in the rates of mammalian karyotype evolution remains unexplained. Environmental effects, overall mutation rates, population size and the activation of mobile elements and retroviruses are among the possible contributory factors (Ferguson-Smith and Trifonov 2007).

1.3 DYNAMICS OF CHROMOSOME EVOLUTION

Along this thesis section it was mentioned that every genome rearrangement study involves identification of the syntenic chromosomal segments between species, and solving a combinatorial puzzle to find a plausible series of genome rearrangements to transform one genome into another. Such studies allow also the localization of **breakpoint regions** which are given by the two boundaries of the syntenic segments (Froenicke 2005), corresponding to

regions where genome synteny has been disrupted by chromosomal rearrangements. A breakpoint or breakpoint region is not a tangible physical entity in a genome; it is an analytical construct arising from the comparison of two genomes (Sankoff 2009). An early work by Ohno (1973) proposed the random breakage model of genomic evolution, postulating that the distribution of chromosome rearrangements breakpoints was uniformly random. The work by Nadeau and Taylor (1984), as well as comparative mapping and sequencing studies among vertebrate species, provided convincing arguments in favor of this model, and the random breakage model become the widely accepted theory of chromosome evolution. When data from the genome sequencing projects was comparatively analyzed by algorithms, it revealed remarkable levels of intrachromosomal rearrangements. The prevalence of short inversions restricted to specific chromosome segments represented a departure from the random breakage model in evidencing several breakpoints in one same region. Besides this, after comparing human and mouse genomes (Pevzner and Tesler 2003a), 281 syntenic blocks were found compared with the 180 known from comparative gene mapping (Nilsson *et al.* 2001). The explanation found was that breakpoint regions between the synteny blocks would have been disrupted an average of 1.9 times each, showing high density of breakpoints over these regions. This suggested an alternate model for chromosomal evolution, termed fragile breakage model (Pevzner and Tesler 2003b, which considers that there are regions (designated hotspots) throughout the mammalian genome prone to breakage and reorganization (Zhao *et al.* 2004, Peng *et al.* 2006). In support of this theory, chromosome breakpoint analyses have identified shared evolutionary breakpoint regions between different species (Murphy *et al.* 2005). Besides, a recent study reveals a high level of reuse of evolutionary breakpoint regions among muroid rodents, further supporting the fragile breakage model of chromosome evolution (Mlynarski *et al.* 2010). An interesting finding was that evolutionary breakpoint regions tend to colocalize with the more commonly occurring human cancer-associated breakpoints (Robinson *et al.* 2006). Furthermore, whole genome alignment studies have shown that evolutionary breakpoints regions are rich in repetitive elements (Murphy *et al.* 2005 Ruiz-Herrera *et al.* 2006), such as segmental duplicated regions, centromeric and telomeric regions. The presence of repetitive sequences at evolutionary breakpoint regions is thought to be related to the role that tandem repeats play as a substrate for non-homologous recombination, thereby promoting chromosomal rearrangements (Froenicke and Lyons 2008).

A segmental duplication involves the duplication of a small portion of chromosomal material (with 90% of similarity) either in tandem or transposed to new locations within the genome. Initial analyses of the human genome sequence have identified a large amount of

tandem as well as interspersed segmental duplications (Bailey *et al.* 2001). These observations raise the possibility that segmental duplications may have played a significant role in gene and genome evolution. Once formed, segmental duplications promote further rearrangement through misalignment and subsequent non-allelic homologous recombination (Stankiewicz and Lupski 2002). Studies in human corroborate this assumption showing that 25-53% of the recurrent breakpoint regions colocalize with human segmental duplications (Armengol *et al.* 2003, Bailey *et al.* 2004). Besides, in primates a strong association of segmental duplications with recurrent chromosomal structural rearrangements and also with disease was also demonstrated (Carbone *et al.* 2006, Marques-Bonet *et al.* 2009).

Centromeres and telomeres have long been recognized as peculiar dynamic regions of chromosomal evolution. The repetitive nature of these regions extends beyond the classically defined boundaries of centromeric and telomeric sequences; such transition regions, termed pericentromeric and subtelomeric DNA, are hotspots for the insertion or retention of repeat sequences. Among primates, there is now overwhelming evidence that blocks of recently duplicated sequence populate subtelomeric and pericentromeric regions (Eichler and Sankoff 2003). Chromosomal fissions probably require the complex regeneration of centromeres and telomeres, once chromatids have to be capped by telomeres and a new centromere have to be created. Gene map alignments indicate a high rate of *de novo* centromere formation (Murphy *et al.* 2005). Because the evolutionary reoccurring breakpoint regions form only a small proportion of a eutherian genome (3%), the colocalization of half of the neo-centromere hotspots with these regions allows us to speculate that an association exists, although only a few neo-centromeres have been observed (Robinson *et al.* 2006). This colocalization might indicate that neo-centromere generation is linked to the eutherian chromosomal plasticity, and that the potential for neo-centromere generation might also be evolutionarily conserved.

Overall, genomes can be considered a mosaic comprising regions of fragility that are prone to reorganization and regions that do not exhibit the same levels of evolutionary plasticity that have been conserved in different lineages during the evolutionary process.

2. GENOMIC COMPARTMENTS AND REPETITIVE SEQUENCES

Chromatin is found in two forms in eukaryotic genomes: euchromatin and heterochromatin. This classification was based on the observation that euchromatic chromosome regions changed their degree of condensation during the cell division cycle, whereas heterochromatic chromosome regions remained highly condensed throughout the majority of the cell cycle (Heitz 1928). In addition to differences in the timing of chromosome condensation, numerous other dissimilarities have been identified between these two genomic compartments. Euchromatin is enriched with unique coding sequences (genes), which are typically transcribed. Heterochromatin, on the other hand, is considered to be gene poor, being primarily composed of arrays of highly repetitive sequences (Hughes and Hawley 2009). Heterochromatin may be either facultative or constitutive. Facultative heterochromatin is found at developmentally regulated *loci*, where the chromatin state can change in response to cellular signals and gene activity, while **constitutive heterochromatin (CH)** occurs as large blocks in regions harboring repetitive sequences such as the pericentromeric regions, interstitial chromosome regions and telomeres (Dimitri *et al.* 2005, Adega *et al.* 2007, Paço *et al.* 2009). Constitutive heterochromatin is a basic component of eukaryotic genomes forming about 5% of the genome in *Arabidopsis thaliana*, 30% in *Drosophila* and humans, 60% in rodents and up to 70–90% in certain nematodes and plants (Sherwood and Patton 1982, *Arabidopsis* genome initiative 2000, Dimitri *et al.* 2005). Constitutive heterochromatin can be revealed by preferential "loss" of DNA from non-heterochromatic regions, achieved by conventional C-banding technique (Pathak and Arrighi 1973). Restriction endonuclease digestion followed by C-banding has shown its ability in demonstrate CH heterogeneity by revealing additional heterochromatic bands, cryptic C-bands (Chaves *et al.* 2004, Adega *et al.* 2005).

Studies primarily conducted in *Drosophila melanogaster* have shown that CH plays different roles in important cellular functions, such as chromosome organization, besides containing essential genes for viability and fertility (Dimitri *et al.* 2009). Pairing of heterochromatic regions is required for the proper segregation of chromosomes that fail to undergo recombination during female meiosis (reviewed in Grewal and Jia 2007).

But how does heterochromatin perform its diverse functions? The formation of heterochromatin requires methylation of histone H3 at lysine 9 and the subsequent recruitment of chromodomain proteins such as heterochromatin protein HP1. Evidence from studies in diverse model systems indicates that heterochromatin serves as a self-assembling framework of histone modifications to recruit effector proteins, which in turn regulate various chromosomal processes (Shimada and Murakami 2010). Given the critical functions of heterochromatic

sequences in both meiosis and mitosis and its rapid change in sequence, it has been hypothesized that differences in either heterochromatic sequences or the proteins that maintain them might indeed play a role in species isolation and thus speciation (Hughes and Hawley 2009).

DIFFERENT TYPES OF REPETITIVE SEQUENCES IN THE GENOME | A significant portion of the eukaryotic genome is comprised by repetitive sequences that can be located in both euchromatin and heterochromatin. Repetitive sequences can be categorized into two main classes considering their primary organization in the genome: interspersed DNA and tandemly repeated DNA (reviewed by Slamovits and Rossi 2002). The first class refers to sequences scattered in the genome, generally known as transposable elements due to their ability of "jumping" to different genomic locations (transposition). These elements are divided in different classes according to their mechanism of transposition (reviewed by Wicker *et al.* 2007): retrotransposons or class I, transpose via an RNA intermediate and as examples are the well-known Alu sequences in humans and the L1 elements in mammals (Capy *et al.* 1997, Furano 2000); the class II are the DNA transposons, and these elements transpose by excision from their location and integration in other genomic sites without an RNA intermediate (Finnegan 1989, Capy *et al.* 1997). The second class of repetitive sequences, the tandemly repeated DNA, includes three distinct groups: microsatellites, minisatellites and satellite DNA. Micro and minisatellites are characterized by short repeat units, ranging from up to 100 bp for microsatellites and 1-5 bp for minisatellites (Charlesworth *et al.* 1994). Array size for both micro and minisatellites varies from 10 to 100 repeat units, and they can be found distributed throughout the genome (Li 1997). Microsatellites appear to be primarily located in euchromatic regions of chromosomes or in the vicinity of genes, such as the human CGG trinucleotide repeats (Riggins *et al.* 1992), but microsatellite arrays can be also often detected in heterochromatin (Gindullis *et al.* 2001). Minisatellites can be found irregularly dispersed in euchromatin and largely clustered in subtelomeric chromosomal regions (Royle *et al.* 1988). **Satellite DNA (satDNA)**, that is the repetitive sequence *in focus* in this thesis, is along with transposable elements, the mainly constituent of constitutive heterochromatin (John 1988, Chaves *et al.* 2004). SatDNA is characterized by long tandem arrays and it is usually present in the genomes in several million copies (Charlesworth *et al.* 1994). Early experiments historically coined the term "satellite DNA", referring to tandemly arranged sequences forming satellite bands which differentiated from the rest of the genomic DNA by density gradient separation (John 1988). Once no protein coding function could be primarily associated with satellite DNAs, it has been early considered as useless genomic elements accumulated as junk (Ohno 1972), or as

sequences representing genomic parasites proliferating independently (Orgel and Crick 1980). Nevertheless, evidences emerged that demonstrated the functional significance of satellite DNA sequences, ranging from chromosome organization and pairing, to cell metabolism and speciation. Studies support these functionalist assumptions concerning the association of satellite DNAs with complex features of eukaryotic chromosomes (e.g. Csink and Henikoff 1998, Henikoff *et al.* 2001, Sullivan *et al.* 2001), which will be highlighted in the next paragraphs.

2.1 SATELLITE DNA FEATURES AND FUNCTION

Satellite DNAs are highly repeated DNA sequences, typically organized as long arrays of head-to-tail linked repeats (Charlesworth *et al.* 1994). The amount of satellite DNA content can sometimes exceed 50% of a species total DNA (Mravinac and Plohl 2010), being this genome fraction responsible for the variation of genome size in some eukaryotes (Gregory *et al.* 2007). The length of the repeating unit (**monomer**) can range from only few base pairs up to more than 1 kb, thus forming arrays that may reach 100 Mb long (reviewed Plohl *et al.* 2008). These lengthy arrays of satDNA form conspicuous blocks of differentially condensed chromatin in the chromosomes, mostly in centromeres but also in telomeres and interstitial positions (Chaves *et al.* 2000, Meles *et al.* 2008). A satDNA family is defined by a specific sequence and length. Different satDNA families can vary greatly in base composition, with some being rich in AT and others in GC. In the mouse genome, for instance, four distinct satDNAs have been characterized, the major satellite (MaSat) and the minor satellite (MiSat) which are AT-rich, and mouse Satellite 3 (MS3) and mouse Satellite 4 (MS4) which are CG-rich (Kuznetsova *et al.* 2005). Despite satDNA variation in nucleotide sequences across species, they share some common features such as monomer length, which is generally between 150–180 bp and 300–360 bp, in both plants and animals (reviewed in Plohl *et al.* 2008). As example, the human alpha-satellite presents a monomer size of 171bp (Manuelidis 1978), the maize CentC and CentO satellites have 156 bp (Birchler *et al.* 2011), while the pig Mc1 satellite presents a monomer of 340 bp (Adega *et al.* 2008). This can be explained by the organization of the satDNA around the nucleossomes. The mentioned sizes constitute the required DNA length to be wrapped around one or two nucleosomes (Henikoff *et al.* 2001). Another feature shared by different satellite DNA families is the presence of a short motif with 17 bp, known as CENP-B box, found in several centromeric satellites, such as human alphoid satellites (Masumoto *et al.* 1989), the mouse minor satellite (Wong and Rattner 1988) and satellites from other species. CENP-B box represents the binding site for centromere protein B (CENP-B) (Kipling and Warburton 1997), being one of the most well characterized satDNA sequence binding protein (Earnshaw *et al.* 1989, Sugimoto *et al.* 1998). Given the conservation of the CENP-B box between diverse mammalian species, a functional

constraint in centromere activity was attributed to CENP-B, being considered essential in human assembly of centromeric-specific chromatin (reviewed by Ugarković 2005). It has been demonstrated that the CENP-B box is required for *de novo* centromere chromatin assembly on human alphoid DNA (Ohzeki *et al.* 2002, Okada *et al.* 2007). A recent study demonstrates that the accumulation of CENP-B-containing satDNA in a neocentromeric region, leads to the increase binding of another centromeric protein, CENP-A (centromeric protein A), eventually leading to a mature centromere that binds more CENP-A (Marshall and Choo 2011). This new proposed model explains, among others, the evolutionary conservation of CENP-B.

Given their primary localization in transcriptionally suppressive heterochromatin, transcriptional activity was not expected for satellite DNAs. Although transcripts of satellite DNAs have been reported in several organisms including vertebrates, invertebrates and plants (e.g. Lee *et al.* 2006, Pathak *et al.* 2006, Wong *et al.* 2007). It has been shown that satellite DNAs are temporally transcribed at particular developmental stages or are differentially expressed in some cell types, tissues or organs in most of the species analyzed (reviewed by Ugarkovic 2005). For instance, mouse gamma satellite DNA is differentially expressed during development of the central nervous system, as well as in the adult liver and testis (Rudert *et al.* 1995). Besides, some satDNA transcripts have shown to be important for epigenetic chromatin modifications, being involved in the initiation of histone H3 methylation, a necessary prerequisite for heterochromatin formation and maintenance (Martienssen 2003). Transcripts of satellite DNAs in the form of small interfering RNAs (siRNA) participate in the epigenetic process of chromatin remodeling and heterochromatin formation (Grewal and Elgin 2007). Some satellite DNA transcripts, particularly from some insects, nematodes and amphibians (Epstein and Gall 1987, Ferbeyre *et al.* 1998, Rojas *et al.* 2000), function as ribozymes with self-cleavage activity, whereas human satellite III transcripts are involved in the recruitment of splicing factors during stress (Chiodi *et al.* 2004). The presented examples suggest an active role for satellite transcripts in several regulatory layers from chromatin modulation to transcription and RNA maturation translation.

2.2 EVOLUTIONARY DYNAMICS OF SATELLITE DNA

Satellite DNA sequences probably arise as the result of large-scale duplication of sequences that are integrated into the genome at a favorable site (Britten and Khone 1968). Once established in the genome how do these sequences spread and amplify? Previously presented features show that satellite DNA seems to be a very distinctive component of the eukaryotic genomes constituting highly dynamic sequences. The variability of satDNAs even among closely

related species, differing in nucleotide sequence and/or copy number of satellite families, results from the rapid turnover of these sequences (reviewed in Plohl *et al.* 2008). Two parameters must be independently considered when referring to the evolution of satellite DNAs, copy number and nucleotide sequence (Urgacovic and Plohl 2002), being copy number alterations the most rapid changes.

Satellite DNAs can vary dramatically in their number of copies among related species. Expansions and contractions of satellite arrays can efficiently change the landscape of repetitive sequences, resulting in the replacement of one dominant satellite repeat (major satellite) with another one less represented (minor satellite) (reviewed in Ugarkovic and Plohl 2002). The mechanisms proposed to be responsible for amplification/deletion of repeated DNA are unequal crossing over, replication slippage and rolling circle amplification (Walsh 1987). Intragenomic identity among units of satellite DNA sequences is on the base of all these mechanisms action. Briefly, unequal crossing over occurs when near, but non-homologous, sites recombine during meiosis, and it have been shown to develop periodicities (Smith 1976); replication slippage takes place when short contiguous repeats cause a mispairing between neighboring repeats during DNA replication (Levinson and Gutman 1987); rolling circle amplification involves intrastrand recombination between repeat units resulting in extrachromosomal DNA circles consisting in several repeat units, that can be lost or integrated in the genome, and in this case is called saltatory amplification (Walsh 1987, Rossi *et al.* 1990).

Rapid evolution of satellite DNA sequences is also possible due to the accumulation of nucleotide changes, usually with a high rate and in a gradual manner (Plohl *et al.* 2008). Mutations in a satellite DNA family accumulate gradually, and depending on the rate of accumulation and spread they can be phylogenetically informative, for example, on the species level, on the level of ecotype-specific variants or on the level of phylogeographic clades (Plohl 2010). In fact, some satDNAs were used to establish phylogenetic relationships in rodents, cetartiodactyla and oyster (Arnason *et al.* 1986, Kunze *et al.* 1999, López-Flores *et al.* 2010, Ostromyshenskiĭ *et al.* 2011), among others. Satellite DNAs divergence among species is quite variable, as some repeats are species-specific, while others are widely conserved, being shared across distantly related species (Mravinac *et al.* 2002, Adega *et al.* 2008). Some satDNA sequences maintain their sequence identity during long evolutionary periods, and copy number changes may not be accompanied by turnover of nucleotide sequences. Recently it has been described a satellite DNA family shared by species belonging to the three main clades of the Class Bivalvia thus representing the oldest described satDNA (540 million years) (Plohl *et al.* 2010). In addition, large-scale changes such as

segmental duplications, as referred previously, play an important role in the rapid evolution of DNA sequences in and around centromeric regions (e.g. Ventura *et al.* 2007).

2.2.1 CONCERTED EVOLUTION

When monomers of a repetitive family are compared between species, higher sequence similarity is found within species than between species. The high repeat homogeneity is suggested to result from non-independent evolution of monomers. This phenomenon is known as **concerted evolution**, and it is promoted by molecular drive, a two-level process in which mutations are homogenized throughout members of a repetitive family, and concomitantly fixed within a group of reproductively linked organisms (Dover 1986, Elder and Turner 1995). Sequence homogenization can be achieved by gene conversion. This mechanism consists in the non-reciprocal transfer of genetic information between similar sequences, in which exchange of flanking DNA is not involved (Schimenti 1999). Other mechanisms responsible for homogenization are transposon-mediated exchange, as well as unequal crossing over and rolling circle replication that, unlike gene conversion, also promote changes in the number of repeats (as mentioned before) (Slamovits and Rossi 2002). The action of these mechanisms between similar sequences on non-homologous chromosomes is related with a specific chromosome configuration during early prophase (Figure 2.1). In this stage all chromosomes migrate into one area of the nucleus and adopt a particular orientation known as the bouquet in which all telomeres attach to the nuclear membrane (Scherthan *et al.* 1996). The bouquet provides the physical opportunity for concerted evolution to occur between similar sequences on non-homologous chromosomes (e.g. Brannan *et al.* 2001). Nevertheless, this "structure" influences the evolution of satDNAs between acrocentric *versus* metacentric chromosomes. This has been demonstrated in humans and more recently in the pig, in which the similarity of centromeric sequences is higher within acrocentric and within metacentric chromosomes than between them (Hirai *et al.* 1999, Adega *et al.* 2008).

The spreading of newly occurring mutations horizontally throughout the members of repetitive family reveals lower efficiency (predominantly with unequal crossover) in bordering regions of the satellite array (Bassi *et al.* 2000, Schueler *et al.* 2005). This fact leads to the prediction that monomers at array ends are more divergent than those located centrally (Stephan 1989). Moreover, considering local and global sequence homogenization, these mechanisms efficiency has been demonstrated to be higher within localized subsets of satellite repeats than between arrays on the same chromosome, homologous and heterologous chromosomes (Dover 1986). This results in adjacent monomers showing a higher degree of sequence similarity than

those retrieved at random, being them often grouped into subsets or subfamilies, defined by specific mutations (Hall *et al.* 2005, Roizès 2006).

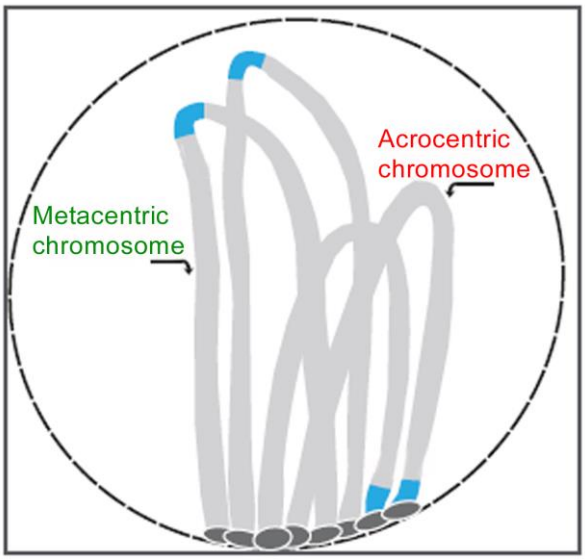


Figure 2.1|The bouquet organization of chromosomes during prophase. All telomeres (dark grey) are attached to the nuclear envelope. The physical proximity between centromeres (blue) may facilitate their sequence homogenization by non-homologous recombination. It's worth noticing that centromeres from acrocentric and metacentric chromosomes are far from each other diminishing the interaction between them (adapted from Cazaux *et al.* 2011).

In some satellites it has been observed a tendency to increase the repeat unit length and complexity. This can be achieved by merging shorter repeat motifs into a higher-order repeat (HOR). HORs result from the homogenization between adjacent monomer variants, in which former monomers become subrepeats or subunits (reviewed in Plohl *et al.* 2008). Since a HOR is a homogenization unit, HORs generally show high level of sequence identity, while substantial sequence divergence is accumulated among constituent subunits. This was demonstrated in human alpha-satellite HORs, which are typically 97–100% identical while subunits within them are only, on average, 70% identical (Roizès 2006).

The concerted evolution model has received considerable support from other repetitive DNA elements, such as rRNA genes (Ganley and Kobayashi 2007, Cazaux *et al.* 2011).

2.2.2 MODELS OF SATELLITE DNA EVOLUTION AND CENTROMERIC FUNCTION

Despite existing in other chromosome locations, generally satDNAs are more representative in the centromeric and pericentromeric regions and its rapid evolution seems to contrast with the high conservation in function of centromeres. The centromere is the chromosomal domain responsible for primordial functions such as kinetochore formation and

sister chromatid cohesion, playing a key role in faithful transmission of chromosomes during mitosis and meiosis (Gonçalves dos Santos Silva *et al.* 2008). While centromeric function is conserved through eukaryotes, the profile of satellite DNAs in this region can be species-specific (Henikoff *et al.* 2001), thus representing a paradox. According to normal expectations of evolutionary biology, a region with such a critical and highly conserved function should have a stable sequence. Given the diversity of rapidly evolving satellite repeats in the centromere, some authors proposed that centromere function and inheritance may be determined by epigenetic determinants (reviewed in Ekwall 2007). Nevertheless, the presence of satellite DNA flanking functional centromeres can be advantageous given the unique dual ability of satDNAs to maintain sequence homogeneity over long stretches of DNA, and simultaneously to change rapidly in evolution. As explained by Plohl *et al.* (2008), these characteristics along with the abundance of satellite repeats may stabilize interactions with DNA-binding proteins and, on the other hand, they can rapidly mutate into a new sequence variant which can better fit protein interactions, once these sequences can present diverse degrees of conservation as highlighted before in this chapter.

Several studies of satDNA from different species revealed diverse patterns of evolution of these sequences in centromeric and pericentromeric regions, which will be presented.

LIBRARY MODEL | It has been suggested that closely related species share a library of satellite DNAs, existing in a hypothetical common ancestor, and the contraction or amplification of the number of copies of those satellites can lead to lineage specific profiles (Figure 2.2) (Fry and Salser 1977). Generally, the **library model** explains the occurrence of species-specific satellite profiles as a consequence of fluctuation in copy number without variance of the sequence (Meštrović *et al.* 1998). A striking example is the insect satellite family PRAT, found in species separated by 50-60 million years and showing low sequence variability ("frozen"), but considerable copy number variation in the species genomes (Mravinac *et al.* 2002). Another well-known example of copy number fluctuations are the recurrent amplifications and deletions of the major satellite RPCS (repetitive PvuII *Ctenomys* sequence) which are characteristic for the rodent species of the genus *Ctenomys* (Slamovits *et al.* 2001). Moreover, recently the library model continues receiving considerable support in different animal models (Bruvo *et al.* 2003, Ellingsen *et al.* 2007, Caraballo *et al.* 2010, Plohl *et al.* 2010).

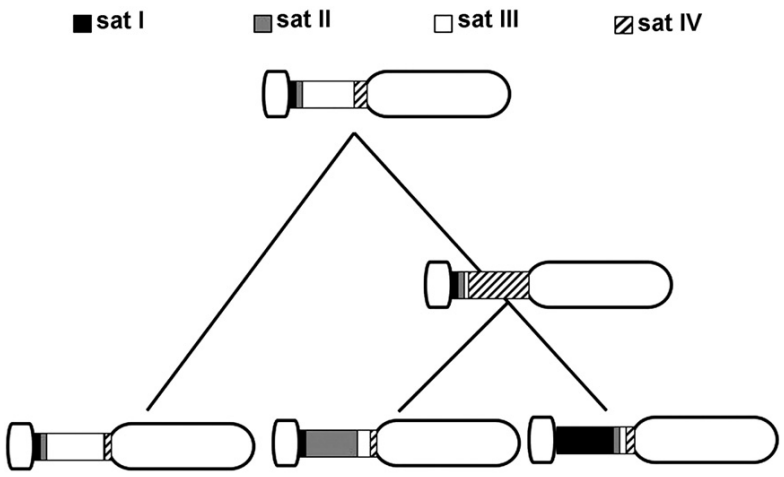


Figure 2.2 | The Library model. Several satDNA families can coexist on chromosomes with different representation (major (major and minor satellites). These families can be differentially amplified resulting in a distinct satellite landscape in the chromosomes, leading to species-specific profiles (from Plohl *et al.* 2008).

FEEDBACK MODEL | Nijman and Lenstra (2001) have proposed a model based on studies in Bovidae explaining satDNA life history. According to the **feedback model** three different phases occur during satDNA life (Figure 2.3). Briefly, in phase I, interactions of homogeneous repeats cause rapid expansions, as well as contractions with saltatory fluctuations in the copy number. These events are favored by the homogeneity between sequences (Positive Feedback). Mutations can lead to new sequence variants, decreasing the recombination events and the homogenization (Negative Feedback). In phase II, mutations and recombination events lead to new variants, evolving independently; finally, phase III, the final phase, is reached when interactions between old monomers and a new satellite DNA family stop. Plus, the selective pressure on maintenance of the genome size eventually leads to the elimination of the older satDNA families (Nijman and Lenstra 2001). This model has been extrapolated to satellite families of several mammalian species (Slamovits *et al.* 2001, Mravinac *et al.* 2005).

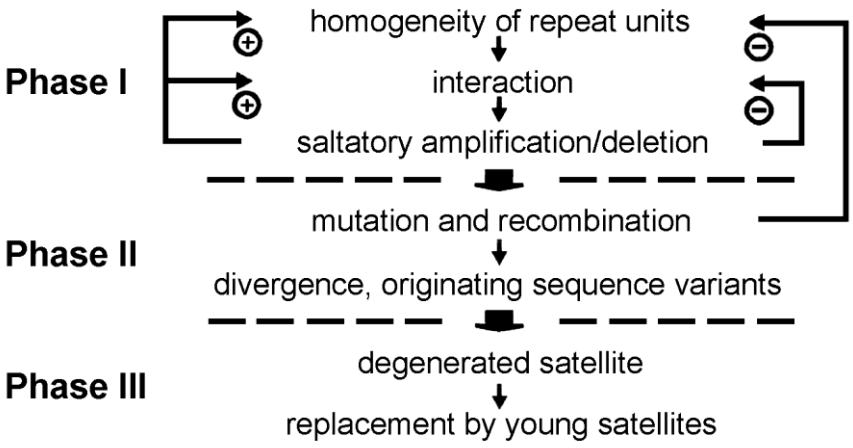


Figure 2.3 | Feedback model. This model predicts three distinct phases in satellite DNA life history (Phase I, II and III). + Positive feedback loop; - Negative feedback loop (adapted from Nijman and Lenstra 2001).

PROXIMAL PROGRESSIVE EXPANSION| Studies on the organization of the centromere in the human X chromosome, revealed another mode for satDNA evolution. A work by Schueler *et al.* (2005) showed that the human alpha-satellite evolves according to **Proximal Progressive Expansion** (Figure 2.4). New satDNA sequences originated by mutations are consecutively added to the centromere, being the older sequences located more distantly (Schueler and Sullivan 2006). This leads to a satDNA age gradient to occur from the centromere into the pericentromeric regions observed in human centromeres (Schueler *et al.* 2005), which was recently pointed as an invaluable tool for phylogenetic studies (Shepelev *et al.* 2009). The described process is facilitated by the low efficacy of the homogenization mechanisms in the terminal monomers of the arrays, as referred previously (Stephan 1986). This occurrence corroborates the predictions of theoretical models which considered that distant satDNA monomers are more divergent than those located in the central region (Smith 1976, Stephan 1989).

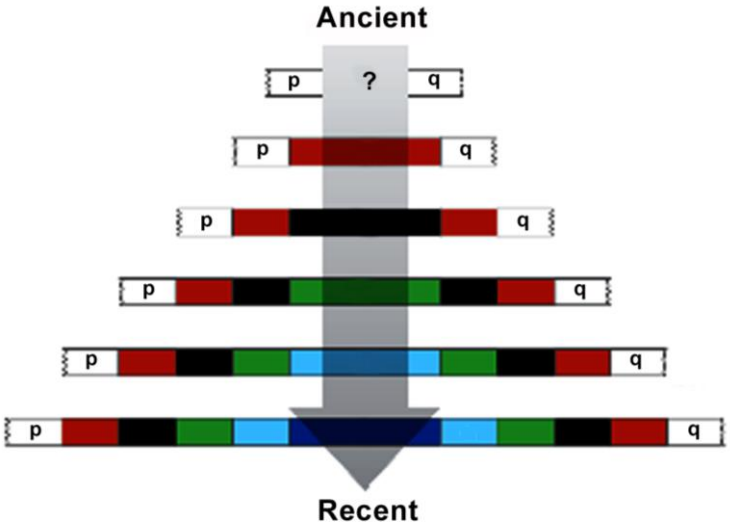


Figure 2.4| Proximal Progressive Expansion mode of evolution of satellite DNA. SatDNA successive additions (colour rectangles) to the centromere along evolution are shown. Each addition of new material moves previous centromeric DNA outward (adapted from Schueler and Sullivan 2006).

CENTROMERE DRIVE MODEL| The existence of fast evolving satDNAs in centromeres is described as a paradox, as explained before in this thesis section. However, it has been demonstrated that not only the profiles of satellite DNAs in this region suffer changes, but also the centromeric proteins were found to diverge (Malik and Henikoff 2001). These observations support the idea that both centromeric satellite sequences and proteins can evolve in concert, being this referred as **centromere drive model** (Henikoff *et al.* 2001). The same idea has been highlighted more recently by Dawe and Henikoff (2006), stating that both DNA and protein evolution drive each other in a centromere (Figure 2.5A), thus providing a stable, but flexible,

system, able to work on genetic and epigenetic platforms and, if necessary, to rescue chromosomal function by forming new centromeres on non-specialized locations. This coevolution may be driven either by changes in satellite DNAs (Malik and Henikoff 2001, Talbert *et al.* 2004), or by satellite repeats competition to better fit the chromatin environment (Dawe and Henikoff 2006). The rapid replacement of DNA sequences without alteration in the binding affinities would constitute an ideal system (Figure 2.5B) (reviewed in Plohl *et al.* 2008).

The divergences in satellite sequences and corresponding proteins accumulated between individuals can cause incompatibilities in hybrids leading eventually to reproductive isolation acting thus as a trigger in the speciation process (e.g. Meetrovic *et al.* 1998, Hall *et al.* 2005). Recent studies show evidences on how satellites can impact chromosomes at a number of different developmental stages and through distinct cellular mechanisms, and can cause postzygotic reproductive isolation (reviewed in Ferree and Prasad 2012).

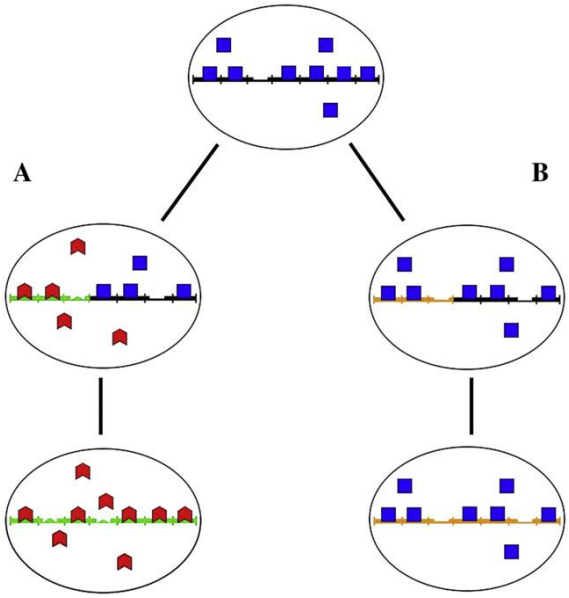


Figure 2.5 | Coevolution of satellite DNA sequences and DNA-binding proteins in the centromeric region. A) A changed satellite monomer variant (green) with modified binding site is able to bind a changed centromeric protein (red) replacing eventually the old protein/satellite pair. B) Rapid changes in satellite DNA profiles without affecting DNA-binding proteins (from Plohl *et al.* 2008).

2.2.3 SATDNA DYNAMICS AND CHROMOSOMAL REARRANGEMENTS

Along this section, several aspects of satDNA have been shown indicating that these sequences constitute a very dynamic component of the mammalian genome, representing an important factor for genomic plasticity (Slamovits and Rossi 2002). Some studies focused on chromosome evolution indicate that satellite DNA can be correlated with the dynamics of the

chromosome restructuring (Garagna *et al.* 2001; Louzada *et al.* 2008). In this sense, satDNA has been claimed to play a role in genomes with particular impact on chromosomal changes related with speciation (Bradley and Wichman 1994, Hartmann and Scherthan 2004). Although considerable evidence from several works support this association, the precise mechanisms remain elusive in many cases (Coghlan *et al.* 2005).

Wichman *et al.* (1991) postulated that rapidly evolving satDNA families promote chromosomal rearrangements by means of their intragenomic movements, occurring among non-homologous chromosomes and between different chromosomal fields, as centromeres, arms and telomeres.

A striking evidence for the correlation between repetitive sequences and chromosome restructuring, come from the analysis of the distribution of the syntenic blocks and evolutionary breakpoint regions between diverse species (highlighted in the previous section of this thesis). Ruiz-Herrera and colleagues (2006) found a high correspondence between human chromosome fragile sites location, the positions of evolutionary breakpoint regions and the distribution of tandem repeats in the human genome. Moreover, the constitutive heterochromatin regions (where satDNA is located) have been previously pointed as hotspots for structural chromosome rearrangements (John 1988, Chaves *et al.* 2004).

Some satellite DNA sequences have been associated with the occurrence of specific chromosome rearrangements along species evolution in particular taxa. An example regards the Robertsonian translocations which constitute a common and frequent form of chromosomal rearrangement occurred in a variety of species along their evolution. Garagna *et al.* (2001) suggested the occurrence of sequence-dependent mechanisms of interchromosomal exchange between satellite blocks (Smith 1976), promoting the Robertsonian and whole-arm reciprocal translocations, based in the orientation and disposition of the satellite repeats in the pericentromeric regions. In fact, the physical and organizational analyses of different satDNA families in a model translocation, like the well known t(1;29) in cattle (e.g. Chaves *et al.* 2003), for example, can greatly increase our understanding on the translocation events during mammalian genome evolution. Also Slamovits *et al.* (2001) verified that chromosome restructuring in the rodent *Ctenomys* was accompanied by the active processes of expansion, contraction and mobilization of the satellite DNA RPCS. In fact, it was shown that karyotypically variable clades underwent substantial contractions and expansions of RPCS copy numbers while chromosomally stable clades exhibit stasis of the RPCS copy number. Moreover, it has been suggested that chromosome fissions would have been accompanied by significant satDNA

expansion events, while the loss of RPCS was hypothetically related with chromosomal fusion events (Slamovits *et al.* 2001, Caraballo *et al.* 2010).

A common feature to many species centromeric satDNAs is the CENP-B box. The presence of this sequence motif possibly plays an important role in recombination events leading to translocations involving the centromeric region (e.g. Volobouev *et al.* 1995). This may occur due to the dimerization ability of CENP-B protein which can lead to misalignment between HOR units of satDNA in non-homologous chromosomes, and also facilitate recombination mediated by nicking activity (Garagna *et al.* 2001).

Adega *et al.* (2009) suggested two different modes of genome restructuring pinpointing the evidences found in diverse mammalian groups. These authors state that satellite DNA dynamic behavior constitutes one of the hallmarks of genome evolution, being the other mode of genome restructuring the chromosomal conservative lineages with satDNA patterns localized only in specific chromosomal fields (Figure 2.6).

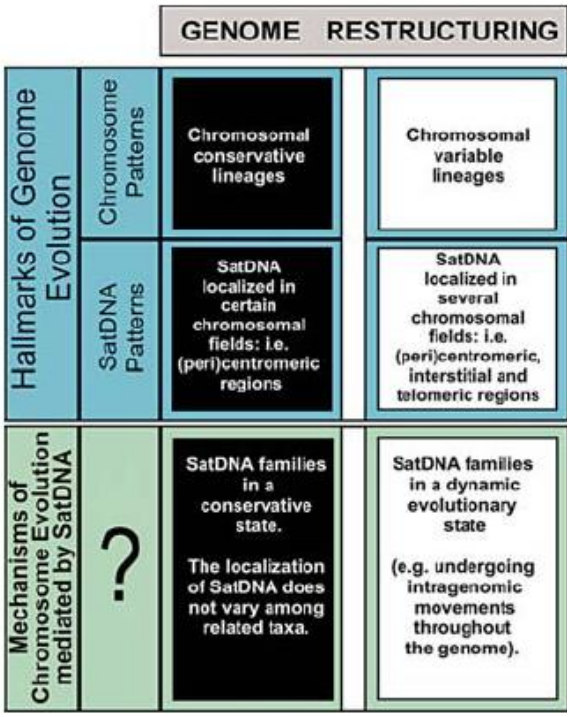


Figure 2.6| The two main outcomes in genome restructuring (from Adega *et al.* 2009).

Clinical cytogenetics and, in particular, cancer cytogenetics can also provide important information about the mechanisms involved in chromosome change induced by satDNA. In tumors, it has been shown that the high density of repetitive DNA in a given region provides "hotspots" for homologous recombination and mediates translocation processes (Kolomietz *et al.* 2002). The analysis of a cat fibrosarcoma by Santos *et al.* (2006) revealed that the amplification

of satellite DNA sequences in the cat chromosomes was related to mitotic instability, which could explain the exhibition of the complex patterns of chromosome aberrations detected.

Evolutionary and clinical evidences show that each genome structure exhibits sensitivity to rearrangements involving certain chromosome regions. However, a common trait in all of these cases is the involvement of repetitive sequences in these rearrangements, either centromeric or interspersed in the genome.

3. CANCER IS AN EVOLUTIONARY PROCESS

Cancer chromosomes present a high level of dynamics and the ability to constantly evolve. This unique characteristic forms the basis of genetic heterogeneity necessary for cancer formation. Therefore, the development of cancer can be considered analogous to the evolution of species being both Darwinian processes based on variation and selection (Greaves and Maley 2012). Cancer progression is based on two fundamental processes, the continuous acquisition of heritable genetic variation in individual somatic cells and natural selection acting on the resultant phenotypic diversity (Stratton *et al.* 2009). This results in the unrestrained proliferation of cells known as **tumor** or neoplasm that may subsequently invade beyond normal tissue boundaries and metastasize to distant organs (**cancer**) (Alberts *et al.* 2008). The earliest ideas about tumor evolution come from Boveri, that over a century ago proposed that tumors originated in chromosomal abnormalities passed on to daughter cells (Boveri 1902, 1914). Later, it has been postulated that most tumors originate from a single cell, and the acquisition of genetic variability within the original clone determinates the tumor progression throughout the selection of more aggressive subclones (Nowell 1976). Accordingly, multistep tumor progression can be portrayed as a succession of clonal expansions, relying on variability between tumor cell subpopulations partly initiated by conditions within the cancer-micro-environment as well as by interactions between tumor cell subpopulations and host cells (reviewed in Tysnes 2010). This is confirmed by the variability in disease presentation and path, found in similarly diagnosed cancers in different patients and in the same cancer at different time periods.

An alternative model for tumor evolution was derived from studies in hematological tumors which revealed the presence of intra-clonal genetic heterogeneity indicative of divergent clonal evolution. These data suggested that clonal architecture might be driven by genetic heterogeneity of propagating or stem cells (reviewed in Greaves 2010). The cancer stem cell concept postulates that similar to the growth of normal tissues, the growth of tumors is fuelled by the presence of stem cells that are capable of self-renewal (reviewed in Clevers 2011). Its worth mentioning that the clonal evolution and cancer stem cell models are not mutually exclusive in cancers that follow a stem cell model, as cancer stem cells would be expected to evolve by clonal evolution (Shackleton *et al.* 2009).

3.1 GENES THAT DRIVE CANCER

Cancer is, in essence, a disease of genes (Vogelstein and Kinzler 2004). In the last years, many important genes responsible for cancer development have been discovered, their mutations precisely identified, and the pathways through which they act characterized. A census of cancer genes was recently updated to 384 genes (almost 2% of genes in the human genome) that are thought to be causally implicated in cancer development when specifically mutated (Santarius *et al.* 2010).

Three important classes of genes can be referred as playing key roles in tumor initiation: oncogenes, tumor suppressor genes and stability genes. Alterations in these genes (e.g. mutations, amplifications or deletions) may lead to a de-coupling of biological mechanisms involved in the regulation of normal cell growth and differentiation (reviewed in Tysnes and Bjerkvig 2007).

ONCOGENES | A proto-oncogene is a gene that under normal conditions controls proper cell growth and differentiation. Mutations in proto-oncogenes (gain-of-function mutations) convert them to oncogenes (Alberts *et al.* 2008). This activation may occur also through chromosomal rearrangement or gene amplification, and usually confers alterations in structure and/or expression level of the oncogene (Vogelstein and Kinzler 2004). These mutations behave in a dominant fashion, meaning that one changed allele can lead to uncontrolled cell proliferation (Alberts *et al.* 2008). Well-known examples of oncogenes are *ERBB2* and *MYCN*.

TUMOR SUPPRESSOR GENES | Tumor suppressor genes are involved in inhibition of cell proliferation and control of cell differentiation in normal cells (Vogelstein and Kinzler 2004). These genes are targeted for inactivation in cancer cells (loss-of-function mutations), however mutations in both alleles of a tumor suppressor gene are generally required to confer a selective advantage to the cell (Alberts *et al.* 2008). This situation commonly arises through the deletion of one allele via a major chromosomal event — such as the loss of an entire chromosome, chromosome arm or region — coupled with an intragenic mutation of the other allele (Knudson 2002). In rare cases, when still one intact allele is retained, the single allele dosage is not enough to suppress tumorigenesis (Santarosa and Ashworth 2004). Well characterized examples of tumor suppressor genes are *PTEN* and *TP53*.

STABILITY GENES | Stability genes keep genetic alterations in the cell to a minimum, and thus, when they are inactivated, mutations in other genes occur at a higher rate (Friedberg 2003). This

class of genes code proteins involved in the mismatch-repair (MMR), nucleotide-excision repair (NER) and basepair-excision repair (BER) responsible for repairing subtle mistakes made during normal DNA replication or induced by exposure to mutagens (reviewed by Yang 2008). This group also includes genes that are responsible for mitotic recombination and chromosomal segregation, such as *BRCA1* and *MSH2* (Vogelstein and Kinzler 2004).

3.2 GENOMIC INSTABILITY UNDERLYING CANCER

Genomic instability is a characteristic of most cancers. This genomic instability can manifest itself as small changes at the nucleotide level affecting cancer-related genes, as gene amplification or as major chromosomal alterations (reviewed in Martin *et al.* 2010). Similarly, epigenetic regulation of genes associated with the maintenance of genomic stability, have also been implicated in cancer development (Issa *et al.* 2007). Another source of tumor associated genomic instability that has been uncovered is the loss of telomeric DNA in many tumors, which generates karyotypic instability and associated amplification and deletion of chromosomal segments (Artandi and DePinho 2010).

The instability of the genome has been suggested to be the means that enable populations of premalignant cells to reach the six biological endpoints acquired during the multistep development of human tumors: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activating invasion and metastasis (Hanahan and Weinberg 2000, 2011). These hallmarks of cancer provide a solid foundation for understanding the biology of cancer (Figure 3.1). Two additional hallmarks of cancer have been suggested, but because neither is yet generalized and fully validated, they are labeled as emerging hallmarks (Hanahan and Weinberg 2011). One involves the capability to modify, or reprogram, cellular metabolism in order to most effectively support tumor proliferation. The second allows cancer cells to evade immunological destruction, in particular by T and B lymphocytes, macrophages and natural killer cells.

Although genomic instability appears to be the driving force of cancer, data from sporadic colorectal cancers showed little evidence of genomic instability in early lesions (Sieber *et al.* 2003). This fact, along with other evidences, raised the question whether genetic instability is an early or late event in the process of tumorigenesis. Despite our current understanding of cancer genomes, it has been difficult to determine if many genetic abnormalities are a cause or consequence of cancer initiation or progression.

Following, some well characterized examples of genome instability and the associated mechanisms are presented.

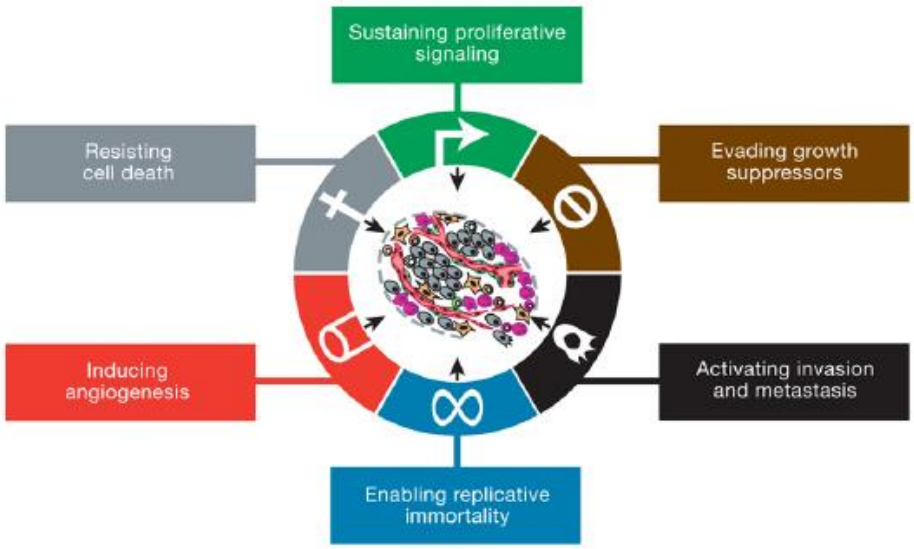


Figure 3.1| The hallmarks of cancer. Distinctive and complementary capabilities that enable tumor growth and metastatic dissemination (from Hanahan and Weinberg 2011).

Whereas human's normal cells contain a specific diploid number of chromosomes that defines the species, some cancer cells are notorious infractors of nature's karyotype stability laws, showing to harbor atypical numbers of chromosomes as well as large-scale structural rearrangements of chromosomes (Figure 3.2). Since Boveri observed abnormal chromosome complements in tumors cells (Boveri 1902, 1914), the role of instability at the chromosome level in tumor initiation and progression has been a central issue in cancer biology. In addition to classical cytogenetic analysis such as G-banding, FISH based methods such as chromosome painting (Ried *et al.* 1998), comparative genomic hybridization (Kallioniemi *et al.* 1992), multiplex FISH (MFISH) (Speicher *et al.* 1996) and spectral karyotyping (SKY) (Schröck *et al.* 1997), enabled researchers to easily identify all the chromosomal segments gained, lost or rearranged in a given cell, determining the genomic complexity of tumor karyotypes with considerable detail.

As previously suggested by Duesberg (2007), the "chaos" at chromosomal level is not just a side effect of malignancy. Although several questions and controversies still remain, recent studies using sophisticated mouse modeling approaches demonstrated that **chromosomal instability** (**CIN**) plays a causative role in a substantial proportion of malignancies (reviewed in Schvartzman *et al.* 2010). One of the consequences of CIN is aneuploidy (Thompson and Compton 2011), known as an imbalance in chromosome number, resulting in losses and gains of whole chromosomes (whole chromosome aneuploidy) or large chromosome parts (segmental aneuploidy) (Geigl *et al.* 2008). Aneuploidy can result in gene dosage changes of thousands of

genes at once and consequently in the corruption of highly conserved proteins involved in the repair or disposal of damaged DNA and segregation of chromosomes (Duesberg *et al.* 2004). Besides, loss of large regions of a chromosome can lead to the inactivation of tumor suppressor genes (Lengauer *et al.* 1998). Aneuploidy may be an initiation event in cancer by promoting the accumulation of mutations in some cancers in the absence of initiation mutations (e.g. Hanks *et al.* 2004, Snape *et al.* 2011). Other studies show aneuploidy as a later event in cancer, being caused by mutations accumulating during cancer progression (Solomon *et al.* 2011). Both aneuploidy and CIN are associated with poor patient prognosis (e.g. Heilig *et al.* 2010) (Figure 3.2).

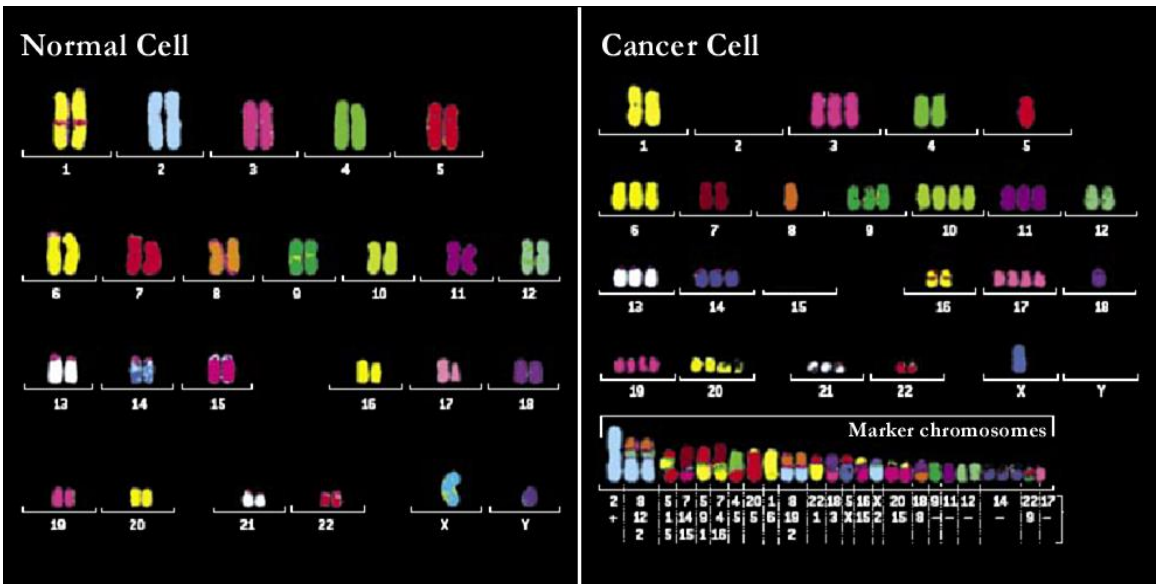


Figure 3.2| Normal and cancer cell karyotypes using multicolor FISH. A normal human cell's chromosome set (left) includes 23 pairs of standard chromosomes, whereas a tumor cell exhibits an irregular karyotype (right). Among the chromosome abnormalities, aneuploidies and structural rearrangements can be seen (forming marker chromosomes) (adapted from Duesberg 2007).

In addition to changes in chromosome numbers, structural alterations in chromosomes (Figure 3.3) are present in a high percentage of both hematological cancers and solid tumors, being the karyotypes of solid tumors far more complex than the ones found in leukemias or lymphomas. These changes may involve multiple chromosomes and multiple breakpoints being regarded as **chromosome structure instability** (**CSI**) (reviewed in Thompson and Compton 2011).

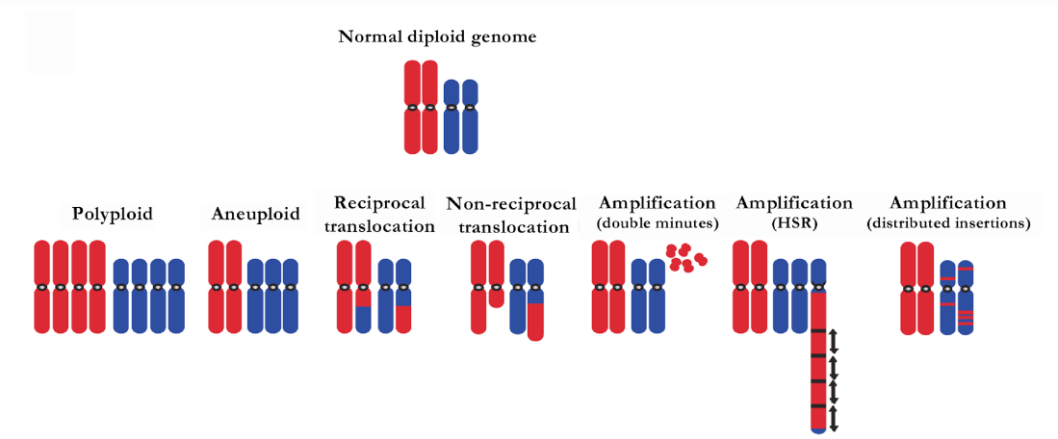


Figure 3.3| Schematic illustration of mechanisms leading to chromosomal alterations. A broad range of chromosomal abnormalities can be present in cancer cells, including altered ploidy, gain or loss of individual chromosomes, structural rearrangements and amplifications (adapted from Albertson *et al.* 2003), that are detailed along this chapter.

MECHANISMS OF CIN| CIN has been shown to result from perturbations of proteins that play key roles in mitosis, including proteins that maintain the mitotic spindle checkpoint and sister chromatid cohesion (Jallepalli *et al.* 2001, Michel *et al.* 2001). The chromosome mis-segregation rate of cancer cells with CIN is about one chromosome every one to five divisions, while in stable diploid cell lines the rate is one chromosome per a hundred cell divisions (Cimini *et al.* 1999, Thompson and Compton 2008). Evidence from human cancer cell lines suggests that the major source of chromosome segregation errors causing CIN is a specific kinetochore–microtubule attachment error called merotely (reviewed in Thompson and Compton 2011). Merotelic attachments are defined by single kinetochores attaching to microtubules emanating from more than one spindle pole (Thompson and Compton 2008). Chromosomes with merotelic attachments align at the metaphase plate once merotely is not detected by the spindle checkpoint (Khodjakov *et al.* 1997). If a cell enters anaphase with a merotelic attachment, the chromatid attached to both poles can segregate to the same daughter cell as its sister resulting in a mis-segregation that produces two aneuploid cells, one with an extra copy of the chromosome and the other with a missing a copy (Cimini *et al.* 2002).

CIN has also been shown to be initiated by tetraploidy, which has been found to precede the development of CIN and aneuploidy in several cancers (Galipeau *et al.* 1996). Different mechanisms can lead to tetraploidy, including cell fusion, mitotic slippage and cytokinesis failure (Holland and Cleveland 2009). In addition to a doubling of the chromosome content, tetraploid cells typically contain twice the normal complement of centrosomes. Centrosomes are involved

in many biological functions as mitotic activity, cell shape, polarity, motility and DNA repair (Wang *et al.* 2004). Aberrant mitotic divisions and chromosome mis-segregation have shown to be promoted by supernumerary centrosomes at a high frequency, and thus, tetraploidy causes an inherently unstable state that acts as a catalyst to promote further aneuploidy and instability (Fujiwara *et al.* 2005).

Recent works show evidence that aneuploidy enhances genetic recombination and defective DNA damage repair (Thompson and Comptom 2010, Sheltzer *et al.* 2011) providing a mechanistic link between aneuploidy and genomic instability. Moreover, aneuploidy has shown to promote the development of high chromosome mis-segregation rates (Solomon *et al.* 2011), what indicates that aneuploidy can cause CIN, and demonstrates that CIN can be a self propagating type of genomic instability.

MECHANISMS OF CSI | Tumor cells are characterized by having structural changes such as deletions, duplications, inversions, isochromosomes, ring chromosomes and translocations (reciprocal and non-reciprocal). Structural chromosome rearrangements can influence tumorigenesis either deregulating expression of specific-target genes or by producing a hybrid, chimeric gene through fusion of parts of two genes on separate chromosomes (Mitelman *et al.* 2007). Gene fusions resulting from translocations are frequent and characteristic from blood cancers, as is the example of the chimeric gene fusion BCR-ABL found in 95% of patients with chronic myeloid leukemia (Kurzrock *et al.* 2003). Recurrent gene fusions appear to be quite rare in solid tumors, although they may also play a role in tumorigenesis in some particular cases (Tomlins *et al.* 2005).

Chromosome structure instability mechanisms are now starting to be disclosed, nevertheless, it seems that CSI can be the result of errors in the DNA damage checkpoints, DNA repair pathways and/or mitotic segregation errors (Thompson and Compton 2011).
Double stranded breaks (DSBs) are among the lesions that cause structural chromosomal instability by initiating rearrangements such as translocations (Richardson and Jasin 2000, Gent *et al.* 2001). Accordingly, mutations in proteins that permit cell cycle progression in the presence of double stranded breaks (e.g. p53, BRCA1, BRCA2, ATM and ATR) may also facilitate CSI (Lengauer *et al.* 1998).

3.2.1 GENE AMPLIFICATION

Gene amplification constitutes a copy number increase of a restricted region of a chromosome arm (Albertson 2006). During normal developmental states, such as oogenesis in *Drosophila*, gene amplification is a characteristic occurring process being strictly developmentally controlled (Tower 2004). Nevertheless, genomic DNA copy number aberrations are frequent in solid tumors and are expected to contribute to tumor evolution (Albertson 2006). The amplification of chromosomal regions might promote tumorigenesis by the activation of proto-oncogenes (Lengauer *et al.* 1998).

The cytogenetic analysis of amplified DNA in mammalian cell lines and tumors has revealed that it can be organized as extrachromosomal copies (called double minutes); in tandem arrays as head-to-tail or inverted repeats within a chromosome (often forming a cytologically visible homogeneously staining region - HSR); or distributed at various locations in the genome (Albertson *et al.* 2003) (Figure 3.3). The size of the amplified region (amplicon) can range from few kilo bases (kb) to several mega bases (Mb) of DNA (Albertson 2006). The mechanism that is likely to contribute for gene amplification initiations is the DNA double-strand break in cells lacking robust checkpoints (Paulson *et al.* 1998, Pipiras *et al.* 1998).

Amplification of a gene may result in an increase in the gene dosage, often leading to up-regulation of gene expression (Schwab 1998, Savelyeva and Schwab 2001). Although a high correlation between amplification and overexpression was demonstrated by several studies, there is not always equivalence between the level of DNA gain and gene expression (Santarius *et al.* 2010). In clinical aspects, amplification may be used as a molecular marker for prognosis as well as in the selection of the most appropriate therapeutic strategy (Al-Kuraya *et al.* 2004). Examples of recurrently amplified oncogenes are the *ERBB2* in breast cancer (Harari and Yarden 2000) and *MYCN* in neuroblastoma (Westermark *et al.* 2011).

ERBB2 GENE | The *ERBB2* gene (human epidermal growth factor receptor 2, also known as *HER-2/neu*) is located in human chromosome 17q and encodes a 185-kDa transmembrane tyrosine kinase growth factor (Schechter *et al.* 1984, Hung and Lau 1999). Deregulated signal transduction can be caused by *ERBB2* mutation, amplification and/or overexpression, leading to cellular immortalization, neoplastic transformation and tumor progression (Harari and Yarden 2000). *ERBB2* represents the best described and most commonly clinically used amplified cancer gene. It has been shown to be amplified in a subset of gastric, lung and salivary carcinomas, but is most commonly amplified in breast carcinoma (20%-30% of invasive breast cancer) (Hynes 2007). The discovery of *ERBB2* gene amplification in primary human breast cancer and its

association with a more aggressive clinical behavior led to early interest in diagnostic and therapeutic applications (Slamon *et al.* 1987). *ERBB2* gene amplification was associated with the high erbB-2 protein overexpression levels found in human carcinomas of the breast, and they were associated with an unfavorable prognosis (Slamon *et al.* 1989). Trastuzumab (Herceptin) is a humanized monoclonal antibody directed against the extracellular domain of the erbB-2 protein that have been found to be effective when in presence of high levels of this protein (Goldenberg 1999). The *ERBB2* gene and erbB-2 protein status (gene amplification/protein overexpression) are considered useful markers for predicting the response to a specific cancer therapy, and analysis of these markers is mandatory for the identification of breast cancer patients that are amenable to trastuzumab treatment (Baselga *et al.* 2001, Slamon *et al.* 2001).

MYCN GENE | The *MYCN* is part of the MYC oncogene family, one of the first cancer related genes discovered (reviewed in Wasylishen and Penn 2010). This gene is localized in human chromosome 2p, and encodes the MYCN transcriptional regulator, predominantly expressed in the developing peripheral neural crest (Zimmerman *et al.* 1986, Zindy *et al.* 2006). *MYCN* is thought to be critical in the tumorigenesis of human neuroblastoma (Weiss *et al.* 1997), with gene amplification found in 20% of these childhood cancers, associated with aggressive disease, rapid progression and poor prognosis (Westermark *et al.* 2011). The initial studies suggested that *MYCN* amplification was specific for neuroblastoma, although latter surveys showed that this mark can also be seen in small cell lung cancer, retinoblastoma, malignant gliomas and peripheral neuroectodermal tumors, although at a much lower incidence (Schwa 2004). As a common feature, all these tumors have neural qualities. Moreover, *Mycn* amplification has also been observed in rat tumors, specifically in uterine endometrial carcinomas (Karlsson *et al.* 2001, Adamovic *et al.* 2005). Amplified *MYCN* has been found only in more aggressive variants of neuroblastoma, having emerged clinically as a powerful independent marker to predict poor patient outcome (reviewed in van Noesel *et al.* 2004). Studies in different human neuroblastoma cell lines have shown that the high-expression level of *MYCN* is considered to be driven by DNA copy number increases, with the most amplified cell lines expressing the highest levels of the gene expression (reviewed in Schwab 1998). Also *MYCN* up-regulation has been associated with human inflammatory breast tumorigenesis, in a study by Bièche *et al.* (2004).

3.3.2 DNA METHYLATION

Epigenetics can be described as a stable alteration in gene expression potential without any change in gene sequence, which takes place during development and cell proliferation

(Kouzarides 2007). The **DNA methylation** is one of the most commonly occurring epigenetic events taking place in the mammalian genome. Besides DNA methylation, epigenetic regulation can be mediated by other molecular mechanisms such as chromatin/histone modifications (reviewed in Cheung and Lau 2005). DNA methylation refers to the covalent post-replicative addition of a methyl group onto the 5th position carbon of the cytosine ring within CpG dinucleotides (Wyatt 1950), being the reaction catalyzed by the enzymes DNA methyltransferases (DNMT) (Goll and Bestor 2005). The methylation of DNA plays a key role in chromatin structure and stable suppression of gene expression (epigenetic silencing), namely in the inactive X chromosome and imprinted genes (Bird 2002). It has been demonstrated that increased methylation in the promoter region of a gene leads to reduced expression, whereas methylation in the transcribed region has a variable effect on gene expression (Singal *et al.* 2002, Dahl *et al.* 2011). DNA methylation can repress gene expression, by means of two main mechanisms: by interfering with the assembly of the transcription machinery and/or by causing a change in chromatin structure via various methyl-CpG binding proteins (e.g. Jones *et al.* 1998, Kaludov and Wolffe 2000).

Much attention has focused on DNA methylation in CpG islands, although the majority of methylated CpG dinucleotides are in fact found within repetitive elements, which in total comprise approximately half of the human genome (reviewed in Wild and Flanagan 2010). It is believed that DNA methylation is an essential mechanism silencing the transcription of these elements to prevent their movement and expansion throughout the genome (Peng and Karpen 2008).

Significant changes in genome-wide DNA methylation have been observed in cultured cancer cells and primary human tumors. For this reason, influence of DNA methylation in cancer has become the topic of intense investigation. As compared with normal cells, the malignant cells show major disruptions in their DNA methylation patterns (Das and Singal 2004). Moreover, it has been proposed that changes in DNA methylation can greatly influence genetic instability (Veigl *et al.* 1998). Hypomethylation and hypermethylation of DNA are relative terms applied to cancer epigenetics and denote less or more methylation than in normal tissues, respectively. Generally, the genome of cancer cells is characterized by global loss of DNA methylation (genome-wide hypomethylation) and regional hypermethylation of CpG islands (Brena *et al.* 2006) (Figure 3.4). Hypomethylation usually occurs in repeated DNA sequences, being the centromeric alpha- satellite, the interspersed Alu and the long interspersed elements (LINE)-1 repeats the most frequently studied DNA cancer hypomethylated repeats (reviewed in Ehrlich 2009). Abnormal DNA hypermethylation at gene promoter CpG islands has been shown

to contribute to transcriptional repression of many genes in different cancers (reviewed in Jones and Baylin 2007) (Figure 3.4).

Unlike genetic changes which cause permanent damage to the genome, changes in DNA methylation patterns are potentially reversible. Therefore, they have been considered targets for therapeutic intervention (Kelly *et al.* 2010). This has led to an intensive search for drugs that target components of the epigenetic machinery. Two examples are inhibitors of DNA methylation, 5-azacytidine and 5-aza-2'-deoxycytidine, which have been known for decades, have now been approved for treatment of patients with myelodysplastic syndrome (reviewed in Dahl *et al.* 2011). 5-Azacytidine and 5-aza-2-deoxycytidine are cytosine analogues that trap all DNA methyltransferases and target them for degradation (reviewed in Issa *et al.* 2007). At low doses that do not inhibit proliferation, these drugs are effective hypomethylating agents and they have shown clinical activity as anticancer agents.

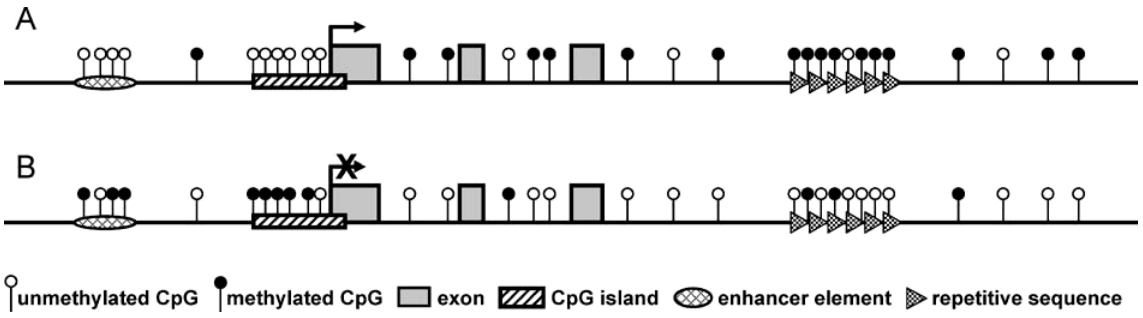


Figure 3.4| Commonly observed DNA methylation changes in cancer. (A) In normal tissues, the majority of CpG islands and regulatory elements are methylation-free, while repetitive sequences and interspersed CpG dinucleotides are heavily methylated. (B) The genome of cancer cells is characterized by global loss of DNA methylation and regional hypermethylation of CpG islands and other gene regulatory regions (from Brena *et al.* 2006).

4. USING RODENTIA SPECIES AS MODEL

This thesis concerns the study of karyotype restructuring along evolutionary processes, such as species evolution and cancer, focusing in the chromosome dynamics featuring at some point particular sequences as satellite DNAs. The chosen species to work were rodents, namely wild species and the well known model organism, the laboratory rat (*Rattus norvegicus*).

Rodentia species have a long tradition as model organisms, and new models keep emerging. The laboratory rat has been extensively used as an animal model for physiology, pharmacology, toxicology, nutrition, behavior, immunology and neoplasia (Aitman *et al.* 2008). Accordingly, in the last decade there has been an extraordinary increase in rat genomic resources (Gibbs *et al.* 2004, Hamta *et al.* 2006). These achievements in rat genome discoveries have been translated to human disease, and contributed to the increasing speed of discovery of new disease genes, pathways and mechanisms (Aitman *et al.* 2008). Also the mouse (*Mus musculus*) has a century of genetic studies, creation of inbred strains, hundreds of spontaneous mutations, practical techniques for random mutagenesis, and, importantly, directed engineering of the genome through transgenic, knockout and knockin techniques (e.g. Yu and Bradley 2001, Bucan and Abel 2002).

The rodents are ubiquitous and occur in all continents (except Antarctica). Rodent's remarkable diversity has always been a challenge when it comes to determine their origins, ways of radiation and times of diversification. A brief review regarding the origins and taxonomy inside the Order Rodentia will be made in the next paragraphs. Here it will be presented the taxonomy followed in this work and highlight the complex phylogenies within this diverse group.

4.1 RODENTS EVOLUTION AND COMPLEX PHYLOGENY

At approximately 62 to 100 Ma, the rodents arose along a branch of the mammalian lineage (Benton and Donoghue 2007) (Figure 4.1) sharing a common ancestor with lagomorphs (rabbits, pikas), forming the clade Glires (Luckett and Hartenberger 1993). Glires share a common ancestry with primates, tree shrews, and the flying lemurs (e.g. Churakov *et al.* 2010). Early studies based in paleontological data and latter based in morphological characters demonstrated that rodent's evolution was monophyletic, assuming a common origin for all species (Luckett and Hartenberger 1993).

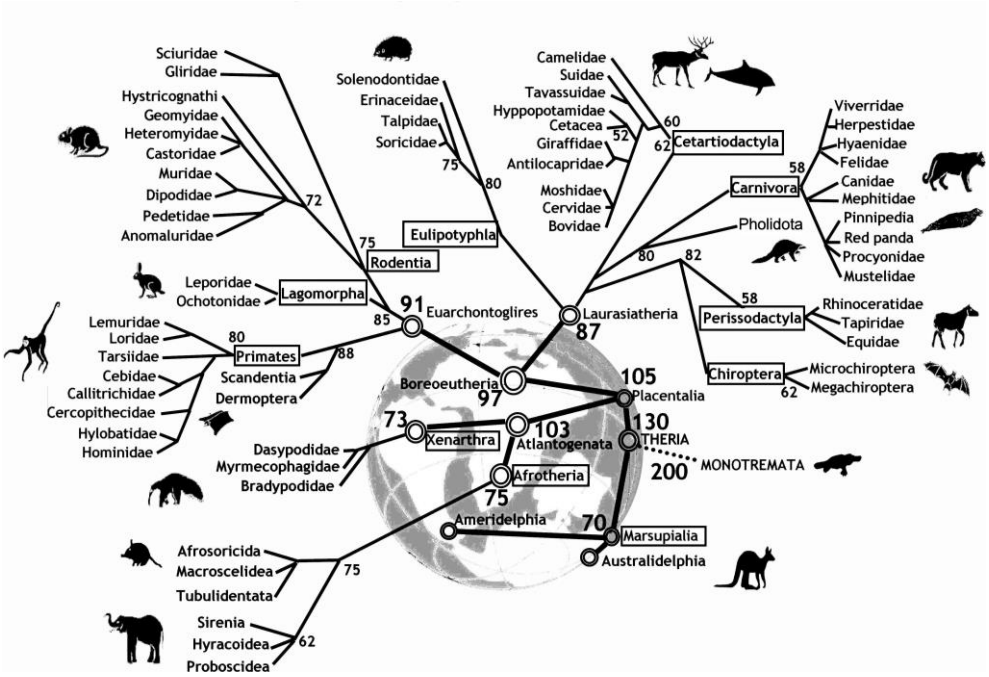


Figure 4.1 | The place of Rodentia in mammal's evolutionary tree. The tree depicts historic divergence relationships among the living orders of mammals. The phylogenetic hierarchy is a consensus view of several decades of molecular genetic, morphological and fossil inference. The rodent's divergence is highlighted in blue. Double rings indicate mammalian supertaxa, numbers indicate the possible time of divergences (adapted from Graphodatsky *et al.* 2011).

Characteristics such as small size, short breeding cycles, and wide variety of foods eaten, probably favored rodents rapid arising and lead them to become one of the most successful mammalian groups, occupying nearly all continents. The order Rodentia represents half of the placental mammals and includes, at least, 2277 species divided into 33 families (Musser and Carleton 2005), being the most abundant and diversified order of living mammals.

Even though several evolutionary relationships within Rodentia have not been easy to determine, the rodents are currently divided in five suborders: Sciuromorpha, Myomorpha (or Myodonta), Anomaluromorpha, Castorimorpha and Ctenohystrica (or Hystricomorpha) (Musser and Carleton 2005). In recent studies, these five suborders were clustered into three main lineages: mouse-related clade (includes Anomaluromorpha, Castorimorpha and Myomorpha), squirrel-related clade (includes Sciuromorpha) and Ctenohystrica (includes Hystricomorpha) (Huchon *et al.* 2007, Montgelard e*t al.* 2008) (Figure 4.2). The mouse and squirrel-related clades were proposed as the most divergent rodent lineages (Montgelard *et al.* 2008, Blanga-Kanfi *et al.* 2009).

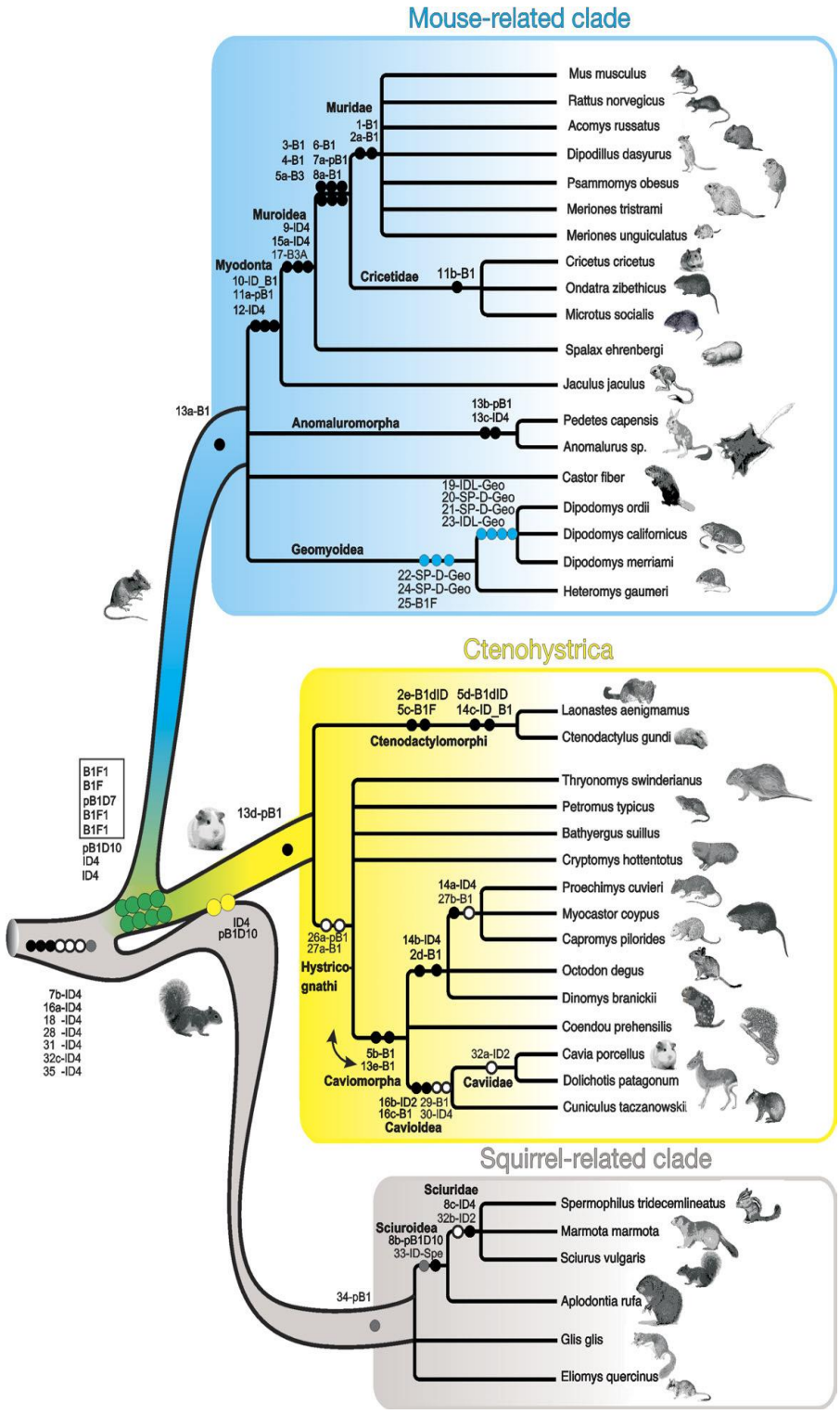


Figure 4.2| Phylogenetic tree of Rodentia. This phylogenetic tree is based on retroposon (SINEs) presence-absence data, and illustrates the classification of rodents in three clades: mouse-related clade, ctenohystrica clade, and squirrel-related clade, as supported by other studies (Huchon *et al.* 2007, Montgelard *et al.* 2008). This study supports an early divergence of a pre-Squirrel-related clade from a common ancestor of a pre-Mouse-related/Ctenohystrica clade, providing some understanding to early speciation in rodents (from Churakov *et al.* 2010).

4.1.1 THE SUPERFAMILY MUROIDEA

The Muroid rodents represent the most diverse and geographically widespread subordinal clade, comprising more than 1500 species with a nearly global distribution (Musser and Carleton 2005). Muroidea members are characterized by myomorphic jaw structure that have lost the upper fourth premolar and have a well-developed anterocone (-id) on the first molar (Flynn *et al.* 1985).

Their presumably rapid radiation left little opportunity for the evolution of distinctive synapomorphies, and so the morphologic phylogenetic studies end up by relying on dental characters which are particularly prone to adaptive convergence. The use of molecular genetic characters assisted in the clarification of the phylogenetic relationships of muroids. Based in molecular studies, six different families were acknowledged to Muroidea: Platacanthomyidae, Spalacidae, Calomyscidae, Nesomyidae, Muridae and Cricetidae (Musser and Carleton 2005). In the next paragraphs only the families (Muridae and Cricetidae) and subfamilies (Murinae, Cricetinae, Neotominae and Arvicolinae), which include the studied species in this work, will be referred in detail.

4.1.1.1 THE MURIDAE FAMILY

The Muridae family constitutes the largest and most diverse family of rodents, comprising about 730 species distributed by 150 genera (Musser and Carleton 2005). According to fossil records, the Muridae originated approximately 54 to 37 Ma, but the large-scale radiation events occurred only between 1.8 to 0.01 Ma (Benton 1997). Murids occupy a broad range of habitats from tropical rain forests to arid deserts and tundra having adopted an equally wide array of lifestyles. It is important to mention that in the literature, the family name Muridae is sometimes used in a broader sense to include all members of the superfamily Muroidea, what doesn't happen in the present study. The phylogenetic analysis at subfamilial level of Muridae has been debated along the years, supported by both paleontological and molecular studies (Senegas and Avery 1998, Jansa and Weksler 2004, Stteppan *et al.* 2004, Steppan *et al.* 2005). Till this date, no cladistic analyses present a broad species sampling among all the families or across extant geographic range, what difficults Muridae complete phylogenetic resolution.

The most representative Murid subfamily, comprising about 561 species, is the Murinae subfamily (Old World mice and rats), which includes the popular animal models, mouse (*Mus musculus*) and rat (***Rattus norvegicus***) (Musser and Carleton 2005).

Within the Old World rodents, the *Praomys* group constitutes a diverse and abundant group, occupying various African biotypes ranging from the equatorial rain forest to Sahelian

savannas (Nicolas *et al.* 2005). This group comprises several genera, which have been divided into two complexes, the jacksoni-complex and the *tullbergi*-complex. ***Praomys tullbergi*** is one of the four species comprising the *tullbergi*-complex (Lecompte *et al.* 2005, Nicolas *et al.* 2005). The basal relationships within *Praomys* group suggest a rapid radiation at about 7 to 9 Ma based on genetic divergence rates calibrated from the fossil record (Lecompte *et al.* 2005). Recently, it was proposed a tribal level of classification within Murinae, and the new name Praomyini tribe nov., which includes *Colomys*, *Heimyscus*, *Hylomyscus*, *Mastomys*, *Myomyscus*, *Praomys*, *Stenocephalemys* and *Zelotomys* (Lecompte *et al.* 2008).

4.1.1.2 THE CRICETIDAE FAMILY

The Cricetidae family is considered by many paleontologists to be ancestral to the Muridae family, with its emergence approximately 58 to 47 Ma according to fossil records (Hartenberger 1998). Some of the species studied in this work belong to the Cricetidae family, and for this reason the following information will only regard three of the subfamilies included in this group: Cricetinae, Arvicolinae and Neotominae (Musser and Carleton 2005).

The Cricetinae subfamily, originated approximately 12 to 5 Ma (Neumann *et al.* 2006), harbors the Old World hamsters. Morphological and paleontological data about this group revealed not enough reliable synapomorphies for phylogenetic resolution (e.g. Hir 1997, Kowalski 2001), and, although molecular studies clarified some aspects of this subfamily systematic (Steppan *et al.* 2004, Neumann *et al.* 2006), the phylogenetic relationships between genera remains unclear. Within Cricetinae subfamily are included the genus *Cricetus* and *Phodopus*. The genus *Cricetus* comprises only one species, ***Cricetus cricetus*** (common hamster) (Nechay 2000). *Cricetus* hamsters' fossils were dated from 2.5 Ma (Hir 1997) but only *Cricetus cricetus* survived areal shifts and local extinction during glacial oscillations. The genus *Phodopus* harbours the small sized hamsters (dwarf hamsters), including ***Phodopus sungorus*** (djungarian hamster). Phylogenetic studies based on molecular data propose that this genus represents the earliest split among all hamsters (Neumann *et al.* 2006).

To the Neotominae subfamily belong the rodents that inhabit North America desert areas (Engel *et al.* 1998). According to fossil data, this subfamily emerged approximately 16 to 9 Ma (Baskin 1989). Among them are the well-known deer mice (*Peromyscus maniculatus*), white-footed mice (*Peromyscus leucopus*), and cactus mouse (***Peromyscus eremicus***). The phylogenetic relationships within this group have been difficult to clarify (Bradley *et al.* 2007).

The subfamily Arvicolinae (lemmings, muskrats, and voles) harbors about 151 species distributed across 28 genera (Musser and Carleton 2005), from which more than 60 are found

within the vole genus *Microtus*, which is thought to have originated 0.5 to 2 Ma (Reppenning 1980, Chaline *et al.* 1999). Within this subfamily is included ***Microtus arvalis*** (common vole). The limited variation in external morphology has been a significant challenge in *Microtus* classification and this has made cytogenetic data important for solving problems of vole taxonomy.

In summary, the phylogenetic relationships in Rodentia are still far from being fully revealed in all their complexity. Along this thesis, the described taxonomy shown in Figure 4.3 was followed. This taxonomy is available in databases, namely "NCBI Taxonomy" (www.ncbi.nlm.nih.gov/taxonomy) and "Tree of Life Web Project" (www.tolweb.org/tree/), which compile and update data from both fossil records and molecular studies.

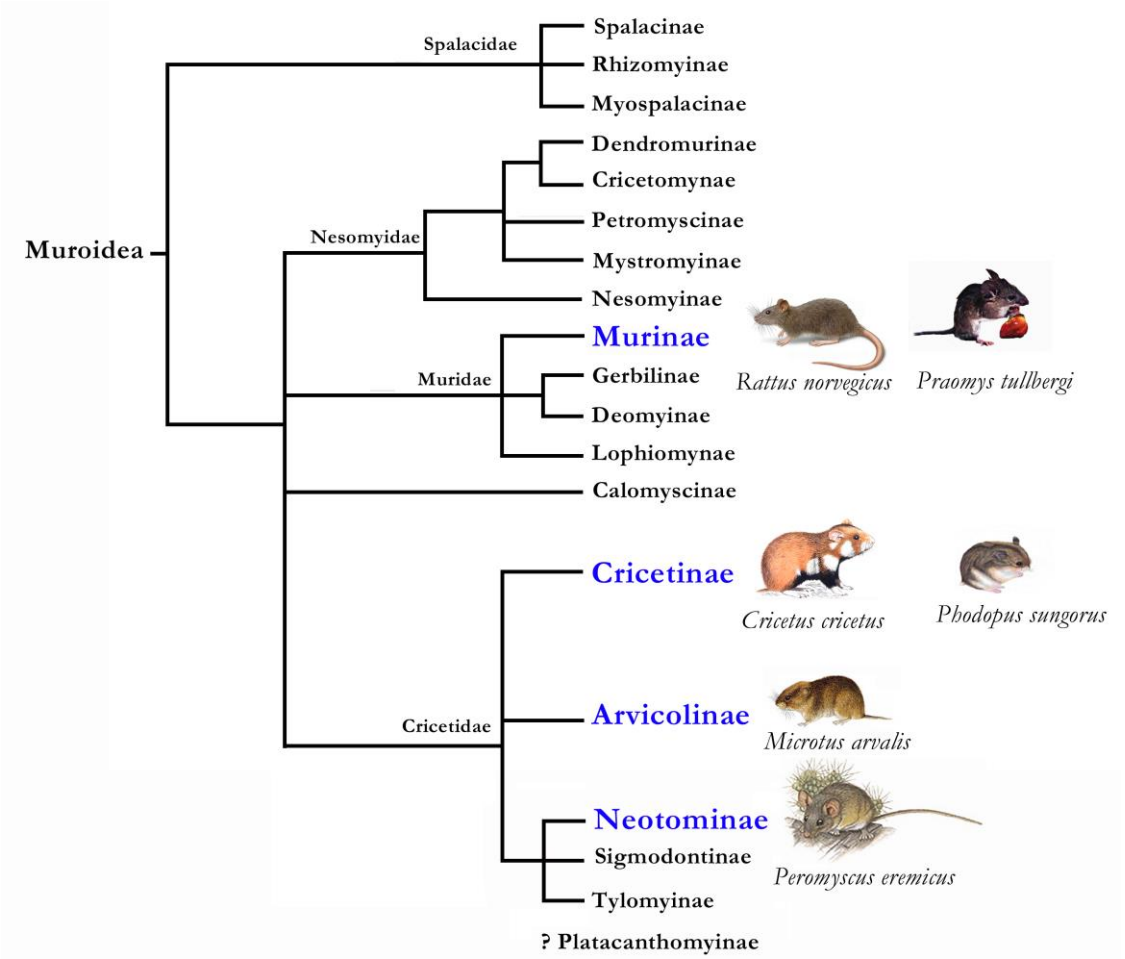


Figure 4.3| Phylogenetic tree of Muroidea superfamily. This tree is based in molecular studies from several authors [Robinson et al. (1997), Michaux and Catzeflis (2000), Michaux *et al.* (2001), Jansa and Weksler (2004) and Steppan *et al.* (2004)] (adapted from Tree of Life Web Project). In the tree are highlighted the studied species in this thesis.

4.2 WHY STUDY RODENTS KARYOTYPE EVOLUTION?

Rodentia order constitutes the most numerous and evolutionarily diverse taxon of mammals. Rodents represent a very karyotypically diverse group, with diploid numbers ranging from 2n=10 to 2n=102 (e.g. Contreras *et al.* 1990, Silva *et al.* 2006). Besides, heterochromatin content and C-value (haploid DNA content) also vary greatly between species (reviewed in Graphodatsky *et al.* 2011).

Compared to humans and other mammals, rodent genomes, specifically Muroidea species, underwent intense chromosome reshufflings, in which many complex structural rearrangements occurred (reviewed in Romanenko *et al.* 2012). This fact makes them preferential animal models for studying the process of karyotype evolution, and so, as more Muroidea species are analyzed by comparative studies, more unraveled the evolutionary events will be. The comparative maps construction, specifically for mouse and rat, was crucial for understanding the human genome functionalities. This allows that information's gathered from high resolution mapping in mouse and rat genomes can be used to make predictions about the orthologous regions of the human genome (e.g. Guigo *et al.* 2003). Besides, biomedical studies of human genes could be complemented by experimental manipulations of corresponding rodent model genes to accelerate functional understanding (Waterston *et al.* 2002).

Many other rodent species have been analyzed by comparative chromosomics with the purpose of tracing the evolution of key traits and determine phylogenies. Two of the species studied, *Cricetus cricetus* and *Peromyscus eremicus*, belong to groups of species that have been used as models. *Peromyscus* species constitute emerging models in diverse areas as ecology, behavior, physiology, reproductive and developmental biology, biochemistry, chromosomal evolution, allozymes, cytogenetics, speciation, biogeography and many other areas of research (Bradley *et al.* 2007). Also hamsters (including *Cricetus cricetus*) have been used in research in areas such as cancer research (Reznik 1977) and neurology (Pévet *et al.* 1987). For this reason, comparative studies are of great importance to be applied to other areas besides evolution, such as physiology, genomics, immunology and oncology, once rodent model organisms are used in biomedical research (e.g. Lee *et al.* 2004).

Furthermore, a robust phylogeny is demanding to interpret the many studies as well as identifying species and clades for potential future biomedical studies. This is particularly interesting in the case of *Praomys tullbergi*, once this species belongs to one of the more difficult and unresolved group, the so called tullbergi-complex, and the comparative studies may shed some light in revealing phylogenetic relationships.

An important source of karyotype variability in rodents is related to heterochromatin amount and its distribution pattern (Volobouev *et al.* 2006). The differences in the amount of CH in rodents genome and their elevated evolutionary rates, makes this group an excellent candidate to study the dynamic behavior of satellite DNA sequences (main constituent of CH), and its possible involvement in the intense genome reshuffling occurred along rodents evolution.

4.3 CELLULAR MODELS FOR CANCER STUDY

As shown along this thesis, cancer constitutes a complex disease that involves a sequence of gene-environment interactions in a progressive process that cannot occur without dysfunction in multiple systems. Understanding the dynamic nature of the cancer genome is important to comprehend the mechanisms of genetic heterogeneity and population diversity, which is the genetic basis for cancer formation.

A major opportunity to increase our knowledge regarding the biology of cancer is associated with the availability of experimental model systems that recapitulate the many forms of this disease. *In vitro* cell culture models play an important role in understanding the molecular basis of cancer, as a majority of researchers actively use *in vitro* models that can be relatively easier for experimentation (Ye *et al.* 2007). Using the evolutionary conservation of gene segments as a guide, animal models, such as rodents, constitute powerful tools to decipher pathways and genes involved in tumorigenesis (e.g. Aitman *et al.* 2008). Moreover, researchers now have access to powerful web servers and databases in which syntenic regions can be easily identified and associated with a great amount of information concerning human genetics.

By studying the biology and pathobiology of the rat allows translating this information to humans, particularly in the context of complex traits. With the advances in molecular genetics, animal models of human diseases are becoming more numerous and more refined.

USING THE RAT FOR MODELLING CANCER | Rat models resulting from chemically induced carcinogenesis of the mammary gland have been extensively used specifically in the study of breast cancer. The carcinogenic compound 7,12-dimethylbenz[a]anthrazene (DMBA) is frequently used to induce such tumors, and DMBA-induced rat mammary tumors and sarcomas are useful cancer models (e.g. Aitman *et al.* 2008). However, some available animal tumor cell lines are often poorly characterized from a cytogenetic/genetic point of view, reducing their usefulness as cell models. In the present thesis the fully cytogenetic and genetic characterization of two rat mammary tumor cell lines was performed and suggested as a cellular model for *Erbb2* study.

5. REFERENCES

Adamovic T, Trosso F, Roshani L, Andersson L, Petersen G, Rajaei S, Helou K, Levan G (2005) Oncogene Amplification in the Proximal Part of Chromosome 6 in Rat Endometrial Adenocarcinoma as Revealed by Combined BAC/PAC FISH, chromosome Painting, Zoo-FISH, and Allelotyping. Genes Chromosomes Cancer 44: 139–153.

Adega F, Chaves R, Guedes-Pinto H (2005) Chromosome restriction enzyme digestion in domestic pig (Sus scrofa) constitutive heterochromatin arrangement. Genes Genet Syst 80(1):49-56.

Adega F, Chaves R, Guedes-Pinto H (2007) Chromosomal evolution and phylogenetic analyses in Tayassu pecari and Pecari tajacu (Tayassuidae): Tales from the constitutive heterochromatin. J Genet 86: 19–26.

Adega F, Chaves R, Guedes-Pinto H (2008) Suiformes orthologous orthogous satellite DNAs as a hallmark of Pecari tajacu and Tayassu pecari (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(1-2):12-20.

Aitman TJ, Critser JK, Cuppen E, Dominiczak A, Fernandez-Suarez XM, Flint J, Gauguier D, Geurts AM, Gould M, Harris PC, Holmdahl R, Hubner N, Izsvák Z, Jacob HJ, Kuramoto T, Kwitek AE, Marrone A, Mashimo T, Moreno C, Mullins J, Mullins L, Olsson T, Pravenec M, Riley L, Saar K, Serikawa T, Shull JD, Szpirer C, Twigger SN, Voigt B, Worley K (2008) Progress and prospects in rat genetics: a community view. Nat Genet 40(5): 516-522.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2008) Molecular Biology of the cell. New York: Garland Science. 1463 p.

Albertson DG (2006) Gene amplification in cancer. Trends Genet 22(8):447-55.

Albertson DG, Collins C, McCormick F, Gray JW (2003) Chromosome aberrations in solid tumors. Nat Genet 34(4):369-76.

Al-Kuraya K, Schraml P, Torhorst J, Tapia C, Zaharieva B, Novotny H, Spichtin H, Maurer R, Mirlacher M, Köchli O, Zuber M, Dieterich H, Mross F, Wilber K, Simon R, Sauter G (2004) Prognostic relevance of gene amplifications and coamplifications in breast cancer. Cancer Res 64(23): 8534-40.

Arabidopsis genome initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–781.

Armengol L, Pujana MA, Cheung J, Scherer SW, Estivill X (2003) Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. Hum Mol Genet 12(17):2201-8.

Arnason U, Manolova Y, Manolov G, Bregula U, Levan A (1986) Satellite DNA in the three C bands of an unusual mouse marker chromosome-A model of chromosomal evolution. Exp Cell Res 164: 256-260.

Artandi SE, DePinho RA (2010) Telomeres and telomerase in cancer. Carcinogenesis 31: 9–18.

Bailey JA, Baertsch R, Kent WJ, Haussler D, Eichler EE (2004) Hotspots of mammalian chromosomal evolution. Genome Biol 5(4):R23.

Bailey JA, Yavor AM, Massa HF, Trask B J, Eichler EE (2001) Segmental duplications: organization and impact within the current human genome project assembly. Genome Res 11: 1005-1017.

Baselga J, Albanell J, Molina MA, Arribas J (2001) Mechanism of action of trastuzumab and scientific update. Semin Oncol 28:4-11.

Baskin JA (1989) The initial origin and diversification of the Neotropical Sigmodontinae (Rodentia: Muridae) – a perspective from the North American fossil record. pp. 263- 264 in Abstracts of papers and posters. Fifth International Theriological Congress, Rome.

Bassi C, Magnani I, Sacchi N, Saccone S, Ventura A, Rocchi M, Marozzi A, Ginelli E, Meneveri R (2000) Molecular structure and evolution of DNA sequences located at the alpha satellite boundary of chromosome 20. Gene 256(1-2):43-50.

Baylin SB, Herman JG (2000) DNA hypermethylation in tumorigenesis: Epigenetics joins genetics. Trends Genet 16:168-174.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS Haussler D (2004) Ultraconserved Elements in the Human Genome Science 304(5675) 1321-1325.

Benton MJ (1997) Vertebrate Palaeontology. Second Edition. Chapman & Hall, London. 288-362 pp.

Benton MJ, Donoghue PC (2007) Paleontological evidence to date the tree of life. Mol Biol Evol 24:26–53.

Bièche I, Lerebours F, Tozlu S, Espie M, Marty M, Lidereau R (2004) Molecular profiling of inflammatory breast cancer: identification of a poor prognosis gene expression signature. Clin Cancer Res 10(20): 6789-6795.

Birchler JA, Gao Z, Sharma A, Presting GG, Han F (2011) Epigenetic aspects of centromere function in plants. Curr Opin Plant Biol 14(2): 217-222.

Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16(1):6-21.

Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry RW, Huchon D (2009) Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. BMC Evol Biol 9:71.

Boveri T (1902) Über mehrpolige mitosen als mittel zur analyse des zellkerns. Verh. Phys Med Ges Würzburg 35: 67–90.

Boveri T (1914) Zur Frage der Entstehung Maligner Tumoren. Germany: Gustav Fischer, p 64.

Bradley RD, Durish ND, Rogers DS, Miller JR, Engstrom MD, Kilpatrick CW (2007) Toward a molecular phylogeny for peromyscus: evidence from mitochondrial cytochrome-b sequences. J Mammal. 8(5):1146-1159.

Bradley RD, Wichman HA (1994) Rapidly evolving repetitive DNAs in a conservative genome: a test of factors that affect chromosomal evolution. Chomosom Res 2:354–360.

Brannan CI, Disteche CM, Park LS, Copeland NG, Jenkins NA (2001) Autosomal telomere exchange results in the the rapid amplification and dispersion and Csf2ra genes in wild-derived mice. Mamm Genome 12:882-886.

Brena RM, Huang TH, Plass C (2006) Quantitative assessment of DNA methylation: Potential applications for disease diagnosis, classification, and prognosis in clinical settings. J Mol Med (Berl) 84(5):365-77.

Britten RJ, Khone DE (1968) Repeated sequences in DNA. Science 161:529-540.

Bucan M, Abel T (2002) The mouse: genetics meets behaviour. Nature Rev Genet 3:114–123.

Capy P, Bazin C, Higuet D, Langin T (1997) Dynamics and evolution of transposable elements. Texas: Landes Bioscience. 197p.

Caraballo DA, Belluscio PM, Rossi MS (2010) The library model for satellite DNA evolution: a case study with the rodents of the genus Ctenomys (Octodontidae) from the Iberá marsh, Argentina. Genética 138(11-12):1201-1210.

Carbone L, Nergadze SG, Magnani E, Misceo D, Cardone MF, Roberto R, Bertoni L, Attolini C, Piras MF, De Jong P, Raudsepp T, Chowdhary BP, Guerin G, Archidiacono N, Rocchi M, Giulotto E (2006) Evolutionary movement of centromeres in horse, donkey, and zebra. Genomics 87:777-782.

Castro MA, Onsten TG, Moreira JC, de Almeida RM (2006) Chromosome aberrations in solid tumors have a stochastic nature. Mutat Res 600:150–164.

Cazaux B, Catalan J, Veyrunes F, Douzery EJ, Britton-Davidian J (2011) Are ribosomal DNA clusters rearrangement hotspots?: a case study in the genus Mus (Rodentia, Muridae). BMC Evol Biol 11:124. doi: 10.1186/1471-2148-11-124.

Chaline J, Brunet-Lecomte P, Motuire S et al (1999) Anatomy of the arvicoline radiation (Rodentia): paleaogeographical, paleaoecological history and evolutionary data. Ann Zool Fennici 36:239–267.

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371(6494):215-20.

Chaves R, Adega F, Heslop-Harrison JS, Guedes-Pinto H, Wienberg J (2003) Complex satellite DNA reshuffling in the polymorphic t(1;29) Robertsonian translocation and evolutionarily derived chromosomes in cattle. Chromosome Res 11(7):641-8.

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

Chaves R, Louzada S, Meles S, Wienberg J, Adega M (2012) Praomys tullbergi (Muridae, Rodentia) genome architecture decoded by comparative chromosome painting with Mus and Rattus. Chromosome Res DOI 10.1007/s10577-012-9304-1.

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.

Cheung P, Lau P (2005) Epigenetic Regulation by Histone Methylation and Histone Variants. Mol Endocrinol 19(3):563-73.

Chiodi I, Corioni M, Giordano M, Valgardsdottir R, Ghigna C, Cobianchi F, Xu RM, Riva S, Biamonti G (2004) RNA recognition motif 2 directs the recruitment of SF2/ASF to nuclear stress bodies. Nucleic Acids Res 32: 4127–4136.

Chowdhary BP, Raudsepp T, Frönicke L, Scherthan H (1998) Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. Genome Res 8(6):577-89.

Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J (2010) Rodent evolution: back to the root. Mol Biol Evol 27(6):1315-26.

Cimini D, Fioravanti D, Salmon ED, Degrassi F (2002) Merotelic kinetochore orientation versus chromosome mono-orientation in the origin of lagging chromosomes in human primary cells. J Cell Sci 115:507–515.

Cimini D, Tanzarella C, Degrassi F (1999) Differences in malsegregation rates obtained by scoring ana-telophases or binucleate cells. Mutagenesis 14:563–568.

Claussen U (2005) Chromosomics. Cytogenet Genome Res 111:101-106.

Clevers H (2011) The cancer stem cell: premises, promises and challenges. Nat Med 17(3):313-9.

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L (2005) Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet 21: 673–682.

Contreras LC, Torresmura JC, Spotorno AE (1990) The Largest Known Chromosome-Number for a Mammal, in a South-American Desert Rodent. Experientia 46:506-508.

Csink AK, Henikoff S (1998) Something from nothing: the evolution and utility of satellite repeats. Trends Genet 14:200–204.

Dahl C, Grønbæk K, Guldberg P (2011) Advances in DNA methylation: 5-hydroxymethylcytosine revisited. Clin Chim Acta 412(11-12):831-6.

Darwin C (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. London: John Murray. 488p.

Dawe RK, Henikoff S (2006) Centromeres put epigenetics in the driver's seat. Trends Biochem Sci 31:662–669.

Dimitri P, Caizzi R, Giordano E, Carmela Accardo M, Lattanzi G, Biamonti G (2009) Constitutive heterochromatin: a surprising variety of expressed sequences. Chromosoma 118(4):419-35.

Dimitri P, Corradini N, Rossi F, Vernì F (2005) The paradox of functional heterochromatin. Bioessays 27(1):29-41.

Dover GA (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117.

Dover GA (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 2:159–165.

Duesberg P (2007) Chromosomal chaos and cancer. Sci Am. 296(5):52-9.

Duesberg P, Fabarius A, Hehlmann R (2004) Aneuploidy, the primary cause of the multilateral genomic instability of neoplastic and preneoplastic cells. IUBMB Life 56: 65–81.

Dutrillaux B, Viegas-Pequignot E, Couturier J (1980) Great homology of chromosome banding of the rabbit (Oryctolagus cuniculus) and primates, including man. Ann Genet 23:22–25.

Ehrlich M (2009) DNA hypomethylation in cancer cells. Epigenomics 1(2):239-59.

Ekwall K (2007) Epigenetic control of centromere behavior. Annu Rev Genet 41:63-81.

Elder JF, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev Biol 70:297–320.

Ellingsen A, Slamovits CH, Rossi MS (2007) Sequence evolution of the major satellite DNA of the genus Ctenomys (Octodontidae, Rodentia). Gene 392(1-2):283-90.

ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489(7414):57-74.

Engel SR, Hogan KM, Taylor JF e Davis SK (1998) Molecular Systematicas and Paleobiogeography of the South American Sigmodontine Rodents. Mol Biol Evol 15(1): 35-49.

Engelbrecht A, Dobigny G, Robinson TJ (2006) Further insights into the ancestral murine karyotype: the contribution of the Otomys-Mus comparison using chromosome painting. Cytogenet Genome Res 112(1-2):126-30.

Epstein LM, Gall JG (1987) Self-cleaving transcripts of a satellite DNA in a newt. Cell 48: 535–543 Ferbeyre G, Smith JM, Cedergren R (1998) Schistosome satellite DNA encodes active hammerhead-ribozymes. Mol Cell Biol 18: 3880–3888.

Ferguson-Smith MA (1997) Genetic analysis by chromosome sorting and painting: phylogenetic and diagnostic applications. Eur J Hum Genet 5(5):253-65.

Ferguson-Smith MA, Trifonov V (2007) Mammalian karyotype evolution. Nat Rev Genet 8(12):950-62.

Ferguson-Smith MA, Yang F, O'Brien PC (1998) Comparative Mapping Using Chromosome Sorting and Painting. ILAR J 39(2-3):68-76.

Ferree PM, Prasad S (2012) How can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. Genet Res Int 2012:430136.

Finnegan D J (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103–107.

Flynn LJ, Jacobs LL, Lindsay EH (1985) Problems in muroid phylogeny: relationship to other rodents and origin of major groups. In: Luckett WP, Hartenberger J-L editors. Evolutionary Relationships Among Rodents. New York: Plenum Press. pp 589– 616.

Friedberg EC (2003) DNA damage and repair. Nature 421:436–440.

Froenicke L (2005) Origins of primate chromosomes - as delineated by Zoo-FISH and alignments of human and mouse draft genome sequences. Cytogenet Genome Res 108(1-3):122-38.

Froenicke L, Caldés MG, Graphodatsky A, Müller S, Lyons LA, Robinson TJ, Volleth M, Yang F, Wienberg J (2006)Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? Genome Res 16(3):306-10.

Froenicke L, Lyons AL (2008) Hotspots of mammalian chromosome evolution. Chichester: John Wiley & Sons.

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

Fujiwara T, Bandi M, Nitta M, Ivanova EV, Bronson RT, Pellman D(2005) Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. Nature 437:1043–7.

Furano AV (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog Nucleic Acid Res Mol Biol 64:255-294.

Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, et al. (1996) 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. Proc Natl Acad Sci U S A 93(14):7081-4.

Ganley ARD, Kobayashi T (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by wholegenome shotgun sequence data. Genome Res 17:184-191.

Garagna S, Marziliano N, Zuccotti M, Searle JB, Capanna E, Redi CA (2001) Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. Proc Natl Acad Sci USA 98(1):171-5.

Geigl JB, Obenauf AC, Schwarzbraun T, Speicher MR (2008) Defining 'chromosomal instability'. Trends Genet 24(2):64-9.

Gent DC, Hoeijmakers JH, Kanaar R (2001) Chromosomal stability and the DNA double-stranded break connection. Nat Rev Genet 2(3):196-206.

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ et al. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428(6982):493-521.

Gindullis F, Desel C, Galasso I, Schmidt T (2001) The large-scale organization of the centromeric region in beta species. Genome Res 11:253–265.

Goldenberg MM (1999) Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. Clin Ther 21:309–318.

Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74:481–514.

Gonçalves dos Santos Silva A. Sakar R, Harizanova J, Guffei A, Mowat M, Garini Y Mai S (2008) Centromeres in cell division, evolution, nuclear organization and disease. Journal of Cellular Biochemistry 104:2040-2058.

Goureau A, Garrigues A, Tosser-Klopp G, Lahbib-Mansais Y, Chardon P, Yerle M (2001) Conserved synteny and gene order difference between human chromosome 12 and pig chromosome 5. Cytogenet Cell Genet 94(1-2):49-54.

Graphodatsky AS (2007) Comparative Chromosomics. Mol Biol 41(3): 361–375.

Graphodatsky AS, Trifonov VA, Stanyon R (2011) The genome diversity and karyotype evolution of mammals. Mol Cytogenet 4:22.

Graphodatsky AS, Yang F, O'Brien PC, Serdukova N, Milne BS, Trifonov V, Ferguson-Smith MA (2000) A comparative chromosome map of the Arctic fox, red fox and dog defined by chromosome painting and high resolution G-banding. Chromosome Res 8(3):253-63.

Greaves M (2010) Cancer stem cells: back to Darwin? Semin Cancer Biol 20(2):65-70.

Greaves M, Maley CC (2012) Clonal evolution in cancer. Nature 481(7381):306-13.

Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD (2007) Eukaryotic genome size databases. Nucleic Acids Res 35:D332-D338.

Grewal SI, Elgin SC (2007) Transcription and RNA interference in the formation of heterochromatin. Nature 447:399–406.

Grewal SI, Jia S (2007) Heterochromatin revisited. Nat Rev Genet 8(1):35-46.

Guigo R, Dermitzakis ET, Agarwal P, Ponting CP, Parra G, Reymond A, Abril JF, Keibler E, Lyle R, Ucla C, Antonarakis SE, Brent MR (2003) Comparison of mouse and human genomes followed by experimental verification yields an estimated 1,019 additional genes. Proc Natl Acad Sci USA 100: 1140–1145.

Hall SE, Luo S, Hall AE, Preuss D (2005) Differential rates of local and global homogenization in centromere satellites from Arabidopsis relatives. Genetics 170(4):1913-27.

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.

Hanahan D, Weinberg RA (2000) The Hallmarks of Cancer Review. Cell 100: 57–70.

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144(5):646-74.

Hanks S, Coleman K, Reid S, Plaja A, Firth H, Fitzpatrick D, Kidd A, Méhes K, Nash R, Robin N, Shannon N, Tolmie J, Swansbury J, Irrthum A, Douglas J, Rahman N (2004) Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. Nat Genet 36(11):1159-61.

Harari D, Yarden Y (2000) Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. Oncogene 19(53):6102-14.

Hartenberger JL (1998) Description de la radiation des Rodentia (Mammalia) du Paléocène supérieur au Miocène; incidences phylogénétiques. C. R. Acad. Sci. Paris Sciences de la terre et des planètes 326: 439–444.

Hartmann N, Scherthan H (2004) Characterization of ancestral chromosome fusion points in the Indian muntjac deer. Chromosoma 112: 213–220.

Heilig CE, Löffler H, Mahlknecht U, Janssen JW, Ho AD, Jauch A, Krämer A (2010) Chromosomal instability correlates with poor outcome in patients with myelodysplastic syndromes irrespectively of the cytogenetic risk group. J Cell Mol Med 14(4):895-902.

Heitz E (1928) Das Heterochromatin der Moose. Jahrb Wiss Botanik 69: 762–818.

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293: 1098–1102.

Hir J (1997) A Comparative study on the dental morphology of the Early Pleistocene Cricetus praeglacialis SCHAUB, 1930 and recent Hungarian Cricetus cricetus L. Folia Historico Naturalia Musei Matraensis 22: 51-79.

Hirai H, Taguchi T, Godwin AK (1999) Genomic differentiation of 18S ribosomal DNA and b-satellite DNA in the hominoid and its evolutionary aspects. Chromosome Res 7:531-540.

Holland AJ, Cleveland DW (2009) Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. Nat Rev Mol Cell Biol 10: 478–487.

Huchon D, Chevret P, Jordan U, Kilpatrick CW, Ranwez V, Jenkins PD, Brosius J, Schmitz J (2007) Multiple molecular evidences for a living mammalian fossil. Proc Natl Acad Sci USA 104(18):7495-7499.

Huges SE, Hawle RS (2009) Heterochromatin: A rapidly evolving species barrier. PLoS Biol 7(10):e1000233.

Hung MC, Lau YK (1999) Basic science of HER-2/neu: a review. Semin Oncol 26(4 Suppl 12):51-9.

Hynes NE (2007) Targeting ERBB receptors in cancer. Recent Results Cancer Res 172:45-57.

Issa JP (2007) DNA methylation as a therapeutic target in cancer. Clin Cancer Res 13(6):1634-7.

Jallepalli PV, Waizenegger IC, Bunz F, Langer S, Speicher MR, Peters JM, Kinzler KW, Vogelstein B, Lengauer C (2001) Securin is required for chromosomal stability in human cells. Cell 105(4):445-57.

Jansa SA, Weksler M (2004) Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. Mol Phylogenet Evol 31(1):256-76.

John, B (1988) The biology of heterochromatin. In: Verma, RS, editors. Heterochromatin: Molecular and structural Aspects. Cambridge: Cambridge University Press. pp. 1–128.

Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128(4):683-92.

Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet 19:187–191.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman FM, Pinkel D (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258: 818–821.

Kaludov NK, Wolffe AP (2000) MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. Nucleic Acids Res 28: 1921–1928.

Karlsson A, Helou K, Walentinsson A, Hedrich HJ, Szpirer C, Levan G (2001) Amplification of Mycn, Ddx1, Rrm2, and Odc1 in rat uterine endometrial carcinomas. Genes Chromosomes Cancer 31(4):345-56.

Kelly TK, De Carvalho DD, Jones PA (2010) Epigenetic modifications as therapeutic targets. Nat Biotechnol 28:1069–78.

Khodjakov A, Cole RW, McEwen BF, Buttle KF, Rieder CL (1997) Chromosome fragments possessing only one kinetochore can congress to the spindle equator. J Cell Biol 136:229–240.

King M (1993) Species evolution. The role of chromosome change. Cambridge: Cambridge University Press. 360 p.

Knudson AG (2002) Cancer genetics. Am J Med Genet 111:96–102.

Kolomietz E, Meyn MS, Pandita A, Squire JA (2002) The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. Genes Chromosomes Cancer 35(2):97-112.

Kouzarides T (2007) Chromatin modifications and their function. Cell 128: 693–705.

Kowalski K (2001) Pleistocene Rodents of Europe. Folia Quaternaria 72: 1–389.

Kunze B, Traut W, Garagna S, Weichenhan D, Redi CA, Winking H (1999) Pericentric satellite DNA and molecular phylogeny in Acomys (Rodentia) Chromosome Res 7(2):131-41.

Kurzrock R, Kantarjian HM, Druker BJ, Talpaz M (2003) Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. Ann Intern Med 138(10):819-30.

Kuznetsova IS, Prusov AN, Enukashvily NI, Podgornaya OI (2005) New type of the mouse centromeric satellite DNA. Chromosome Res 13:9–25.

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

Lecompte E, Denys C, Granjon L (2005) Confrontation of morphological and molecular data: The Praomys group (Rodentia, Murinae) as a case of adaptive convergences and morphological stasis. Mol Phylogenet Evol 37(3):899-919.

Lee HR, Neumann P, Macas J, Jiang J (2006) Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. Mol Biol Evol 23: 2505–2520.

Lee JS, Chu IS, Mikaelyan A, Calvisi DF, Heo J, Reddy JK, Thorgeirsson SS (2004) Application of comparative functional genomics to identify best-fit mouse models to study human cancer. Nat Genet 36(12):1306-11.

Lemskaya NA, Romanenko SA, Golenishchev FN, Rubtsova NV, Sablina OV, Serdukova NA, O'Brien PC, Fu B, Yiğit N, Ferguson-Smith MA, Yang F, Graphodatsky AS (2010) Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). III. Karyotype relationships of ten Microtus species. Chromosome Res 18(4):459-71.

Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. Nature 396(6712):643-9.

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4(3):203-21.

Li T, O'Brien PC, Biltueva L, Fu B, Wang J, Nie W, Ferguson-Smith MA, Graphodatsky AS, Yang F (2004) Evolution of genome organizations of squirrels (Sciuridae) revealed by cross-species chromosome painting. Chromosome Res 12(4):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(3-4):270-6.

Li WM (1997) Molecular Evolution. Sunderland: Sinauer Associates. 432 p.

López-Flores I, Ruiz-Rejón C, Cross I, Rebordinos L, Robles F, Navajas-Pérez R, de la Herrán R (2010) Molecular characterization and evolution of an interspersed repetitive DNA family of oysters. Genetica 138(11-12):1211-9.

Louzada S, Paço 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(8):1149-55.

Luckett WP, Hartenberger J-L (1993) Monophyly or polyphyly of the order Rodentia: possible conflict between morphological and molecular interpretations. J Mammal Evol 1(2):127-147.

Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromerespecific histone in Drosophila. Genetics 157:1293–1298.

Manuelidis L (1978) Chromosomal localization of complex and simple repeated human DNAs. Chromosoma 66(1):23-32.

Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, Hillier LW, Jiang Z, Baker C, Malfavon-Borja R, Fulton LA, Alkan C, Aksay G, Girirajan S, Siswara P, Chen L, Cardone MF, Navarro A, Mardis ER, Wilson RK, Eichler EE (2009) A burst of segmental duplications in the genome of the African great ape ancestor. Nature 457(7231):877-881.

Marshall OJ, Choo KH (2011) Putative CENP-B paralogues are not present at mammalian centromeres. Chromosoma. 121(2):169-79.

Martienssen RA (2003) Maintenance of heterochromatin by RNA interference of tandem repeats. Nat. Genet. 35:213–214.

Martin S A, Hewish M, Lord CJ, Ashworth A (2010) Genomic instability and the selection of treatments for cancer. J Pathol 220: 281–289.

Meles S, Adega F, Guedes-Pinto H, Chaves R (2008) The karyotype and sex chromosomes of Praomys tullbergi (Muridae, Rodentia): a detailed characterization. Micron 39 (5): 559–568.

Meštrović N, Plohl M, Mravinac B, Ugarković Đ (1998) Evolution of satellite DNAs from the genus Palorus—experimental evidence for the library hypothesis. Mol Biol Evol 15:1062–106.

Michaux J, Reyes A, Catzeflis F (2001) Evolutionary History of the Most Speciose Mammals: molecular Phylogeny of Muroid Rodents. Mol Biol Evol 18(11): 2017-2031.

Michel LS, Liberal V, Chatterjee A, Kirchwegger R, Pasche B, Gerald W, Dobles M, Sorger PK, Murty VV, Benezra R (2001) MAD2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells. Nature 409(6818):355-9.

Mitelman F, Johansson B, Mertens F (2007) The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 7(4):233-45.

Mlynarski EE, Obergfell CJ, O'Neill MJ, O'Neill RJ (2010) Divergent patterns of breakpoint reuse in Muroid rodents. Mamm Genome 21(1-2):77-87.

Montgelard C, Forty E, Arnal V, Matthee CA (2008) Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. BMC Evol Biol 8:321.

Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF *et al.* (2202) Initial sequencing and comparative analysis of the mouse genome. Nature 420(6915):520-62.

Mravinac B, Plohl M (2010) Parallelism in evolution of highly repetitive DNAs in sibling species. Mol Biol Evol 27: 1857–1867.

Mravinac B, Plohl M, Mestrović N, Ugarković D (2002) Sequence of PRAT satellite DNA "frozen" in some Coleopteran species. J Mol Evol 54(6):774-83.

Müller S, Hollatz M, Wienberg J (2003) Chromosomal phylogeny and evolution of gibbons (Hylobatidae). Hum Genet 113(6):493-501.

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'brien SJ, Pevzner PA, Lewin HA (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309:613–617.

Murphy WJ, Pevzner PA, O'Brien SJ (2004) Mammalian phylogenomics comes of age. Trends Genet 20(12): 631-639.

Musser GG, Carleton M D (2005) Superfamily Muroidea. In: Wilson DE, Reeder DM, editors. Mammal species of the world: a taxonomic and geographic reference. Baltimore: Johns Hopkins University Press. pp 894–1531.

Nadeau JH, Taylor BA (1984) Lengths of chromosomal segments conserved since divergence of man and mouse. Proc Natl Acad Sci USA, 81:814–818.

Nash WG, O'Brien SJ (1982) Conserved regions of homologous G-banded chromosomes between orders in mammalian evolution: carnivores and primates. Proc Natl Acad Sci USA 79:6631–6635.

Nechay G (2000) Status of Hamsters Cricetus cricetus, Cricetus migratorius, Mesocricetus newtoni and other hamster species in Europe. Nature and Environment Series 106.

Neumann K, Michaux J, Lebedev V, Yigit N, Colak E, Ivanova N, Poltoraus A, Surov A, Markov G, Maak S, Neumann S e Gattermann R (2006) Molecular phylogeny of the Cricetinae subfamily based on the mitochondrial cytochrome b and 12S rRNA genes and the nuclear vWF gene. Molecular Phylogenetics and Evolution 39: 135–148.

Nicolas V, Verheyen E, Verheyen W, Hulselmans J, Dillen M , Akpatou B, Dudu A, Wendelen W, Colyn M (2005) Systematics of African lowland rainforest Praomys (Rodentia, Muridae) based on molecular and craniometrical data. Zool J Linn Soc 145(4) 539–553.

Nie W, Wang J, Su W, Wang D, Tanomtong A, Perelman PL, Graphodatsky AS, Yang F (2012) Chromosomal rearrangements and karyotype evolution in carnivores revealed by chromosome painting. Heredity 108(1):17-27.

Nijman IJ and Lenstra JA (2001) Mutation and recombination in cattle satellite DNA: A feedback model for the evolution of satellite DNA repeats. J Mol Evol 52(4):361-71.

Nilsson S, Helou K, Walentinsson A, Szpirer C, Nerman O, Ståhl F (2001)Rat-mouse and rat-human comparative maps based on gene homology and high-resolution zoo-FISH. Genomics 74(3):287-98.

Nowell PC (1976) The clonal evolution of tumor cell populations. Science 194: 23–28.

O'Brien SJ, Menotti-Raymond M, Murphy WJ, Nash WG, Wienberg J, Stanyon R, Copeland NG, Jenkins NA, Womack JE, Graves JA (1999) The promise of comparative genomics in mammals. Science 286:458–481.

Ohno S (1972) So much "junk" DNA in our genome. Brookhaven Symp. Biol 23:366–370.

Ohno S (1973) Ancient linkage groups and frozen accidents. Nature 244:259-262.

Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. Nature 284: 604–607.

Ostromyshenskiĭ DI, Kuznetsova IS, Golinishchev FN, Malikov VG, Podgornaia OI (2011) Satellite DNA as a phylogenetic marker: case study of three genera of the Murinae subfamily. Tsitologiia 53(7):564-71.

Paço A, Adega F, Guedes-Pinto H, Chaves R (2009) Hidden heterochromatin: Characterization in the Rodentia species Cricetus cricetus, Peromyscus eremicus (Cricetidae) and Praomys tullbergi (Muridae). Genet Mol Biol 32(1):58-68.

Pathak D, Srivastava J, Premi S, Tiwari M, Garg LC, Kumar S, Ali S (2006) Chromosomal localization, copy number assessment, and transcriptional status of BamHI repeat fractions in water buffalo *Bubalus bubalis*. DNA Cell Biol 25:206–214.

Pathak S, Arrighi FE (1973) Loss of DNA following Cbanding procedures. Cytogenet Cell Genet 12:414-422.

Paulson TG, Almasan A, Brody LL, Wahl GM (1998) Gene amplification in a p53-deficient cell line requires cell cycle progression under conditions that generate DNA breakage. Mol Cell Biol 18:3089–3100.

Peng JC, Karpen GH (2008) Epigenetic regulation of heterochromatic DNA stability. Curr Opin Genet Dev 18(2):204-11.

Peng Q, Pevzner PA, Tesler G (2006) The fragile breakage versus random breakage models of chromosome evolution. PLoS Comput Biol 2:e14.

Perelman PL, Graphodatsky AS, Serdukova NA, Nie W, Alkalaeva EZ, Fu B, Robinson TJ, Yang F (2005) Karyotypic conservatism in the suborder Feliformia (Order Carnivora). Cytogenet Genome Res 108(4):348-54.

Pévet P, Bujis R, Masson-Pévet M (1987) Effect of pinealectomy and constant high level of circulating melatonin or of 5-methoxytryptamine on vasopressinergic innervation in the brain of the European hamster (Cricetus cricetus, L). J Neural Transm 70:287-294.

Pevzner P, Tesler G (2003a) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc Natl Acad Sci USA 100:7672-7677.

Pevzner P, Tesler G (2003b) Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. Genome Res 13(1):37-45.

Pipiras E, Coquelle A, Bieth A, Debatisse M (1998) Interstitial deletions and intrachromosomal amplification initiated from a double-strand break targeted to a mammalian chromosome. EMBO J 17:325–333.

Plohl M (2010) Those mysterious sequences of satellite DNAs. Periodicum Biologorum UDC 112(4):403–410.

Plohl M, Petrović V, Luchetti A, Ricci A, Satović E, Passamonti M, Mantovani B (2010) Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. Heredity 104(6):543-51.

Plohl M, Luchetti A, Mestrović N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. Gene 409(1-2):72-82.

Reece R (2004) Analysis of genes and genomes. England: John Wiley and Sons. 490p.

Reppenning CA (1980) Faunal exchanges between Siberia and North America. Canadian J Anthropol 1:37-44.

Reznik G (1977) Experimental carcinogenesis in the respiratory tract using the European field hamster (Cricetus cricetus L.) as a model. Fortschr Med. 5(43):2627-31.

Richardson C, Jasin M (2000) Frequent chromosomal translocations induced by DNA double-strand breaks. Nature 405:697–700.

Ried T, Schröck E, Ning Y, Wienberg J (1998) Chromosome painting: a useful art. Hum Mol Genet 7(10):1619-26.

Riggins GJ, Lokey LK, Chastian JL, Leiner HA, Sherman SL, Wilkinson KD, Warren ST (1992) Human genes containing polymorphic trinucleotide repeats. Nat Genet 2:186–191.

Robinson M, Catzeflis F, Briolay J e Mouchiroud D (1997) Molecular phylogeny of rodents, with special emphasis on murids: Evidence from nuclear gene LCAT. Mol Phylogenet Evol 8: 423-434.

Robinson TJ, Ruiz-Herrera A (2008) Defining the ancestral eutherian karyotype: a cladistic interpretation of chromosome painting and genome sequence assembly data. Chromosome Res 16(8):1133-41.

Robinson TJ, Ruiz-Herrera A, Froenicke L (2006) Dissecting the mammalian genome-new insights into chromosomal evolution. Trends Genet 22(6):297-301.

Robinson TJ, Yang F (2012) Molecular cytogenetics: karyotype evolution, phylogenomics and future prospects. Heredity 108(1):1-3.

Roizès G (2006) Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. Nucleic Acids Res 34(6):1912-24.

Rojas A, Vazquez-Tello A, Ferbeyre G, Venanzetti F, Bachmann L, Paquin B, Sbordoni V, Cedergren R (2000) Hammerhead-mediated processing of satellite pDo500 family transcripts from Dolichopoda cave crickets. Nucleic Acids Res 28: 4037–4043.

Rokas A, Holland PW (2000) Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol 15(11):454-459.

Romanenko SA, Perelman PL, Serdukova NA, Trifonov VA, Biltueva LS, Wang J, Li T, Nie W, O'Brien PC, Volobouev VT, Stanyon R, Ferguson-Smith MA, Yang F, Graphodatsky AS (2006) Reciprocal chromosome painting between three laboratory rodent species.Rodents Mamm Genome 17(12):1183-1192.

Romanenko SA, Perelman PL, Trifonov VA, Graphodatsky AS (2012) Chromosomal evolution in Rodentia. Heredity (Edinb) 108(1):4-16.

Romanenko SA, Volobouev VT, Perelman PL, Lebedev VS, Serdukova NA, Trifonov VA, Biltueva LS, Nie W, O'Brien PC, Bulatova NSh, Ferguson-Smith MA, Yang F, Graphodatsky AS (2007) Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroidea, Rodentia) inferred from chromosomal painting and banding comparison. Chromosome Res 15:283-297.

Rossi MS, Reig OA, Zorzópulos J (1990) Evidence for rolling-circle replication in a major satellite DNA from the South American rodents of the genus Ctenomys. Mol Biol Evol 7(4):340-50.

Royle NJ, Clarkson RE, Wong Z, Jeffreys AJ (1988) Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics 3:352–360.

Rudert F, Bronner S, Garnier JM, Dollé P (1995) Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mamm Genome 6:76–83.

Ruiz-Herrera A, Castresana J, Robinson TJ (2006) Is mammalian chromosomal evolution driven by regions of genome fragility? Genome Biol 7(12):R115.

Sankoff D (2009) The where and wherefore of evolutionary breakpoints. J Biol 8(7):66.

Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS (2010) A census of amplified and overexpressed human cancer genes. Nat Rev Cancer 10(1):59-64.

Santarosa M, Ashworth A (2004) Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. Biochim Biophys Acta 1654:105–122.

Santos S, Chaves R, Adega F, Bastos E, Guedes-Pinto H (2006) Amplification of the major satellite DNA family (FA-SAT) in a cat fibrosarcoma might be related to chromosomal instability. J Hered 97(2):114-8.

Savelyeva L, Schwab M (2001) Amplification of oncogenes revisited: from expression profiling to clinical application. Cancer Lett 167(2):115-23.

Schechter AL, Stern DF, Vaidyanathan L, Decker SJ, Drebin JA, Greene MI, Weinberg RA (1984) The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. Nature 312: 513-516.

Scherthan H, Cremer T, Arnason U, Weier HU, Lima-de-Faria A, Fronicke L (1994) Comparative chromosome painting discloses homologous segments in distantly related mammals. Nat Genet 6:342-347.

Scherthan H, Weich S, Schwegler H, Heyting C, Härle M, Cremer T (1996) Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing. J Cell Biol 134:1109-1125.

Schimenti JC (1999) Mice and the role of unequal recombination in gene-family evolution. Am J Hum Genet 64(1):40-5.

Schröck E, Veldman T, Padilla-Nash H, Ning Y, Spurbeck J, Jalal S, Shaffer LG, Papenhausen P, Kozma C, Phelan MC, Kjeldsen E, Schonberg SA, O'Brien P, Biesecker L, du Manoir S, Ried T (1997) Spectral karyotyping refines cytogenetic diagnostics of constitutional chromosomal abnormalities. Hum Genet 101:255–62.

Schueler MG, Dunn JM, Bird CP, Ross MT, Viggiano L et al. (2005) Progressive proximal expansion of the primate X chromosome centromere. Proc Natl Acad Sci USA 102(30):10563-8.

Schueler MG, Sullivan BA (2006) Structural and functional dynamics of human centromeric chromatin. Annu Rev Genomics Hum Genet 7:301-13.

Schvartzman JM, Sotillo R, Benezra R (2010) Mitotic chromosomal instability and cancer: mouse modelling of the human disease. Nat Rev Cancer 10(2):102-15.

Schwa M (2004) MYCN in neuronal tumours. Cancer Lett 204(2):179-87.

Schwab M (1998) Amplification of oncogenes in human cancer cells. Bioessays 20(6):473-9.

Schwab M (1999) Oncogene amplification in solid tumors. Seminars in Cancer Biology 9(4):319-325.

Senegas F, Avery DM (1998) New evidence for the murine origins of the Otomyinae (Mammalia, Rodentia) and the age of Bolt's Farm (South Africa). S Afr J Sci 94:503–507.

Shackleton M, Quintana E, Fearon ER, Morrison SJ (2009) Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell 138(5):822-9.

Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, Humpton TJ, Brito IL, Hiraoka Y, Niwa O, Amon A (2011) Aneuploidy drives genomic instability in yeast. Science 333(6045):1026-30.

Shepelev VA, Alexandrov AA, Yurov YB, Alexandrov IA (2009) The evolutionary origin of man can be traced in the layers of defunct ancestral alpha satellites flanking the active centromeres of human chromosomes. PLoS Genet 5(9):e1000641.

Sherwood SW, Patton JL(1982) Chromosome evolution in pocket gophers (genus Thomomys). II.Variation in cellular DNA content. Chromosoma 85:163–179.

Shimada A, Murakami Y (2010) Dynamic regulation of heterochromatin function via phosphorylation of HP1-family proteins. Epigenetics 5(1):30-3.

Shore D (2001) Telomeric chromatin: Replicating and wrapping up chromosome ends. Curr Opin Genet Dev 11:189-198.

Sieber OM, Heinimann K, Tomlinson IP (2003) Genomic instability-The engine of tumorigenesis? Nat Rev Cancer 3: 701–708.

Silva MJ, Patton JL, Yonenaga-YassudaY (2006) Phylogenetic relationships and karyotype evolution in the sigmodontine rodent Akodon (2n = 10 and 2n = 16) from Brazil. Genetics and Molecular Biology 29(3):469-474.

Singal R, Wang SZ, Sargent T, Zhu SZ, Ginder GD (2002) Methylation of promoter proximal-transcribed sequences of an embryonic globin gene inhibits transcription in primary erythroid cells and promotes formation of a cell type-specific methyl cytosine binding complex. J Biol Chem 277: 1897-1905.

Sitnikova NA, Romanenko SA, O'Brien PC, Perelman PL, Fu B, Rubtsova NV, Serdukova NA, Golenishchev FN, Trifonov VA, Ferguson-Smith MA, Yang F, Graphodatsky AS (2007) Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). I. The genome homology of tundra vole, field vole, mouse and golden hamster revealed by comparative chromosome painting. Chromosome Res 15(4):447-56.

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235(4785):177-82.

Slamon DJ, Godolphin W, La Jones B, Holt JA, Wang SG, Keith DE (1989) Studies of the ERBB proto-oncogene in human breast and ovarian cancer. Science 244:707–712.

Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344:783-792.

Slamovits CH, Rossi M S (2002) Satellite DNA: agent of chromosomal evolution in mammals. A review. J. Neotrop. Mammal 9(2):297-308.

Slamovits CH, Cook JA, Lessa EP, Rossi MS (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(9):1708-19.

Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. Science 191:528-535.

Snape K, Hanks S, Ruark E, Barros-Núñez P, Elliott A, Murray A, Lane AH, Shannon N, Callier P, Chitayat D, Clayton-Smith J, Fitzpatrick DR, Gisselsson D, Jacquemont S, Asakura-Hay K, Micale MA, Tolmie J, Turnpenny PD, Wright M, Douglas J, Rahman N (2011) Mutations in CEP57 cause mosaic variegated aneuploidy syndrome. Nat Genet 43(6):527-9.

Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, Toretsky JA, Rosenberg SA, Shukla N, Ladanyi M, Samuels Y, James CD, Yu H, Kim JS, Waldman T (2011) Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science 333(6045):1039-43.

Speicher MR, Gwyn Ballard S, Ward DC (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nat Genet 12:368–75.

Stankiewicz P, Lupski JR (2002) Genome architecture, rearrangements and genomic disorders. Trends Genet 18(2):74-82.

Stanyon R, Stone G, Garcia M, Froenicke L (2003) Reciprocal chromosome painting shows that squirrels, unlike murid rodents, have a highly conserved genome organization. Genomics 82(2):245-9.

Stanyon R, Yang F, Morescalchi AM, Galleni L (2004) Chromosome painting in the long-tailed field mouse provides insights into the ancestral murid karyotype. Cytogenet Genome Res 105:406–411.

Stephan W (1989) Tandem-repetitive noncoding DNA: forms and forces. Mol Biol Evol 6:198–212.

Steppan S, Adkins R, Anderson J (2004) Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst Biol 53(4):533-53.

Steppan SJ, Adkins RM, Spinks PQ e Hale C (2005) Multigene phylogeny of the World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. Molecular Phylogenetics and Evolution 37: 370-388.

Stratton M R, Campbell P J, Futreal PA (2009) The cancer genome. Nature Rev 458:719-723.

Sullivan BA, Blower MD, Karpen GH (2001) Determining centromere identity: cyclical stories and forking paths. Nat Rev Genet 2:584–596.

Sumner AT (2003) Chromosomes, the karyotype and evolution. In: Chromosomes Organization and Function. United Kingdom: Blackwell Science. pp 194-205.

Swier VJ, Bradley RD, Rens W, Elder FFB, Baker RJ (2009) Patterns of Chromosomal Evolution in Sigmodon, Evidence from Whole Chromosome Paints. Cytogenet Genome Res 125:54–66.

Talbert PB, Bryson TD, Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. J Biol 3:18–69.

Thompson SL, Compton DA (2008) Examining the link between chromosomal instability and aneuploidy in human cells. J Cell Biol 180:665–672.

Thompson SL, Compton DA (2010) Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. J Cell Biol 188(3):369-81.

Thompson SL, Compton DA (2011) Chromosomes and cancer cells. Chromosome Res 19(3): 433–44.

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun X-W, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie J E, Shah RB, Pienta KJ, Rubin M A, Chinnaiyan AM (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310:644–648.

Tower J (2004) Developmental gene amplification and origin regulation. Annu Rev Genet 38:273–304.

Tysnes BB, Bjerkvig R (2007) Cancer initiation and progression: Involvement of stem cells and the microenvironment. Biochim Biophys Acta 1775(2):283-297.

Tysnes BB (2010) Tumor-initiating and -propagating cells: cells that we would like to identify and control. Neoplasia 12(7):506-15.

Ugarkovic D (2005)Functional elements residing within satellite DNAs. EMBO Rep 6(11):1035-9.

Ugarković D, Plohl M (2002) Variation in satellite DNA profiles – causes and effects. EMBO J 21(22): 5955-5959.

van Noesel MM, Versteeg R: Pediatric neuroblastomas: genetic and epigenetic 'Danse Macabre'. Gene 325: 1-15, 2004.

Ventura K, O'Brien PCM, Yonenaga-Yassuda Y, Ferguson-Smith MA 2009 Chromosome homologies of the highly rearranged karyotypes of four Akodon species (Rodentia, Cricetidae) resolved by reciprocal chromosome painting: the evolution of the lowest diploid number in rodents. Chromosome Res 17:1063–1078.

Ventura M, Antonacci F, Cardone MF, Stanyon R, D'Addabbo P, Cellamare A, Sprague LJ, Eichler EE, Archidiacono N, Rocchi M (2007) Evolutionary formation of new centromeres in macaque. Science 316, 243–246.

Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. Nat Med 10(8):789-799.

Volobouev V, Vogt N, Viegas-Péquignot E, Malfoy B, Dutrillaux B (1995) Characterization and chromosomal location of two repeated DNAs in three Gerbillus species. Chromosoma 104(4):252-9.

Walsh JB (1987) Persistence of tandem arrays: implications for satellite DNA and simple-sequence DNAs. Genetics 115:553-567.

Wang Q, Hirohashi Y, Furuuchi K, Zhao H, Liu Q, Zhang H, Murali R, Berezov A, Du X, Li B, Greene MI (2004) The centrosome in normal and transformed cells. DNA Cell Biol 23: 475–489.

Wasylishen AR, Penn LZ (2010) Myc: the beauty and the beast. Genes Cancer 1(6):532-41.

Weinberg J, Fronicke L, Stanyon R (2000) Insights into mammalian genome organization and evolution by molecular cytogenetics. In: Clark M, editors. Comparative Genomics. USA: Kluwer academic Publishers. pp 207-244.

Weiss WA, Aldape K, Mohapatra G, Feuerstein BG, Bishop JM (1997) Targeted expression of MYCN causes neuroblastoma in transgenic mice. EMBO J 16: 2985–2995.

Westermark UK, Wilhelm M, Frenzel A, Henriksson MA (2011) The MYCN oncogene and differentiation in neuroblastoma. Semin Cancer Biol 21(4):256-66.

Wichman HA, Payne CT, Ryder OA, Hamilton MJ, Maltbie M, Baker RJ (1991) Genomic distribution of heterochromatic sequences in equids: implications to rapid chromosomal evolution. J Hered 82(5):369-77.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973-82.

Wienberg J (2004) The evolution of eutherian chromosomes. Curr Opin Genet Dev 14 (6):657–666.

Wienberg J (2005) Fluorescence in situ hybridization to chromosomes as a tool to understand human and primate genome evolution. Cytogenet Genome Res 108(1-3):139-60.

Wienberg J, Jauch A, Stanyon R, Cremer T (1990) Molecular cytotaxonomy of primates by chromosomal in situ suppression hybridization. Genomics 8(2):347-50.

Wild L, Flanagan JM (2010) Genome-wide hypomethylation in cancer may be a passive consequence of transformation. Biochim Biophys Acta 1806(1):50-7.

Wong LH, Brettingham-Moore KH, Chan L, Quach JM, Anderson MA, Northrop EL, Hannan R, Saffery R, Shaw ML, Williams E, Choo KH (2007) Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res 17(8):1146-60.

Wurster DH, Benirschke K (1970) Indian Muntjac, Muntiacus muntjak: a deer with a low diploid number. Science 168, 1364-1366.

Wyatt, G. R. (1950). Occurrence of 5-methylcytosine in nucleic acids. Nature 166:237–238.

Yang F, O'Brien PC, Milne BS, Graphodatsky AS, Solanky N, Trifonov V, Rens W, Sargan D, Ferguson-Smith MA (1999)A complete comparative chromosome map for the dog, red fox, and human and its integration with canine genetic maps. Genomics 62(2):189-202.

Yang W (2008) Structure and mechanism for DNA lesion recognition. Cell Res 18(1):184-97.

Ye CJ, Liu G, Bremer SW, Heng HH (2007) The dynamics of cancer chromosomes and genomes. Cytogenet Genome Res 118(2-4):237-46.

Ye CJ, Stevens JB, Liu G, Bremer SW, Jaiswal AS, Ye KJ, Lin MF, Lawrenson L, Lancaster WD, Kurkinen M, Liao JD, Gairola CG, Shekhar MP, Narayan S, Miller FR, Heng HH (2009) Genome based cell population heterogeneity promotes tumorigenicity: the evolutionary mechanism of cancer. J Cell Physiol 219(2):288-300.

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13: 335–340.

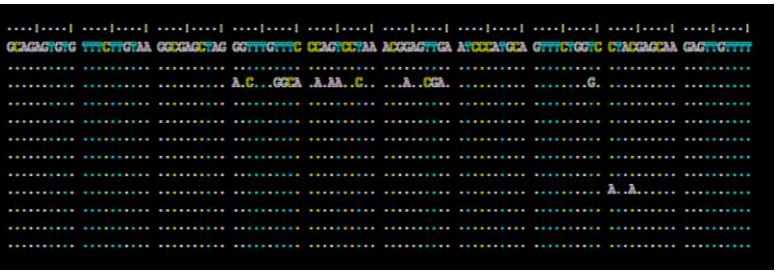
Yu Y, Bradley A (2001) Engineering chromosomal rearrangements in mice. Nature Rev. Genet. 2:780–790.

Zhao S, Shetty J, Hou L, Delcher A, Zhu B, Osoegawa K, de Jong P, Nierman WC, Strausberg RL, Fraser CM (2004) Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. Genome Res 14:1851-1860.

Zimmerman KA, Yancopoulos GD, Collum RG, Smith RK, Kohl NE, Denis KA, Nau MM, Witte ON, Toran-Allerand D, Gee CEMinna JD, Alt FW (1986) Differential expression of myc family genes during murine development. Nature 319: 780–783.

Zindy F, Knoepfler PS, Xie S, Sherr CJ, Eisenman RN, Roussel MF (2006) N-Myc and the cyclin-dependent kinase inhibitors p18Ink4c and p27Kip1 coordinately regulate cerebellar development. Proc Natl Acad Sci USA 103: 11579–11583.

CHAPTER II |



RESULTS AND DISCUSSION

1. COMPARATIVE CHROMOSOMICS

Rodents constitute a very abundant and diverse order within mammalian, characterized by an impressive radiation leading to the high number of species observed today. Once some rodents are used extensively in biomedical research, this has stimulated interest in the study of this group. High record rates of karyotype evolution are found in muroid rodents, making them the perfect organisms for studying chromosome evolution.

Cross-species chromosome painting is currently the method of choice for comparative cytogenetic studies in rodents, used in the construction of a high number of comparative chromosome maps. The intense focus on building comparative maps was due to the fact that these maps provide an unprecedented opportunity to use multispecies analysis as a tool for inferring karyotype evolution.

In this section it will be presented two papers presenting the construction of comparative maps for three Muroid species, *Praomys tullbergi* (PTU, Muridae), *Cricetus cricetus* (CCR) and *Peromyscus eremicus* (PER) (Cricetidae). These studies were developed with the objective of investigate the number and nature of the rearrangements that modelled these species' karyotype evolution. The use of *Mus musculus* (MMU) and *Rattus norvegicus* (RNO) paint probes permitted the identification of the syntenic associations for each species and the comparison with syntenic associations identified in previous studies for other species. This analysis culminated in the determination of the related clades, the suggestion of an Ancestral Muroidea Karyotype (AMK) and the identification of the major rearrangements occurred along evolution of these species. An analysis of the co-localization of constitutive heterochromatin (CH) with the breakpoint sites was also performed. As major outcomes can be pointed: 1) the elaboration of high resolution comparative maps (MMU, RNO and HSA syntenies) for the three species; the detection of intrachromosomal rearrangements due to the combined use of two sets of paint probes; 2) PER possesses a very conserved genome while PTU and CCR present more derivative karyotypes; 3) a high percentage of the identified breakpoints co-localize with CH.

1.1 *Praomys tullbergi* (Muridae, Rodentia) genome architecture decoded by comparative chromosome painting with *Mus* and *Rattus*

Chromosome Res (2012) 20:673–683
DOI 10.1007/s10577-012-9304-1

Praomys tullbergi (Muridae, Rodentia) genome architecture decoded by comparative chromosome painting with *Mus* and *Rattus*

Raquel Chaves · Sandra Louzada · Susana Meles · Johannes Wienberg · Filomena Adega

Received: 11 May 2012 / Revised: 28 June 2012 / Accepted: 29 June 2012 / Published online: 31 July 2012

Abstract The order Rodentia and in particular the Muridae are characterised by extremely high rates of chromosome evolution and remarkable chromosome diversity. The *Praomys* group (Murinae, Muridae and Rodentia) constitutes a diverse and abundant group divided into two complexes, the *jacksoni* complex and the *tullbergi* complex which includes the species *Praomys tullbergi*. Comparative chromosome painting using the two index genomes, *Mus musculus* and *Rattus norvegicus*, was performed resulting in a high resolution chromosome map for *P. tullbergi*. The combined use of rat and mouse probes and the assistance of the assembly of all the available sequencing data from Ensembl genome browser allowed a great dissection of *P. tullbergi* genome, the detection of inversion events and ultimately the refinement of *P. tullbergi* comparative map. A key achievement was the reconstruction of a high precision Muroidea ancestral karyotype (Muridae/Cricetidae and Murine) based in a broad species analysis combining previous reported comparative maps together with the presented data. This permitted the reconstruction of the evolutionary history of chromosome changes since the ancestral Muroidea genome and enlightened the phylogenetic relationships with the related species mouse and rat. The analysis of constitutive heterochromatin and its co-localisation with the identified evolutionary breakpoints regions was performed suggesting the involvement of repetitive sequences in the chromosome rearrangements that originated the present *P. tullbergi* genome architecture.

Keywords Chromosome evolution · Chromosome map of *Praomys tullbergi* · evolutionary breakpoint regions · constitutive heterochromatin

Responsible Editor: Herbert Macgregor.

Electronic supplementary material The online version of this article (doi:10.1007/s10577-012-9304-1) contains supplementary material, which is available to authorized users.

R. Chaves (✉) · S. Louzada · S. Meles · F. Adega
Centre of Genomics and Biotechnology,
Institute for Biotechnology and Bioengineering,
University of Trás-os-Montes and Alto Douro
(IBB/CGB-UTAD),
Vila Real, Portugal
e-mail: rchaves@utad.pt

J. Wienberg
Chrombios GmbH,
Raubling, Germany

J. Wienberg
Department Biologie II,
Anthropologie und Humangenetik,
Ludwig-Maximilians-University,
Munich, Germany

Abbreviations

AMK	Ancestral Muroidea karyotype
CH	Constitutive heterochromatin
DOP-PCR	Degenerated oligonucleotide primed-PCR
FACS	Fluorescence-activated cell sorter
FITC	Fluorescein isothiocyanate
GTG	G-banding by trypsin using Giemsa
HSA	*Homo sapiens*
MMU	*Mus musculus*
PTU	*Praomys tullbergi*
RNO	*Rattus norvegicus*
TAMRA	Tetramethyl-6-carboxyrhodamine

Introduction

The *Praomys* group (Rodentia, Muridae and Murinae) is one of the most diverse and abundant of Old World rodents. This group consists of several genera, among them *Praomys*, which is divided into two complexes, the *jacksoni* complex and the *tullbergi* complex. *Praomys tullbergi* is one of the four species comprising the *tullbergi* complex (Lecompte et al. 2005; Nicolas et al. 2005).

The Muridae is the single most diverse family of living mammals. Extremely high rates of chromosome evolution and remarkable chromosome diversity (wide variation in diploid numbers, rearrangements, localisation and amount of heterochromatin, occurrence of supernumerary B chromosomes, etc.) are features of the order Rodentia and Muridae rodents in particular (rats, mice, voles, hamsters, jerboas and many related other species) (Graphodatsky et al. 2000; Murphy et al. 2001; Romanenko et al. 2012). This high karyotype diversity has made myomorph rodents difficult to study using simple comparative analyses based on chromosome banding alone. The high number of Muridae species hindered a more broad-spectrum analysis of this family and their subfamilies in a single study. For this reason, the different studies have focused on one or a few species in a single published report.

In 2008, Lecompte et al. estimated the timing of key cladogenic events for the African murine diversity (Lecompte et al. 2008). In this broad sampling analysis, the radiation of the '*Praomys* group', considered one of the most diverse lineages, seems to have occurred somewhat at 7.6 Ma. The phylogenetic relationships within Murinae and the exact distribution ranges of these taxa are unclear. Previous studies point to the genus *Apodemus* as the sister group of the *Praomys* group by the *cytb* phylogeny (Lecompte et al. 2002). Chevret et al. (1994) proposed the genera *Mus* and *Malacomys* in polytomy with the *Praomys* group, based on DNA/DNA hybridization experiments. In 2004, Jansa and Weksler suggested *Mus* as the sister group of the *Praomys* group based on the molecular analysis of the IRBP gene (Jansa and Weksler 2004). In 2005, analyses based on morphological characters and sequence data from a mitochondrial and a nuclear gene performed by Lecompte et al. (2005) did not reached a clear determination of the sister taxa of the *Praomys* group, but all the candidates belong to one of the Eurasian taxa, namely *Apodemus*, *Rattus* or *Mus*. According to these authors, the discrepancies between molecular and morphological results are probably a reflex of the numerous convergences, as well as variations in the rates of morphological evolution among lineages. More recently, in 2008, Lecompte et al., based on mitochondrial *cyt b* gene and two nuclear gene fragments (IRBP exon 1 and GHR), point the genus *Praomys* as being monophyletic, as in previous molecular and morphological analyses (Lecompte et al. 2005). In this phylogenetic analysis, the *Praomys* group lineage appears as a sister group of the *Mus* lineage, but only with moderate support (77 % BP; 0.69 PP).

Dobigny et al. (2002, 2003) suggested that in the case of African murids, karyotype differences provide some of the most reliable diagnostic criteria. However, cytogenetic information of the recognised *Praomys* species is restricted to G- and C-banding (Lyons et al. 1977; 1980; Lee and Martin 1980, Capanna and Corti 1982; Baker et al. 1988; Qumsiyeh et al. 1990; Meles et al. 2008). Also in 2008, Meles et al. characterised the *P. tullbergi* karyotype by G- and C-banding and also analysed the distribution of LINEs and telomeric sequences. This work revealed that *P. tullbergi* karyotype is completely composed of acrocentric autosomes and a submetacentric X. In various rodents, karyotypes of closely related species just differ by reshuffling of chromosomes through Robertsonian translocations (Gropp and Winking 1981; Savic and Nevo 1990; Nachman and Searle 1995; Gava and Freitas 2003; Britton-Davidian et al. 2005). Robertsonian

fusions and fissions are the main forces that changed the organisation of genomes in various other mammals. This was thoroughly analysed in chromosome painting studies for Ursidae (bears) (Nash et al. 1998), Canidae (dogs) (Nash et al. 2001; Nie et al. 2012), Bovidae (Robinson and Ropiquet 2011) and many other mammals. In *P. tullbergi*, however, compared to mouse and rat chromosomes, other changes that are not Robertsonian can be suggested from the banded chromosomes (Meles et al. 2008). Only for few chromosomes that homologous patterns can be found between mouse, rat and *Praomys*. Thus, there should have been further mechanisms that changed their genome organisation during the estimated 7–9 Ma of radiation.

In this work, we assembled a high resolution chromosome painting map for *P. tullbergi* compared with the two index genomes *Mus musculus* and *Rattus norvegicus* for which the genomic DNA sequence is available. The objective of the present work was to perform an improved characterization of *P. tullbergi* genome organisation and thus to reconstruct the evolutionary history of chromosome changes with the related species mouse and rat. We therefore applied comparative chromosome painting (Wienberg et al. 1990) that allows the generation of global comparative genome maps at a cytogenetic resolution of about 5 Mbp (Wienberg et al. 2000; Froenicke et al. 2003). These maps could then be further compared to the existing higher resolution maps based on sequence data (Synteny viewer application in the Ensembl database; http://www.ensembl.org/).

Previous studies which focused on chromosome evolution indicate that the genome repetitive fraction can be correlated with the dynamics of the chromosome restructuring (Garagna et al. 2001). Accordingly, constitutive heterochromatin regions have been pointed as 'hotspots' for structural chromosome rearrangements (John 1988; Chaves et al. 2004). In this work, all the data related to the constitutive heterochromatin (CH) characterization of *P. tullbergi* were assembled in the constructed comparative map. The obtained levels of co-localisation between CH and the identified evolutionary breakpoints suggest a potential role of repetitive sequences in the chromosome rearrangements that originated *P. tullbergi* genome architecture along evolution.

Materials and methods

Cell culture and chromosome preparation

Chromosomes from *P. tullbergi* (PTU) were prepared from a fibroblast cell culture using standard procedures. In general, for karyotyping, *Praomys* chromosomes were ordered following Meles et al. (2008), according to their relative size. Chromosome numbering for mouse (*M. musculus*, MMU) and rat (*R. norvegicus*, RNO) paint probes was the same as previously described (Stanyon et al. 1999).

Chromosome painting probes and probe labelling

The paints (mouse and rat chromosome-specific probes) for this study are commercially available probes from Chrombios GmbH and were kindly provided by Dr. A. Kofler and R. Kofler from the flow sorting facility at Chrombios and were obtained from flow sorting spleen cell chromosomes on a FACS Vantage (BD) as described before (Rabbitts et al. 1995). Chromosome-specific probes were made by DOP-PCR from the flow-sorted chromosomes by using 6MW (CCGACTCGAGNNNNNNATGTGG) (rat probes) and F/S (CGGACTCGAGNNNNNNTACACC) (mouse probes) PCR universal primers and amplification conditions as previously described (Rabbitts et al. 1995). In a secondary PCR, the same PCR primers were also used to label the chromosome paints with either digoxigenin–2′-deoxyuridine, 5′-triphosphate (dUTP) or biotin–dUTP (Roche Molecular Biochemicals).

Fluorescent in situ hybridization and image processing

In situ hybridization of rat and mouse painting probes to *P. tullbergi* chromosome preparations was done as previously described (Wienberg et al. 1997). After hybridization and washing of the slides, digoxigenin-11-dUTP- or biotin-16-dUTP- (Roche Molecular Biochemicals) labelled chromosome paints were detected with antidigoxigenin-5′ TAMRA (Roche, Molecular Biochemicals) or avidin–FITC (Vector Laboratories), respectively.

Digital images were obtained with an Axiocam camera coupled to a Zeiss ImagerZ microscope and analysed by AxioVision software (Zeiss).

Digitised photos were prepared for printing including contrast and colour optimization in Adobe Photoshop.

In silico analysis

The 'Synteny' application from Ensembl database (http://www.ensembl.org/; v67—May 2012) was used to access the syntenies between the genomes mouse/rat, mouse/human and rat/human (Supplementary S1a and S1b). These data were used to confirm and further increase the painting data resolution using mouse and rat probes and for refining the *P. tullbergi* map by inferring human syntenies in this genome.

Results

Pattern of chromosome rearrangements

The karyotype of *P. tullbergi* is composed of 34 chromosomes with all acrocentric autosomes. The Y is a small acrocentric and the X is a large submetacentric chromosome (Qumsiyeh et al. 1990; Capanna et al. 1996; Meles et al. 2008). Figure 1 shows various hybridization images obtained with mouse and rat paint probes on *P. tullbergi* chromosomes. The hybridization results are summarised in Fig. 2, where they are all mapped onto high resolution chromosomes of the haploid karyotype of *P. tullbergi*. The rat probes disclosed 33 syntenic segments and the mouse probes, at least 44 syntenic segments in *P. tullbergi* autosomes. When including probable bulk heterochromatin insertions (see below), the rat probes delineated at least 61 segments and the mouse paints 66 segments in the *P. tullbergi* karyotype (autosomes).

In some *P. tullbergi* chromosomes, synteny was entirely conserved compared to rat [RNO4 (PTU5), RNO5 (PTU7), RNO6 (PTU10), RNO10 (PTU11) and RNO20 (PTU16), X] and mouse [MMU3 (PTU9) and MMU18 (PTU13), X]. All other *P. tullbergi* chromosomes were further rearranged. Nine out of the 16 *Praomys* autosomes showed homologies to two or more rat chromosomes/chromosome segments. Some chromosomes showed extensive rearrangements such as *P. tullbergi* chromosomes 1 and 4 (both show four homologous segments with rat paints).

The comparison using mouse paints indicated even more rearrangements: 14 out of the 16 *Praomys* autosomes showed to be composed by two or more homologous mouse chromosomes/chromosome segments. Again, some chromosomes showed extensive rearrangements such as *P. tullbergi* chromosomes 1, 2 and 8 (all show four or more homologous segments with mouse paints).

As for the comparison of mouse and rat (Stanyon et al. 1999) numerous translocations, none of the Robertsonian type was observed. Some patterns indicate further inversions after the translocation event (for example the pattern of rat paints 7 and 19 and mouse paints 10 and 8 on *P. tullbergi* chromosome 1). On some chromosomes, homologous segments are disrupted by segments that do not hybridise with either rat or mouse paints (for example on chromosome 2, four segments; on chromosome 5, five segments). The great majority of these segments, not present in the mouse and rat genome, are of heterochromatic nature, as previously revealed by C-banding (Meles et al. 2008; Paço et al. 2009). Figure 2 also shows a map for the constitutive heterochromatin bands obtained by classical C-banding sequentially to G-banding and heterochromatin identified by specific fluorochrome staining (Meles et al. 2008) and in situ RE digestion (Paço et al. 2009).

Syntenic associations in *P. tullbergi* delineated with mouse and rat paints

Whenever chromosomes rearrange during evolution, they form new 'syntenic associations' which can be used in comparative studies to trace ancestral vs. commonly derived chromosome patterns. *M. musculus* chromosome paints delineated the following syntenic associations on *P. tullbergi* chromosomes: MMU1/4 (PTU7), MMU1/16 (PTU14), MMU2/10 (PTU11), MMU2/13 (PTU8), MMU3/5 (PTU15), MMU5/6 (PTU5), MMU5/11 (PTU3), MMU5/13 (PTU4, PTU15), MMU7/17 (PTU2), MMU7/19 (PTU2), MMU8/16 (PTU8), MMU9/13 (PTU6), MMU10/8/10 (PTU1), MMU10/17 (PTU2, PTU16), MMU11/14 (PTU3), MMU11/17 (PTU11), MMU12/17 (PTU10), MMU13/15/13/15 (PTU4), MMU13/16 (PTU8), MMU16/17 (PTU11) and MMU17/1/17 (PTU12). *Praomys* syntenic associations delineated by the *R. norvegicus* paints were the following: RNO1/2 (PTU4), RNO1/7 (PTU4), RNO2/12 (PTU4, PTU15), RNO3/7 (PTU1), RNO7/19/7 (PTU1), RNO8/17 (PTU6), RNO9/13/9 (PTU12), RNO11/13 (PTU14), RNO11/16 (PTU8), RNO11/17

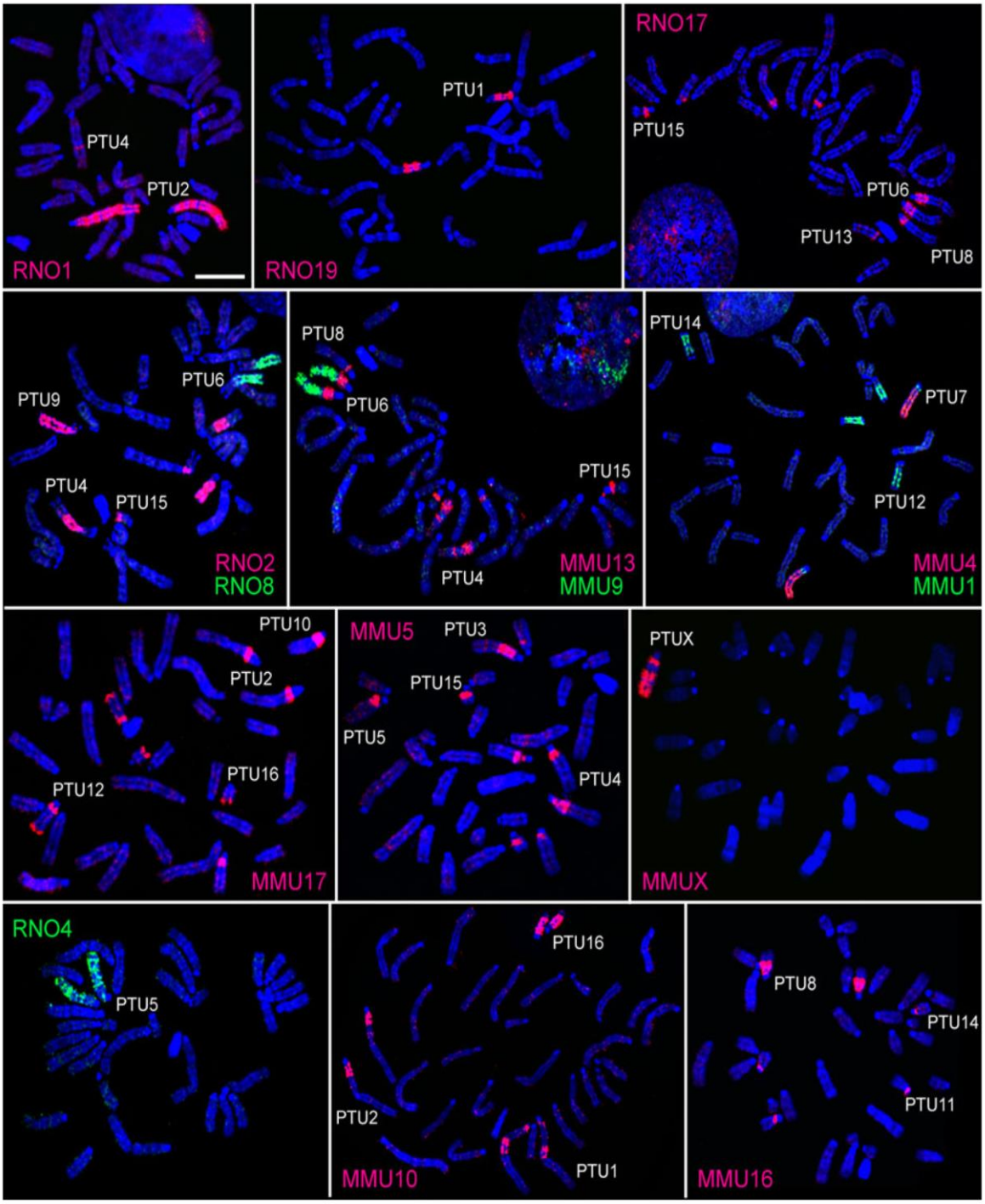


Fig. 1 Representative images of in situ hybridization with specific *Rattus norvegicus* (*RNO*) and *Mus musculus* (*MMU*) paint probes onto *P. tullbergi* chromosomes. The paint probes are indicated in the corner of each picture segment. The target chromosomes are numbered close to the hybridization signals scale bar = 5 μm

(PTU8), RNO12/17 (PTU15), RNO14/16 (PTU3), RNO15/16 (PTU3) and RNO17/18 (PTU13) (Fig. 2).

A test of consistency of rat and mouse painting in *P. tullbergi* chromosomes was performed (supplementary S2), where the common derived syntenic associations in *Praomys*/mouse vs. rat and *Praomys*/rat vs. mouse were examined. From the 21 syntenic associations delineated by MMU chromosome paints in *P. tullbergi* genome, 11 are common between PTU and RNO (MMU1/4, MMU2/13, MMU5/6, MMU5/11,

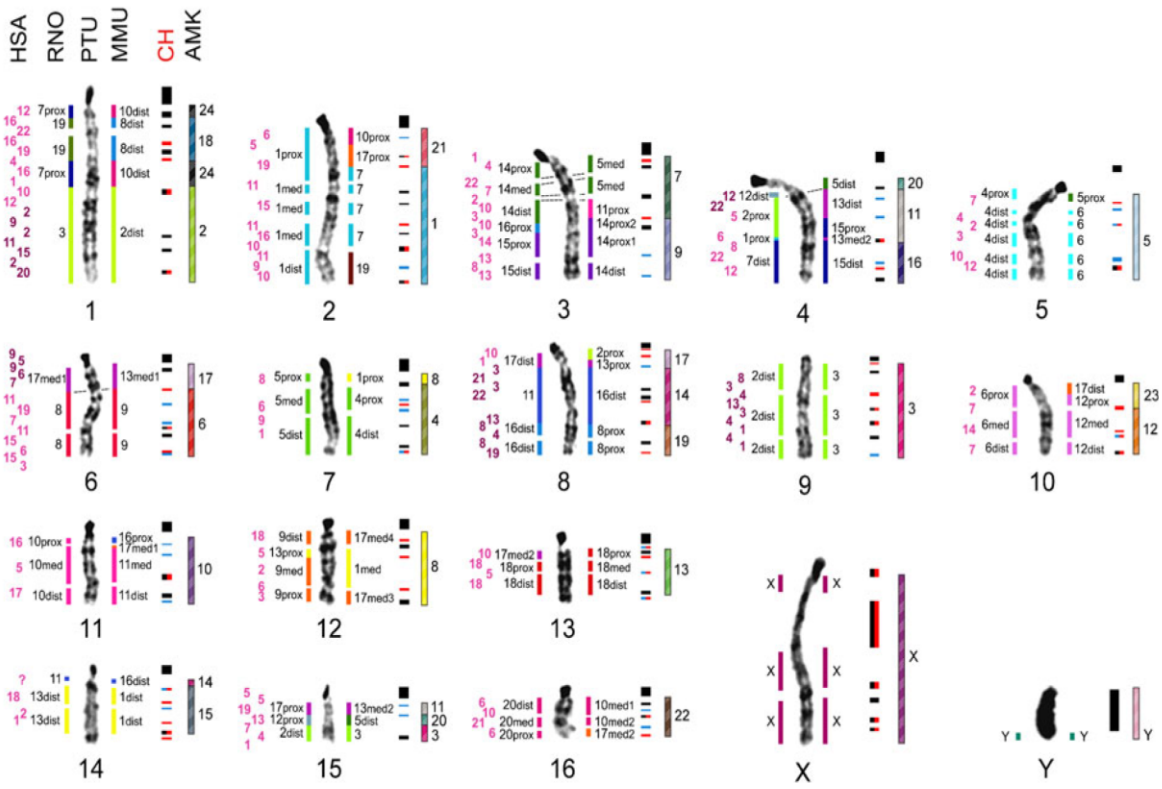


Fig. 2 Comparative chromosome map of *P. tullbergi* displaying the chromosomal homologies to *Mus musculus* (*MMU*), *Rattus norvegicus* (*RNO*), *Homo sapiens* (*HSA*) and the ancestral Muroidea karyotype (*AMK*) in the haploid karyotype of the GTG-banded *P. tullbergi* (*PTU*) chromosomes. In respect to human homologies, two shades of pink were used: in *pale pink*, the segments in which the order is known; in *dark pink*, the order of the segments is also known but is not possible to disclose if the block is inverted. Constitutive heterochromatin (*CH*) bands are also shown, namely C-banding sequential to G-banding (*black*), to in situ RE digestion (*blue*) and to specific fluorochromes staining (*red*)

MMU7/17, MMU7/19, MMU10/17, MMU12/17, MMU13/15, MMU16/17 and MMU17/1/17), and from the 14 associations obtained by RNO chromosome paints, only three (RNO 9/13, RNO15/16 and RNO17/18) are common between PTU and MMU.

Discussion

In comparative chromosome painting, contiguous chromosome segment combinations are used as phylogenetic characters to infer evolutionary relationships by discriminating between ancestral and derived chromosome forms (Wienberg et al. 2000). The present experiments delineate extreme genomic reshuffling between closely related species which up to now has only been reported for some few other rodents (e.g. Guilly et al. 1999; Rambau and Robinson 2003; Romanenko et al. 2007a, b; Hass et al. 2008; Badenhorst et al. 2011) and gibbons (primates) (Jauch et al. 1992).

Moreover, chromosome painting has been instrumental in transferring linkage information from 'map-rich' taxa to 'map-poor' taxa (e.g. Yang et al. 2000; Cavagna et al. 2002). This is especially true for the present experiments. The comparative paintings of mouse and rat probes to all individual chromosomes of *P. tullbergi* resulted in two consistent comparative maps correlating the genomes of these three rodents (Fig. 2). Rat and mouse genomic DNA sequencing data provided by the Ensembl genome browser, however, also provided the tool to suggest strong hypotheses about ancestral and derived chromosome forms at a resolution (base pair level) not possible up to now and using the human genomic DNA sequences as outgroup.

Phylogenetic analysis

Although intrachromosomal rearrangements usually escape detection by chromosome painting, the pattern

provided by some probes and syntenic associations with rat and mouse probes allowed the detection of inversion events in *P. tullbergi*. Rat paints disclosed two inversions on the *Praomys* genome and mouse paints three inversions: RNO7/19 and MMU8/10 on PTU1, RNO9/13 and MMU1/17 on PTU12 and MMU13/15 on PTU4. For the inversion on PTU1, both mouse and rat paints show the same pattern indicating a derived inversion in *P. tullbergi* after the divergence from mouse and rat. Curiously, a similar inversion was also detected in *Microtus oeconomus* from Cricetidae (Arvicolinae) (Sitnikova et al. 2007). The second inversion involving the syntenic association 1/17 is present in the *Rattus* genome, (RNO9, Nilsson et al. 2001) and in all the members of Muridae studied until now, with the exception of the *Mus* group (Matsubara et al. 2003; Veyrunes et al. 2006), some *Apodemus* members (Matsubara et al. 2004) and one Arvicanthis species (*Otomys irroratus*, Engelbrecht et al. 2006). This inversion also exists in one of the Cricetidae studied (*Ellobius lutescens*, Romanenko et al. 2007b) and is considered an ancestral form and may have been lost in the *Mus* lineage (Stanyon et al. 2004). Finally, the inversion observed with the mouse MMU13/15 paints was also found in *Coelomys pahari* (*Mus* subgenera, Veyrunes et al. 2006), the majority of the *Apodemus* karyotypes (Matsubara et al. 2004, not detected in Stanyon et al. 2004) and *O. irroratus* (Arvicanthis group, Engelbrecht et al. 2006) and should be considered the ancestral form for the Muridae, lost by some *Mus* subgenera and the *Rattus* group. Interestingly, the complex form of the inversion (double inversion 13/15/13/15) is found not only in *P. tullbergi*, but also in the Cricetidae *Cricetulus griseus* (Yang et al. 2000) which may be due to convergence supported by its absence in the majority of the studied genomes.

The presented results indicate that there are some autosomes and large autosomal segments that are conserved between the three species: mouse, rat and *Praomys*. The data show that *Praomys*/rat conserved more than twice of their syntenies (seven autosomes) than *Praomys*/mouse (only three that are also conserved between *Praomys*/rat). More interestingly, *Praomys*/rat shares nine commonly derived chromosome rearrangements that are not present in mouse, while *Praomys*/mouse shares only two that are not present in the rat. These data strongly suggest that *Praomys* and rat have retained more conserved regions (in comparison with *Praomys* and mouse), what is consistent with the notion that these are probably ancestral traits.

An integrated and detailed analysis of all the comparative chromosome painting works performed until now gave us a comprehensive picture of the chromosome phylogenetic relationships that unite the rodents already analysed by cross-species chromosome painting. The syntenic associations MMU2/10, MMU3/5, MMU8/16 and MMU11/14 are only found on the *P. tullbergi* genome, and for this reason, it should be considered species specific and derived forms, at least, until more species will be examined. All the other detected syntenic associations are shared by two or more species/families and should be considered common derived traits. At least one member of the two groups, *Rattus* (Muridae) and Arvicolinae (Cricetidae), shares the highest number of conserved syntenic associations: *M. oeconomus*, *Ellobius lutencens* (both Cricetidae, Arvicolinae) and *R. norvegicus*. Within the Muridae, *Rattus* and *Praomys* seem more conserved because of the highest number of common syntenic associations. This interpretation contrasts with some results obtained by molecular approaches alone (Steppan et al. 2004). However, Lee and Martin (1980) also suggested a close phylogenetic relationship between *Praomys natalensis* and *R. norvegicus*, using conventional G- and C-banding and Ag-NOR analysis. Curiously, *M. oeconomus* and *E. lutencens*, two species from the Cricetidae subfamily Arvicolinae, also share with *P. tullbergi* the same number of syntenic associations than *R. norvegicus* does (11 syntenic associations). This observation, together with the high number of syntenic associations also shared by some other members of this family, suggests the Cricetidae and specially Arvicolinae as a sister group of *P. tullbergi* what points to the fact that even though the high number of shuffles undergone in the evolution of these genomes, they may not be so phylogenetically apart as previously suggested.

Finally, the use of two index genomes in a same species together with the in silico analysis performed in Ensembl database allowed not only to construct two high resolution maps of *P. tullbergi*, and thus to gain insights about the genome architecture of one more member of this large and interesting family, but also to infer the human syntenies in this genome (Fig. 2). All the human syntenic associations 3/21 (PTU8), 4/8 (PTU8), 10/12 (PTU1 and PTU5), 12/22 (PTU4) and 16/19 (PTU1), already present on MMU and RNO genomes, were also observed in *P. tullbergi*; therefore, the high genome

repatterning occurred in this species genome apparently did not cause any disruption in these ancestral human syntenic associations.

Ancestral karyotype

The chromosome painting experiments with both mouse and rat paints and the assistance of the Ensembl genome browser allowed us to reconstruct a Muroidea ancestral karyotype (Muridae/Cricetidae and Murine). For this, homologous mouse and rat chromosome segments in *P. tullbergi* were assigned according to the position on the respective chromosome as proximal (prox), medium (med) and distal (dist) relatively to the centromere. The suggested ancestral karyotype is shown in Fig. 3 and displays the following chromosomes, composed from homologous mouse chromosome segments: 7prox/7dist/19, 2dist, 3, 4, 5prox/6, 9, 5med/11prox, 1prox + med/17med3 + med4, 14prox2/14prox1 + dist, 16prox/17med1/11med + dist, 13dist/15prox/13med2, 12med + dist, 18, 16dist, 1dist, 15dist, 13prox + med1/2prox, 8dist, 8prox, 5dist, 10prox/17prox, 17med2/10med, 12prox/17dist, 10dist, X, Y.

This karyotype would have $2n=50$ which is four chromosomes less than proposed by Stanyon et al. in 2004 ($2n=54$), two chromosomes less than the one suggested by Engelbrecht et al. (2006) ($2n=52$) and two more chromosomes than proposed by Romanenko et al. (2006) ($2n=48$). We propose that the segment homologous to 16dist of MMU would be a single chromosome. The other difference, also supported by other authors (Veyrunes et al. 2006), follows the assumption that the association (with inversion) 13dist/15prox/13med is an ancestral state and is a unique chromosome in the Muroidea ancestor. The proposed ancestral karyotype differs from that very recently suggested by Romanenko et al. (2012) which has $2n=52$ chromosomes. Lasting contrast to this karyotype, we propose that there was no composition 8/2/13 or 5/14, since the syntenic associations 2/8 and 5/14 are found only in Cricetidae and therefore should be considered derived traits for this group. Also, the ancestral chromosome originated by the inversion 1/17/1 is not considered ancestral since this chromosome form is observed mainly in Muridae species.

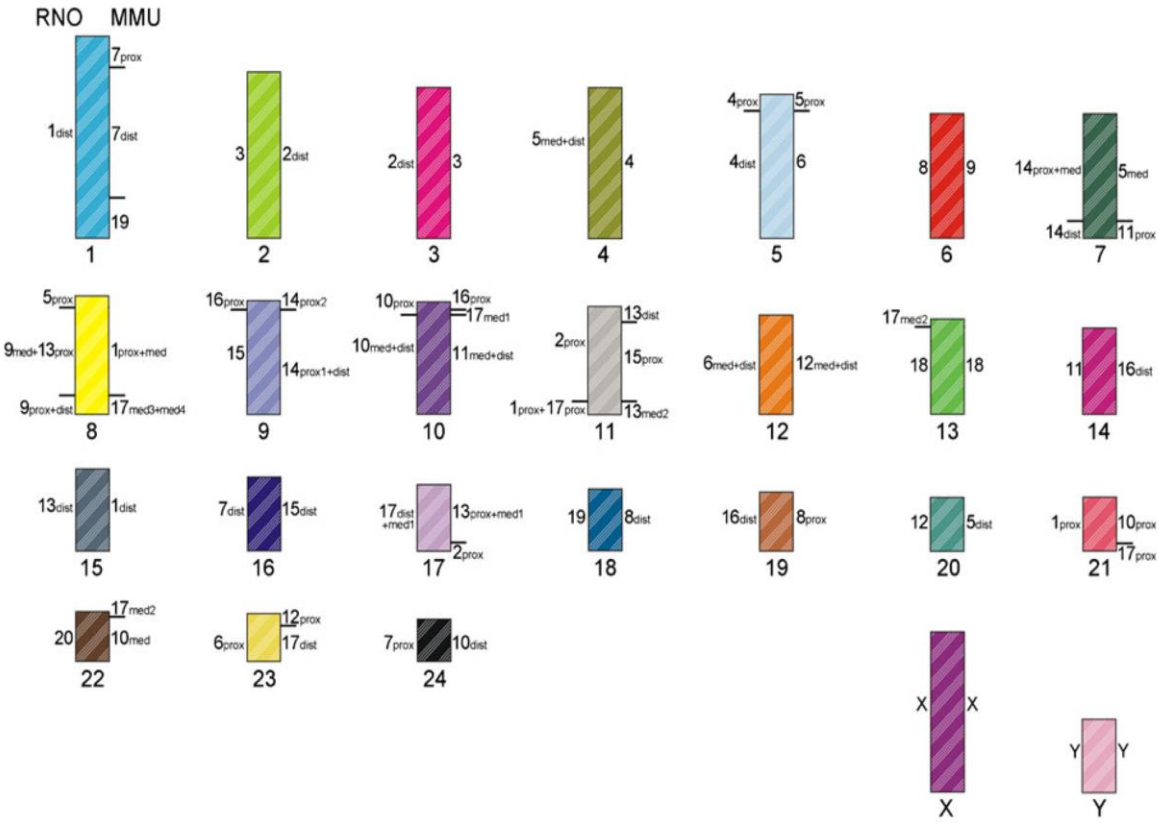


Fig. 3 Putative AMK with *Mus musculus* (*MMU*) and *Rattus norvegicus* (*RNO*) homologies

Starting from this proposed ancestral karyotype, it should be possible to reconstruct the chromosome evolutionary events that the *Praomys* genome experienced from the ancestral putative genome. There are about 32 autosomal conserved segments between the ancestor and *Praomys* genomes. A minimum of 23 major rearrangements separate the Muroidea ancestor and *P. tullbergi*, including, at least, 7 fissions, 15 translocations (tandem translocations and/or Robertsonian translocations accompanied by centromere repositioning) and 1 inversion. The mouse karyotype experienced 26 rearrangements after the divergence from the ancestral form, while in the rat, only 18 rearrangements occurred. Seven (RNO3, 4, 8, 10, 14, 19 and 20) of the 20 autosome pairs of the rat are apparently entirely conserved from the ancestral karyotype. The relative high conservation of the rat karyotype compared to other rodents would make rat paint probes a well-suited tool for chromosome studies in rodents.

Constitutive heterochromatin accumulation at the breakpoints

In this work, we also assembled all the data related to the CH characterization of *P. tullbergi* in the homology map (Fig. 2) using C-banding sequential to G-banding, to in situ RE digestion (Paço et al. 2009) and to specific fluorochromes staining (Meles et al. 2008). The combination of these methodologies allowed the identification of 126 heterochromatic bands, displaying centromeric, interstitial and telomeric locations. The majority of the detected CH co-localises with evolutionary breakpoint regions of *P. tullbergi* karyotype identified by comparative chromosome painting with rat and mouse probes. Among the *P. tullbergi* breakpoint regions revealed by *R. norvegicus* genome (45), about 70 % co-localise with CH sequences. For the mouse paints, 74 % of the breakpoints detected (49) are composed by CH. Finally, the assessment to the putative ancestral Muroidea karyotype (AMK) allowed inferring the breakpoint regions in *P. tullbergi* chromosomes; in a total of 16 breakpoints, 15 of them co-localise with C-bands (94 %).

The combination of the index genomes *R. norvegicus* and *M. musculus* allowed a further extrapolation of the breakpoints between *P. tullbergi* and the human genome. This in silico analysis increased largely the resolution of the map, especially of the largest syntenic blocks detected by the index rodent's genomes. For instance, the large syntenic block disclosed by RNO3 and MMU2 on PTU1 could be resolved in four smaller conserved syntenies with the human genome, all of them flanked by CH. This was observed for the great majority of the chromosomes. These results strongly suggest the involvement of repetitive sequences in the chromosome rearrangements that originated *P. tullbergi* genome architecture along evolution. The dynamic nature of repetitive sequences has been outlined in several studies (e.g. Plohl 2010), revealing that they constitute an important factor for genomic plasticity (e.g. Adega et al. 2009).

Acknowledgments This work was supported by the project POCI/BIA-BCM/58541/2004, a research position on Animal Genomics of the 'Sistema Científico e Tecnológico Nacional-Ciência 2007' and the PhD grants SFRH/BD/25813/2005 and SFRH/BD/41576/2007, all from the Science and Technology Foundation (FCT) from Portugal. We are deeply grateful to Dr. Vitaly Volobouev for providing the cell cultures of *P. tullbergi*.

References

Adega F, Guedes-Pinto H, Chaves R (2009) Satellite DNA in the karyotype evolution of domestic animals—clinical considerations. Cytogenet Genome Res 126(1–2):12–20

Badenhorst D, Dobigny G, Adega F et al (2011) Chromosomal evolution in Rattini (Muridae, Rodentia). Chromosome Res 19(6):709–727

Baker RJ, Qumsiyeh MB, Rautenbach IL (1988) Evidence for eight tandem and five centric fusions in the evolution of the karyotype of *Aethomys namaquensis* A. Smith (Rodentia: Muridae). Genetica 76(3):161–169

Britton-Davidian J, Catalan J, Ramalhinho MG et al (2005) Chromosomal phylogeny of Robertsonian races of the house mouse on the island of Madeira: testing between alternative mutational processes. Genet Res 86(3):171–183

Capanna E, Corti M (1982) Reproductive isolation between two chromosomal races of *Mus musculus* in the Rhaetian Alps (Northern Italy). Mammalia 46:107–109

Capanna E, Codjia JTC, Chrysostome C, Civitelli MV (1996) Les chromosomes des Rongeurs du Bénin (Afrique de l'Ouest): 3 Murinae. Atti della Accademia dei lincei classe di scienze fisiche matematiche e naturali rendiconti lincei scienze fisiche e naturali 8(9):25–37

Cavagna P, Stone G, Stanyon R (2002) Black rat (*Rattus rattus*) genomic variability characterized by chromosome painting. Mammal Genome 13(3):157–163

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(3):315–325

Chevret P, Granjon L, Duplantier J-M, Denys C, Catzeflis FM (1994) Molecular phylogeny of the *Praomys* complex (Rodentia: Murinae): as study based on DNA/DNA hybridization experiments. Zool J Linn Soc Lond 112(4):425–442

Dobigny G, Baylac M, Denys C (2002) Geometric morphometrics, neural networks and diagnosis of sibling *Taterillus* (Rodentia, Gerbillinae). Biol J Linn Soc 77(3):319–327

Dobigny G, Granjon L, Aniskine V, Ba K, Volobouev V (2003) A new sibling *Taterillus* species from West Africa. Mamm Biol 68(5):299–316

Engelbrecht A, Dobigny G, Robinson TJ (2006) Further insights into the ancestral murine karyotype: the contribution of the *Otomys–Mus* comparison using chromosome painting. Cytogenet Genome Res 112(1–2):126–130

Froenicke L, Wienberg J, Stone G, Adams L, Stanyon R (2003) Towards the delineation of the ancestral eutherian genome organization: comparative genome maps of human and the African elephant (*Loxodonta africana*) generated by chromosome painting. Proc Biol Sci 270(1522):1331–1340

Garagna S, Marziliano N, Zuccotti M, Searle JB, Capanna E, Redi CA (2001) Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. Proc Natl Acad Sci USA 98(1):171–175

Gava A, Freitas TR (2003) Inter and intra-specific hybridization in tuco-tucos (*Ctenomys*) from Brazilian coastal plains (Rodentia: Ctenomyidae). Genetica 119(1):11–17

Graphodatsky AS, Sablina OV, Meyer MN, Malikov VG, Isakova EA, Trifonov VA et al (2000) Comparative cytogenetics of hamsters of the genus *Calomyscus*. Cytogenet Cell Genet 88(3–4):296–304

Gropp A, Winking H (1981) Robertsonian translocations: cytology, meiosis, segregation patterns and biological consequences of heterozygozity. In: Berry RJ (ed) Biology of the house mouse. Academic, London, pp 141–181

Guilly M-N, Fouchet P, De Chamisso P, Schmitz A, Dutrillaux B (1999) Comparative karyotype of rat and mouse using bidirectional chromosome painting. Chromosome Res 7 (3):213–221

Hass I, Sbalqueiro IJ, Müller S (2008) Chromosomal phylogeny of four Akodontini species (Rodentia, Cricetidae) from Southern Brazil established by Zoo-FISH using *Mus musculus* (Muridae) painting probes. Chromosome Res 16 (1):75–88

Jansa SA, Weksler M (2004) Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. Mol Phylogenet Evol 31 (1):256–276

Jauch A, Wienberg J, Stanyon R et al (1992) Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. Proc Natl Acad Sci 89(18):8611–8615

John B (1988) The biology of heterochromatin. In: Verma RS (ed) Heterochromatin: molecular and structural aspects. Cambridge University Press, Cambridge, pp 1–128

Lecompte E, Granjon L, KerbisPeterhans JC, Denys C (2002) Cytochrome b-based phylogeny of the *Praomys* group (Rodentia, Murinae): a new African radiation? CR Biol 325(7):1–14

Lecompte E, Nicolas V, Colyn M, Denys C, Vitaly V (2005) Description of the karyotype of *Heimyscus fumosus* and of several other murids from the Mount Doudou área (Gabon). Belg J Zool 135(1):21–25

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 MR, Martin LK (1980) *Mastomys* (=*Praomys*) *natalensis* is not a *Rattus* (Mammalia: Rodentia): karyological evidence. Cytogenetic Cell Genet 28(1–2):95–103

Lyons NF, Green CR, Gordon DH, Walters CR (1977) G-banding chromosome analysis of *Praomys natalensis* (Smith) (Rodentia, Mammalia). I. 36 chromosome population. Heredity 38:197–200

Lyons NF, Gordon DH, Green CA (1980) G-banding chromosome analysis of species a of the *Mastomys natalensis* complex (Smith, 1834) (Rodentia, Muridae). Genetica 54 (2):209–212

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 (1):57–64

Matsubara K, Nishuda-Umehara C, Tsuchiya K, Nukaya D, Matsuda Y (2004) Karyotypic evolution of *Apodemus* (Muridae, Rodentia) inferred from comparative FISH analyses. Chromosome Res 12(4):383–395

Meles S, Adega F, Guedes-Pinto H, Chaves R (2008) The karyotype and sex chromosomes of *Praomys tullbergi* (Muridae, Rodentia): a detailed characterization. Micron 39(5):559–568

Murphy WJ, Stanyon R, O'Brien SJ (2001) Evolution of mammalian genome organization inferred from comparative gene mapping. Genome Biol 2(6):1–8

Nachman MW, Searle JB (1995) Why is the house mouse karyotype so variable? Trends Ecol Evol 10(10):397–402

Nash WG, Wienberg J, Ferguson-Smith MA, Menninger JC, O'Brien SJ (1998) Comparative genomics: tracking chromosome evolution in the family Ursidae using reciprocal chromosome painting. Cytogenet Cell Genet 83(3–4):182–192

Nash WG, Menninger JC, Wienberg J, Padilla-Nash HM, O'Brien SJ (2001) The pattern of phylogenomic evolution of the Canidae. Cytogenet Cell Genet 95(3–4):210–224

Nicolas V, Verheyen E, Verheyen W et al (2005) Systematics of African lowland rainforest *Praomys* (Rodentia, Muridae) based on molecular and craniometrical data. Zool J Linn Soc-Lond 145(4):539–553

Nie W, Wang J, Su W et al (2012) Chromosomal rearrangements and karyotype evolution in carnivores revealed by chromosome painting. Heredity 108(1):17–27

Nilsson S, Helou K, Walentinsson A, Szpirer C, Nerman O, Stahl F (2001) Rat-mouse and rat-human comparative maps based on gene homology and high-resolution zoo-FISH. Genomics 74(3):287–298

Paço A, Adega F, Guedes-Pinto H, Chaves R (2009) Hidden heterochromatin: characterization in the Rodentia species *Cricetus cricetus*, *Peromyscus eremicus* (Cricetidae) and *Praomys tullbergi* (Muridae). Genet Mol Biol 32(1):58–68

Plohl M (2010) Those mysterious sequences of satellite DNAs. Periodicum Biologorum UDC 112(4):403–410

Qumsiyeh MB, King SW, Arroyo-Cabrales J et al (1990) Chromosomes and protein in morphologically similar species of *Praomys* sensu lato (Rodentia, Muridae). J Hered 81(1):58–65

Rabbitts P, Impey H, Heppell-Parton A et al (1995) Chromosome specific paints from a high resolution flow karyotype of the mouse. Nat Genet 9(4):369–375

Rambau RV, Robinson TJ (2003) Chromosome painting in the African four-striped mouse *Rhabdomys pumilio*: detection of possible murid specific contiguous segment combinations. Chromosome Res 11(2):91–98

Robinson TJ, Ropiquet (2011) Examination of hemiplasy, homoplasy and phylogenetic discordance in chromosomal evolution of the Bovidae. Syst Biol 60(4):439–450

Romanenko SA, Perelman P, Serdukova N et al (2006) Reciprocal chromosome painting between three laboratory rodent species. Mamm Genome 17(12):1183–1192

Romanenko SA, Sitnikova NA, Serdukova NA et al (2007a) Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). II. The genome homology of two mole voles (genus *Ellobius*), the field vole and golden hamster revealed by comparative chromosome painting. Chromosome Res 15(7):891–897

Romanenko SA, Volobouev V, Perelman PL et al (2007b) Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroidea, Rodentia) inferred from chromosomal painting and banding comparison. Chromosome Res 15(3):283–297

Romanenko SA, Perelman PL, Trifonov VA, Graphodatsky AS (2012) Chromosomal evolution in Rodentia. Heredity (Edinb) 108(1):4–16

Savic I, Nevo E (1990) The Spalacidae: evolutionary history, speciation, and population biology. In: Nevo E, Reig AO (eds) Evolution of subterranean mammals at the organismal and molecular levels. Liss, New York, pp 129–153

Sitnikova NA, Romanenko SA, O'Brien PCM et al (2007) Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). I. The genome homology of tundra vole, field vole, mouse and golden hamster revealed by comparative chromosome painting. Chromosome Res 15(4):447–456

Stanyon R, Yang F, Cavagna P et al (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(3–4):150–155

Stanyon R, Yang F, Morescalchi AM, Galleni L (2004) Chromosome painting in the long-tailed field mouse provides insights into the ancestral murid karyotype. Cytogenet Genome Res 105(2–4):406–411

Steppan S, Adkins R, Anderson J (2004) Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst Biol 53(4):533–553

Veyrunes F, Dobigny G, Yang F et al (2006) Phylogenomics of the genus *Mus* (Rodentia; Muridae): extensive genome repatterning is not restricted to the house mouse. Proc R Soc B 273(1604):2925–2934

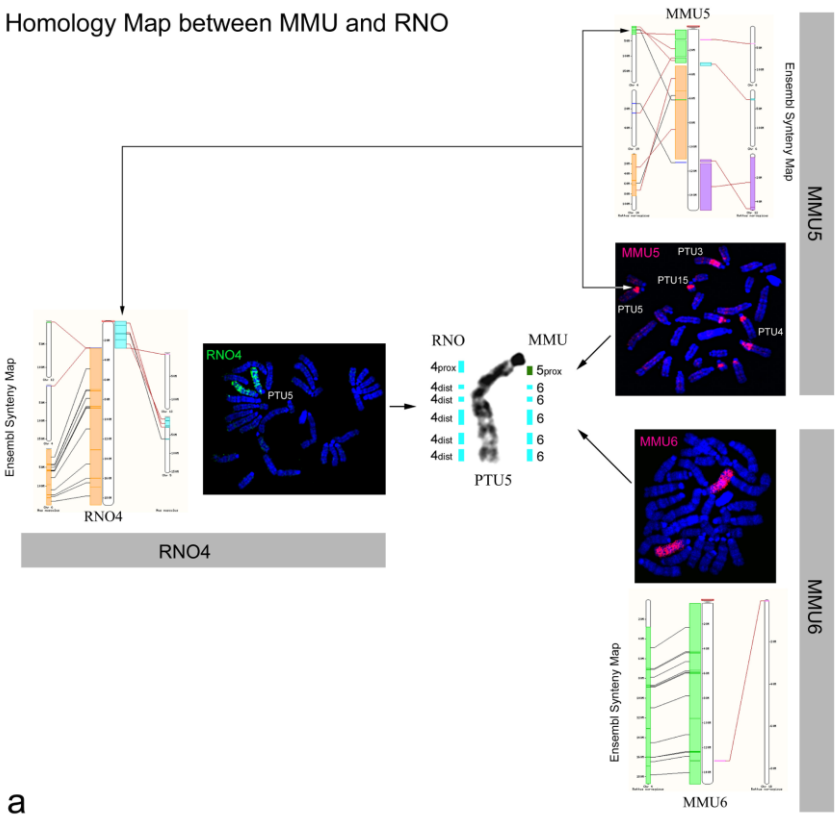
Wienberg J, Jauch A, Stanyon R, Cremer T (1990) Molecular cytotaxonomy of primates by chromosomal in situ suppression hybridization. Genomics 8(2):347–350

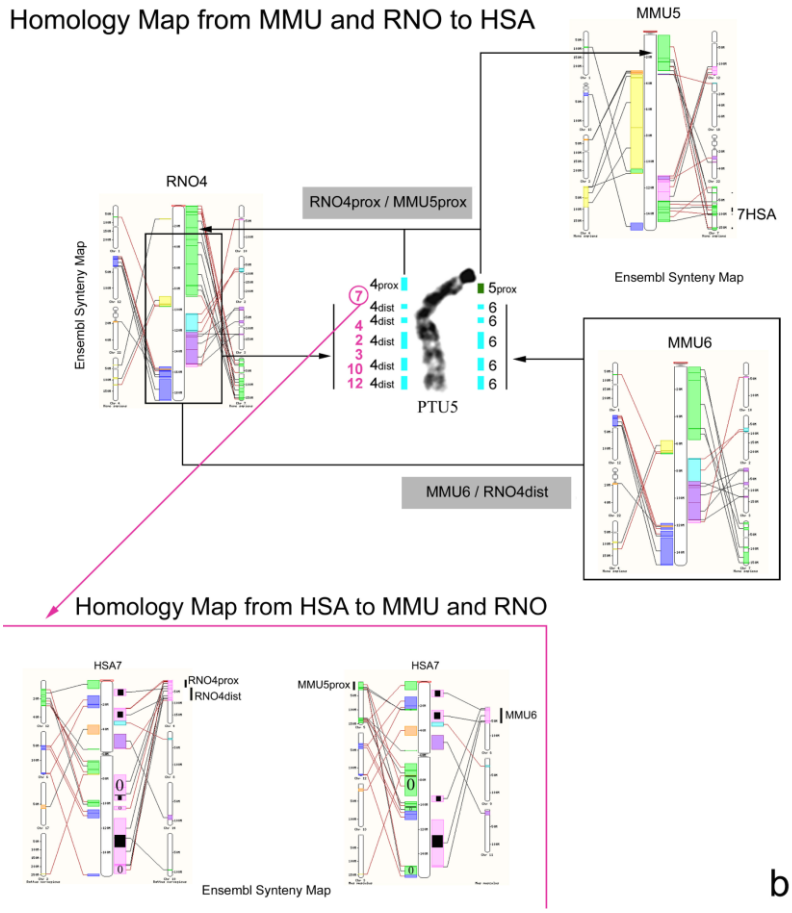
Wienberg J, Stanyon R, Nash WG et al (1997) Conservation of humans vs. feline genome organization revealed by reciprocal chromosome painting. Cytogenet Cell Genet 77(3–4):211–217

Wienberg J, Fronicke L, Stanyon R (2000) Insights into mammalian genome organization and evolution by molecular cytogenetics. In: Clark MS (ed) Comparative genomics. Kluwer Academic, Norwell

Yang F, O'Brien PCM, Ferguson-Smith MA (2000) Comparative chromosome map of the laboratory mouse and Chinese hamster defined by reciprocal chromosome painting. Chromosome Res 8(3):219–227

SUPLEMENTARY FIGURES AND TABLES





S1: Schematic presentation of the integrative in silico analysis (Ensembl sequencing data from the Synteny tool with our data) that led to the inference of the rat (RNO) and mouse (MMU) chromosome homologies to human (HSA). The example for chromosome PTU5 is shown. **S1a.** The use of both index species (rat and mouse) paint probes onto *Praomys tullbergi* (PTU) chromosomes allowed the specific delineation of homologies between RNO and MMU. In this specific case RNO4 to MMU5 and MMU6. With the assistance of Ensembl Synteny assemblies it was then possible to unequivocally determine the exact homology chromosome segment between each synteny block of *Mus musculus* and *Rattus norvegicus*. This data integration also allowed to designate as proximal, median and distal the different syntenic blocks regarding its position in relation to the centromere. In the example shown, RNO4prox is homologous to MMU5prox and RNO4dist to the entire RNO6. **S1b**. The accuracy of this analysis allowed to further assign the homologies between each rodent index species and human. And from this standpoint to *P. tullbergi* chromosomes. RNO4prox/MMU5prox showed to be homologous to HSA7. RNO4dist/MMU6 showed to be homologous to HSA7, HSA4, HSA2, HSA3, HSA10 and HSA12. Highlighted is HSA7, whose homologies to MMU and RNO shown to confirm the results above. This is evident when comparing segments marked with black squares or circles in both Ensembl synteny maps (HSA7 to RNO and MMU): black squares correspond to RNO4dist/MMU6 and circles correspond to RNO4prox/MMU5prox (PDF 2,033 kb).

MMU syntenic associations	PTU	RNO
MMU 1/4	+	+
MMU 1/16	+	
MMU 2/10	+	
MMU 2/13	+	+
MMU 3/5	+	
MMU 5/6	+	+
MMU 5/11	+	+
MMU 5/13	+	
MMU 7/17	+	+
MMU 7/19	+	+
MMU 8/16	+	
MMU 9/13	+	
MMU 10/8/10	+	
MMU 10/17	+	+
MMU 11/14	+	
MMU 11/17	+	
MMU 12/17	+	+
MMU 13/15	+	+
MMU 13/16	+	
MMU 16/17	+	+
MMU 17/1/17	+	+

RNO syntenic associations	PTU	MMU
RNO 1/2	+	
RNO 1/7	+	
RNO 2/12	+	
RNO 3/7	+	
RNO 7/19/7	+	
RNO 8/17	+	
RNO 9/13/9	+	+
RNO 11/13	+	
RNO 11/16	+	
RNO 11/17	+	
RNO 12/17	+	
RNO 14/16	+	
RNO 15/16	+	+
RNO 17/18	+	+

S2: Test of consistency of rat and mouse painting in *P. tullbergi* chromosomes showing the common derived syntenic associations in PTU/MMU vs. RNO and PTU/RNO vs. MMU (PDF 5 kb).

1.2 A high-resolution comparative chromosome map of *Cricetus cricetus* and *Peromyscus eremicus* reveals the involvement of constitutive heterochromatin in breakpoint regions

A high-resolution comparative chromosome map of *Cricetus cricetus* and *Peromyscus eremicus* reveals the involvement of constitutive heterochromatin in breakpoint regions

A Vieira-da-Silva[1,*], S Louzada[1,*], F Adega[1], R Chaves[1,2,+]

* AV-da-S and SL contributed equally to the present work

[1]Institute for Biotechnology and Bioengineering, Centre of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro (CGB-UTAD/IBB),Vila Real, Portugal.

[2]Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Apdo 1013, 5001-801 Vila Real, Portugal

+ Author for correspondence E-mail: rchaves@utad.pt

ABSTRACT | Compared to humans and other mammals, rodent genomes, specifically Muroidea species, underwent intense chromosome reshufflings, in which many complex structural rearrangements occurred. This fact makes them preferential animal models for studying the process of karyotype evolution. Here we present the first combined chromosome comparative maps between the index species *Mus musculus* and *Rattus norvegicus* and two Cricetidae species - *Cricetus cricetus* and *Peromyscus eremicus*. The methodology applied was the comparative chromosome painting, using mouse and rat paint probes, together with in silico analysis from Ensemble genome browser database. This study allowed inferring the evolutionary events (inter- and intrachromosomal rearrangements) that occurred in these karyotypes since the putative ancestral Muroidea genome as well as the detection of *C. cricetus* and *P. eremicus* evolutionary breakpoint regions. The analysis of the colocalization between constitutive heterochromatin regions and evolutionary breakpoint regions in each genome was performed. Our results suggest the involvement of constitutive heterochromatin regions in these species karyotype restructuring, despite the different level of conservation of *C. cricetus* (derivative) and *P. eremicus* (conserved) genomes.

Key words: comparative chromosome map; *Cricetus cricetus*; *Peromyscus eremicus*; breakpoint regions; constitutive heterochromatin

INTRODUCTION

Comparative studies have started with the main purpose of clustering the extant species in groups. First works rely on morphologic features but with the technological improvements more precise analysis became available, such as paleontological, molecular and cytogenetic data. Beyond resolving phylogenies, the obtained information allowed new approaches such as weaving the path of evolutionary events in the genomes (e.g. infer probable evolutionary chromosomal rearrangements and pinpoint the localization of breakpoint regions) (Veyrunes et al., 2006; Ramsdell et al., 2008). These studies reveal also important biomedical information since the evolutionary large-scale genome rearrangements (translocations, inversions and deletions of chromosomal regions of several megabases in length), characteristic of genomic instability required to evolution, are also observed in many different human disease states (Longo et al., 2009). For instance, the most commonly occurring human cancer-associated breakpoint regions tend to colocalize with evolutionary breakpoint regions (EBr) (Robinson et al., 2006).

Whole genome alignment studies have shown that EBr (region between two syntenic segments) are rich in repetitive elements (Murphy et al., 2005, Ruiz-Herrera et al., 2006). Moreover, the constitutive heterochromatin (CH) regions have been previously pointed as hotspot for structural chromosome rearrangements (John, 1988; Louzada et al., 2008; Adega et al., 2009). The association between EBr and repetitive sequences is thought to be related to the role that repeat sequences have in predisposing the region where they are located (mainly in CH) to large-scale chromosomal instability, playing as a substrate for non-homologous recombination, thereby promoting chromosomal rearrangements (Froenicke and Lyons, 2008; Adega et al., 2009).

The Muroidea species (Rodentia) are a preferential mammalian group for evolutionary studies (Romanenko et al., 2007a). This superfamily comprises the larger, the most diverse and evolutionary successful mammalian species (Murphy et al., 2001; Romanenko et al., 2007a), with a wide variation in diploid numbers and heterochromatin nature. Muroid species genomes are characterized to have undergone a rapid genomic evolution (Steppan et al., 2004) involving many complex structural rearrangements (e.g. Nadeau and Taylor, 1984; Nilsson et al., 2001; Kent et al., 2003; Ramsdell et al., 2008), hindering the phylogeny of this group. The majority of rodent phylogenetic studies that have been performed are morphological (e.g. Miller and Gidley, 1911; Catzeflis et al., 1992), paleontological (e.g. Chaline et al., 1977; Jacobs et al., 1989; Hugueney and Mein, 1993), molecular (e.g. Catzeflis et al., 1993; Dubois et al., 1996, 1999; Robinson et al., 1997; Jansa et al., 1999; Michaux et al., 2001; Adkins et al., 2003) and cytogenetic (by classical methods) (e.g. Veyrunes et al., 2004). However, a molecular cytogenetic method with high resolution,

Comparative Chromosome Painting (CCP), became the most used (reviewed in Romanenko et al., 2012).

The present study concerns two Cricetidae (Rodentia, Muroidea) species, the common hamster (*Cricetus cricetus*, CCR) and the cactus mouse (*Peromyscus eremicus*, PER) belonging to the Cricetinae and Neotominae subfamilies, respectively (Musser and Carleton, 2005). These two species have been used in biomedical research such as neurology (Glasper and DeVries, 2005) and neuroendocrinology (Hanon et al., 2010).

In this work we provide a high-resolution detailed comparison between the index genomes, *Mus musculus* (MMU) and *Rattus norvegicus* (RNO), and the two Cricetidae species, using CCP. Here we report for the first time the combined use of mouse and rat paint probes supported by an in silico analysis using the mouse and rat sequences available in Ensemble genome browser, resulting in the refinement of CCR and PER comparative maps resolution. Also, a detailed analysis of CH and its colocalization with EBr is shown. Our data strongly suggest the involvement of CH in this species karyotype restructuring.

MATERIALS AND METHODS

Cell culture and metaphase preparation

The metaphase preparations were obtained with standard cell culture procedures from fibroblast cell lines of *Cricetus cricetus* (female) and *Peromyscus eremicus* (male). Chromosome nomenclature of *C. cricetus* (2n=22) chromosomes follows Gamperl et al. (1976) and nomenclature of *P. eremicus* (2n=48) chromosomes is according to The Committee for Standardization of Chromosomes of *Peromyscus* (1977). Chromosome number for rat (*Rattus norvegicus*, RNO) and mouse (*Mus musculus*, MMU) paint probes was according Stanyon et al. (1999).

Chromosome specific paint probes

The paints for this study (mouse and rat chromosome-specific probes) are commercially available probes from Chrombios GmbH and were kindly provided by Dr. A. Kofler and R. Kofler from the flow sorting facility at Chrombios. Chromosome-specific probes were prepared by primary DOP-PCR amplification (Rabbits et al., 1995) using the universal primers 6MW (for RNO paints) and F/S (for MMU paints) followed by labelling DOP-PCR with the incorporation of digoxigenin-dUTP (Roche Molecular Biochemicals).

Fluorescent in situ Hybridization (FISH)

The chromosome specific paint probes from mouse and rat were in situ hybridized to *C. cricetus* and *P. eremicus* chromosomes according to Wienberg et al. (1997). The most stringent post-hybridization wash was 50% formamide/2xSSC at 37ºC, and in the probe detection it was used antidigoxigenin-5'TAMRA (Roche, Molecular Biochemicals).

Image capture and processing

FISH images were observed in a Zeiss ImagerZ microscope, coupled to an Axiocam digital camera with AxioVision software (version Rel. 4.5 – Zeiss), which allowed the image capture. Digitised photos were prepared for printing in Adobe Photoshop (version 7.0); image optimization included contrast and colour adjustments that affected the whole image equally.

In silico analysis

The *in silico* analysis was performed according with Chaves et al. (2012). The syntenies between the genomes Mouse/Rat were accessed using the "Synteny" viewer application from Ensembl database (http://www.ensembl.org/; release 68 - July 2012). These data were used to confirm the painting results using mouse and rat probes and further increase the comparative maps resolution. For the mouse and rat chromosome segments delineated in *C. cricetus* and *P. eremicus* karyotypes maps it was used the nomenclature proximal (prox), medium (med) and distal (dist).

RESULTS

A high-resolution comparison of both *Cricetus cricetus* (CCR) and *Peromyscus eremicus* (PER) karyotypes was carried out using *Mus musculus* (MMU) and *Rattus norvegicus* (RNO) chromosomes and supported by MMU and RNO Ensembl genome browser data. *Cricetus cricetus* (2n=22) karyotype is composed by one acrocentric, five metacentric and four submetacentric autosomal chromosome pairs, a metacentric X and a submetacentric Y (Matthey, 1952). In turn, *Peromyscus eremicus* (PER, 2n=48) karyotype is constituted exclusively by submetacentric chromosomes, including the X (large submetacentric) and the Y (small submetacentric) chromosomes (Pathak et al., 1973; Deaven et al., 1977). Paint probes specific for each mouse (MMU1-19, X, Y) and rat (RNO1-20, X, Y) chromosomes were successfully cross-hybridized, delimiting the homologous chromosomal segments in CCR and PER, by in situ hybridization. In figure 1 it can be observed some of the resulting hybridization images obtained with MMU and RNO paint probes.

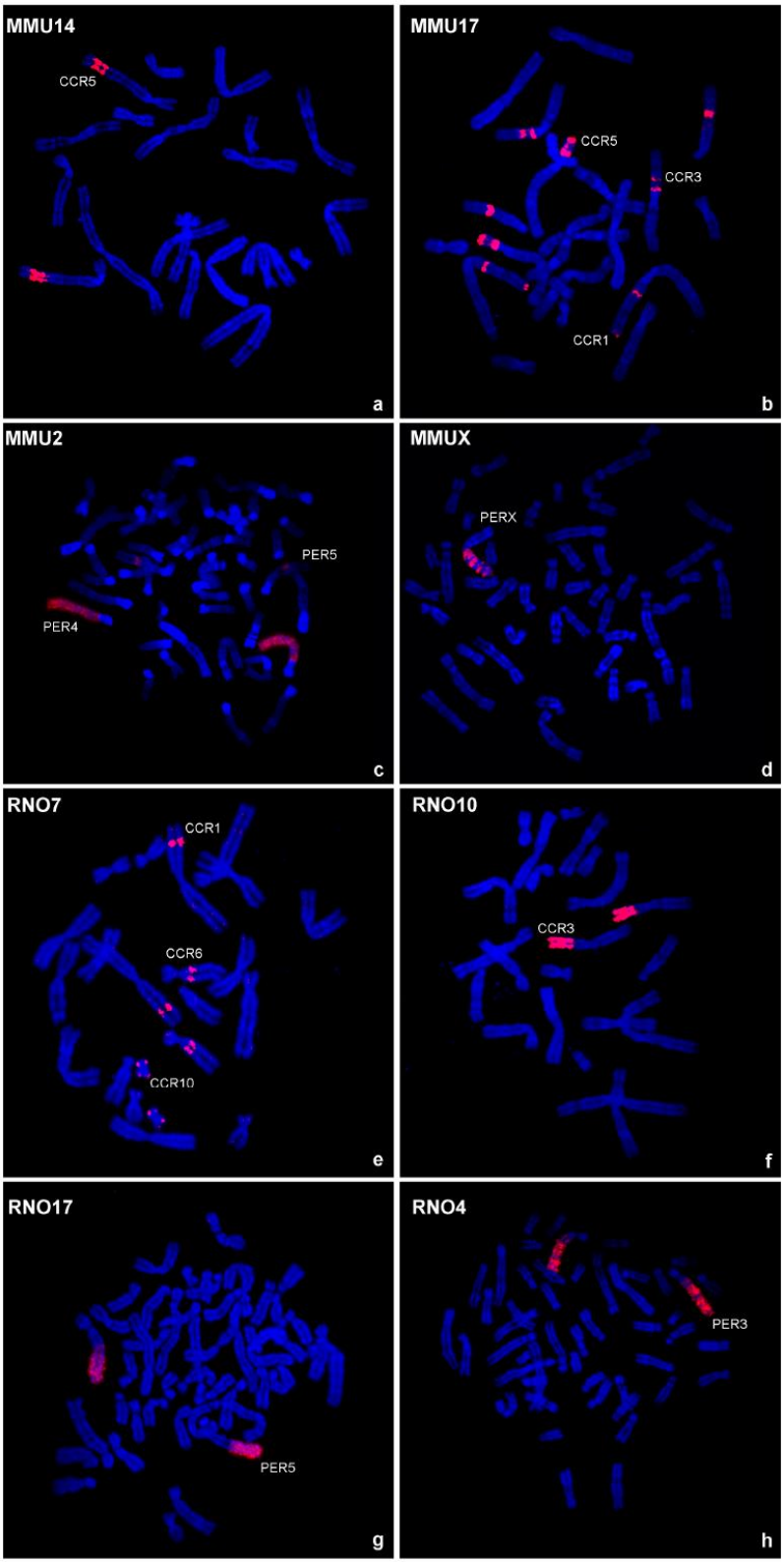


Figure 1 - Representative images of fluorescent *in situ* hybridization of different MMU (a,b,c,d) and RNO (e,f,g,h) paint probes on CCR (a,b,e,f) and PER (c,d,g,h) metaphasic chromosomes. In each image is referred the paint probe used (in the top left at the corner) and chromosomal localization in the species.

Comparative Chromosome Painting in *Cricetus cricetus*

Mus musculus and *Rattus norvegicus* paint probes highly segmented the *Cricetus cricetus* genome revealing respectively, 44 and 32 homologous segments in its autosomes (fig. 2 and supplementary table S1). When considering the intrachromosomal segments, the mouse probes delineated 77 syntenic segments, while the rat probes 67 syntenic segments (e.g. MMU4 delineated three segments in CCR1 and RNO3 is divided into four segments in CCR3). Regarding the sex chromosomes, X chromosome from *Mus musculus* and *Rattus norvegicus* each originated three conserved segments on CCRX p arm.

Detailed analysis of the mouse comparative results showed that MMU19 was the only probe homologous in toto to a single CCR chromosome region (CCR2). Five MMU paints (MMU3, MMU4, MMU9, MMU14 and MMU18) each hybridized also in only one CCR chromosome, originating, however, at least two signals on these chromosomes. The most rearranged MMU paint in CCR genome was MMU17, delineating seven syntenic segments on four CCR chromosomes. The other MMU probes painted two, three or four CCR chromosomes, revealing one or more conserved segments on these chromosomes. 26 syntenic associations of mouse homologous chromosomal segments were observed in CCR genome: MMU1/10, MMU1/15, MMU1/17 (thrice), MMU2/8, MMU2/13, MMU2/17, MMU3/10, MMU4/18, MMU5/6, MMU5/9, MMU5/11, MMU5/14, MMU5/16, MMU6/10, MMU7/8, MMU7/19, MMU8/11, MMU8/17, MMU10/17 (twice), MMU10/18 (twice), MMU11/17, MMU12/16, MMU12/17 (twice), MMU13/15, MMU13/17 and MMU16/17.

Regarding the rat comparative results, it was observed that two RNO chromosomes were conserved in toto in CCR genome (RNO12 and RNO15), each hybridizing in a single block in CCR4 and CCR5, respectively. Other nine rat chromosomes (RNO3, RNO5, RNO8, RNO10, RNO11, RNO13, RNO14, RNO17 and RNO19) hybridized in toto in one CCR chromosome, although originating more than one syntenic segment (e.g. RNO5 on CCR1). Also RNO16 and RNO18 paint probes hybridized in one CCR chromosome but appear disrupted by a different RNO segment. RNO1 and RNO7 were the more rearranged ones: RNO1 painted nine regions spread in three different CCR chromosomes and RNO7 produced four painting signals in three different CCR chromosomes. All the other RNO paints (RNO2, RNO4, RNO6, RNO9 and RNO20) provided one, two or more conserved segments in two CCR chromosomes. Nineteen syntenic associations of rat homologous chromosomal segments were observed in CCR genome: RNO1/2, RNO1/7, RNO1/18, RNO1/19, RNO2/7, RNO3/6, RNO4/7, RNO5/18, RNO6/10, RNO6/13, RNO7/9, RNO8/12, RNO9/16, RNO9/20, RNO11/12, RNO14/16, RNO15/16, RNO17/19 and RNO18/20 (twice).

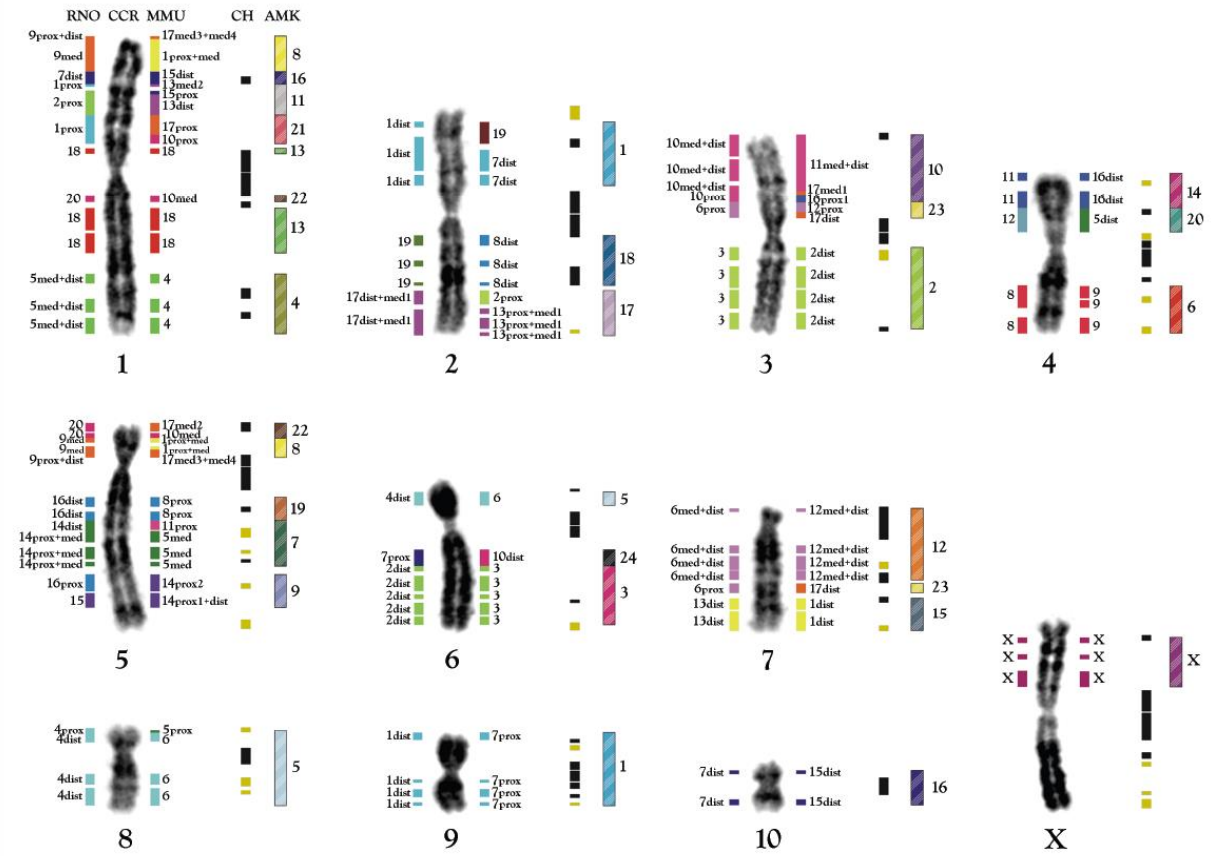


Figure 2- Comparative chromosome map of *Cricetus cricetus* (CCR). The correspondence with the haploid set of CCR and the homologous segments of *Rattus norvegicus* (RNO) (at the left) and *Mus musculus* (MMU) (at the right). The constitutive heterochromatin (CH) location is shown (according to Paço et al. 2009) as well the correspondence between CCR and ancestral Muroidea karyotype (AMK) (Chaves et al. 2012) homologies. The CH black blocks represent classical C-bands and yellow blocks represent cryptic C-bands. (dist-distal, med-median, prox-proximal).

Comparative Chromosome Painting in *Peromyscus eremicus*

The number of syntenic segments delineated by *M. musculus* and *R. norvegicus* paints in *Peromyscus eremicus* autosomes was 38 and 29 syntenic segments, respectively (fig. 3 and supplementary table S1). These numbers increase to a total of 65 (for MMU) and 62 (for RNO) when accounting the intrachromosomal segments (e.g. MMU4 delineated 4 segments in PER2 and RNO1 delineated 6 segments in PER 1). Hybridization of MMUX and RNOX resulted in eight conserved segments each in *P. eremicus* X chromosome, seven of which are located at the q arm and one in the p arm, near the centromeric region. Regarding the Y chromosome, both index species paint probes highlighted only one small terminal region of PERY.

Mouse comparative results showed that nine MMU chromosomes (MMU3, MMU4, MMU6, MMU7, MMU9, MMU14, MMU16, MMU18 and MMU19) hybridized in toto in PER

autosomes. Although, and with exception for MMU18, that hybridized in a single block, they were segmented in several syntenic segments, being MMU6 the chromosome that promoted the higher number of syntenic segments, five segments in PER3. Six MMU paints (MMU2, MMU8, MMU11, MMU12, MMU13 and MMU15) each hybridized in two PER chromosomes. MMU1, MMU5 and MMU10 produced signals in three PER chromosomes and MMU17 revealed a total of five conserved segments in four different *Peromyscus eremicus* chromosomes. The twelve syntenic associations delineated by *Mus musculus* chromosome paints were: MMU1/10, MMU1/17, MMU2/8, MMU2/13, MMU5/6, MMU5/11, MMU7/19, MMU10/17 (twice), MMU11/17, MMU12/17, MMU13/15 and MMU16/17.

Concerning the RNO paint results, 13 hybridize in toto to only one PER chromosome/chromosome segment, from which RNO12, RNO17 and RNO18 produced one continuous segment while the others (RNO3, RNO4, RNO5, RNO8, RNO10, RNO11, RNO13, RNO14, RNO15 and RNO19) delineated more than one syntenic segment. Six rat paint probes (RNO2, RNO6, RNO7, RNO9, RNO16 and RNO20) each produced signals in two PER chromosomes and RNO1 hybridized in three different PER chromosomes. *Rattus norvegicus* paint probes delineated 5 syntenic associations in *Peromyscus eremicus* genome: RNO1/2, RNO7/20, RNO9/20, RNO15/16 and RNO17/19.

Evolutionary Breakpoint regions versus Constitutive Heterochromatin

CCR and PER comparative maps allowed the detection of the evolutionary breakpoint regions (EBr), outlined by MMU, RNO and the putative ancestral Muroidea karyotype in each species (AMK according to Chaves et al., 2012) (figs. 2 and 3; supplementary table S2). It was possible to detect 67/57 breakpoint regions in CCR genome and 44/39 in PER genome displayed by MMU and RNO paint probes, respectively in each species. Additionally is shown, in figures 2 and 3, the constitutive heterochromatin (CH) regions in each species chromosomes. Previous studies identified a total of 56 C-bands in CCR and 112 C-bands in PER autosomes (summarized in Paço et al., 2009). These bands display different locations in both genomes, namely centromeric, interstitial and telomeric. The putative association between the detected breakpoint regions and the CH was analyzed, elucidating about the CH dynamic in both genomes during evolutions. The data is assembled in supplementary table S2.

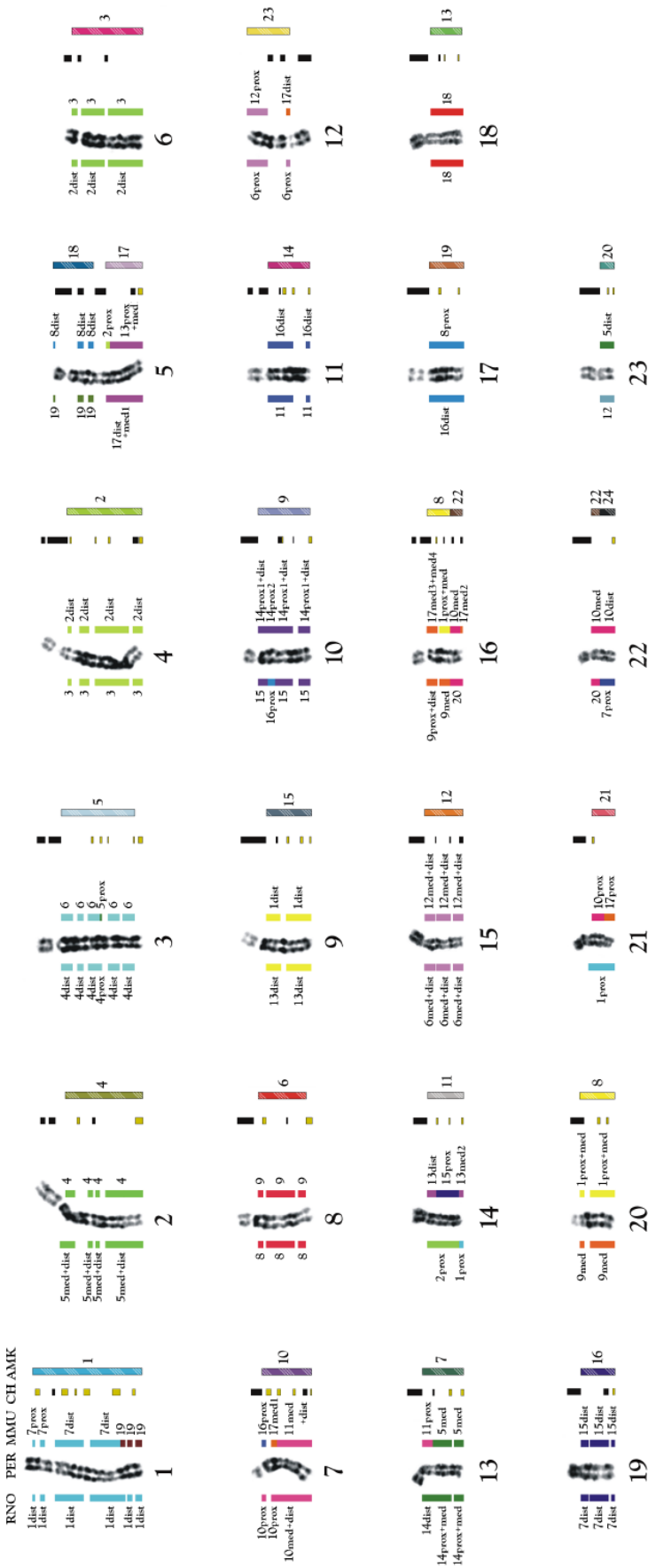

RNO
PER
MMU
CH
AMK
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Figure 3 - Comparative chromosome map of *Peromyscus eremicus* (PER). The correspondence with the haploid set of PER and the homologous segments of *Rattus norvegicus* (RNO) (at the left) and *Mus musculus* (MMU) (at the right). The constitutive heterochromatin (CH) location is shown (according to Paço et al. 2009) as well the correspondence between PER and ancestral Muroidea karyotype (AMK) (Chaves et al. 2012) homologies. The CH black blocks represent classical C-bands and yellow blocks represent cryptic C-bands. (dist-distal, med-median, prox-proximal)

DISCUSSION

Comparative analysis of *C. cricetus* and *P. eremicus*

Several comparative studies have been performed using *Mus musculus* (MMU) paint probes in Cricetidae and Muridae families. On the other hand, paint probes from *Rattus norvegicus* (RNO) have only been used for comparative studies in few Muridae species (Chaves et al., 2012; Romanenko et al., 2012). In the present study both paint probes (MMU and RNO) were used in the construction of the comparative maps for the Cricetidae species *Cricetus cricetus* and *Peromyscus eremicus*. The use of both index genomes' has increased the resolution of the *C. cricetus* (CCR) and *P. eremicus* (PER) maps, once they allowed the detection of intrachromosomal rearrangements. Further resolution was obtained by the use of the available sequence data for MMU and RNO genomes present in Ensembl genome browser, as shown in previous study (Chaves et al., 2012).

Comparative chromosome painting results (fig. 1) revealed a significant number of syntenic segments disclosed by the mouse and rat genomes in both species studied. RNO cross-species painting experiments revealed 29 regions of homology in *Peromyscus eremicus* and 32 in *Cricetus cricetus* autosomes. MMU chromosome-specific painting probes segmented even more PER and CCR genomes, originating 38 and 44 homologous segments, respectively. These numbers increase when we take in account that a significative amount of the syntenic blocks are disrupted by constitutive heterochromatin (CH) (figs. 2 and 3). Therefore, in PER and CCR autosomes, RNO paint probes revealed 62 (in PER) and 67 (in CCR) regions of homology and MMU paint probes originated 65 (in PER) and 77 (in CCR) homologous segments (supplementary tables S1). When comparing these results with analogous studies, it is possible to denote that they constitute, along with *Praomys tullbergi* (Chaves et al., 2012), the higher number of syntenies ever revealed among Muroidea species.

To weave the evolutionary path of the events occurred in the genomes of CCR and PER, MMU syntenic associations were analyzed. Supplementary tables S3 and S4 summarize all the MMU syntenic associations found in CCR and PER genomes and its analysis (presence/absence) in all the other Muroidea species studied so far. It is possible to verify that none is specific of CCR and PER, being all present in, at least, one more species. Ten MMU syntenic associations revealed in both CCR and PER genomes - MMU1/17, MMU2/13, MMU5/6, MMU5/11,

MMU7/19, MMU10/17, MMU11/17, MMU12/17, MMU13/15 and MMU16/17 - are also found in most Cricetidae and Muridae species, suggesting that they constitute a symplesiomorphic character, already present in the ancestral Muroidea karyotype (AMK). These results support the recently AMK proposed by Chaves et al. (2012), once they consider the associations MMU11/17 and MMU16/17 as ancestral conditions (like Romanenko et al., 2012) and not the association MMU11/16, as it is proposed by other authors (Stanyon et al., 2004; Engelbrecht et al., 2006; Romanenko et al., 2006). Also because Chaves et al. (2012), considers the association MMU5/6 an ancestral condition, which is not considered by Romanenko et al. (2012). The other MMU syntenic associations found in CCR (16) and in PER (3) are likely apomorphic, not present in the AMK, being considered as derivative for these and for other species (Cricetidae and/or Muridae) (supplementary tables S3 and S4).

The high number of ancestral conditions of *P. eremicus* genome revealed the occurrence of a small number of evolutionary large-scale genome rearrangements, evidencing that this species has a conserved genome. This fact suggests that the 12 MMU syntenic associations common between *Peromyscus eremicus* and *Peromyscus maniculatus* (PMA) (supplementary table S4) are ancestral conditions for the *Peromyscus* genus. CCR genome is more derivative than PER, given the greater number of apomorphic MMU syntenic assotiations. Cricetulus griseus (CGR) showed to be the closest related species to *Cricetus cricetus*, once they share 24 MMU syntenic associations (supplementary table S3). This data is supported by G-banding (Gamperl et al., 1976) and molecular cytogenetic studies (CCR/CGR/MAU comparative map) (Romanenko et al., 2006, 2007a). Besides CGR, the more related species to CCR belongs to Arvicolinae and Neotominae (both Cricetidae), however, it is closer to Arvicolinae species, sharing a large number of derivative MMU associations. This confirms the molecular data, which states that Cricetidae first split into two main branches, Cricetinae-Arvicolinae and Sigmodontinae-Neotominae at approximately 19.6 Mya, and only afterwards the four presently known subfamilies were originated (Steppan et al., 2004).

CCR and PER Evolutionary Rearrangements

One of the main goals of comparative studies is the delineation of chromosomal rearrangements from an ancestral karyotype. Regarding PER, a minimum of six major rearrangements occurred since the Muroidea ancestor proposed by Chaves et al. (2012), namely three fusions, two fissions and one intrachromosomal inversion. It is worth mentioning that without the RNO paint probes the intrachromosomal rearrangement in AMK9 (inversion revealed by RNO15 and RNO16 paint probes, fig. 3) would not be detected, showing the

importance and significance of the use of both index genomes. Besides, pericentric inversion(s), centromere repositioning and heterochromatin repatterning were events also involved in the evolution of PER genome. Greenbaum and Baker (1978) proposed the primitive karyotype for the *Peromyscus* genus as being composed by chromosomes with heterochromatin restricted to the centromeric regions, suggesting that the heterochromatic short arms, that characterize several *Peromyscus* lineages (including *Peromyscus eremicus*), have been added afterwards and independently in the species. Although *P. eremicus* chromosome short arms are almost entirely heterochromatic, some MMU syntenic segments were found at this region, e.g. in PER5 (fig. 3). This can be explained throughout the occurrence of a pericentric inversion along with heterochromatin additions or centromere repositioning. Other evidence for the occurrence of intrachromosomal inversions and/or heterochromatin repatterning is the presence of CH between breakpoint regions detected simultaneously by the two index genomes (e. g. PER6 and PER15, fig. 3).

Cricetus cricetus genome evolutionary path was more complex (fig. 2), given the high number of rearrangements observed since the AMK, at least 28 major rearrangements - twenty fusions, six fissions, one inversion and one pericentric inversion. Like Romanenko et al. (2007a), we suggest that fusions, fissons and inversions were the predominant rearrangements occurred on CCR genome evolution; however, our results showed a higher number of fusions and fissions than the obtained by the referred authors. Furthermore, the occurrence of heterochromatin repatterning certainly happens during the CCR genomic evolution. Accordingly, Gamperl et al. (1976) argued the occurrence of CH additions, mainly in the centromeric regions and our results suggest also that has occurred CH elimination (see below). Pericentric inversion(s) and/or centromere repositioning is likely also to have occurred in the evolution of CCR genome.

Concerning sex chromosomes, both MMUX and RNOX paint probes delineated eight homologous segments in PERX, being seven of them located in the q arm and only one located in the p arm (next to the centromere). This suggests the occurrence of a pericentromeric inversion in PERX, not detected in the related species *P. maniculatus* (Mlynarski et al. 2008). MMUY and RNOY revealed only one homologous segment in *P. eremicus* Y chromosome. In *Cricetus cricetus*, RNOX and MMUX paint probes painted three regions in the euchromatic region of CCRX p arm (the X ancestral chromosome).

This study has shown that CCR and PER genomes are closer to RNO genome than to the genome of MMU. RNO karyotype can be converted in CCR karyotype by, at least, 31 rearrangements (19 fusions, 9 fissions, 2 inversions and 1 pericentric inversion), but MMU karyotype requires at least 50/53 rearrangements (27/30 fusions, 18/21 fissions, 4/1 inversion(s) and 1 pericentric inversion) to be transformed in CCR karyotype. A total of 14 rearrangements (5

fusions, 8 fissions and 1 inversion) are necessary to convert RNO in PER karyotype, and 31/32 (12/13 fusions, 16/17 fissions and 3/2 inversions) to transform MMU in PER.

Constitutive heterochromatin and evolutionary breakpoint regions

Several studies discuss the involvement of CH regions in evolutionary chromosomal rearrangements, suggesting that these regions represent hotspots for chromosomal reshuffling (e.g. Yunis and Yasmineh, 1971; Peacock et al., 1982; John, 1988; Louzada et al., 2008; Adega et al. 2009). Taking this into account, a detailed analysis of CH and its colocalization with evolutionary breakpoint regions (EBr) was accomplished (figs. 2 and 3, supplementary table S2). Concerning PER genome, the majority of the EBr revealed by the AMK and by RNO and MMU probes are colocalized with PER C-bands, reinforcing the putative association EBr/CH. Although, as mentioned, *P. eremicus* karyotype harbors high amounts of CH (Paço et al., 2009) detected by the presence of 112 C-bands and from these, only few colocalize with EBr, fact that supports the idea that additions of CH seems to be a feature of the karyotypic evolution within the *Peromyscus* lineage, as proposed by other authors (Greenbaum and Baker, 1978).

In what respects to CCR genome the majority of CH is located in a breakpoint region, although not all EBr colocalize with CH. This fact may be explained by the occurrence of constitutive heterochromatin elimination at the fusion sites during CCR karyotype evolution. Loss of heterochromatin at the fusion breakpoints have been associated to tandem translocations, possibly to stabilize the rearrangement and ensure its fixation in the speciation process (Elder, 1980; Yang et al., 1997). In fact, our results showed that fusions were the most frequent rearrangements occurred in CCR genome evolution.

Our results also reveal that in the majority of the ancestral Muroidea chromosomes (AMC) that undergone a fission, at least one of the fissioned parts is located near a centromere in both CCR and PER genomes. An association between centromeres and EBr reuse has been found (Murphy et al., 2005; Longo et al., 2009; Cazaux et al., 2011). It seems that these evolutionary breakages preferentially occur at sites of ancestral centromeres or neocentromeres in independent lineages; or, alternatively, reused breakpoint regions are unstable chromosomal sites that, after breakage, will tend to form a new centromere or telomere (Murphy et al., 2005).

In summary, a detailed overview of the two Cricetidae species genome architectures was performed. Taking in account that several breakpoint regions have been characterized as rich in repetitive sequences (Longo et al. 2009), the main constituents of constitutive heterochromatin (Petrović and Plohl 2005; Adega et al. 2009), we analyzed the correlation EBr/CH. The high level

of colocalization revealed in both genomes suggests the involvement of constitutive heterochromatin regions in these species karyotype restructuring despite the different level of karyotypic evolution, namely, conserved in *Peromyscus eremicus* and derivative in *Cricetus cricetus*.

ACKNOWLEDGMENTS

This work was supported by the project POCI/BIA-BCM/58541/2004, a research position on Animal Genomics of the "Sistema Científico e Tecnológico Nacional - Ciência 2007" and a PhD grants SFRH/BD/25813/2005 and SFRH/BD/41942/2007, all from the Science and Technology Foundation (FCT) from Portugal. We are grateful to Dr. Vitaly Volobouev for providing the Rodentia species cell cultures.

REFERENCES

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.

Adkins RM, Walton AH, Honeycutt RL (2003) Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. Mol Phylogenet Evol 26: 409–420.

Catzeflis FM, Aguilar J-P, Jaeger J-J: Muroid rodents (1992) Phylogeny and evolution. Trends Ecol Evol 7: 122–126.

Catzeflis FM, Dickerman AW, Michaux J, Kirsch JAW: DNA hybridization and rodent phylogeny; in Szalay FS, Novacek MJ, McKenna MC (eds): Mammal Phylogeny, pp 159 –172 (Springer-Verlag, New York, 1993).

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

Cazaux B, Catalan J, Veyrunes F, Douzery E, Britton-Davidian J (2011) Are ribosomal DNA clusters rearrangement hotspots? A case study in the genus *Mus* (Rodentia, Muridae). BMC Evolutionary Biology 11:124.

Chaline J, Mein P, Petter F(1977) Les grandes lignes d'une classification évolutive des Muroidea. Mammalia 41: 245–252.

Chaves R, Louzada S, Meles S, Adega F (2012) *Praomys tullbergi* genome (chromosome) architecture decoded by the related genomes *Mus* and *Rattus*. 20(6):673-683.

Deaven LL, Vidal-Rioja L, Jett JH, Hsu TC (1977) Chromosomes of *Peromyscus* (Rodentia, Cricetidae). VI. The genomic size. Cytogenet Cell Genet 19: 241–249.

Dubois J-Y, Rakotondravony D, Hänni C, Sourrouille P, Catzeflis FM (1996) Molecular evolutionary relationships of three genera of Nesomyinae, endemic rodent taxa from Madagascar. J Mamm Evol 3: 239–260.

Dubois J-Y, Catzeflis FM, Beintema JJ (1999) The phylogenetic position of "Acomyinae" (Rodentia, Mammalia) as sister group of a Murinae + Gerbillinae clade: evidence from the nuclear ribonuclease gene. Mol Phylogenet Evol 13: 181–192.

Elder FFB (1980) Tandem fusion, centric fusion, and chromosomal evolution in the cotton rat, genus Sigmodon. Cytogenet Cell Genet 26: 199–210.

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

Froenicke L, Lyons AL: Hotspots of mammalian chromosome evolution; in Encyclopedia of Life sciences (ELS) (John Wiley & Sons, Ltd, Chichester, 2008).

Gamperl R, Vistorin G, Rosenkranz W (1976) A comparative analysis of the karyotypes of *Cricetus cricetus* and Cricetulus griseus. Chromosoma 55: 259-265.

Glasper ER, DeVries AC (2005) Social structure influences effects of pair-housing on wound healing. Brain Behav Immun 19: 61-68.

Greenbaum IF, Baker RJ (1978) Determination of the primitive karyotype for *Peromyscus*. J Mammal 59: 820–834.

Hanon EA, Routledge K, Dardente H, Masson-Pévet M, Morgan PJ, Hazlerigg DG (2010) Effect of Photoperiod on the Thyroid-Stimulating Hormone Neuroendocrine System in the European Hamster (*Cricetus cricetus*). Journal of Neuroendocrinology 22: 51–55.

Hass I, Sbalqueiro IJ, Müller S (2008) Chromosomal phylogeny of four Akodontini species (Rodentia, Cricetidae) from Southern Brazil established by Zoo-FISH using *Mus musculus* (Muridae) painting probes. Chromosome Res 16: 75-88.

Hass I, Müller S, Artoni RF, Sbalqueiro IJ (2011) Comparative chromosome maps of neotropical rodents Necromys lasiurus and Thaptomys nigrita (Cricetidae) established by ZOO-FISH. Cytogenet Genome Res 135: 42–50.

Hugueney M, Mein P (1993) A comment on the earliest Spalacinae (Rodentia, Muroidea). J Mamm Evol 1: 215–223.

Jacobs LL, Flynn LJ, Downs WR: Neogene rodents of southern Ásia; in Black CC, Dawson MR (eds): Papers on Fossil Rodents in Honor of Albert Elmer Wood. Science, Series No. 33, pp 157–177 (Nat. Hist. Mus. Los Angeles County, Los Angeles, 1989).

Jansa SA, Goodman SM, Tucker PK (1999) Molecular phylogeny and biogeography of the native rodents of Madagascar (Muridae: Nesomyinae): a test of the single-origin hypothesis. Cladistics 15: 253–270.

John B: The biology of heterochromatin; in Verma RS (ed): Heterochromatin: molecular and structural aspects, pp 1–128 (Cambridge University Press, Cambridge, (1988).

Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D: Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. Proc Natl Acad Sci USA 100: 11484-11489 (2003).

Lemaitre C, Tannier E, Gautier C, Sagot M-F (2008) Precise detection of rearrangement breakpoints in mammalian chromosomes. BMC Bioinformatics 9:286.

Longo MS, Carone DM, NISC Comparative Sequencing Program, Green ED, O'Neill MJ, O'Neill RJ (2009) Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. BMC Genomics 10: 334.

Louzada S, Paço A, Kubickova S, Adega F, Guedes-Pinto H, et al (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.

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 analyses. 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.

Matthey R (1952) Chromosomes de Muridae (Microtinae et Cricetinae). Chromosoma 5: 113-138.

Michaux J, Reyes A, Catzeflis F (2001) Evolutionary history of the most speciose mammals: molecular phylogeny of muroid rodents. Mol Biol Evol 18: 2017–2031.

Miller GS, Gidley JW (1911) Synopsis of supergeneric groups of rodents. J Wash Acad Sci 8: 431–448.

Mlynarski EE, Obergfell CJ, Rens W, O'Brien PCM, Ramsdell CM, et al (2008) Peromyscus maniculatus – Mus musculus chromosome homology map derived from reciprocal cross species chromosome painting. Cytogenet Genome Res 121: 288-292.

Murphy WJ, Stanyon R, O'Brien SJ (2001) Evolution of mammalian genome organization inferred from comparative gene mapping. Genome Biol 2: 1-8.

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, et al (2005) Dynamics of Mammalian Chromosome Evolution Inferred from Multispecies Comparative Maps. Science 309: 613-617.

Musser G, Carleton MD: Superfamily Muroidea; in Wilson DE, Reeder DM (eds): Mammal Species of the World: a Taxonomic and Geographic Reference, 3rd ed, pp 894–1531 (Johns Hopkins University Press, Baltimore, Maryland, 2005).

Nadeau JH, Taylor BA (1984) Lengths of chromosomal segments conserved since divergence of man and mouse. Proc Natl Acad Sci USA 81: 814-818.

Nakamura T, Kuroiwa A, Nishida-Umehara C, Matsubara K, Yamada F, Matsuda Y (2007a) Comparative chromosome painting map between two Ryukyu spiny rat species, Tokudaia osimensis and Tokudaia tokunoshimensis (Muridae, Rodentia). Chromosome Res 15: 799–806.

Nakamura T, Matsubara K, Yasuda SP, Tsuchiya K, Matsuda Y (2007b) Chromosome homology between mouse and three Muridae species, Millardia meltada, Acomys dimidiatus and Micromys minutus, and conserved chromosome segments in murid karyotypes. Chromosome Res 15: 1023–1032.

Nilsson S, Helou K, Walentinsson A, Szpirer C, Nerman O, Ståhl F (2001) Rat–Mouse and Rat–Human Comparative Maps Based on Gene Homology and High-Resolution Zoo-FISH. Genomics 74: 287–298.

Paço A, Adega F, Guedes-Pinto H, Chaves R: Hidden heterochromatin (2009) Characterization in the Rodentia species *Cricetus cricetus*, *Peromyscus eremicus* (Cricetidae) and *Praomys tullbergi (*Muridae). Genet Mol Biol 32: 58-68.

Pathak S, Hsu TC, Arrighi FE (1973)Chromosomes of Peromyscus (Rodentia, Cricetidae). IV. The role of heterochromatin in karyotypic evolution. Cytogenet Cell Genet 12: 315–326.

Peacock WJ, Dennis ES, Gerlach WL (1982) DNA sequence changes and speciation. Prog Clin Biol Res 96: 123–142.

Petrović V, Plohl M (2005) Sequence divergence and conservation in organizationally distinct subfamilies of Donax trunculus satellite DNA. Gene 362: 37-43.

Rabbitts P, Impey H, Heppell-Parton A, Langford C, Tease C, et al (1995) Chromosome specific paints from a high resolution flow karyotype of the mouse. Nat Genet 9: 369–375.

Rambau RV, Robinson TJ (2003) Chromosome painting in the African four-striped mouse Rhabdomys pumilio: detection of possible murid specific contiguous segment combinations. Chromosome Res 11: 91-98.

Ramsdell CM, Lewandowski AA, Glenn JLW, Vrana PB, O'Neill RJ, Dewey MJ (2008) Comparative genome mapping of the deer mouse (*Peromyscus maniculatus*) reveals greater similarity to rat (*Rattus norvegicus*) than to the lab mouse (Mus musculus). BMC Evol Biol 8: 65.

Robinson M, Catzeflis F, Briolay J, Mouchiroud D (1997) Molecular Phylogeny of Rodents, with Special Emphasis on Murids: Evidence from Nuclear Gene LCAT. Mol Phylogenet Evol 8: 423-434.

Robinson TJ, Ruiz-Herrera A, Froenicke L (2006) Dissecting the mammalian genome – new insights into chromosomal evolution. Trends Genet 22: 297-301.

Romanenko SA, Perelman PL, Serdukova NA, Trifonov VA, Biltueva LS, et al (2006) Reciprocal chromosome painting between three laboratory rodent species. Mamm Genome 17: 1183-1192.

Romanenko SA, Volobouev VT, Perelman PL, Lebedev VS, Serdukova NA, et al (2007a) Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroidea, Rodentia) inferred from chromosomal painting and banding comparison. Chromosome Res 15: 283-297.

Romanenko SA, Sitnikova NA, Serdukova NA, Perelman PL, Rubtsova NV, et al (2007b)Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). II. The genome homology of two mole voles (genus Ellobius), the field vole and golden hamster revealed by comparative chomosome painting. Chromosome Res 15: 891-897.

Romanenko SA, Perelman PL, Trifonov VA, Graphodatsky AS: Chromosomal evolution in Rodentia. Heredity (Edinb) 108: 4-16 (2012).

Ruiz-Herrera A, Castresana J, Robinson TJ (2006) Is mammalian evolution driven by regions of genome fragility? Genome Biol 7:R115.

Sitnikova NA, Romanenko SA, O'Brien PCM, Perelman PL, Fu B, et al (2007) Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). I. The genome homology of tundra vole, field vole, mouse and golden hamster revealed by comparative chromosome painting. Chromosome Res 15: 447-456.

Stanyon R, Yang F, Cavagna P, O'Brien PCM, Bagga M, Ferguson-Smith MA, 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, Yang F, Morescalchi AM, Galleni L (2004) Chromosome painting in the long-tailed field mouse provides insights into the ancestral murid karyotype. Cytogenet Genome Res 105:406-411.

Steppan SJ, Adkins RM, Anderson J (2004) Phylogeny and Divergence-Date Estimates of Rapid Radiations in Muroid Rodents Based on Multiple Nuclear Genes. Syst Biol 53: 533–553.

The Committee for Standardization of Chromosomes of *Peromyscus*: Standardized karyotype of deer mice, *Peromyscus* (Rodentia) (1977) Cytogenet Cell Genet 19: 38–43.

Veyrunes F, Catalan J, Sicard B, Robinson TJ, Duplantier J-M, et al (2004)Autosome and sex chromosome diversity among the African pygmy mice, subgenus Nannomys (Muridae; Mus). Chromosome Res 12: 369–382.

Veyrunes F, Dobigny G, Fengtang Y, O'Brien PCM, Catalan J, et al (2006) Phylogenomics of the genus Mus (Rodentia; Muridae): extensive genome repatterning is not restricted to the house mouse. Proc R Soc B 273: 2925–2934.

Wienberg J, Stanyon R, Nash WG, O'Brien PC, Yang F, et al (1997) Conservation of humans vs. feline genome organization revealed by reciprocal chromosome painting. Cytogenet Cell Genet 77: 211-217.

Yang F, O'Brien PCM, Wienberg J, Ferguson-Smith MA (1997) 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 PCM, Ferguson-Smith MA (2000) Comparative chromosome map of the laboratory mouse and Chinese hamster defined by reciprocal chromosome painting. Chromosome Res 8: 219-227.

Yunis JJ, Yasmineh WG (1971) Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation. Science 174: 1200–1209.

SUPLEMENTARY TABLES

Table S1 – Number of syntenic segments revealed by MMU and RNO paint probes.

MMU paints	CCR	PER	RNO paints	CCR	CCR
1	3(5)	3(5)	**1**	3(9)	3(8)
2	2(5)	2(4)	**2**	2(6)	2(4)
3	1(5)	1(3)	**3**	1(4)	1(4)
4	1(3)	1(4)	**4**	2(4)	1(5)
5	3(5)	3(4)	**5**	1(3)	1(4)
6	2(4)	1(5)	**6**	2(6)	2(5)
7	2(6)	1(4)	**7**	3(4)	2(4)
8	2(5)	2(4)	**8**	1(2)	1(3)
9	1(3)	1(3)	**9**	2(3)	2(4)
10	3(4)	3(3)	**10**	1(3)	1(2)
11	2(2)	2(2)	**11**	1(2)	1(2)
12	2(5)	2(4)	**12**	1(1)	1(2)
13	2(5)	2(3)	**13**	1(2)	1(2)
14	1(2)	1(2)	**14**	1(3)	1(2)
15	2(4)	2(4)	**15**	1(1)	1(3)
16	2(3)	1(2)	**16**	1(3)	2(2)
17	4(7)	4(5)	**17**	1(2)	1(1)
18	1(3)	1(1)	**18**	1(3)	1(1)
19	1(1)	1(3)	**19**	1(3)	1(3)
			20	2(3)	2(2)
X	1(3)	1(8)	**X**	1(3)	1(8)
Y	-	1(1)	**Y**	_	1(1)

The number displayed in the columns of the table means the number of chromosome(s) where MMU/RNO paints hybridized and the number in parentheses means the number of segments that the respective probe outlined.

Table S2 – Colocalization between constitutive heterochromatin and evolutionary breakpoint regions (delineated by both index species and by the ancestral Muroidea karyotype), in *C. cricetus* and *P. eremicus*.

	EBr	**Colocalization EBr/CH(%)**	**Colocalization CH/EBr(%)**
C. cricetus	delineated by AMK	57%	34%
	delineated by MMU	49%	77%
	delineated by RNO	49%	68%
P. eremicus	delineated by AMK	67%	2%
	delineated by MMU	71%	30%
	delineated by RNO	64%	25%

EBr – Evolutionary breakpoint regions; CH- Constitutive heterochromatin. The constitutive heterochromatin was accessed by counting the C-bands. EBr/CH- Percentage of evolutionary breakpoint regions that are colocalized with C-bands; CH/EBr- Percentage of C-bands that are colocalized with evolutionary breakpoint regions.

Table S3 – MMU syntenic associations present in *C. cricetus* genome and their presence/absence in other Muroidea genomes.

CRICETIDAE															MURIDAE																			
SubFamily Cricetinae			SubFamily Neotominae		SubFamily Sigmodontinae							SubFamily Arvicolinae			SubFamily Deomynae	SubFamily Murinae																		
(Yang et al., 2000; Romanenko et al., 2006, 2007a)			(this paper; Mlynarski et al., 2008)		(Hass et al., 2008, 2011)							(Romanenko et al., 2007b; Sitnikova et al., 2007)			(Nakamura et al., 2007b)	*Millardia* group? (Nakamura et al., 2007b)	*Micromys* group? (Nakamura et al., 2007b)	*Praomys* group? (Chaves et al., 2012)	*Rattus* group (Cavagna et al., 2002; Ensembl database*)		*Mus* group (Matsubara et al., 2003; Veyrunes et al., 2006)			*Arvicanthis* Group (Rambau and Robinson, 2003; Engelbrecht et al., 2006)		*Apodemus* group (Matsubara et al., 2004; Nakamura et al., 2007a)								
CCR (MMU syntenic associations)	CGR	MAU	PER	PMA	ACU	AMO	APA	ASE	OFL	NLA	TNI	MOE	ELU	ETA	ADI	MME	MMI	PTU	RNO	RRA	CPA	NMA	MPL	OIR	RPU	TOS	TTO	APE	AGU	ASY	AAG	ASE	ASP	AAR
MMU1/10	✓	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU1/15	✓	✓	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-
MMU1/17	✓	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU2/8	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU2/13	✓	✓	✓	✓	-	-	-	-	-	-	-	✓	✓	-	✓	-	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU2/17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-
MMU3/10	✓	✓	-	-	-	-	✓	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU4/18	✓	✓	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU5/6	-	-	✓	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	✓	✓	-	✓	✓	✓	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	-
MMU5/9	✓	-	-	-	-	-	-	-	-	✓	-	-	-	-	✓	-	-	-	✓	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-
MMU5/11	✓	-	✓	✓	-	-	-	-	-	-	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU5/14	✓	✓	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU5/16	✓	✓	-	-	-	-	-	-	-	-	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU6/10	✓	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU7/8	✓	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU7/19	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU8/11	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU8/17	✓	-	-	-	-	✓	-	-	-	✓	-	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU10/17	✓	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU10/18	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU11/17	✓	-	✓	✓	-	-	-	-	-	-	-	✓	✓	✓	-	-	-	✓	✓	-	-	-	-	-	✓	-	-	✓	✓	✓	✓	✓	✓	✓
MMU12/16	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU12/17	✓	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU13/15	✓	✓	✓	✓	-	-	-	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	✓	✓	✓	✓	✓	✓	✓	✓
MMU13/17	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU16/17	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	✓	-	-	-	✓	✓	✓	-	-	✓	-	-	✓	-	✓	✓	✓	✓	✓	✓	✓
26 (shared syntenies)	24	14	12	12	8	2	2	1	2	4	2	11	14	11	9	4	7	9	11	10	7	5	4	5	6	7	7	9	9	9	9	9	9	8

Table S4 - MMU syntenic associations present in *P. eremicus* genome and their presence/absence in other Muroidea genomes.

CRICETIDAE																MURIDAE
SubFamily Neotominae		SubFamily Cricetinae			SubFamily Sigmodontinae							SubFamily Arvicolinae			SubFamily Deomynae	
(Mlynarski et al., 2008		(this paper; Yang et al., 2000; Romanenko et al., 2006, 2007a)			(Hass et al., 2008, 2011)							(Romanenko et al., 2007b; Simikova et al., 2007)			(Nakamura et al., 2007b)	
PER (MMU syntenic associations)	PMA	CCR	CGR	MAU	ACU	AMO	APA	ASE	OFL	NLA	TNI	MOE	ELU	ETA	ADI	
MMU1/10	✓	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	-	
MMU1/17	✓	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	✓	
MMU2/8	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	
MMU2/13	✓	✓	✓	✓	-	-	-	-	-	-	-	✓	✓	-	✓	
MMU5/6	✓	✓	-	-	✓	-	-	-	✓	-	-	-	-	-	-	
MMU5/11	✓	✓	✓	-	-	-	-	-	-	-	-	✓	✓	✓	✓	
MMU7/19	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
MMU10/12	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	
MMU10/17	✓	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	-	
MMU11/17	✓	✓	✓	-	-	-	-	-	-	-	-	✓	✓	✓	-	
MMU12/17	✓	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	✓	
MMU13/15	✓	✓	✓	✓	-	-	-	-	-	-	-	✓	✓	✓	✓	
MMU16/17	✓	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	✓	-	
13 (shared syntenies)	12	12	9	7	7	1	1	1	2	1	2	10	10	9	6	

	MURIDAE																		
	SubFamily Murinae																		
	Millardia group?	*Micromys* group?	*Praomys* group?	*Rattus* group		*Mus* group			*Arvicanthis* group		*Apodemus* group								
	(Nakamura et al., 2007b)	(Nakamura et al., 2007b)	(Chaves et al., 2012)	(Cavagna et al., 2002; Ensembl database)*		(Matsubara et al., 2003; Veyrunes et al., 2006)			(Rambau and Robinson, 2003; Engelbrecht et al., 2006)		(Matsubara et al., 2004; Nakamura et al., 2007a)								
PER (MMU syntenic associations)	MME	MMI	PTU	RNO	RRA	CPA	NMA	MPL	OIR	RPU	TOS	TTO	APE	AGU	ASY	AAG	ASE	ASP	AAR
MMU1/10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU1/17	✓	✓	✓	✓	✓	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU2/8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU2/13	-	✓	✓	✓	✓	✓	-	-	✓	-	-	-	-	-	-	-	-	-	-
MMU5/6	-	✓	✓	-	✓	✓	✓	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	-
MMU5/11	✓	-	✓	✓	✓	✓	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU7/19	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU10/12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMU10/17	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU11/17	-	-	✓	✓	-	-	-	-	-	✓	-	-	✓	✓	✓	✓	✓	✓	✓
MMU12/17	✓	✓	✓	-	✓	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MMU13/15	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	✓	✓	✓	✓	✓	✓	✓	✓
MMU16/17	-	-	✓	✓	✓	-	-	✓	-	-	✓	-	✓	✓	✓	✓	✓	✓	✓
13 (shared syntenies)	5	6	9	8	9	6	4	4	5	5	7	7	9	9	9	9	9	9	8

For both presented tables (✓) represents Presence and (-) represents Absence of the syntenic association.
PER - *Peromyscus eremicus*; PMA - *Peromyscus maniculatus*; CCR – *Cricetus cricetus*; CGR - *Cricetulus griseus*; MAU - *Mesocricetulus auratus*; ACU - *Akodon cursor*; AMO - *Akodon montensis*; APA - *Akodon paranaensis*; ASE - *Akodon* serrensis; OFL - *Olygoryzomis flavescens*; NLA – *Necromys lasiurus*; TNI – *Thaptomys nigrita*; MOE - *Microtus oeconomus*; ELU – *Ellobius lutescens*; ETA – *Ellobius talpinus*; ADI – *Acomys dimidiatus*; MME – *Millardia meltada*; MMI – *Micromys minutus*; PTU - *Praomys tullbergi*; RNO - *Rattus norvegicus*; RNA - *Rattus rattus*; CPA – *Coelomys pahari*; NMA - *Nannomys mattheyi*; MPL - *Mus platythrix*; OIR - *Otomys irroratus*; RPU - *Rabdhomys pumilio*; TOS - *Tokudaia osimensis*; TTO – *Tokudaia tokunoshimensis*; APE - *Apodemus peninsulae*; AGU - *Apodemus gurkha*; ASY - *Apodemus sylvaticus*; AAG - *Apodemus agrarius*; ASE - *Apodemus semotus*; ASP - *Apodemus speciosus*; AAR - *Apodemus argenteus*
* - Ensembl database (http://www.ensembl.org/; release 68 - July 2012)

2. SATELLITE DNA

The characterization of satellite DNA sequences has shown that these highly repetitive sequences are responsible for genome plasticity and its dynamics has been correlated with chromosome restructuring.

With the objective of investigating such interesting characteristics of satDNA and their involvement in rodent karyotype restructuring, different sequences were isolated from the genome of rodent species which had few studies in this thematic. This investigation resulted in two papers. Two repetitive sequences – CCR4/10sat and PMSat - were isolated *de novo,* using laser microdissection, from *Cricetus cricetus* and *Peromyscus eremicus* genome, respectively.

Several studies in satDNA have shown that these sequences evolve in concerted fashion and present specific modes of evolution promoted by their ability to change in copy number and to mobilize through the genome. CCR4/10sat orthologous sequences were identified in *P. eremicus* genome assuming different chromosome localization. It is proposed that these sequences moved throughout the genome by means of intragenomic movements, passing from a centromeric location, as in *C. cricetus*, and assuming a scattered pattern in *Peromyscus eremicus*.

PMSat orthologous sequences were found in *C. cricetus*, *P. sungorus* and *M. arvalis* genomes, presenting a high interspecies sequence identity but differing significantly in copy number. These characteristics seem to indicate that its evolutionary pathway occurred throughout copy number variation, culminating in different profiles. An integrated analysis of this satDNA dynamics with species evolutionary rearrangements was performed.

The major outcomes of these works were: 1) the isolation of repetitive sequences in the genomes of CCR and PER; 2) the identification of orthologous sequences in related species which places them in a common ancestor; the analysis of the different characteristics presented in each species (different location or differences in copy number) revealing a highly dynamic behavior.

2.1. Different evolutionary trails in the related genomes *Cricetus cricetus* and *Peromyscus eremicus* (Rodentia, Cricetidae) uncovered by orthologous satellite DNA repositioning

Micron 39 (2008) 1149–1155

Contents lists available at ScienceDirect

Micron

journal homepage: www.elsevier.com/locate/micron

Different evolutionary trails in the related genomes *Cricetus cricetus* and *Peromyscus eremicus* (Rodentia, Cricetidae) uncovered by orthologous satellite DNA repositioning

Sandra Louzada [a], Ana Paço [a], Svatava Kubickova [b], Filomena Adega [a], Henrique Guedes-Pinto [a], Jiri Rubes [b], Raquel Chaves [a,*]

[a] *Institute for Biotechnology and Bioengineering, Centre of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGB-UTAD), Apdo 1013, 5001-801 Vila Real, Portugal*
[b] *Veterinary Research Institute, Hudcova 70, 621 32 Brno, Czech Republic*

ARTICLE INFO

Article history:
Received 9 April 2008
Received in revised form 23 May 2008
Accepted 24 May 2008

Keywords:
Satellite DNA
Rodentia
Orthologous sequences
Constitutive heterochromatin
Genome evolutionary restructuring

ABSTRACT

Constitutive heterochromatin comprises a substantial fraction of the eukaryotic genomes and is mainly composed of tandemly arrayed satellite DNAs (satDNA). These repetitive sequences represent a very dynamic and fast evolving component of genomes. In the present work we report the isolation of *Cricetus cricetus* (CCR, Cricetidae, Rodentia) centromeric repetitive sequences from chromosome 4 (CCR4/10sat), using the laser microdissection and laser pressure catapulting procedure, followed by DOP-PCR amplification and labelling. Physical mapping by fluorescent *in situ* hybridisation of these sequences onto *C. cricetus* and another member of Cricetidae, *Peromyscus eremicus*, displayed quite interesting patterns. Namely, the centromeric sequences showed to be present in another *C. cricetus* chromosome (CCR10) besides CCR4. Moreover, these almost chromosome-specific sequences revealed to be present in the *P. eremicus* genome, and most interestingly, displaying a ubiquitous scattered distribution throughout this karyotype. Finally and in both species, a co-localisation of CCR4/10sat with constitutive heterochromatin was found, either by classical C-banding or C-banding sequential to *in situ* endonuclease restriction.

The presence of these orthologous sequences in both genomes is suggestive of a phylogenetic proximity. Furthermore, the existence of common repetitive DNA sequences with a different chromosomal location foresees the occurrence of an extensive process of karyotype restructuring somehow related with intragenomic movements of these repetitive sequences during the evolutionary process of *C. cricetus* and *P. eremicus* species.

1. Introduction

A substantial proportion of the higher eukaryotic genome consists of constitutive heterochromatin (CH) preferentially found in (peri)centromeric regions (see Corradini et al., 2007; Rossi et al., 2007), although telomeric and interstitial positions have also been described in different species (see Adega et al., 2007; Meles et al., 2008). This genomic fraction is mainly composed of highly repetitive sequences of satellite DNA (satDNA) (John, 1988; Chaves et al., 2004b), organised into long and uninterrupted tandem arrays of more or less well defined repeat units (Charlesworth et al., 1994).

In a general way, eukaryotic satDNA sequences are characterised by highly dynamic molecular behaviour, promoted by concerted evolution, which leads to rapid change between repeat sequences of different species, throughout sequence modification, amplification of new variants during speciation, and intragenomic movements (Ugarkovic and Plohl, 2002; Hamilton et al., 1992). This characteristic pattern of occurrence allows that some taxonomic groups enclose specific satDNA sequences, these sometimes being species-specific (Jobse et al., 1995; Nijman and Lenstra, 2001). Simultaneously, it is also recognized that different satDNA families can coexist in the same genome, forming a satDNA library (Salser et al., 1976; Fry and Salser, 1977). In some taxa, however, it has been observed that the evolution of satDNA families proceeds slowly (Mravinac et al., 2002; Li and Kirby, 2003; Cafasso et al., 2003), meaning that species separated by several million years may share orthologous repetitive sequences (Robles et al., 2004; Li and Kirby, 2003; Mravinac et al., 2002; Cafasso et al., 2003; Adega et al., 2008). These few cases of repetitive sequences conservation highlight the complex behaviour of this genome fraction.

* Corresponding author. Tel.: +351 259 350 841.
E-mail address: rchaves@utad.pt (R. Chaves).

0968-4328/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.micron.2008.05.008

The molecular analysis of repetitive sequences and their physical mapping in chromosomes of different species has shown its value as evolutionary markers (see Saffery et al., 1999; Lander et al., 2001; Ugarkovic and Plohl, 2002).

Given the dynamics of the majority of satDNA families, it is believed that these repetitive sequences play an important role in mammal genome evolution by promoting chromosomal rearrangements (see Wichman et al., 1991; Reig et al., 1992; Schluter et al., 1997; Slamovits et al., 2001). In accordance, several works discuss the involvement of constitutive heterochromatin (CH) regions in the occurrence of chromosomal evolution, suggesting that these regions act as hotspots that preferentially enable structural chromosome rearrangements (Yunis and Yasmineh, 1971; Peacock et al., 1982; John, 1988; Chaves et al., 2004b). Recent studies focused on molecular characterization of the breakpoint regions (see Garagna et al., 2001; Li et al., 2000; Locke et al., 2003; Kehrer-Sawatzki et al., 2005; Ruiz-Herrera et al., 2006) have demonstrated that the location of evolutionary breakpoint regions is coincident with the location of regions rich in repetitive sequences.

The C-banding technique is extremely useful in the identification of chromosomes' CH, however the location of CH determined by this technique, and the distribution of satDNA sequences ascertained by *in situ* hybridisation, are often, but not always coincident (reviewed by John, 1988). *In situ* restriction endonuclease (RE) digestion with sequential C-banding proved to be very useful in the understanding of the mechanisms involved in CH evolution in different genomes (see Gosálvez et al., 1997; Leitão et al., 2004; Chaves et al., 2004b; Adega et al., 2007). Moreover, this technique allows the identification of CH bands not always detected by classical C-banding, the cryptic C-bands (see Chaves et al., 2004b; Adega et al., 2005, 2007).

Two rodent species, the common hamster *Cricetus cricetus* (CCR), and the cactus mouse *Peromyscus eremicus* (PER) (Rodentia: Cricetidae), displaying diploid chromosome numbers of 22 and 48 chromosomes, respectively, were studied. *C. cricetus* enclose a nearly meta/submetacentric karyotype, whose CH seems to be mostly found at the (peri)centromeric regions, exhibiting in the majority of the chromosomes two very large CH blocks (Gamperl et al., 1976; unpublished data). *P. eremicus* exhibits a very distinct karyotype organisation, this being constituted solely by submetacentric chromosomes. This karyotype also displays great amounts of CH, especially located at the (peri)centromeric regions, the p arms of the majority of the chromosomes almost being entirely heterochromatic (Pathak et al., 1973; Deaven et al., 1977; unpublished data). Several cryptic C-bands were previously identified in both species chromosomes, by *in situ* RE digestion with sequential C-banding (unpublished data).

In the present work we report the isolation of CCR centromeric repetitive sequences using the laser microdissection and laser pressure catapulting procedure. The physical mapping of these sequences onto *C. cricetus* and *P. eremicus* chromosomes revealed very distinct patterns. The existence of common repetitive DNA sequences displaying different chromosomal locations in these two related genomes is discussed.

2. Materials and methods

2.1. Chromosome preparations

Metaphase chromosomal spreads were prepared from fibroblast cell lines of the rodents' species *C. cricetus* and *P. eremicus*, both part of the cell and tissue collection housed at the Department of Systematics and Evolution, Muséum National d'Histoire Naturelle – MNHN (Paris, France). Standard cell culture from both species was performed according to the method described by Chaves et al. (2004a). The nomenclature of *C. cricetus* ($2n = 22$) and *P. eremicus* ($2n = 48$) chromosomes is according to Gamperl et al. (1976) and the Committee for standardization of chromosomes of Peromyscus (1977), respectively.

2.2. GTD-banding

Air-dried slides were aged at 65 °C for 5 h or overnight and then submitted to standard G-banding procedures with trypsin (Sumner et al., 1971). As the chromosome slides proceeded sequentially to C-banding, they were fixed with formaldehyde. Briefly, dried slides were placed in 1 × PBS solution (2 × 5 min) followed by fixation in 3% formaldehyde (Sigma)/1× PBS (room temperature) for 20 min. Afterwards, the slides were dehydrated for 5 min in 70%, 90% and 100% chilled ethanol and air-dried. DAPI was used for staining (instead of routine Giemsa) in order to obtain a better contrast (Chaves et al., 2002). The inversion of DAPI colour in Adobe Photoshop revealed the chromosome G-banding pattern (GTD-banding, G-bands by trypsin with DAPI).

2.3. In situ RE digestion

Air-dried slides were aged at 65 °C for 6 h and then submitted to *in situ* RE digestion. The four restriction enzymes used (AluI, BamHI, PstI and RsaI) were diluted in buffers indicated by the manufacturer (Invitrogen Life Technologies), and final concentrations of 30 U per 100 μl were obtained. A total of 100 μl were placed on the slides and covered with coverslips. The slides were incubated in a moist chamber for 16 h at 37 °C. Control slides were also prepared according to the aforementioned procedures but only with buffer. The slides were washed in distilled water and air-dried. Once these slides proceeded to C-banding, they were fixed in formaldehyde, as described above for GTD-banding. Finally the slides were stained with DAPI (the inversion of the DAPI colour revealed the RE-banding).

2.4. CBP-banding sequential to G-bands or RE-bands

C-banding technique was performed sequentially to G-banding or RE-banding, being performed after distaining the slides. CBP-banding [C-bands by Barium Hydroxide using Propidium Iodide (PI)] was done according to the standard procedure of Sumner (1972) with slight modifications. Briefly, the slides were submitted to hydrochloric acid (0.1 M) 20 min, barium hydroxide (5% solution) 7 min and 2× saline solution citrate (2× SSC: 0.3 mol/ NaCl, 0.03 mol/l sodium citrate) at 60 °C for 40 min. The slides where then counterstained with PI (1.5 μl/ml).

2.5. Microdissection and probe preparation

The PALM MicroLaser system (P.A.L.M. GmbH, Bernried, Germany) was used for chromosome dissection and collection. The referred system consists of a 337-nm nitrogen laser coupled to the light path of an inverted microscope (Olympus) and focused through an oil immersion objective (100× magnification), with high numerical aperture to yield a spot size of less than 1 μm in diameter. About 10 chromosome centromeres from *C. cricetus* chromosome 4 were microdissected and catapulted by a single laser pulse directly into the cap of a PCR tube, to which 2 μl PCR oil had been placed. The microdissected material was then dissolved in 20 μl 10 mmol/l Tris–HCl pH 8.8 in the cap, placed in the respective tube and submitted to centrifugation. Probes were generated and labelled with digoxigenin-11-d'UTP (Roche, Molecular Biochemicals) by DOP-PCR, as described by Kubickova et al. (2002).

2.6. Fluorescent in situ hybridisation

Chromosome preparations from *C. cricetus* and *P. eremicus* were aged at 65 °C overnight. Afterwards they were fixed using 3% formaldehyde (Sigma)/1× PBS (room temperature) and dehydrated with chilled ethanol, as described previously (before C-banding procedure). Hybridisation was carried out in a moist chamber in 2× SSC and 50% formamide at 37 °C (overnight for *C. cricetus* slides and during two days for *P. eremicus*), and the most stringent post-hybridisation wash was 50% formamide/2× SSC at 37 °C, allowing sequences with more than 77% similarity to remain hybridised. Digoxigenin-labelled probes were detected with anti-digoxigenin-5'TAMRA (Roche, Molecular Biochemicals).

2.7. Chromosome observation

Chromosomes were observed with a Zeiss Axioplan 2 Imaging microscope, coupled to an Axiocam digital camera with AxioVision software (version Rel. 4.5 – Zeiss). Digitised photos were prepared for printing in Adobe Photoshop (version 5.0); contrast and colour optimization were the functions used and affected the whole of the image equally.

3. Results

3.1. Isolation of CCR4/10sat sequences

A centromeric probe was isolated from *C. cricetus* chromosome 4, using the laser microdissection and laser pressure catapulting procedure (Kubickova et al., 2002) which permitted a precise and efficient cut and collection of ten CCR4 centromeres. The isolated material was submitted to DOP-PCR, which enabled the amplification and labelling of the sequences mixture present in CCR4 centromere.

3.2. Physical analysis in C. cricetus and P. eremicus

Physical mapping of CCR4/10sat in the genomes of the Cricetidae species, *C. cricetus* and *P. eremicus,* was carried out by fluorescent *in situ* hybridisation. The hybridisation signals obtained in both species chromosomes displayed very different outcomes. In *C. cricetus* chromosome preparations, a strong hybridisation signal in the (peri)centromeric region of chromosomes CCR4 and CCR10 was observed. As discerned in Fig. 1a, all the chromosomes, except the two referred to above, apparently

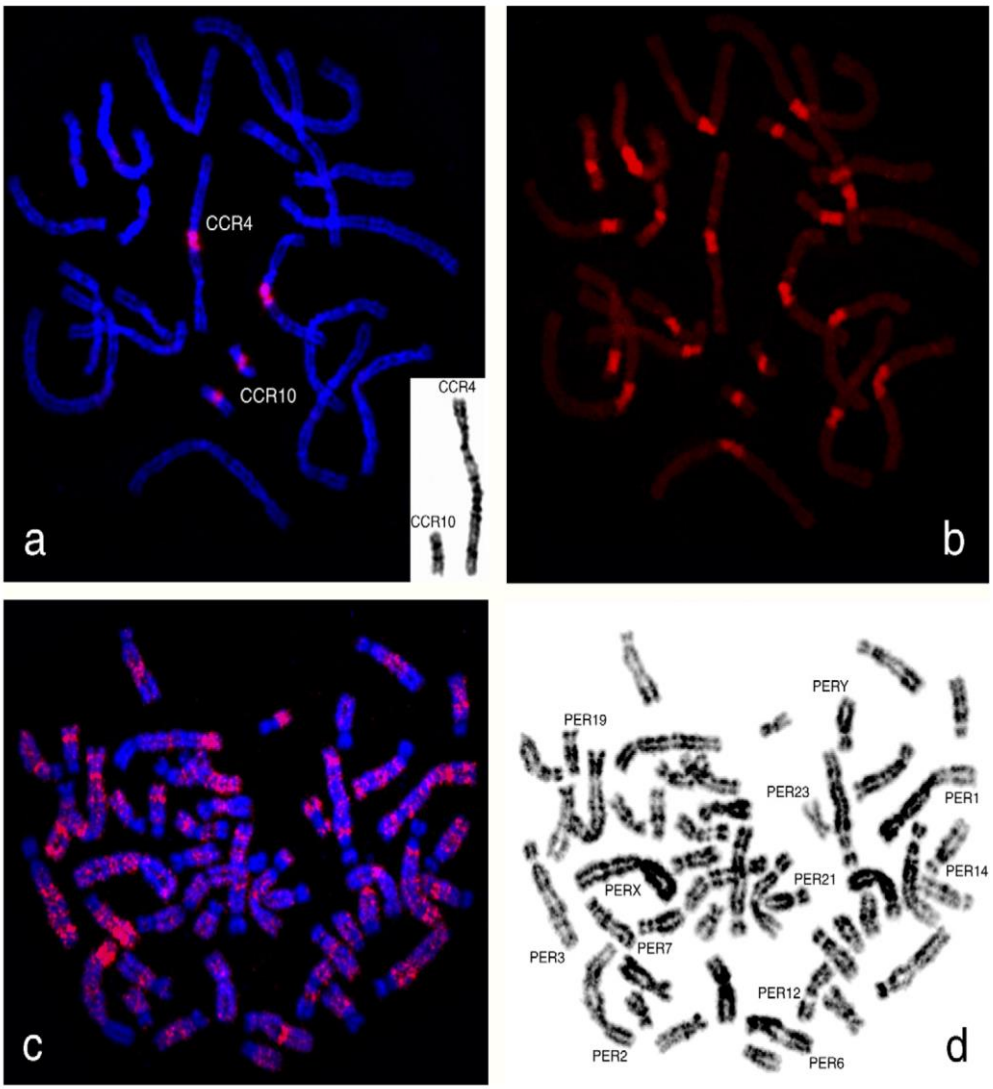


Fig. 1. Representative *in situ* hybridisation of CCR4/10sat sequences from *Cricetus cricetus* (CCR) onto *C. cricetus* chromosomes. CCR4 and CCR10 were DAPI inverted for chromosome identification (a). The same metaphase after C-banding (b). *In situ* hybridisation of CCR4/10sat onto *Peromyscus eremicus* chromosomes (c). The same metaphase was DAPI inverted for chromosome identification, where some of the most interesting chromosomes are identified (d).

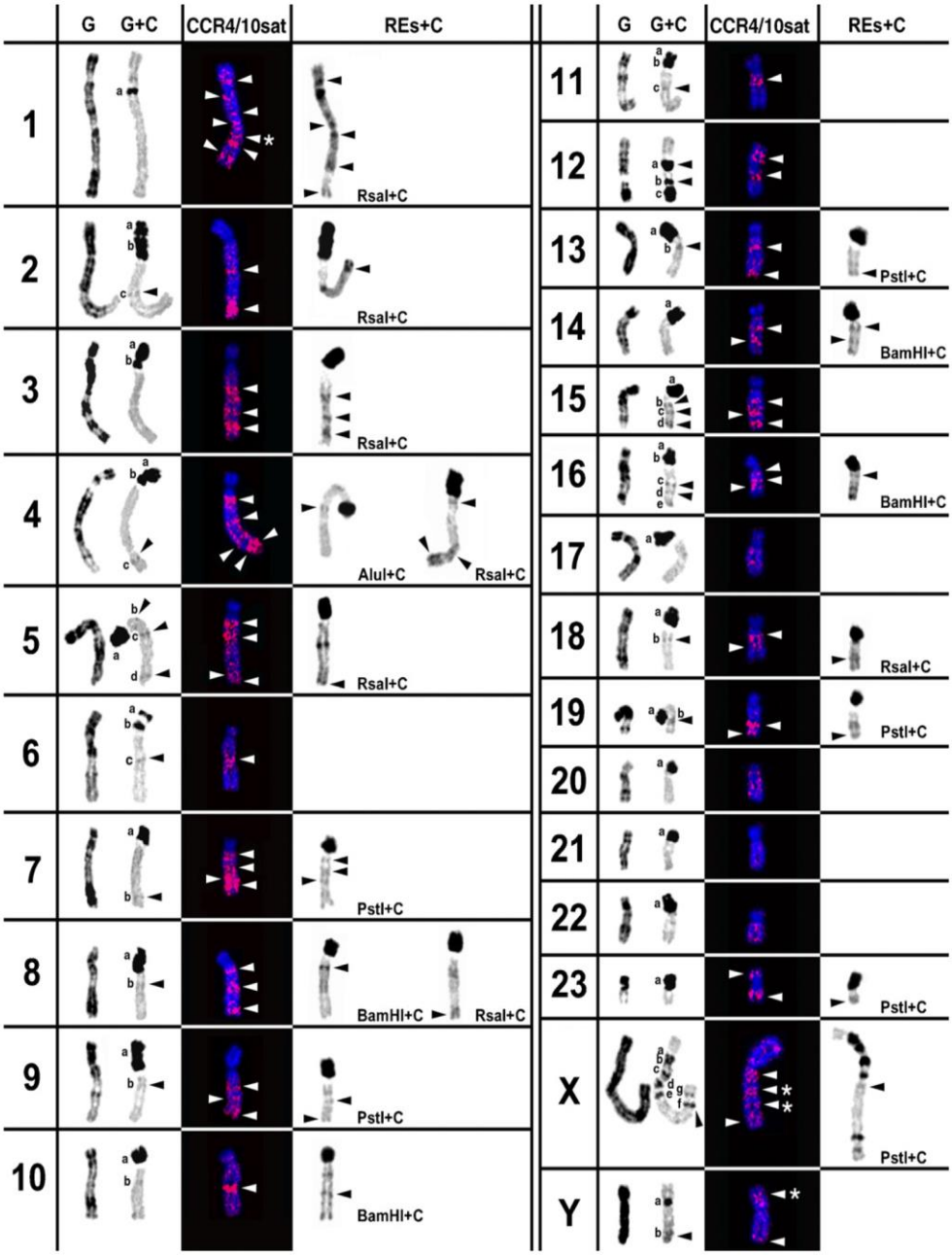


Fig. 2. Table showing the *in situ* hybridisation pattern of CCR4/10sat in *Peromyscus eremicus* chromosomes. G- and C-banding of each *P. eremicus* chromosomes are shown in the left column. The letters (a–g) represent the C-bands according to order of appearance in each chromosome. In the other columns it is possible to observe the hybridisation pattern of the CCR4/10sat probe and the constitutive heterochromatin bands produced by *in situ* restriction endonuclease digestion followed by C-banding (RE + C-banding). The black arrowheads indicate classical and cryptic C-bands that co-localize with CCR4/10sat sequences. The white arrowheads evidence the CCR4/10sat signal in *P. eremicus* chromosomes.

lack hybridisation signal, suggesting that the isolated sequences are specific for CCR chromosomes 4 and 10, indicating a certain chromosome specificity. Moreover, when analysing this species' constitutive heterochromatin evidenced by classical C-banding (Fig. 1b), it is possible to verify that the isolated probe co-localizes with (peri)centromeric CH, suggesting its repetitive nature.

The chromosomal distribution of the CCR4/10sat probe in *P. eremicus* genome revealed interesting results. Although no hybridisation signal was detected at the chromosomes (peri)centromeric regions, an interspersed hybridisation pattern was observed in almost all *P. eremicus* chromosomes (Fig. 1c), except for chromosomes PER17, PER20, PER21 and PER22 with an apparent absence of hybridisation signal (Fig. 2). Moreover, when karyotypes were built, a banding like pattern was consistently observed in each of the chromosome pairs.

A detailed comparison between CCR4/10sat hybridisation signals and *P. eremicus* chromosome C-banding pattern (unpublished data) was performed. In the aforementioned work, the CH from the chromosomes of the two species in analysis was characterised in detail using a panel of four restriction endonucleases (*in situ* RE digestion) and sequential C-banding. The identification of cryptic C-bands, in addition to the CH bands revealed by classical C-banding, allowed the ascertainment of a

coincident location of the CCR4/10sat probe and *P. eremicus* constitutive heterochromatin.

Fig. 2 presents a detailed analysis of CCR4/10sat physical mapping in *P. eremicus* chromosomes. In the left column of this figure, it is possible to observe the G- and C-banding of each chromosome pair, which are the controls for the subsequent comparative analysis of the several REs and sequential C-banding (RE + C-banding). C-bands in this column were identified by letters (a–g), according to their order of appearance in each chromosome. In the other columns, CCR4/10sat and REs + C, the hybridisation pattern of the CCR4/10sat probe (white arrowheads) and the action of the different REs used (AluI, BamHI, PstI, RsaI) with sequential C-banding in *P. eremicus* chromosomes, are presented respectively. The C-bands that co-localize with CCR4/10sat sequences are evidenced by black arrowheads.

As can be observed in Fig. 2, the genomic distribution of the CCR4/10sat sequences in this species' chromosomes co-localizes with the distribution of CH revealed by classical C-banding and/or C-banding sequential to *in situ* RE digestion (RE + C-banding). A more detailed analysis reveals that in chromosomes PER6, PER11, PER12, PER15 and PERY, these sequences are only co-localized with CH bands revealed by classical C-banding (black arrowheads in the control chromosomes column). Nevertheless, in other chromosomes (PER1, PER3, PER10, PER14 and PER23), these sequences only co-localize with cryptic C-bands (black arrowheads in RES + C column), evidenced by the digestion with one or more RE, from the panel of enzymes used, namely AluI, BamHI, PstI and RsaI. It should be mentioned that the hybridisation signals co-localized with cryptic C-bands, might correspond to the same band revealed by several REs; however in this figure, only the action of one RE for each corresponding C-band is shown as an example. For instance, in PER3 the CCR4/10sat signal closer to the centromere corresponds to a cryptic C-band revealed by more than one RE used, though only the band revealed by RsaI is presented. Finally, concerning chromosomes PER2, PER4, PER5, PER7, PER8, PER9, PER13, PER16, PER18, PER19 and PERX, the disclosed hybridisation signals correspond simultaneously to C-bands evidenced by classical C-banding and C-banding after *in situ* RE digestion. In some specific chromosomes, PER1, PERX and PERY (bands marked with an (*)), the CCR4/10sat signal did not correspond to any cryptic C-band nor to any C-band observed in control chromosomes. Besides this, it was also observed that several C-bands, revealed by classical C-banding or RE + C-banding do not correspond to CCR4/10sat hybridisation signals. The majority of these bands are located in the short arms of *P. eremicus* chromosomes.

4. Discussion

In the present work the isolation of *C. cricetus* centromeric sequences from chromosome 4 is reported, using the laser microdissection and laser pressure catapulting procedure. The physical mapping of these sequences was performed in *C. cricetus* and in the related species *P. eremicus* (Rodentia, Cricetidae) chromosomes, by fluorescent *in situ* hybridisation. The presence of these sequences in *P. eremicus* implies their existence in a common ancestor, indicating that these sequence variants can be considered as orthologous. When analyzing the hybridisation signals, it was observed a high correspondence among the chromosomal location of CCR4/10sat and the constitutive heterochromatin of both species chromosomes, suggesting the repetitive nature of these orthologous sequences.

In *C. cricetus* it was possible to observe a strong hybridisation signal in the (peri)centromeric region of CCR4 and CCR10 chromosomes (Fig. 1a). These signals co-localize with (peri)centromeric heterochromatin, evidenced by classical C-banding (Fig. 1b). This feature suggests the presence of this satellite family (or variants of the sequence) in these two chromosomes, since these sequences were isolated only from CCR4, but also hybridises in CCR10 centromeric region.

According to the obtained results, CCR4/10sat seems to be almost chromosome-specific, making this the first report on chromosome-specific sequences in *C. cricetus* (as far as we know). Other chromosome-specific sequences have been described in different Rodentia species, namely *Mus musculus* (Boyle and Ward, 1992), *Rattus norvegicus* (Essers et al., 1995), *Cricetulus griseus* (Fátyol et al., 1994) and *Mesocricetus auratus* (Yamada et al., 2006), the last two in this list belonging to the same family and subfamily (Cricetidae, Cricetinae) of *C. cricetus*.

In *P. eremicus* chromosomes, no hybridisation signal was detected at (peri)centromeric regions. However, a scattered pattern can easily be observed in the great majority of *P. eremicus* chromosomes, whose location is mainly coincident with CH bands, revealed by classical C-banding or RE + C-banding (Fig. 2).

On the other hand, and as noted, not all CH bands revealed by classical and RE + C-banding in PER chromosomes (mainly at the short arms) showed correspondence to CCR4/10sat sequences, implying the occurrence of different repetitive sequences. This is not surprising, since different satDNA sequences can coexist in the same genome. Specifically, in *P. eremicus*, several satDNA families were already identified in the chromosomes' short arms (Hazen et al., 1977; Hamilton et al., 1992).

Although belonging to the same family, *C. cricetus* and *P. eremicus* comprise different subfamilies, Cricetinae and Neotominae, respectively. The presence of the same repetitive sequences in the genomes of these related species imply its existence, at least, in the common ancestor of the two subfamilies. Fossil records and molecular data suggest the origin of Cricetinae and Neotominae subfamilies in the middle Miocene (Baskin, 1989; McKenna and Bell, 1997; Neumann et al., 2006), implying that the sequences here isolated date, at least, from this epoch (approximately 16–11 million years ago). To determine whether these sequences are older, or if they are present in other Cricetidae subfamilies, an extension of this analysis to other related genomes, will certainly shed light on the evolutionary history of these repetitive sequences and simultaneously, contribute to the clarification of the phylogenetic relationships of the species sharing them.

The CCR4/10sat orthologous sequences present however, a different chromosome location in the two species. According to the results obtained and according to parsimony, it is proposed that these sequences had originally a (peri)centromeric location, as the observed condition in *C. cricetus* chromosomes, that later assumed a scattered pattern, as observed in *P. eremicus* karyotype. Recently, Adega et al. (2008) found similar results in Suidae vs. Tayassuidae families, in cross-species physical mapping of orthologous satellite DNA sequences that revealed a completely different chromosomal location, reflecting a high level of karyotypes divergence after the radiation of each family. Also in Rodentia, species possessing more primitive karyotypes have satellite DNA sequences at (peri)centromeric regions, whereas derived karyotypes also revealed interstitial and full arm localisations (Hamilton et al., 1990; Rossi et al., 1995). The CCR4/10sat sequences repositioning and amplification in *P. eremicus* chromosomes can be explained by intragenomic movements of the satellite DNA sequences. One of the processes that can readily explain the expansion of heterochromatin within the same or different chromosomal fields (e.g. telomeres and interstitial locations) is "saltatory amplification" that can be mediated by several mechanisms (e.g. rolling circle amplification, unequal crossing-over, among others). Amplification events in

satellite DNA sequences were also found in other Rodentia genomes (see Hamilton et al., 1990; Rossi et al., 1995).

Another important feature is that satellite DNA and other repetitive DNA sequences can be the "active agent" of chromosomal evolution in mammals, being the genetic factors responsible for genomic plasticity and therefore, higher rates of chromosomal mutation (Slamovits and Rossi, 2002). According to several authors, satDNA sequences play an important role in mammal genome evolution by promoting chromosomal rearrangements, due to the rapid evolution of this repetitive fraction by means of their intragenomic movements among nonhomologous chromosomes and between different chromosomal fields (see Wichman et al., 1991; Reig et al., 1992; Schluter et al., 1997; Slamovits et al., 2001). Following this reasoning, it can be proposed that the chromosomal rearrangements occurred during *P. eremicus* karyotype restructuring may be the consequence of the presence and high dynamics of these repetitive sequences. A similar study was performed in the rodent genus *Ctenomys* by Rossi et al. (1995) regarding RPCS (repetitive PvuII *Ctenomys* sequence). These authors observed that *Ctenomys* species possessing karyotypes closer to the hypothesized ancestral karyotype, exhibit (peri)centromeric heterochromatin (containing RPCS), whereas the most derived karyotypes also display interstitial and entire length short arm heterochromatic blocks. Therefore, concerning the high dynamic of the CCR4/10sat sequences in *P. eremicus* genome, compared with its orthologous variant in *C. cricetus*, it can be suggested that *P. eremicus* karyotype is more derivative, being characterised by the occurrence of a higher number of complex chromosomal rearrangements. This idea is also supported by Comparative Chromosome Painting results (unpublished data) that indicate this genome as highly restructured by the occurrence of complex chromosome rearrangements. Furthermore, the construction of comparative maps between these species and other Cricetidae rodents would certainly elucidate this hypothesis.

Acknowledgements

This work was supported by the project POCI/BIA-BCM/ 58541/ 2004, the PhD grants SFRH/BD/25813/2005, SFRH/BD/41574/2007 of the Science and Technology Foundation (FCT) from Portugal and also by the Czech Ministry of Agriculture (MZE 0002716201). We are deeply grateful to Dr. Vitaly Volobouev for providing the Rodentia species cell cultures.

References

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

Adega, F., Chaves, R., Guedes-Pinto, H., 2007. Constitutive heterochromatin characterization of White-lipped and Collared peccaries (Tayassuidae). J. Genet. 86, 19–26.

Adega, F., Chaves, R., Guedes-Pinto, H., 2008. Suiformes orthologous satellite DNAs as a hallmark of *Pecari tajacu* and *Tayassu pecari* (Tayassuidae) evolutionary rearrangements. Micron 39, 1281–1287.

Baskin, J.A., 1989. The initial origin and diversification of the Neotropical Sigmodontinae (Rodentia: Muridae) – a perspective from the North American fossil record. In: Nevo, E., Reig, O. (Eds.), Proceedings of Fifth International Theriological Congress, Rome.

Boyle, A.L., Ward, D.C., 1992. Isolation and initial characterization of a large repeat sequence element specific to mouse chromosome 8. Genomics 12, 517–525.

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

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

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

Chaves, R., Frönicke, L., Guedes-Pinto, H., Wienberg, J., 2004a. Multidirectional chromosome painting between the Hirola antelope (*Damaliscus hunteri*, Alcelaphini, Bovidae), sheep and human. Chromosome Res. 12, 495–503.

Chaves, R., Santos, S., Guedes-Pinto, H., 2004b. 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.

Committee for standardization of chromosomes of *Peromyscus*, 1977. Standard katyotype of the deer mice, *Peromyscus* (Rodentia). Cytogenet. Cell. Genet. 19, 38–43.

Corradini, N., Rossi, F., Giordano, E., Caizzi, R., Vern, F., Dimitri, P., 2007. *Drosophila melanogaster* as a model for studying protein-encoding genes that are resident in constitutive heterochromatin. Heredity 98, 3–12.

Deaven, L.L., Vidal-Rioja, L., Jett, J.H., Hsu, T.C., 1977. Chromosomes of *Peromyscus* (rodentia, cricetidae). VI. The genomic size. Cytogenet. Cell Genet. 19 (5), 241–249.

Essers, J., de Stoppelaar, J.M., Hoebee, B., 1995. A new rat repetitive DNA family shows preferential localization on chromosome 3, 12 and Y after fluorescence *in situ* hybridization and contains a subfamily which is Y chromosome specific. Cytogenet. Cell Genet. 69, 246–252.

Fátyol, K., Cserpán, I., Praznovszky, T., Keresö, J., Hadlaczky, G., 1994. Cloning and molecular characterization of a novel chromosome specific centromere sequence of Chinese hamster. Nucl. Acids Res. 22 (18), 3728–3736.

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.

Gamperl, R., Vistorin, G., Rosenkranz, W., 1976. A comparative analysis of the karyotypes of *Cricetus cricetus* and *Cricetulus griseus*. Chromosoma 55, 259–265.

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. U.S.A. 98, 171–175.

Gosálvez, J., López-Fernández, C., Goyanes, R., Mezzanotte, V., 1997. Chromosome differentiation using nucleases: an overview. In: Henriques-Gil, N., Parker, J.S., Puertas, M.J. (Eds.), Chromosomes Today. Chapman & Hall, London, pp. 23–49.

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 (5), 321–329.

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

Hazen, M.W., Kuo, M.T., Arrighi, F.E., 1977. Genome analysis of *Peromyscus* (Rodentia, Cricetidae). Chromosoma 64, 133–142.

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.

John, B., 1988. The biology of heterochromatin. In: Verma, R.S. (Ed.), Heterochromatin: Molecular and structural Aspects. Cambridge University Press, Cambridge, pp. 1–128.

Kehrer-Sawatzki, H., Schreiner, B., Tanzer, S., Platzer, M., Muller, S., Hameister, H., 2005. Molecular characterization of the pericentric inversion that causes differences between chimpanzee chromosome 19 and human chromosome 17. Am. J. Hum. Genet. 71, 375–388.

Kubickova, S., Cernohorska, H., Musilova, P., Rubes, J., 2002. The use of laser microdissection for the preparation of chromosome-specific painting probes in farm animals. Chromosome Res. 10, 571–577.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Leitão, A., Chaves, R., Santos, S., Guedes-Pinto, H., Boudry, P., 2004. Restriction enzyme digestion chromosome banding in *Crassostrea* and Ostrea species: comparative karyological analysis within ostreidae. Genome 47 (5), 781–788.

Li, Y.C., Lee, C., Sanoudou, D., Hsu, T.H., Li, S.Y., Lin, C.C., 2000. Interstitial colocalization of two cervid satellite DNAs involved in the genesis of the Indian muntjac karyotype. Chromosome Res. 8, 363–373.

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.

Locke, D.P., Archidiacono, N., Misceo, D., Cardone, M.F., Deschamps, S., Roe, B., Rocchi, M., Eichler, E.E., 2003. Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. Genome Biol. 4, R50.

McKenna, M.C., Bell, S.K., 1997. Classification of Mammals: Above the Species Level. Columbia University Press, New York, pp. 114–211.

Meles, S., Adega, F., Guedes-Pinto, H., Chaves, R., 2008. The karyotype and sex chromosomes of *Praomys tullbergi* (Muridae, Rodentia): a detailed characterization. Micron 39 (5), 559–568.

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.

Neumann, K., Michaux, J., Lebedev, V., Yigit, N., Colak, E., Ivanova, N., Poltoraus, A., Surov, A., Markov, G., Maak, S., Neumann, S., Gattermann, R., 2006. Molecular phylogeny of the Cricetinae subfamily based on the mitochondrial cytochrome *b* and 12S rRNA genes and the nuclear vWF gene. Mol. Phyl. Evol. 39, 135–148.

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.

Pathak, S., Hsu, T.C., Arrighi, F.E., 1973. Chromosomes of *Peromyscus* (Rodentia, Cricetidae). IV. The role of heterochromatin in karyotypic evolution. Cytogenet. Cell Genet. 12, 315–326.

Peacock, W.J., Dennis, E.S., Gerlach, W.L., 1982. DNA sequence changes and speciation. Prog. Clin. Biol. Res. 96, 123–142.

Reig, O.A., Massarini, A.L., Ortells, M.O., Barros, M.A., Tiranti, S.I., Dyzenchauz, F.J., 1992. New karyotypes and C-banding of subterranean rodents of the genus *Ctenomys* (Caviomorpha, Octodontidae) from Argentina. Mammalia 56, 603–623.

Robles, F., Herrán, R., Ludwig, A., Ruiz, R.C., Ruiz, R.M., Garrido-Ramos, M.A., 2004. Evolution of ancient satellite DNAs in sturgeon genomes. Gene 338, 133–142.

Rossi, M.S., Redi, C.A., Viale, G., Massarini, A.L., Capanna, E., 1995. Chromosomal distribution of the major satellite DNA of South American rodents of the genus *Ctenomys*. Cytogenet. Cell Genet. 69, 179–184.

Rossi, F., Moschetti, R., Caizzi, R., Corradini, N., Dimitri, P., 2007. Cytogenetic and molecular characterization of heterochromatin. Gene models in *Drosophila melanogaster*. Genetics 175, 595–607.

Ruiz-Herrera, A., Castresana, J., Robinson, T.J., 2006. Is mammalian chromosomal evolution driven by regions of genome fragility? Genome Biol. 7, R115.

Saffery, R., Earle, E., Irvine, D.V., Kalitsis, P., Choo, K.H., 1999. Conservation of centromere protein in vertebrates. Chromosome Res. 7, 261–265.

Salser, W.S., Bowen, S., Browne, D., el-Adli, F., Fedoroff, N., Fry, K., Heindell, H., Paddock, G., Poon, R., Wallace, B., Whitcome, P., 1976. Investigation of the organization of mammalian chromosomes at the DNA sequence level. Fed. Proc. 35 (1), 23–35.

Schluter, D., Price, T., Mooers, A., Ludwig, D., 1997. Likelihood of ancestor states in adaptative radiation. Evolution 51, 1699–1711.

Slamovits, C.H., Cook, J.A., Lessa, E.P., Rossi, M.S., 2001. Recurrent amplifications and delections of satellite DNA accompanied chromosomal diversification in South American tucos-tucos (genus *Ctenomys*, Rodentia: Octodontidae): a phylogenetic approach. Mol. Biol. Evol. 18, 1708–1719.

Slamovits, C.H., Rossi, M.S., 2002. Satellite DNA: agent of chromosomal evolution in mammals. A review. Mastozoologia Neotropical/J. Neotrop. Mammal 9 (2), 297–308.

Sumner, A.T., 1972. A simple technique for demonstrating centromeric heterochromatin. Exp. Cell Res. 75, 304–306.

Sumner, A.T., Evans, H.J., Buckland, R.A., 1971. New technique for distinguishing between human chromosomes. Nature 232, 31–32.

Ugarkovic, D., Plohl, M., 2002. Variation in satellite DNA profiles-causes and effects. EMBO J. 21, 5955–5959.

Wichman, H.A., Payne, C.T., Ryder, O.A., Hamilton, M.J., Maltbie, M., Baker, R.J., 1991. Genomic distribution of heterochromatic sequences in equids: implications to rapid chromosomal evolution. J. Hered. 82, 369–377.

Yamada, K., Kamimura, E., Kondo, M., Tsuchiya, K., Nishida-Umehara, C., Matsuda, Y., 2006. New families of site-specific repetitive DNA sequences that comprise constitutive heterochromatin of the Syrian hamster (*Mesocricetus auratus*, Cricetinae, Rodentia). Chromosoma 115, 36–49.

Yunis, J.J., Yasmineh, W.G., 1971. Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation. Science 174, 1200–1209.

2.2. An ancient satellite DNA in *Peromyscus* genome that evolves by copy number fluctuation: does the sequence matters?

An ancient satellite DNA in *Peromyscus* genome that evolves by copy number fluctuation: does the sequence matters?

S Louzada [1], A Vieira-da-Silva[1], S Kubickova[2], F Adega[1], J Rubes[3], R Chaves[1,2]

[1]Institute for Biotechnology and Bioengineering, Centre of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro (CGB-UTAD/IBB),Vila Real, Portugal.

[2]Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Apdo 1013, 5001-801 Vila Real, Portugal

[2]Veterinary Research Institute, Brno, Czech Republic.

Keywords: Satellite DNA, Laser Microdissection, Copy number, Library model, Rodentia.

ABSTRACT | Satellite DNAs (satDNA) are tandemly arrayed repeated sequences present in eukaryotic genomes, which play an important role in genome evolution, hence, identification and analysis of satDNA regions are vital to the complete understand of genome structure and function. Here we describe the isolation of a novel satellite DNA family (PMSat) from the rodent *Peromyscus eremicus* (Cricetidae, Rodentia) using laser microdissection. Orthologous PMSat satDNA were isolated and charactrized in three species belonging to the Cricetidae family, namely, *Cricetus cricetus*, *Phodopus sungorus* and *Microtus arvalis*. Moreover PMSat was also found in species belonging to different phyla indicating that this constitutes the oldest satDNA described so far. Despite the taxonomic distribution, PMSat revealed extremely high interspecies sequence homology and the absence of fixed species-specific mutation. Different number of copies of this sequence was found in the species studied indicating the evolution of specific PMSat composition by copy number changes. The remarkably high evolutionary sequence conservation between the analyzed species along with the preservation of this sequence in the genome, even in very low copy number in some of the species, strongly suggests functional significance.

INTRODUCTION

Tandemly repeated DNAs, which include satellite DNA (satDNA), constitute a highly repetitive component and significant fraction of the eukaryotic genomes that can be present in up to several million of copies (Charlesworth et al. 1994). The long arrays of satDNA form prominent blocks of constitutive heterochromatin (CH) (John 1988), typically found at pericentromeric regions, being also reported in interstitial and telomeric positions (Adega et al. 2007; Meles et al. 2008; Paço et al. 2009; Petrović et al. 2009). Due to their peculiar repeat structure these sequences remained the final frontier in genome assembly and annotation (Alkan et al. 2011). For this reason, particular studies reporting the isolation and characterization of the satellite repeats constitute a way of better understand genomes evolution and function. Some satDNA sequences exhibit high variability affecting monomer size, nucleotide sequence, chromosome organization and location (reviewed by Plohl et al. 2008). This comes as the result of satDNA dynamic molecular behavior promoted by concerted evolution, in which different mechanisms of DNA turnover lead to rapid intraspecific homogenization of occurring changes (e.g. Dover 1986; Elder and Turner 1995). However, the landscape of satDNAs can be modified and differ between species without necessarily implying changes in the monomer sequence. In the last years diverse studies report the existence of satellite DNA sequences preserved between evolutionary distant species, indicating possible functionality (Li and Kirby 2003; Mravinac et al. 2005; Komissarov et al. 2011). Transcripts from satellite containing regions have been identified in multiple organisms (Stimpson and Sullivan 2010), and proteins that interact with those transcripts have been described (reviewed in Hall et al. 2012). Moreover, the functional significance of satellite DNA sequences have been demonstrated to range from chromosome organization and pairing, to cell metabolism and speciation (e.g. Martienssen 2003; Wong *et al.* 2007; Grewal and Elgin 2007).

Peromyscus eremicus (PER) is a wild rodent species, commonly known as cactus mouse, that belongs to the Cricetidae family. Species from this genus have received special attention, and have been used as model organisms (e.g. Glasper and DeVries 2005; Shorter et al. 2012). With a diploid number of 48 chromosomes, *P. eremicus* presents a karyotype comprised solely by submetacentric chromosomes (The Committee for standardization of chromosomes of *Peromyscus*, 1977). Constitutive heterochromatin (CH) is present in great amount, and is localized pericentromerically, being the p-arms of the majority of the chromosomes almost entirely heterochromatic (see Pathak et al. 1973; Deaven et al. 1977).

In this work we describe the characterization of a novel satellite DNA family (PMSat) isolated by laser microdissection from the genome of the rodent *Peromyscus eremicus*. We show that

PMSat is a highly homogeneous satellite present in other rodent species, but exhibiting striking differences in genomic abundance between the rodent species analyzed. The occurrence of PMSat in species belonging to different phyla was also detected and indicates that this constitutes the oldest satDNA described so far.

MATERIALS AND METHODS

Chromosome preparations and Genomic DNA extraction

Fibroblast cell lines of the species *Peromyscus eremicus* (PER), *Cricetus cricetus* (CCR)*, Phodopus sungorus* (PSU) and *Microtus arvalis* (MAR) were used for obtaining metaphase chromosomal spreads and genomic DNA. Genomic DNA was extracted using the JETQUICK DNA kit (Genomed), according with manufacturers' recommendations.

The nomenclature of *P. eremicus* (2n=48) chromosomes is according to the Committee for standardization of chromosomes of *Peromyscus* (1977).

Microdissection, DOP-PCR amplification, Cloning and Sequencing

Chromosome dissection and collection was carried out using PALM MicroLaser system (P.A.L.M. GmbH, Bernried, Germany) coupled to the light path of an inverted microscope (Olympus). About 10 randomly chosen chromosome centromeres from *P. eremicus* were microdissected and catapulted by a single laser pulse directly into the cap of a PCR tube, containing 2 μL PCR oil. 20 μL of 10 mmol/L Tris-HCl pH 8.8 was added to the cap, and the sample was spun down by centrifugation. DOP-PCR was performed as described by Kubickova et al. (2002). Briefly, the reaction mixture contained 60 mM Tris-HCl (pH 8.8), 15 mM $(NH_4)_2SO_4$, 3.5 mM $MgCl_2$, 0.2 mM each dNTPs, 1.6μM 6MW primer (5` - CCGACTCGAGNNNNNNATGTG G - 3`, Telenius 1992), 0.05% W-1 (Invitrogen Life Technologies) and 2U Taq polymerase (Invitrogen Life Technologies) in a final volume of 40 μL.

Primary DOP-PCR product was reamplified and afterwards purified using the High Pure PCR Product Purification kit (Roche Diagnostics). The product was ligated to pDrive vector (QIAGEN) and transformed into QIAGEN EZ Competent cells following the manufacturers' guidelines (QIAGEN PCR Cloning ^{plus} Kit). Recombinant clones screening was performed by PCR using the vector primers M13.

For dot-blot analysis, amplification products (1.5 μL) were spotted onto two copies of nylon filters and hybridized with two digoxigenin-11-dUTP probes (Roche Diagnostics): specific primary DOP-PCR and genomic DNA, both under standard conditions (Maniatis et al. 1982) at

42°C for 16 h. Anti-digoxigenin-11-dUTP antibody coupled to alkaline phosphatase (Roche Diagnostics) and BM purple AP substrate (Roche Diagnostics) were used for signal detection. The presence of specific repeats in each clone was estimated by comparison of the intensity of the hybridization signals. Positive clones were labeled by PCR and checked by *in situ* hybridization onto *P. eremicus* chromosome metaphases. The selected clones were then sequenced in both directions using M13 primers.

Fluorescent in situ Hybridization

*C*hromosome preparations were aged at 65ºC overnight and dehydrated with 100% chilled ethanol for 10 min. Hybridization was carried out in a moist chamber at 37ºC (overnight), and the most stringent post-hybridization wash was 50% Formamide/2×SSC at 37ºC. Biotin-16-dUTP (Roche Diagnostics) labeled probes were detected with FITC conjugated with avidin (Vector Laboratories). Chromosomes were observed with a Zeiss Axioplan 2 Imaging microscope, coupled to an Axiocam digital camera with AxioVision software (version Rel. 4.5 – Carl Zeiss). Digitised photos were prepared for printing in Adobe Photoshop (version 5.0). Contrast and colour optimization were the functions used and affected the whole of the image equally.

CBP-Banding

CBP-banding [C-bands by barium hydroxide using Propidium Iodide(PI)] was done according to the standard procedure of Sumner (1972) with slight modifications. This technique was performed sequentially to *in situ* hybridization, after distaining the slides. Shortly, the slides were submitted to hydrochloric acid (0.1 M) 20 min, barium hydroxide (5% solution) 7 min and 2×saline solution citrate (2×SSC: 0.3 mol/L NaCl, 0.03 mol/L sodium citrate) at 60ºC for 40 min. The slides where then counterstained with PI (1.5 μL/mL).

RE digestion and Southern blot hybridization

Genomic DNA from *P. eremicus* was digested with the restriction endonucleases (RE) ApaI, AvaI, HaeIII and RsaI and the resulting fragments were separated in a 1.2% agarose gel and blotted onto a Nylon membrane Hybond™-N^+ (Amersham, GE Healthcare) according with the manufactures" procedures. Membranes were probed with the microdissection isolated sequence, previously labeled by PCR with digoxigenin-11-dUTP (Roche Diagnostics). Hybridization was performed overnight in hybridization solution at 68ºC. Positive signals were visualized using chemiluminescent CDP-Star system (Roche Diagnostics).

PCR amplification

Two sets of PCR primers were designed based in the isolated and sequenced *P. eremicus* satellite sequences, using Primer3 web-based interface (primer3.sourceforge.net) (Rozen and Skaletsky 2000). The two primer combinations were the following: PMSat1 forward 5'-CGA CTC GAG TGG GTT ATG TG3'/ PMSat1 reverse 5'-GTA TGT GGG GTC CGA ACA GT-3') and PMSat1forward / PMSat2 reverse 5'-CCG ACT CGA GGA AGG TAT GT-3'). Genomic DNA from *P. eremicus*, *C. cricetus*, *P. sungorus* and *M. arvalis* was used for PCR amplification, performed as follows: one denaturation cycle at 94ºC (5 min), 35 cycles at 94ºC denaturation (1 min), 58ºC annealing (1 min), 72ºC extension (45 sec) and one final extension cycle 72ºC (5 min).

Cloning and Sequencing

PCR Amplification products from the species *P. eremicus*, *C. cricetus*, *P. sungorus* and *M. arvalis* and HaeIII digestion from *P. eremicus*, were extracted from the agarose gel and purified using QIAquick PCR Purification Kit (QIAGEN). Afterwards Fast DNA End Repair (FERMENTAS Life Science) was performed for blunting and phosphorylation of DNA ends, and subsequently they were linked into the SmaI site of plasmid pUC19 vector (FERMENTAS Life Science) and used to transform DH5α competent cells (Invitrogen Life Technologies). Clones were screened using the β-galactosidase blue-white colour system, and selected clones were labeled with digoxigenin-11-d'UTP (Roche Diagnostics), confirmed by *in situ* hybridization onto *P. eremicus* chromosome metaphases and then sequenced in both directions using M13 primers.

Sequence Analysis

For sequence characterization of the new satellite in *P. eremicus* genome, seven clones were analyzed, resulting from different approaches. For this reason clone designation contains the name of the species, the initial letter of the isolation method (or enzyme name) and the clone number. Two clones were isolated from **m**icrodissection PER**m**40 and PER**m**57, three from **P**CR amplification, PER**p**25, PER**p**45 and PER**p**62, and two clones were isolated from RE digestion with **HaeIII** are PER **HaeIII**A and PER **HaeIII**B. Clones from all the other species were isolated from PCR amplification, and their designation includes the species abbreviation, and isolation method (example CCRpB1). All sequence analysis was performed for the monomer region of the isolated clones, and incomplete monomers were completed with N letter representing all possible nucleotides A/C/G/T (according with IUPAC nomenclature).

Clone chromatograms and sequence alignment analysis were performed throughout BIO EDIT sequence alignment editor (Hall 1999). A BLAST search of the isolated sequences, against nucleotide sequences present in Genbank (Release 187.0, December 2011) and ENSEMBLE (Release 65, December 2011) databases, was accomplished, and the sequences that showed homology were retrieved from Genbank and analyzed together with the clone sequences isolated by us. Dotmatcher tool from Emboss (http://emboss.sourceforge.net) was used for Dot plot analysis with a threshold of 50%.

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5 (Tamura, Peterson, Stecher, Nei, and Kumar 2011). Distance tree was built using Neighbour Joining (NJ), distances were calculated according to Kimura's two-parameter method, and 1000 bootstraps replicas were performed to access the statistical support of each node.

Sequences have been deposited in genbank under the Accession numbers: GQ902036, KC351938, KC351939, KC351940, KC351941, KC351942, KC351943, KC351944, KC351945, KC351946, KC351947, KC351948, KC351949, KC351950, KC351951, KC351952, KC351953, KC351954 (Table S1).

Satellite DNA copy number quantification (absolute and relative) by TaqMan assay

For PMSat quantification we designed a TaqMan specific assay mix (primers/probe) using Primer Express® Software v3.0 (Life Technologies Applied Biosystems) based in PMSat consensus sequences of all species studied. The PCR primers PMSat F (5'^{196}CATGCAGTTTCTGGTCCTACGA3') and PMSat R (5'^{261} TGGGAGCGCAAGAGTGACT3') were located between the positions 196 and 261 of the 66 bp product. The probe [5'^{219}(FAM)CAAGAGTTGTTTTCTGGTAAGA3' (NFQ)] had the fluorescent reporter dye, 6-carboxy-fluorescein (FAM) located at the 5' end of the probe and the non-fluorescent quencher (NFQ) is located at the 3' end.

For PMSat absolute quantification in *Peromyscus eremicus* genome the standard curve method was performed. A 10-fold serial dilution series of the plasmid DNA standard (PERm40), ranging from 1×10^9 to 1×10^5 copies, was used to construct the standard curve (5 points series dilutions). The concentration of the plasmid PERm40 was measured using the NanoDrop ND-1000 (NanoDrop Technologies) equipment and the corresponding plasmid copy number was calculated using the following equation:

$$\text{DNA (}\textit{copy number}\text{)} = \frac{6{,}023 \times 10^{23} \text{ (}\textit{copy number}\text{/mol)} \times \text{DNA amount (g)}}{\text{DNA length (bp)} \times 660 \text{ (g/mol/bp)}}$$

Where:

Avogadros number = $6{,}023 \times 10^{23}$ molecules (*copy number*) / 1 mol

Average molecular weight of a double-stranded DNA molecule = 660 g/mol/bp

In the respective formula the recombinant plasmid DNA length is 4315 bp (pDrive vector 3851 bp and the insert 464 bp).

C_T values in each dilution were measured using real-time qPCR with the TaqMan specific assay described above to generate the standard curve for PMSat. Briefly, the standard curve includes a plot of the C_T values versus the log concentration of the PERm40 standard. For *Peromyscus eremicus* genomic DNA, the unknown total DNA sample, was obtained by interpolating its C_T value against the standard curve. We used 0,5 ng of *Peromyscus eremicus* genomic DNA in the PCR reaction. These reactions were performed for a total of 20 μL with 1.25 μL of the primer/probe assay mixture and 12.5 μL of TaqMan® Genotyping Master Mix (Life Technologies Applied Biosystems). This experiment was carried out in the StepOne real-time PCR system (Life Technologies Applied Biosystems), where the samples were subjected to an initial denaturation at 95ºC (10 min), and then to 40 cycles at 95ºC 15 sec followed by 60ºC 1 min. All reactions were performed in triplicate, and negative controls (without DNA) were also run.

The StepOne software (version 2.2.2, Life Technologies Applied Biosystems) was used to generate the standard curve and to data analysis. Only standard curves with the following parameters were considered to be typically acceptable: $R^2 > 0.99$ and slopes between −3.1 and −3.6 giving reaction efficiencies between 90 and 110%.

The absolute quantification of PMSat allows determining the copy number of this sequence in *Peromyscus eremicus* genome that corresponds to 0,5 ng which, in turn, comprises 152 haploid genomes. The mass of *Peromyscus eremicus* haploid genome was obtained in Genome Size database (http://www.genomesize.com/) as being 3,3 pg ($3{,}3 \times 10^{-3}$ ng).

For PMSat quantification within the other species genomes (*Cricetus cricetus*, *Phodopus sungorus* and *Microtus arvalis*) we used a relative quantification real-time PCR approach, being the *Peromyscus eremicus* genome the reference sample. We used the same PMSat TaqMan assay described for the absolute quantification and as reference assay the Glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Rn01749022_g1).

For comparative analysis, the PCR reactions were performed with 0,5 ng of genomic DNA. Mixture reactions and real-time PCR conditions were the same already described. All reactions were performed in triplicate, and negative controls (without template) were run for each master mix.

StepOne software version 2.2.2 (Life Technologies Applied Biosystems) was applied for comparative analysis, and the quantification was normalized with GAPDH gene. The $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) was used to calculate fold changes in the amount of PMSat

in the different species. Results are shown as the $\log_{10}$ of $2^{-\Delta\Delta CT}$ PMSat copy number in *C. cricetus*, *M. arvalis* and *P. sungorus* relative to *P. eremicus* (calibrator sample).

Student's t-test was used to compare the data obtained. Values were expressed as the mean ± SD, and differences were considered statistically significant at $p<0.05$, representing the 95% confidence interval.

RESULTS

Isolation of a novel satellite DNA sequence from P. eremicus genome

Using the laser microdissection of *P. eremicus* chromosomes and subsequent procedures, two clones were obtained, with 464 bp (PERm40) and 463 bp (PERm57), sharing 99.8% identical nucleotide positions. Dot plot allowed the identification of two internal repeats (Figure 1) with 42 bp each, showing to be 100% identical. This feature gave some indication that the isolated sequence could in fact represent more than one monomer. Besides, another trait was identified in these sequences; in particular two inverted repeat motifs (31 bp) nearby the two internal repeats (Figure 1), suggesting a complex arrangement of this sequence in this species genome.

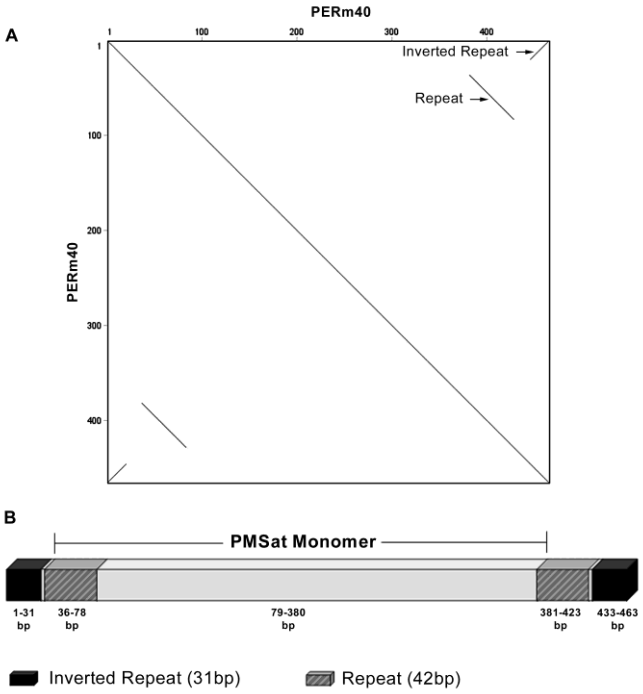


Figure 1 – Sequence internal structural characteristics. Dot plot diagram of the clone PERm40 compared to itself, showing two internal repeats (42bp length) and two inverted repeats (31bp length) (A). Schematic representation of the satellite detected features (B) with indication of the monomer.

Physical mapping in *P. eremicus* chromosomes

The physical mapping of PERm40 and PERm57 clones in *P. eremicus* genome was performed throughout fluorescent *insitu* hybridization. This sequence revealed to be present in all chromosomes being highly representative in this species genome (Figure 2A). Particularly, this sequence displays a pericentromeric location in all PER autosomes, being PER1 the only autosome presenting hybridization only at that region. In the other chromosomes this sequence is also found at the entire p-arm (see for instance PER2 and PER9) or at the terminal portion (see PER6) (see Figure 2B for detail). Regarding the sex chromosomes, the clones hybridized to both chromosomes pericentromeric region, but in the X chromosome it hybridizes also to distinct bands in the p-arm (see Figure 2B for detail). After performing sequential C-banding (Figure 2C), it was evidenced a co-localization of the repetitive sequence with *P. eremicus* constitutive heterochromatin (CH) (Figure 2C). However not all PER CH is composed by this sequence as it can be seen by the CH band in the long arm of X chromosome where no hybridization signal was found (Figure 2D).

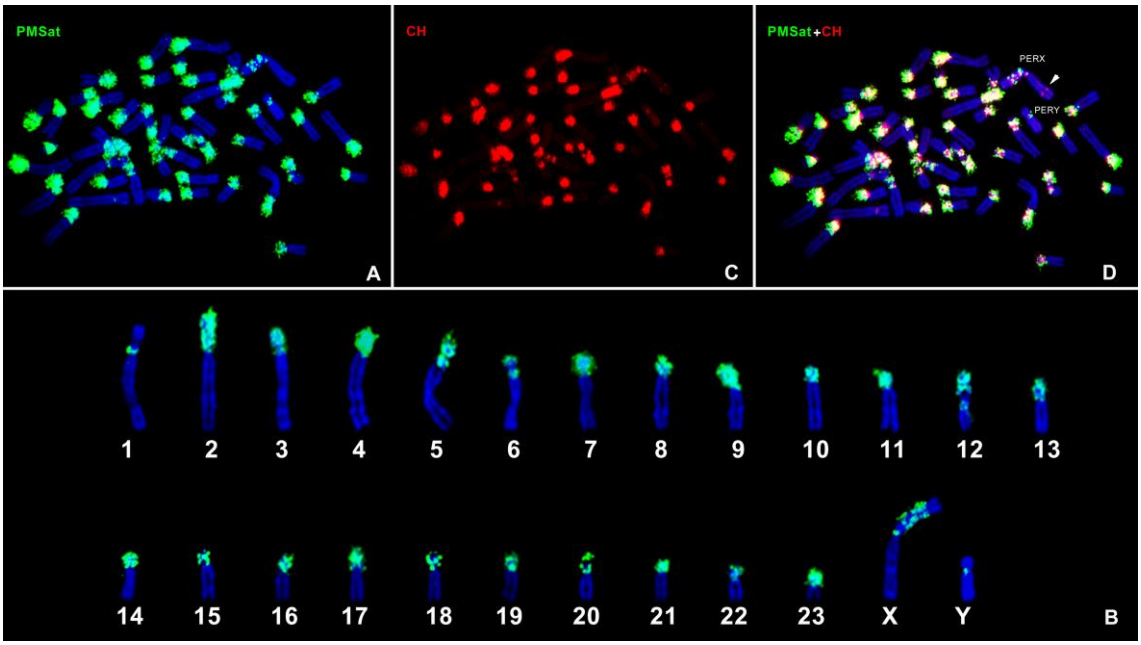


Figure 2 – Physical mapping of PMSsat on *Peromyscus eremicus* chromosomes. Representative *in situ* hybridization presenting the chromosomal localization of PMSat (clone PERm40) (A). Same metaphase after sequential C-banding (B). Overlapping of PMSat hybridization signal with C-banding (C). The arrowhead indicates a chromosomal region containing CH but no PMSat signal. Haploid karyotype arrangement of *P. eremicus* chromosomes showing PMSat hybridization signal (D).

Restriction enzyme digestion and Southern blot analysis

In order to investigate the genome organization of the isolated sequence, southern blot analysis was carried out using digested genomic DNA from *P. eremicus*, separated by conventional agarose gel electrophoresis and hybridized with the PERm40 clone. The restriction enzymes used - ApaI, RsaI, AvaI, and HaeIII - have respectively, none, one, two and three recognition sites in the isolated sequence. After restriction enzyme digestion, bands were observed only for HaeIII and RsaI (Figure 3), with approximately 345 bp and 690 bp. Besides these, for HaeIII bands with about 300 bp and 600 bp were also obtained (Figure 3). After southern hybridization a common band with approximately 345 bp was obtained for all the used enzymes (monomer), and other bands can also be seen with 345bp periodicity: 690bp (dimer), 1035bp (trimer) and 1380bp (tetramer). Besides, in AvaI, HaeIII and RsaI other bands than the 345bp multiples were obtained after southern hybridization (Figure 3), suggesting the presence of monomer variants that present nucleotide changes in the recognition site for those enzymes.

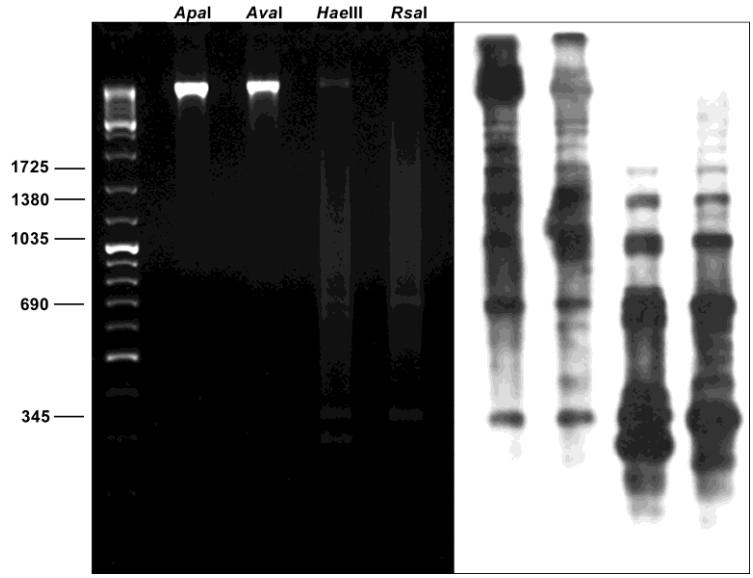


Figure 3 – Restriction enzyme Digestion and Southern blot. Electrophoresis separation of P. eremicus genomic DNA after digestion with ApaI, RsaI, AvaI, and HaeIII having respectively none, one, two and three recognition sites in the PMSat sequence (shown on the left). The corresponding southern blot obtained after hybridization with PMSat (clone PERm40) is shown on the right.

The presence of a ladder hybridization pattern indicates the tandem organization of this novel sequence, characteristic for satellite sequences. The 345bp band from HaeIII was isolated from

the agarose gel and subsequently cloned and sequenced, and two clones showing similarity with PER isolated sequence were selected (PER **HaeIII**A and PER **HaeIII**B).

All the performed analysis showed that the isolated sequence corresponds to a satellite DNA family highly represented in *P. eremicus* genome, which was named PMSat - **P**eromyscus **M**ajor **Sat**ellite.

Isolation and sequence analysis of PMSat in other Cricetidae species

In order to verify the existence of PMSat orthologous sequences in the genome of related species, specific primers were designed and used in PCR amplification experiments. The targeted species *Cricetus cricetus* (CCR), *Phodopus sungorus* (PSU) and *Microtus arvalis* (MAR), are all from Cricetidae family, but belong to different subfamilies (Musser and Carleton 2005) than *Peromyscus eremicus*. PCR amplification was performed also for *P. eremicus*. Electrophoresis of the PCR products showed amplification products for all the species (data not shown), that were subsequently cloned and sequenced. Clones that contained PMSat satellites were selected for analysis. The number of clones obtained for each species was: four clones for *C. cricetus* and *P. sungorus* and three clones for *P. eremicus* and *M. auratus* (Table S1).

Blast results and *in silico* analysis

Using the BLAST search algorithm the consensus sequence for PMSat monomer was compared against different databases. Homology with our sequence was obtained with whole-genome shotgun contigs from *Bombus terrestris* (large earth bumblebee) and *Lytechinus variegates* (green sea urchin). BLAST search revealed also a hit with a cDNA sequence from *Peromyscus maniculatus bairdii* (prairie deer mouse) (Table S1).

Sequence analysis

In total, seven sequences were isolated from *P. eremicus* genome: two by microdissection, two by *Hae*III endonuclease digestion, and three by PCR (Table S1). Sequences' sizes range from 343-464bp with an average AT content of 55% (TableS1). A multiple alignment was performed with all the sequences, however, analysis was restricted to the region corresponding to the satellite monomer region (345bp), limited by the *Hae*III restriction sites (highlighted in Figure 4). Overall, the differences found between sequences are mainly nucleotide substitutions, existing though two deletion/insertion sites (positions 71 and 131). The consensus sequence presents 47 variable positions between all clones, corresponding 46 out of those to variations in the *Hae*III monomers (Figure 4). Besides, PER *Hae*IIIA and PER *Hae*IIIB sequences share 11 common nucleotide substitutions. Generally, sequences show high similarity and some clones display

100% nucleotide identities (see Table 1). PER *Hae*IIIA and PER *Hae*IIIB are the ones with less homology with all the other sequences (Figure 4 and Table 2), and even lower between them than with the other sequences.

PMSat orthologous sequences isolated in the other Cricetidae species displayed sizes ranging from 387-446bp with an average AT content of 54% in *C. cricetus* and *M. arvalis* and 55% in *P. sungorus* (Table S1). All the monomers were aligned with *P. eremicus* consensus sequences and sequence variability analysis was performed (Figure 5). Nucleotide substitutions were the main type of changes found, with only two insertion/deletion sites (positions 9 and 40, Figure 5). When comparing all sequences to PER consensus sequence, it is possible to identify 36 variable positions, corresponding 19 to PSU monomers and 14 to CCR monomers (Figure 5). This corresponds to a similarity of 95% with CCR, 94% with PSU and 100% with MAR. The most divergent monomers are PSUpR1 and CCRpB1. The interspecies similarity found was very high, demonstrating a high conservation of monomer sequence between species. Regarding the orthologous PMSat sequences retrieved from Genbank, belonging to *Bombus terrestris* and *Lytechinus variegates*, they represent a part of the monomer sequence respectively 91% and 90%. These sequences present a homology of 76% and 78% with PMSat monomer (Table S1).

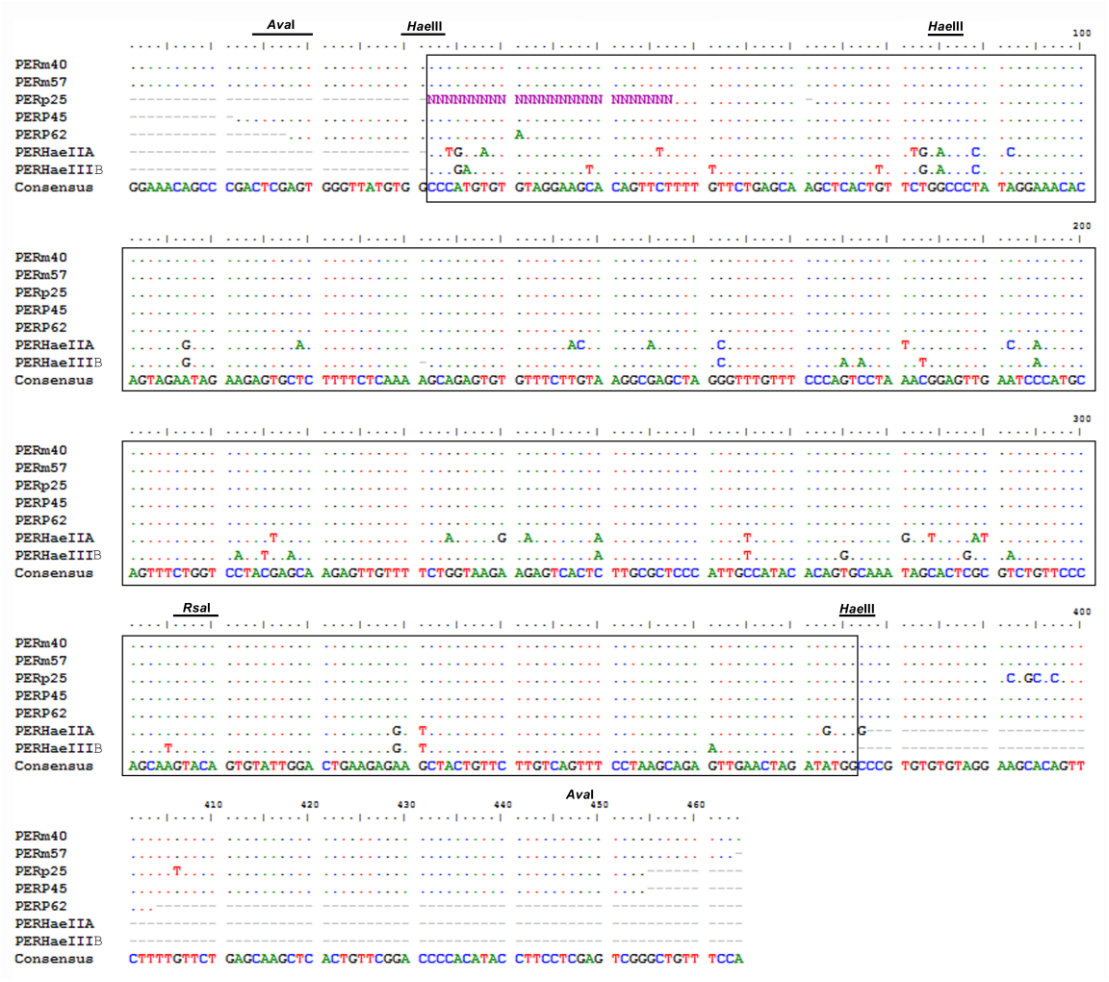


Figure 4 – Alignment of *P. eremicus* PMSat isolated clones. The sequence in the box corresponds to PMSat monomer. Only differences in sequences are indicated, while the positions of sequence identity are represented by a dot. Part of the PERp25 clone sequence was completed using the letter N (according to IUPAC). Enzyme restriction sites are indicated.

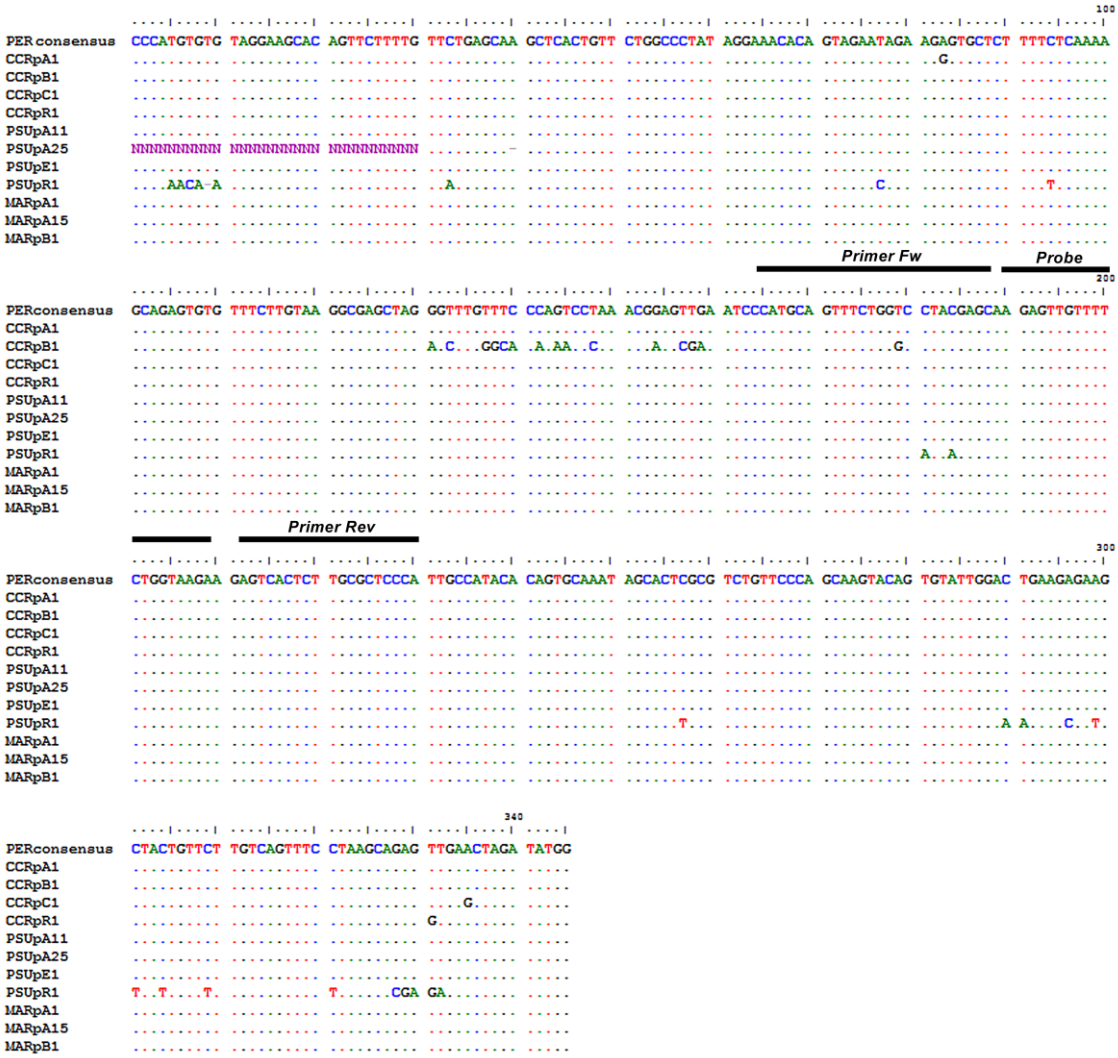


Figure 5 - Alignment of PMSat monomer sequence. The image shows the alignment between PMSat monomer consensus sequence of PER with other species PMSat clones. Only differences in sequences are indicated, while the positions of sequence identity are represented by a dot. Part of the PSUpA25 clone sequence was completed using the letter N (according to IUPAC). The primer/probe used in quantification experiments is indicated.

Table 1- Matrix of sequence identity of PMSat isolated monomer region in *P. eremicus* based in the alignment of Figure 4.

	PERm40	**PERm57**	**PERp45**	**PERp62**	**PERp25**	**PER *HaeIII*A**	**PER *HaeIII*C**
PERm40		1,000	1,000	0,997	0,997	0,910	0,921
PERm57	1,000		1,000	0,997	0,997	0,910	0,921
PERp45	1,000	1,000		0,997	0,997	0,910	0,921
PERp62	0,997	0,997	0,997		0,997	0,907	0,918
PERp25	0,997	0,997	0,997	0,997		0,912	0,922
PER HaeIIIA	0,910	0,910	0,910	0,907	0,912		0,895
PER HaeIIIC	0,921	0,921	0,921	0,918	0,922	0,895	

Results presented for the clone PERp25 refer to the nucleotide identities from positions 58-376.

Table 2- List of species and number of clones analysed, weight haploid genome, percentage of A-T and quantification of PMSat satellite family in the genome.

Species	**Clones**	**Similarity (%)**	**PMSat copy number analysis**
P. eremicus	7		$\geq 1{,}73 \times 10^6$ copies/genome **Relative quantification**
C. cricetus	4	95%	- 1×10^6 fold
P. sungorus	4	94%	- 6×10^6 fold
M. arvalis	3	100%	- 4×10^6 fold

Homology percentage refers to the number of identical nucleotide positions compared with PER consensus sequence (Figure 5). The amount of PMSat in CCR, MAR and PSU is presented as fold change relative to the amount determined for PER.

In order to highlight this conservation, a phylogenetic tree was constructed based in the multiple alignments of the different species monomers (either isolated by us and sequences retrieved in Genbank), using Neighbor-Joining method. This analysis is based in the calculation of evolutionary distances as number of base substitutions per site. As it can be seen in Figure 6, two main branches can be observed dividing the rodent species from the arthropod *B. terrestris* and echinoderm *L. variegates* species. These were placed in a separate branch meaning that their sequences show more homology between them than with the other species monomers. Among the rodent species analyzed there is no species based branch separation or cluster, what is in agreement with the interspecies high similarity found for PMSat orthologs. *P. eremicus Hae*III monomers were positioned in a separate branch meaning that these are closely related between them than with the other monomers. This is in accordance with previous sequence analysis showing that they share common nucleotide substitutions (Figure 4). Besides, the monomers

CCRpR1 and PSUpR1 were clustered together and separated from the other species, once they share a nucleotide position (position 331, Figure 5).

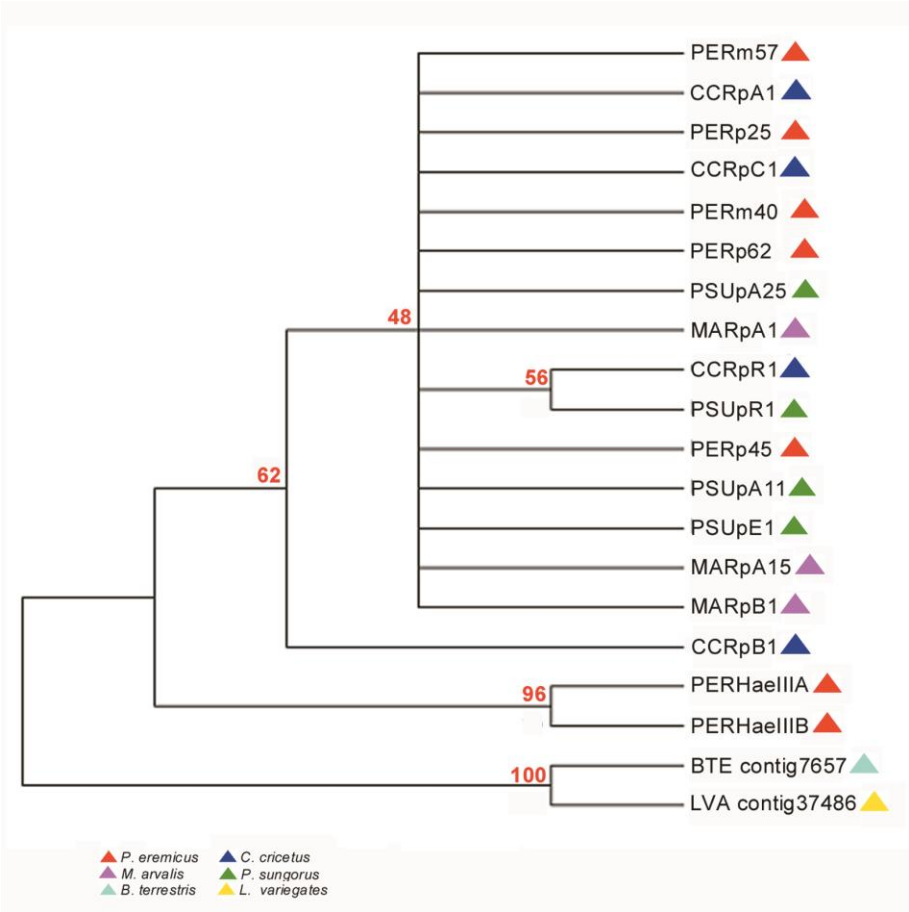


Figure 6 - Evolutionary tree based in PMSat monomer sequences. Evolutionary relationships were inferred using the Neighbor-Joining method. Bootstrap values (calculated for 1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Kimura two-parameter method and are in the units of the number of base substitutions per site. Branch lengths represent the evolutionary distances used to infer the phylogenetic tree.

PMSat satellite DNA copy number analysis

Although also found in other Cricetidae genomes, physical mapping of PMSat and southern blot hybridization was only accomplished in PER (Figure 2). For the other 3 species, CCR, PSU or MAR, this was not observed (data not shown), what could be due to differences in the copy number of this satDNA in the different genomes. Satellite DNA quantification was thus performed using a new methodology based in real-time quantitative PCR allied to TaqMan chemistry. A specific assay, made by two primers and a probe, was designed based in PMSat consensus sequence in a conserved region (shown in Figure 5). Both absolute and relative quantification experiments were performed. Despite the high similarity between monomers

found, PMSat family may include other divergent monomers that escaped our detection. For this reason, the number of copies estimated is considered the minimal number of copies existing in *P. eremicus* genome.

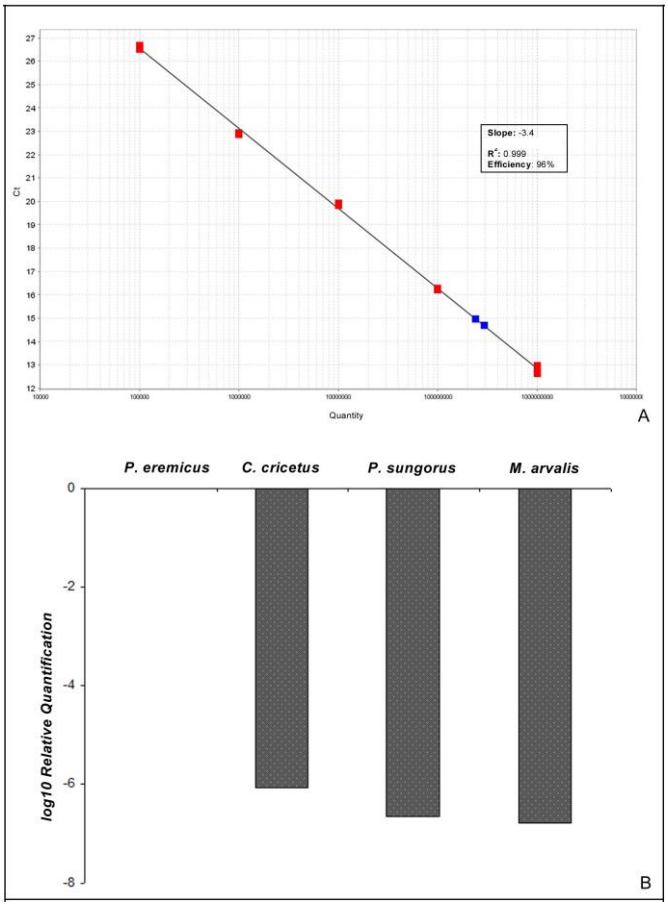


Figure 7- PMSat quantification. Standard calibration curve (A) used in the absolute quantification of PmSat copy number in *P. eremicus* genome. In blue are represented the standard dilutions and in red the PmSat samples. Relative quantification (represented as $\log_{10}$) of PmSat in CCR, PSU and MAR (B), using PER as control sample, showing lower copy number in those species.

Absolute quantification revealed that at least 20% of *P. eremicus* genome is comprised by PMSat (Figure 7A), corresponding to at least 8.7×10^5 copies per genome. Relative quantification showed that the amount of this satellite family in the other species genome is much lower than in *P. eremicus*, specifically about 10^6-fold lower in copy number (Figure 7B, Table 2), being all the presented results statistically significant values ($p<0.05$).

DISCUSSION

Since the early beginnings of understanding the genome structure, satellite DNA has always revealed to be enigmatic and controversial. The initial idea that these sequences represented a useless part of the genomes ("junk") has been contested by new evidences that illuminated the significant role that these repeats play in genomes. For this reason the discovery and characterization of newly satDNAs are necessary and revealed to contribute for a better understanding of genomes evolution and function.

Novel satellite DNA family in *Peromyscus eremicus*

Here we report for the first time the isolation and characterization (nucleotide sequence and genome organization) of a satellite DNA family from the rodent *Peromyscus eremicus,* named PMSat, using the laser microdissection procedure (Kubickova et al. 2002). This methodology revealed to be effective in isolating centromeric repetitive sequences in previous studies (Li et al. 2005, Pauciullo et al. 2006), including rodents (Louzada et al. 2008).

Our results show that PMSat represents a novel satellite DNA family present in the constitutive heterochromatin of *P. eremicus* genome*,* particularly in large pericentromeric blocks, and also present in the entire p-arm of the majority of the autosomes (Figure 2A, 2D). Regarding the sex chromosomes, this sequence is located at the pericentromeric region and, in the X chromosome, PMSat is also located in interstitial bands. Southern blot hybridization revealed a ladder like pattern with a periodicity of about 345bp for all the enzymes (Figure 3), indicating that this repeated sequence is arrayed in tandem with a monomer of 345bp, confirmed after sequence analysis. Studies regarding the rodent *Peromyscus eremicus* repetitive genome fraction are scarce. Previous works rely in cytogenetic analysis and describe the presence of satellite DNA in the pericentromeric region and in some chromosome short arms of *P. eremicus* (Hazen 1977; Hamilton et al 1992), as well as in interspersed locations (Louzada et al. 2008), although no sequence is available. This work provides information of the first satellite DNA in this species for which sequence data is available.

Orthologous PMSat are present in other species and suggests functional constraints

In the present work we isolated orthologous PMSat sequences in the genomes of species belonging to Cricetidae family (phylum Craniata), *C. cricetus* (CCR), *P. sungorus* (PSU) and *M. arvalis* (PSU). Besides, a BLAST search against available data from genome sequencing projects allowed to find PMSat orthologous sequences also in phylogenetically distant species such as *Bombus terrestris* (BTE, phylum arthropoda) and *Lytechinus variegates* (LVA, phylum echinodermata). This

suggests that this satDNA family may also be present in further species. The existence of this sequence in species belonging to three different phyla (Arthropoda, Craniata and Ecnhinodermata) shows that PMSat dates back to the Pre-cambrian period and to the Bilateria ancestral, indicating that this satellite originated at about 635 million years ago (Edgecombe et al. 2011). This makes PMSat the oldest satellite DNA family described so far, that along with the bivalves BIV 160 with 540 million years (Plohl et al 2010a) and the sturgeon PstI with 100 million years (Robles et al. 2004), constitute ancient satellites.

It has been predicted that rapid evolution of satellite DNA sequences occurs by means of concerted evolution, where monomers evolve in a non-independent manner with homogenization of the accumulated mutations throughout members of the repetitive family, followed by its fixation within a group (Dover 1982; Dover 1986). This would result in the accumulation of divergences among satellite sequences in isolated groups of organisms in short evolutionary periods (reviewed by Plohl 2008). On the contrary, the data presented here shows a remarkably high homology of the PMSat sequence isolated in the different rodentia species when compared with *P. eremicus* (PER), ranging from 94% in PSU to 100% in MAR (Figure 5, Table 2). Lower homology (76% to 78%) was obtained when compared with the arthropod and echinoderm partial sequences. Phylogenetic tree construction revealed the division of two main branches, one harbouring the rodent species and other with BTE and LVA. For the rodent's no specific clustering pattern of the studied satellite was observed; the monomers were not grouped into distinct clusters and couldn't be separated into species-specific branches (Figure 6). This shows the absence of complete fixation of any nucleotide variant since the probable presence of this satDNA in the genome of a hypothetical ancestor of the rodent species. The variability existing between these and the remaining analyzed species indicates that PMSat accumulated some variation since the Bilateria ancestral till the rodentia ancestral, but from that point (about 19 million years ago according with Steppan et al. 2004) this sequence seems to be kept highly conserved. The inexistence of species-specific mutations, and very high homology of the majority of the PMSat monomers from the rodents analyzed can be explained by the occurrence of non-concerted evolution, as observed for other satellite DNAs (e.g. Plohl et al. 2010a).

The comparison of sequence conservation between species, have been extensively used in the identification of functional elements within large complex genomes (Pennachio et al. 2001, 2006). The protein-coding regions of the mouse and human genomes are about 85% identical, and they shared a common ancestral approximately 80 million years ago (Jegga and Aronow, 2006). The same level of conservation is found for the PMSat orthologs in rodents, and slightly lower conservation is obtained when we analyse the orthologs present in species that shared a

common ancestral at 635 million years ago with PER (Edgecombe et al. 2011). Inside the monomeric unit of a satellite family, short sequence segments can act as motifs involved in putative functional interactions, meaning that the information is not necessarily contained in the whole monomer sequence. For this reason, parts of the monomer sequence or particular nucleotide positions evolve under different mutation rates (reviewed in Plohl 2010). Uneven distribution of variability was observed in satellites from different organisms (e.g. Romanova et al. 1996, Petrovic and Plohl 2005). In the light of these facts our results strongly suggest that PMSat is under evolutionary constraints in the species studied and probably constitutes a functional element in their genomes. To support our hypothesis about the putative functionality of this satellite DNA family, we found a cDNA sequence highly similar to the PMSat in the rodent species *Peromyscus maniculatus biirdi* (Table S1), showing that this sequence is also present in a sister species of *Peromyscus eremicus* and proving that its transcribed, at least in this species. Transcripts of satellite DNAs have been identified and involved in diverse genomic functions (Chiodi et al. 2004; Grewal and Elgin 2007). In resume, in the future it will be mandatory to depict the satellite transcripts of PMSat once this data strongly suggest that this satellite DNA is effectively transcribed.

Using a new methodology based in real-time quantification, PMSat number of copies was determined in *P. eremicus* genome, and relative quantification was performed to the other rodent studied species. Real-time PCR has been previously used to access repetitive sequences copy number, combining the use of standard primers with SYBR Green I chemistry (e.g. Navajas-Perez et al. 2009). In the present work we applied, an innovative methodology using specific TaqMan probe/primers. Some drawbacks can be anticipated for this method once is PCR based, and for this reason some variation can remain hidden, or some bias that distort the observed representation of the satDNA variability can be introduced by the primer specificity. To overcome this scenario the primer/probe were designed for the highest conserved region (between all species) within PMSat. Absolute quantification showed that this satellite comprises at least 20% of *P. eremicus* genome (Figure 7, Table 2), and relative quantification revealed a much lower copy number of PMSat in the other species genomes (about 10^6-fold fewer copies) when compared with *P. eremicus* (Figure 7, Table 2). This explains why PMSat was not detected in all the other species using hybridization methods such as FISH and southern blot. Once this satellite DNA presents a high interspecies sequence identity and a wide range in number of copies, it can be hypothesized that its evolutionary pathway occurred throughout copy number fluctuation and that a prominent amplification of this satellite occurred in *Peromyscus eremicus*. The mechanisms proposed to be responsible for amplification of repeated DNA are unequal crossing over,

replication slippage and rolling circle amplification (Walsh 1987). The variation in copy number of PMSat in PER may be associated with this species karyotype evolution. A high degree of karyotypic conservation within *Peromyscus* genus has been observed, being the species karyotype variations attributable to heterochromatin additions and pericentric inversions (Robbins and Baker, 1981; Rogers et al., 1984). PMSat satellite DNA, as part of *P. eremicus* constitutive heterochromatin, surely contributed to the CH repatterning in this species. This satellite DNA dynamics by means of amplification events may have lead to the CH additions, resulting in the present large CH blocks enriched by PMSat.

In summary, this work presents the molecular and cytogenetic characterization of a novel satellite DNA family, PMSat, highly represented in the rodent *P. eremicus* genome. Orthologous sequences were isolated from related Cricetidae species, and found to be present in one arthropod and one echinoderm species, revealing high sequence homology**,** and making it the oldest satellite DNA sequence described so far. We have found evidences that this high sequence conservation is related with a possible functional activity in the genome. This satellite seems to have been amplified in *P. eremicus,* representing a significant part of its genome. This work constitutes an important starting point as further and deeper study of this interesting sequence will certainly shed some light into the enigmatic nature of satellite DNA.

ACKNOWLEDGEMENTS

This work was supported by the project POCI/BIA-BCM/ 58541/2004, the PhD grants SFRH/BD/25813/2005 and SFRH/BD/41942/2007 and a research position on Animal Genomics of the "Sistema Científico e Tecnológico Nacional – Ciência 2007", all from the Science and Technology Foundation (FCT) from Portugal and also by the Czech Ministry of Agriculture (MZE 0002716202).

REFERENCES

Adega F, Chaves R, Guedes-Pinto H. 2007. Constitutive heterochromatin characterization of White-lipped and Collared peccaries (Tayassuidae). J. Genet. 86:19–26.

Alkan C, Cardone MF, Catacchio CR, Antonacci F, O'Brien SJ, Ryder OA, Purgato S, Zoli M, Della Valle G, Eichler EE, Ventura M (2010) Genome-wide characterization of centromeric satellites from multiple mammalian genomes. Genome Res. 21(1):137-45.

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

Chiodi I, Corioni M, Giordano M, Valgardsdottir R, Ghigna C, Cobianchi F, Xu RM, Riva S, Biamonti G (2004) RNA recognition motif 2 directs the recruitment of SF2/ASF to nuclear stress bodies. Nucleic Acids Res 32: 4127–4136.

Committee for standardization of chromosomes of *Peromyscus*. 1977. Standard katyotype of the deer mice, *Peromyscus* (Rodentia). Cytogenet. Cell Genet. 19: 38–43.

Deaven LL, Vidal-Rioja L, Jett JH, Hsu TC. 1977. Chromosomes of *Peromyscus* (rodentia, cricetidae). VI. The genomic size. Cytogenet. Cell Genet. 19:241–249.

Dover GA (1982) Molecular drive: a cohesive mode of species evolution. Nature 299: 111–117.

Dover GA. 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet. 2:159–165.

Edgecombe GD, Giribet G, Dunn CW, Hejnol A, Kristensen RM, Neves RC, Rouse GW, Worsaae K, Sørensen MV (2011) Higher-level metazoan relationships: recent progress and remaining questions. Org Divers Evol 11(2):151-172 .

Elder JF, Turner BJ. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. 70:297–320.

Glasper ER, Devries AC. 2005. Social structure influences effects of pair-housing on wound healing. Brain Behav Immun 19(1):61-8.

Grewal SI, Elgin SC. 2007. Transcription and RNA interference in the formation of heterochromatin. Nature 447:399–406.

Hall LE, Mitchell SE, O'Neill RJ. 2012. Pericentric and centromeric transcription: a perfect balance required. Chromosome Res 20(5):535-46.

Hamilton MJ, Hong GH, Wichman A. 1992. Intragenomic movement and concerted evolution of satellite DNA in *Peromyscus*: evidence from *in situ* hybridization. Cytogenet. Cell Genet. 60:40–44.

Hazen MW, Kuo MT, Arrighi FE. 1977. Genome analysis of *Peromyscus* (Rodentia, Cricetidae). Chromosoma 64:133–142.

Jegga AG, Aronow BJ. 2006. Evolutionary conserved noncoding DNA. Encyclopedia of Life Sciences. DOI: 10.1002/9780470015902.a0006126.pub2.

John B. 1988. The biology of heterochromatin. In: Verma RS (ed) Heterochromatin: Molecular and structural Aspects. Cambridge University Press, Cambridge, pp. 1–128.

Komissarov AS, Gavrilova EV, Demin SJ, Ishov AM, Podgornaya OI. 2011. Tandemly repeated DNA families in the mouse genome. BMC Genomics 12:531.

Kubickova S, Cernohorska H, Musilova P, Rubes J. 2002. The use of laser microdissection for the preparation of chromosome-specific painting probes in farm animals. Chromosome Res. 10:571–577.

Li YC, Cheng YM, Hsieh LJ, Ryder OA, Yang F, Liao SJ, Hsiao KM, Tsai FJ, Tsai CH, Lin CC. 2005. Karyotypic evolution of a novel cervid satellite DNA family isolated by microdissection from the Indian muntjac Y-chromosome. Chromosoma 114:28-38.

Li YX, Kirby ML. 2003. Coordinated and conserved expression of alphoid repeat and alphoid repeat-tagged coding sequences. Dev Dyn 228(1):72-81.

Livak K, Schmittgen T. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods 25: 402–408.

Louzada S, Paço 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.

Martienssen RA. 2003. Maintenance of heterochromatin by RNA interference of tandem repeats. Nat. Genet. 35:213–214.

Meles S, Adega F, Guedes-Pinto H, Chaves R. 2008. The karyotype and sex chromosomes of *Praomys tullbergi* (Muridae, Rodentia): a detailed characterization. Micron 39:559–568.

Mravinac B, Plohl M, Ugarković D .2005. Preservation and high sequence conservation of satellite DNAs suggest functional constraints. J Mol Evol 61(4):542-50.

Musser GG, Carleton M D. 2005. Superfamily Muroidea. In: Wilson DE, Reeder DM, editors. Mammal species of the world: a taxonomic and geographic reference. Baltimore: Johns Hopkins University Press. pp 894–1531.

Navajas-Pérez R, Quesada del Bosque ME, Garrido-Ramos MA. 2009. Effect of location, organization, and repeat-copy number in satellite-DNA evolution. Mol. Genet. Genomics 282(4):395-406.

Paço A, Adega F, Guedes-Pinto H, Chaves R. 2009. Hidden heterochromatin: Characterization in the Rodentia species *Cricetus cricetus*, *Peromyscus eremicus* (Cricetidae) and *Praomys tullbergi* (Muridae). Genet. Mol. Biol. 32:58-68.

Pathak S, Hsu TC, Arrighi FE. 1973. Chromosomes of *Peromyscus* (Rodentia, Cricetidae). IV. The role of heterochromatin in karyotypic evolution. Cytogenet. Cell Genet. 12:315–326.

Pauciullo A, Kubickova S, Cernohorska H, Petrova K, di Berardino D, Ramunno L, Rubes J. 2006. Isolation and physical localization of new chromosome-specific centromeric repeats in farm animals. Veterinarni Medicina 51 224–231.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444: 499–502.

Pennacchio LA, Olivier M, Hubacek JA, Cohen JC, Cox DR, Fruchart JC, Krauss RM, Rubin EM. 2001. An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* 294: 169–173.

Petrović V, Pérez-García C, Pasantes JJ, Šatović E, Prats E, Plohl M. 2009. A GC-rich satellite DNA and karyology of the bivalve mollusk *Donax trunculus:* a dominance of GC-rich heterochromatin. Cytogenet Genome Res 124:63–71.

Plohl M, Luchetti A, Meštrović N, Mantovani B. 2008. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene 409:72–82.

Plohl M, Petrović V, Luchetti A, Ricci A, Satović E, Passamonti M, Mantovani B. 2010a. Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. Heredity 104(6):543-51.

Plohl M. 2010b. Those mysterious sequences of satellite DNAs. Periodicum Biologorum 112(4) 403–410.

Robbins LW, Baker RJ. 1981. An assessment of the nature of chromosomal rearrangements in 18 species of *Peromyscus* (Rodentia: Cricetidae). Cytogenet. Cell Genet. 31:194–202.

Robles F, de la Herran R, Ludwig A, Ruiz Rejon C, Ruiz Rejon M, Garrido-Ramos MA. 2004. Evolution of ancient satellite DNAs in sturgeon genomes. Gene 338: 133–142.

Rogers DS, Greenbaum IF, Gunn SJ, Engstrom MD. 1984. Cytosystematic value of chromosomal inversion data in the genus Peromyscus (Rodentia, Cricetidae). J. Mammal. 65: 457–465.

Romanova LY, Deriagin GV, Mashkova TD, Tumeneva IG, Mushegian AR, Kisselev LL, Alexandrov IA. 1996. Evidence for selection in evolution of alpha satellite DNA: the central role of CENP-B/pJ alpha binding region. J Mol Biol 261(3):334-40.

Rozen, S. and H. Skaletsky. 2000. Primer3 on the WWW for General Users and for Biologist Programmers, p. 365-386. In Krawetz, S. and S. Misener (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology, Humana Press, Totowa, NJ.

Shorter KR, Crossland JP, Webb D, Szalai G, Felder MR, Vrana PB. 2012. Peromyscus as a Mammalian epigenetic model. Genet Res Int 2012:179159.

Steppan S, Adkins R, Anderson J. 2004. Phylogeny and Divergence-Date Estimates of Rapid Radiations in Muroid Rodents Based on Multiple Nuclear Genes. Syst. Biol. 53:533–553.

Stimpson KM, Sullivan BA. 2010. Epigenomics of centromere assembly and function. Curr Opin Cell Biol 22(6):772–780.

Sumner AT. 1972. A simple technique for demonstrating centromeric heterochromatin. Exp Cell Res 75:304–306.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28(10):2731-9.

Walsh JB. 1987. Persistence of tandem arrays: implications for satellite DNA and simple-sequence DNAs. Genetics 115:553-567.

Wong LH, Brettingham-Moore KH, Chan L, Quach JM, Anderson MA, Northrop EL, Hannan R, Saffery R, Shaw ML, Williams E, Choo KH. 2007. Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res 17(8):1146-60.

SUPLEMENTARY TABLE

Table S1 – Summary of the analysis in all PMSat isolated clones and Genbank sequences.

			Clones		Monomer		
Phylum	**Species**	**%Homology**	**Designation**	**Isolation**	**Length (bp)**	**% AT**	**Access. number**
Craniata	*P. eremicus*		PERm40	Microdissection	464	54	GQ902036
			PERm57	Microdissection	463	54	KC351938
			PERp25	PCR	396	55	KC351941
			PERp45	PCR	443	54	KC351942
			PERp62	PCR	386	55	KC351943
			PERHaeIIIA	RE HaeIII	346	56	KC351939
			PERHaeIIIB	RE HaeIII	343	57	KC351940
	C. cricetus	95	CCRpA1	PCR	446	54	KC351944
			CCRpB1	PCR	399	54	KC351945
			CCRpC1	PCR	442	54	KC351946
			CCRpR1	PCR	411	54	KC351947
	M. arvalis	94	MARpA1	PCR	430	54	KC351948
			MARAp15	PCR	431	54	KC351949
			MARpB1	PCR	430	54	KC351950
	P. sungorus	100	PSUpA11	PCR	387	54	KC351951
			PSUpA25	PCR	392	54	KC351952
			PSUpE1	PCR	443	54	KC351953
			PSUpR1	PCR	425	57	KC351954
				Genbank Retrieved Sequences			
	P. mani bairdii	83	PMA	-	310	-	GH487254.1
Arthropoda	*B. terrestris*	76	BTE	-	315	53	AELG01007655.1
Echrinodermata	*L. variegates*	78	LVA	-	311	53	AGCV01272891.1

3. CANCER CHROMOSOMES AND CELL LINE MODELLING

Cancer chromosomes, present a high level of dynamics and the ability to constantly evolve. Understanding the dynamic nature of the cancer genome is important to comprehend the mechanisms of genetic heterogeneity and population diversity, which is the genetic basis for cancer formation. For this purpose it was performed, the genetic/cytogenetic characterization of two DMBA-induced rat mammary tumor cell lines, being this work described in the following paper.

Breast cancer cell lines have shown to be valuable tools in the investigation of the role of genomic alterations in cancer progression and as important resources for the discovery of new breast cancer genes. In particular DMBA-induced rat tumor models proved to be useful models for studying hormone-dependent breast cancer. The characterization of the rat mammary tumor sister cell lines HH-16 cl.2/1 and HH-16.cl.4 was performed using a combination of multiple techniques, such as, G-banding, chromosome painting, BAC FISH, real-time reverse transcription quantitative PCR (RT-qPCR) and RNA FISH. The cytogentic analysis showed distinct karyotypic changes for both cell lines, namely very different levels of ploidy. Chromosome rearrangements in these cell lines lead to an increase in copy number of two oncogenes *Mycn* and *Erbb2*. These genes expression was assessed showing *Erbb2* overexpression in HH-16.cl.4, but not in the sister cell line HH-16 cl.2/1. Moreover this gene expression appears to be affected by global genome demethylation, after cells treatment with 5-Aza-2'-Deoxicitidine. This work illustrates: 1) the karyotype restructuring effect in cancer progression by affecting cancer-related genes; 2) besides, despite having the same initial genetic background the studied cell lines culminated in different outcomes, suggesting different mechanisms involved in tumor progression; 3) the development of suitable *in vitro* models of human breast cancer is of crucial importance in the study of cancer and, consequently, in the development of new therapeutics; 4) the obtained results provide a platform for future studies on tumor progression and encourage the use of these cell lines as a model. We are confident that this work will contribute to the validation of this cellular model and to its use in future studies.

3.1. Defining the Sister Rat Mammary Tumor Cell Lines HH-16 cl.2/1 and HH-16.cl.4 as an *In Vitro* Cell Model for *Erbb2*

OPEN ACCESS Freely available online

PLoS one

Defining the Sister Rat Mammary Tumor Cell Lines HH-16 cl.2/1 and HH-16.cl.4 as an *In Vitro* Cell Model for *Erbb2*

Sandra Louzada[1], Filomena Adega[1], Raquel Chaves[1,2]*

1 Center of Genomics and Biotechnology, Institute for Biotechnology and Bioengineering, University of Trás-os-Montes and Alto Douro (IBB/CGB-UTAD), Vila Real, Portugal, **2** Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

Abstract

Cancer cell lines have been shown to be reliable tools in genetic studies of breast cancer, and the characterization of these lines indicates that they are good models for studying the biological mechanisms underlying this disease. Here, we describe the molecular cytogenetic/genetic characterization of two sister rat mammary tumor cell lines, HH-16 cl.2/1 and HH-16.cl.4, for the first time. Molecular cytogenetic analysis using rat and mouse chromosome paint probes and BAC/PAC clones allowed the characterization of clonal chromosome rearrangements; moreover, this strategy assisted in revealing detected breakpoint regions and complex chromosome rearrangements. This comprehensive cytogenetic analysis revealed an increase in the number of copies of the *Mycn* and *Erbb2* genes in the investigated cell lines. To analyze its possible correlation with expression changes, relative RNA expression was assessed by real-time reverse transcription quantitative PCR and RNA FISH. *Erbb2* was found to be overexpressed in HH-16.cl.4, but not in the sister cell line HH-16 cl.2/1, even though these lines share the same initial genetic environment. Moreover, the relative expression of *Erbb2* decreased after global genome demethylation in the HH-16.cl.4 cell line. As these cell lines are commercially available and have been used in previous studies, the present detailed characterization improves their value as an *in vitro* cell model. We believe that the development of appropriate *in vitro* cell models for breast cancer is of crucial importance for revealing the genetic and cellular pathways underlying this neoplasy and for employing them as experimental tools to assist in the generation of new biotherapies.

Citation: Louzada S, Adega F, Chaves R (2012) Defining the Sister Rat Mammary Tumor Cell Lines HH-16 cl.2/1 and HH-16.cl.4 as an *In Vitro* Cell Model for *Erbb2*. PLoS ONE 7(1): e29923. doi:10.1371/journal.pone.0029923

Editor: Eliana Saul Furquim Werneck Abdelhay, Instituto Nacional de Câncer, Brazil

Received October 13, 2011; **Accepted** December 6, 2011; **Published** January 10, 2012

Funding: This work was supported by the project POCI/BIA-BCM/58541/2004, the PhD grant SFRH/BD/25813/2005 and a research position on Animal Genomics of the "Sistema Científico e Tecnológico Nacional - Ciência 2007", all from the Science and Technology Foundation (FCT) from Portugal. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rchaves@utad.pt

Introduction

Breast cancer is one the most commonly occurring cancers among women and has been described as a molecularly heterogeneous disease. Genetic studies of breast cancer rely on the use of primary tumors, paraffin-embedded samples or cell lines. Breast cancer cell lines present the great advantage of being readily available, and the full characterization of cell line models has been shown to provide valuable insights regarding the degree of complexity of the polygenetic etiology of breast cancer and the biological mechanisms that characterize this disease [1]. Chemically induced carcinogenesis of the rat mammary gland has been used extensively to investigate breast cancer. In rat models, the carcinogenic compound 7,12-dimethylbenz[a]anthrazene (DMBA) is frequently used to induce tumors, and DMBA-induced rat mammary tumors and sarcomas are useful cancer models [2,3,4]. Using the evolutionary conservation of gene segments as a guide, animal models, such as the rat, constitute powerful tools to decipher pathways and genes involved in tumorigenesis [4]. Moreover, researchers now have access to powerful web servers and databases in which syntenic regions can be easily identified and associated with a great amount of information regarding human and rat genetics. The available animal tumor cell lines are often poorly characterized from a cytogenetic/genetic point of view, reducing their usefulness as cell models.

Here, we present the molecular cytogenetic/gene expression characterization of two DMBA-induced rat mammary tumor cell lines: the HH-16 cl.2/1 fibrosarcoma cell line and the HH-16.cl.4 adenocarcinoma cell line. The choice of these cell lines was based on two factors: first, the reliability of both cell lines as models has been demonstrated in investigations of the effects of glucocorticoid hormones on cell morphology and proliferation and the stability of cultured rat cells after infection with Moloney murine sarcoma virus [5,6,7]; second, these cell lines are commercially available to the entire scientific community, and when they are properly characterized, they may constitute reliable cell models for breast cancer research.

Performing a chromosome count constitutes a mandatory step in the cytogenetic characterization of cell lines, allowing an overview of their genetic variability and stability. Of the two investigated cell lines, only HH-16 cl.2/1 presents low polyploidy levels, indicating a certain degree of stability, and for this reason, detailed cytogenetic characterization was restricted to this cell line. The methodology used in this study included fluorescent *in situ* hybridization with rat and mouse chromosome paint probes to identify chromosomal rearrangements, complemented with BAC/PAC clones that assisted in the accurate detection of the

breakpoint regions of the rearrangements as well as complex chromosome abnormalities. The increase in the number of copies (determined with specific BAC clones) of the *Mycn* and *Erbb2* genes detected in this analysis was of particular note. The development and progression of cancer are characterized by a variety of genetic modifications in mechanisms that control genome stability, including alterations in oncogenes [8]. *ERBB2* oncogene amplification constitutes one of the most important genetic alterations associated with human breast cancer and was found to be correlated with poor patient prognosis by Slamon and colleagues [9]. *MYCN* oncogene amplification is characteristic of human neuroblastomas, being found in 20% of these childhood cancers, and has been observed to be involved in breast tumorigenesis, with up-regulation being detected in inflammatory breast cancer [10]. In the present study, the amplification status of the rat counterpart *Erbb2* and *Mycn* genes was analyzed in the HH-16 cl.2/1 and HH-16.cl.4 rat cell lines by fluorescent *in situ* hybridization, and the expression of these genes was assessed by real-time reverse transcription quantitative PCR (RT-qPCR) complemented and validated with an RNA fluorescent *in situ* hybridization (RNA FISH) analysis.

Abnormal patterns of DNA methylation have been found in several types of human cancer. DNA hypermethylation may result in gene expression silencing and loss of protein function as well as being associated with cancer progression [11]. Currently, epigenetic therapies aim to restore hypomethylation and to reverse gene silencing induced by hypermethylation [12]. A cytosine analogue established as a potent inhibitor of DNA methylation, 5-Aza-2′-Deoxicitidine (decitabine), [13] has been used in both preclinical models and in cancer patients [14]. However, global demethylation effects in tumor cells treated with this agent remain poorly understood. Early studies suggest that the loss of DNA methylation is a common event in tumorigenesis [15,16]. To evaluate global genome demethylation effects on gene expression in the studied rat tumor cell lines, cells were treated with 5-Aza-2′-Deoxicitidine, and *Mycn* and *Erbb2* expression was subsequently determined.

The cytogenetic and genetic characterization of the HH-16 cl. 2/1 and HH-16.cl.4 rat mammary cell lines, complemented with expression profiling analysis of the *Mycn* and *Erbb2* oncogenes and verification of the influence of global demethylation on the expression of these genes validates the use of these cell lines as models for breast cancer research.

Materials and Methods

Cell culture and chromosome preparation

The HH-16 cl.2/1 and HH-16.cl.4 cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). Both cell lines were established from ascitic fluid of the same female Sprague-Dawley rat with a mammary tumor produced by injection of cultured cells from a DMBA-induced mammary tumor. When injected into rats, HH-16.cl.2/1 cells have been found to produce fibrocarcinomas while HH-16.cl.4 cells generate adenocarcinomas. Both cell lines were grown in RPMI 1640 supplemented with 10% FCS, 1% 200 mM L-Glutamine and 1% of a Penincilin-Streptomycin antibiotic mixture (all from Gibco, Life Technologies). The HH-16.cl.4 cell medium was also supplemented with 1% 100 mM Sodium Pyruvate MEM (Gibco, Life Technologies). Both cultures were passaged at confluence using 0.25% trypsin (1×) with EDTA in Hanks' balanced salt solution (Gibco, Life Technologies). For both cell lines, metaphase chromosomes were obtained by treatment with colcemide (10 µg/ml, Invitrogen, Life Technologies) for 45 minutes followed by hypotonic solution (0,05 M KCl, 30 minutes, 37°C) and fixation with methanol:acetic acid (3:1), and the samples were then dropped onto microscope slides.

GTD-banding

Air-dried slides from the HH-16 cl.2/1 cell line were aged at 65°C overnight and then subjected to standard G-banding procedures with trypsin [17]. DAPI was used for staining (instead of routine Giemsa staining) to obtain a better contrast [18]. Inversion of the DAPI color in Adobe Photoshop (version 7.0) revealed the chromosome G-banding pattern (GTD-banding, G-bands revealed by trypsin with DAPI).

Chromosome painting

Chromosome paint probes from *Rattus norvegicus* (RNO) and *Mus musculus* (MMU) were kindly provided by Dr. Johannes Wienberg and Dra. Andrea Kofler from Chrombios GmbH, Germany. Chromosome-specific probes were labeled by DOP-PCR using the universal primers 6MW (for RNO paints) and F/S (for MMU paints) together with incorporation of digoxigenin-11-dUTP (Roche) or biotin-16-dUTP (Roche).

Fluorescent *in situ* hybridization experiments were performed according to [19]. RNO paint probes were hybridized to chromosomes from both the HH-16 cl.2/1 and HH-16.cl.4 cell lines while MMU paint probes were only hybridized to HH-16 cl.2/1 chromosomes. The most stringent post-hybridization wash was 50% formamide/2×SSC at 37°C, and probe detection was performed using antidigoxigenin-5′TAMRA (Roche) and FITC conjugated with avidin (Vector Laboratories).

Probe construction from BAC/PAC clones and FISH

BAC and PAC clones were obtained from the BACPAC Resources Center from Children's Hospital Oakland Research Institute (http://bacpac.chori.org/). The acquired clones were RP31-262B4, CH230-208E5, RP31-202O5, RP31-039D3, CH230-10B5 (for rat chromosome 6); CH230-174M18, CH230-9A5, CH230-215E5, CH230-27O13, CH230-165C24, CH230-117H20 (for rat chromosome 15); and CH230-162I16, CH230-276G18 and CH230-305O21 (rat *Erbb2* predicted clones). DNA from the clones was purified using QUIAGEN Plasmid Purification Kit as recommended by the manufacture (QIAGEN) and labeled with tetramethyl-rhodamine-5-dUTP (Roche) by Nick Translation (Abbott) for 2 hours at 15°C. Labeled probes were precipitated with an excess of sonicated normal rat genomic DNA and dissolved in hybridization solution. FISH procedures were performed as described in the Chromosome Painting section using chromosome preparations of the HH-16 cl.2/1 and HH-16.cl.4 cell lines.

For rat *Erbb2*, three BAC clones were selected *in silico* using the NCBI Map Viewer online resource (http://www.ncbi.nlm.nih.gov/mapview/) and then tested for the presence of *Erbb2* and mapped by FISH (see Figure S1). Briefly, a rat *Erbb2* genomic sequence obtained from the Ensembl database (http://www.ensembl.org/) was used to design specific primers for the amplification of this gene in the three BAC clones. PCR was performed with purified plasmid DNA from rat BAC clones (as described above), and PCR products with the predicted sizes were excised from 1.2% agarose gels, purified and sequenced. FISH procedures were performed as described above.

FISH image capture, processing and analysis

Chromosomes were observed using a Zeiss AxioImager Z1 microscope, and images were captured using an Axiocam MRm

digital camera with LSM 510 software (version 4.0 SP2). Digitized photos were prepared in Adobe Photoshop (version 7.0); image optimization included contrast and color adjustments that affected the whole image equally. Karyotypes were constructed following the nomenclature for rat chromosomes described by Levan [20], and chromosome rearrangements were described according to ISCN (2009) [21].

Gene amplification criteria

Gene amplification was calculated based on the ratio between the number of gene signals and the number of chromosomes harboring that gene. *Mycn* amplification was defined for *Mycn*/RNO6≥2 and *Erbb2* amplification by *Erbb2*/RNO10≥2, with 2 being the cut-off value for both. Rat PAC clone RP31-202O5 was used to identify the *Mycn* gene, rat BAC clone CH230-162I16 allowed detection of *Erbb2*, and rat paint probes were used to identify chromosomes 6 and 10. Additional copies of each gene, detected by FISH at levels equal to or no more than 4-fold higher (when compared with normal gene number) were considered to be a *Mycn* or *Erbb2* gain.

RNA isolation and reverse transcription quantitative real-time PCR

Total RNA from rat cell lines was isolated using the mirVana Isolation Kit (Ambion) following the manufacturer's recommendations. Expression experiments were performed using the TaqMan® RNA-to-CT^{TM} 1-Step Kit (Applied Biosystems). The TaqMan Gene Expression Assay Mixes (primer/probe sets) used were beta-actin (Rn00667869_m1) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Rn01749022_g1) as reference genes and *Mycn* (Rn01473353) and *Erbb2* (Rn00566561_m1) as targets (all assays were from Applied Biosystems). The 20 µl reactions included 2 µl of RNA sample (50 ng/µl), 1 µl of the primer/probe assay mixture, 10 µl of PCR Master Mix, 0.5 µl of RT enzyme mix (Applied Biosystems) and 6.5 µl of DEPC-treated water. The reactions were carried out in a 96-well optical plate at 48°C for 15 min and 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. PCR was carried out in the ABI 7500 Fast Real Time PCR system (Applied Biosystems). All reactions were performed in triplicate, and negative controls (without template) were run for each master mix. SDS software version 1.4 (Applied Biosystems) was applied for comparative analysis, and the relative expression level was normalized with multiple reference genes. The $2^{-\Delta\Delta CT}$ method [22] was used to calculate fold changes in the expression levels of the genes of interest using a control RNO sample as a calibrator. Expression fold changes≥3 were considered relevant.

Statistical analysis

Student's t-test was used to compare the data obtained. Values were expressed as the mean ± SD, and differences were considered statistically significant at $p<0.05$, representing the 95% confidence interval of the mean expression level.

RNA fluorescent *in situ* hybridization

RNA FISH was performed using the QuantiGene ViewRNA plate-based assay kit (Panomics) following the manufacturer's recommendations with some modifications. Briefly, HH-16 cl.2/1 and HH-16.cl.4 cells were grown on polysine coated glass slides, fixed using 8% formaldehyde, dehydrated in ethanol (50%–70%–100%) and held at 4°C overnight. Then, cells were rehydrated, permeabilized and hybridized as recommended, except that protease digestion was optimized for each cell line. The RNA target was human *ERBB2* (Panomics), and the reference RNA was human/rat/mouse 18S RNA (Panomics). Confocal fluorescence images were captured on an LSM 510 META with a Zeiss Axio Imager Z1 microscope and LSM 510 software (version 4.0 SP2). For each scan, the same microscope settings were employed for all images to normalize the results. The lasers used were as follows: argon (488 nm) set at 12.9%, helium–neon (543 nm) set at 50.8% and Diode (405 nm) set at 9.9%. The pinhole was set to 96 µm (1.02 airy units) for argon laser, 102 µm (0.98 airy units) for helium–neon laser and 112 µm for the Diode laser using a 63× objective. Images were captured at a scan speed of 5 (3.30 µs) with 1 µm thick Z sections and processed using the "3D Viewer" plug-in for ImageJ. Twenty slide fields were randomly selected and analyzed by counting the number of signals in each cell.

5-Aza-2′-Deoxicitidine demethylation

For global genome demethylation, the media for the HH-16 cl.2/1 and HH-16.cl.4 rat cell lines were supplemented with different concentrations of 5-Aza-2′-Deoxicitidine (Sigma) (3 µM, 10 µM and 30 µM) for 72 hours. Every 24 hours, the medium was changed, followed by the addition of 5-Aza-2′-Deoxicitidine. After the 72 h period, a sample of the cells was collected for RNA extraction, and remaining cells were allowed to grow without drug treatment for another 72 hours, after which they were also subjected to RNA extraction. Additionally, the HH-16 cl.2/1 and HH-16.cl.4 cell lines were grown without 5-Aza-2′-Deoxicitidine as controls.

Results

HH-16 cl.2/1 and HH-16.cl.4 morphological features and ploidy

Phase contrast microscopy analysis of the HH-16 cl.2/1 cell line revealed a fibroblastoid cell morphology, with the cells growing in a criss-cross pattern (Figure 1A). The HH-16.cl.4 line presented distinct cell morphology, with epitheloid-shaped cells growing in monolayer (Figure 1B).

Chromosome number analysis of the HH-16 cl.2/1 rat mammary fibrosarcoma cell line was carried out throughout the examination of 75 cells. The results show that this cell line presents a near diploid karyotype (Figure 1C), with 2n = 42 being the normal chromosome number for this species. The HH-16 cl.2/1 modal chromosome number is 40–41 (2n = 39–43 is the ploidy referenced in the available cell line description in the DSMZ database), and the polyploidy levels of this line are reduced (less than 3%), with only two cells being observed with a nearly tetraploid karyotype, containing 79 and 82 chromosomes. Chromosome number analysis was also performed for the HH-16.cl.4 rat mammary tumor cell line based on examination of 75 cells. This cell line presents a nearly tetraploid karyotype (Figure 1D) with a modal number of 79–80 (4n = 79–84 is the ploidy referenced in the available cell line description in the DSMZ database). When compared with the sister cell line, HH-16.cl.4 shows a wider range of cells with different chromosome numbers, with approximately 9% of cells being observed to have a nearly triploid karyotype (60–68 chromosomes) and 2% of cells exhibiting a nearly heptaploid karyotype (151–155 chromosomes). This variability in ploidy might be reflected in significant levels of karyotypic heterogeneity within this cell line, which are indicative of a higher order of complexity and instability when compared with the HH-16 cl.2/1 cell line. This observation restricted large-scale cytogenetic characterization to only the HH-16 cl.2/1 cell line, which apparently presents a more "stable" karyotype.

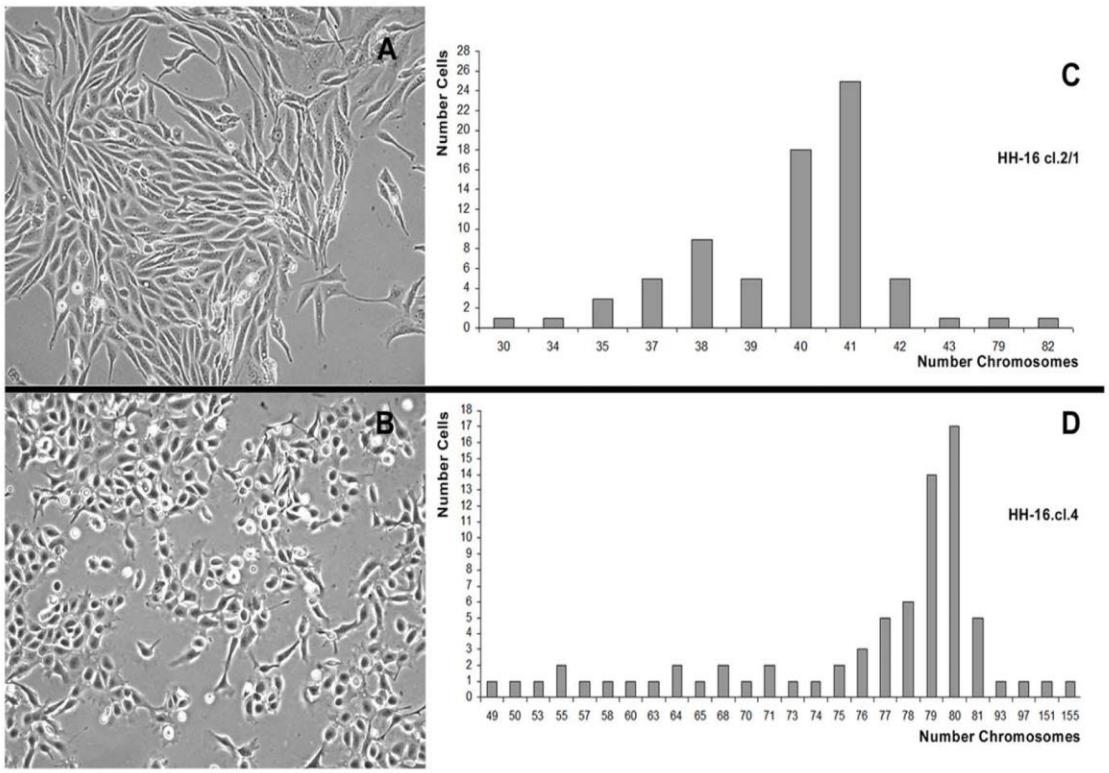


Figure 1. Morphology (×10) and ploidy of HH-16 cl.2/1 and HH-16.cl.4 cells. HH-16 cl.2/1 cell line presenting a fibroblastoid cell morphology with the cells growing in a criss-cross pattern (A), and HH-16.cl.4 cells morphology showing epitheloid shaped cells (B). Chromosome count analysis revealed a near-diploid karyotype with low level of polyploidy in HH-16 cl.2/1 (C) while a wide range of different chromosome numbers were observed in HH-16.cl.4, being the most representative the near-tetraploid karyotype (D).
doi:10.1371/journal.pone.0029923.g001

Cytogenetic Characterization

Identification of clonal chromosome rearrangements. A combination of G-banding and fluorescent *in situ* hybridization was used in the cytogenetic characterization of clonal rearrangements for HH-16 cl.2/1. Paint probes for each rat chromosome (RNO1-20, X) and from mice (MMU19) were successfully hybridized to HH-16 cl.2/1 cell line chromosomes (Figure 2A–E), revealing a total of 13 rearrangements, both numeric and structural in character, involving chromosomes RNO1, RNO3, RNO4, RNO6, RNO7, RNO11, RNO13, RNO15, RNO18, RNO19 and RNOX. Three numerical changes were observed, involving a whole chromosome gain (+1) and two losses (−X, −18), with X chromosome monosomy being one of the most representative rearrangements. The rat chromosomes associated with greater numbers of rearrangements were RNO1, RNO6, RNO15 and RNO19. More structural than numerical aberrations were observed, and derivative chromosomes resulting from translocations were the predominant structural abnormalities. The most frequent structural chromosome rearrangements identified using this approach were as follows: t(3;11)(p12;p12), der(4;15)(q10;p10), der(7)t(1;7)(q51;q36), del(13)(p13) and der(19)t(6;19). Almost all rearrangements were unbalanced, involving gains and losses of chromosome segments. G-banding analysis allowed us to determine that the region of chromosome 1 involved in the rearrangement der(7)t(1;7)(q51;q36) was the terminal region. To confirm this analysis, the MMU19 paint probe was used because it is syntenic to this region in the rat. This approach confirmed that the region presented by the derivative chromosome is 1qter→1q51 (Figure 2E).

High-resolution chromosome rearrangement characterization and identification of breakpoint regions. To refine the cytogenetic characterization, a total of 8 BAC and 3 PAC clones were hybridized to HH-16 cl.2/1 cell line chromosomes. The selected clones contained regions of rat chromosomes 6 (RP31-262B4, CH230-208E5, RP31- 202O5, RP31-039D3, CH230-10B5) and 15 (CH230-174M18, CH230-9A5, CH230-215E5, CH230-27O13, CH230-165C24, CH230-117H20), which were physically mapped in a previous study [23]. The BAC/PAC results allowed the identification of the breakpoint regions of the derivative chromosomes involving RNO6 to der(19)t(6;19)(q14;q12), der(19)t(4;19)(q31;p11)t(6;19)(q14;q12) and der(18;19)t(18;19)(p10,q10)t(6;19)(q14;q12), assigning the location of the breakpoint in all of these chromosomes to band 6q14, above the region included within clone RP31-262B4 (Figure 3). Concerning the analysis of RNO15, BAC mapping allowed the identification of these breakpoint chromosome regions involved in the der(4;15)(q10;p10) and der(15)del(15)(p11)t(1;15)(q12;q24) (Figure 4). Regarding the first derivative chromosome, it was possible to verify that it involved the entire chromosome 15p arm in a whole-arm translocation with chromosome 4 (Figure 4A). Concerning der(15)del(15)(p11)t(1;15)(q12;q24), BAC clones assisted in the identification of chromosome regions 15p11 and 15q24, which were involved in the formation of the derivative chromosome (Figure 4B–D). Moreover, the BAC analysis allowed the detection of a complex rearrangement in chromosome 15

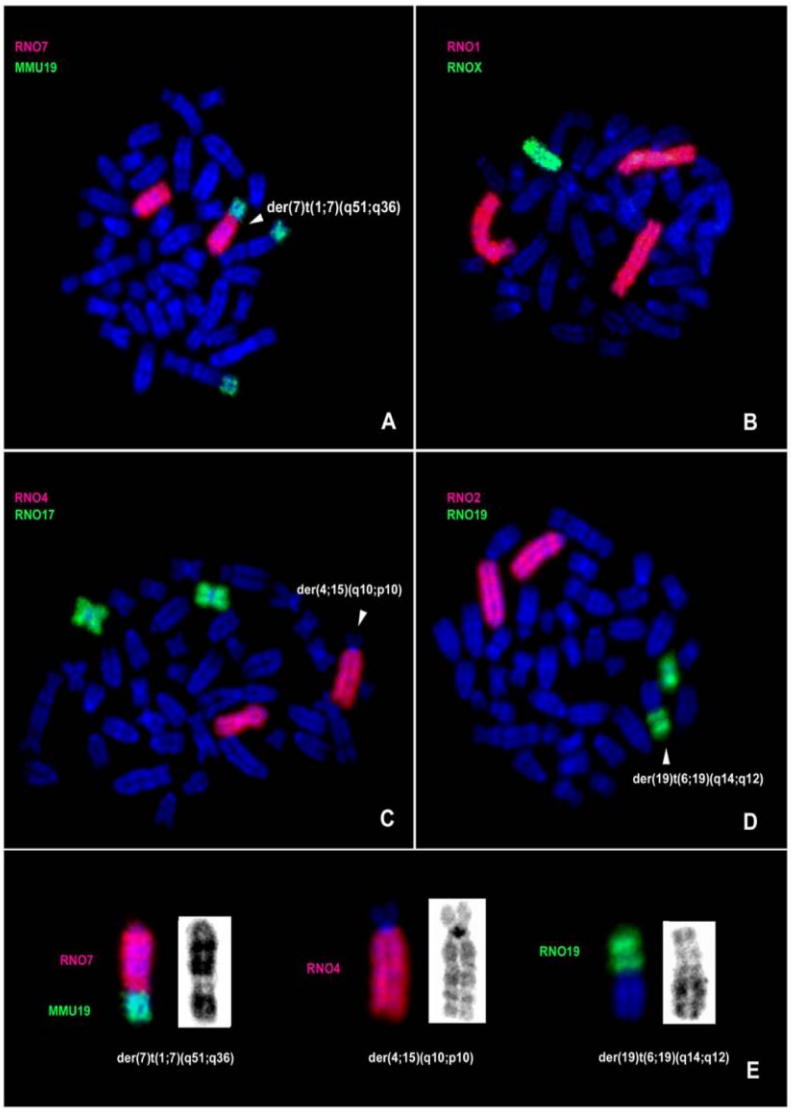


Figure 2. Molecular cytogenetic characterization of HH-16 cl.2/1 clonal chromosome rearrangements. Representative images of *in situ* hybridization with RNO and MMU paint probes onto HH-16 cl.2/1 metaphases (A–D), highlighting the derivative chromosomes. Derivative chromosomes are shown in detail (FISH and GTD) (E).
doi:10.1371/journal.pone.0029923.g002

that was not detectable using chromosome painting alone. After physical mapping of all of the BAC clones, two of them (CH230-117H20 and CH230-9A5) were found to have assumed different cytogenetic positions than expected (Figure 4E). The type of structural rearrangement that would most likely explain these results is a pericentric inversion. However, the remaining BAC clones used in this chromosome analysis were shown to assume the expected locations, meaning that the region between CH230-117H20 and CH230-9A5 maintained its expected order, suggesting a more complex rearrangement. We suggest the occurrence of a second pericentric inversion event involving two other breakpoints. Nevertheless, we cannot discard other possible events leading to the observed derivative chromosome. An interesting characteristic was that both RNO15 homologs, der(4;15)(q10;p10) and der(15)del(15)(p11)t(1;15)(q12;q24), present this configuration.

Integration of all of the FISH data allowed the construction of an HH-16 cl.2/1 composite karyotype based on the analysis of 64 cells:

30~42,X,-X, +1 t(3;11)(p12;p12),der(4;15)(q10;p10),der(7)t(1;7)(q51;q36),del(13)(p13), der(15)del(15)(p11)t(1;15)(q12;q24,der(15)-inv(15)(p14~p16q23~q25)inv(15)(p12~p14q22~q23)×2, −18, der(18)t(1;18)(q11;q12.3), der(18;19)t(18;19)(p10,q10)t(6;19)(q14;q12), der(19)t(6;19)(q14;q12), der(19)t(4;19)(q31;p11)t(6;19)(q14;q12)[cp64]

Reconstruction of HH-16 cl.2/1 cell line clonal evolution

The chromosomal structural abnormalities t(3;11)(p12;p12), del(13)(p13), and der(15)inv(15)(p14~p16q23~q25)(p12~p14q22~q23) ×2 as well as the numeric change –X were observed in all of the cells analyzed (64 cells), suggesting a monoclonal origin of the tumor cell line. The other most frequent chromosomal abnormalities were der(19)t(6;19)(q14;q12), found in 45 cells;

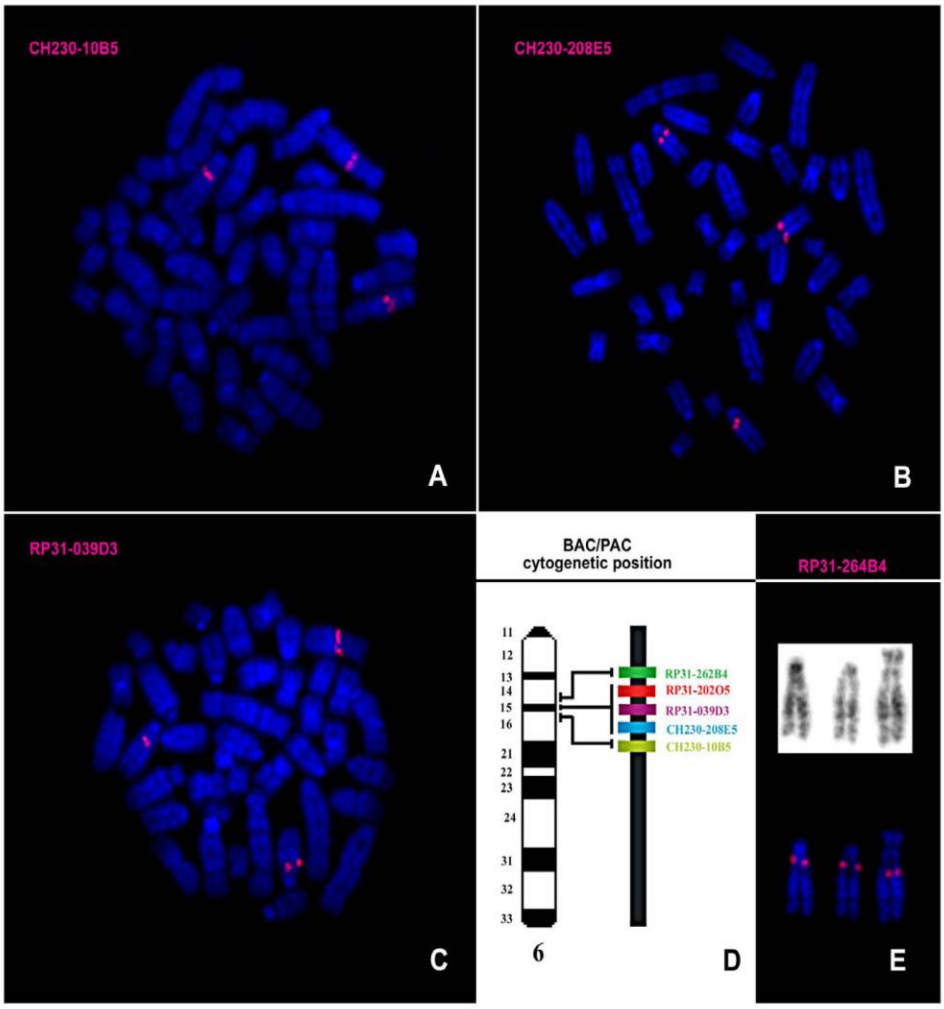


Figure 3. Molecular characterization of the rearrangements involving RNO6 using BAC/PAC clones. Representative images of *in situ* hybridization with BAC/PAC clones onto HH-16 cl.2/1 metaphases (A–C). Chromosome map of the region from bands 6q14 to 6q16, showing the relative positions of the clones used in this study (not to scale) (D). GTD and RP31-262B4 hybridization on the two normal RNO6 and one derivative chromosome of a rearranged cell (E).
doi:10.1371/journal.pone.0029923.g003

der(4;15)(q10;p10), found in 23 cells; and der(7)t(1;7)(q51;q36) observed in 20 of the 64 cells analyzed. This analysis permitted the identification of different cell subclones (Table 1), and comparison of these subclones allowed inferring ancestral rearrangements as well as a tentative reconstruction of the clonal evolution that occurred during tumor progression. The rearrangements present in all cells were considered to be part of the ancestral clone (as shown in Figure 5), from which several branches diverged during tumor progression (karyotype formulas presented in Table 1).

In silico analysis of breast cancer-related genes present in breakpoint regions

All of the identified breakpoint regions resulting from clonal chromosome rearrangements in the HH-16 cl.2/1 cell line are summarized in Figure S2. An *in silico* analysis using data from the Rat Genome Database (http://rgd.mcw.edu; assembly RGSC 3.4) and Ensembl (http://www.ensembl.org/; assembly RGSC 3.4) permitted screening of the breakpoint regions of the ancestral structural rearrangements for the presence of breast cancer-related genes (summarized in Table S1). With the exception of breakpoint bands 11p12, 15q22, 15q23, 15q24 and 15q25, all of the other breakpoints contain genes in the rat genome with human homologs that have been associated with breast cancer in humans.

Mycn and *Erbb2* analysis

Gene amplification. Unlike most gene amplification studies using FISH, the present analysis was performed in metaphase chromosomes instead of interphase nuclei. This approach was advantageous, as it allowed a clear view of aneuploidies and chromosome rearrangements involving regions harboring the studied genes to be obtained. RNO6 painting and the rat PAC clone RP31- 202O5 were used to access the amplification status of *Mycn*. RP31-202O5 was earlier confirmed to contain *Mycn* gene and mapped to RNO 6q15.3-16 [23] and was also used in this work for the accurate identification of HH-16 cl.2/1 breakpoint regions. During HH-16 cl.2/1 cytogenetic characterization, it was possible to verify that this gene was present in three copies

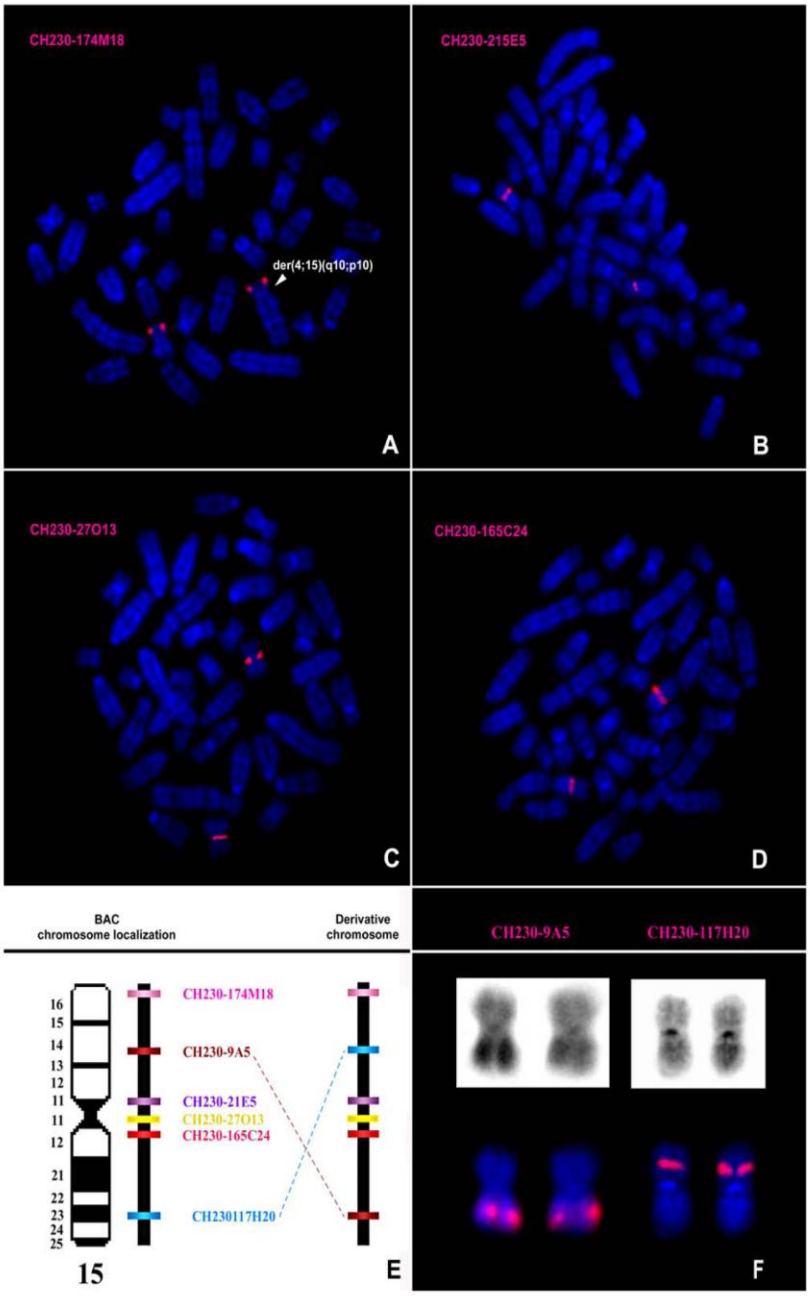


Figure 4. Molecular characterization of the rearrangements involving RNO15 using BAC clones. Representative images of *in situ* hybridization with the BAC clones onto HH-16 cl.2/1 metaphases (A–D). Chromosome map of RNO15 showing the relative positions of the clones used in this study, and the respective clone positions in the rearranged chromosome (not to scale) (E). GTD and CH230-9A5 and CH230-117H20 hybridization on the derivative chromosomes of a rearranged cell (F).
doi:10.1371/journal.pone.0029923.g004

distributed among two intact RNO6 chromosomes and in the derivative chromosomes der(19)t(6;19)(q14;q12), der(18;19)t(18;19)(p10,q10)t(6;19)(q14;q12),der(19)t(6;19)(q14;q12) and der(19)t(4;19)(q31;p11)t(6;19)(q14;q12), indicating that a partial trisomy of RNO6 was involved in a translocation. The derivative chromosomes were not considered in the estimation of RNO6 for the *Mycn*/RNO6 calculation. As can be seen in Table 2, the most representative ratio was 1.5 (84.6%), corresponding to three *Mycn* signals distributed among two normal RNO6 chromosomes and one derivative chromosome (Figures 6A and 6B). *Mycn* was not considered to be amplified in this cell line, while a *Mycn* gain was considered to have occurred. The derivative chromosome der(19)t(6;19)(q14;q12) presenting *Mycn* was found in the majority of cells analyzed and, thus, was considered to represent an ancestral rearrangement (Figure 5). This finding raised the question of its importance in tumor initiation and progression, as this extra copy of

Table 1. Karyotypic formulas of the subclones (A to H) presently found in HH-16 cl.2/1 cell line.

Subclone	Karyotypic formulas
A	38~42,X,-X,t(3;11)(p12;p12),der(7)t(1;7)(q51;q36),del(13)(p13),der(15)inv(15)(p14~p16q23~q25)inv(15)(p12~p14q22~q23)×2, der(19)t(4;19)(q31;p11)t(6;19)(q14;q12) [6]
B	35~42,X,-X,t(3;11)(p12;p12),der(7)t(1;7)(q51;q36),del(13)(p13),der(15)inv(15)(p14~p16q23~q25)inv(15)(p12~p14q22~q23)×2, der(19)t(6;19)(q14;q12) [14]
C	30~41, X,-X,t(3;11)(p12;p12),der(4;15)(q10;p10),del(13)(p13),der(15)del(15)(p11)t(1;15)(q12;q24), der(15)inv(15)(p14~p16q23~q25)inv(15)(p12~p14q22~q23)×2,der(19)t(6;19)(q14;q12) [9]
D	30~41,X,-X,t(3;11)(p12;p12),der(4;15)(q10;p10),del(13)(p13),der(15)inv(15)(p14~p16q23~q25)inv(15)(p12~p14q22~q23)×2, der(19)t(6;19)(q14;q12) [14]
E	30~42,X,-X,t(3;11)(p12;p12),del(13)(p13),der(15)inv(15)(p14~p16q23~q25)inv(15)(p12~p14q22~q23)×2, der(19)t(6;19)(q14;q12) [11]
F	39~41,X,-X,t(3;11)(p12;p12),del(13)(p13), der(15)inv(15)(p14~p16q23~q25)inv(15)(p12~p14q22~q23)×2, der(18)t(1;18)(q11;q12.3),der(19)t(6;19)(q14;q12) [3]
G	38~42,X,-X,t(3;11)(p12;p12),del(13)(p13),der(15)inv(15)(p14~p16q23~q25)inv(15)(p12~p14q22~q23)×2, der(18;19)t(18;19)(p10,q10)t(6;19)(q14;q12),−18 [4]
H	42,X,-X,t(3;11)(p12;p12),del(13)(p13),der(15)inv(15)(p14~p16q23~q25)inv(15)(p12~p14q22~q23)×2, der(19)t(6;19)(q14;q12),+1 [3]

The karyotype formulas correspond to the different subclones identified in HH-16 cl.2/1 cell line allowing the identification of ancestral chromosome rearrangements and to deduce the hypothetic clonal evolution shown in Figure 5.

doi:10.1371/journal.pone.0029923.t001

Mycn was present in the ancestral clone. Regarding the HH-16.cl.4 cell line, all of the cells analyzed were characterized by a ratio of 1, presenting four *Mycn* signals distributed among four RNO6 chromosomes (data not shown). *Mycn* was also not amplified in the HH-16.cl.4 cell line.

To investigate *Erbb2* gene amplification, the RNO10 paint probe and CH230-162I16 rat BAC clone were used for FISH experiments with HH-16 cl.2/1 and HH-16.cl.4 chromosomes. This BAC clone was selected from a total of three clones acquired that were validated by PCR isolation followed by sequencing, with this clone being the only found to contain the *Erbb2* gene (Figure S1). CH230-162I16 was mapped by FISH for the first time in this study, and it was assigned to RNO 10q32.1, which is the cytogenetic position of rat *Erbb2* determined by Koelsch in 1998 [24]. According to the criteria used, no *Erbb2* amplification was detected in the HH-16 cl.2/1 cell line. The analysis revealed an *Erbb2*/RNO10 ratio of 1 in all analyzed cells (Table 2), corresponding to the presence of two *Erbb2* signals distributed among two RNO10 chromosomes (data not shown). Among the HH-16.cl.4 cells analyzed (Table 2), the most representative *Erbb2*/RNO10 ratio was 1.7 (88.6% of cells), which is near the cut-off value. In these cells, five *Erbb2* signals can be seen to be distributed among one intact RNO10 (one *Erbb2* signal) and two derivative RNO10 chromosomes with a duplication involving *Erbb2* loci (two *Erbb2* signals) (Figures 6C and 6D). As five *Erbb2* signals were observed, an *Erbb2* gain was considered to have occurred. An RNO10 polysomy was verified.

RNA expression analysis. The levels of expression of the *Mycn* and *Erbb2* genes in HH-16 cl.2/1 and HH-16.cl.4 were determined by one-step real-time RT quantitative PCR (RT-qPCR), complemented and validated by RNA FISH (for *Erbb2*). Figure 7 shows the relative RT-qPCR quantification in terms of the fold change in *Erbb2* and *Mycn* RNA expression for both cell lines, which was normalized using multiple reference genes and is given relative to a calibrator (control RNO sample). All of the expression values presented in the graph were considered statistically significant following analysis using Student's *t*-test with a p value$<$0.05. Regarding *Mycn*, despite the statistical significance of the results, the fold changes in gene expression were low. For the HH-16.cl.4 cell line, a gain of 1.6 was verified, while for HH-16 cl.2/1, the expression value was below control sample expression (0.7 SD$\pm$0.06), corresponding to 1.4 times less expression than the control sample (Table 2). Only the results for *Erbb2* showed significant expression level changes, especially in HH-16.cl.4. The increase in *Erbb2* expression in HH-16 cl.2/1 was 2.6 fold (close to the cut-off value), and in HH-16.cl.4, *Erbb2* was expressed at a level 10.7 times higher than in the control sample (Table 2). HH-16 cl.2/1 *Erbb2* expression was approximately 4 times lower than in the sister cell line HH-16.cl.4, with significant expression only being found in the HH-16.cl.4 rat mammary cell line.

Evaluation of *Erbb2* expression was also performed using RNA fluorescent *in situ* hybridization in HH-16 cl.2/1 and HH-16.cl.4, validating the RT-qPCR analysis. This procedure allowed the visualization of *Erbb2* mRNA in individual cells of both cell lines (Figures 8A and 8B). The number of signals *per* cell was counted in 20 slide fields for each rat cell line, resulting in a total of 483 cells being analyzed for HH-16 cl.2/1 and 321 cells being analyzed for HH-16.cl.4. The results are displayed as the percentages of cells with total *Erbb2* signals falling between four numerical intervals: [1–5], [6–10], [11–30] and [+30]. Figure 8C shows that 84.5% of the HH-16 cl.2/1 cells present 1–5 *Erbb2* signals, and 60.4% of HH-16.cl.4 cells present 11–30 *Erbb2* signals, with these intervals being the most representative for each cell line. The mean number of signals *per* cell was 3.3 for HH-16 cl.2/1 and 15.2 for HH-16.cl.4 (Figure 8D and Table 2). These results show that there was higher expression of *Erbb2* in HH-16.cl.4 than in HH-16.cl.2/1, with 4.6 times higher expression being observed in the former cell line than in that latter, with is in accordance with the RT-qPCR data. The advantage of this methodology is the use of single cell analysis, which showed a wide range of expression in the cells of both cell lines. In addition to the expression analysis, RNA FISH permitted us to examine the sub-cellular localization of *Erbb2* mRNA. In both rat cell lines, *Erbb2* displayed cytoplasmic localization.

Influence of 5-Aza-2′-Deoxicitidine global demethylation on *Mycn* and *Erbb2* RNA expression

Both cell lines were treated with 5-Aza-2′-Deoxicitidine for a period of 72 h, after which RNA was extracted and used to evaluate *Erbb2* and *Mycn* expression levels by means of RT-qPCR. These experiments were normalized with multiple reference genes

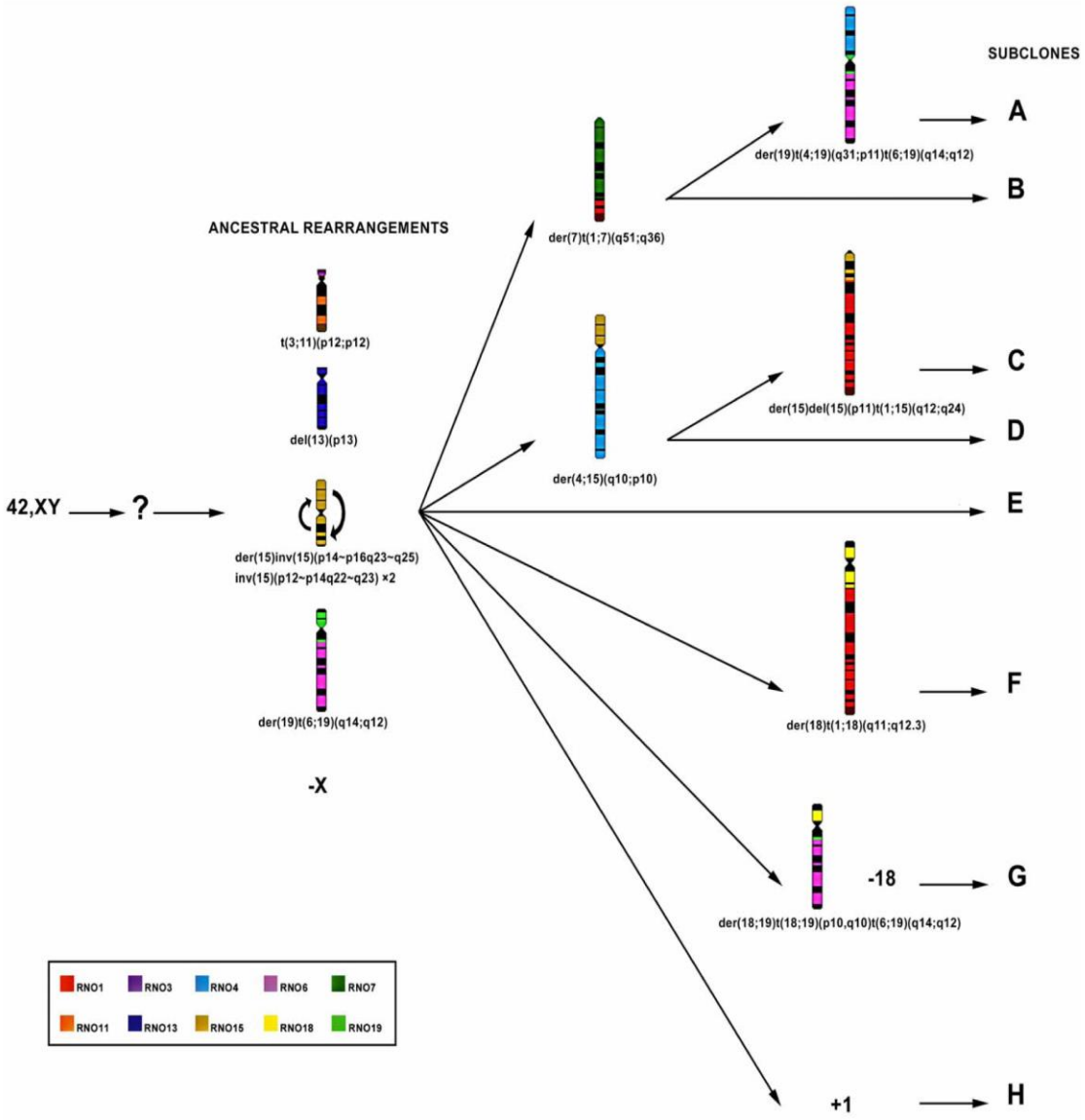


Figure 5. Chromosome reconstruction of the clonal evolution in HH-16 cl.2/1 tumor cell line. Diagram showing the hypothetic clonal evolution of HH-16 cl.2/1 chromosomes. In the diagram are shown numerical and structural clonal rearrangements. Ideograms represent all structural clonal rearrangements. Each rat chromosome is represented by a different color according to the legend. Subclones A–H are presently found in the cell line (respective karyotype formulas are shown in Table 1).
doi:10.1371/journal.pone.0029923.g005

(beta-actin and GAPDH) using RNA from cell lines that were not treated with 5-Aza-2′-Deoxicitidine as a control for calculating relative expression. No significant changes in *Mycn* expression were registered for either cell line. Statistically significant results based on Student's *t*-test (p value$<$0.05) were only obtained for *Erbb2* expression in HH-16.cl.4. *Erbb2* expression decreased after treatment with 5-Aza-2′-Deoxicitidine at a concentration of 3 μM in HH-16.cl.4 cells (Figure 9). This expression decrease, although significant ($p<0.05$), was not high when compared with the untreated cells. Moreover, *Erbb2* expression continued to decrease, even after the removal of the drug. These results show that global genomic demethylation only affects the expression of the *Erbb2* gene in the HH-16.cl.4 cell line, whereas it appears to have no effect on *Mycn* expression in either cell line.

Discussion

A major opportunity to increase our knowledge regarding the biology of breast cancer is associated with the availability of experimental model systems that recapitulate the many forms of this disease. Recent studies have described the genetic characterization of breast cancer cell lines, showing their value in the investigation of the role of genomic alterations in cancer progression and as a resource for the discovery of new breast cancer genes [25,26]. Rat cancer models, such as DMBA-induced rat tumors, have been found to be useful models for studying hormone-dependent breast cancer [1].

Here we present, for the first time, the genetic/cytogenetic characterization of two DMBA-induced rat mammary tumor cell

Table 2. *Mycn* and *Erbb2* amplification and expression results for HH-16 cl.2/1 and HH-16.cl.4.

	Mycn		*Erbb2*		
Cell line	***Mycn*/RNO6**	**Expression Fold Change (±SD)**	***Erbb2*/RNO10**	**Expression Fold Change (±SD)**	**RNA in situ signal Mean (±SD)**
HH-16 cl.2/1	1.5 (84.6%)	−1.4 (±0.06)	1 (100%)	+2.6 (±0,1)	3.3 (±0.9)
	1.0 (15.4%)				
HH-16.cl.4	1.0 (100%)	+1.6 (±0.2)	1.7 (88.6%)	+10.7(±1,2)	15.2 (±3.6)
			1.5 (5.6%)		
			1.3 (2.9%)		
			1.0 (2.9%)		

Mycn and *Erbb2* amplification results were calculated as the *Mycn*/RNO6 and *Erbb2*/RNO 10 ratios, respectively (values between brackets represent the percentage of analyzed cells showing that result). Expression levels were accessed by RT-qPCR and RNA FISH (values between brackets represent the standard deviation).
doi:10.1371/journal.pone.0029923.t002

lines, HH-16 cl.2/1 and HH-16.cl.4, which share the same genetic origin. These cell lines exhibit very distinct cytogenetic characteristics, beginning with different levels of ploidy. While HH-16.cl.4 cell line presents a nearly tetraploid karyotype, showing a wide range of cells with different chromosome numbers and levels of ploidy (Figure 1D), a nearly diploid karyotype with low levels of polyploidy can be found in HH-16 cl.2/1 (Figure 1C). This finding might be indicative of a higher order of complexity and chromosomal instability (CIN) of HH-16.cl.4, which is described as the presence of ploidy changes as well as high levels of aneuploidy [27]; these phenomena have been shown to have a direct causal role in tumorigenesis [28]. Additionally, heterogene-

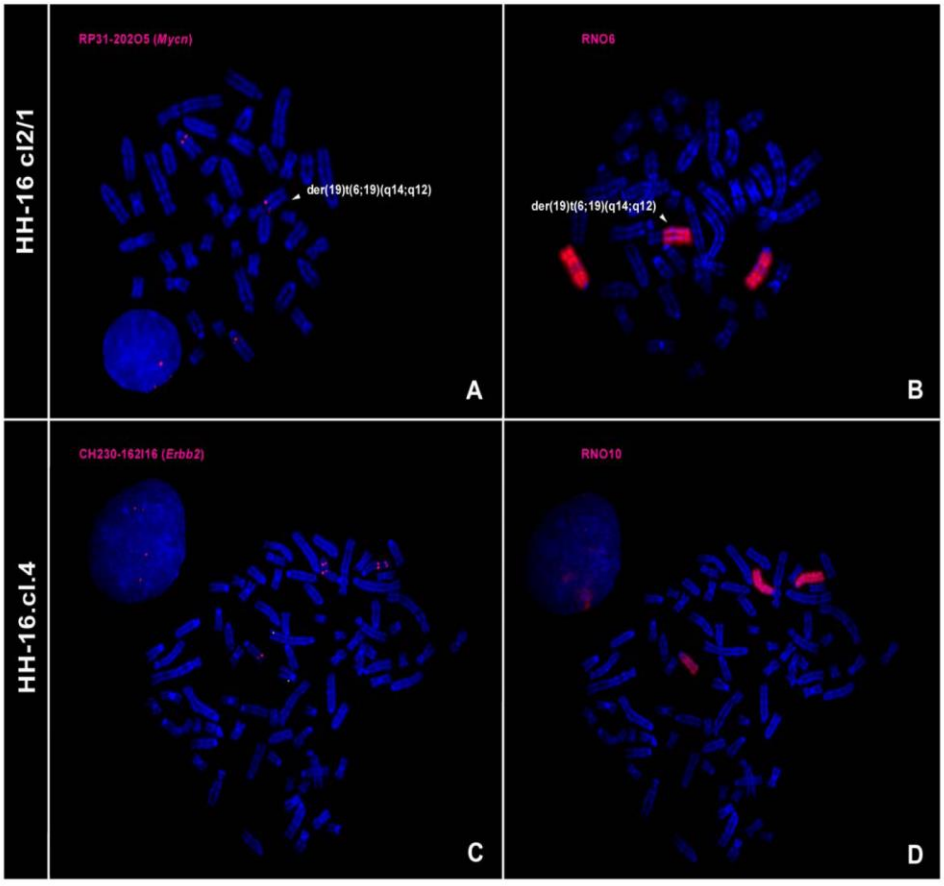


Figure 6. FISH results for *Mycn* and *Erbb2* amplification analysis in HH-16 cl.2/1 and HH-16.cl.4cell lines. Images show *Mycn* hybridizes in three chromosomes (A) two RNO6 and a derivative chromosome, der(19)t(6;19)(q14;q12) (B). Two *Erbb2* signals are present in chromosome 10 (C) identified with RNO10 paint probe hybridization (D).
doi:10.1371/journal.pone.0029923.g006

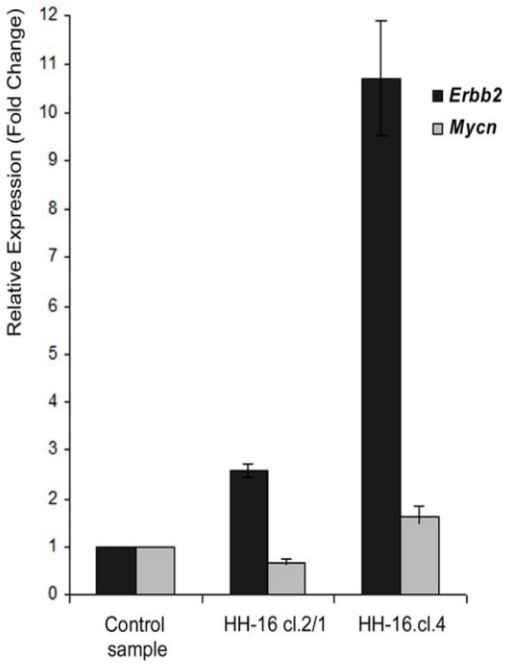


Figure 7. Relative expression of *Erbb2* and *Mycn* in the HH-16 cl.2/1 and HH-16.cl.4 cell lines. Expression results were obtained by reverse transcription quantitative real time PCR, normalized with the reference genes beta-actin and GAPDH and compared with a control sample. Data is presented as mean corresponding to fold change relative to the control sample ($p<0.05$). Error bars represent ±SD.
doi:10.1371/journal.pone.0029923.g007

ity reflects the existence of different tumor clones as well as a large number of apparently random chromosome changes, or so-called "cytogenetic noise". For this reason, performing genome-wide cytogenetic characterization did not appear to be promising, and cytogenetic analysis was limited to the identification of relevant chromosome rearrangements associated with specific gene expression changes.

For the cytogenetic characterization of the HH-16 cl.2/1 cell line a multi approach was used which included G-banding, chromosome painting using rat and mouse probes and BAC/PAC clones hybridization. Clonal chromosome rearrangements were characterized (Figure 2) and specific breakpoint regions were identified (Figure 3 and 4). Few studies on the cytogenetic characterization of rat cell lines have been performed, particularly using rat or mouse paint probes. However, there have been some reports addressing rat tumor cell lines indicating RNO1 [3,29,30], RNO3 [29], RNO6 [31] and RNO15 [32,33] as recurrent and/or relevant chromosomes related to the tumorigenesis/tumor progression in mammary fibrosarcomas, endometrial adenocarcinomas and lung cancer. Also in our study rearrangements in those chromosomes have been identified, as it is the case of the complex rearrangement involving RNO15 (Figure 4), only detected using the combination of varied cytogenetic tools. For this derivative chromosome we propose the occurrence of a double inversion as previously found in Acute Myeloid Leukemia karyotypes (e.g., [34,35,36]), and its presence in both RNO15 homologues can be explained by the loss of the normal chromosome, followed by the duplication of the abnormal homolog [37,38,39]. An interesting finding was the loss of an entire X chromosome which was present in all subclones identified. X chromosome loss has been described in numerous human cancer cases corresponding to the inactive X copy (e.g., [40,41]) identified by a detectable Barr body (classic characteristic of X chromosome inactivation) [42,43]. During our analysis of HH-16 cl.2/1 cells interphase nuclei, no Barr bodies were found in the X chromosome territory identified with the rat X paint probe (data not shown). This finding provides evidence that the X chromosome present in this cell line is the active X chromosome.

Assembly of the obtained data allowed us to deduce the clonal evolution of this tumor, which is illustrated in Figure 5. This diagram allows easy visualization of the ancestral and recent rearrangements, as well as providing an overview of the microevolutionary processes that have occurred in the progression of this tumor cell line. Analyses of karyotype clonal evolution have been performed previously in rats [32,44], showing its relevance in the investigation of tumor progression. Moreover, the existence of ancestral structural chromosome abnormalities suggests a relevant role for these rearrangements in providing a selective advantage to this tumor cell line. An *in silico* analysis was performed focused on the breakpoint regions of the ancestral structural chromosome rearrangements and demonstrated that almost all of the breakpoint regions contain genes in the rat genome for which the human homolog has been associated with breast cancer (Table S1). This finding is relevant once translocations can lead to altered gene activity either through the formation of a chimeric gene product with cell transforming properties, or by juxtaposition of an oncogene with a foreign activator element [45].

In the cytogenetic characterization of HH-16 cl.2/1, the *Mycn* extra copy number was of particular note, especially because this characteristic was present in all of the cells analyzed and was considered to represent an ancestral condition. This observation raised the possibility of relevance of the *Mycn* gene in mammary tumor initiation and progression for both cell lines (once they are related). *MYCN* is part of a large family of oncogenes found to be amplified in human neuroblastomas and is correlated with aggressiveness and a negative prognosis in this type of pediatric cancer (reviewed by [46]). *Mycn* amplification has also been observed in rat tumors, specifically in uterine endometrial carcinomas [31,47], however, the available literature does not include any investigation of *MYCN* amplification status in breast cancer. Overall, *Mycn* amplification was not detected in the HH-16 cl.2/1 or in HH-16.cl.4 cell lines, but an *Mycn* gain was found in HH-16 cl.2/1 (Figure 6 and Table 2). Additional copies of *MYCN* equal or less than 4-fold detected by FISH were considered as an *MYCN* gain, following a study on neuroblastoma [48].

The other gene analyzed in the present study was *Erbb2*. In humans, *ERBB2* gene amplification constitutes one of the most important genetic alterations associated with human breast cancer and was first correlated with poor patient prognosis by Slamon and colleagues [9]. Hence, no *Erbb2* amplification was found in the HH-16 cl.2/1 cell line, while for HH-16.cl.4 a low level of amplification was detected (Table 2). Chromosome painting data showed that *Erbb2* gain resulted from a chromosome alteration involving *Erbb2* gene locus resulting in its duplication (Figure 6C). Amplified DNA can be observed in various forms, including double minutes or amplified regions on a chromosome or distributed across the genome [49]. This gene gain may act as a precursor to further *Erbb2* amplification, or it may represent an alternative pathway for activating the oncogenic potential of this gene.

Generally, gene amplification has been associated with overexpression of the amplified gene(s) [49], although this correlation is not absolute. Both genes expression (*Mycn* and *Erbb2*) was accessed by RT-qPCR in the present work. *Mycn* RNA

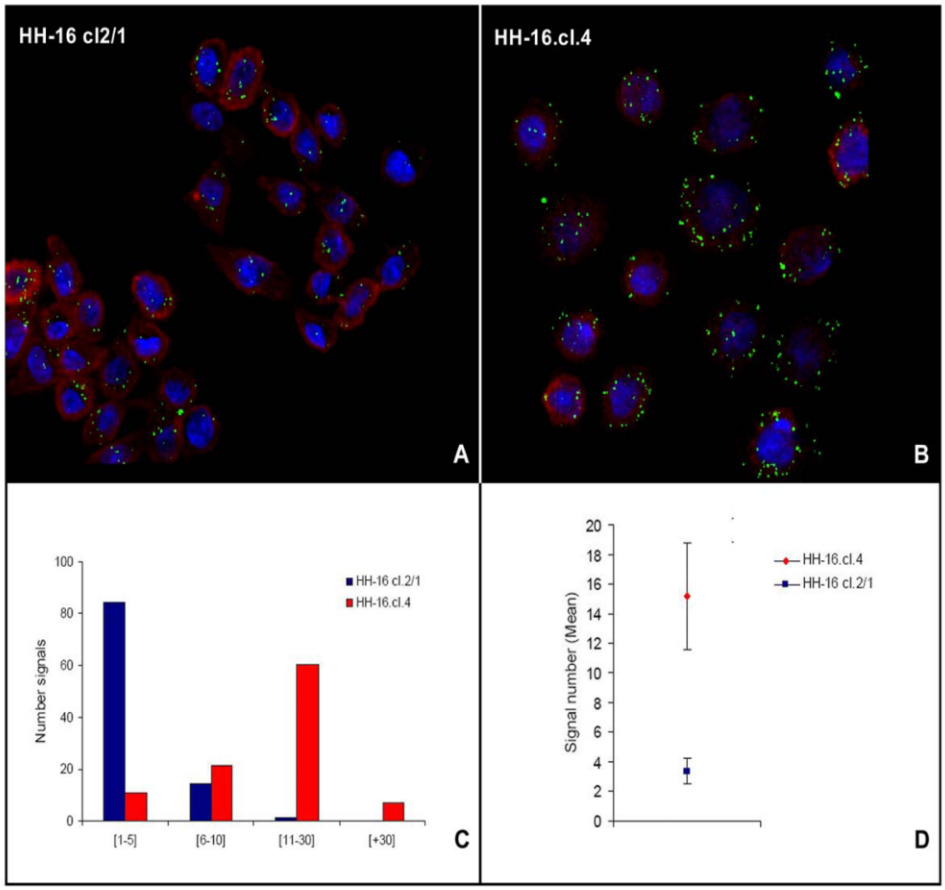


Figure 8. Expression analysis of *Erbb2* by RNA fluorescent *in situ* hybridization. RNA fluorescent *in situ* hybridization of *Erbb2* mRNA (green) and ribosomal 18S (red) used as reference, in HH-16 cl.2/1 (A) and HH-16.cl.4 (B) cell lines. The number of signals distributed by 4 intervals (C) and the mean number of signals for each cell line (error bars represents ±SD) (D) clearly showed differences in *Erbb2* expression between the two cell lines, being considerably higher in the HH-16.cl.4 cell line.
doi:10.1371/journal.pone.0029923.g008

expression status showed no evidence for considerable expression changes which is in accord with the absence of gene amplification detected (Figure 7 and Table 2). These results also show that the *Mycn* gain corresponding to the three *loci* presented in the HH-16 cl.2/1 cell line was not reflected in an RNA expression change. With respect to *ERBB2*, the most frequently used method to determine its expression in breast cancer is immunohistochemistry (protein quantification) [50]. In human invasive duct carcinomas of the breast, erbB-2 protein overexpression is particularly frequent, and in most cases, this overexpression is caused by *ERBB2* gene amplification and associated with an unfavorable prognosis [9,51]. Trastuzumab (Herceptin) is a humanized monoclonal antibody directed against the extracellular domain of the erbB-2 protein [52] that have been found to be effective when in presence of high levels of this protein [53,54]. The *ERBB2* gene and erbB-2 protein status (gene amplification/protein overexpression) are considered useful markers for predicting the response to a specific cancer therapy, and analysis of these markers is mandatory for the identification of breast cancer patients that are amenable to trastuzumab treatment. In addition to immunohistochemistry, other methods have proven reliable in determining *ERBB2* expression status, such as real-time reverse transcription quantitative PCR [55,56]. In the present study, a 3-fold increase in expression was considered to represent a significant expression change [57]. Relevant RNA expression changes for *Erbb2* were detected only for HH-16.cl.4 (10.7-fold increase) (Figure 7 and Table 2). This result correlates with the *Erbb2* gene gain, suggesting that the amplification, while low, may have played a role in the overexpression of *Erbb2* RNA in this cell line, although it may not be the only mechanism involved. The involvement of human chromosome 17 (harbors *ERBB2*) polysomy in erbB-2 protein expression has been discussed with some controversy [58]; however, some authors point to it as the cause of *ERBB2* overexpression [59,60]. This cell line presents different levels of ploidy, and most of the cells analyzed present three copies of RNO10 (Figure 6D), suggesting the possible correlation of this chromosome copy number with the observed *Erbb2* expression levels. Another possible explanation is transcriptional regulation, which could have promoted the accumulation of *Erbb2* mRNA in the absence of high levels of amplification. Moreover, both older and more recent studies show that *ERBB2* RNA overexpression does not always correspond to erbB-2 protein overexpression, suggesting the existence of post-transcriptional regulation of *ERBB2* [61,62], which shows the relevance of using RT-qPCR in routine assessment of *ERBB2* overexpression in human breast cancer in the clinical laboratory setting.

RNA FISH was used to measure *Erbb2* expression, complementing and validating the results of the RT-qPCR analysis. RNA

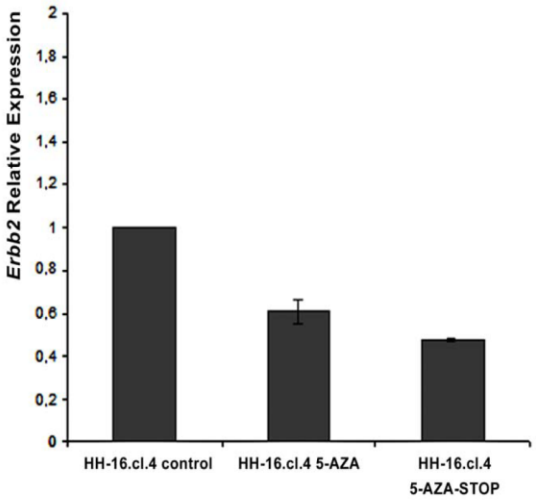


Figure 9. Relative expression analysis of *Erbb2* in HH-16.cl.4 cells after treatment with 5-Aza-2-Deoxicitidine. Relative expression analysis of *Erbb2* in HH-16.cl.4 cells treated with 5-Aza-2′-Deoxicitidine (HH-16.cl.4 5-AZA) and in HH-16.cl.4 cells after stopping the treatment with 5-Aza-2-Deoxicitidine (HH-16.cl.4 5-AZA-STOP). HH-16.cl.4 cells that were not treated with 5-Aza-2′-Deoxicitidine served as control (HH-16.cl.4 control). Data is presented as mean corresponding to fold change relative to control sample ($p<0.05$). Error bars represent ±SD.
doi:10.1371/journal.pone.0029923.g009

fluorescent *in situ* hybridization is advantageous because it allows analysis of spatial gene expression patterns at a single-cell resolution [63,64,65]. This approach allowed clear visualization and semi-quantification of mRNA molecules in the cytoplasm, allowing quantification of the expression of *Erbb2* in both cell lines. The RNA FISH data strongly supported the RT-qPCR expression results, showing higher expression of *Erbb2* in HH-16.cl.4 (4.6 times greater) compared with the sister cell line HH-16 cl.2/1 (Figure 8 and Table 2), demonstrating to be an excellent technology when applied either alone or together with other technique.

Interestingly, the expression of *Erbb2* in the HH-16.cl.4 rat cell line appears to be affected by global genome demethylation. In the present study, HH-16 cl.2/1 and HH-16.cl.4 cells were treated with 5-Aza-2′-Deoxicitidine, promoting global genome demethylation. Statistically significant results were obtained for the *Erbb2* gene in the HH-16.cl.4 cell line, although the variation was not especially large (Figure 9). It has been demonstrated that *ERBB2* gene is overexpressed and unmethylated (in its promoter) in tumors and tumor cell lines, such as ovarian tumoral tissues and MCF-7 cell line [66,67]. A similar study to ours was performed in a rat chondrosarcoma cell line, in which an increase in *Erbb2* expression was found after global genome demethylation [68]. Intriguingly, our data shows a decrease in *Erbb2* expression after 5-Aza-2′-Deoxicitidine treatment. While in the rat chondrosarcoma cell line, *Erbb2* promoter unmethylation seems to be the main cause for *Erbb2* overexpression, our data suggests a different pivotal epigenetic mechanism underlying the expression of this gene. Candidate negative regulators of *Erbb2* might be non-coding RNAs that for instance promote the degradation of transcripts [69]; or even other less understood epigenetic mechanisms such as splicing regulation [70] can explain our results. Our findings emphasize that future studies are mandatory to reveal the exact epigenetic events involved in the regulation of *Erbb2* expression, and that HH-16.cl.4cell line is an excellent tool to complete this task.

The cell lines used in the present work were generated simultaneously from the DMBA-induced rat mammary tumor [5], but despite having the same initial genetic background, fibroblastoid H-16 cl.2/1 cell line apparently reflect mesenchymal cells of the stromal part of the tumor, while the epitheloid HH-16.cl.4 cell line display epithelial origin. The cell lines different lineage, associated with the higher chromosomal instability revealed by HH-16.cl.4 (explaining the *Erbb2* overexpression here observed), suggests different mechanisms involved in tumor progression of both cell lines. In fact, HH-16.cl.4 exhibits a mainly tetraploid number of chromosomes. Tetraploidy can arise through a number of mechanisms, including cell fusion, mitotic slippage and cytokinesis failure [71]. In addition, tetraploid cells typically contain twice the normal complement of centrosomes that promote aberrant mitotic divisions and chromosome missegregation at a high frequency. Moreover, tetraploidy has been shown to initiate chromosomal instability and has been found to precede the development of CIN and aneuploidy in several cancers (e.g. [72,73]). On the other hand, in the fibroblastoid H-16 cl.2/1 cell line, chromosome structure instability (CSI) seems to be the distinguishing feature, whose mechanisms are now starting to be disclosed [74]. Nevertheless, it seems that CSI can be the result of errors in the DNA damage checkpoints, DNA repair pathways, and/or mitotic segregation errors. However, mutations in proteins that permit cell cycle progression in the presence of double stranded breaks (e.g. p53, BRCA1, BRCA2, ATM and ATR) may also facilitate CSI [75].

In conclusion, molecular cytogenetics, gene expression profiling and examination of the influence of global demethylation on gene expression were used to characterize two rat mammary cell lines, H-16 cl.2/1 and HH-16.cl.4. All the presented results provide a platform for future studies on tumor progression and encourage the use of these cell lines as a model. In particular this study highlights H-16 cl.2/1 and HH-16.cl.4 potential as models for studying *Erbb2* associated mechanisms and as experimental tools to assist in the generation of new biotherapies.

We believe that the development of capable *in vitro* models of human breast cancer is of crucial importance in the study of cancer and, consequently, in the development of new therapeutics. We are confident that his work has contributed to the validation of this cellular model and to its use in future studies.

Supporting Information

Figure S1 Representative images of the *in situ* hybridization of putative *Erbb2* BAC clones onto RNO metaphases. Both CH230-276G18 (A) and CH230-305O21 (B) hybridize in different locations than the *Erbb2* position determined by [24]. Only CH230-162I16 hybridizes at the cytogenetic position of *Erbb2* in *Rattus norvegicus* (10q32.1) (C). PCR amplification of *Erbb2* in the three clones (D). Only for CH230-162I16 BAC clone the expected 350 bp band is observed.
(PDF)

Figure S2 Chromosomal location of the clonal rearrangements breakpoint regions in HH-16 cl.2/1 cell line. Clonal rearrangements breakpoint regions in HH-16 cl.2/1cell line are displayed in the rat ideogram [20]. Each type of rearrangement originated by the breakpoints is identified by a specific color.
(PDF)

Table S1 *In silico* analysis of breast cancer related genes present in the most representative rat breakpoint regions, and its correspondent human homolog.
(PDF)

Acknowledgments

The authors are deeply grateful to Dr. Johannes Wienberg and Dra. Andrea Kofler (Chrombios GmbH, Germany) for providing the paint probes from *Rattus norvegicus* and *Mus musculus*. We thank to Bioportugal-Químico, Farmacêutica, Lda for technical support in the expression assays.

Author Contributions

Conceived and designed the experiments: SL FA RC. Performed the experiments: SL FA RC. Analyzed the data: SL FA RC. Contributed reagents/materials/analysis tools: FA RC. Wrote the paper: SL. Final revision of the paper: FA RC.

References

1. Vargo-Gogola T, Rosen JM (2007) Modelling breast cancer: one size does not fit all. Nat Rev Cancer 7: 659–672.
2. Haag JD, Hsu LC, Newton MA, Gould MN (1996) Allelic imbalance in mammary carcinomas induced by either 7,12-dimethylbenz[*a*]anthracene or ionizing radiation in rats carrying genes conferring differential susceptibilities to mammary carcinogenesis. Mol Carcinog 17: 134–143.
3. Sjöling A, Lindholm H, Samuelson E, Yamasaki Y, Watanabe TK, et al. (2001) Analysis of chromosomal aberrations involving chromosome 1q31→q53 in a DMBA-induced rat fibrosarcoma cell line: amplification and overexpression of Jak2. Cytogenetic Cell Genetic 95: 202–209.
4. Aitman TJ, Critser JK, Cuppen E, Dominiczak A, Fernandez-Suarez XM, et al. (2008) Progress and prospects in rat genetics: a community view. Nat Genet 40(5): 516–522.
5. Steffen M, Scherdin U, Duvigneau C, Hölzel F (1988) Glucocorticoid-induced alterations of morphology and growth of fibrosarcoma cells derived from 7,12-dimethylbenz(a)anthracene rat mammary tumor. Cancer Res 48: 212–7218.
6. Scherdin U, Steffen M, Dietel M, Boecker W, Breindl M, et al. (1990) Elevated expression of v-mos is correlated with altered differentiation of carcinoma cells. Oncogene 5(11): 1619–1627.
7. Steffen M, Scherdin U, Vértes I, Boecker W, Dietel M, et al. (1992) Karyotype instability and altered differentiation of rat sarcoma cells after retroviral infection. Genes Chromosomes Cancer 4(1): 46–57.
8. O'Connell P (2003) Genetic and cytogenetic analyses of breast cancer yield different perspectives of a complex disease. Breast Cancer Res Treat 78: 347–357.
9. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, et al. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. pp 235177–235182.
10. Bièche I, Lerebours F, Tozlu S, Espie M, Marty M, et al. (2004) Molecular profiling of inflammatory breast cancer: identification of a poor-prognosis gene expression signature. Clin Cancer Res 10(20): 6789–6795.
11. Herman JG, Baylin SB (2003) Gene silencing in cancer in association with promoter hypermethylation. N Engl JMed 349: 2042–2054.
12. Mund C, Brueckner B, Lyko F (2006) Reactivation of epigenetically silenced genes by DNA methyltransferase inhibitors: basic concepts and clinical applications. Epigenetics 1: 7–13.
13. Christman JK (2002) 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21: 5483–5495.
14. Stresemann C, Lyko F (2008) Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int J Cancer 123: 8–13.
15. Lapeyre JN, Becker FF (1979) 5-Methylcytosine content of nuclear DNA during chemical hepatocarcinogenesis and in carcinomas which result. Biochem Biophys Res Commun 87: 698–705.
16. Feinberg AP, Gehrke CW, Kuo KC, Ehrlich M (1988) Reduced genomic 5-methylcytosine content in human colonic neoplasia. Cancer Res 48: 1159–1161.
17. Seabright M (1971) A rapid banding technique for human chromosomes. Lancet 2: 971–972, 1971.
18. Chaves R, Adega F, Santos S, Guedes-Pinto H, Heslop-Harrison JS (2002) *In situ* hybridization and chromosome banding in mammalian species. Cytogenet Genome Res 96: 113–116.
19. Wienberg J, Stanyon R, Nash WG, O'Brien PC, Yang F, et al. (1997) Conservation of human vs. feline genome organization revealed by reciprocal chromosome painting. Cytogenet Cell Genet 77(3–4): 211–7.
20. Levan G (1974) Nomenclature for G-bands in rat chromosomes. Hereditas 77: 37–52.
21. ISCN (2009) An international System for Human Cytogenetic Nomenclature, L.G. Shaffer, M.L. Slovak, L.J. Campbell (eds); S. Karger, A. G. Basel 2009.
22. Livak K, Schmittgen T (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods 25: 402–408.
23. Hamta A, Adamovic T, Samuelson E, Helou K, Behboudi A, et al. (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.
24. Koelsch BU (1998) Assignment1 of Erbb2 to rat chromosome band10q32.1 by in situ hybridization. Cytogenet Cell Genet 81: 182.
25. Kao J, Salari K, Bocanegra M, Choi YL, Girard L, et al. (2009) Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One 4(7): e6146.
26. Kadota M, Yang HH, Gomez B, Sato M, Clifford RJ, et al. (2010) Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines. PLoS One 5(2): e9201.
27. Bayani J, Selvarajah S, Maire G, Vukovic B, Al-Romaih K, et al. (2006) Genomic mechanisms and measurement of structural and numerical instability in cancer cells. Semin Cancer Biol 17(1): 5–18.
28. Schvartzman JM, Sotillo R, Benezra R (2010) Mitotic chromosomal instability and cancer: mouse modelling of the human disease. Nat Rev Cancer 10: 102–115.
29. Popescu NC, Greiner JW (1999) Recurrent alterations of the short arm of chromosome 3 define a tumor supressor region in rat mammary tumor cells. Carcinogenesis 20(10): 2033–2036.
30. Walentinsson A, Sjöling A, Helou K, Klinga-Levan K, Levan G (2000) Genomewide assessment of genetic alterations in DMBA-induced rat sarcomas: cytogenetic, CGH, and allelotype analyses reveal recurrent DNA copy number changes in rat chromosomes 1, 2, 4, and 7. Genes Chromosomes Cancer 28(2): 184–95.
31. Adamovic T, Trossö F, Roshani L, Andersson L, Petersen G, et al. (2005) Oncogene Amplification in the Proximal Part of Chromosome 6 in Rat Endometrial Adenocarcinoma as Revealed by Combined BAC/PAC FISH, chromosome Painting, Zoo-FISH, and Allelotyping. Genes Chromosomes Cancer 44: 139–153.
32. Dano L, Guilly MN, Dutrillaux B, Chevillard S (2001) Clonal evolution of a radon-induced rat lung tumor. Cancer Genet Cytogenet 125: 52–58.
33. Hamta A, Talebbeigy F (2010) Reccurent regional allelic imbalance in chromosome 15 in rat endometrial adenocarcinomas. Yakhteh Medical Journal 12: 59–72.
34. Lee C (1999) Double inversion (3)(q21q26) and monossomy 7 in a case of Acute Myeloid Leukemia. Cancer Genet Cytogenet 111: 99–102.
35. Lahortiga I, Vázquez I, Agirre X, Larrayoz MJ, Vizmanos JL, et al. (2004) Molecular heterogeneity in AML/MDS patients with 3q21q26 rearrangements. Genes Chromosomes Cancer 40: 179–189.
36. Toydemir R, Rowe L, Hibbard M, Salama M, Shetty S (2010) Cytogenetic and molecular characterization of double inversion 3 associated with a cryptic BCR-ABL1 rearrangement and additional genetic changes. Cancer Genet Cytogenet 201: 81–87.
37. Walter TA, Morgan R, Ondreyco S, Sandberg AA (1990) Apparent duplication of inv(3) in myelodysplastic syndrome in acute leukemia. Am J Hematol 33: 210–215.
38. Ohyashiki K, Murakami T, Ohyashiki JH, Kodama A, Sakai N, et al. (1992) Double 2q-anomaly in myelodysplastic syndrome. Cancer Genet Cytogenet 58: 174–176.
39. Hirata J, Umemura T, Nishimura J, Sadamura S, Sato H, et al. (1993) Double t(1;17)(p36;p11) in a megakaryocytic crisis of chronic myelogenous leukemia with variant t(5;9;22). Cancer Genet Cytogenet 65: 107–110.
40. Brown C (1996) Role of the X Chromosome in Cancer. J Natl Cancer Inst 88(8): 480–482.
41. Sirchia SM, Ramoscelli L, Grati FR, Barbera F, Coradini D, et al. (2005) Loss of the Inactive X Chromosome and Replication of the Active X in BRCA1-Defective and Wild-type Breast Cancer Cells. Cancer Res 65(6): 2139–2146.
42. Savino A, Koss LG (1971) The evaluation of sex chromatin as a prognostic factor in carcinoma of the breast. A preliminary report. Acta Cytol 15(4): 372–4.
43. Perry M (1972) Evaluation of breast tumor sex chromatin (Barr body) as an index of survival and response to pituitary ablation. Br J Surg 59(9): 731–4.
44. Kerler R, Rabes H (1996) Karyotype evolution of the clonal rat liver cell line CL 52 during progression in vitro and in vivo. Cancer Genet Cytogenet 87: 140–147.
45. Strefford JC, Lane TM, Hill A, LeRoux L, Foot NJ, et al. (2006) Molecular characterization of the t(1;15)(p22;q22) translocation in the prostate cancer cell line LNCaP. Cytogenet Genome Res 112: 45–52.
46. Wasylishen AR, Penn LZ (2010) Myc: the beauty and the beast. Genes Cancer 1(6): 532–41.
47. Karlsson A, Helou K, Walentinsson A, Hedrich HJ, Szpirer C, et al. (2001) Amplification of Mycn, Ddx1, Rrm2, and Odc1 in rat uterine endometrial carcinomas. Genes Chromosomes Cancer 31(4): 345–56.
48. Souzaki R, Tajiri T, Teshiba R, Higashi M, Kinoshita Y, et al. (2011) The genetic and clinical significance of MYCN gain as detected by FISH in neuroblastoma. Pediatr Surg Int 27(3): 231–6.

49. Albertson DG (2006) Gene amplification in cancer. Trends Genet 22(8): 447–455.
50. Ross J, Fletcher J, Bloom K, Linette G, Stec J, et al. (2003) HER-2/neu Testing in Breast Cancer. Am J Clin Pathol 120(Suppl 1): S53–S71.
51. Slamon DJ, Godolphin W, La Jones B, Holt JA, Wang SG, et al. (1989) Studies of the ERBB proto-oncogene in human breast and ovarian cancer. Science 244: 707–712.
52. Goldenberg MM (1999) Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. Clin Ther 21: 309–318.
53. Baselga J, Albanell J, Molina MA, Arribas J (2001) Mechanism of action of trastuzumab and scientific update. Semin Oncol 28: 4–11.
54. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, et al. (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344: 783–792.
55. Zhang X, Hashemi SS, Yousefi M, Ni J, Wang Q, et al. (2008) Aberrant c-erbB2 expression in cell clusters overlying focally disrupted breast myoepithelial cell layers: a trigger or sign for emergence of more aggressive cell clones? Int J Biol Sci 4(5): 259–69.
56. Li-Ping Z, Da-Lei Z, Jian H, Liang-Quan X, Ai-Xia X, et al. (2010) Proto-oncogene c-erbB2 initiates rat primordial follicle growth via PKC and MAPK pathways. Reprod Biol Endocrinol 8: 66.
57. Bièche I, Onody P, Laurendeau I, Olivi M, Vidaud D, et al. (1999) Real-time reverse transcription-PCR assay for future management of ERBB2-based clinical applications. Clin Chem 45: 1148–56.
58. Glazyrin A, Shen X, Blanc V, Eliason JF (2007) Direct detection of herceptin/trastuzumab binding on breast tissue sections. J Histochem Cytochem 55(1): 25–33.
59. Wang S, Saboorian MH, Frenkel EP, Haley BB, Siddiqui MT, et al. (2002) Aneusomy 17 in breast cancer: its role in ERBB2 protein expression and implication for clinical assessment of ERBB2 status. Mod Pathol 15: 137–145.
60. Ma Y, Lespagnard L, Durbecq V, Paesmans M, Desmedt C, et al. (2005) Polysomy 17 HER-2/neu status elaboration in breast cancer: effect on daily practice. Clin Cancer Res 11(12): 4393–4399.
61. Mrhalová M, Kodet R, Kalinová M, Hilská I (2003) Relative quantification of ERBB2 mRNA in invasive duct carcinoma of the breast: correlation with ERBB-2 protein expression and ERBB2 gene copy number. Pathol Res Pract 199(7): 453–461.
62. Cuadros M, Talavera P, López FJ, García-Peréz I, Blanco A, et al. (2010) Real-time RT-PCR analysis for evaluating the Her2/neu status in breast cancer. Pathobiology 77(1): 38–45.
63. van de Corput MP, Dirks RW, van Gijlswijk RP, van Binnendijk E, Hattinger CM, et al. (1998) Sensitive mRNA detection by fluorescence in situ hybridization using horseradish peroxidase-labeled oligodeoxynucleotides and tyramide signal amplification. J Histochem Cytochem 46(11): 1249–1259.
64. Dirks RW, Molenaar C, Tanke HJ (2003) Visualizing RNA molecules inside the nucleus of living cells. Methods 29: 51–57.
65. Itzkovitz S, van Oudenaarden A (2011) Validating transcripts with probes and imaging technology. Nat Methods 8(4 Suppl): S12–S19.
66. Hattori M, Sakamoto H, Satoh K, Yamamoto T (2001) DNA demethylase is expressed in ovarian cancers and the expression correlates with demethylation of CpG sites in the promoter region of c-erbB-2 and survivin genes. Cancer Lett 169(2): 155–64.
67. Kawasaki H, Taira K (2004) Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. Nature 431(7005): 211–217.
68. Hamm CA, Xie H, Costa FF, Vanin EF, Seftor EA, et al. (2009) Global Demethylation of Rat Chondrosarcoma Cells after Treatment with 5-Aza-29-Deoxycytidine Results in Increased Tumorigenicity. PLoS ONE 4(12): e8340.
69. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. Genes Dev 20: 515–24.
70. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, et al. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479(7371): 74–9.
71. Holland AJ, Cleveland DW (2009) Boveri revisited: chromosomal instability, neuploidy and tumorigenesis. Nat Rev Mol Cell Biol 10: 478–487.
72. Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, et al. (1996) 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. Proc Natl Acad Sci U S A 93(14): 7081–4.
73. Olaharski AJ, Sotelo R, Solorza-Luna G, Gonsebatt ME, Guzman P, et al. (2006) Tetraploidy and chromosomal instability are early events during cervical carcinogenesis. Carcinogenesis 27(2): 337–743.
74. Thompson SL, Compton DA (2011) Chromosomes and cancer cells. Chromosome Res 19(3): 433–44.
75. Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. Nature17 396(6712): 643–649.

SUPLEMENTARY FIGURES AND TABLES

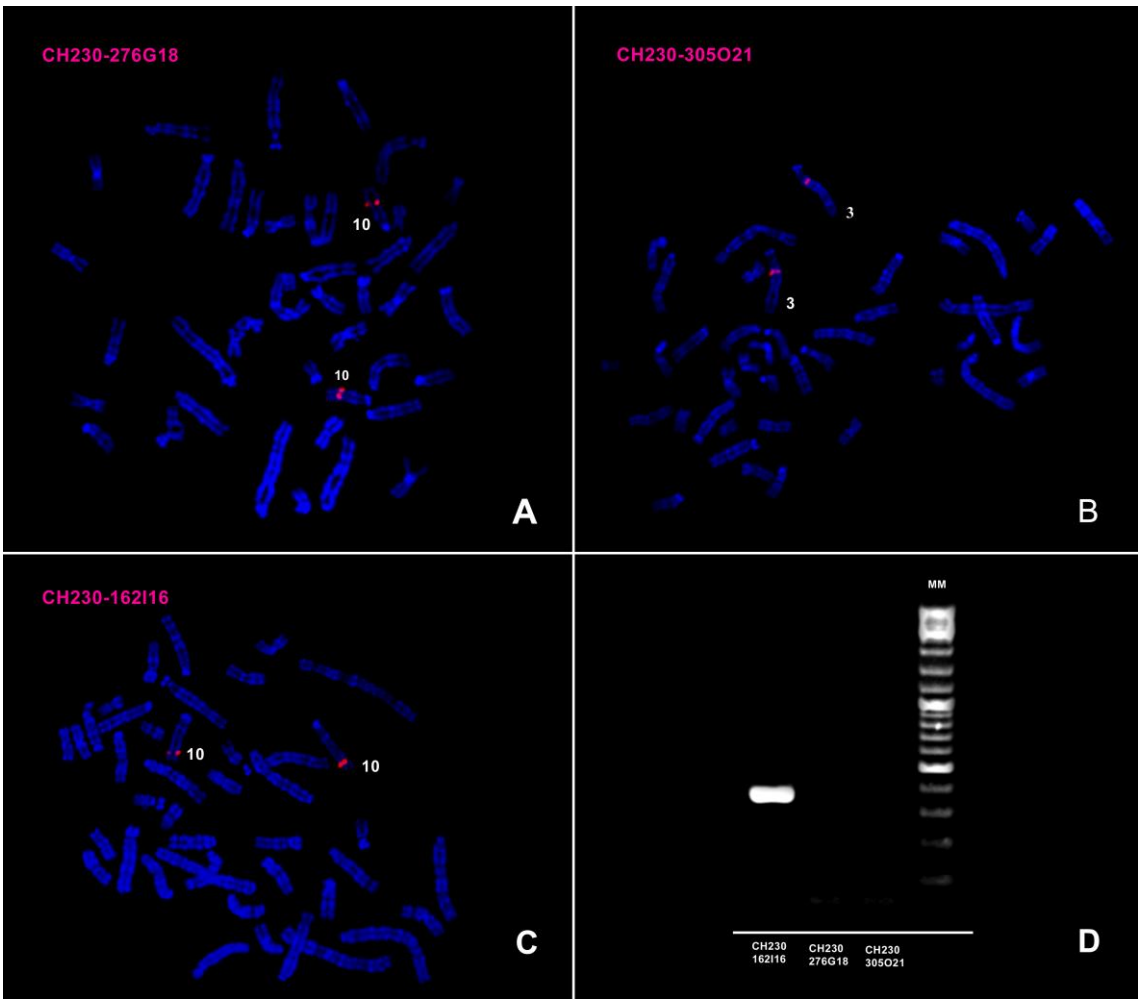


Figure S1- Representative images of the *in situ* hybridization of putative Erbb2 BAC clones onto RNO metaphases. Both CH230-276G18 (A) and CH230-305O21 (B) hybridize in different locations than the Erbb2 position determined by [24]. Only CH230-162I16 hybridizes at the cytogenetic position of Erbb2 in *Rattus norvegicus* 10q32.1 (C). PCR amplification of Erbb2 in the three clones (D). Only for CH230-162I16 BAC clone the expected 350 bp band is observed.

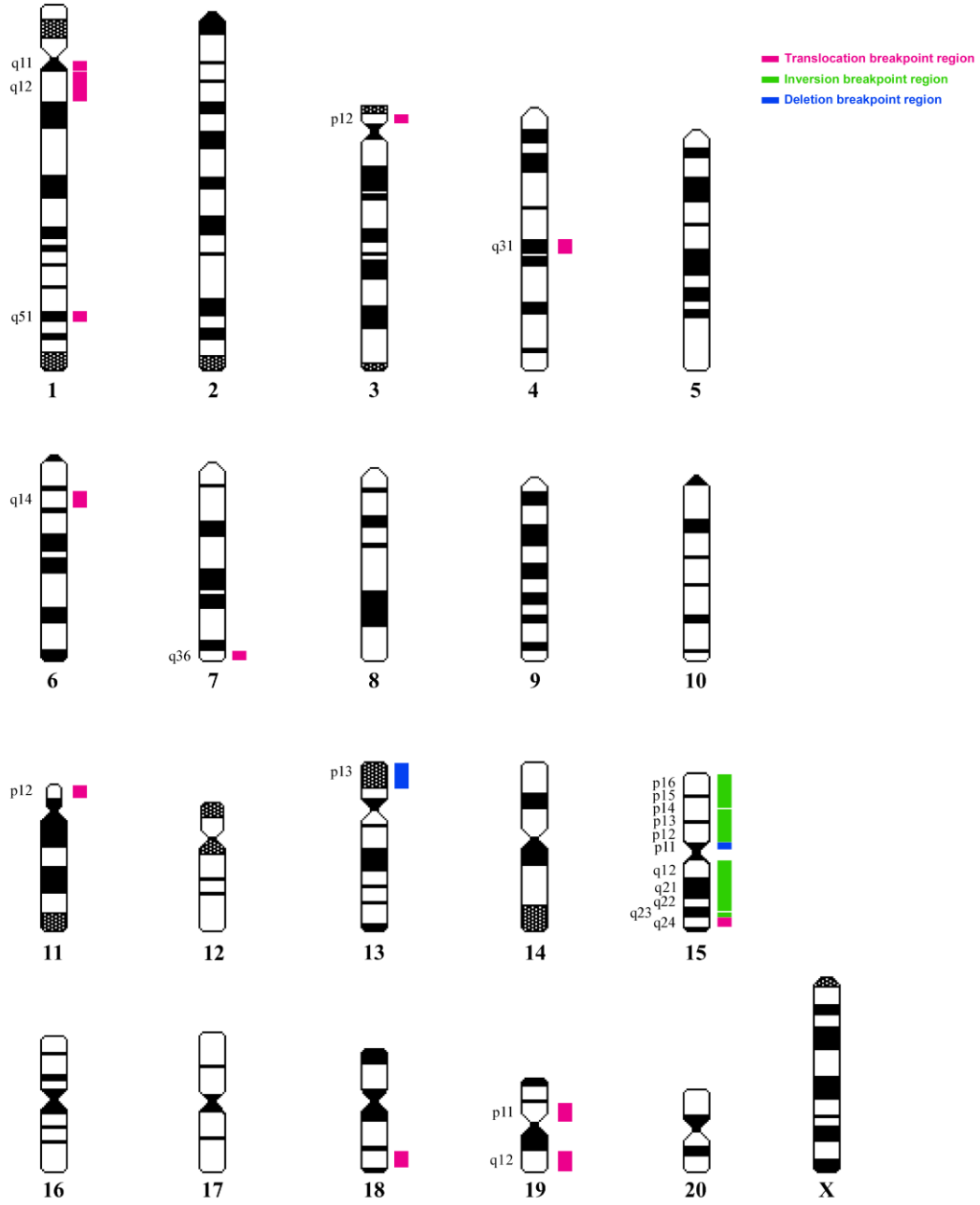
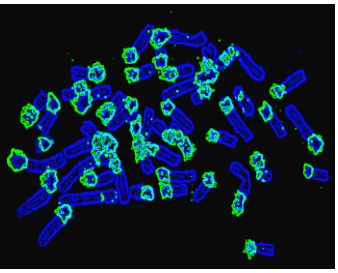


Figure S2- Chromosomal location of the clonal breakpoint regions in HH-16 cl.2/1 cell line. Clonal breakpoint regions in HH-16 cl.2/1cell line are displayed in the rat ideogram [20]. Each type of rearrangement originated by the breakpoints is identified by a specific color.

Table S1: *In silico* analysis of breast cancer related genes present in the most representative rat breakpoint bands, and its correspondent human
homolog.

RNO		Human		
Breakpoint region	**Gene**	**Homolog gene**	**Cytogenetic position**	**Gene Definition**
3p12	Rxra	RXRA	9q34.2	retinoid X receptor, alpha
	Ptges	PTGES	9q34.11	prostaglandin E synthase
6q14	Rhob	RHOB	2p24.1	ras homolog gene family, member B
	Ncoa1	NCOA	2p23.3	nuclear receptor coactivator 1
	Cad	CAD	2p23.3	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase
13p13	Bcl2	BCL2	18q21.33	B-cell CLL/lymphoma 2
15p12	Ctsb	CTSB	8p23.1	cathepsin B
	Gnrh1	GNRH1	8p21.2	gonadotropin-releasing hormone 1 (luteinizing-releasing hormone)
	Ptk2b	PTK2B	8p21.2	PTK2B protein tyrosine kinase 2 beta
15p13	Mmp14	MMP14	14q11.2	matrix metallopeptidase 14 (membrane-inserted)
15p14	Bmp4	BMP4	14q22.2	bone morphogenetic protein 4
	Apex	APEX	14q11.2	APEX nuclease (multifunctional DNA repair enzyme) 1
	Ang1	ANG1	14q11.2	angiogenin, ribonuclease, RNase A family, 5
15p16	Anxa7	ANXA7	10q22.2	annexin A7
	Thrb	THRB	3p24.2	thyroid hormone receptor, beta
19q12	Cdh1	CDH1	16q22.1	Cadherin 1, type 1, E-cadherin (epithelial)
	Zfhx3	ZFHX3	16q22.2	Zinc finger homeobox 3
	Bcar1	BCAR1	16q23.1	Breast cancer anti-estrogen resistance 1
	Cdh13	CDH13	16q23.3	Cadherin 13, H-cadherin (heart)

CHAPTER III |



GENERAL DISCUSSION AND FUTURE PERSPECTIVES

1. GENERAL DISCUSSION

Genomes are constantly changing, evolving either by point changes in the nucleotide sequence or by chromosome rearrangements marking major evolutionary steps. The main objective of this thesis was the study of karyotype restructuring in rodent species, both during species evolution and cancer. Specifically and regarding species evolution it was performed the analysis in two different perspectives: single copy architecture through the construction of comparative maps, and analysis of the repetitive fraction with the characterization of satellite DNA sequences. Concerning cancer it was made the characterization of clonal chromosome rearrangements and their influence in gene changes (amplification, expression and methylation status) in two rat tumor commercial cell lines.

In this thesis section it will be performed an integrated discussion of all the results presented, divided in five parts: 1.1) Comparative chromosome study in Rodentia, 1.2) Characterization and evolution of satellite DNA in Cricetidae; 1.3) The role of CH and satDNA in Muroids karyotype restructuring; 1.4) Cancer chromosomes and cell lines as model and 1.5) Concluding remarks.

1.1 COMPARATIVE CHROMOSOME STUDY IN RODENTIA

Rodents constitute a very karyotypically diverse order, with muroids presenting record high rates of karyotype evolution (Romanenko *et al.* 2006). To understand the dynamics of chromosomal evolution it is fundamental to know how genomes are organized and which types

of chromosomal rearrangements are implicated in macroevolutionary events and speciation. Great advances were achieved throughout molecular cytogenetic data and availability of partially or fully sequenced genomes, leading to the identification of conserved chromosomal syntenies, syntenic associations and ultimately to hypothesized ancestral karyotypes (Graphodatsky *et al.* 2011, Ruiz-Herrera *et al.* 2012). Plus, such studies have permitted reliable phylogenetic reconstructions using genomic data. The present thesis describes the construction of high resolution chromosome maps of three Rodentia species, specifically one Muridae: *Praomys tullbergi* (PTU), and two Cricetidae: *Cricetus cricetus* (CCR) and *Peromyscus eremicus* (PER). The assembling of these comparative maps permitted the disclosure of their genome architectures, as well as the delineation of the chromosome evolutionary history since the common Muroidea ancestor.

Identification of the chromosomal rearrangements that shaped the genomes of PTU, CCR and PER

Painting probes from *Mus musculus* (MMU) were used in the comparative chromosome painting experiments allowing to identify syntenic associations, and the further integration and comparison of our data with the available from all the species studied so far, elucidating the most probable sister clades. Besides, the complementation of this map with the use of *Rattus norvegicus* (RNO) paint probes and the integration of sequencing data from both index genomes available in Ensembl database (http://www.ensembl.org/; release 66 - Feb 2012) permitted to further refine this investigation by disclosing intrachromosomal rearrangements (inversion events) that would escape detection by chromosome painting and using only one genome. The number of syntenic segments disclosed in the performed studies (Table 1) revealed to be high (when we consider the intrachromosomal segments) comparing with previous works (e.g. Romanenko *et al.* 2006). A number close to these could only be observed for the comparison of MMU and RNO in a blastZ-Net Analysis available in Ensemble database.

The analysis of the shared syntenic associations allowed determining the closest clade to each of the studied species. Regarding *Praomys tullbergi*, *Rattus* group (*R. norvegicus and R. rattus*) seems to be the closest Muridae genomes to that species, and *Microtus oeconomus* (Arvicolinae) the closer Cricetidae species. Remarkably, Lee and Martin (1988) suggested also a close phylogenetic relationship between the related species *Praomys natalensis* and *Rattus norvegicus*, using conventional G- and C-banding and Ag-NOR's analysis. Concerning *C. cricetus*, the higher number of syntenic associations is shared with *Cricetulus griseus, CGR* (Cricetinae). These results support G-banding data (Gamperl *et al.* 1976) and molecular cytogenetic studies (*Cricetus cricetus*/*Cricetulus griseus*/*Mesocricetus auratus* comparative map) (Romanenko *et al.* 2006, 2007).

Table 1| Summary of the results obtained with painting experiments and constitutive heterochromatin analysis.

	Syntenic Segments		Colocalization EBr/CH (%)			Colocalization CH/EBr (%)		
	RNO	MMU	RNO	MMU	AMK	RNO	MMU	AMK
P. tullbergi	33 (61)	44(66)	70%	74%	94%	-	-	-
C. cricetus	32(67)	44(77)	49%	49%	57%	68%	77%	34%
P. eremicus	29(62)	38(65)	64%	71%	67%	25%	30%	2%

MMU- *Mus musculus;* RNO- *Rattus norvegicus;* AMK- Ancestral Muroid Karyotype; EBr- Evolutionary breakpoint; CH- constitutive heterochromatin. The numbers between brackets indicate the number of syntenic segments when accounting for the intrachromsomal segments. The constitutive heterochromatin was accessed by counting the C-bands. EBr/CH- Percentage of evolutionary breakpoint regions that are colocalized with C-bands; CH/EBr- Percentage of C-bands that are colocalized with evolutionary breakpoint regions.

Finally, *P. eremicus* seems to be closely related with *Peromyscus maniculatus* and *C. cricetus*, both Cricetidae species. The similarity with *Peromyscus maniculatus* is not surprising, once all the *Peromyscus* species have similar karyotypes (Greenbaum *et al.* 1994). Most interesting seems to be the highest similarity with CCR (Cricetinae) than with Sigmodontinae subfamily species, what is not in accordance with molecular data, which suggests the highest phylogenetic proximity between Neotominae (*Peromyscus* family) and Sigmodontinae among the Cricetidae subfamilies (Steppan *et al.* 2004). Once few studies have been performed in Sigmodontinae species and those results were not clear, one have to be cautious when withdrawing conclusion regarding phylogenetic relationships involving this group.

One important outcome presented in this thesis was the delineation of a putative Ancestral Muroidea Karyotype (AMK), based in the analysis of the MMU syntenic associations outlined by several works and supported by the results of the species studied. Besides, it was also used the Ensembl database assisting in the reconstruction of a high precision AMK by the identification of the proximal, medium and distal regions of chromosomes, allowed by the combined use of MMU and RNO index genomes. The AMK proposed here has a diploid number 2n=50, and was suggested based in the comparison of shared syntenic MMU associations in 31 species (20 Muridae and 11 Cricetidae), using a cladistic analysis. This karyotype has four chromosomes less than the one proposed by Stanyon *et al.* in 2004 (2n=54), two chromosomes less than the one suggested by Engelbrecht *et al.* (2006) (2n=52) and two more chromosomes than proposed by Romanenko *et al.* (2006) (2n=48). In the proposed AMK the segment homologous to 16dist of MMU would be a single chromosome. The association (with an inversion) 13dist/15prox/13med is an ancestral state and is a unique chromosome in the Muroidea ancestor, in accordance with Veyrunes *et al.* 2006. Moreover, the proposed ancestral karyotype differs from that very recently suggested by Romanenko *et al.* (2012) which has 2n=52

chromosomes, not presenting the composition 8/2/13 and 5/14, once these were found only in Cricetidae and therefore should be considered derived traits for this group. Moreover the ancestral chromosome originated by the inversion 1/17/1 is not considered ancestral since this chromosome form is observed mainly in Muridae species.

One of the main goals of comparative studies is the delineation of chromosomal changes from an ancestral karyotype, determining the type of chromosomal rearrangements that modelled the karyotype of the studied species. From the presented study, *P. eremicus* revealed a high number of ancestral syntenic associations, revealing the occurrence of a small number of evolutionary large-scale rearrangements (fusion/fission), evidencing that this species has a conserved genome. *C. cricetus* and *P. tullbergi* revealed the occurrence of a high number of rearrangements, mainly translocation events. In all the studied species, inversion events were identified. Similar results were obtained for PER and CCR using different painting probes in a study by Romanenko *et al.* (2007). Finally in PER and CCR, constitutive heterochromatin repatterning was also evident and corroborated by earlier studies (Pathak *et al.* 1973, Gamperl *et al.* 1976).

1.2 CHARACTERIZATION AND EVOLUTION OF SATELLITE DNA IN CRICETIDAE

Genomes can be considered a mosaic comprising regions of fragility that are prone to reorganization and regions that do not exhibit the same levels of plasticity. Particularly, constitutive heterochromatin regions are thought to be "hotspots" for structural chromosome rearrangements (John 1988, Chaves *et al.* 2004) and its constituents, satellite DNAs, are most likely the responsible for promoting genomic plasticity and consequently high rates of chromosome rearrangements (Slamovits and Rossi 2002). This occurs, most probably, due to the fact that tandem repeats promote non-homologous chromosomal recombination, thereby promoting chromosomal reorganization (Froenicke and Lyons 2008). Furthermore tandem repeats have been shown to be translated into noncoding RNAs, which revealed to facilitate chromosome changes, creating novel karyotypes and thus affecting chromosome evolution (Brown *et al.* 2010).

With the objective of investigating such interesting characteristics of satDNA and their involvement in karyotype restructuring, different sequences were isolated from the genome of some rodent species and were analyzed along the presented thesis. Moreover, after this analysis the gathered information could be integrated with the performed comparative maps providing a global picture of these genomes' evolution.

Cricetidae species share orthologous satellite DNAs with different modes of evolution

In the present thesis, two examples of repetitive DNA dynamics and their implication in karyotype architecture in rodents (Cricetidae) were observed. Both sequences were isolated *de novo,* using laser microdissection from the genomes of *C. cricetus* (CCR4/10sat) and *P. eremicus* (PMSat). This approach revealed to be effective in the isolation of repetitive sequences as previously described by other authors (Li *et al.* 2005, Pauciullo *et al.* 2006). The fact that these sequences are shared by different species, which places them in a common ancestor, conciliated with the fact that they present different locations (CCR4/10sat) or copy number variations (PMSat) in the analyzed species, indicates a highly dynamic behavior. The estimated age for CCR4710sat is, at least, 19 million years old, according with the estimated time of divergence of the different Cricetidae subfamilies (Steppan *et al.* 2004). On the other hand, PMSat was detected not only in Rodentia, but also in species belonging to distant groups (Artropoda and Echinodermata), indicating that this constitutes the oldest satDNA described so far, with at least 635 million of years according to Edgecombe *et al.* (2011). Despite being present in the same species (CCR and PER) both sequences disclose very different modes of evolution. In the case of CCR4/10sat, the most parsimonious scenario seems to be its presence at (peri)centromeric regions in the ancestral karyotype of Cricetidae, as the observed condition in *Cricetus cricetus* chromosomes, being afterwards these sequences moved throughout the genome by means of intragenomic movements, assuming a scattered pattern in *Peromyscus eremicus*. Similar results were obtained by Adega *et al.* (2008) in Suidae vs. Tayassuidae families, with orthologous satellite DNA sequences that revealed a completely different chromosomal location, reflecting a high level of karyotypes divergence after the radiation of each family. The dissimilar patchwork of the CCR4/10sat satellite in the two species (CCR and PER) could have been established by mechanisms involved in the sequence homogenization during concerted evolution. Good candidates are the saltatory amplification mechanisms consisting in excision, rolling circle replication, and reinsertion of extrachromosomal circular DNA, known for participating in the turnover of tandem repeats (Navratilova *et al.* 2008, Cohen and Segal 2009), and that can promote the relocation of satDNA sequences. Other mechanism likely to promote tandem repeat dynamics in different organisms is transposon-mediated exchange (Palomeque *et al.* 2006, Macas *et al.* 2009) which can contribute to the plasticity and also to the scattering of repetitive DNAs.

Regarding PMSat, orthologous sequences were isolated in different Cricetidae species and were also detected in the genome of phylogenetic distant species following a BLAST search against available data from genome sequencing projects. High interspecies sequence identity was observed and a wide variation in number of copies among the analyzed rodent species (*P. eremicus*,

C. cricetus, *Phodopus sungorus* and *Microtus arvalis*). These characteristics seem to indicate that its evolutionary pathway occurred throughout copy number fluctuation, culminating in different profiles (Fry and Salser 1977). Mechanisms such unequal crossing over, replication slippage and rolling circle amplification are proposed to be responsible for the amplifications/deletions of repeated DNA (Walsh 1987). The amplification and contraction events may not be accompanied by changes in the satDNA nucleotide sequences, as it has been demonstrated in other studies (Bruvo *et al.* 2003, Ellingsen *et al.* 2007), what is in accordance with the interspecific sequence homology found. The inexistence of species-specific mutations, and very high homology for the majority of the PMSat monomers from the rodents analyzed can be explained by the occurrence of non-concerted evolution, as observed for other satellite DNAs (e.g. Plohl *et al.* 2010a). Our results strongly suggest that PMSat is under evolutionary constraints in the species studied and probably constitutes a functional element in their genomes. Supporting this theory, transcripts from a similar sequence were found in *Peromyscus* sister species, *Peromyscus maniculatus biirdi*. This work constitutes an important starting point as it will be very interesting to perform transcription studies in this satellite.

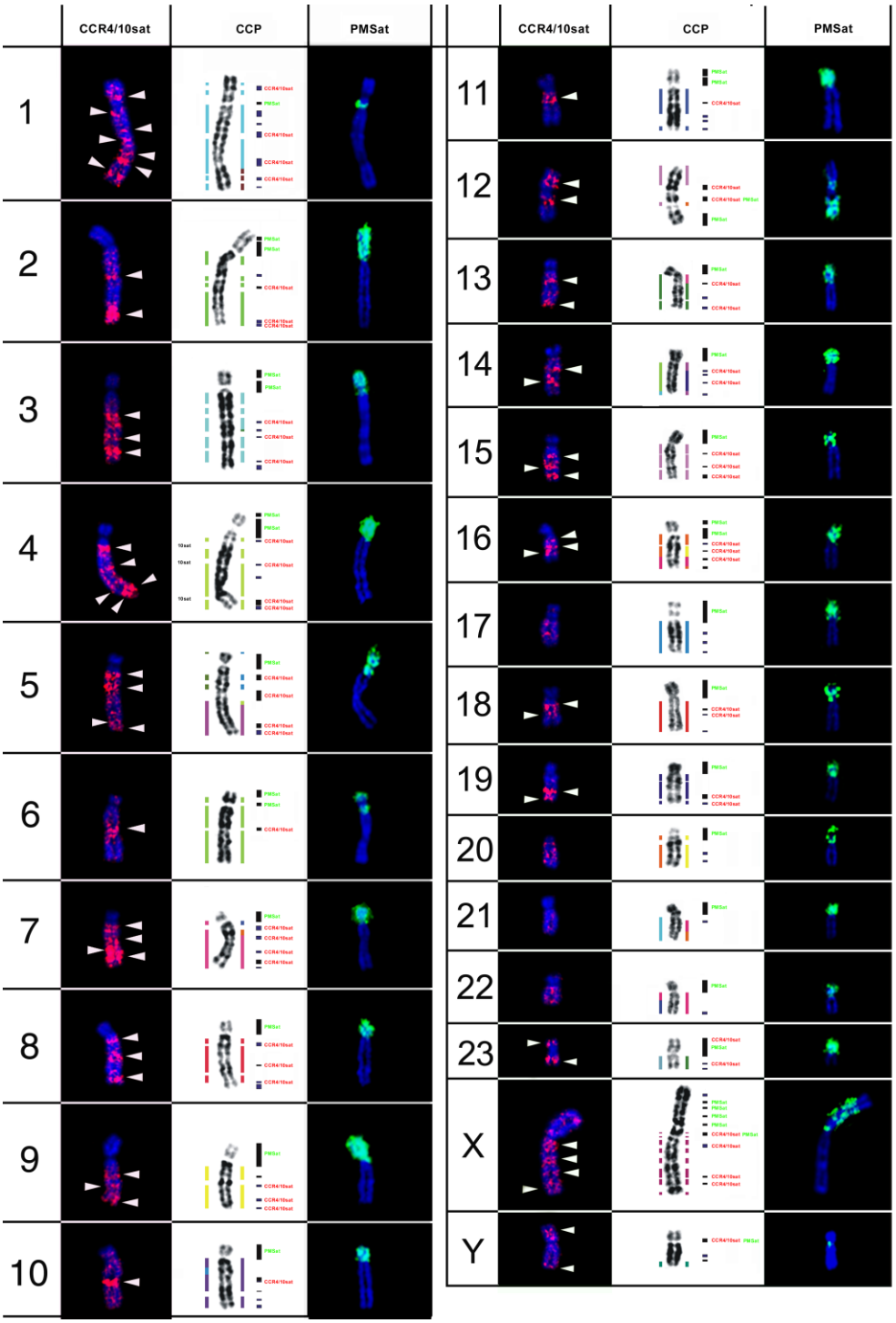
1.3 THE ROLE OF CH AND SATDNA IN MUROIDS KARYOTYPE RESTRUCTURING

When comparing constitutive heterochromatin, satellite DNA and the comparative chromosome painting data, a clear association stands out (Table 2). The integration analyses of the obtained results permitted to identify different evolutionary patterns of the rodents' genomes in analysis, particularly in the Cricetidae subfamilies.

Karyotype restructuring in muroids can be related with CH

Comparative maps permit the identification of the chromosomal homologous segments between species, and also the identification of the regions where genome synteny has been disrupted by chromosomal rearrangements, i.e., the breakpoint regions. The demonstration of breakpoints reuse during chromosome evolution (Murphy *et al.* 2005), revealed a predisposition of certain genomic regions to instability (hotspots) that leaded to the formalization of the fragile breakage model (Pevzner and Tesler 2003). The presence of repetitive sequences at evolutionary breakpoint regions has been shown by whole genome alignment studies (Murphy *et al.* 2004, Ruiz-Herrera *et al.* 2006), besides as already was referred CH has been considered a hotspot for structural chromosome rearrangements (Yunis and Yasmineh, 1971, John, 1988, Chaves *et al.* 2004).

Table 2| Summary table showing the *in situ* hybridization pattern of CCR4/10sat and PMSat in *Peromyscus eremicus* chromosomes.



In the centre column are shown the comparative results (syntenic segments identified), CH mapping and localization of the satellites in PER chromosomes. The white arrowheads evidence the CCR4/10sat signal in *P. eremicus* chromosomes.

During the present work an analysis of the co-localization of CH and the identified evolutionary breakpoint regions for *Praomys tullbergi*, *Cricetus cricetus* and *Peromyscus eremicus* was performed. In order to elucidate about the CH dynamic in these genomes during evolution, it was analyzed the percentage of EBr that are colocalized with CH, and also the percentage of C-bands that are located at breakpoint regions. The values are summarized in Table 1, and as it can be seen, substantial values were obtained for all the analyzed species. These results suggest that the CH and specifically, the repetitive sequences under analysis could, in fact, be somehow involved (maybe as drivers) in the chromosome rearrangements occurred during karyotype evolution of these Rodentia species. The localization of the isolated satellites CCR4/10sat and PMSat in *P. eremicus* and *C. cricetus* chromosomes are summarized in Figure 1. An integrated discussion regarding this correlation will be carried in the next paragraphs.

Cricetidae evolutionary trails are related with satellite DNA dynamics

Rapid variation of satellite DNAs have been hypothesized as important source of chromosomal instability and consequently karyotype evolution (Slamovits *et al.* 2001). In the light of these assertions, it seems that CCR4/10 sat and PMSat were implicated in the genome restructuring of the species studied. Repetitive sequences with shared patterns in all the analyzed species are classified as conservative, while those present only in some species and absent or in low copy number in others, are classified as rapidly evolving (Wichman *et al.* 1991).

According to Volobouev *et al.* (2006), an important source of karyotype variability in rodents is related to the heterochromatin amount and distribution pattern. Within *Peromyscus* genus, the interspecies differences in heterochromatin amount expressed by the quantity of nuclear DNA reaches 36% (Deaven *et al.* 1977). Also, the karyotype of the genus is notable once it comprises a constant diploid chromosome number of 2n=48 for the more than 50 species comprising it (Committee for standardization of chromosomes of *Peromyscus* 1977).

The species of the genus *Peromyscus* constitute a striking example of karyotype evolution by heterochromatic arm additions/deletions (Pathak *et al.* 1973), along with pericentric inversions and centromere repositioning (Robbins and Baker 1981, Rogers *et al.* 1984). The comparative studies results presented along in this thesis corroborate such suggestions, with *P. eremicus* presenting a very similar karyotype to the AMK, with few restructurings, among them some inversions and CH additions. It must be referred that the technique used has the drawback of not revealing intrachromosomal rearrangements, preventing the identification of other possibly existing inversion events.

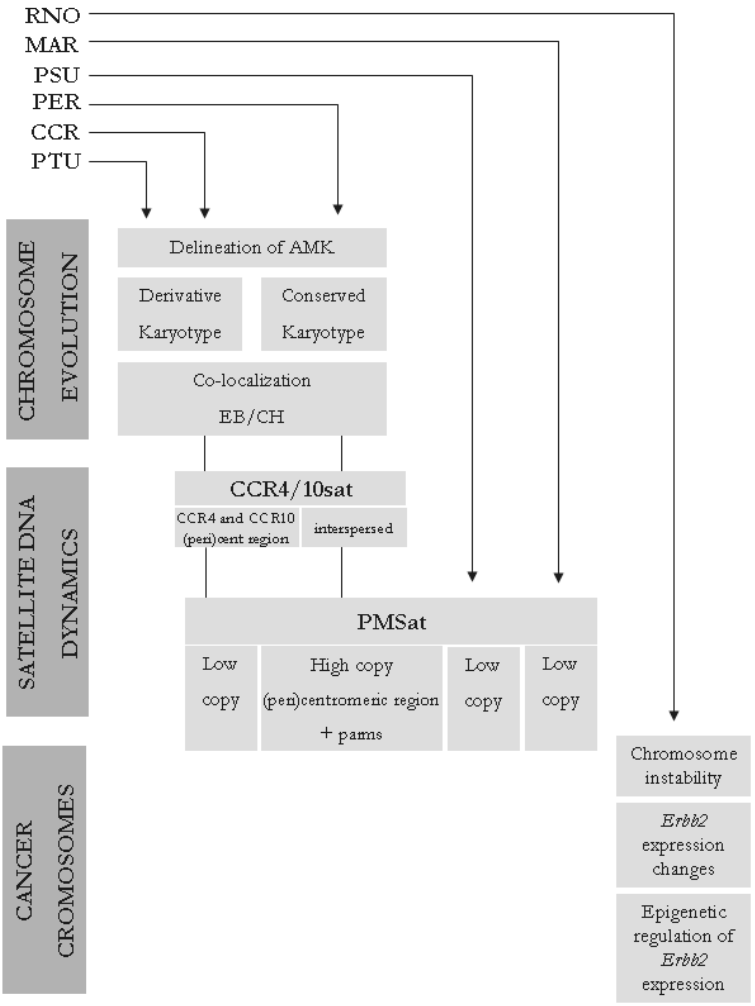


Figure 1| Schematic diagram summarizing the different topics studied, the species involved in each topic and some of the main outcomes.

Attending to the mechanisms enumerated above (saltatory amplification and transposon-mediated exchange), it seems feasible to propose that the rearrangements occurred in *Peromyscus* lineage (mainly inversions) could have been facilitated by the presence of repetitive sequences (regions of "fragility") (Ruiz Herrera *et al.* 2006), namely CCR4/10 sa. As already referred, rapidly evolving satDNA families promote chromosomal rearrangements by means of their intragenomic movements among nonhomologous chromosomes and between different chromosomal fields, as centromeres, arms and telomeres Wichman *et al.* (1991).

PMSat satellite DNA also seems to be involved in *P. eremicus* karyotype evolution, particularly regarding CH repatterning in this species. As part of PER constitutive

heterochromatin, this satellite DNA dynamics by means of amplification events seem to have led to the CH additions, resulting in the present large CH blocks, which extend to the entire short arms of the majority of PER chromosomes. The selective amplification of this satellite over other possibly existing ones might be related with functional constraints, or heterochromatic environment imposed constraints (Plohl *et al.* 2008). Previous studies in this genus suggest that heterochromatin additions seem to have been favoured along species evolution, and for that reason it was assumed that it confers some selective advantage to the organisms (Greenbaum and Baker 1978). Table 2 shows an integrated view of the comparative results, along with CH and the satDNAs (CCR4/10sat and PMSat) mapping. In this table it can be seen that some of the breakpoints identified co-localize with CH (as referred) and simultaneously with satellite DNA, illustrating the possible role of these sequences in PER karyotype restructuring.

Regarding the other species studied in this thesis, *Cricetus cricetus* (our results) and *Phodopus sungorus*, their chromosome evolution is characterized by the occurrence of fusion events (centric and/or tandem fusions), while robertsonian translocations (centric fusion translocation) were the predominant rearrangement in the karyotype evolution of *Microtus arvalis* (studied elsewhere, Romanenko *et al.* 2007, Romanenko *et al.* 2012). Li *et al.* (2000) demonstrated the involvement of two satDNAs of the muntjac deer genome in tandem fusions of chromosomes. Also the correlation between fusion evolutionary events and satellite DNA has been well documented (e.g. Garagna *et al.* 2001, and reviewed in Adega *et al.* 2009). Moreover, Chaves *et al.* (2003) highlighted the organization of different satDNA families in the t(1;29) translocation in cattle. Following these examples it can be proposed that CCR4/10sat (located in the centromeres of CCR4 and CCR10) can be implicated in the fusion events occurred during *C. cricetus* evolutionary pathway. Fusions have been also associated to the loss of centromeric and telomeric constitutive heterochromatin at the fusion breakpoints, possibly to stabilize those rearrangements and ensure its fixation in during the speciation process (Shi *et al.* 1980, Iannuzzi *et al.* 1987). The presented work in PMSat shows evidences of contraction events occurred during PMSat evolution in *C. cricetus*, *P. sungorus* and *M. arvalis*, once this sequence exists in the genome of these species, but in a much lower amount than in PER. Such satellite mode of evolution may have contributed to the occurrence of chromosome fusions in *C. cricetus*, *P. sungorus* and *M. arvalis* species, as has been shown for the rodent *Ctenomys* (Slamovits *et al.* 2001). Such elimination could have resulted in the decrease of copy number of this satellite in the mentioned species to a residual amount, enough to be detected by molecular techniques but undetectable by hybridization ones (southern blot and FISH). Regarding *C. cricetus* a question arises: Why the selection against PMSat in striking contrast with CCR4/10sat that still highly represented in CCR chromosomes? The elimination of satellite

copies can be attributed, for example, to selective pressure on the maintenance of the size of the genome as proposed by Nijman and Lenstra (2001). So, the loss of PMSat from CCR chromosomes has been probably balanced by amplification of other repeated DNA sequences, as CCR4/10sat, thus maintaining the heterochromatic features of these regions.

The organizational and structural properties of repetitive sequences, and the general features of its dynamics (copy number variations and intragenomic movements) analyzed and correlated with the chromosome architecture in the Rodentia genomes analyzed, allowed to support the idea proposed in previous works, that constitutive heterochromatin and satDNA play an important role in chromosome evolution. Moreover, this role seems to rely (at least) on the ability of the satellites to change its copy number and to mobilize through the genome.

1.4 CANCER CHROMOSOMES AND CELL LINES AS MODEL

The last part of this thesis was dedicated to the genetic/cytogenetic characterization of two related DMBA-induced rat mammary tumor cell lines: HH-16 cl.2/1 and HH-16.cl.4. This analysis allowed major outcomes that will be succinctly described.

Rat tumor cell lines are characterized by distinct karyotypic changes and instability in the chromosomes

As mentioned before, tumor progression is characterized by the successive accumulation of genetic changes in cells, which gives them the ability to proliferate without restrictions. Genomic instability was considered to contribute to the tumor stepwise progression, occurring as small changes at the nucleotide level affecting cancer-related genes, as gene amplification or as gross chromosomal alterations (reviewed in Martin *et al.* 2010). The cytogenetic analysis of both cell lines revealed significant changes in their karyotypes, suggesting the presence of chromosomal instability (CIN) and chromosome structure instability (CSI). Ploidy changes were the major feature of HH-16.cl.4 cell line presenting a nearly tetraploid karyotype, with a wide range of cells showing different chromosome numbers. In contrast, a nearly diploid karyotype with low levels of polyploidy was found in HH-16 cl.2/1, what allowed a detailed characterization of the structural rearrangements in this cell line.

It is proposed that most tumors arise from a single cell of origin, and tumor progression results from acquired genetic variability within the original clone, allowing sequential selection of the more aggressive subclones (reviewed in Tysnes 2010). HH-16 cl.2/1 clonal rearrangements were identified and allowed tracing a putative clonal evolution for this tumor. A diagram was constructed which permitted the visualization of the ancestral and recent rearrangements, as well

as providing an overview of the microevolutionary processes that have occurred in the progression of this tumor cell line (Figure 5 of Paper V).

The instability in the chromosomes influenced cancer-related genes, specifically their expression

It has been demonstrated that chromosome structure instability can influence tumorigenesis by deregulating expression of specific target genes or by promoting gene fusion (reviewed in Thompson and Compton 2011). An *in silico* analysis was performed focused on the breakpoint regions of the ancestral structural chromosome rearrangements identified in HH-16 cl.2/1, demonstrating that almost all of the breakpoint regions contain genes in the rat genome for which the human homolog has been associated with breast cancer. This information can direct future studies, when using this available and now characterized cell line.

Moreover, the cytogenetic characterization of HH-16 cl.2/1 and HH-16.cl.4 revealed an increased number of copies of two oncogenes, *Mycn* and *Erbb2*. According with the defined criteria for gene amplification, neither of the genes is considered amplified, but a *Mycn* gain was considered to have occurred in HH-16 cl.2/1 and an *Erbb2* gain was considered to have occurred in HH-16.cl.4. In the present case it was clear the association of chromosome rearrangements/karyotype restructuring in tumor progression, specifically by causing changes in oncogene copy number. As shown previously in this thesis, gene amplification has been associated with overexpression of the amplified gene(s) (Albertson 2006), although this correlation is not absolute. Relative RNA expression was assessed by real-time reverse transcription quantitative PCR (RT-qPCR) and RNA FISH. Our results show that the *Mycn* gain in the HH-16 cl.2/1 cell line was not reflected in an RNA expression change. Alternatively, *Erbb2* was found to be overexpressed in HH-16.cl.4, in particular a 10.7-fold increase was determined by RT-qPCR. RNA FISH also revealed to be a reliable and advantageous way of access gene expression, corroborating the results from RT-qPCR and showing a higher expression in HH-16.cl.4 than in the sister cell line HH-16 cl.2/1. Several hypothesis were outlined to explain *Erbb2* overexpression in HH-16.cl.4, some of them related with the observed chromosome changes: 1) the referred *Erbb2* gain, while low, may have played a role in the overexpression of *Erbb2* RNA in this cell line, although it may have not been the only mechanism involved; 2) the polysomy of RNO10, once some authors point polysomy (in HSA17, chromosome location of *ERBB2* in Human) as the cause of *ERBB2* overexpression in humans (Wang *et al.* 2002, Ma *et al.* 2005); 3) epigenetic regulation, namely transcriptional regulation which could have promoted the accumulation of *Erbb2* mRNA in the absence of high levels of amplification.

Erbb2 expression is influenced by methylation in HH-16.cl.4 cell line

The influence of DNA methylation in cancer has become the topic of intense investigation, once cancer cells have been shown to have major changes in their DNA methylation patterns, when compared with normal cells (Das and Singal 2004).

In the presented work it was verified that the expression of *Erbb2* in the HH-16.cl.4 rat cell line appears to be affected by global genome demethylation, after cells treatment with 5-aza-2'-deoxicitidine. Interestingly, in a similar study to ours performed in a rat chondrosarcoma cell line, an increase in *Erbb2* expression was found after global genome demethylation (Hamm *et al.* 2009), while our data shows a decrease in *Erbb2* expression after 5-Aza-2'-deoxicitidine treatment. Different epigenetic mechanisms underlying the expression of this gene seem to be pivotal in each of these cell lines. While in the rat chondrosarcoma cell line, *Erbb2* promoter unmethylation seems to be the main cause for *Erbb2* overexpression, our data suggests the action of negative regulators of *Erbb2* expression. Our findings emphasize that future studies are mandatory to reveal the exact epigenetic events involved in the regulation of *Erbb2* expression, and that HH-16.cl.4 cell line is an excellent tool to complete this task.

Same origin but different outcomes suggest different mechanisms involved in tumor progression of HH-16 cl.2/1 and HH-16.cl.4

The cell lines used in this work were generated simultaneously from the DMBA-induced rat mammary tumor (Steffen *et al.* 1998), but despite having the same initial genetic background, they display quite different cell lineages: HH-16 cl.2/1 reflect mesenchymal cells of the stromal part of the tumor, and HH-16.cl.4 cell line display an epithelial origin. This fact along with the higher chromosomal instability revealed by HH-16.cl.4 (explaining the *Erbb2* overexpression observed), suggests different mechanisms involved in tumor progression of each of the cell lines. In HH-16.cl.4, tetraploidy associated mechanisms seem to be responsible for the observed karyotypic features by promoting the initiation of chromosomal instability (CIN) as it has been found in several cancers (e.g. Olaharski *et al.* 2006). On the other hand, in the fibroblastoid H-16 cl.2/1 cell line, chromosome structure instability (CSI) seems to be the distinguishing feature, whose mechanisms are now starting to be disclosed (Thompson and Compton 2011).

HH-16 cl.2/1 and HH-16.cl.4 as cellular models

We believe that the performed characterization of these two rat mammary cell lines, throughout molecular cytogenetics, gene expression profiling and examination of the influence of global demethylation on gene expression, provides a platform for future studies on tumor

progression and encourage the use of these two cell lines as a model. In particular, this study highlights HH-16 cl.2/1 and HH-16.cl.4 potential as models for studying *Erbb2* associated mechanisms and as experimental tools to assist in the generation of new biotherapies. We are confident that his work will contribute to the validation of this cellular model and to its use in future studies.

1.5 CONCLUDING REMARKS

Following the analogy by Peng *et al.* (2006), if a genome is compared to a continental landform, genome rearrangements would comprise the evolutionary "earthquakes" that dramatically change the landscape. In the present thesis the comparative maps of three rodent species, *Praomys tullbergi, Cricetus cricetus* and *Peromyscus eremicus*, were constructed based in the comparative chromosome painting results using mouse and rat probes and refined using Ensembl available data. Based on the obtained results, the main evolutionary rearrangements occurred in each species karyotype were identified, and *P. eremicus* genome stand out as possessing the most conserved genome, while *P. tullbergi* and *C. cricetus* revealed to be more derivative. One major outcome was the proposal of an Ancestral Muroidea Karyotype presenting a diploid number 2n=50. Going back to the starting analogy, a fundamental question in studies of chromosome evolution is whether the evolutionary "earthquakes" are happening along evolutionary "faults" (hotspots of rearrangements). A significant correlation between the identified evolutionary breakpoint regions in the studied species and the constitutive heterochromatin location in the chromosomes was observed. This is in accordance with other studies that propose that constitutive heterochromatin regions constitute hotspots for chromosomal rearrangements.

The major constituent of constitutive heterochromatin is satellite DNA. Two satellite DNA families (CCR4/10sat and PMSat) were isolated *de novo* by microdissection in *C. cricetus* and *P. remicus*, revealing distinct evolutionary trails. While CCR4/10sat dynamics seems to be throughout intragenomic movements (displaying different chromosome locations in the different species), PMSat seems to have changed by copy number variations (striking dissimilarity in abundance in the different species). PMSat constitutes the oldest satDNA family described so far, existing in the genome of distantly related species. Despite that, high sequence conservation was observed, what strongly suggests a possible functional activity of this sequence in the genome. These works constitute a good example of satellite DNA duality, once they seem to influence the karyotypes restructuring by its dynamic behavior (location and copy number change) but also its sequence can be evolutionarily conserved, suggesting a functional role.

The karyotype restructuring is also a feature of tumor progression. The genetic/cytogenetic characterization of two related DMBA-induced rat mammary tumor commercial cell lines (HH-16 cl.2/1 and HH-16.cl.4) was accomplished. The cytogenetic characterization of HH-16 cl.2/1 allowed the identification of clonal rearrangements, the reconstruction of tumor clonal progression and the *in silico* assignment of homologous human genes associated with breast cancer to the rearrangement breakpoint regions.

The analysis of the gene amplification and expression levels revealed *Erbb2* overexpression in HH-16 cl.4, 715 but not in its sister cell line HH-16 cl.2/1. This result, associated with the fact that these cell lines shared a similar initial genetic environment, highlights their potential as models for studying Erbb2-associated mechanisms and as experimental tools to assist in the generation of new biotherapies.

All the described work highlights the karyotype restructuring occurred in Rodentia species, showing that genomes evolution involves multifactor events that are parallel between species evolution and tumor progression.

2. FUTURE PERSPECTIVES

Despite all the studies focusing in the evolutionary events that shaped genomes during evolution, all the mechanisms involved and sequences implicated are far from being discovered. Rodents revealed to be excellent models for studies involving this subject. In order to pursuit the knowledge regarding the effect of karyotypic changes in species evolution, different research lines can be suggested:

- Considering that Rodentia order is the largest group within mammals, there are still many more species that need to be studied by comparative chromosomics. The many more comparative studies, the more unveiled will be the events that took place during the rapid rodents' evolution, the more accurate will be the ancestral karyotypes delineated and the better understood will be the dynamics of genomes evolution;

- Chromosome rearrangements shape the genomes not only during species evolution but also during tumor progression. Understanding the dynamic nature of the tumor genome will be the key to understand the mechanism of genetic heterogeneity and population diversity, which is the genetic basis for tumor formation.

- Once constitutive heterochromatin and its constituents, the repetitive sequences, have proven to play a significant role in genomes' architecture, it will be important to continue isolating and characterizing more sequences. Particularly satDNAs have revealed diverse evolution patterns showing that many relevant features are yet to be fully comprehended. Moreover the satDNa family isolated, PMSat, reveals to be promising in future functional studies (e.g. at the transcription level). These will allow to better understanding the evolution and dynamics of repetitive sequences and their implication in karyotype restructuring during evolution and cancer;

- There were found regions in the species and cancer genomes analysed more prone to the occurrence of rearrangements. These fragile sites need to be identified and the nature of the underlying sequences studied to better understand their fragility and their specific location in the genome.

- The development of capable *in vitro* models is of crucial importance to the study of cancer and, consequently, to the development of new therapeutics. This can be achieved by the full cytogenetic/genetic characterization of tumor cell lines already available commercially.

Taking this work as departure point, this and other interesting research lines can be followed. After all, science is a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions.

3. REFERENCES

Adega F, Chaves R, Guedes-Pinto H (2008) Suiformes orthologous orthogous satellite DNAs as a hallmark of Pecari tajacu and Tayassu pecari (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(1-2):12-20.

Albertson DG (2006) Gene amplification in cancer. Trends Genet22(8):447-55.

Brown JD, Mitchell SE, O'Neill RJ (2011) Making a long story short: noncoding RNAs and chromosome change. Heredity 108(1):42-9.

Bruvo B, Pons J, Ugarković Đ, Juan C, Petitpierre E, Plohl M (2003) Evolution of low-copy number and major satellite DNA sequences coexisting in two Pimelia species-groups (Coleoptera). Gene 312: 85–94.

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.

Cohen S, Segal D (2009) Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. Cytogenet Genome Res. 124:327–338.

Committee for standardization of chromosomes of Peromyscus (1977) Standard katyotype of the deer mice, Peromyscus (Rodentia). Cytogenet Cell Genet 19: 38–43.

Das PM, Singal R (2004) DNA methylation and cancer. J Clin Oncol 22(22):4632-42.

Deaven LL, Vidal-Rioja L, Jett JH, Hsu TC (1977) Chromosomes of Peromyscus (Rodentia, Cricetidae). VI. The genomic size. Cytogenet Cell Genet 19:241–249.

Edgecombe GD, Giribet G, Dunn CW, Hejnol A, Kristensen RM, Neves RC, Rouse GW, Worsaae K, Sørensen MV (2011) Higher-level metazoan relationships: recent progress and remaining questions. Org Divers Evol 11(2):151-172.

Ellingsen A, Slamovits CH, Rossi MS (2007) Sequence evolution of the major satellite DNA of the genus Ctenomys (Octodontidae, Rodentia). Gene 392(1-2):283-90.

Engelbrecht A, Dobigny G, Robinson TJ (2006) Further insights into the ancestral murine karyotype: the contribution of the Otomys–Mus comparison using chromosome painting. Cytogenet Genome Res 112(1–2):126–130.

Froenicke L, Lyons AL (2008) Hotspots of mammalian chromosome evolution. Chichester: John Wiley & Sons.

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

Gamperl R, Vistorin G, Rosenkranz W (1976) A comparative analysis of the karyotypes of Cricetus cricetus and Cricetulus griseus. Chromosoma 55: 259-265.

Garagna S, Marziliano N, Zuccotti M, Searle JB., Capanna E, and Redi CA (2001) Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. Proc Natl Acad Sci USA 98:171–175.

Graphodatsky AS, Trifonov VA, Stanyon R (2011) The genome diversity and karyotype evolution of mammals. Mol Cytogenet 4:22.

Greenbaum IF, Baker RJ (1978) Determination of the primitive karyotype for Peromyscus . J Mammal 59:820–834.

Hamm CA, Xie H, Costa FF, Vanin EF, Seftor EA, Sredni ST, Bischof J, Wang D, Bonaldo MF, Hendrix MJ, Soares MB (2009) Global Demethylation of Rat Chondrosarcoma Cells after Treatment with 5-Aza-29-Deoxycytidine Results in Increased Tumorigenicity. PLoS ONE 4(12): e8340.

Iannuzzi L, Di Berardino D, Gustavsson I, Ferrara L, Di Meo GP (1987) Centromeric loss in translocation of centric fusion type in cattle and water buffalo. Hereditas. 106(1):73-81.

John, B (1988) The biology of heterochromatin. In: Verma, RS, editors. Heterochromatin: Molecular and structural Aspects. Cambridge: Cambridge University Press. pp. 1–128.

Lee MR, Martin LK (1980) Mastomys (=Praomys) natalensis is not a Rattus (Mammalia: Rodentia): karyological evidence. Cytogenetic Cell Genet 28(1–2):95–103.

Li YC, Cheng YM, Hsieh LJ, Ryder OA, Yang F, Liao SJ, Hsiao KM, Tsai FJ, Tsai CH, Lin CC(2005) Karyotypic evolution of a novel cervid satellite DNA family isolated by microdissection from the Indian muntjac Y-chromosome. Chromosoma 114:28-38.

Li YC, Lee C, Sanoudou D, Hseu TH, Li SY, Lin CC (2000) Interstitial colocalization of two cervid satellite DNAs involved in the genesis of the Indian muntjac karyotype. Chromosome Res 8(5):363-73.

Ma Y, Lespagnard L, Durbecq V, Paesmans M, Desmedt C, Gomez-Galdon M, Veys I, Cardoso F, Sotiriou C, Di Leo A, Piccart MJ, Larsimont D (2005) Polysomy 17 HER-2/neu status elaboration in breast cancer: effect on daily practice. Clin Cancer Res 11(12): 4393–4399.

Macas J, Koblizkova A, Navratilova A, Neumann P (2009) Hypervariable 3' UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. Gene 448:198–206.

Martin S A, Hewish M, Lord CJ, Ashworth A (2010) Genomic instability and the selection of treatments for cancer. J Pathol 220: 281–289.

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'brien SJ, Pevzner PA, Lewin HA (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309:613–617.

Murphy WJ, Pevzner PA, O'Brien SJ (2004) Mammalian phylogenomics comes of age. Trends Genet 20(12): 631-639.

Navratilova A, Koblizkova A, Macas J (2008) Survey of extrachromosomal circular DNA derived from plant satellite repeats. BMC Plant Biol. 8:90.

Nijman IJ and Lenstra JA (2001) Mutation and recombination in cattle satellite DNA: A feedback model for the evolution of satellite DNA repeats. J Mol Evol 52(4):361-71.

Olaharski AJ, Sotelo R, Solorza-Luna G, Gonsebatt ME, Guzman P, Mohar A, Eastmond DA (2006) Tetraploidy and chromosomal instability are early events during cervical carcinogenesis. Carcinogenesis 27(2): 337–743.

Palomeque T, Antonio Carillo J, Munõz-López M, Lorite P (2006) Detection of a mariner-like element and a miniature invertedrepeat transposable element (MITE) associated with the heterochromatin from ants of the genus Messor and their possible involvement for satellite DNA evolution. Gene 371:194–205.

Pathak S, Hsu TC, Arrighi FE (1973) Chromosomes of Peromyscus (Rodentia, Cricetidae). IV. The role of heterochromatin in karyotypic evolution. Cytogenet Cell Genet 12: 315–326.

Pathak S, Hsu TC, Arrighi FE (1973) Chromosomes of Peromyscus (Rodentia, Cricetidae). IV. The role of heterochromatin in karyotypic evolution. Cytogenet Cell Genet 12:315–326.

Pauciullo A, Kubickova S, Cernohorska H, Petrova K, di Berardino D, Ramunno L, Rubes J (2006) Isolation and physical localization of new chromosome-specific centromeric repeats in farm animals. Veterinarni Medicina 51 224–231.

Peng Q, Pevzner PA, Tesler G (2006) The fragile breakage versus random breakage models of chromosome evolution. PLoS Comput Biol 2(2):e14.

Pevzner P, Tesler G (2003) Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. Genome Res13:37-45.

Plohl M (2010) Those mysterious sequences of satellite DNAs. Periodicum Biologorum 112(4):403–410.

Plohl M, Luchetti A, Mestrović N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene 409(1-2):72-82.

Robbins LW, Baker RJ (1981) An assessment of the nature of chromosomal rearrangements in 18 species of Peromyscus (Rodentia: Cricetidae). Cytogenet Cell Genet 31:194–202.

Rogers DS, Greenbaum IF, Gunn SJ, Engstrom MD (1984). Cytosystematic value of chromosomal inversion data in the genus Peromyscus (Rodentia, Cricetidae). J Mammal 65: 457–465.

Romanenko SA, Perelman PL, Serdukova NA, Trifonov VA, Biltueva LS, Wang J, Li T, Nie W, O'Brien PC, Volobouev VT, Stanyon R, Ferguson-Smith MA, Yang F, Graphodatsky AS (2006) Reciprocal chromosome painting between three laboratory rodent species.Rodents Mamm Genome 17(12):1183-1192.

Romanenko SA, Perelman PL, Trifonov VA, Graphodatsky AS (2012) Chromosomal evolution in Rodentia. Heredity (Edinb) 108(1):4-16.

Romanenko SA, Volobouev VT, Perelman PL, Lebedev VS, Serdukova NA, Trifonov VA, Biltueva LS, Nie W, O'Brien PC, Bulatova NSh, Ferguson-Smith MA, Yang F, Graphodatsky AS (2007) Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroidea, Rodentia) inferred from chromosomal painting and banding comparison. Chromosome Res 15:283-297.

Ruiz-Herrera A, Castresana J, Robinson TJ (2006) Is mammalian chromosomal evolution driven by regions of genome fragility? Genome Biol 7(12):R115.

Ruiz-Herrera A, Farré M, Robinson TJ (2012) Molecular cytogenetic and genomic insights into chromosomal evolution. Heredity108(1):28-36.

Shi L, Ye Y, Duan X (1980) Comparative cytogenetic studies on the red muntjac, Chinese muntjac and their F1 hybrids. Cytogenet Cell Genet 26:22–27.

Slamovits CH, Cook JA, Lessa EP, Rossi MS (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(9):1708-19.

Slamovits CH, Rossi M S (2002) Satellite DNA: agent of chromosomal evolution in mammals. A review. J. Neotrop. Mammal 9(2):297-308.

Stanyon R, Yang F, Morescalchi AM, Galleni L (2004) Chromosome painting in the long-tailed field mouse provides insights into the ancestral murid karyotype. Cytogenet Genome Res 105(2–4):406–411.

Steffen M, Scherdin U, Duvigneau C, Holzel F (1988) Glucocorticoid-induced alterations of morphology and growth of fibrosarcoma cells derived from 7,12-dimethylbenz(a)anthracene rat mammary tumor. Cancer Res 48: 212–7218.

Steppan S, Adkins R, Anderson J (2004) Phylogeny and Divergence-Date Estimates of Rapid Radiations in Muroid Rodents Based on Multiple Nuclear Genes. Syst Biol 53:533–553.

Thompson SL, Compton DA (2011) Chromosomes and cancer cells. Chromosome Res 19(3): 433–44.

Tysnes BB (2010) Tumor-initiating and -propagating cells: cells that we would like to identify and control. Neoplasia 12(7):506-15.

Veyrunes F, Dobigny G, Yang F, O'Brien PC, Catalan J, Robinson TJ, Britton-Davidian J. (2006) Phylogenomics of the genus Mus (Rodentia; Muridae): extensive genome repatterning is not restricted to the house mouse. Proc R Soc B 273(1604):2925–2934.

Volobouev VT, Gallardo MH, Graphodatsky AS (2006) Order Rodentia In: O'Brien SJ, Menninger JC, Nash WG. Atlas of mammalian chromosomes. Hoboken: John Wiley & Sons, Inc. pp 173-176.

Walsh JB (1987) Persistence of tandem arrays: implications for satellite DNA and simple-sequence DNAs. Genetics 115:553-567.

Wang S, Hossein Saboorian M, Frenkel EP, Haley BB, Siddiqui MT, Gokaslan S, Hynan L, Ashfaq R (2002) Aneusomy 17 in breast cancer: its role in ERBB2 protein expression and implication for clinical assessment of ERBB2 status. Mod Pathol 15: 137–145.

Wichman HA, Payne CT, Ryder OA, Hamilton MJ, Maltbie M, Baker RJ (1991) Genomic distribution of heterochromatic sequences in equids: implications to rapid chromosomal evolution. J Hered 82(5):369-77.