Universidade de Trás-os-Montes e Alto Douro

Repetitive DNA Sequences in Rodentia Genomes

Its involvement in chromosome architecture reshuffling and in genome functionality

Tese de Doutoramento em Genética Molecular Comparativa e Tecnológica

Ana Isabel do Paço Teixeira

Orientadora: Professora Doutora Raquel Chaves

Co-Orientadora: Doutora Filomena Adega

Vila Real, 2014

Universidade de Trás-os-Montes e Alto Douro

Repetitive DNA Sequences in Rodentia Genomes

Its involvement in chromosome architecture reshuffling and in genome functionality

Tese de Doutoramento em Genética Molecular Comparativa e Tecnológica

Ana Isabel do Paço Teixeira

Orientadora: Professora Doutora Raquel Chaves

Co-Orientadora: Doutora Filomena Adega

Composição do Júri:

Luis Herculano Melo de Carvalho

Isabel Marques Carreira

Raquel Maria Garcia dos Santos Chaves

José Eduardo Lima Brito

Agostinho Antunes

Vila Real, 2014

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 Operacional Potencial Humano – POPH", co-financiado pelo Fundo Social Europeu (POPH/FSE) e por fundos nacionais (POPH-QREN).

“A ciência será sempre uma busca e jamais uma descoberta. É uma viagem, nunca uma chegada”.

Karl Popper

À minha família

AGRADECIMENTOS

Ao terminar esta etapa do meu percurso académico, olho para trás e vejo o quanto aprendi. Não adquiri apenas conhecimento! As pessoas com quem convivi e estabeleci laços de amizade foram imprimindo em mim novas formas de pensar e de ver o mundo. A pessoa que hoje sou não é mais que a soma de tudo o que vivi e consegui aprender. Desta forma, não posso deixar de prestar o meu mais profundo agradecimento a todos aqueles que contribuíram para a minha formação como pessoa, e que surpreendentemente não têm real consciência de como foram agentes ativos neste meu processo de crescimento.

À UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO, na pessoa do magnífico Reitor Professor Doutor ***António Augusto Fontainhas Fernandes***, instituição que me acolheu sempre durante a minha formação Académica, que se iniciou nos finais de 1998 ao ingressar na Licenciatura de Engenharia Zootécnica, continuando em 2004, com a realização de um curso em Genética Molecular Comparativa e Tecnológica (Pós-Graduação/Mestrado/Doutoramento). Olhando para trás, cerca de 35% da minha vida está relacionada com esta Universidade e por este motivo é imperativo agradecer-lhe. Sinto realmente que estou em casa na UTAD! Irei para sempre lembrar-me de todos os espaços verdes, locais encantados e misteriosos, onde consegui sentir paz e podia sempre recorrer quando precisava de calma. Irei recordar o cantar dos pássaros, principal razão pela qual ia tantas vezes ao nascer do sol para a Universidade, do orvalho ou a geada nas folhas, dos esquilos a saltar de árvore em árvore como se tivessem ventosas nas patas, da cidade de toupeiras ao pé do lago e do coaxar das rãs pela tardinha nos dias quentes. Mas principalmente do Pedrinhas, o meu local preferido, onde ficava maravilhada com as flores dos abrunheiros dos jardins ou das magnólias. Adorava o mês de Maio no Pedrinhas, com os seus dias longos e quentes, existindo aqui o conforto na sombra das grandes árvores sempre que o sol queimava. Mas também a sua igreja à noite, nos dias curtos de Dezembro, envolta em nevoeiro, sobressaindo as luzes dos candeeiros que a iluminavam. Irei realmente sentir saudades……

À ESCOLA DE CIÊNCIAS DA VIDA E DO AMBIENTE, na pessoa do Professor Doutor ***António Luis Herculano Melo de Carvalho***, instituição que me acolheu durante o período que realizei trabalho de investigação referente ao meu Mestrado/Doutoramento.

À COMISSÃO PERMANENTE DO CONSELHO CIENTÍFICO DA ESCOLA DE CIÊNCIAS DA VIDA E DO AMBIENTE, na pessoa do seu atual Presidente, Professor Doutor Victor Machado Reis, pela aceitação da intenção de Doutoramento.

AO CENTRO DE GENÓMICA E BIOTECNOLOGIA DESTA UNIVERSIDADE, na pessoa da Professora Doutora ***Arlete Mendes Faia***, expresso o meu profundo agradecimento pela disponibilização de meios para a realização desta tese.

ÀS COMISSÕES DE CURSO EM GENÉTICA MOLECULAR COMPARATIVA E TECNOLÓGICA, atualmente na pessoa da Professora Doutora ***Paula Lopes***, pela disponibilidade e prontidão que sempre demonstraram na resolução de qualquer problema que surgisse.

À PROFESSORA DOUTORA RAQUEL CHAVES, minha orientadora, por me ter dado o possibilidade de participar na sua linha de investigação e com isso ter aprendido tudo o que sei hoje. Ter podido dar aulas e orientar estágios. Para além dos conhecimentos científicos que adquiri durante estes últimos sete anos, aprendi também consigo a ter os pés sempre assentes na terra. Agradeço-lhe sempre que me chamou à razão! Por vezes, estava tão envolta no trabalho que não me distanciava o suficiente para perceber com mais clareza as minhas falhas. E esteve sempre lá para me chamar a atenção. Irei sempre recordar a sua lucidez e inteligência. Foi a pessoa que mais contribuiu para o meu crescimento, pois os seus ensinamentos estão muito para além da minha vida académica.

À PROFESSORA DOUTORA FILOMENA ADEGA, minha co-orientadora e mestre, dedico um agradecimento muito especial. Em todos os momentos decisivos esteve sempre comigo! Sempre que não sabia como continuar era a ela que recorria. Agradeço-lhe pela alegria e pela sua preocupação com o meu bem-estar. Agradeço-lhe sobretudo também pela exigência, pois sem ela não teria aprendido tanto. Olhando para tudo que passamos juntas, não lhe posso só chamar co-orientadora, mas amiga, uma grande amiga! Pois sempre foi verdadeira comigo, disse-me sempre o que pensava, mesmo que eu não o quisesse ouvir! E isso mostra o quanto se preocupa comigo!

Ao Doutor MIROSLAV PLOHL e Doutora NEVENKA MEŠTROVIĆ, do Instituto Ruđer Bošković (Croácia, Zagreb) por me terem recebido no seu laboratório e me terem ensinado

tanto. Obrigada também pela amizade e preocupação com o meu bem-estar. Agradeço igualmente aos restantes membros da sua equipa, **Eva Šatović**, **Martina Pavlek** e **Brankica Mravinac** com quem tive o prazer de trabalhar no período de tempo que passei lá.

AOS MEUS COLEGAS DE LABORATÓRIO, amigos para a vida e espero que para sempre. Refiro particularmente a Ana Borges, Ana Vieira da Silva, Cláudia Baptista, João Coutinho, Jorge Pereira, Karine Bouilly, Sandra Louzada, Sara Santos e Susana Meles (ordem alfabética). Tudo o que diga nunca chegará para expressar a estas pessoas o quanto as admiro e me são necessárias. E por não ter palavras suficientemente amplas que abranjam o agradecimento que lhes gostaria de prestar, dedico-lhes este poema de Vinícius de Morais.

Tenho amigos que não sabem o
quanto são meus amigos.
Não percebem o amor que lhes
devoto e a absoluta
necessidade que tenho deles.
A amizade é um sentimento mais
nobre do que o amor,
eis que permite que o objeto dela
se divida em outros afetos,
enquanto o amor tem intrínseco o ciúme,
que não admite a rivalidade.
E eu poderia suportar,
embora não sem dor,
que tivessem morrido todos os
meus amores, mas enlouqueceria
se morressem todos os meus amigos!
Até mesmo aqueles que não percebem
o quanto são meus amigos e o quanto
a minha vida depende de suas existências ....

Vinícius de Morais (amigos)

Admiro particularmente a força da **Ana Borges**. Mesmo quando tudo lhe está a correr menos bem, consegue sempre incentivar e animar os outros. Defende arduamente quem ela considera amigo, e fico mesmo feliz por estar incluída nesse grupo.

Aprecio a sensibilidade da **Ana Vieira da Silva**. Até hoje, sem considerar a minha família, acho que foi a pessoa que me conheceu e compreendeu melhor. Foi com ela que descobri o gosto por viajar e conhecer novas culturas e modos de vida. Foi também com ela que vivi os momentos mais divertidos destes últimos oito anos. Esteve sempre do meu lado em tudo o que vivi de bom e mau. È uma pessoa particular, muito diferente de qualquer outra que conheço. Vê a bondade em toda a gente, o que por vezes a leva a desiludir-se.

Gosto da pessoa prática que é a **Cláudia Baptista**. Embora não conviva com ela todos os dias (o que lamento), segui seriamente todos os conselhos que ela me deu, principalmente no que diz respeito à escrita da tese e dos artigos. Penso que ela não tem consciência de quão seriamente ouvia o que ela me dizia.

Admiro a alegria e a autoconfiança do **João Coutinho**. Consegue sempre ter uma resposta para tudo! Sem se aperceber ajudou-me mil e uma vezes! Só o simples facto de me fazer rir tornava tudo muito mais fácil.

Para sempre vou lembrar-me do **Jorge Pereira** e da **Sandra Louzada**, meus grandes amigos. Adoro-os, são pessoas excecionais e confio neles mais do que em mim mesma. São realmente boas pessoas, em quem não consigo ver defeitos. Agradeço aos dois a ajuda que me deram em Inglaterra e particularmente ao Jorge a ajuda que também me deu na Croácia.

Agradeço à **Karine Bouilly** a amizade, a alegria e o apoio que sempre me deu quando precisava de tomar uma decisão. Irei para sempre recordar a sua companhia nas viagens que fizemos juntas. Foi mesmo divertido!

Admiro na **Sara Santos** a vontade altruísta que tem de ajudar os outros. Essa vontade é tão grande que por vezes se esquece dela própria. Tem sempre uma boa ideia para resolver qualquer problema, mas acho que ela não tem consciência do quanto me ajudou.

Agradeço à **Susana Meles** e **Ana Cristina Silva**, pela amizade e toda a disponibilidade que mostraram em ajudar-me. Com a Susana já vivi muitas experiências. Estivemos juntas desde o princípio e por isso lhe agradeço. À Ana Cristina, gostava de agradecer a alegria e o carinho que me transmitiu nos meus últimos dias no laboratório.

Ao **João Castro**, **Raquel Matoso** e **Carlota Conceição**, agradeço a simpatia e amizade. Estiveram comigo pouco tempo, mas tornaram a minha vida bem mais alegre.

Agradeço também à **Daniela Ferreira** e **Ana Escudeiro.** São a nova força do laboratório e por isso as reuni no mesmo grupo. Gostava de lhes agradecer por toda a ajuda que sempre me ofereceram.

AOS PROFESSORES DOUTORES JOSÉ EDUARDO LIMA-BRITO E ESTELA BASTOS, meus Professores na parte letiva deste curso, agradeço a simpatia e prontidão em ajudar sempre que necessário.

É impossível negar que na vida não passamos de simples folhas ao sabor do vento. Há um ano atrás, nunca suspeitaria que estaria agora a trabalhar na Universidade de Évora. Um novo

desafio nesta fase já por si só tão difícil do meu doutoramento. Por isso, não posso deixar de agradecer à **professora Solange Oliveira** pela confiança que depositou em mim e por me ter dado a oportunidade de trabalho que eu precisava. Agradeço também às doutoras **Ana Alexandre** e **Marta Laranjo**, por mostrarem confiança em mim e me incentivarem sempre. São realmente muito atenciosas comigo! Expresso também o meu agradecimento à dona **Gertrudes Mariano**, à professora **Rosário Félix**, ao professor **José Manuel Martins** e à doutora **Carla Varandas**, por serem sempre tão simpáticos para mim e terem ajudado na minha integração aquando da minha chegada a Évora. Mesmo muito obrigada!

Felizmente também, neste meu percurso por Évora fiz amigos a quem já devo muito! Pessoas que diminuem as minhas enormes saudades de casa, do Norte, de Vila Real!!! Que me fazem rir e dizer disparates, o que me faz tão bem! Refiro assim o **Fabrício Macedo, Fernando Eliziário, Clarisse Brígido, Rodrigo Silva**, **Mónica Portella** e **Wellma Pedra** (ordem alfabética). À Clarisse e ao Rodrigo, devo um agradecimento muito especial! São o meu apoio aqui, as pessoas que tem paciência de ouvir os meus problemas, que me tentam puxar para cima sempre que me sinto a bater no fundo! Fazem-me já muita falta e sem vocês tudo seria muito mais difícil. Muito, muito obrigada por tudo! Espero conseguir retribuir-lhes todo o esse carinho!

À MINHA FAMÍLIA, as pessoas que melhor me conhecem e que sempre me apoiaram durante toda a minha vida, não só nesta etapa, mas em todas as outras que fui superando. Agradeço o esforço que os meus pais, pessoas simples, filhos de lavradores, fizeram para que eu e a minha irmã estudássemos. Agradeço o estímulo, a confiança e o carinho que me deram todos os dias. Lamento no entanto, não poder dar-lhes a alegria de ter conseguido ter uma vida estável e por isso ficarem mais descansados em relação ao meu futuro. À minha irmã, e minha melhor amiga, agradeço o carinho, os conselhos, os estímulos, o incentivo, a confiança e a amizade. Agradeço também a ela e ao meu cunhado (Nuno) o privilégio de ser madrinha da Beatriz, a minha sobrinha. É uma menina feliz e espero sinceramente ser um bom exemplo para ela.

A todos, o meu muito obrigada, embora seja impossível pagar tudo quanto vos devo!

RESUMO

Uma das evidências mais claras revelada pelos projetos de sequenciação de genomas eucariotas, foi o seu elevado nível em sequências repetitivas de DNA, sendo a extraordinária variação no tamanho do genoma encontrada entre *taxa* atribuída à amplificação e eliminação diferencial destas famílias de sequências. No entanto, apesar da sua abundância, a(s) função(ões) que as sequências repetitivas desempenham nos genomas sempre esteve envolta em grande mistério, pois ao contrário dos genes, nunca foi atribuída a estas sequências a capacidade de codificar proteínas (pelo menos proteínas que não estejam envolvidas na sua própria replicação e integração no genoma, como é o caso dos Elementos Transponíveis). Por esta razão, estas sequências foram inicialmente designadas como "DNA lixo", às quais nenhuma função era atribuída. Atualmente, as sequências repetitivas ganharam a merecida atenção e são agora consideradas como uma fração essencial dos genomas eucariotas, sendo reconhecidas como elementos reguladores importantes e também sendo implicadas na ocorrência de rearranjos cromossómicos, tendo assim uma função importante na evolução de genomas. Este foi precisamente o objetivo principal deste trabalho, contribuir para a compreensão da importância que as sequências repetitivas têm na evolução de genomas eucariotas. Com este propósito, aqui foi analisada a fração repetitiva de cinco espécies de roedores Cricetidae/Muridae, que apresentam cariótipos muito distintos, sendo nomeadamente estudadas sequências em tandem e dispersas: sequências DNA satélite (SatDNAs), Sequências Teloméricas Intersticiais (ITS) e Retrotransposões LINE-1. Uma análise detalhada sobre a distribuição e a natureza molecular da Heterocromatina Constitutiva (HC) foi também realizada para estes genomas de roedores.

A integração de todos os dados obtidos permitiu entender como os cinco genomas estudados evoluíram, bem como reconstruir os rearranjos cromossómicas ocorridos, nos quais as sequências repetitivas estão inquestionavelmente implicadas. De facto, todos os resultados obtidos aqui convergem para esta mesma conclusão. Nomeadamente, foi observada uma forte associação entre a distribuição e heterogeneidade da HC com o percurso evolutivo seguido por estes cariótipos. Uma análise detalhada feita neste trabalho, para duas das espécies estudadas, incidindo sobre a localização de regiões de "breakpoint" evolutivo e HC, revelou uma elevada coincidência entre estas regiões. Outros trabalhos, também focados na evolução cromossómica das restantes três espécies em análise aqui, relataram uma relação semelhante entre regiões de "breakpoint" e HC. Portanto, as sequências repetitivas localizadas na

heterocromatina parecem estar intimamente envolvidas na ocorrência de rearranjos, quer promovendo diretamente reorganizações cromossómicas e/ou porque correspondem a regiões frágeis propensas à quebra. A análise de sequências repetitivas localizadas em regiões HC, especificamente satDNAs (sequências exclusivamente heterocromáticas), ITS (sequências localizadas principalmente nas regiões de HC) e LINE-1 (sequências frequentemente localizadas na HC), sugere realmente as sequências repetitivas como uma força motriz para a ocorrência de rearranjos cromossómicos. A natureza repetitiva das diferentes classes de sequências repetidas estudadas, por si só, favorece eventos de recombinação entre sequências homólogas em regiões não-homólogas, que podem culminar em rearranjos cromossómicos. No entanto, o papel na origem de reorganizações cromossómicas é particularmente sugerido para satDNAs, devido ao seu modo de evolução (evolução concertada), geralmente caracterizado por rápidas alterações na sua sequência de nucleótidos, variações no número de cópias e/ou movimentos intragenómicos, que resultam da ocorrência de diferentes eventos de recombinação (como "unequal crossing-over" ou "rolling circle replication/reinsertion"), podendo induzir quebras cromossómicas.

Adicionalmente, para além de sua importante função na reestruturação dos genomas, os resultados obtidos neste trabalho atribuem também outras funções às sequências repetitivas. Uma análise mais relacionada com a atividade transcricional de alguns satDNAs estudados, suporta o papel destas sequências em muitas outras funções, como no controlo da expressão génica, na remodelação da cromatina, na resposta celular ao stress e na função centromérica. As sequências LINE-1 têm também importantes funções no controlo da expressão génica, estando envolvidas no "imprinting" de genes e na inativação do cromossoma X. Assim, apesar de inicialmente consideradas sequências inertes, à luz de todos estes dados, é impossível negar que sequências repetitivas são cruciais para o bom funcionamento e evolução dos genomas eucariotas, destronando aos nossos olhos a importância dada no passado apenas a sequências codificantes. É agora realmente difícil de entender como estas sequências, tão abundantes nos genomas eucariotas, podem ter sido consideradas desnecessárias, só porque não lhes foi atribuída uma capacidade em codificar proteínas. Afinal, as sequências que codificam proteínas representam apenas uma pequena parte dos genomas (~ 1,5 % do genoma humano).

A presente tese resultou na elaboração de sete artigos que estão publicados, submetidos ou em preparação para submissão em revistas científicas internacionais indexadas.

PALAVRAS-CHAVE: Sequências Repetitivas de DNA, Evolução Cromossómica, Repetições em Tandem Repeats, Long Interspersed Nuclear Elements-1, Rodentia.

ABSTRACT

One of the clearest evidences that emerged from the eukaryotic genome sequencing projects was the high content in repetitive DNA sequences that these genomes harbour, being the extraordinary genome size variation found between *taxa* attributed to the differential amplification and deletion of these sequence families. However, despite its abundance, the role(s) that these sequences play in the genomes has always been shrouded in great mystery, as unlike genes, it was never assigned to them the ability to code proteins (at least proteins that are not involved in their own replication and genomic integration, as is the case of Transposable Elements). For this reason, these sequences were initially designated as "junk" DNA, with no function assigned. Presently, these sequences have won the deserved respect and are now regarded as a crucial fraction of eukaryotic genomes, recognized as important regulatory elements and also as being implicated in the occurrence of chromosomal rearrangements, with an important role in genome evolution. This was, precisely, the main goal of this work: to contribute to the understanding of the repetitive sequences significance in the evolution of eukaryotic genomes. For this purpose, it was analysed the repetitive genomic fraction of five Cricetidae/Muridae Rodentia species, with very distinct karyotypes, regarding tandem and dispersed repeats: Satellite DNAs (satDNAs), Interstitial Telomeric Sequences (ITSs) and LINE-1 Retrotransposons. A detailed analysis about the distribution and molecular nature of the Constitutive Heterochromatin (CH) of these rodent genomes was also performed.

The integration of all data allowed to understand how the five studied genomes evolved and to reconstruct the chromosomal evolutionary events elapsed, where the repetitive sequences were unquestionably involved. Indeed, all the results obtained here converge to this same conclusion. Namely, a strong association was observed between both the distribution and the level of CH heterogeneity with the evolutionary pathway that these karyotypes followed. In fact, for two of these species, a detailed analysis on the location of evolutionary breakpoint and CH regions revealed a very high coincidence between them. Other works focused on the evolution of the other three species, reported a similar relationship. Therefore, the repeats located in heterochromatin seem to be highly involved in the occurrence of chromosomal rearrangements, either by promoting directly chromosome reorganizations and/or because correspond to fragile regions prone to chromosome breakage. The analysis of the repeats located in CH regions performed here, namely satDNAs (exclusively

heterochromatic), ITSs (mainly located in CH regions) and LINE-1 (frequently located in CH), really suggest the repetitive sequences as a driving force in the occurrence of chromosomal rearrangements. The repetitive nature *per se* of the different classes of repeats studied favours recombinational events between homologous sequences in non-homologous regions, which may culminate in chromosomal restructurings. Nevertheless the role in the origin of chromosomal reorganizations is particularly proposed for satDNAs, due to its characteristic evolutionary mode (concerted evolution) generally marked by rapid sequence mutations, copy number variations and/or intragenomic movements, driven by different recombinational events (as unequal crossing-over or rolling circle replication/reinsertion), that may induce chromosome breakage.

Additionally, beyond its important function in genome restructuring, the data obtained in this work also suggest other roles to repetitive sequences. An analysis devoted to the transcriptional activity of some of the studied satDNAs supports the role of these sequences in many other functions, as in control of gene expression, chromatin remodelation, cellular response to stress and centromeric function. LINE-1 sequences as well have important functions in control of gene expression, acting in gene imprinting and in X-chromosome inactivation. Thereby, despite initially considered useless genomic elements, in the light of all this data, it is impossible to deny that repetitive sequences are crucial for proper functioning and evolution of eukaryotic genomes, dethroning to our view the importance given in the past just to the protein-coding sequences. It is really now difficult to understand how these sequences, so abundant in eukaryotic genomes, may have been considered unnecessary, just because a coding capacity was not reported. After all, the protein-coding sequences only account for a tiny part of genomes (~1,5% of the human genome).

The present thesis resulted in the elaboration of seven articles that are published, submitted or in preparation for submission to indexed international scientific journals.

KEY WORDS: Repetitive DNA Sequences, Chromosomal Evolution, Tandem Repeats, Long Interspersed Nuclear Elements-1, Rodentia.

GENERAL INDEX

INDEX OF FIGURES

LIST OF PUBLICATIONS

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

Article 1:

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.

Article 2:

Paço A, Adega F, Guedes-Pinto H and Chaves R (2009). The hidden heterochromatin: characterization in the Rodentia species *Cricetus cricetus, Peromyscus eremicus* (Cricetidae) and *Praomys tullbergi* (Muridae). *Genetics and Molecular Biology* 32 (1): 58-68.

Article 3:

Paço A, Chaves R, Vieira-da-Silva A and Adega F (2012). The involvement of repetitive sequences in the remodelling of karyotypes: the *Phodopus* genomes (Rodentia, Cricetidae. *Micron* 46: 27-34.

Article 4:

Paço A, Adega F, Chaves R (2014). Line-1 Retrotransposons: from "parasite" sequences to Functional Elements. *Submitted to Journal of Applied Genetics.*

Article 5:

Paço A, Adega F, Meštrović N, Plohl M, Chaves R (2014). High-resolution organization of repetitive DNA sequences in *Phodopus roborovskii* and *P. sungorus* (Cricetidae, Rodentia). *In preparation.*

Article 6:

Paço A, Adega F, Meštrović N, Ploh M, Chaves R (2014). Evolutionary story of a satellite DNA sequence from *Phodopus sungorus* (Rodentia, Cricetidae). *In preparation.*

Article 7:

Paço A, Adega F, Chaves R (2014). Quest for the functional significance of a satellite DNA sequence from *Peromyscus eremicus* (Cricetidae, Rodentia). *In preparation*

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

1- Paço A, Adega F, Guedes Pinto H, Volobouev V and Chaves R (2007). Constitutive heterochromatin characterization of the rodents *Peromyscus eremicus, Cricetus cricetus* and *Praomys tullbergi*. Chromosome Research 15(2): 41-42.

2- Meles S, Paço A, Filomena A, Guedes Pinto H and Chaves R (2008). Detailed Constitutive Heterochromatin Map for *Praomys tullbergi* (Rodentia, Muridae) karyotype. Chromosome Research 16: 1035.

3- Louzada S, Vieira-da-Silva A, Paço A, Svatva K, 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.

4- Paço A, Adega F, Guedes Pinto H, Volobouev V and Chaves R (2008). Molecular analysis and Physical distribution of LINE-1 sequences in three rodent species, *Cricetus cricetus, Peromyscus eremicus* (Cricetidae family) and *Praomys tullbergi* (Muridae family). Chromosome Research 16: 1047.

5- Paço A, Adega F, Guedes-Pinto H and Chaves R (2009). Repetitive sequences: a possible source of genome Remodelling. Chromosome Research 17 (Suppl 1): 166.

6- Paço A, Meštrovic N, Adega F, Plohl M and Chaves R (2010). Different organization patterns of a satellite DNA sequence in closely related species, *Phodopus sungorus* and *Peromyscus eremicus* (Rodentia, Cricetidae). Chromosome Research 18: 727.

7- Paço A, Adega F, Meštrovic N, Silva A, Plohl M and Chaves R (2011). Construction of repetitive DNA libraries from the genomes of two hamster species, *Phodopus sungorus* and *Phodopus roborovskii* (Cricetinae). Chromosome Research 19 (Suppl 1): S195.

From the obtained results are also deposited several DNA sequences in Genbank:

1- *Cricetus cricetus* LINE-1 partial sequence (Accession: **HQ386006**).

2- *Peromyscus eremicus* LINE-1 partial sequence (Accession: **HQ386007**).

3- *Praomys tullbergi* LINE-1 partial sequence (Accession: **HQ386008**).

4- *Phodopus roborovskii* Satellite DNA (PROsat) complete monomer variant clone 1 (Accession: **KJ649144**).

5- *Phodopus roborovskii* Satellite DNA (PROsat) incomplete monomer clone 2 (Accession: **KJ649145**).

6- *Phodopus* Repetitive DNA (PsatDNA) clone 1 (Accession: **KJ649146**).

7- *Phodopus sungorus* Chromosome 1 Repetitive DNA (PSUchr1sat) clone 1 (Accession: **KJ649147**).

8- *Phodopus sungorus* Centromeric Repetitive DNA (PSUcentSat) clone 1 (Accession: **KJ649148**).

ABBREVIATIONS

A, C, T, G Adenine, cytosine, thymine, guanine

bp Base pair

CCP Comparative Chromosome Painting

CENP-A Centromere protein A

CENP-B Centromere protein B

CH Constitutive Heterochromatin

cDNA Complementary DNA

DNA Deoxyribonucleic acid

DIRs *Dictyostelium* intermediate Repeat sequences

EN Endonuclease

FA-SAT *Felis catus* satellite DNA

FISH Fluorescent in situ hybridization

het-ITSs Heterochromatic Interstitial Telomeric Sequences

HIV-1 Human Immunodeficiency Virus

HOR Higher-order repeat

HP1 Heterochromatin protein 1

HSF1 Heat Shock Factor 1

ITSs Interstitial Telomeric Sequences

Kb Kilo bases

LINEs Long Interspersed Nuclear elements

LINE-1 Long Interspersed Nuclear elements 1

My Million years

Mb Mega bases

mRNA Messenger RNA

RE Restriction Endonuclease

RT Reverse transcriptase

ORFs Open Reading Frame

RNA Ribonucleic acid

rRNA Ribosomal RNA

satDNA Satellite DNA

siRNA Small interfering RNAs

s-ITSs Short Interstitial Telomeric Sequences

RISC RNA-induced silencing complex

RIST RNA-induced transcriptional silencing complex

RNP Ribonucleoprotein Particle

UTR's Untranslated Regions

TE's Transposable Elements

LTR Long Terminal Repeats

Non-LTR Non Long Terminal Repeats

SIV Simian Immunodeficiency Virus

TPRT Target Primed Reverse Transcription

TSDs Target Site Duplications

ABBREVIATIONS OF SPECIES' NAME

CCR *Cricetus cricetus*

CGR *Cricetulus griseus*

MAU *Mesocricetus auratus*

MMU *Mus musculus*

PER *Peromyscus eremicus*

PRO *Phodopus roborovskii*

PSU *Phodopus sungorus*

PTU *Praomys tullbergi*

RNO *Rattus norvegicus*

Chapter I

Introduction

CHAPTER I. INTRODUCTION

In eukaryotic cells, genomic DNA is folded with histone and non-histone proteins forming the chromatin (Luger et al. 1997). During cell division, the condensed chromatin fibers are nominated as chromosomes, being this term firstly introduced in 1888 by Waldeyer (Zacharias 2001). The entire chromosome set, properly organized, of a species is known as karyotype. Organisms from different species present dissimilar karyotypes and this variation is due to numerical and structural chromosomal reorganizations (Sumner 2003). Some of the most spectacular karyotype differences occur among rodents, even among populations of the same species. The majority of the house mouse individuals have a karyotype consisting of 40 acrocentric chromosomes, however some populations, often in isolated places such as in Alpine valleys, presents a low chromosomal number as 2n = 22, by the formation of metacentric chromosomes from two acrocentric (Nachman and Searle 1995).

An important feature of the eukaryotic genomes is its high content in repetitive DNA sequences, whose functions are not yet completely understood. Nevertheless, several are the roles proposed to this genomic fraction, as its involvement in chromosomal rearrangements and consequently, in genome evolution (e.g. Moran and Gilbert 2002, Adega et al. 2009). This was, precisely, the main goal of this work; contribute to the understanding of the repetitive sequences significance in the evolution of genomes. For this purpose, several families of repeats were analysed in detail in various Rodentia species with very distinct karyotypes, namely Satellite DNAs, Interstitial Telomeric Sequences and LINE-1 Retrotransposons. A detailed analysis about the distribution and molecular nature of the Constitutive Heterochromatin for these rodent genomes was also performed.

To introduce the theme, a literature review embracing the current knowledge about the repetitive fraction in the eukaryotic genome, regarding its molecular characteristics, evolution mode and possible functions was undertaken and is presented in this chapter. The Results and Discussion chapter will be presented as individual papers being some of them already published, one submitted and others in preparation for submission. A general discussion is presented at the end, integrating and correlating all the data achieved.

I.1- Chromatin organization

In the eukaryotic genomes chromatin can be categorized as euchromatin or heterochromatin. Heitz (1928) was the first proposing this classification, based in the observation that throughout the cell cycle a chromatin fraction alters their degree of condensation (euchromatin), while the other fraction remains highly condensed (heterochromatin). Later, other features distinguishing these genomic compartments were identified. The euchromatin is enriched in unique coding sequences, while the heterochromatic fraction is referred to as gene poor, being mainly composed by repetitive DNA sequences such as clusters of satellite DNA and transposable elements (described in the next sections of this chapter) (Grewal and Jia 2007, Pezer and Ugarković 2008a). Besides, euchromatin and heterochromatin present characteristic histone-modification marks. The first is characterized by histone H4 acetylation and methylation of histone H3 at lysine 4, while the second is distinguished by the hypoacetylation and methylation of histone H3 at lysine 9 (Nakayama et al. 2001). The heterochromatin can also be considered as constitutive or facultative. The Constitutive Heterochromatin (CH) can occur as large blocks or discrete bands in different regions of the chromosome (detected by C-banding), however it is mainly found in large blocks near the centromere (Corradini et al. 2007, Probst and Almouzni 2008). Facultative heterochromatin is usually found at developmentally regulated *loci* where the chromatin state is changed in response to cellular signals and gene activity (Grewal and Jia 2007, Enukashvily and Ponomartsev 2013).

Several are the works reporting the ability of heterochromatin to propagate under epigenetic control to nearby DNA sequences, where a repression of these sequences occur in a process known as silencing (reviewed in Grewal and Jia 2007). This epigenetic control requires the methylation of histone H3 at lysine 9 and the subsequent association of chromodomain proteins, such as heterochromatin protein HP1, allowing histone modifications and chromatin assembly (Eymery et al. 2009). This process reduces the accessibility of the involved DNA sequences for being transcribed or for recombination (reviewed in Grewal and Jia 2007). Moreover, it was also reported that the heterochromatic surrounding of centromeres is necessary for their function, ensuring sister chromatid cohesion and kinetochore formation (e.g. Bernard et al. 2001, Nonaka et al. 2002, Obuse et al. 2004, Hall et al. 2012). The association between histone H3 methylated at lysine 9 and chromodomain proteins (e.g. HP1), allows the recruitment of cohesin complex that promote sister chromatid cohesion at

pericentromeric regions. HP1 facilitates as well the recruitment of kinetochore proteins such as Mis12 (reviewed in Grewal and Jia 2007). Heterochromatic rich regions are also considered as "hotspots" for the occurrence of structural chromosome rearrangements (John 1988, Chaves et al. 2004, Adega et al. 2009), which is mainly justified by the high molecular dynamics of the repetitive DNA sequences located in these regions (Adega et al. 2009). Therefore, the analysis of the heterochromatic fraction in a genome, regarding specifically its chromosome location and molecular heterogeneity (evaluation of repetitive sequences diversity located in heterochromatin), can be extremely important to understand the impact of repetitive sequences in genome evolution (e.g. Chaves et al. 2004, Adega et al. 2009).

I.1.1- Repetitive DNA sequences

Repetitive DNA sequences correspond to DNA elements that are present in multiple copies in a genome (e.g. Jurka et al. 2007). The accumulation of data from the 214 eukaryotic nuclear genomes completely sequenced up to now (data reviewed in January 2014, in http://www.ncbi.nlm.nih.gov/genome/browse/), shows that the repetitive sequences represent a significant portion of these genomes. Moreover, the differential amplification and deletion of their various families contribute significantly to the extraordinary genome size variation found between taxa (Petrov 2001, Boulesteix et al. 2006, Pritham 2009, Venner et al. 2009, Devos 2010, Sun et al. 2011), from 0.02 to 130 Gb within the Animal kingdown (Gregory 2011).

Based on the genomic organization of their copies, repetitive DNA sequences are classified as either tandem or dispersed repeats (reviewed in Slamovits and Rossi 2002, Jurka et al. 2007, Richard et al. 2008), as it is shown in figure 1. The first main type of repeats is organized in arrays of copies that can occur at few or in many different chromosomal locations. The second class, dispersed repeats, consist of sequences whose copies are highly scattered through the genome (Strachan and Read 2004). Each of these two classes of repetitive sequences can be itself divided into several subclasses (Figure 1). Here will be focused the largest of its subclasses, excluding the genic repetitive DNA sequences families. Within tandem repeats, three distinct subtypes exhibiting different properties, genomic distributions and length of copies and arrays are recognized: satellites, minisatellites and microsatellites (reviewed in Slamovits and Rossi 2002, Strachan and Read 2004). The dispersed repeats are represented by various types of transposable elements, presenting the

ability to transpose within the genome (Kazazian 2004, Jurka et al. 2007) as can be observed in figure 1.

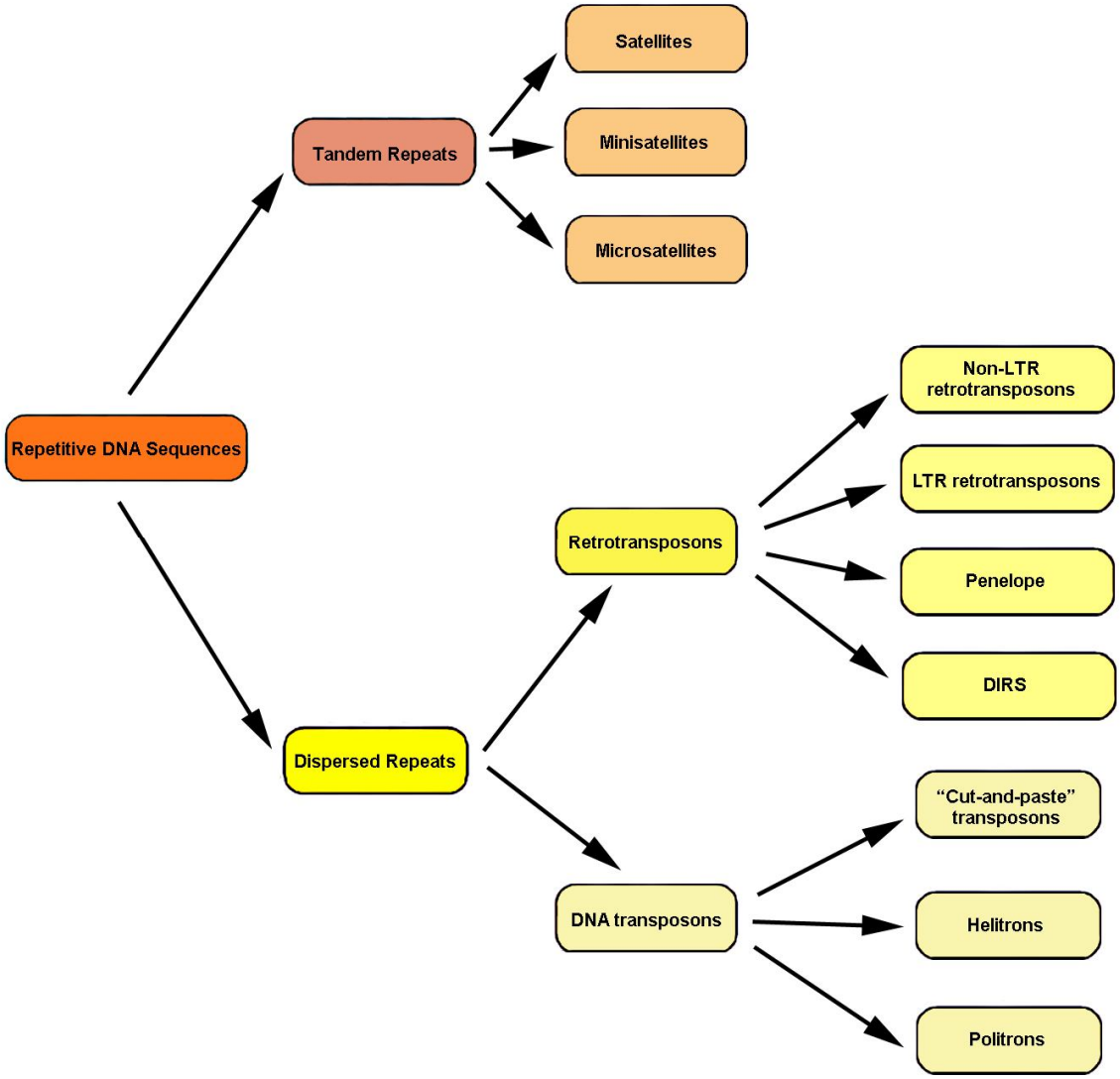


Figure 1- Repetitive DNA sequences in eukaryotic genomes. This schematization collects the information from several works (Slamovits and Rossi 2002, Feschotte and Pritham 2007, Jurka et al. 2007, Kapitonov and Jurka 2008, Richard et al. 2008, Kapitonov et al. 2009, Rebollo et al. 2012). Here there are only represented the largest subclasses of tandem and dispersed repeats, not including the tandem paralogue genes, ribosomal genes (tandem organization), retropseudogenes, transfer RNA genes or dispersed paralogue genes (dispersed organization).

I.1.1.1- Tandem repeats

Structurally, tandem repeats are formed by a sequential arrangement of copies, positioned one after the other (arrays). Two possible repeat orientations can be found in the genomes, head-to-tail repeats (direct repeats) and head-to-head repeats (inverted repeats) (Richard et al. 2008). Comparing the three subclasses of tandem repeats, the satellite DNA sequences have by far the largest arrays of copies, and this is the main feature that allows its differentiation from the micro and minisatellites (Plohl et al. 2008).

I.1.1.1.1- Satellite DNA

Satellite DNA sequences (satDNAs) correspond to highly tandemly repeated sequences that can be present in several millions of copies in a genome, organized into long arrays in the heterochromatic regions (Charlesworth et al. 1994). Indeed, satDNAs are the main constituent of Constitutive Heterochromatin (Ugarković and Plohl 2002, Chaves et al. 2004), being preferentially found in and around centromeres, but also assuming interstitial and terminal **positions** (reviewed in Adega et al. 2009). Historically, the first **isolations** of satDNAs were achieved by experiments with gradient centrifugation originating satellite bands, what justifies its name (Szybalski 1968). Currently, the more widespread strategy for the isolation of these repeats is the digestion of genomic DNA with restriction endonucleases, followed by sequence analysis of prominent cloned bands. This approach continued to be used even after the burst of large scale genome sequencing projects, since the high repetitive nature of satellites imposes serious limitations in assembling tandemly repeated motifs into large contigs, remaining these sequences underrepresented in outputs of genome sequencing projects (reviewed in Plohl 2010).

SatDNA repeat units (monomers) show a great **variation in size**, ranging from five nucleotides in human satellite III as well in some *Drosophila* satellites (Borstnik et al. 1994), up to several hundreds of base pairs (e.g. Modi 1993). However, it is considered that the preferential monomer length is 140-180 bp and 300-360 bp, since many satDNAs monomers in both plants and animals present these lengths (Schmidt and Heslop-Harrison 1998, Henikoff et al. 2001). In the same genome, several unrelated satDNA families presenting characteristic monomer extents and sequences can coexist, sharing only two common features, tandem arrangement of monomer repeats and heterochromatic localization (e.g.

Meštrović et al. 1998, Plohl 2010). The satellite DNA **contribution to the total genomic content** varies significantly among species, exceeding sometimes 50% of the total genomic DNA (Elder and Turner 1995, Schmidt and Heslop-Harrison 1998), and consequently these are involved in the enormous variation of genome size in eukaryotes (Doolittle and Sapienza 1980, Cavalier-Smith 1985, Gregory et al. 2007), as referred previously.

It is generally accepted that satDNAs follow the principles of **concerted evolution** (e.g. Palomeque and Lorite 2008, Plohl et al. 2008, Meštrović et al. 2013), indicating a non-independent evolution of satellite monomers within a genome, that results in an **intraspecific homogenization** of satDNAs (Elder and Turner 1995). This evolution mode is promoted by molecular drive, a complex process in which monomer mutations are spread or eliminated in the satellite arrays leading to repeats homogeneity (Figure 2), and concomitantly to its fixation in the individuals of a population (reviewed in Plohl 2010). This sequence homogenization occurs through **mechanisms** of non-reciprocal transfer within and between chromosomes (as gene conversion, unequal crossing-over and rolling circle replication/reinsertion that are dependent of intragenomic identity among satDNA monomers, and also transposon mediated exchange) (Walsh 1987, Thompson-Stewart et al. 1994, Elder and Turner 1995, Dover 2002). These homogenization mechanisms seem to act more efficiently within localized subsets of satellite monomers, decreasing their **efficiency** when occurring between different arrays on the same chromosome, homologous or heterologous chromosomes (Figure 2). This result on different rates of local and global sequence homogenization, showing **adjacent monomers** a higher degree of sequence similarity than those retrieved at random (reviewed in Plohl et al. 2008). In accordance, different works reported that when adjacent monomers are homogenized together may originate a new composite higher order repeat (HOR) unit, in which original monomers become subunits (e.g. Willard and Waye 1987, Warburton and Willard 1990, Acosta et al. 2010). These large complex repeats generally show a high level of sequence identity, accumulating the

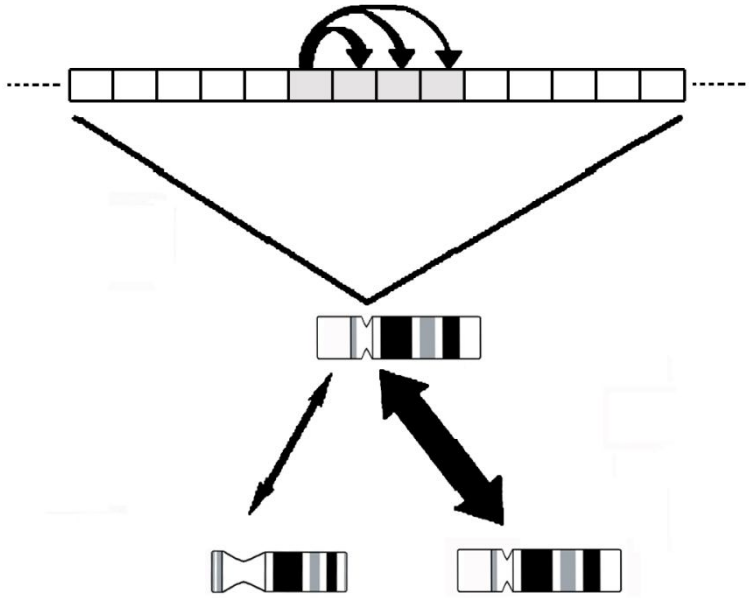


Figure 2- Concerted evolution. Homogenization of a "mutated" satellite monomer within an array. Double head arrows are correlated with the homogenization efficiency that is higher between homologous than non-homologous chromosomes. Adapted from Plohl et al. (2008).

constituent subunits a substantial sequence divergence (Palomeque and Lorite 2008). An example was reported for human alpha-satellite, presenting HORs highly homogeneous (97-100% of similarity), and internal subunits (alpha-satellite monomers) ~70% identical (Willard and Waye 1987, Roizes 2006). Based also on the different rates of sequence homogenization, theoretical models predicted that satDNA monomers at the array ends are more divergent than those located centrally (Smith 1976, Stephan 1989). The low efficiency of homogenization mechanisms in bordering regions of the satellite arrays leads to mutation accumulations in peripheral monomers (Mashkova et al. 1998, Schueler et al. 2005). The subsequent amplification of these monomers can originate novel satDNAs, as schematized in figure 3.

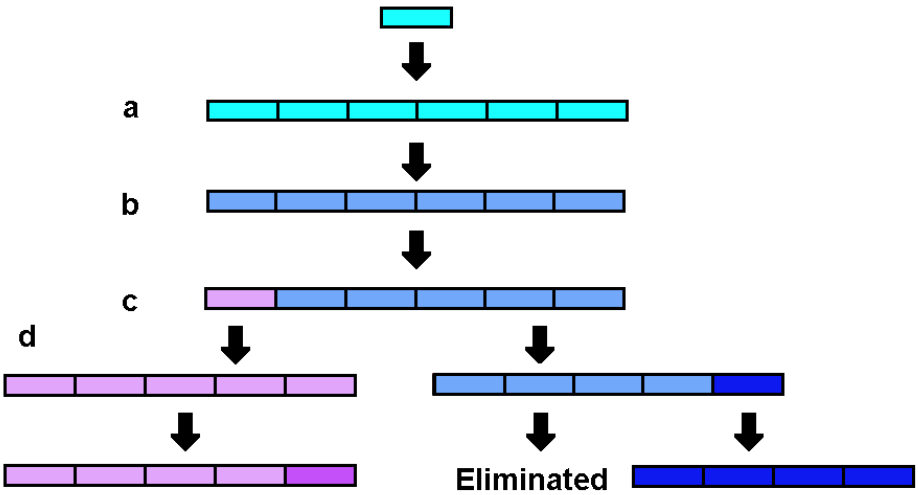


Figure 3- Satellite DNA evolution and diversification. (a) SatDNA is formed by tandem amplification of a DNA sequence. (b) After amplification, sequence alterations and its homogenization lead to satellite sequence divergence relatively to the original sequence presented in **a**. (c) Due to the low efficiency of homogenization mechanisms in bordering regions of the satellite arrays, the amplification of highly mutated peripheral monomers can originate a new satDNA, and the process might repeat itself again. (d) During this evolutionary process, copy number of monomers can change significantly, what may eventually lead to extinction of some satDNAs. Adapted from Plohl (2010).

The described evolution mode of satDNAs results in **species-specific satellite** profiles due to differences in nucleotide sequence, monomer size, copy number variation or chromosome location (Charlesworth et al. 1994, Slamovits and Rossi 2002, Ugarković and Plohl 2002). Generally satDNAs present a highly dynamic molecular behaviour, being often reported satellite sequences only conserved in species belonging to a restricted taxonomic group (e.g. Martinez-Lage et al. 2005), a species (e.g. Stitou et al. 1999) or even a chromosome (e.g. Fátyol et al. 1994). Interestingly, despite the fact that the majority of satDNAs studied so far correspond to rapidly evolving genome components, some of these repeats seem to remain **conserved during long evolutionary** periods, with almost unaltered nucleotide sequences in the different genomes (e.g. Mravinac et al. 2002, Robles et al. 2004, Plohl et al. 2010). These few cases of satDNAs nucleotide sequence conservation highlight the complex behaviour of this genome fraction.The most extreme examples described until now are the mollusc BIV160 satellite family and the rodent PMsat, with about respectively 500 (Plohl et al. 2010) and 635 million years (My) (Louzada et al. 2014 *submitted for publication*). The basis for extreme conservation of some satDNAs is poorly understood. One assumption is that conservation can be a consequence of selection constraints imposed on satellite sequences, and/or it can be the result of slowing down mutation rates (e.g. Robles et al. 2004, Meštrović et al. 2006, Plohl et

al. 2010). The conservation of the satDNA nucleotide sequences can be also predicted by the concerted evolution mode, if "non desirable" mutations are preferentially eliminated instead of being spread throughout the satellite monomers (Dover and Flavell 1984, Ohta and Dover 1984, Strachan et al. 1985).

SatDNAs can **vary dramatically in their number of copies** among related species, and this high variation can be explained by the occurrence of molecular mechanisms involved in the homogenization of satDNA repeats (concerted evolution), namely unequal crossing-over and rolling circle replication/reinsertion (Slamovits and Rossi 2002). The variation in copy number of a set of satDNAs shared by related species was originally explained through the library model. This model postulates that related species share a satellite collection inherited from the ancestral genome, and in each species each satellite may suffer reduction or amplification in copy number, resulting in species-specific satDNA profiles (Fry and Salser 1977), as can be observed in figure 4.

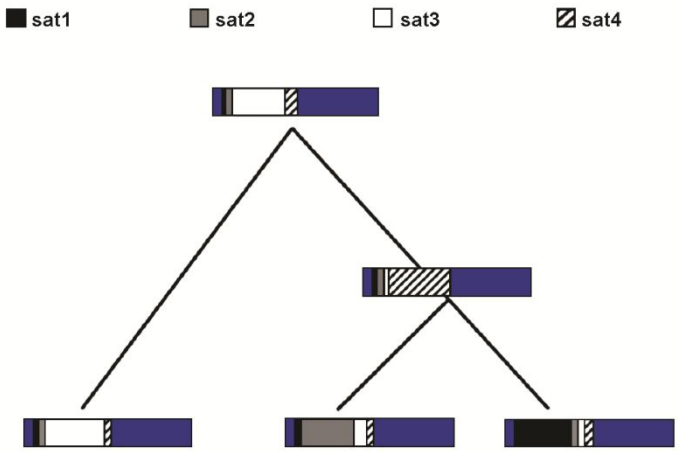


Figure 4- Library model. In a genome several satDNA families can coexist, however, one of these families often exhibits a higher copy number being considered as the major satellite. Adapted from Plohl et al. (2008).

As follows, one or few satDNAs can become highly represented in a species, whereas others remain as low copy number repeats. An experiment supporting this model was carried out by Meštrović et al. (1998) when studying different satDNAs shared by insect species from the genus *Palorus*. In this last work it was also demonstrated that satDNA copy number variations does not implies nucleotide sequence alterations. Species-specific profiles of satDNAs could result only by the variation of repeats copy number in the different genomes, independently of sequence nucleotide conservation (Meštrović et al. 1998, Mravinac et al. 2002, Bruvo et al. 2003).

The concerted evolution mode of satDNAs can also justify a distinct chromosomal location presented by orthologous satDNA sequences in related genomes (Hamilton et al. 1990, Slamovits and Rossi 2002). The homogenization of repeats between different chromosomal fields (within and between chromosomes) by molecular events as unequal crossing-over, rolling circle replication/reinsertion and transposon mediated exchange, can explain the different distribution of a satDNA in the genomes (Hamilton et al. 1990).

I.1.1.1.2- Micro and minisatellites

Micro and minisatellites are tandem repeated sequences composed by short repeat units, being the classification in one of these two categories mostly based on the length of their copies (reviewed by Richard et al. 2008). The size of the repetition motifs in microsatellites varies from 1 to 6 base pairs (bp) and therefore can be classified as mono, di, tri, tetra, penta or hexanucleotide repeats (e.g. (A)13, (GT)8, (GAT)7, (CTAG)6, (CATTG)5, (GGATCC)4). The classical microsatellites present only a single type of repeat unit, exhibiting an array size with up to 100 copies. However, these sequences can present more than a single type of repetitions and thus be named compound microsatellites (Schlötterer and Harr 2001).

Minisatellites present repeat units with up to tens of nucleotides and array sizes that may vary from 10 to 100 copies (reviewed in Slamovits and Rossi 2002). Both microsatellites and minisatellites are distributed throughout the genome, nevertheless minisatellites are also characterized by its (sub)telomeric location (Li 1997, Strachan and Read 2004), representing the telomeric repeats the major family of minisatellites (Strachan and Read 2004).

I.1.1.1.2.1- Telomeric repeats

Telomeres correspond to specialized nucleoprotein complexes (made up of DNA and proteins) that constitute the natural ends of eukaryotic linear chromosomes, displaying important roles, such as protection of the chromosome *termini* from degradation during DNA replication, prevention of deleterious end-to-end chromosome fusions, and participation in intranuclear chromosome positioning and segregation during cell division (Blackburn 2001). In the majority of eukaryotes, telomeric DNA is based on tandem arrays of simple short motifs with about 5-10 bp in length. The telomeric DNA of vertebrates and of some Bilateria

species, such as Mollusca, Annelida and Echinodermata consists of TTAGGG tandem repetitions (Ruiz-Herrera et al. 2008).

In addition to its characteristic terminal position, blocks of telomeric DNA repeats were already found at internal sites on the chromosomes of several vertebrate species, known as interstitial telomeric sequences (ITSs) (e.g. Meyne et al. 1990, Liu and Fredga 1999). The first cytogenetic evidence for the presence of ITSs in the karyotypes of vertebrate species was achieved by Meyne and collaborators (1990), using Fluorescent *in situ* Hybridization (FISH), that identified large ITS blocks preferentially located at the (peri)centromeric regions of the chromosome's. According to their sequence organization and genomic location, two different types of ITSs can be identified in mammalian genomes, large blocks of heterochromatic ITSs (het-ITSs) and short ITSs (s-ITSs). The first type of ITSs presents several hundred kb of telomeric-like DNA, mainly located at heterochromatic (peri)centromeric regions. The s-ITSs length ranges from a few to a few hundred bp, and these repeats can be found in (peri)centromeric and interstitial regions of chromosomes (Ruiz-Herrera et al. 2008). Het-ITSs have been described in several mammalian species, such as primates of the genus *Eulemur* (Go et al. 2000), marsupials (Metcalfe et al. 2007), carnivores (Wurster-Hill et al. 1988), cetartiodactyls (Vermeesch et al. 1996), perissodactyls (Santani et al. 2002), chiropterans (Finato et al. 2000) and rodents (Bertoni et al. 1996, Ventura et al. 2006, Rovatsos et al. 2011). Outside the mammalian, this type of ITSs were also reported in amphibians (Wiley et al. 1992), reptiles (Pellegrino et al. 1999), fishes (Abuín et al. 1996) and birds (Nanda et al. 2002). Regarding s-ITSs, it is believed that they are probably present in all mammalian genomes (Ruiz-Herrera et al. 2008).

The origin of ITSs (het-ITSs or s-ITSs) is attributed to the occurrence of chromosomal rearrangements during karyotypes evolution (Figure 5), namely fusions (e.g. Slijepcevic 1998, Li et al. 2000) or pericentric inversions (e.g. Rovatsos et al. 2011). Specifically, the origin of (peri)centromeric ITSs are commonly explained by the occurrence of robertsonian-like fusions between acrocentric chromosomes, without loss of telomeric sequences. Besides, it is also suggested that s-ITSs can be generated through the insertion of telomeric DNA during the repair of double strand breaks (Nergadze et al. 2007, Ruiz-Herrera et al. 2008). The likely amplification of telomeric repeats through molecular mechanisms, such as unequal crossing-over, replication slippage or gene conversion, have been suggested to explain the large size presented by the het-ITS blocks (Ruiz-Herrera et al. 2008), as schematized in figure 5.

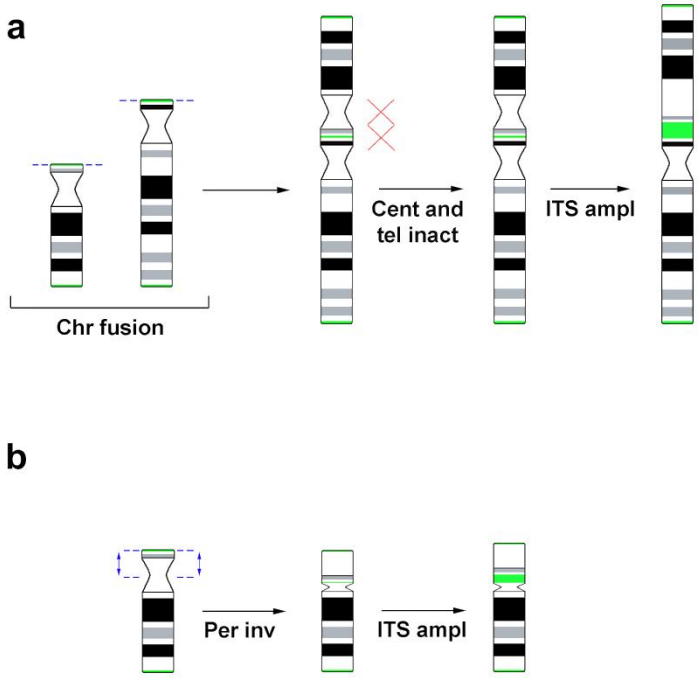


Figure 5- Possible mechanisms to explain the origin of large ITS blocks. Green blocks represent telomeric sequences. The interrupted blue lines correspond to the breaks required for the occurrence of chromosomal reorganizations. (a) Origin of ITS blocks after the occurrence of fusion events. This process requires the inactivation of one centromere and the telomeres in the fusion point, for the correct segregation of the rearranged chromosome. (b) Origin of ITS blocks after the occurrence of a pericentric inversion. Chr fusion – Chromosome fusion, Cent and tel inact – Centromere and telomere inactivation, Per inv – pericentric inversion, ITS ampl – Interstitial telomeric sequences amplification. ITS amplification possibly occurs through molecular mechanisms as unequal crossing-over. This schematization collects informations from several works (Meyne et al. 1990, Slijepcevic 1998, Liu and Fredga 1999, Ruiz-Herrera et al. 2008, Rovatsos et al. 2011).

I.1.1.1.3- Genomic impact and potential roles of tandem repeats

Presently, repetitive sequences are gaining the respect deserved and are regarded as a crucial fraction of eukaryotic genomes, to which important functions have been assigned, like its recognition as important regulatory elements and also its involvement in the reorganization of genomes (e.g. Richard et al. 2008, Plohl 2010, Zhu and Pao et al. 2011, Hall et al. 2012, Enukashvily and Ponomartsev 2013). In fact, the extent of their regulatory importance is currently being explored in detail by The ENCODE Project Consortium, which have started with the human and mouse genomes (Mouse Encode Consortium 2012, The Encode Project Consortium 2012).

One of earliest functions suggested to satDNAs, if not the first, was its involvement in the centromeric activity, which was mainly based in their preferential (peri)centromeric location. In this regard, satDNAs localized within and around centromeres attract a considerable attention, having been proposed a role for these repeats in kinetochore assembling, spindle microtubule attachment and sister chromatid cohesion (e.g. Csink and Henikoff 1998, Henikoff et al. 2001, Sullivan et al. 2001). Additionally to its preferential location, the association of centromeric satDNAs with centromeric proteins also supports the involvement of these repeats in the centromeric functions (e.g. Plohl et al. 2008). The alpha-satellite in the human genome presents a sequence motif of 17 bp long, named CENP-B box, which is able to bind with the CENP-B protein, probably facilitating the kinetochore formation (e.g. Okada et al. 2007, Meštrović et al. 2013). Sequence motifs similar to the CENP-B box were also found in satDNAs from various organisms (e.g. Canapa et al. 2000, Lorite et al. 2004, Mravinac et al. 2005, Meštrović et al. 2013). Moreover, the molecular nature of satellite repeats led also to the suggestion that these sequences are the preferential form of DNA in functional centromeres and their flanking regions, because these sequences gather two features at the same time, sequence homogeneity over long DNA segments but also the potential to change very rapidly in time (e.g. Plohl et al. 2008, Plohl 2010). This dualism is considered particularly important for the interactions centromeric DNA sequences/proteins, since these interactions must remain stable; but to retain this stability over time the DNA sequence must also have the potential to coevolve with the rapidly changing protein component (Dawe and Henikoff 2006). Thus, the homogenization of monomer mutations within a (peri)centromeric satDNA array, which bring more efficiency for binding centromeric proteins, seems a crucial event in centromeric function (reviewed in Plohl et al. 2008). In accordance, it was assumed that both satDNA and protein evolution drive each other in the centromere, providing a stable but flexible system essential for centromeric activity (Dawe and Henikoff 2006). Nevertheless, the confirmation and completely recognition of the satDNA contribution to centromeric activity was achieved by the finding of transcription of these repeats, that shown that satDNA transcripts are structural elements of the functional centromere/kinetochore complex (e.g. Hall et al. 2012, Enukashvily and Ponomartsev 2013). Moreover, the satDNA centromeric transcripts are actually implicated in CENP-A deposition, a histone H3 variant found only at active centromeres (Bergmann et al. 2012). These discoveries add a new dimension to our current view of how (peri)centromeric sequences

participate in the formation, maintenance and function of specific structures such as centromeres.

Studies in the last decades shed light on a previous dark area where transcription was not accepted as a trait of satDNAs, mainly because these repeats are embedded in tightly packed heterochromatin. Presently, we all are aware of the fact that these repeats are transcribed, resulting in non-coding RNAs (e.g. Wong et al. 2007, Vourc'h and Biamonti 2011, Hall et al. 2012, Pezer and Ugarković 2012, Enukashvily and Ponomartsev 2013). In fact, the transcription of satDNAs seems to be a general phenomenon, having been reported satellite transcripts in several organisms including vertebrates, invertebrates and plants (Pezer and Ugarković 2012). So far, however, little is known about basic mechanisms of satDNA expression, its regulation and the function of all the generated transcripts (Pezer and Ugarković 2008b, Vourc'h and Biamonti 2011, Enukashvily and Ponomartsev 2013). Concerning the current knowledge, satDNA transcripts are usually heterogeneous in size and the transcription can proceed in both DNA strands or be strand-specific (Rudert et al. 1995, Rouleux-Bonnin et al. 1996). Some transcripts are present as polyadenylated RNA in the cytoplasm while others are found exclusively in the nucleus (Trapitz et al. 1988, Bonaccorsi et al. 1990). Besides, the transcription of some satDNAs is associated with the differentiation and development, and can be gender, age or tissue-specific, which provide clues indicating that these transcripts exhibit a regulatory role (Pezer and Ugarković 2012, Enukashvily and Ponomartsev 2013).

In addition to the centromeric functions, satDNA transcripts seem to have other and diverse roles. It has been shown that long single-stranded polyadenylated transcripts of human satellite III are directly involved in the cellular response to stress (Valgardsdottir et al. 2005). Some long satDNA transcripts can also function as ribozymes with self-cleavage activity (reviewed in Ugarković 2005). Moreover, diverse works reported as well that some satDNA transcripts can act as precursors of small interfering RNAs (siRNAs), which are recognized as having an important role in chromatin remodelation, leading to the heterochromatin formation and maintaining, and in the control of gene expression (reviewed in Vourc'h and Biamonti 2011, Enukashvily and Ponomartsev 2013). The mechanisms of chromatin modifications by siRNAs derived from satDNAs have been extensively studied in *Schizosaccharomyces pombe*, but it is believed that these mechanisms are conserved in other organisms as *Drosophila*, plants and, although with some doubts, in mammals (Pezer and Ugarković 2012, Enukashvily and Ponomartsev 2013). Double stranded RNAs of the *S. pombe* pericentromeric

satDNA are processed into siRNAs by the endonuclease dicer (Usakin et al. 2007, Eymery et al. 2009, Pezer and Ugarković 2012). After, these siRNAs are associated with a complex of proteins (RIST complex), enabling this complex to identify nascent centromeric/pericentromeric transcripts (by siRNA sequence complementarily), to which it will associate. This association targets these regions for chromatin modifications, as the methylation of histone H3 at lysine 9, providing the formation of the self-sustaining closed chromatin state (Usakin et al. 2007). This heterochromatic state of the centromeric/pericentromeric regions is needed for proper centromere functioning, being required for kinetochore formation and sister chromatid cohesion ensuring (e.g. Hall et al. 2012), as referred previously. In this regard, chromatin modifications by satDNA siRNAs are other evidence that supports the high importance of these repeats in the centromeric activity. To the control of gene expression by siRNAs derived from satDNAs, these siRNAs need to associate with a different protein complex (RISC complex), allowing the cleavage of gene transcripts complementary to siRNAs in a process named post-transcriptional gene regulation (Buckingham 2003, Ugarković 2005). The existence of several human coding mRNAs that contain alpha-like satellite repeats as part of their 5' or 3' UTRs, indicates that their expression can be controlled by siRNAs derived from alpha-satellite repeats (Li and Kirby 2003). Alternatively to long strand transcripts, the role of satDNA siRNAs in response to stress conditions was also reported for some eukaryotic organism, as plants and insects (Sunkar et al. 2007, Pezer and Ugarković 2012).

Other generally accepted function of satDNAs is its involvement in genomic instability, acting as source of chromosomal rearrangements, which has a great impact on the genesis of genetic diseases as cancer (Santos et al. 2006), but also in genome evolution (e.g. Wichman et al. 1991, Bradley and Wichman 1994, Rossi et al. 1995, Garagna et al. 1997, Slamovits and Rossi 2002, Adega et al. 2009, Zhu and Pao et al. 2011). This role of satDNAs is justified mostly by the high molecular dynamic of these repeats (reviewed in Slamovits and Rossi 2002), having been reported the location of satDNAs at the breakpoint regions of chromosomes (Adega et al. 2009). The high repetitive nature of these repeats and its dynamic molecular behaviour based in an evolutionary mode mediated through mechanisms of non-reciprocal transfer within and between chromosomes can create the opportunity for the occurrence of chromosomal rearrangements (e.g. Slamovits and Rossi 2002, Froenicke and Lyons 2008). Recently, the analysis of satDNA transcription reinforced the involvement of satDNAs in the origin of chromosomal rearrangements, being observed an induction of

genomic instability by the overexpression of satDNAs (Ting et al. 2011, Zhu and Pao et al. 2011).

Like satDNAs, micro and minisatellites are also frequently associated with genomic instability (reviewed in Richard et al. 2008). Different studies reported an enrichment of micro and minisatellites at breakpoint regions (e.g. Bertoni et al. 1996, Bolzán and Bianchi 2006, Ruiz-Herrera et al. 2006) what reflects an involvement of these sequences in the occurrence of chromosomal rearrangements (Ruiz-Herrera et al. 2006, Richard et al. 2008). Further, studies about a particular kind of minisatellite repeats, the ITS blocks, point these as unstable regions, representing hotspots for chromosome fusions/fissions, tandem fusions or inversions (e.g. Ashley and Ward 1993, Nanda et al. 2002, Farré et al. 2009). This was particularly suggested for het-ITSs, due to their propensity to form secondary structures, which is favoured by its large length, creating the opportunity for recombination events that may induce breakage (Balajee et al. 1994, Fernández et al. 1995, Nergadze et al. 2007).

I.1.1.2- Dispersed repeats (transposable elements)

The first study about transposable elements (TEs) was performed by Barbara McClintock in 1950. Since that time, and as more genomes are sequenced, higher is the diversity of the TEs families described. Considering the information collected in some fully sequenced genomes, as human, mouse and rat, it is possible to estimate that repetitive DNA derived from TEs comprises from 40% to almost half of these genomes (International Human Genome Sequencing Consortium 2001, Mouse Genome Sequencing Consortium 2002, Rat Genome Sequencing Project Consortium 2004). However, these estimates can be fairly conservative, with considerable amounts of more ancient elements being undetectable due to their divergence from the consensus sequences used for its detection.

Despite their high diversity and abundance, all eukaryotic TEs fall into two basic types: retrotransposons (class I) and DNA transposons (class II) (reviewed in Jurka et al. 2007, Kapitonov and Jurka 2008, Rebollo et al. 2012). Retrotransposons are transposed through an RNA intermediate mechanism (retrotransposition), generating new copies of these elements that insert into novel genomic locations (Jurka et al. 2007). Unlike retrotransposons, DNA transposons are transposed by moving their DNA copies from one chromosomal location to another without any RNA intermediate, with no requirement of copy number increase (Jurka et al. 2007, Wicker et al. 2007).

A classification based on the mode of retrotransposition allows distinguishing four classes of retrotransposons: long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons, Penelope and DIRS retrotransposons (*Dictyostelium* intermediate repeat sequence) (Jurka et al. 2007, Kapitonov et al. 2009). Eukaryotic DNA transposons belong to three classes: "cut-and-paste" transposons, helitrons and polintons. All these TEs are represented by autonomous and non-autonomous variants. Whereas an autonomous element encodes the enzymes that allow their own transposition, a non-autonomous element transposes using the enzyme machinery encoded by an autonomous variant (e.g. Kazazian 2004, Wicker et al. 2007).

Here we describe in more detail the molecular characteristics of autonomous non-LTR retrotransposons, commonly referred to as Long Interspersed Nuclear Elements (LINEs). These elements contain one or two ORFs that encode the enzymes required for retrotransposition (reviewed in Jurka et al. 2007, Han 2010). Some present a single ORF that codifies for a protein with both endonuclease and reverse transcriptase domains. Others present two ORFs, ORF1 and ORF2, that encode for the ORF1 and ORF2 proteins

respectively (e.g. Kazazian 2004, Lee et al. 2010). The role of ORF1 protein is not yet completely clear because the aminoacid sequence predicted for its polypeptide (40 KDa) lacks homology with any known functional protein (see Hohjoh and Singer 1996, Martin 2006) however, it is regarded as a non-specific nucleic acid binding protein with nucleic acid chaperone activity (e.g. Martin 2010). By contrast, the role of ORF2 protein is best known, having its 146 KDa multifunctional polypeptide simultaneously endonuclease and reverse transcriptase activities, crucial for retrotransposition (e.g. Dewannieux and Heidmann 2005, Doucet et al. 2010), presenting, although, a cysteine-rich domain with a still unknown function (Dai et al. 2011). Flanking the ORFs, a full length LINE also contains untranslated regions (UTRs), the 5' and 3' UTRs, quite variable in these elements. The internal promoter (RNA polymerase II promoter) of functional non-LTR elements is present in the 5'UTR (reviewed in Han 2010).

Based on the structural features and phylogeny of reverse transcriptase's, LINEs can be classified into five groups: R2, L1, RTE, I and Jockey, which can be subdivided into 28 clades (see Kapitonov et al. 2009). The most active LINEs group identified within the sequenced mammalian genomes are the L1 elements (LINE-1), comprising about 17%, 20% and 23% of the human, mouse and rat genomes, respectively (International Human Genome Sequencing Consortium 2001, Mouse Genome Sequencing Consortium 2002, Rat Genome Sequencing Project Consortium 2004). In fact, LINE-1 is the dominant group of transposable elements in these genomes (Pascale et al. 1990, Furano et al. 2004). In mice and rats, it was suggested that in the past 10 My the genome size has increased 10 to 20% due to LINE-1 insertions (Pascale et al. 1990).

I.1.1.2.1- Mammalian LINE-1 retrotransposons

A complete and transpositionally active mammalian LINE-1 sequence has approximately 6 to 7 kb long, containing a 5' UTR with an internal promoter, two ORFs (ORF1 and ORF2), a 3'-UTR that presents a G-rich polypurine tract, ending in a A-rich region (e.g. Furano and Usdin 1995, Kazazian 2000, Deininger and Batzer 2002, Moran and Gilbert 2002, Weiner 2002, Lee et al. 2010). However, the sequence of these elements presents particular features in different genomes (Figure 6). For example, unlike the human LINE-1 elements, in mouse and rat LINE-1, 5' UTR have a region with tandemly repeated monomers of approximately 200 and 600 bp (respectively), which are situated upstream of a single copy non-monomeric

sequence (Adey et al. 1991, 1994, Severynse et al. 1992, DeBerardinis and Kazazian 1999). The link of mouse 5' LINE-1 monomers to reporter genes shows that these monomers present promoter activity, and that the increase of the monomer number at the 5' UTR region raises the level of LINE-1 transcription (DeBerardinis and Kazazian 1999).

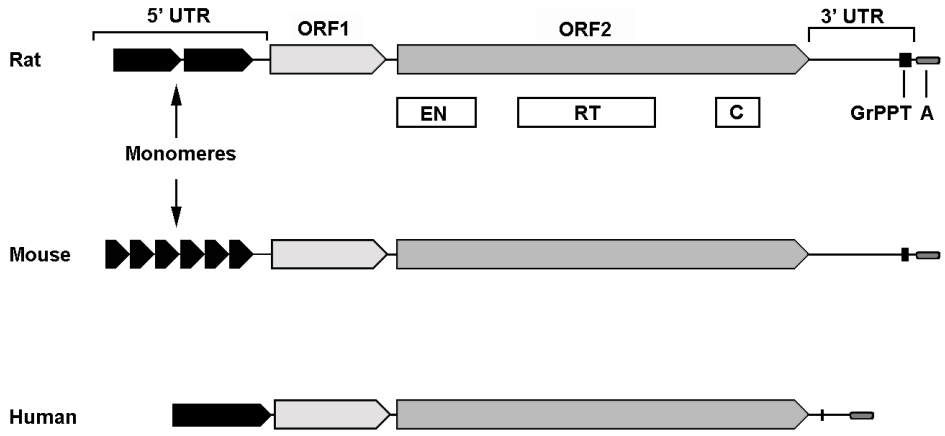


Figure 6- Schematic representation of rat, mouse and human LINE-1 retrotransposon. (EN) Endonuclease domain, (RT) Reverse transcriptase domain, (C) cysteine-rich domain, (GrPPT) G-rich polypurine tract, (A) A-rich region. Adapted from Furano et al. (2004).

Contrary to what was thought initially, the LINE-1 retrotransposition is not restricted to the germline, having been found LINE-1 translation products in some somatic tissues, such as the vascular endothelia of human male gonads (Ergun et al. 2004), rat cardiomyocytes and endothelial cells (Lucchinetti et al. 2006). Like other LINEs it is believed that LINE-1 retrotransposition involves a Target Primed Reverse Transcription (TPRT) process (Belancio et al. 2008). The most generally accepted steps involved in the mechanism of LINE-1 retrotransposition are resumed in the figure 7. The first step corresponds to the transcription of a full length active LINE-1 retrotransposon. The resultant RNA is exported to the cytoplasm, where ORFs are translated and form a ribonucleoprotein particle (RNP) with the LINE-1 RNA that encode them (Han and Boeke 2005) (Figure 7a). After, the RNP is imported to the nucleus, beginning what is specifically designated of TPRT (Figure 7b). During this process, the LINE-1 endonuclease domain of the ORF2 protein recognizes a consensus 5'-TTTTAA-3' sequence in the target site and introduces a nick between the T and A nucleotides of the minus strand. The resulting free 3'end of the target site base pair with the

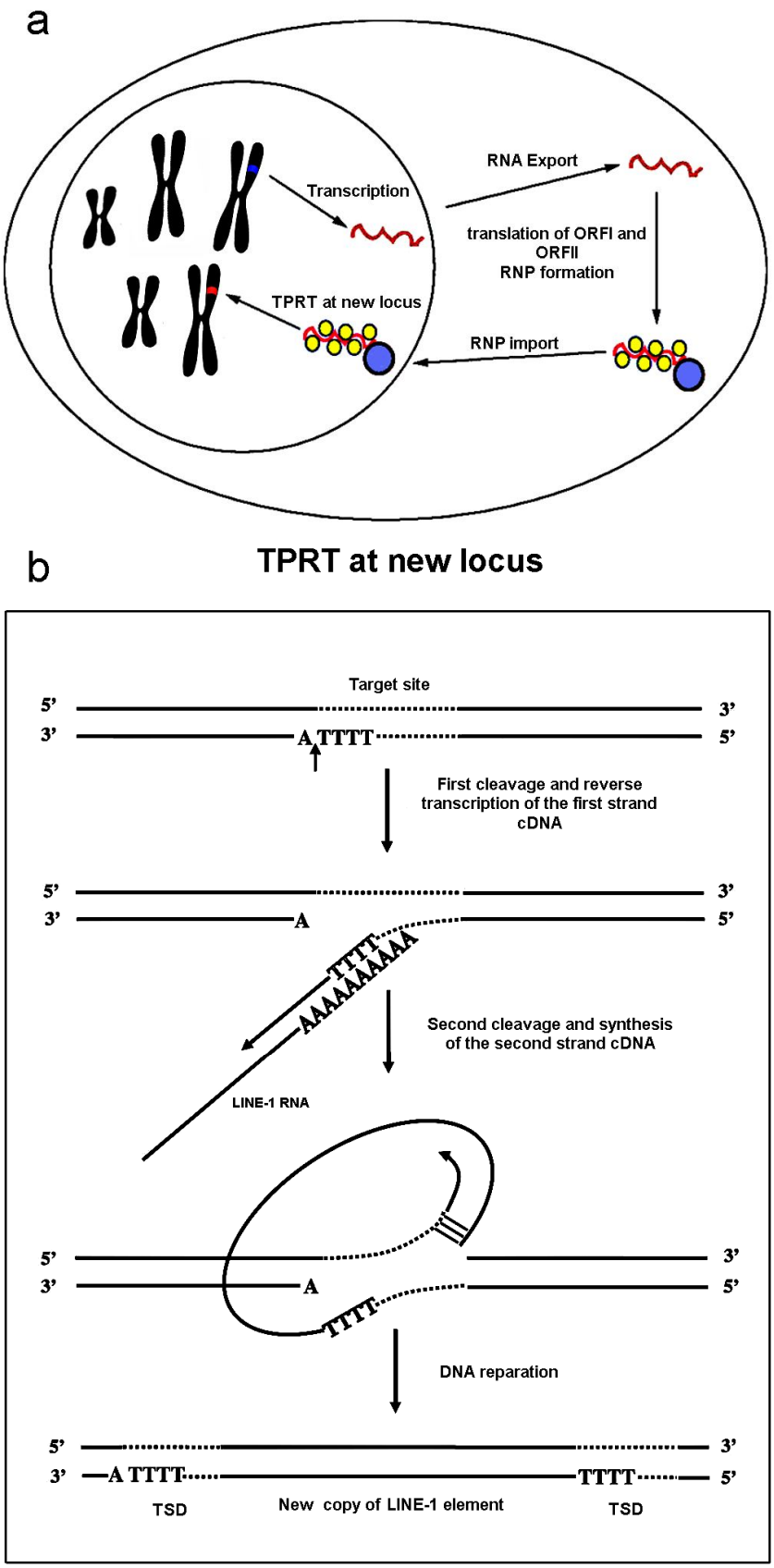


Figure 7- LINE-1 retrotransposition mechanism. (a) Life cycle of LINE-1 retrotransposon. (b) Target Primed Reverse Transcription (TPRT) process. Target Site Duplication (TSD). Adapted from Han and Boeke (2005) and Ding et al. (2006).

poly(A) tail of the LINE-1 mRNA and serves as a primer for the first strand cDNA synthesis, by the ORF2 reverse transcriptase that uses the LINE-1 mRNA as template. The details for the end of the LINE-1 integration process are not well defined yet. At some point during LINE-1 integration, a second cleavage is introduced at the plus strand of the target site that will prime the second strand cDNA synthesis using the first strand cDNA as template, by an unknown polymerase activity. Finally, the two nicks in the cellular DNA are repaired to complete the LINE-1 integration event, originating target site duplications (TSDs) (Ding et al. 2006, Belancio et al. 2008).

It is recognized that the LINE-1 insertions often fail to include the 5' end of the retrotransposon (promoter region), generating defective copies 5' truncated, which remain in the genome without autonomous mobile capacity (retrotransposition inability). These truncated insertion events can be the result of the low LINE-1 reverse transcriptase processivity (Ostertag and Kazazian 2001). Therefore, there are frequently found in the genomes 5' truncated elements and only a very limited copy number of full length elements. Additionally, the retrotransposition can also generate other variants of LINE-1 elements. LINE-1 transcripts are commonly submitted to processes of premature polyadenylation and different splicing events, originating spliced and/or partially rearranged copies of the template element (Belancio et al. 2008). But more importantly, is the possible origin of novel replication competent LINE-1 sequence variants during retrotransposition, displaying the ability to pass on to all subsequent copies the new acquired sequence alteration, generating a new family of retrotransposons (e.g. Mayorov et al. 1999, Casavant et al. 2000, Boissinot and Furano 2001, Furano et al. 2004). From out of 500000 LINE-1 elements estimated in the human genome, only approximately 7000 are full length copies, and of those only 80 to 100 are probably active for retrotransposition (Brouha et al. 2003). This suggests that the LINE-1 evolution is intimately related with the retrotransposition mechanism, since only a minority of LINE-1 elements in a genome are competent for producing copies ("master" templates) and consequently to pass their sequence features over time. During LINE-1 evolution, these templates will be replaced by a small number of the recently originated elements (Casavant et al. 1996, 1998). The elements that had the ability to replace its predecessor's enclose selective advantages for LINE-1 retrotransposition. This is probably achieved by the acquisition of novel and more competent transcriptional regulatory sequences, as the promoter region (Adey et al. 1994, Goodier et al. 2001).

Regarding specifically their genomic distribution, the mammalian LINE-1 sequences seem to present a non-random location. In several eutherian mammalian species, it was shown that these retrotransposons are preferentially located at the AT rich isochores (G-positive bands, late-replicating DNA) (Boyle et al. 1990), as reported for human (Korenberg and Rykowski 1988), rabbit (Waters et al. 2004), mouse (Boyle et al. 1990), rodents from the genus *Taterillus* (Muridae) (Dobigny et al. 2002) and from the genus *Cryptomys* (Bathyergidae) (Deuve et al. 2006), all belonging to the Superorder Euarchontoglires (Murphy et al. 2001, Delsuc et al. 2002). Nevertheless, this AT banding pattern was not observed in species belonging to other mammalian Superorders as Laurasiatheria, Xenarthra or Afrotheria (Thomsen and Miller 1996, Parish et al. 2002, Waters et al. 2004).

LINE-1 elements were also claimed to accumulate in regions presenting low rates of recombination (e.g. Boissinot et al. 2001, Graham and Boissinot 2006), in regions with monoallelically expressed genes (see Allen et al. 2003, Walter et al. 2005), and in the X-chromosome, in comparison with autosomes (Korenberg and Rykowski 1988, Boyle et al. 1990). This suggests that LINE-1 elements can be involved in gene expression and in the inactivation of the X-chromosome (Lyon 1998). These topics will be discussed in more detail ahead in this chapter.

I.1.1.2.2- Genomic impact and potential roles of mammalian LINE-1 retrotransposons

Traditionally, as referred for the remaining genomic repetitive fraction, the occurrence of LINE-1 in mammalian genomes has been explained by the selfish DNA hypothesis (Orgel and Crick 1980). Presently however, this idea is no longer accepted and several important functions are addressed to these retrotransposons, as its role in regulation of gene expression (e.g. Yang et al. 1998, Han and Boeke 2004, Muotri et al. 2007, Akagi et al. 2008), X-chromosome inactivation (Lyon 1998, 2006, Bailey et al. 2000), genome reorganization (e.g. Boissinot et al. 2006, Song and Boissinot 2007, Kolb et al. 2009, Longo et al. 2009) and progression of early embryogenesis (e.g. Vitullo et al. 2011).

The fact that transposable elements (TEs) can influence host gene expression was first recognized more than 50 years ago, mostly because they present a noticeable ability to produce mutations when integrating at new genomic sites (Rebollo et al. 2012). The integration of these elements in the coding or regulatory sequences of genes may produce silencing or changes in the expression pattern of these genes (McDonald 1995, Capy et al.

1997). Indeed, several genetic diseases have been associated with the insertion of LINE-1 retrotransposons (reviewed in Belancio et al. 2008). Nevertheless, despite the vast majority of LINE-1 insertions are either neutral or deleterious to their host its inclusions into new locations may also be advantageous, promoting gene evolution and the codification of more efficient protein variants. One of the most highly publicized discoveries of this nature is the resistance to HIV-1 (Human Immunodeficiency Virus) infection in owl monkeys, that present an altered TRIM5 gene with a cyclophilin A domain acquired by LINE-1 retrotransposition (Sayah et al. 2004). The binding of this cyclophilin domain to the HIV-1 viral capsid lead to a disruption of the infection process, through a not yet very clear mechanism (Frausto et al. 2013). These primates are however permissive to other immunodeficiency virus, as Simian Immunodeficiency Virus (SIV) (Sayah et al. 2004).

Still in the scope of the LINE-1 role in control of gene expression, it is also important to refer the probable involvement of LINE-1 in the control of monoallelically genes expression, which is supported by the preferential and non-random location of these sequences in regions with monoallelically expressed genes (Allen et al. 2003, Walter et al. 2005).

A controversial issue is the possible involvement of LINE-1 retrotransposons in X-chromosome inactivation, hypothesized for the first time by Lyon (1998), based in high accumulations of these sequences in human and mouse X-chromosomes in comparison with the autosomes (Korenberg and Rykowski 1988, Boyle et al. 1990). Nevertheless, this LINE-1 preferential distribution is not observed in all mammalian studied species (e.g. Deuve et al. 2006), which difficult the determination of the LINE-1 exact role in X-chromosome inactivation.

Like tandem repetitive sequences, LINE-1 retrotransposons are also associated with the occurrence of chromosomal rearrangements (e.g. Hedges and Deininger 2007, Venner et al. 2009). These sequences are arguably one of the most significant dynamic forces operating on the mammalian genomes, with studies pointing to a great impact on genome evolution (e.g. Martin et al. 2005, Wallace et al. 2008, Lee et al. 2010, Lupski 2010, Martin 2010). LINE-1 retrotransposition events generate homologous sequences in non-homologous regions providing opportunities for recombination on misaligned chromosomes, which may result in chromosome rearrangements (e.g. Moran and Gilbert 2002, Boissinot et al. 2006, Hedges and Deininger 2007, Song and Boissinot 2007). A much larger fraction of retrotransposons related diseases in human results from recombination mutations than from insertional mutations (Deininger and Batzer 1999).

The LINE-1 retrotransposons correspond also to key DNA elements in the progression of the early embryogenesis. Several works have shown that the reverse transcriptase, encoded by LINE-1 retroelements in human and mouse cells, display a high importance in progression of early embryogenesis (e.g. Pittoggi et al. 2003, Beraldi et al. 2006, Spadafora 2008). Moreover, most, if not all, of the reverse transcriptase activity required for preimplantation development is encoded by LINE-1 (Vitullo et al. 2011).

I.1.1.3- Tandem and dispersed repeats evolutionary relationship

Tandem and dispersed repetitive sequences have mostly been investigated independently, however there are growing evidences that transposable elements (TEs) are involved at various stages of tandem repeats evolution, namely in its origin, homogenization and genome dispersion (e.g. Rossi et al. 1993, Batistoni et al. 1995, Kapitonov et al. 1998, Kapitonov and Jurka 1999, Cheng and Murata 2003, Inukai 2004, López-Flores et al. 2004, Macas et al. 2009, Smýkal et al. 2009).

The molecular mechanisms proposed to explain the amplification and homogenization of satDNA monomers by concerted evolution are generally accepted (e.g. unequal crossing-over and rolling circle replication). On the contrary, questions regarding the origin of the first repetitions from which satellites evolved are poorly understood (reviewed in Slamovits and Rossi 2002). In a genome, new satellite sequences could evolve from the preexisting pool of satDNAs, or could also be originated *de novo* from non satDNA sequences (Smith 1976, Kapitonov et al. 1998). Theoretical models and computer simulations suggested that satellite units could be generated from a wide spectrum of non satDNA and propagated into an array by unequal crossing-over (Smith 1976). The best known example comes from several reports indicating a role of TEs in the origin/expansion of some tandem repeats (Rossi et al. 1993, Batistoni et al. 1995, Heikkinen et al. 1995, Kapitonov et al. 1998, Kapitonov and Jurka 1999, Cheng and Murata 2003, López-Flores et al. 2004, Macas et al. 2009). The sequence similarity between some satDNAs and TEs supports the probable evolutionary relationship between them (e.g. Kapitonov et al. 1998). Amplifications of part of a TE sequence by unequal crossing-over between homologous dispersed repeats could correspond to the first step for the origin of tandem duplications. Mutational changes, followed by successive rounds of crossing-over homogenization, can justify the divergence observed between the emergent satDNA and the original TE, presenting only conserved parts of their sequences (Wong and

Choo 2004). Moreover, it is also believed that TEs are involved in the origin of some sequence motifs that characterize satDNAs, as the CENP-B box identified in different satellites, presenting this sequence motif strong similarity with a DNA transposon family (Plohl 2010).

Besides the role suggested to the tandem repeats origin, the TEs were also implicated in satDNA homogenization and dispersion in the genomes (e.g. Palomeque and Lorite 2008, Macas et al. 2009), having the transposition been considered one of the homogenization mechanisms in the process of concerted evolution (Dover 2002). The abundance of TEs in heterochromatic (peri)centromeric regions of a wide range of species, being these regions also mainly built by satDNAs (Enukashvily and Ponomartsev 2013), certainly facilitates the dispersion of these highly tandem repeats by transposition. Regarding particularly LINE-1 retrotransposons, diverse are the works reporting a location of these elements in centromeric regions of different mammalian species chromosomes (e.g. Mayorov et al. 1996, Waters et al. 2004, Marchal et al. 2006, Acosta et al. 2008), suggesting that these retrotransposons are probably intermingled with satDNAs, forming a complex structure (e.g. Mayorov et al. 1996, Marchal et al. 2006). This intricate organization pattern of repetitive sequences in the centromeric regions eventually favour the dispersion of satDNAs to other genomic locations by LINE-1 retrotransposition, since frequently these elements allow the transduction of flanking non-LINE-1 DNA to new genomic locations. This is a consequence of the LINE-1 polyadenylation signal weakness that is occasionally bypassed in favor of a stronger signal downstream, or when a promoter located upstream of the retrotransposon is used to transcribe this retroelement (e.g. Goodier et al. 2001, Babushok et al. 2007, Cordaux and Batzer 2009). The transcript containing the retrotransposon along with the extra genomic sequence is integrated back into the genome via retrotransposition, resulting in the duplication and genomic dispersion of the LINE-1 flanking sequences (Cordaux and Batzer 2009), as satDNA. Concerning the micro and minisatellites, a considerable part of these repeats in the eukaryotic genomes are embedded within mobile elements which probably facilitate their amplification and dispersion in the genomes (Inukai 2004, Smýkal et al. 2009), explaining its preferential dispersed chromosomal location.

I.2- Rodentia phylogeny: radiation and biogeography

Rodents present an ample distribution occupying a broad range of habitats from humid tropical forests to arid deserts and tundra, adopting equally a wide range of life styles, as semi-aquatic, arboreal, scansorial and fossorial. It is estimated that the origin of this Order (Rodentia) occurs in the late Cretaceous, at ~82,6 My (OneZoom Tree of Life Explorer database, http://www.onezoom.org/index.htm, updated December 2013), and since that time rodents have undergone a remarkable radiation leading to the elevated number of species observed today (Huchon et al. 2002, Benton and Donoghue 2007), comprising about 42% of all living mammalian species (Carleton and Musser 2005). Some rodent species are used extensively in biomedical research and this has stimulated the interest in the study of this group. Namely, the role that certain rodent species play as vectors of human disease has given high importance to the studies of their ecology and phylogeny (Hughes and Friedman 2000). Furthermore, it also exist a high interest in understanding why this group is so much more diverse than any other mammalian clade, sustaining one of the highest *net* speciation rates among vertebrates.

The phylogeny of rodents has been one of the most intractable problems in mammalogy, because their rapid radiation left little opportunities for the evolution of unique synapomorphies and the systematic is largely based on dental characters, which are particularly prone to adaptive convergence (Feldhamer et al. 2003). At the moment, modern taxonomy recognizes 5 suborders of rodents, Sciuromorpha, Myomorpha, Castorimorpha, Hystricomorpha and Anomaluromorpha (reviewed in Romanenko et al. 2012), as it is presented in figure 8. Nearly one third of all rodent species are classified in the suborder Myomorpha making this taxon particularly attractive for evolutionary studies. The species studied in this work belong to two large Myomorpha families Cricetidae and Muridae, included in the Muroidea superfamily (Steppan et al. 2004).

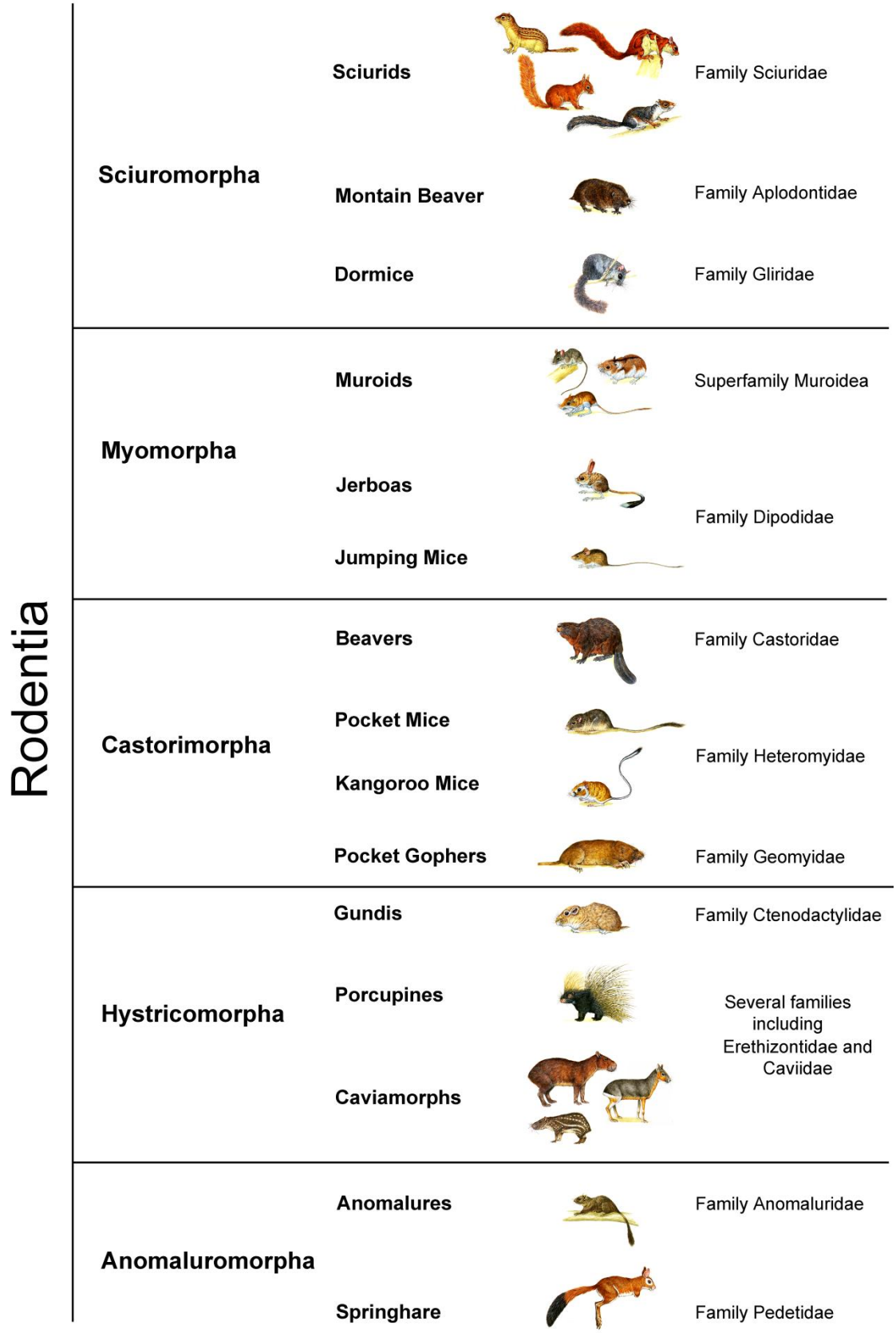


Figure 8- Rodentia suborders. Classification according to data from Tree of Life Web Project, mainly based in works of Robinson et al. (1997), Michaux and Catzeflis (2000), Michaux et al. (2001), Jansa and Weksler (2004), Steppan et al. (2004, 2005).

Within the superfamily Muroidea (muroids) all the species descendants from the most recent common ancestor of the Calomyscinae, Nesomyidae, Cricetidae and Muridae are grouped in the Eumuroida clade (Tree of Life Web Project, http://www.tolweb.org/tree/), a taxonomic group defined by Steppan et al. (2004). According to molecular data, it is assumed that the origin of the Eumuroida species took place near the border between the Miocene and Oligocene (~24-26 My) (Steppan et al. 2004). Besides the Eumuroida, all mouse-like fossorial rodents of the family Spalacidae are also classified in the Muroidea superfamily (Norris et al. 2004), as can be observed in figure 9.

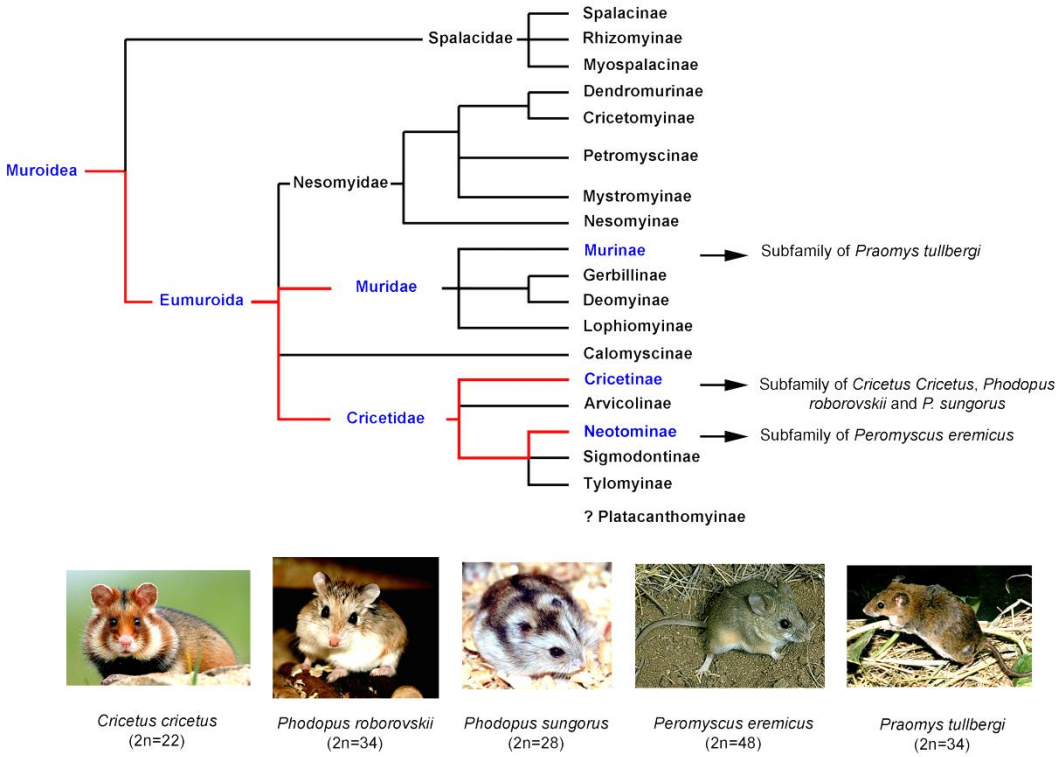


Figure 9- Phylogenetic tree of the superfamily Muroidea. This tree results from the compilation of data from Robinson et al. (1997), Michaux and Catzeflis (2000), Michaux et al. (2001), Jansa and Weksler (2004), Steppan et al. (2004, 2005). Additional studies are needed to determine the appropriate position of the Platacanthomyinae. This figure also presents in more detail the taxonomic framework of the five species studied in this work (red lines and blue names). Adapted from Tree of Life Web Project.

According to estimates of molecular taxonomists, Muridae and Cricetidae rodents diverged at ~17 My ago (Robinson et al. 1997). The Cricetidae species studied here are classified in two subfamilies, Cricetinae (*Cricetus cricetus*, *Phodopus roborovskii* and *Phodopus sungorus*) and Neotominae (*Peromyscus eremicus*). Concerning the family Muridae, the

species studied in this work was *Praomys tullbergi*, classified in the subfamily Murinae (Tree of Life Web Project), as can be observed in figure 9.

I.2.1- Subfamily Cricetinae: *Cricetus* and *Phodopus* genus

The Cricetinae subfamily comprises the hamsters, small rodents that originally live in natural or semi-natural steppes in Europe and Asia (Nechay 2000). Most of Cricetinae genera are described from the late Miocene (McKenna and Bell 1997, Kowalski 2001) (~5.3 - 11.6 My), being the origin of these rodents often associated with a taxon that lived in the northern hemisphere during the early and middle Miocene (~11.6 - 23 My), the Democricetodontini (Fahlbusch 1969). The Cricetinae species diversification is related with the spread of steppe and open woodlands in Europe and Asia, due to increasingly drier climatic conditions that took place at ~8 - 10 My ago, as also the tectonic changes such as further uplift of the Tibetian Plateau (~10 My) that provides physical barriers as well as climatic changes across Euroasia. The spread of deserts across Asia during the Pliocene (~2.5 - 5.3 My) could also have facilitated further diversification within major Cricetinae clades (Neumann et al. 2006).

Some discrepancies occur between the informations obtained by paleontological and molecular data for the *Cricetus* genus. Fossil records of *Cricetus* species were found from the beginning of the Pleistocene (~2.5 My), however, molecular data place the origin of the genus *Cricetus* sooner in the upper Pliocene (~ 2.5 - 3.6 My) (Neumann et al. 2006). Currently, this genus presents only one species, *Cricetus cricetus* (Figure 9), inhabiting Eurasia from Belgian to Siberian (Nowak 1991).

Based on molecular (Neumann et al. 2006) and cytogenetic data (Romanenko et al. 2007) it was proposed that the species from the genus *Phodopus* represent the earliest split among all species of the Cricetinae subfamily. It is estimated that the separation between the genus *Phodopus* and the other Cricetinae genera occurred at 8.5 – 9 My, in the later Miocene (Neumann et al. 2006). *Phodopus* genus comprises three hamster species [two of these species are studied in this work, *P. roborovskii* and *P. sungorus* (Figure 9)], inhabiting cold and deserts regions of Central Asia (Carleton and Musser 2005).

I.2.2- Subfamily Neotominae: *Peromyscus* genus

The subfamily Neotominae includes many of New World mices and rats that live almost exclusively in North America. These rodents are closely related to other subfamilies of rodents, Sigmodontinae and Tylomyinae, and because of that some authors classify them in a single subfamily (Steppan et al. 2004). Fossil records suggest that the Neotominae descend from the North American genus *Copemys* that lived during the Miocene, at ~9 - 16 My ago (Baskin 1986). To this subfamily belongs the *Peromyscus* genus. The phylogenetic relationships within this genus are poorly understood, however, it is recognized that the diversification of *Peromyscus* species is related with the occurrence of geological events and climatic alterations in North America, resulting in the creation of new habitats. *Peromyscus eremicus* was the species studied in this work and it inhabits desert regions of North America (Riddle et al. 2000).

I.2.3- Subfamily Murinae: *Praomys* genus

The Murinae subfamily (Old World mice and rats) corresponds to the largest subfamily of mammals, comprising the most commonly used laboratory species, mouse (*Mus musculus*) and rat (*Rattus norvegicus*) (Carleton and Musser 2005). These rodents seem to have originated in Southeast Asia and then rapidly expanded out of these regions (Steppan et al. 2005). Watts and Baverstock (1995) concluded that much of the murine radiation took place as a consequence of range expansion across the Old World (Africa, Asia and Europe) followed by formation of geographic barriers to gene flow and any subsequent dispersal. Tropical Africa, Southeast Asia, and Australia/New Guinea (Australo-Papua), correspond to the main centres of murine diversity, each area displaying its own characteristic groups that may represent clades (Watts and Baverstock 1995). There are accepted, at least, two groups in Africa, the *Praomys* group and the *Arvicanthine* group (Steppan et al. 2005).

Paleontological data indicate that the genera from the *Praomys* group were originated from the Miocene genus *Progonomys*, which inhabited regions of the Southeast Asia (Jacobs 1978). Presently, the phylogenetic relationships within the *Praomys* group are poorly resolved, however it is accepted that this group include the *Praomys* genus (Lecompte et al. 2008). Nicolas et al. (2005), according to morphological and molecular data, considered that the diversification of species from the genus *Praomys* occurred during the adaptive radiation

that took place in the Pliocene (~2.5 - 5.3 My), as consequence of climatic changes in Africa. To this genus belongs *Praomys tullbergi*, which inhabits African intertropical regions (Lecompte et al. 2005).

I.3- Karyotype features of rodents

The analysis of rodent karyotypes and the ascertainment of its evolutionary chromosome changes may benefit rodent phylogenetic resolution, since the conservation of these changes offer strong signatures of common ancestry (e.g. Robinson et al. 2008). In a general perspective, the karyotypes of rodent species present extreme variability in the diploid chromosome number, ranging from 2n = 10 in *Akodon sp.* (Silva and Yonenaga-Yassuda 1998) to 2n = 102 in *Tympanoctomys barrerae* (Svartman et al. 2005). This indicates that rodent genomes experienced a rapid chromosomal evolution, compared to other mammals (Veyrunes et al. 2007). In some species it is also possible to find B-chromosomes (Trifonov et al. 2002), chromosomal polymorphisms, unusual sex chromosome systems (e.g. *Ellobius lutescens*) (e.g. Koop et al. 1983, Fagundes et al. 1998, Vogel et al. 1998) and a significant heterochromatic variation (Patton and Sherwood 1982, Graphodatsky 1989, Svartman et al. 2005). This high karyotipic diversity is accompanied by a similar variation in the rates of karyotype evolution that can be found in the phylogenetic tree of Rodentia. For instance, it was proposed that the karyotypes of Sciuridae are highly conserved and close to the hypothesized ancestral Rodentia karyotype (Beklemisheva et al. 2011), with 50 - 48 chromosomes (Graphodatsky et al. 2008, Beklemisheva et al. 2011). In sharp contrast, the Myomorph karyotypes are extensively reorganized (Yang et al. 2000, Romanenko et al. 2007, Sitnikova et al. 2007). Even within a single family it is not unusual to find genera with low levels of reorganization and those whose genomes are extensively rearranged (e.g. Veyrunes et al. 2006).

I.3.1- Muroidea karyotypes evolution

The Muroidea karyotypes' divergence from the putative ancestral state involves centric fusions, fissions, addition of heterochromatin and a great number of inversions. Based solely in mouse chromosome syntenies, it was suggested that the ancestral Muroidea karyotype presented a diploid number of 2n = 52 (Romanenko et al. 2012). However in a more recent work that considered both mouse and rat chromosome syntenies, the diploid chromosome number for the Muroidea ancestor was revised for 2n = 50 (Chaves et al. 2012).

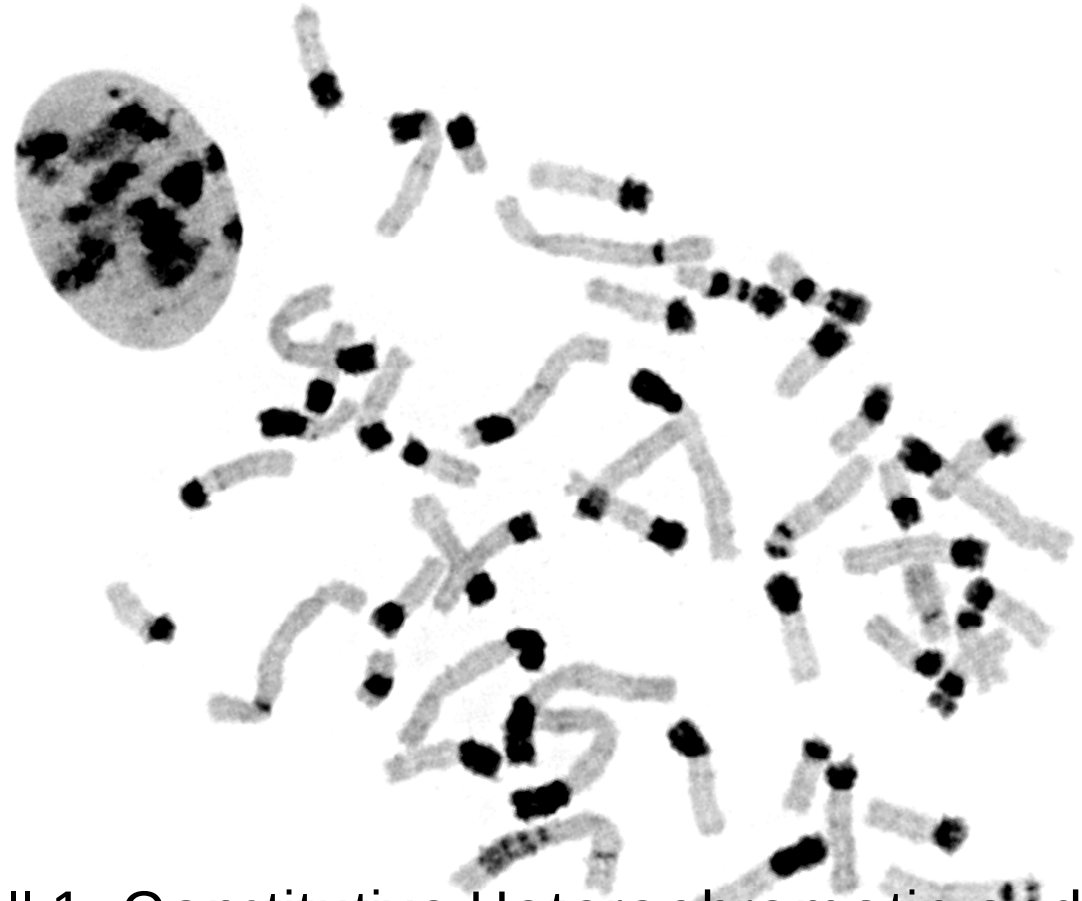
The evolution of the Cricetidae and Muridae genus where are classified the species studied here, *Cricetus*, *Phodopus*, *Peromyscus* and *Praomys*, is marked by the preferential occurrence

of specific chromosomal rearrangements. A recent work (Romanenko et al. 2012) reported that fusion and fission events were the predominant chromosomal rearrangements shaping the karyotypes of *Cricetus* and *Phodopus* species. Within *Peromyscus* a high degree of karyotypic conservation is observed, presenting all species 2n = 48, varying the number of chromosomal arms from 52 to 92 by the result of heterochromatin additions and pericentric inversions (Robbins and Baker 1981, Rogers et al. 1984). According to different works, it seems reasonable to consider that the karyotype of *Peromyscus eremicus* is close to the putative ancestral Muroidea karyotype (Romanenko et al. 2007, Vieira-da-Silva and Louzada et al. *unpublished data*). Cytogenetic data within *Praomys* genus are still very scarse, but it is suggested that *Praomys tullbergi* karyotype evolved from the Muroidea ancestral mostly by translocations (tandem and robertsonian translocations) and fissions (Chaves et al. 2012).

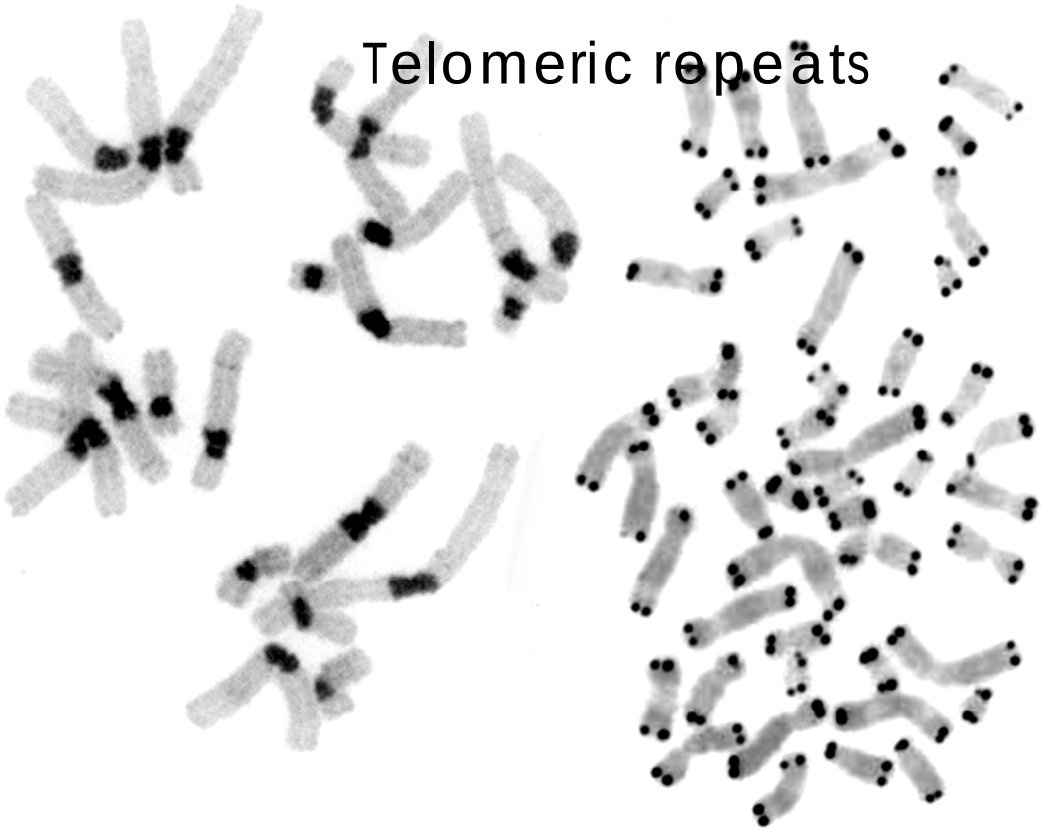
The five species studied in the present work present very distinct karyotypes, what makes them good models for studying the genome evolution of muroids. The karyotype of *Cricetus cricetus* has 22 chromosomes, being the first description performed by Matthey (1952). This karyotype comprises five meta/submetacentric, four submetacentric and one acrocentric chromosome pairs. The X and Y-chromosomes are large meta/submetacentric (Gamperl et al. 1976). The karyotype of *Phodopus roborovskii* comprises 34 chromosomes, and was firstly outlined by Vorontsov and Krjukova (1969). In this karyotype it is possible to observe six meta/submetacentric, six submetacentric and four acrocentric chromosome pairs. The X-chromosome is submetacentric and the Y-chromosome is acrocentric (O'Brien et al. 2006). *Phodopus sungorus* exhibits a diploid chromosome number of 2n = 28 and the karyoptype of this species was firstly studied by Matthey (1960). In this karyotype it is possible to identify six meta/submetacentric chromosome, five submetacentric and two acrocentric pairs. The X-chromosome is submetacentric and the Y-acrocentric (O'Brien et al. 2006). Comparing both *Phodopus* species, it was suggested that *P. roborovskii* retains a more primitive karyotype than *P. sungorus*, differing from the *Phodopus* ancestor by three and six rearrangements, respectively (Schmid et al. 1986, Romanenko et al. 2007). The karyotype of *Peromyscus eremicus* exhibits 48 chromosomes, as referred, and was firstly characterized by Hsu and Arrighi (1966). In this species, all of the chromosomes are submetacentric (O'Brien et al. 2006). The karyotype of *Praomys tullbergi* has 34 chromosomes in which all of the autosomes are acrocentric, the Y-chromosome is a small acrocentric and the X-chromosome is a large submetacentric (Meles et al. 2007). The first report of this karyotype was performed by Matthey (1958).

Chapter II

Results and Discussion



II.1- Constitutive Heterochromatin and Telomeric repeats



In this section are presented two papers where a detailed analysis of Constitutive Heterochromatin (CH) in chromosomes of five rodent species, *Cricetus cricetus*, *Peromyscus eremicus*, *Phodopus roborovskii*, *Phodopus sungorus* (Cricetidae) and *Praomys tullbergi* (Muridae) was performed. For this CH analysis, chromosomes were submitted to *in situ* restriction endonuclease digestion with sequential C-banding. This methodology proved to be a useful technique in improving the study of heterochromatin (Gosálvez et al. 1997, Pieczarka et al. 1998), allowing the characterization of CH in terms of its chromosome location and molecular nature, what gives a perception about the CH molecular heterogeneity in these genomes (Rocco et al. 2002, Schmid et al. 2002, Chaves et al. 2004, Adega et al. 2005). Moreover, this technique allows the identification of CH bands not detected by conventional C-banding, named cryptic C-bands (Chaves et al. 2004, Adega et al. 2005, 2007). The CH analysis data obtained here for *C. cricetus*, *P. eremicus* and *P. tullbergi* was used in other works performed in the host lab (Chaves et al. 2012, Vieira-da-Silva and Louzada et al. *unpublished data*), which found that a high percentage of the identified breakpoint regions in these species' chromosomes are coincident with CH regions. Particularly for the two studied *Phodopus* species, it was verified a high CH coincidence with evolutionary breakpoint regions identified in the karyotypes. For these two species chromosomes, it was also performed an analysis on the telomeric repeats distribution. Both species present telomeric sequences located at chromosome ends (as expected) and interstitially (ITS), as short or large ITS blocks. The number and degree of ITSs amplification varies greatly in the two hamsters, indicating independent evolutionary events for these repeats in each genome.

The main goal of these works was to contribute for the understanding of the repetitive sequences role in the occurrence of chromosomal rearrangements. All the obtained results clearly suggest an involvement of the repetitive genomic fraction in karyotypes reshaping, allowing also establishing evolutionary considerations on the studied species chromosomes.

II.1.1

Hidden heterochromatin: Characterization in the Rodentia species *Cricetus cricetus*, *Peromyscus eremicus* (Cricetidae) and *Praomys tullbergi* (Muridae)

Genetics and Molecular Biology, 32, 1, 58-68 (2009)

www.sbg.org.br

Research Article

Hidden heterochromatin: Characterization in the Rodentia species *Cricetus cricetus*, *Peromyscus eremicus* (Cricetidae) and *Praomys tullbergi* (Muridae)

Ana Paço, Filomena Adega, Henrique Guedes-Pinto and Raquel Chaves

Institute for Biotechnology and Bioengineering, Centre of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

Abstract

The use of *in situ* restriction endonuclease (RE) (which cleaves DNA at specific sequences) digestion has proven to be a useful technique in improving the dissection of constitutive heterochromatin (CH), and in the understanding of the CH evolution in different genomes. In the present work we describe in detail the CH of the three Rodentia species, *Cricetus cricetus*, *Peromyscus eremicus* (family Cricetidae) and *Praomys tullbergi* (family Muridae) using a panel of seven REs followed by C-banding. Comparison of the amount, distribution and molecular nature of C-positive heterochromatin revealed molecular heterogeneity in the heterochromatin of the three species. The large number of subclasses of CH identified in *Praomys tullbergi* chromosomes indicated that the karyotype of this species is the more derived when compared with the other two genomes analyzed, probably originated by a great number of complex chromosomal rearrangements. The high level of sequence heterogeneity identified in the CH of the three genomes suggests the coexistence of different satellite DNA families, or variants of these families in these genomes.

Key words: constitutive heterochromatin, *in situ* restriction endonuclease digestion, Rodentia.

Received: April 14, 2008; Accepted: July 14, 2008.

Introduction

Constitutive heterochromatin (CH) is a ubiquitous and abundant component of eukaryotic genomes that accounts for ~30% of the genome in humans and up to 50% in the kangaroo rat (*Dipidomys ordii*) (Singer, 1982; Dimitri *et al.*, 2004, 2005; Rossi *et al.*, 2007). The similarity in the genetic and molecular properties of CH among plants and animals, led to the traditional view of this genome fraction as a "genomic wasteland" or a repository of "junk" DNA (John, 1988). Nowadays this idea is becoming obsolete; in fact, in the past two decades molecular genetics studies have implicated CH in important cellular functions, in a remarkable structural and functional basis (Dimitri *et al.*, 2004, 2005; Corradini *et al.*, 2007; Rossi *et al.*, 2007). Constitutive heterochromatin can occur as large blocks or discrete C-positive bands in any part of a chromosome, but is most commonly found in large blocks near the centromere (Corradini *et al.*, 2007; Probst and Almouzni, 2008). Satellite DNA, the main constituent of this genomic fraction, usually occurs in the centromeric region of chromosomes (Chaves *et al.*, 2000), but is also frequently found at telomeres (Shore, 2001). The occurrence of CH at interstitial positions is much less common, although large blocks of interstitial CH have been found in the large chromosomes of some insects (John *et al.*, 1985), plants (Bauchan and Hossain, 1999) and some mammals (Santos *et al.*, 2004; Adega *et al.*, 2007; Meles *et al.*, 2008). Although present in almost all eukaryotes, the sequence and chromosomal organization of CH is not well conserved among species. Indeed, there is strong evidence for the sharing of homologous satellite DNA sequences by closely related species (Waye and Willard, 1989; Jobse *et al.*, 1995; Lee *et al.*, 1999; Saito *et al.*, 2007), with species-specific sequences of satellite DNA occurring in almost all taxonomic groups (Slamovits and Rossi, 2002). It seems reasonable to accept that the presence of CH facilitates the occurrence of

Send correspondence to Raquel Chaves. Institute for Biotechnology and Bioengineering, Centre of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal.
E-mail: rchaves@utad.pt.

chromosome rearrangements, as it is in accordance with several authors that consider CH as hotspots for structural chromosome rearrangements (Yunis and Yasmineh 1971; Peacock *et al.*, 1982; John, 1988; Chaves *et al.*, 2004b). Wichman *et al.* (1991) postulated that rapidly evolving families or variants of satellite DNA can promote chromosomal rearrangements via of their intragenomic movements among non-homologous chromosomes and between different chromosomal regions such as centromeres, arms and telomeres. Sequences of CH can be easily detected by the preferential "loss" of DNA from non-C-band regions of chromosomes (Comings, 1973; Pathak and Arrighi, 1973), achieved by conventional C-banding, involving depurination and denaturation of chromosomal DNA (Arrighi and Hsu, 1971; Sumner, 1972) followed by its extraction during incubation in a saline solution (Holmquist and Dancis, 1979; Verma and Babu, 1995). Nevertheless other analytical methodologies are indispensable when a detailed molecular characterization of CH is the central issue. The use of *in situ* restriction endonuclease (RE) digestion proved to be a very useful technique in improving the dissection of CH, and in the understanding of the CH evolution in different genomes (Gosálvez *et al.*, 1997; Pieczarka *et al.*, 1998). Besides the ability of REs followed by C-banding in demonstrating the C-heterochromatin heterogeneity (Rocco *et al.*, 2002; Schmid *et al.*, 2002; Chaves *et al.*, 2004b; Adega *et al.*, 2005). In this work, we used seven restriction endonucleases followed by C-banding to study the heterochromatin of three Rodentia species, *Cricetus cricetus*, *Peromyscus eremicus* (family Cricetidae) and *Praomys tullbergi* (family Muridae). In rodents' chromosomes, *in situ* REs digestion was only applied without sequential C-banding and only in *Microtus savii* (Galleni *et al.*, 1992), species from the genus *Reithrodontomys* (Van Den Bussche *et al.*, 1993) (family Muridae) and from the genus *Ctenomys* (family Octodontidae) (García *et al.*, 2000a, 2000b). The approach used here allowed a detailed CH characterization in terms of its location, detection of different CH subclasses, and revelation of its molecular composition.

Materials and Methods

Chromosome preparations

The material analyzed consisted of chromosomal preparations of *Cricetus cricetus* (CCR), *Peromyscus eremicus* (PER) and *Praomys tullbergi* (PTU), prepared from fibroblast cell lines obtained from the cell and tissue collection maintained at the Department of Systematics and Evolution, Muséum National d'Histoire Naturelle (MNHN), Paris. Standard cell culture was followed as described elsewhere (Chaves *et al.*, 2004a) in order to prepare fixed chromosome spreads.

GTD-banding

Air dried slides were aged at 65 °C for 5 h or overnight and then subjected to standard G-banding procedures with trypsin (Seabright, 1971). The same slides were subsequently fixed with formaldehyde and used for C-banding (Chaves *et al.*, 2002). Slides were stained with DAPI for a better contrast. The inversion of the DAPI color in Adobe Photoshop (version 7.0) revealed the chromosomes G-banding pattern (GTD-banding, G-bands by trypsin with DAPI).

In situ RE digestion

Air dried slides were aged at 65 °C for 6 h and then submitted to *in situ* restriction enzyme (RE) digestion. The seven restriction enzymes used (AluI, ApaI, BamHI, DraI, HaeIII, PstI and RsaI) were diluted in buffers indicated by the manufacturer (Invitrogen Life Technologies) to give final concentrations of 30U per 100 µL of solution. One hundred microliters of the desired solution was placed on slides that were then covered with coverslips and incubated in a moist chamber for 16 h at 37 °C. Control slides were incubated only with buffer under the same conditions. Prior to C-banding, the slides were fixed with formaldehyde. Finally, the slides were stained with DAPI (the inversion of the DAPI color revealed the RE-banding). The residual bands obtained after the endonuclease digestion were suitable for chromosome identification and karyotype organization.

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

The C-banding technique was performed sequentially to G-bands or to RE banding and was carried out after distaining the slides. CBP-banding (C-bands by barium hydroxide using propidium iodide) was done using the standard procedure of Sumner (1972), but with propidium iodide as counterstain. The results presented below are representative of at least 35 metaphases from at least five independent experiments done for each endonuclease.

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). Digitized photos were prepared for printing in Adobe Photoshop (version 7.0); contrast and color optimization were the functions used and affected the whole of the image equally.

Results and Discussion

The karyotype of *Cricetus cricetus* has 22 chromosomes, being the first description performed by Matthey (1952). This karyotype is composed by five meta/submetacentric chromosome pairs, four submetacentric and one acrocentric, being the X chromosome a large meta/submetacentric. The karyotype of *Peromyscus eremicus* has 48 chromosomes, in agreement with the initial description by Hsu and Arrighi (1966). In this species, all of the chromosomes are submetacentric, being the X chromosome a large submetacentric and the Y a small submetacentric. The karyotype of *Praomys tullbergi* has 34 chromosomes in which all of the autosomes are acrocentric, the Y chromosome is a small acrocentric and the X chromosome is a large submetacentric (Matthey, 1958; Qumsiyeh *et al.*, 1990; Capanna *et al.*, 1996; Meles *et al.*, 2008). The first description of this karyotype was reported by Matthey (1958). The action of all seven different REs and REs+Cbanding on *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi* chromosomes are presented in Figures 1,

2 and 3, respectively. The residual bands seen after digestion with endonucleases AluI, ApaI, BamHI, DraI, HaeIII, PstI and RsaI (left column for each enzyme shown in Figures 1-3) are mainly G-like and suitable for chromosome identification. Although each restriction endonuclease was expected to yield a specific banding pattern, in practice most of the banding patterns overlapped. Nevertheless some endonucleases (*e.g.* ApaI, PstI and RsaI in chromosomes of *Cricetus cricetus*, BamHI, PstI and RsaI in *Peromyscus eremicus* and HaeIII, PstI and RsaI in *Praomys tullbergi*) produced a higher banding contrast. AluI was, perhaps, the used enzyme that produced the smallest number of bands but the higher contrast banding pattern. It is important to refer that the banding patterns produced by each RE are reproducible and can be used in sequential experiment procedures without loss of chromosome morphology (Chaves *et al.*, 2002; Adega *et al.*, 2005). In a general overview, the C-positive heterochromatin (Figures 1-3, right chromosome in each column, showing control C-banding and RE+C-banding) is mainly found at the centromeres of most chromosomes, although some C-bands can also be seen at interstitial and telomeric locations. In the individuals analyzed, some heterochromatin polymorphism of minor significance were detected, *i.e.*, variation in the banding patterns of homologous chromosomes of the same pair, as also reported for pig (Adega *et al.*, 2005) and some Tayassuidae species (Adega *et al.*, 2007) chromosomes. The heterochromatin polymorphisms detected in the chromosomes of the studied species were not considered for the analysis relatively to the characterization of CH here presented, because they might not be representative of the population. At least three major classes of CH were identified in the species studied in this work: (peri)centromeric, interstitial and telomeric (Figures 1-3). With RE+C-banding treatment, these major C-positive heterochromatin blocks could be discriminated in at least 26 C-positive heterochromatin subclasses in the autosomal complement of *Cricetus cricetus* [seven in (peri)centromeric regions, 13 in interstitial regions and six in telomeric regions] and three C-positive heterochromatin subclasses in the CCRX chromosome [one (peri)centromeric and two in interstitial regions] (cf. Figure 1). In *Peromyscus eremicus* chromosomes (Figure 2), the RE+C-banding treatment discriminated at least 26 C-positive heterochromatin subclasses in the autosomal complement [seven in (peri)centromeric regions, 13 in interstitial regions and six in telomeric regions], three C-positive heterochromatin subclasses in the PERX chromosome (one in the centromeric region and two in interstitial regions) and two in the PERY chromosome (one centromeric and one subtelomeric). Finally, in *Praomys tullbergi*, the RE+C-banding treatment (Figure 3) discriminated the major C-positive heterochromatin blocks into at least 45 C-positive heterochromatin subclasses in the autosomal complement (two in centromeric regions, 35 in interstitial

regions and eight in telomeric regions), four C-positive heterochromatin subclasses in the PTUX chromosome (one in the centromeric region and three in interstitial regions) and three in the PTUY chromosome (one in the centromeric region and two in interstitial regions).

Constitutive Heterochromatin (C-positive heterochromatin) characterization in *Cricetus cricetus*

Control experiment (G+C-banding) show that all the chromosomes of *Cricetus cricetus* exhibit large (peri)centromeric C-bands that in most cases consist of two blocks of CH (exception goes to CCR7, CCR8 and CCR10 chromosomes which show only one block of CH). Notice the very large centromeric CH block of the only acrocentric chromosome of the karyotype, CCR7. All the chromosomes except CCR3, CCR8 and CCR10 exhibit interstitial C-positive heterochromatin.

Telomeric C-bands can be seen on chromosomes CCR1, CCR3, CCR5, CCR6, and CCR9. Incubation of this species chromosomes with restriction endonucleases followed by C-banding revealed C-bands heterogeneity (Figure 1), being verified that (peri)centromeric, interstitial or telomeric C-bands present a different molecular nature, exhibiting different restriction patterns when submitted to the same panel of REs. This is not surprising as similar results have been reported for other species (Babu, 1988; Fernández-García *et al.*, 1998; Chaves *et al.*, 2004b; Adega *et al.*, 2005, 2007). The arrowheads in Figure 1 indicate C-bands revealed only after RE treatment (cryptic C-bands). Of the endonucleases used here, BamHI+C-banding was the one that produced

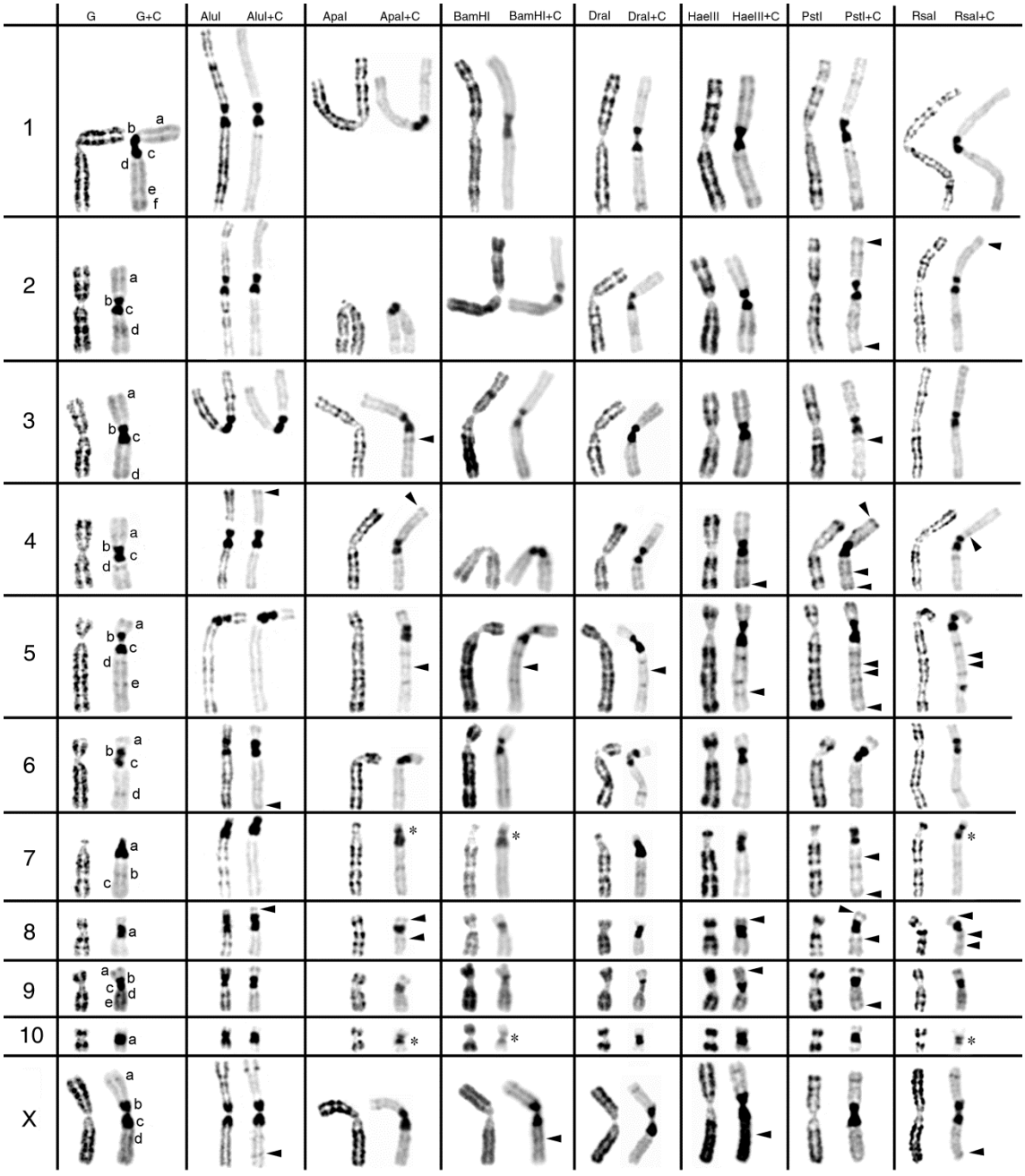


Figure 1 - Table resume of *in situ* restriction endonuclease digestion of *Cricetus cricetus* chromosomes (2n = 22) and sequential C-banding. Control G and C-banding of *Cricetus cricetus* chromosomes are shown on the left column. The other columns show the bands produced by the seven restriction endonucleases before and after C-banding. The letters (a-f) represent the C-bands according to their order of appearance in each chromosome. Arrowheads indicate C-positive heterochromatin bands only revealed by previous RE treatment. Asterisks indicate extra C-bands produced by the splitting of a control C- band after endonuclease digestion+C-banding.

the most evident effect in CH sequences of the *Cricetus cricetus* chromosomes. See for instance chromosomes CCR7, CCR8, CCR9 and CCR10, being observed less intense bands in comparison with the control chromosomes. This enzyme, along with ApaI+C-banding and RsaI+C-banding, produ-

ced the partition of the (peri)centromeric CH band at chromosomes CCR7 and CCR10 into two distinct CH blocks, thus revealing the occurrence of two instead of one (peri)centromeric CH block [bands identified with an asterisk in Figure 1]. Some enzymes seem to have a drastic effect resulting in a more accentuated contrast pattern in the (peri)centromeric regions of some chromosomes. See, for example, chromosomes CCR1 and CCR6 with DraI+C-banding, CCR5 with BamHI+C-banding, CCR9 with BamHI+C-banding and DraI+C-banding.

Constitutive heterochromatin (C-positive heterochromatin) characterization in *Peromyscus eremicus*

In the control experiment (G+C-banding) the majority of *Peromyscus eremicus* chromosomes exhibit large (peri)centromeric C-bands (Figure 2,

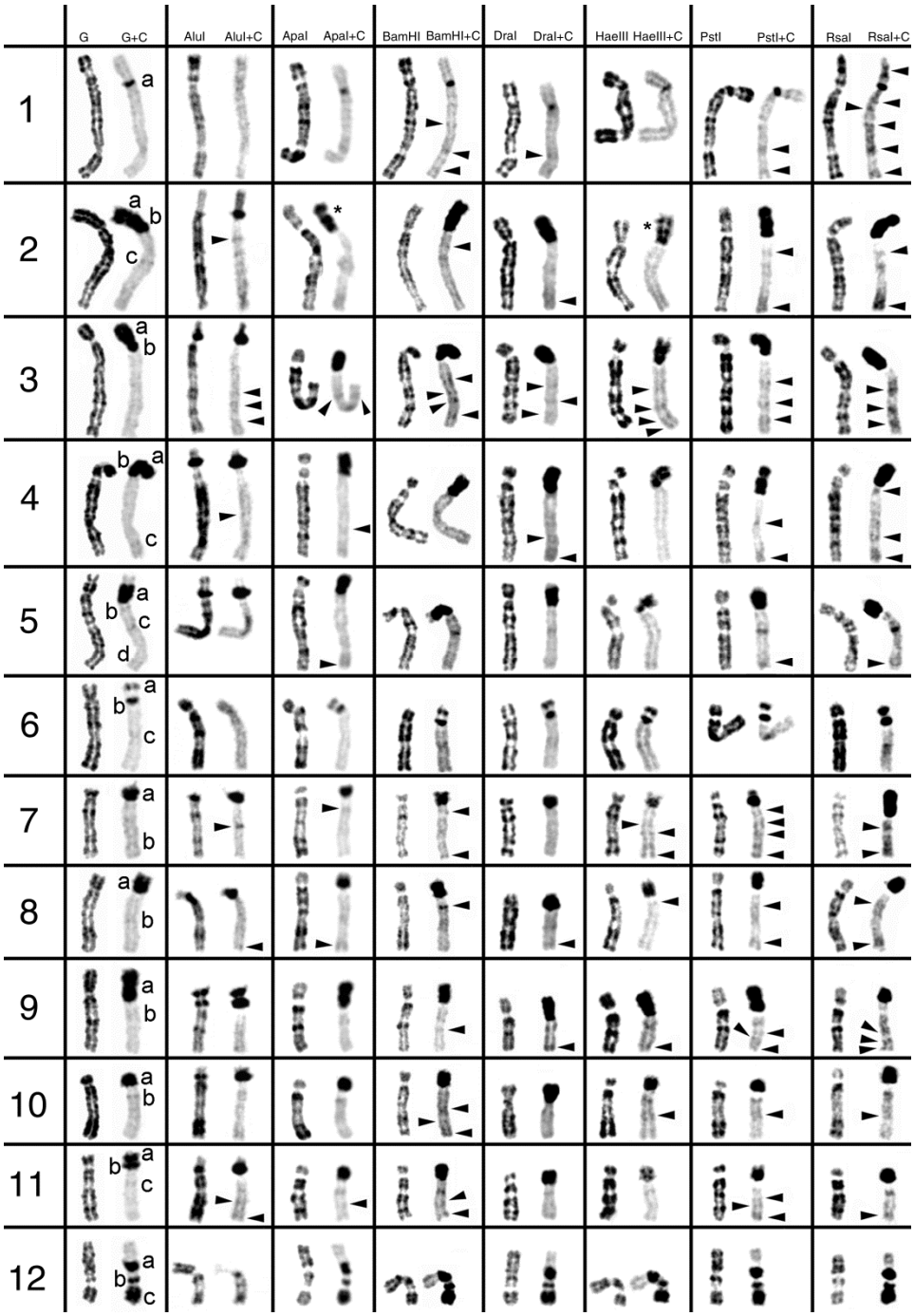


Figure 2 - Table resume of *in situ* restriction endonuclease digestion of *Peromyscus eremicus* chromosomes (2n = 48) and sequential C-banding. Control G- and C-banding of *Peromyscus eremicus* chromosomes are shown on the left column. The other columns show the bands produced by the seven restriction endonucleases, before and after C-banding. The letters (a-g) represent the C-bands according to their order of appearance in each chromosome. Arrowheads indicate C-positive heterochromatin bands only

revealed by previous RE treatment. Asterisks indicate extra C-bands produced by the splitting of a control C- band after endonuclease digestion+C-banding.

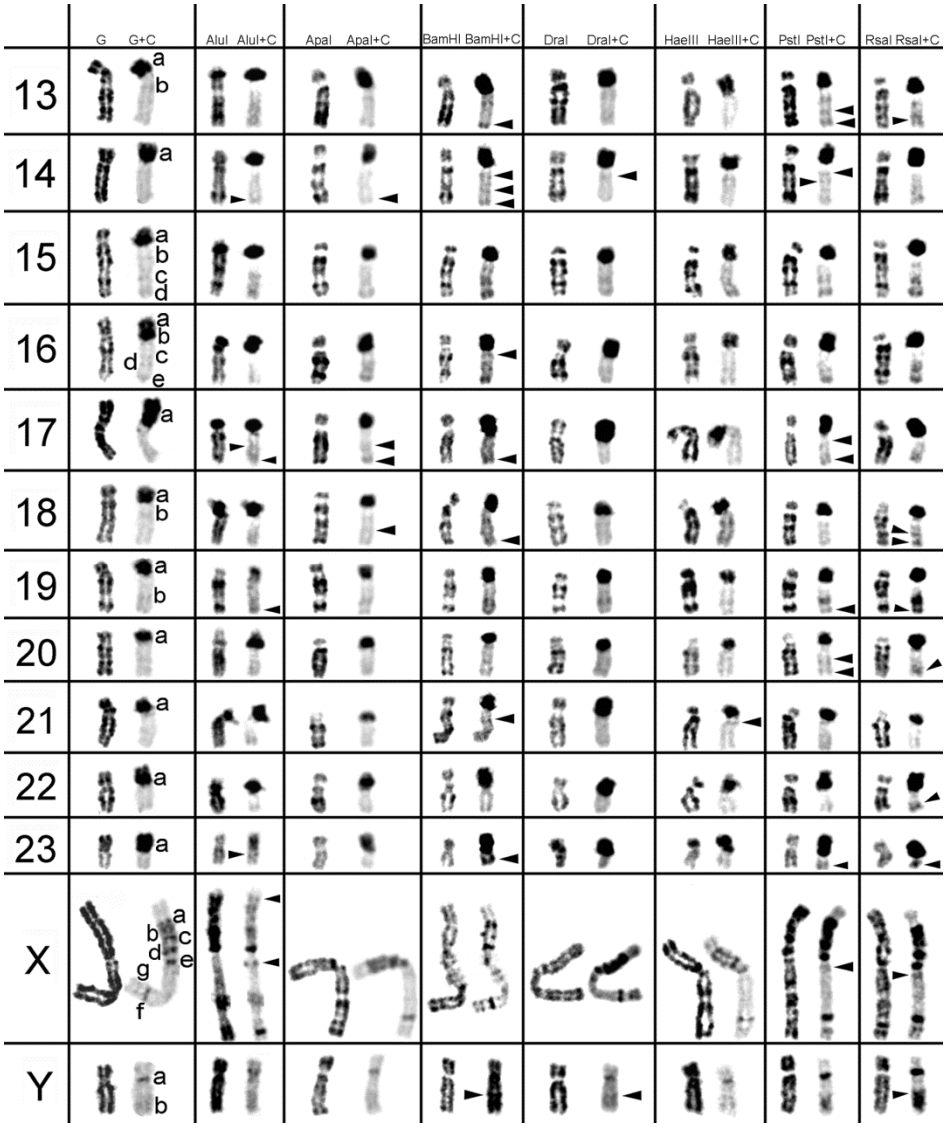


Figure 2 (cont.) - Table resume of *in situ* restriction endonuclease digestion of *Peromyscus eremicus* chromosomes (2n = 48) and sequential C-banding. Control G- and C-banding of *Peromyscus eremicus* chromosomes are shown on the left column. The other columns show the bands produced by the seven restriction endonucleases, before and after C-banding. The letters (a-g) represent the C-bands according to their order of appearance in each chromosome. Arrowheads indicate C-positive heterochromatin bands only revealed by previous RE treatment. Asterisks indicate extra C-bands produced by the splitting of a control C- band after endonuclease digestion+C-banding.

left column), and in some of these chromosomes, the C banding spreads from the centromeric region to the p arm telomere, apparently covering all the p arm, *e.g.*, chromosomes PER9 and PER17. In some chromosomes, this band seems to be split in two C-bands, one clearly centromeric and the other covering the chromosome p arm (chromosomes PER2, PER3 and PER4). Chromosomes PER11 and PER16 display two well-defined bands of (peri)centromeric CH, although this may have been an artifact caused by the small size of the p arms. Chromosomes PER1 and PERY apparently display the lowest amount of heterochromatin in control G+Cbanding, showing PER1 only a small centromeric CH band. The situation observed in the PERY is not usual for most of the mammals' species, once this chromosome is usually the more heterochromatic of the whole complement. Some of the chromosomes exhibit C-bands at interstitial locations, presenting chromosome PERX the highest number of these bands (at least six). Telomeric C-bands can be observed in some chromosomes of this species, *e.g.*, PER6, PER11, PER12, PER15 and PER16 (Figure 2). When C-banding was applied after *in situ* REs digestion to the chromosomes of this species, it was possible to verify that its CH shows some degrees of heterogeneity (Figure 2). The arrowheads in Figure

2 indicate C-bands revealed only after treatment with endonucleases (cryptic C-bands). From the REs used in this work, RsaI+Cbanding, PstI+C-banding and BamHI+C-banding, were the enzymes that revealed the greatest number of CH bands not previously detected by the control G+C-banding. In a general analysis, AluI was the enzyme that produced the most divergent effects on the CH of *Peromyscus eremicus* chromosomes. In some cases, such as in chromosomes PER1 and PER6, some C-bands seem to have undergone a greater reduction or even have, apparently disappeared when compared with control experiment, while in other cases, such as chromosomes PER7 and PER16, the CH was apparently unaffected by treatment with this enzyme. The p arms CH of PER2, PER3 and PER4 chromosomes are particularly interesting in what respects to its molecular nature. In these heterochromatic arms the CH reveals a high heterogeneity, what is verified by the different restriction patterns produced by the enzymes at these CH regions. For instance in the p arm of PER 2 there were recognized two C-bands in the control G+C-banding; after AluI+C restriction a lesser intensity of one of these bands was observed and ApaI+C-banding and HaeIII+C-banding seem to reveal an extra C-band, by splitting one of the previous in two [bands evidenced with an asterisk (*) in Figure 2].

Constitutive heterochromatin (C-positive heterochromatin) characterization in *Praomys tullbergi*

From the studied species, *Praomys tullbergi* (Figure 3) is the one whose chromosomes exhibit the lower amount of centromeric CH in the control experiment (G+Cbanding). In some chromosomes, centromeric CH is almost as abundant as interstitial CH, in contrast to the observed for the majority of the chromosomes from the other species here analyzed. However, the chromosomes PTU5 and PTU10 in the control experiment, present a small centromeric CH band and apparently do not reveal interstitial bands. The majority of the chromosomes display several interstitial CH bands, presenting the chromosomes PTU1 and PTU2 the greatest number of these bands. Telomeric C-bands are clearly distinguishable in some chromosomes, *e.g.*, chromosomes PTU10, PTU12 and PTU15. The PTUX chromosome presents three distinct classes of CH, centromeric, interstitial and telomeric. PTUY chromosome exhibits a centromeric band and two interstitial C-bands. When C-banding was applied after *in situ* REs digestion to the chromosomes of this species, it was possible to verify that its CH shows some degrees of heterogeneity, just as it was described for the other two rodent species studied in this work. AluI+C-banding produced the higher contrast between the centromeric *versus* interstitial/telomeric CH classes; digestion with AluI greatly decreased the interstitial/telomeric CH while, simultaneously, evidenced the centromeric heterochromatin. See for instance, chromosomes PTU15 or PTU16, whose centromeres showed in the control G+C-banding an almost absence of CH, and after the AluI+C-banding the centromeres showed large centromeric CH blocks. Digestion with DraI seems to highlight the telomeric CH after C-banding, *e.g.*, chromosomes PTU7 and PTU16. RsaI+C-banding seems to produce the most similar results with the control G+C-banding, however also discriminating cryptic C-bands, such as the ones observed in chromosomes PTU4, PTU5, PTU10 or PTUX. Other endonucleases also disclosed cryptic C-bands, especially DraI+C-banding, BamHI+C-banding or HaeIII+Cbanding. These special bands are very interesting from the CH molecular nature point of view, since their disclosure is probably dependent on sequence modifications (not yet clearly understood) induced by the REs, leading for instance, to an increase of the stain capacity to bind a specific chromosome region (Gosálvez *et al.*, 1997; Nieddu *et al.*, 1999; Chaves *et al.*, 2004b). Whatever the mechanism behind these sequences modification, RE digestion triggers it, revealing "hidden" C-bands. Curiously, and from several different works in different species, these sequences not detected by classical C-banding have proven to correspond to clinical (*Sus scrofa*, Adega *et al.*, 2005) or evolutionary breakpoints (Tayassuidae, Adega *et al.*, 2007).

Inter-species constitutive heterochromatin (C-positive heterochromatin)

A general comparison of the amount, distribution and molecular nature of C-positive heterochromatin in the three Rodentia species, suggests that the CH of these karyotypes is extremely different. Evidence comes from the detailed combined analysis of the different REs+C-banding patterns disclosed on the karyotypes of these species. The application of a seven REs panel to the chromosomes of three different rodent species, *Cricetus cricetus*, *Peromyscus eremicus* (Cricetidae) and *Praomys tullbergi* (Muridae), allowed a characterization of its CHand the recognition of its molecular heterogeneity. These results are a clear reflex of the different C-positive heterochromatin composition of these karyotypes, possible to observe by the different REs actions on the respective chromosome's bands. *Cricetus cricetus* has an almost entirely meta/submetacentric

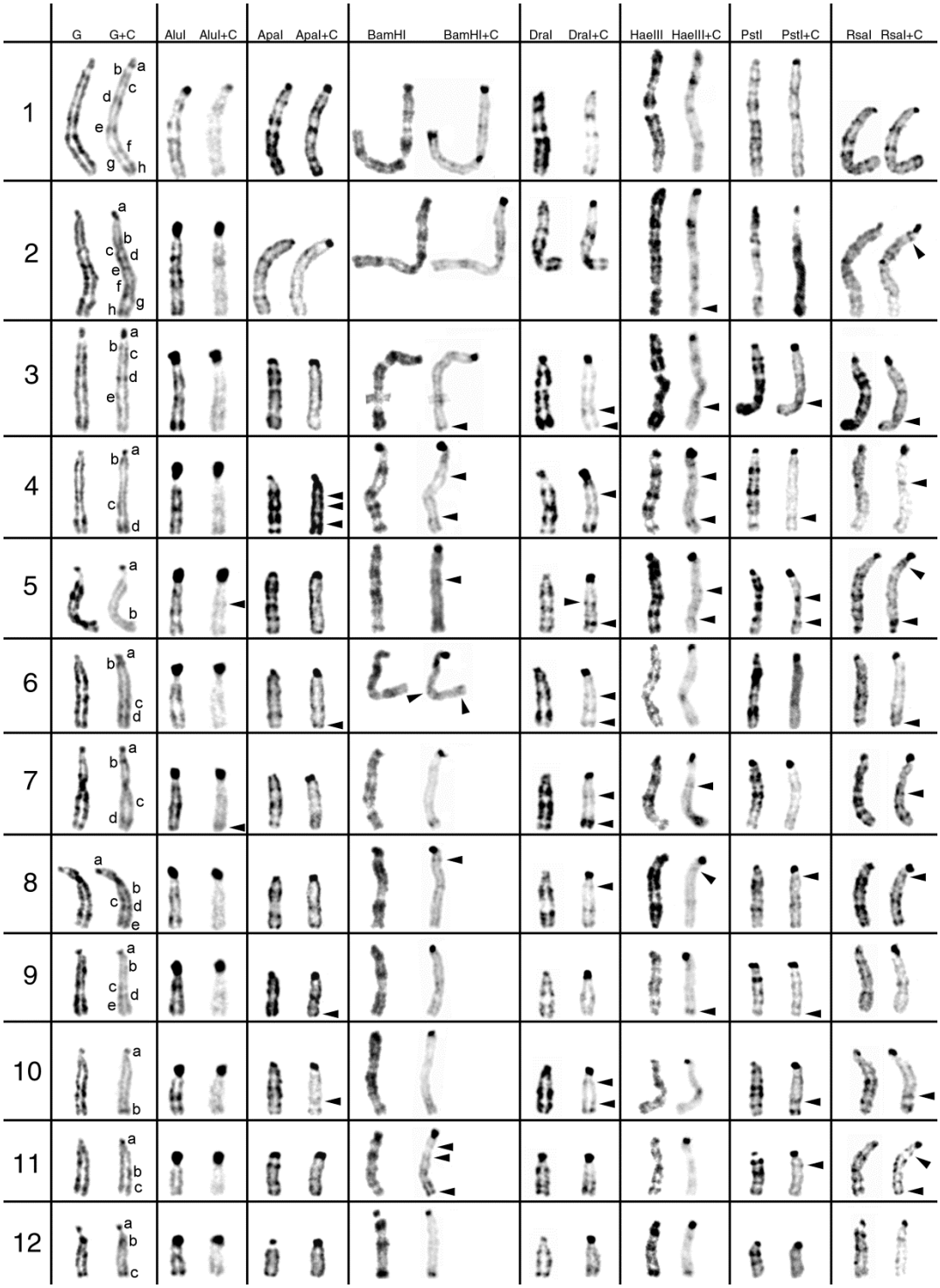


Figure 3 - Table resume of *in situ* restriction endonuclease digestion of *Praomys tullbergi* chromosomes (2n = 34) and sequential C-banding. Control G and C-banding of *Praomys tullbergi* chromosomes are shown on the left column. The other columns show the bands produced by the seven different restriction endonucleases, before and after C-banding. The letters (a-h) represent the C-bands according to their order of appearance in each chromosome. Arrowheads indicate C-positive heterochromatin bands only revealed by previous RE treatment.

karyotype (with only one acrocentric pair), with the CH primarily located in (peri)centromeric region. Most of the chromosomes in this species exhibit two very large blocks at (peri)centromeric location, which suggested the occurrence of dicentric Robertsonian translocations or, alternatively, heterochromatin additions during the course of this karyotype evolution. The other Cricetidae species, *Peromyscus eremicus*, has a very distinct karyotype that comprises only submetacentric chromosomes. This karyotype also displays great amounts of CH, especially located at

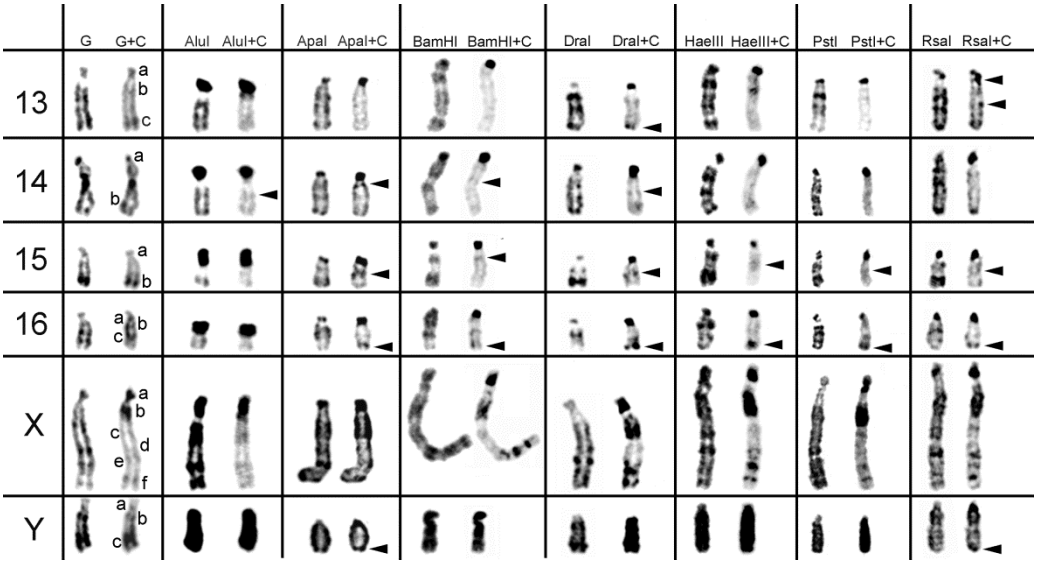


Figure 3 (cont.) - Table resume of *in situ* restriction endonuclease digestion of *Praomys tullbergi* chromosomes (2n = 34) and sequential C-banding. Control G- and C-banding of *Praomys tullbergi* chromosomes are shown on the left column. The other columns show the bands produced by the seven different restriction endonucleases, before and after C-banding. The letters (a-h) represent the C-bands according to their order of appearance in each chromosome. Arrowheads indicate C-positive heterochromatin bands only revealed by previous RE treatment.

the (peri)centromeric regions, being the p arms of some chromosomes composed entirely by this repetitive component of the genome. The heterochromatin of p arms revealed a great heterogeneity, what implies a different molecular composition, which is certainly indicative of the coexistence of different satellite DNA families or variants at these chromosome regions. The species *Praomys tullbergi*, with a complete acrocentric autosome complement, it is the one whose chromosomes exhibit the lower amount of centromeric CH in the control experiment (G+C-banding), and in some cases, interstitial heterochromatin is almost as abundant as centromeric heterochromatin. This uniform and scattered distribution, together with the higher number of CH subclasses identified in *Praomys tullbergi* chromosomes (52 subclasses) suggests that this species has a more derivative karyotype than the other two genomes analyzed, probably originated by a great number of complex chromosomal rearrangements. This is based on the assumption that heterochromatic rich regions act as hotspots for the occurrence of chromosome rearrangements (Yunis and Yasmineh, 1971; Peacock *et al.*, 1982, John, 1988; Chaves *et al.*, 2004b), either by promoting the chromosome structural rearrangements that reshape karyotypes or by being fragile regions prone to chromosome breakage, and consequently to chromosome rearrangement, representing remnants of these events. The suggestion that the karyotype of *Praomys tullbergi* was originated by the occurrence of a high number of complex chromosomal rearrangements is supported by the work of Meles *et al.* (2008), where it was detected telomeric interstitial sequences in several chromosome arms of this species, probably the result of tandem fusions. Finally, it is worth mentioning the value of *in situ* RE digestion with sequential C-banding as an alternative tool for the study of Rodentia chromosomes CH, especially when other techniques are not available, as fluorescent *in situ* hybridization with different repetitive sequences.

Acknowledgments

This work was supported by a project (POCI/BIABCM/58541/2004) and a PhD grant (SFRH/BD/ 41574/2007) from the Science and Technology Foundation (FCT), Portugal. We are deeply grateful Dr. Vitaly Volobouev for providing the cell cultures of the Rodentia species.

References

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

Arrighi FE and Hsu TC (1971) Localization of heterochromatin in human chromosomes. Cytogenetics 10:81-86.

Babu A (1988) Heterogeneity of heterochromatin of human chromosomes as demonstrated by restriction endonuclease treatment. In: Verma RS (ed) Heterochromatin: Molecular and Structural

Aspects. 1st edition. Cambridge University Press, New York, pp 250-275.

Bauchan GR and Hossain MA (1999) Constitutive heterochromatin DNA polymorphisms in diploid *Medicago sativa* ssp. *falcata*. Genome 42:930-935.

Capanna E, Codjia JTC, Chrysostome C and Civitelli MV (1996) Les chromosomes des rongeurs du Benin (Afrique de l'Ouest): 3 Murinae. Rend Fis Acc Lincei 8:25-37.

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, Adega F, Santos S, Guedes-Pinto H and Heslop-Harrinson JS (2002) *In situ* hybridization and chromosome banding in mammalian species. Cytogenet Genome Res 96:113-116.

Chaves R, Frönicke L, Guedes-Pinto H and 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 and 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.

Comings DE (1973) Biochemical mechanisms of chromosome banding and color banding with acridine orange. In: Casperson T and Zeck L (eds) Chromosome Identification - Techniques and Applications in Biology and Medicine. Academic Press, New York, pp 292-306.

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

Dimitri P, Corradini N, Rossi F and Verní F (2004) The paradox of functional heterochromatin. BioEssays 27:28-41.

Dimitri P, Verní F, Mei E, Rossi F and Corradini N (2005) Transposable elements as artisans of the heterochromatic genome. Cytogenet Genome Res 110:165-172.

Fernández-García JL, Martínez-Trancón M, Rabasco A and Padilla JA (1998) Characterization of the heterochromatic chromosome regions in sheep. Genes Genet Syst 73:45-50.

Galleni L, Stanyon R, Tellini A, Giordano G and Santini L (1992) Karyology of the Savi pine vole, *Microtus savii* (De Sélys- Longchamps, 1838) (Rodentia, Arvicolidae): G-, C-, DA/DAPI, and AluI-bands. Cytogenet Cell Genet 59:290-292.

García L, Ponsá M, Egozcue J and GarcíaM (2000a) Comparative chromosomal analysis and phylogeny in four *Ctenomys* species (Rodentia, Octodontidae). Biol J Linn Soc 69:103-120.

García L, Ponsá M, Egozcue J and GarcíaM(2000b) Cytogenetic variation in *Ctenomys perrensi* (Rodentia, Octodontidae). Biol J Linn Soc 71:615-624.

Gosálvez J, López-Fernández C, Goyanes R and Mezzanotte V (1997) Chromosome differentiation using nucleases: An overview. In: Henriques-Gil N, Parker JS and Puertas MJ (eds) Chromosomes Today. Chapman & Hall, London, pp 23-49.

Holmquist GP and Dancis B (1979) Telomere replication, kinetochore organizers, and satellite DNA evolution. Proc Natl Acad Sci USA 76:4566-4570.

Hsu TC and Arrighi FE (1966) Chromosomal evolution in the genus *Peromyscus* (Cricetidae, Rodentia). Cytogenetics 5:355-359.

Jobse C, Buntjer JB, Haagsma N, Breukelman HJ, Beintema JJ and Lenstra JA (1995) Evolution and recombination of bovine DNA repeats. J Mol Evol 41:277-283.

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

John B, King M, Schweizer D and Mendelak M (1985) Equilocality of heterochromatin distribution and heterochromatin heterogeneity in acridid grasshoppers. Chromosoma 91:185-200.

Lee C, Stanyon R, Lin CC and Ferguson-Smith MA (1999) Conservation of human gamma-X centromeric satellite DNA among primates with an autosomal localization in certain Old World monkeys. Chromosome Res 7:43-47.

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

Matthey R (1958) Les chromosomes et la position systématique de quelques Murinae africains (Mammalia-Rodentia). Acta Trop 15:97-117.

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

Nieddu M, Rossino R, Pichiri G, Rocchi M, Setzu MD and Mezzanotte R (1999) The efficiency of *in situ* hybridization on human chromosomes with alphoid DNAs is enhanced by previous digestion with AluI and TaqI. Chromosome Res 7:593-602.

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

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

Pieczarka JC, Nagamachi CY, Muniz JAPC, Barros RMS and Mattevi MS (1998) Analysis of constitutive heterochromatin of *Aotus* (Cebidae, Primates) by restriction enzyme and fluorochrome bands. Chromosome Res 6:77-83.

Probst AV and Almouzni G (2008) Pericentric heterochromatin: Dynamic organization during early development in mammals. Differentiation 76:15-23.

Qumsiyeh MB, King SW, Arroyo-Cabrales J, Aggundey IR, Schlitter DA, Baker RJ and Morrow KJ (1990) Chromosomal and protein evolution in morphologically similar species of *Praomys* sensu lato (Rodentia, Muridae). *J Hered* 81:58-65.

Rocco L, Morescalchi MA, Costagliola D and Stingo V(2002) Karyotype and genome characterization in four cartilaginous fishes. Gene 295:289-298.

Rossi F, Moschetti R, Caizzi R, Corradini N and Dimitri P (2007) Cytogenetic and molecular characterization of heterochromatin gene models in *Drosophila melanogaster*. Genetics 175:595-607.

Saito Y, Edpalina RR and Abe S (2007) Isolation and characterization of salmonid telomeric and centromeric satellite DNA sequences. Genetica 131:157-166.

Santos S, Chaves R and Guedes-Pinto H (2004) Chromosomal localization of the major satellite DNA family (FA-SAT) in the domestic cat. Cytogenet Genome Res 107:119-122.

Schmid M, Haaf T, Steinlein C, Nanda I and Mahony M (2002) Chromosome banding in Amphibia: XXV. Karyotype evolution and heterochromatin characterization in Australian *Mixophyes* (Anura, Myobatrachidae). Cytogenet Genome Res 97:239-253.

Seabright M (1971) A rapid banding technique for human chromosomes. Lancet 2:971-972.

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

Singer MF (1982) Highly repeated sequences in mammalian genomes. Int Rev Cytol 76:67-112.

Slamovits CH and Rossi MS (2002) Satellite DNA: Agent of chromosomal evolution in mammals. A review. J Neotrop Mammal 9:297-308.

Sumner AT (1972) A simple technique for demonstrating centromeric heterochromatin. Exp Cell Res 75:304-306.

Van Den Bussche RA, Honeycutt RL and Baker RJ (1993) Restriction endonuclease digestion patterns of harvest mice (*Reithrodontomys*) chromosomes: A comparison to G-bands, C-bands, and *in situ* hybridization. Genetica 87:141-149.

Verma RS and Babu A (1995) Human Chromosomes – Principles and Techniques. 2nd edition. McGraw-Hill, New York, 419 pp.

Waye JS and Willard HF (1989) Concerted evolution of alpha satellite DNA: Evidence for species specificity and a general lack of sequence conservation among alphoid sequences of higher primates. Chromosoma 98:273-279.

Wichman HA, Payne CT, Ryder OA, Hamilton MJ, Maltbie M and Baker RJ (1991) Genomic distribution of heterochromatic sequences in equids: Implications to rapid chromosomal evolution. J Hered 82:369-377.

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

Associate Editor: Yatiyo Yonenaga-Yassuda

II.1.2

The involvement of repetitive sequences in the remodelling of karyotypes: the *Phodopus* genomes (Rodentia, Cricetidae)

Micron 46 (2012) 27–34

Contents lists available at SciVerse ScienceDirect

Micron

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

The involvement of repetitive sequences in the remodelling of karyotypes: The Phodopus genomes (Rodentia, Cricetidae)

A. Paço[a,1], R. Chaves[a,b,1], A. Vieira-da-Silva[a,1], F. Adega[a,*]

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

[b] Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

ARTICLE INFO

Article history:
Received 20 September 2012
Received in revised form
28 November 2012
Accepted 29 November 2012

Keywords:
Phodopus
Constitutive heterochromatin
Interstitial telomeric sequences
Karyotype evolution

ABSTRACT

In this work we characterize the *Phodopus roborovskii* and *Phodopus sungorus* karyotypes, describing the constitutive heterochromatin and the telomeric repeats distribution in these chromosomes. In the two species, (peri)centromeric, interstitial and (sub)telomeric C-bands were revealed, presenting a very high colocalization with evolutionary breakpoint regions identified in these karyotypes. Also both species present telomeric sequences located interstitially (ITS), as short ITS blocks or as large ITS blocks, mainly at the (peri)centromeric heterochromatic regions. The number and degree of ITSs amplification varies greatly in the two hamsters, indicating independent evolutionary events of these repeats in each genome. The combination of the data provided interesting insights about the genome organization of these hamster species, also allowing establishing evolutionary considerations on their chromosomes. The obtained results clearly suggest an involvement of the repetitive genomic fraction in the reshaping of *P. roborovskii* and *P. sungorus* karyotypes.

1. Introduction

Constitutive heterochromatin (CH) represents a significant fraction of the higher eukaryotic genomes (Dimitri et al., 2005; Rossi et al., 2007), accounting in some genomes for up to 50% (as in the kangaroo rat, *Dipidomys ordii*) (Singer, 1982). This genomic fraction can occur as large blocks or discrete C-bands in any part of a chromosome, but it is most commonly found in large blocks near the centromere (Corradini et al., 2007; Probst and Almouzni, 2008). CH regions are easily detected in the chromosomes by conventional C-banding procedures (Comings, 1973; Pathak and Arrighi, 1973), however, the *in situ* restriction endonuclease (RE) digestion of the chromosomes followed by C-banding proved to be a very useful technique in improving the analysis of CH (Gosálvez et al., 1997; Pieczarka et al., 1998). This technique allows the identification of C-bands undetected by conventional C-banding, named cryptic C-bands (e.g. Chaves et al., 2004b; Adega et al., 2005, 2007; Paço et al., 2009). Diverse works suggest the involvement of heterochromatin rich regions in chromosomal rearrangements' breakpoints, acting as hotspots for karyotypes' reshaping during evolution (Yunis and Yasmineh, 1971; Peacock et al., 1982; John, 1988; Chaves et al., 2004b), either by promoting the chromosome rearrangements, as is the case of satellite DNA (see Wichman et al., 1991; Slamovits et al., 2001;Adega et al., 2009) and/or, as fragile regions prone to chromosome breakage and consequently to the structural rearrangement, thus representing remnants of these events (e.g. Ruiz-Herrera et al., 2006).

The telomeric DNA of vertebrates and of some Bilateria species, such as Mollusca, Annelida and Echinodermata (Vitkova et al., 2005) consists of tandem repetitions of TTAGGG hexamers. In addition to its characteristic terminal position, blocks of telomeric repeats were already found at internal sites in the chromosomes of several mammalian species, known as interstitial telomeric sequences (ITS) (e.g. Meyne et al., 1990; Liu and Fredga, 1999; Rovatsos et al., 2011). The most common non-telomeric positions correspond to (peri)centromeric regions, where ITSs can present large arrays up to hundred of kilobases within or at the margins of the CH, named heterochromatic ITSs (het-ITSs) (Meyne et al., 1990; Ruiz-Herrera et al., 2008). Other short stretches of ITSs (s-ITSs) can be variably located at internal sites of chromosomes and are presumably present in all vertebrate species (Ruiz-Herrera et al., 2008). The origin and function of these telomeric sequences are still unclear but, as the other repeats here in analysis, it is believed that ITSs are probably

* Corresponding author. Tel.: +351 259350841.
E-mail addresses: apaco@utad.pt (A. Paço), rchaves@utad.pt (R. Chaves), ana.vieiradsilva@gmail.com (A. Vieira-da-Silva), filadega@utad.pt (F. Adega).

0968-4328/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.micron.2012.11.010

associated to chromosomal rearrangements as these are fragile sites and recombination hotspots (Lee et al., 1993; Finato et al., 2000; Ruiz-Herrera et al., 2008). For large ITSs, a noticeable molecular instability was reported, due to their propensity to form secondary structures, creating the opportunity for recombination events that might induce breakage (Balajee et al., 1994; Fernández et al., 1995; Nergadze et al., 2007). These breaks probably resulting in chromosomal reorganizations, such as inversions, fusions or fissions, redistributing telomeric repeats throughout the genomes (Ruiz-Herrera et al., 2008).

The *Phodopus* genus comprises three hamster species, *Phodopus campbelli*, *Phodopus roborovskii* and *Phodopus sungorus* (Carleton and Musser, 1993). The karyotype of *P. roborovskii* is composed by 34 chromosomes, and was first described by Vorontsov and Krjukova (1969). *P. campbelli* and *P. sungorus* species display a diploid chromosome number of 28 chromosomes, described for the first time by Matthey, 1960. Based on molecular (Neumann et al., 2006) and cytogenetic data (Romanenko et al., 2007) it was proposed that the species of the genus *Phodopus* represent the earliest split among all species of the Cricetinae subfamily. Comparing the species studied here, it is suggest that *P. roborovskii* retain a more primitive karyotype than *P. sungorus*, differing from the common *Phodopus* ancestor by three and six rearrangements, respectively (Schmid et al., 1986; Romanenko et al., 2007).

In this work we describe the CH and the telomeric repeats distribution in the karyotypes of *P. roborovskii* and *P. sungorus*. The assembling of the data provided some interesting insights about the genome organization of these hamsters, also allowing establishing evolutionary considerations about their chromosomes.

2. Material and methods

2.1. *Chromosome preparations and genomic DNA extraction*

Fixed chromosome preparations and genomic DNA from *P. roborovskii* (PRO) and *P. sungorus* (PSU) were obtained from fibroblast cell cultures using standard cell culture procedures described elsewhere (Chaves et al., 2004a), and the JETQUICK DNA kit (Genomed), respectively.

2.2. *GTD-banding*

Air dried slides were aged at 65 °C overnight and then subjected to standard G-banding procedures with trypsin (Seabright, 1971). Slides were stained with DAPI for a better contrast. The inversion of the DAPI in Adobe Photoshop (version 7.0) revealed the chromosomes G-banding pattern (GTD-banding, G-bands by trypsin with DAPI). The same slides were subsequently fixed with formaldehyde and used for C-banding (Chaves et al., 2002).

2.3. *In situ RE digestion*

Air dried slides were aged at 65 ºC for 6 h and then submitted to in situ restriction endonuclease (RE) digestion. The four restriction endonucleases used (AluI, DraI, HaeIII and RsaI) were diluted in buffers indicated by the manufacturer (Invitrogen LifeTechnologies) to final concentrations of 30 U per 100 μL of solution. One hundred microliters of the working solution was placed on slides that were then covered with coverslips and incubated in a moist chamber for 16 h at 37 ºC. Control slides were incubated only with buffer under the same conditions. Prior to C-banding, the slides were fixed with formaldehyde. Finally, the slides were stained with DAPI (the inversion of the DAPI revealed the RE-banding). The residual bands obtained after the endonuclease digestion were suitable for chromosome identification and karyotype organization.

2.4. *Physical mapping of telomeric sequences on chromosomes of P. roborovskii and P. sungorus*

The physical mapping of the telomeric sequences on the studied species chromosomes was carried out followed FISH procedures described by Schwarzacher and Heslop-Harrison (2000). The telomeric probe (TTAGGG)n was generated and labelled with biotin-16-dUTP (Roche Diagnostics), as described by Ijdo et al. (1991). The most stringent post-hybridization wash was 50% formamide/2×SSC at 42 ºC. Biotin-labelled probes were detected by FITC-avidin (Vector Laboratories).

2.5. CBP-banding sequential to G-banding, in situ RE digestion and fluorescent *in situ* hybridization

C-banding performed sequentially to G-banding, in situ RE banding or Fluorescent in situ Hybridization (FISH) was carried out after distaining the slides. CBP-banding (C-bands by barium hydroxide using propidium iodide) was done using the standard procedure of Sumner (1972), but with propidium iodide as counterstain. Briefly, the slides were submitted to routine C-banding: hydrochloric acid (0.2 M) for 20 min, barium hydroxide (5% solution) for 7 min and 2 × SSC (saline sodium citrate) at 60 ºC, 40 min. The results presented below are representative of, at least, 35 metaphases from five independent experiments for each endonuclease.

2.6. *Capture and preparation of images*

Chromosomes and interphase nuclei were observed in a Zeiss Axioplan Z1 microscope, and images were captured using an Axiocam MRm digital camera with LSM 510 software (version 4.0 SP2). Digitized photos were prepared for printing in Adobe Photoshop (version 7.0); contrast and colour optimization were the functions used and affected the whole image equally.

3. Results

3.1. *Constitutive heterochromatin distribution*

In the present work we analysed in detail the distribution of constitutive heterochromatin (CH) in the chromosomes of the two hamster species, *P. roborovskii* (PRO) and *P. sungorus* (PSU). For this analysis we performed C-banding after digesting the chromosomes with a panel of four restriction endonucleases (REs). This approach allowed the identification of several C-bands unidentified by classical C-banding, cryptic C-bands (see supplementary Figs. A.1 and A.2). In Fig. 1 it is possible to observe representative chromosome preparations of P. sungorus submitted to classical C-banding procedures (C-banding performed after G-banding, Fig. 1a and b) and C-banding performed after endonuclease DraI digestion (Fig. 1c and d). The cryptic C-bands are identified by black arrowheads (Fig. 1e). In a general overview, according to its location, at least three major classes of CH were identified in the studied species, (peri)centromeric, interstitial and (sub)telomeric, without showing any preferential representation. The majority of the chromosomes of both species present several interstitial C-bands. Notice PSU2 highlighted at the bottom of this figure, that presents five interstitial classical C-bands and 5 cryptic C-bands evidenced after the action

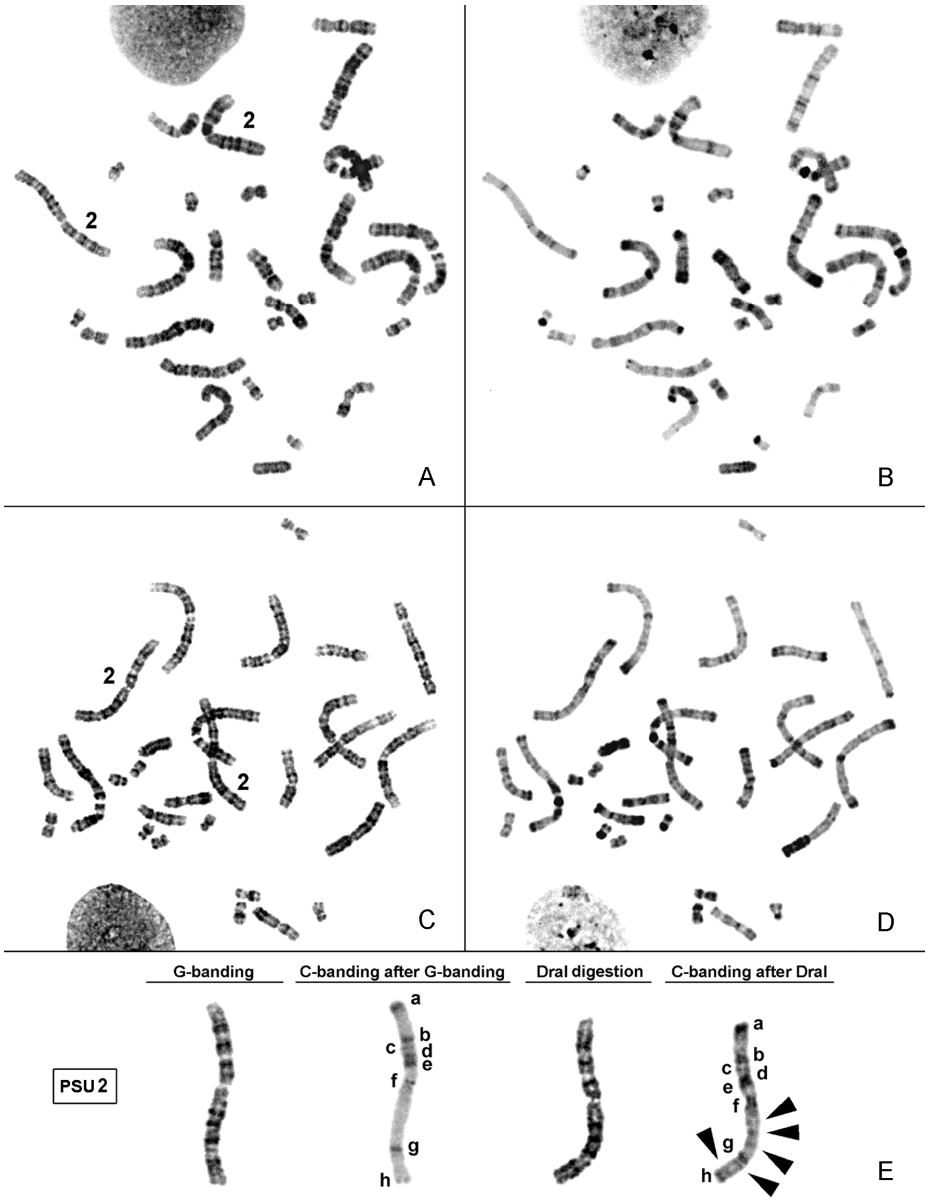


Fig. 1. Analysis of the endonuclease DraI on constitutive heterochromatin (CH) of *P. sungorus* chromosomes. (A) G-banding. (B) C-banding performed after G-banding (classical C-banding). (C) Banding produced by in situ chromosomes' digestion with DraI. (D) C-banding performed after DraI digestion. (E) CH detailed analysis performed for *P. sungorus* chromosome 2 (PSU2). Black arrowheads point C-bands unidentified by classical C-banding (cryptic C-bands).

of DraI (Fig. 1e).

3.2. *Physical distribution of telomeric sequences*

All the chromosomes of *P. roborovskii* and *P. sungorus* showed discrete telomeric hybridization signals at the telomeric regions (Fig. 2a and d). In *P. roborovskii* several chromosomes also display blocks of interstitial telomeric sequences (ITSs) located at the (peri)centromeric regions (PRO4, PRO5, PRO9, PRO10, PRO12, PRO13, PRO14, PRO15 and PRO16, Fig. 2a) and/or interstitial regions (PRO12 and PRO13, Fig. 2a). Similarly, in *P. sungorus*, several autosomal pairs and the X chromosome presents blocks of ITSs located at the (peri)centromeric regions (PSU3, PSU4, PSU6, PSU7, PSU11, PSU12 and PSUX, Fig. 2d). Attention goes to the large size of the ITS blocks observed in the chromosomes of P. roborovskii (Fig. 2a) and in one chromosome pair of P. sungorus (PSU11, Fig. 2d).

In interphase nuclei of *P. roborovskii* is notorious the "cluster-like pattern" displayed by the large ITS blocks (Fig. 3a). In *P. sungorus*, only two ITS clusters could be identified in the analysed nuclei (Fig. 3d, evidenced by arrowheads). Discrete telomere hybridization signals were also observed in *P. roborovskii* and *P. sungorus* nuclei (Fig. 3a and d), which certainly correspond to the telomere sequences present at the terminal chromosome regions and also to the short ITS blocks (in *P. sungorus*). In these two hamster species, all the ITSs identified seem to be located within heterochromatic regions, as can be observed in the metaphasic chromosomes (Fig. 2c and f, yellow signals) and in the interphase nuclei (Fig. 3c and f, yellow signals).

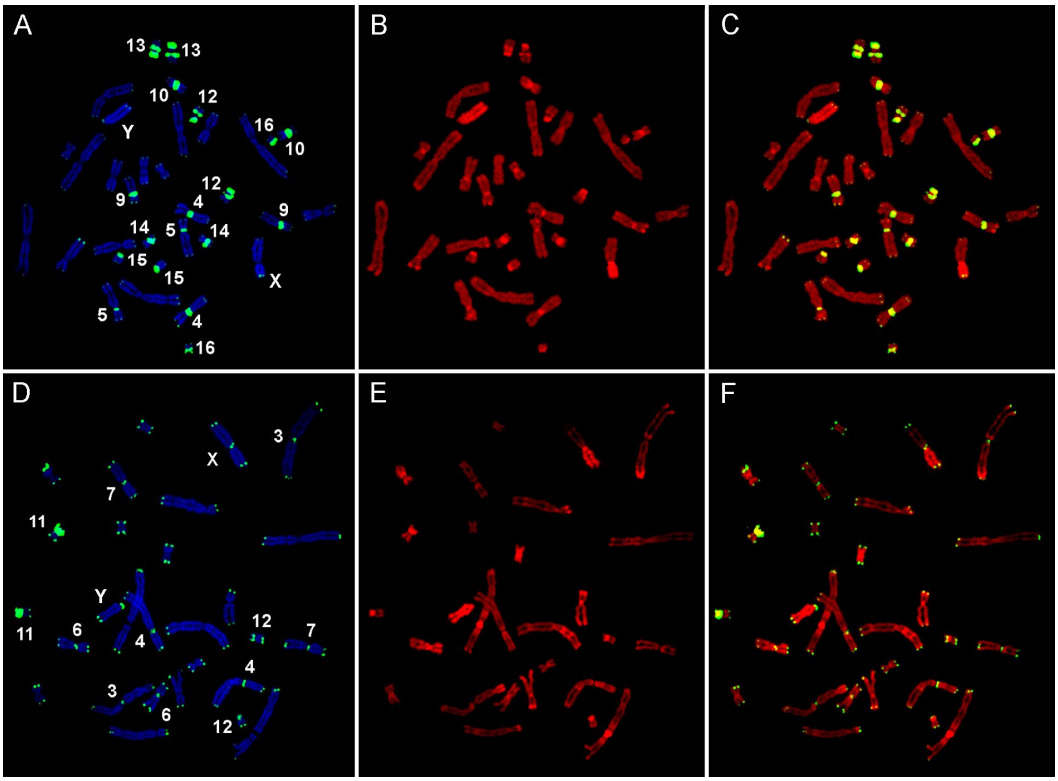


Fig. 2. Representative *in situ* hybridization of telomeric repeats (TTAGGG)n in *Phodopus roborovskii* and *P. sungorus* chromosomes. (A) FISH with telomeric sequences **Q3** (green signals) in metaphase chromosomes of *P. roborovskii*. The chromosomes were counterstained with DAPI (blue). (B) The same metaphase after C-banding with Propidium Iodide used as counterstain. (C) Overlapping of FISH and sequential C-banding in *P. roborovskii* chromosomes. (D) FISH with telomeric sequences (green signals) in metaphase chromosomes of *P. sungorus*. (E) The same metaphase after C-banding. (F) Overlapping of FISH and sequential C-banding in *P. sungorus* chromosomes.

4. Discussion

4.1. *Constitutive heterochromatin analysis*

In the current work, we present the distribution of constitutive heterochromatin (CH) in the chromosomes of the two hamster species *P. roborovskii* (PRO) and *P. sungorus* (PSU), using in situ restriction endonuclease (RE) digestion followed by C-banding. All the C-bands identified in this analysis are schematized in Figs. 4 and 5 and these could be found at (peri)centromeric, interstitial and (sub)telomeric regions, without any preferential location. Various are the works proposing the involvement of heterochromatin rich regions in the occurrence of chromosomal rearrangements (Yunis and Yasmineh, 1971; Peacock et al., 1982; John, 1988; Chaves et al., 2004b), either because of its dynamic nature and/or because these are fragile regions prone to chromosome breakage. In this way, the observed scattered CH distribution in *P. roborovskii* and *P. sungorus* genomes could indicate the involvement of this genomic fraction in the chromosomal restructurings that modulated these karyotypes, representing remnants of these events. To assess the possible involvement of CH with the karyotype evolution of the *Phodopus* species, we compared the CH distribution with the location of evolutionary breakpoint regions (regions between two syntenic blocks, according to Murphy et al., 2005). For this analysis we used the comparative chromosome map of *P. roborovskii* with the genomes of *Cricetulus griseus* and *Mesocricetus auratus* (Fig. 4), inferred in this work from Romanenko et al. (2007) data, and the comparative map of *P. sungorus* performed by Romanenko et al. (2007) (Fig. 5). This analysis revealed approximately 95% of coincidence between the breakpoint regions and the CH in P. roborovskii (Fig. 4) and 100% in P. sungorus (Fig. 5), thus supporting our hypothesis that CH was involved in the reshaping of these karyotypes during evolution.

4.2. *Telomeric repeat sequences analysis*

As expected, all the chromosomes of P. roborovskii and *P. sungorus* present telomeric repeats at the telomeres (Fig. 2). In addition, both species display blocks of interstitial telomeric sequences (ITSs) located mainly in the heterochromatic (peri)centromeric regions (heterochromatic ITS, het-ITSs) (Fig. 2a and d). This type of ITSs have been described in several mammalian species, such as carnivores (Wurster-Hill et al., 1988), cetartiodactyls (Vermeesch et al., 1996), chiropterans (Finato et al., 2000), primates of the genus *Eulemur* (Go et al., 2000), perissodactyls (Santani et al., 2002), marsupials (Metcalfe et al., 2007) and rodents (Bertoni et al., 1996; Ventura et al., 2006; Rovatsos et al., 2011; present work). Outside the mammalian group, het-ITSs were also reported in amphibians (Wiley et al., 1992), fishes (Abuín et al., 1996), reptiles (Pellegrino et al.,1999) and birds (Nanda et al., 2002).

The ITS blocks within a given genome can vary in size. Accordingly, in P. roborovskii all ITSs identified occur as large blocks (Fig. 2a), but in *P. sungorus* the majority are short ITS (s-ITSs) and only one large ITS block (see PSU11, Fig. 2d). In contrast to large ITS blocks, s-ITSs are probably present in all mammalian genomes, however their FISH detection depends probably on its sequence copy number (Ruiz-Herrera et al., 2008).

Regarding the origin of these telomeric sequences, it

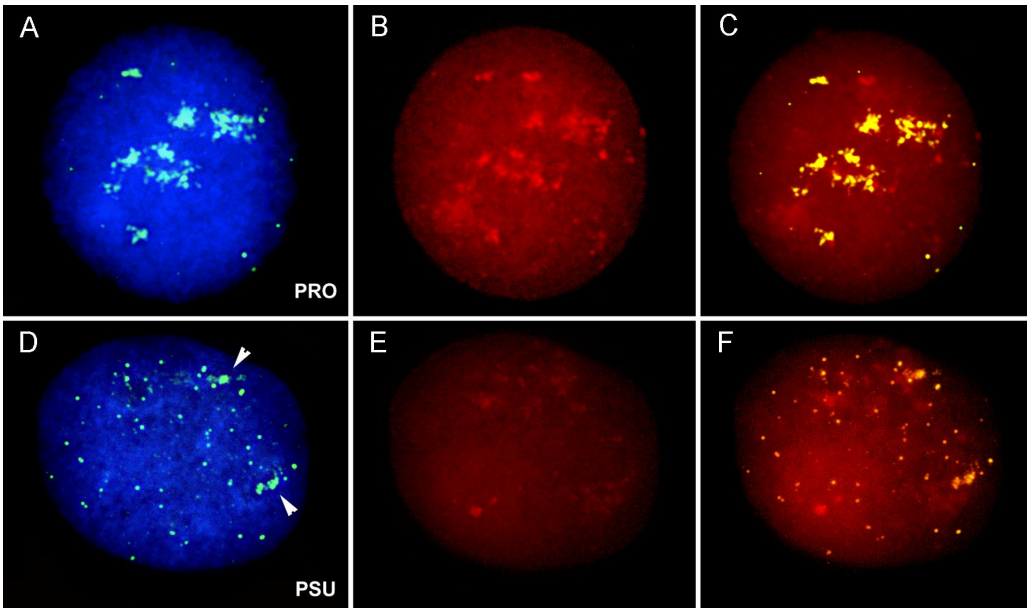


Fig. 3. Representative *in situ* hybridization of telomeric repeats (TTAGGG)n in *Phodopus roborovskii* and *P. sungorus* interphase nuclei. (A) FISH with telomeric sequences (green signals) in interphase nucleus of *P. roborovskii*. The chromatin was counterstained with DAPI (blue). (B) The same nucleus after C-banding with Propidium Iodide used as counterstain. (C) Overlapping of FISH and sequential C-banding in a *P. roborovskii* nucleus. (D) FISH with telomeric sequences (green signals) in interphase nucleus of *P. sungorus*. (E) The same nucleus after C-banding. (F) Overlapping of FISH and sequential C-banding in *P. sungorus* nucleus.

has been proposed that ITSs could result from chromosome rearrangements occurring during karyotype evolution, representing remnants these rearrangements, as robertsonian-like fusions (Slijepcevic,1998), tandem chromosome fusions (Li et al., 2000) or pericentric inversions (Rovatsos et al., 2011). In this context, we can suggest that some ITS blocks identified in the two Phodopus genomes could have its origin in chromosomal rearrangements, as chromosome fusions. The lower chromosome number of P. roborovskii and P. sungorus genomes (2n = 34 and 28, respectively), comparing to the hypothetical ancestral Muroidea (2n = 50, Chaves et al., 2012), Cricetinae and *Phodopus* (2n = 48 and 40, respectively, Romanenko et al., 2007) karyotypes supports this. Other chromosomal rearrangements, such as pericentric inversions, may also justify the presence of ITS blocks at the (peri)centromeric regions of P. roborovskii and *P. sungorus* chromosomes. Fig. 6 schematizes the possible origin of some ITSs observed in the chromosomes of these species, taking into consideration the hypothetical ancestral Muroidea karyotype (AMK, Chaves et al., 2012) and also

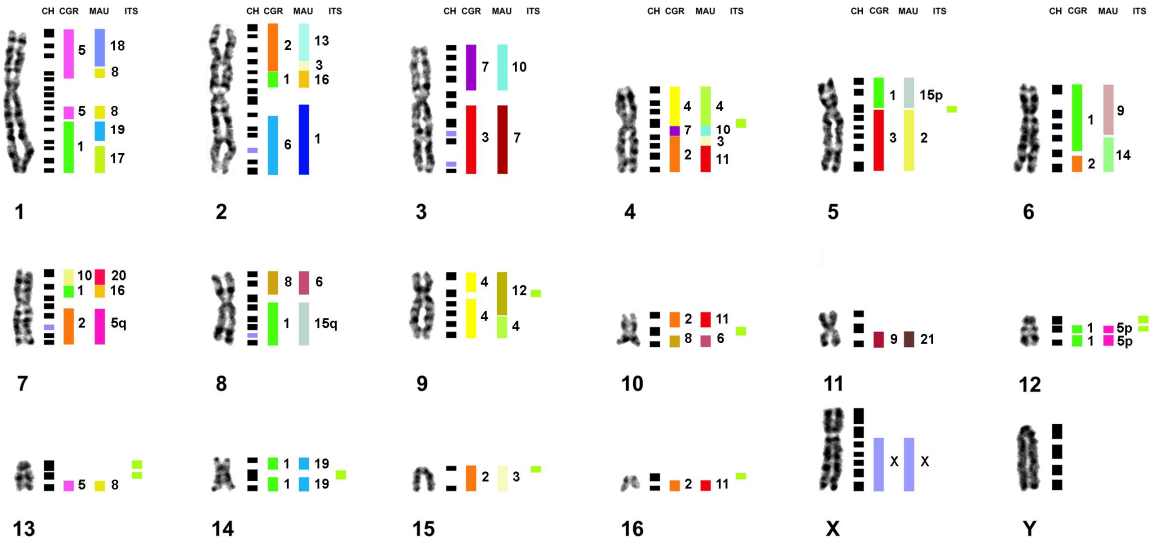


Fig. 4. *Phodopus roborovskii* comparative chromosome map displaying the chromosomal sintenies to *C. griseus* and *M. auratus*. The chromosome map of *Phodopus roborovskii* (PRO) was inferred in the present work from Romanenko et al. (2007) data. The chromosomal syntenies with *C. griseus* (CGR) and *M. auratus* (MAU) are indicated as coloured blocks (second and third columns) in the haploid karyotype of the GTG-banded PRO chromosomes. The black blocks on the right of each chromosome represent classical C-bands and blue blocks represent cryptic C-bands. The green blocks (fourth column) represent interstitial telomeric sequences (ITS).

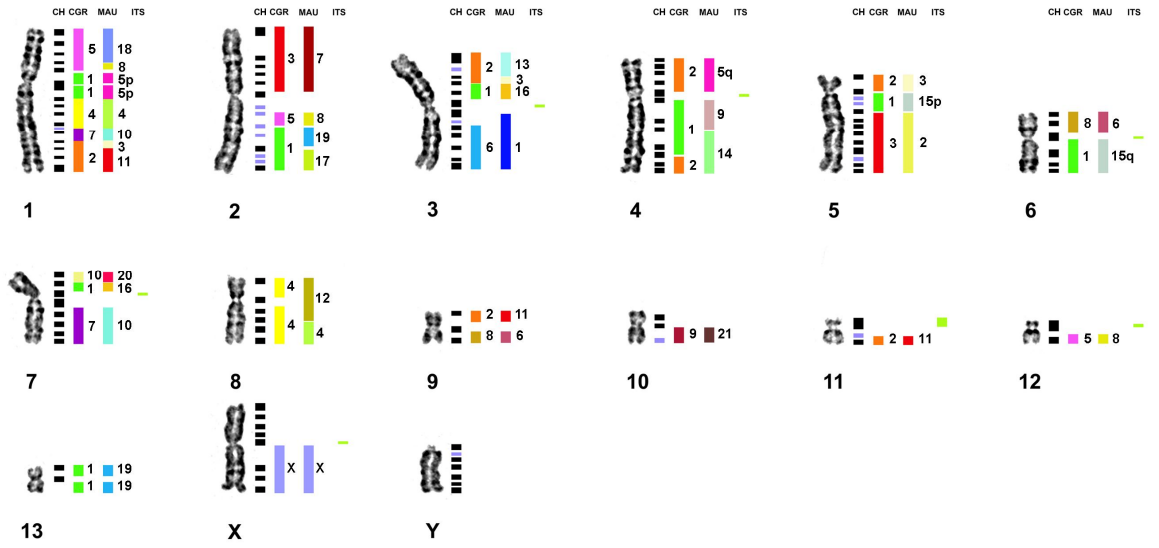


Fig. 5. *P. sungorus* comparative chromosome map displaying the chromosomal syntenies to *C. griseus* and *M. auratus*. The chromosome map of P. *sungorus* (PSU) was assigned from that of Romanenko et al. (2007). The chromosomal syntenies with *C. griseus* (CGR) and *M. auratus* (MAU) are indicated as coloured blocks (second and third columns) in the haploid karyotype of the GTG-banded PSU chromosomes. The black blocks on the right of each chromosome represent classical C-bands and blue blocks represent cryptic C-bands. The green blocks (fourth column) represent interstitial telomeric sequences (ITS).

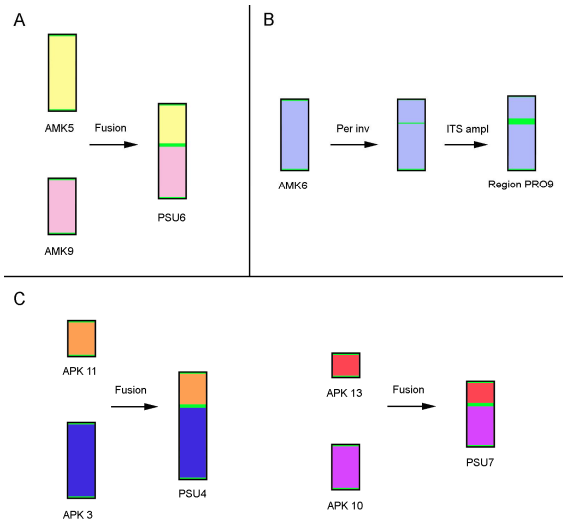


Fig. 6. Possible origin of ITS blocks in *Phodopus roborovskii* and *P. sungorus* chromosomes. Green blocks represent telomeric sequences. (A) Origin of the ITS block in chromosome 6 of P. sungorus (PSU6) by fusion of two ancestral Muroidea chromosomes (AMK5 and AMK9). (B) Origin of the ITS block of *P. roborovskii* chromosome 9 (PRO9) after the occurrence of a pericentric inversion. (C) Origin of ITS blocks in PSU4 and PSU7 chromosomes by fusion of two ancestral *Phodopus* chromosomes (APK11 and APK3, APK 13 and APK10). Chromosome fusion – fusion, pericentric inversion – Per inv, interstitial telomeric sequences amplification – ITS ampl. The Muroidea ancestral chromosomes, homologous to *P. roborovskii* and *P. sungorus* chromosomes, were identified according to Romanenko et al. (2007) and Chaves et al. (2012) data. The *Phodopus* ancestral chromosomes, homologous to *P. roborovskii* and *P. sungorus* chromosomes, were identified according to the ancestral *Phodopus* karyotype presented as supplementary material.

the hypothetical ancestral *Phodopus* karyotype (APK, present work, see supplementary Fig. B.1). Considering the ITS pattern analysed here and the sintenies between *P. sungorus* chromosomes and the AMK, we suggest the occurrence of a chromosome fusion to originate the biarmed PSU6, with the retention of telomeric sequences (Fig. 6a). The ITS block observed in PRO9 could have been originated by a pericentric inversion of an ancestral chromosome, since this region of PRO9 corresponds to a single ancestral Muroidea chromosome (according to Chaves et al., 2012), as can be observed in Fig. 6b. According to Romanenko et al. (2007) data, PSU4 and PSU7 could only result from fusion of two chromosomes of the APK (see supplementary Fig. A.1). Based on this, the (peri)centromeric ITS blocks observed in these *P. sungorus* chromosomes most probably originated by the retention of telomeric repeats at the time of the fusion (Fig. 6c).

The large ITS blocks that seem to occupy the entire short arm of PRO12, PRO13, PRO15, PRO16 and PSU11, from the (peri)centromeric region to the telomere border, could be the result of telomeric repeats amplification (by unequal crossing-over, replication slippage, gene conversion and/or rolling circle amplification, etc.), as described for other tandemly repeated sequences (Elder and Turner, 1995). This amplification could take place in each genome after the split of *Phodopus* or in the ancestral chromosomes, as it seems to be the case of PRO16 and PSU11, homologous *in toto* (see Romanenko et al., 2007). In this context, we believe in the occurrence of ITSs amplification or alternatively, in reductions and progressive degeneration/elimination of large ITS blocks in each *Phodopus* genome. Nevertheless, ITSs relocations also seem a plausible scenario. These molecular events certainly occurred in the two *Phodopus* genomes independently, and together with the occurred chromosomal rearrangements are responsible for the different ITS pattern observed in both species. This independent ITS evolution is easily observed in several chromosome pairs of *P. roborovskii* and *P. sungorus* that present a different ITS pattern despite the homology in toto (Figs. 4 and 5) (PRO2/PSU3, PRO8/PSU6, PRO9/PSU8, PRO10/PSU9, PRO13/PSU12, PRO14/PSU13, homologies according to the work of Romanenko et al., 2007).

After analyzing several nuclei from P. roborovskii (Fig. 3), it was evident that the large ITS blocks present a "cluster-like pattern", similar to the organization of satellite DNA (satDNA) sequences in the interphase nucleus (Manuelidis et al., 1982). In *P. sungorus*, two ITS clusters could be observed (Fig. 3d), probably corresponding to the large ITS block of PSU11. C-banding performed after

FISH revealed that all the ITS blocks (in metaphase chromosomes and interphase nuclei) are localized in heterochromatic regions. Taking this observation in consideration, is most probable that the ITS sequences could be integrated in repeat units of a satDNA. FA-SAT from Felis catus is an example of this kind of satDNA families (Fanning, 1987; Santos et al., 2004). If it is also the case for Phodopus, the amplification and relocation of satDNA may also have lead to the amplification and relocation of the telomeric sequences (by unequal crossing-over, for example), resulting in the large and diverse telomeric blocks of both *Phodopus* studied. If so, the evolution of ITSs is associated with the one from satDNA repeats.

4.3. *Conclusions*

The assembling of all the findings reported in this work clearly suggests an involvement of the repetitive genomic fraction in the reshaping of *P. roborovskii* and *P. sungorus* karyotypes. A high colocalization of CH (approximately 95% in *P. roborovskii* and 100% in *P. sungorus*) with the evolutionary breakpoint regions in these two species indicates the involvement of CH with the chromosomal rearrangements that reorganized these karyotypes. In both species telomeric sequences located interstitially (ITS) were also identified. The different ITS patterns observed in the two *Phodopus* species are the result of different molecular events: chromosomal rearrangements and/or the high dynamic nature of the ITSs itself. This last one can be explained by the amplification, reduction, progressive degeneration/elimination and relocation of the ITS blocks. This ITSs' molecular dynamics could be a source of chromosomal restructurings through its molecular unstable nature. As follows these results points to a simultaneous involvement of all these repeats in the Phodopus karyotype's reorganization and evolution.

Acknowledgements

This work was supported by the project 342 POCI/BIABCM/ 58541/2004, a PhD grant (SFRH/BD/41574/2007) 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.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/ j.micron.2012.11.010.

References

Abuín, M., Martínez, P., Sánchez, L., 1996. Localization of the repetitive 352 telomeric sequence (TTAGGG)n in four salmonid species. Genome 39, 1035–1038.

Adega, F., Chaves, R., Guedes-Pinto, H., 2005. Chromosome restriction enzyme digestion in domestic pig (*Sus scrofa*), constitutive heterochromatin 355 arrangement. Genes and Genetic Systems 80, 49–56.

Adega, F., Chaves, R., Guedes-Pinto, H., 2007. Constitutive heterochromatin 357 characterization of white-lipped and collared peccaries (Tayassuidae). Journal of Genetics 86,19–26.

Adega, F., Guedes-Pinto, H., Chaves, R., 2009. Satellite DNA in 360 the karyotype evolution of domestic animals – clinical considerations. Cytogenetic and Genome Research 126,12–20.

Balajee, A.S., Oh, H.J., Natarajan, A.T., 1994. Analysis of restriction enzyme-induced chromosome aberrations in the interstitial telomeric repeat sequences of CHO and CHE cells by FISH. Mutation Research 307, 307–313.

telomeric sequence in vertebrate chromosomes. Chromosoma 99, 3–10.

Murphy, W.J., Larkin, D.M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., Beever, J.E., Chowdhary, B.P., Galibert, F., Gatzke, L., Hitte, C., Meyers, S., Milan, D., Ostrander, E.A., Pape,

Bertoni, L., Attolini, C., Faravelli, M., Simi, S., Giulotto, E., 1996. Intrachromosomal telomere-like DNA sequences in Chinese hamster. Mammalian Genome 7, 853–855.

Carleton, M.D., Musser, G.G., 1993. Family muridae. In: Wilson, D.E., Reeder, D.M. (Eds.), Mammal Species of the World: A Taxonomic and Geographic Reference. 2nd ed. Smithsonian Institution Press, Washington and London, pp. 1189–1532.

Chaves, R., Adega, F., Santos, S., Guedes-Pinto, H., Heslop-Harrinson, J.S., 2002. *In situ* hybridization and chromosome banding in mammalian species. Cytogenetic and Genome Research 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 Research 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.

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

Comings, D.E., 1973. Biochemical mechanisms of chromosome banding and color banding with acridine orange. In: Casperson, T., Zeck, L. (Eds.), Chromosome Identification – Techniques and Applications in Biology and Medicine. Academic Press, New York, pp. 292–306.

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.

Dimitri, P., Verní, F., Mei, E., Rossi, F., Corradini, N., 2005. Transposable elements as artisans of the heterochromatic genome. Cytogenetic and Genome Research 110, 165–172.

Elder Jr., J.F., Turner, B.J., 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Quarterly Review of Biology 70, 297–320.

Fanning, T., 1987. Origin and evolution of a major feline satellite DNA. Journal of Molecular Biology 197, 627–634.

Fernández, J.L., Gosálvez, J., Goyanes, V., 1995. High frequency of mutagen-induced chromatid exchanges at interstitial telomere-like DNA sequence blocks of Chinese hamster cells. Chromosome Research 3, 281–284.

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

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

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.

Ijdo, J.W., Wells, R.A., Baldini, A., Reeders, S.T., 1991. Improved telomere detection using a telomere repeat probe (TTAGGG)n generated by PCR. Nucleic Acids Research 19 (17), 478.

John, B., 1988. The biology of heterochromatin. In: Verma, R.S. (Ed.), Heterochromatin: Molecular and Structural Aspects. Cambridge University Press, Cambridge, pp. 1–128.

Lee, C., Sasi, R., Lin, C.C., 1993. Interstitial localization of telomeric DNA sequences in the Indian muntjac chromosomes: further evidence for tandem chromosome fusions in the karyotypic evolution of the Asian muntjacs. Cytogenetics and Cell Genetics 63, 156–159.

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 Research 8, 363–373.

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

Manuelidis, L., Langer-Safer, P.R., Ward, D.C., 1982. High-resolution mapping of satellite using biotin-labeled DNA probes. Journal of Cell Biology 95, 619–625.

Matthey, R., 1960. Chromosomes, hétêrochromosomes, et cytology comparée des Cricetinar Paléarctiques (Rodentia). Caryologia 13, 199–223.

Metcalfe, C.J., Eldridge, M.D., Johnston, P.G., 2007. Mapping the distribution of the telomeric sequence (T2AG3) n in the Macropodoidea (Marsupialia) by fluorescence *in situ* hybridization, II. The ancestral 2n = 22 macropodid karyotype. Cytogenetic and Genome Research 116, 212–217.

Meyne, J., Baker, R.J., Hobart, H.H., Hsu, T.C., Ryder, O.A., Ward, O.G., Wiley, J.E., Wurster-Hill, D.H., Yates, T.L., Moyzis, R.K., 1990. Distribution of non-telomeric sites of the (TTAGGG) n G., Parker, H.G., Raudsepp, T., Rogatcheva, M.B., Schook, L.B., Skow, L.C., Welge, M., Womack, J.E., O'Brien, S.J., Pevzner, P.A., Lewis, H.A., 2005. Dynamics of mammalian chromosome evolution inferred from multi-species comparative maps. Science 309, 613–617.

Nanda, I., Schrama, D., Feichtinger, W., Haaf, T., Schartl, M., Schmid, M., 2002. Distribution of telomeric (TTAGGG)(n) sequences in avian chromosomes. Chromosoma 111, 215–227.

Nergadze, S.G., Santagostino, M.A., Salzano, A., Mondello, C., Giulotto, E., 2007. Contribution of telomerase RNA retrotranscription to DNA double-strand break repair during mammalian genome evolution. Genome Biology 8, R260.

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. Molecular Phylogenetics and Evolution 39, 135–148.

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). Genetics andMolecular Biology 32 (1), 58–68.

Pathak, S., Arrighi, F.E., 1973. Loss of DNA following C-banding procedures. Cytogenetics and Cell Genetics 12, 414–422.

Peacock, W.J., Dennis, E.S., Gerlach, W.L., 1982. DNA sequence changes and speciation. Progress in Clinical and Biological Research 96,123–142.

Pellegrino, K.C., Rodrigues, M.T., Yonenaga-Yassuda, Y., 1999. Chromosomal evolution in the Brazilian lizards of genus *Leposoma* (Squamata, Gymnophthalmidae) from Amazon and Atlantic rain forests: banding patterns and FISH of telomeric sequences. Hereditas 131,15–21.

Pieczarka, J.C., Nagamachi, C.Y., Muniz, J.A.P.C., Barros, R.M.S., Mattevi, M.S.,1998. Analysis of constitutive heterochromatin of *Aotus* (Cebidae, Primates) by restriction enzyme and fluorochrome bands. Chromosome Research 6, 77–83.

Probst, A.V., Almouzni, G., 2008. Pericentric heterochromatin: dynamic organization during early development in mammals. Differentiation 76,15–23.

Rovatsos, M.Th., Marchal, J.A., Romero-Fernández, I., Fernández, F.J., Giagia- Athanosopoulou, E.B., Sánchez, A., 2011. Rapid, independent, and extensive amplification of telomeric repeats in pericentromeric regions in karyotypes of arvicoline rodents. Chromosome Research 19, 869–882.

Romanenko, S.A., Sitnikova, N.A., Serdukova, N.A., Perelman, P.L., Rubtsova, N.V., Bakloushinskaya, I.Y., Lyapunova, E.A., Just, W., Ferguson-Smith, M.A., Yang, F., Graphodatsky, A.S., 2007. Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroideia, Rodentia) inferred from chromosomal painting and banding comparison. Chromosome Research 15, 283–297.

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 Biology 7, R115.

Ruiz-Herrera, A., Nergadze, S.G., Santagostino, M., Giulotto, E., 2008. Telomeric repeats far from the ends: mechanisms of origin and role in evolution. Cytogenetic and Genome Research 122, 219–228.

Santani, A., Raudsepp, T., Chowdhary, B.P., 2002. Interstitial telomeric sites and NORs in Hartmann's zebra (*Equus zebra hartmannae*) chromosomes. Chromosome Research 10, 527–534.

Santos, S., Chaves, R., Guedes-Pinto, H., 2004. Chromosomal localization of the major satellite DNA family (FA-SAT) in the domestic cat. Cytogenetic and Genome Research 107, 119–122.

Schmid, M.T., Haaf, T., Weis, H., Schempp, W., 1986. Chromosomal homologies in hamster species of the genus Phodopus (Rodentia, Cricetinae). Cytogenetics and Cell Genetics 43, 168–173.

Schwarzacher, T., Heslop-Harrison, J.S., 2000. Practical *In situ* Hybridization. Bios, Oxford.

Seabright, M., 1971. A rapid banding technique for human chromosomes. Lancet 2, 971–972.

Singer, M.F., 1982. Highly repeated sequences in mammalian genomes. International Review of Cytology 76, 67–112.

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. Molecular Biology and Evolution 18, 1708–1719.

Slijepcevic, P., 1998. Telomeres and mechanisms of Robertsonian fusion. Chromosoma 107, 136–140.

Sumner, A.T., 1972. A simple technique for demonstrating centromeric heterochromatin. Experimental Cell Research 75, 304–306.

Ventura, K., Silva, M.J., Fagundes, V., Christoff, A.U., Yonenaga-Yassuda, Y., 2006. Non-telomeric sites as evidence of chromosomal rearrangement and repetitive (TTAGGG)n arrays in heterochromatic and euchromatic regions in four species of Akodon (Rodentia, Muridae). Cytogenetic and Genome Research 115, 169–175.

Vermeesch, J.R., De Meurichy, W., Van Den Berghe, H., Marynen, P., Petit, P., 1996. Differences in the distribution and nature of the interstitial telomeric (TTAGGG)n sequences in the chromosomes of the Giraffidae, okapi (Okapia johnstoni), and giraffe (*Giraffa camelopardalis*): evidence for ancestral telomeres at the okapi polymorphic rob(5; 26) fusion site. Cytogenetics and Cell Genetics 72, 310–315.

Vitkova, M., Král, J., Traut, W., Zrzavý , J., Marec, F., 2005. The evolutionary origin of insect telomeric repeats (TTAGG)n. Chromosome Research 13, 145–156.

Vorontsov, N.N., Krjukova, E.P., 1969. *Phodopus przhewalskii* species nova, a new species of desert hamsters (Cricetinae, Cricetidae, Rodentia) from the Zaissan Basin. In: Vorontsov, N.N. (Ed.), The Mammals. Academy of Sciences of the USSR, Novosibirsk, pp. 102–104.

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. Journal of Heredity 82, 369–377.

Wiley, J.E., Meyne, J., Little, M.L., Stout, J.C., 1992. Interstitial hybridization sites of the (TTAGGG) n telomeric sequence on the chromosomes of some North American hylid frogs. Cytogenetics and Cell Genetics 61, 55–57.

Wurster-Hill, D.H., Ward, O.G., Davis, B.H., Park, J.P., Moyzis, R.K., Meyne, J., 1988. Fragile sites, telomeric DNA sequences, B chromosomes, and DNA content in raccoon dogs, *Nyctereutes procyonoides*, with comparative notes on foxes, coyote, wolf, and raccoon. Cytogenetics and Cell Genetics 49, 278–281.

Yunis, J.J., Yasmineh, W.G., 1971. Heterochromatin, satellite DNA, and cell function. Structural DNA of eukaryotes may support and protect genes and aid in speciation. Science 174, 1200–1209.

Supplementary data

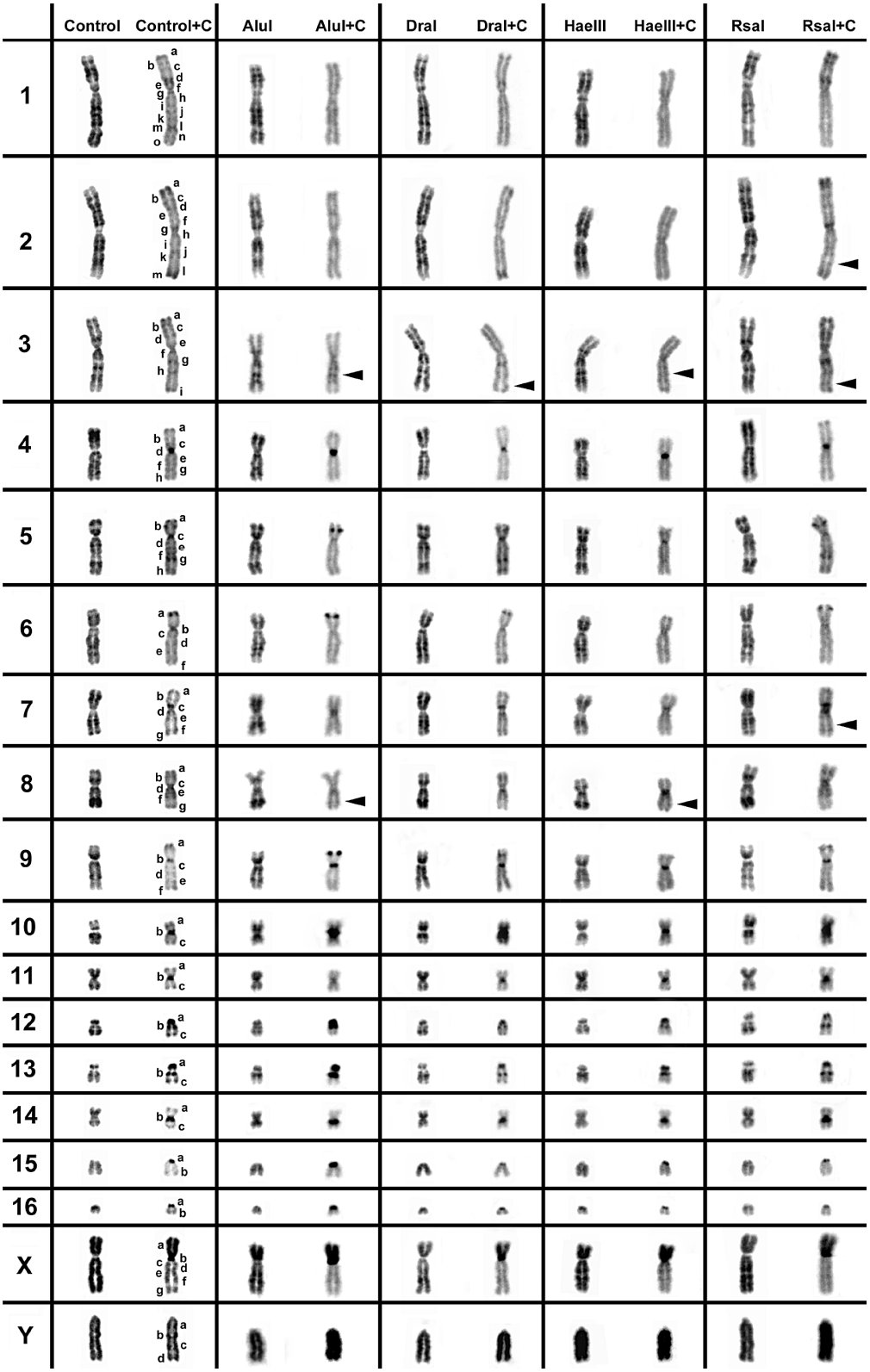


Fig. A.1. *In situ* restriction of *Phodopus roborovskii* chromosomes and sequential C-banding. Control G- and C-banding of *Phodopus roborovskii* chromosomes (2n = 34) are shown on the left column. The other columns show the bands produced by each of

the four restriction endonucleases (REs) used, before and after C-banding. The letters (a-g) represent the C-bands according to their order of appearance in each chromosome. Arrowheads indicate C-positive heterochromatin bands only revealed by previous RE treatment (cryptic C-bands).

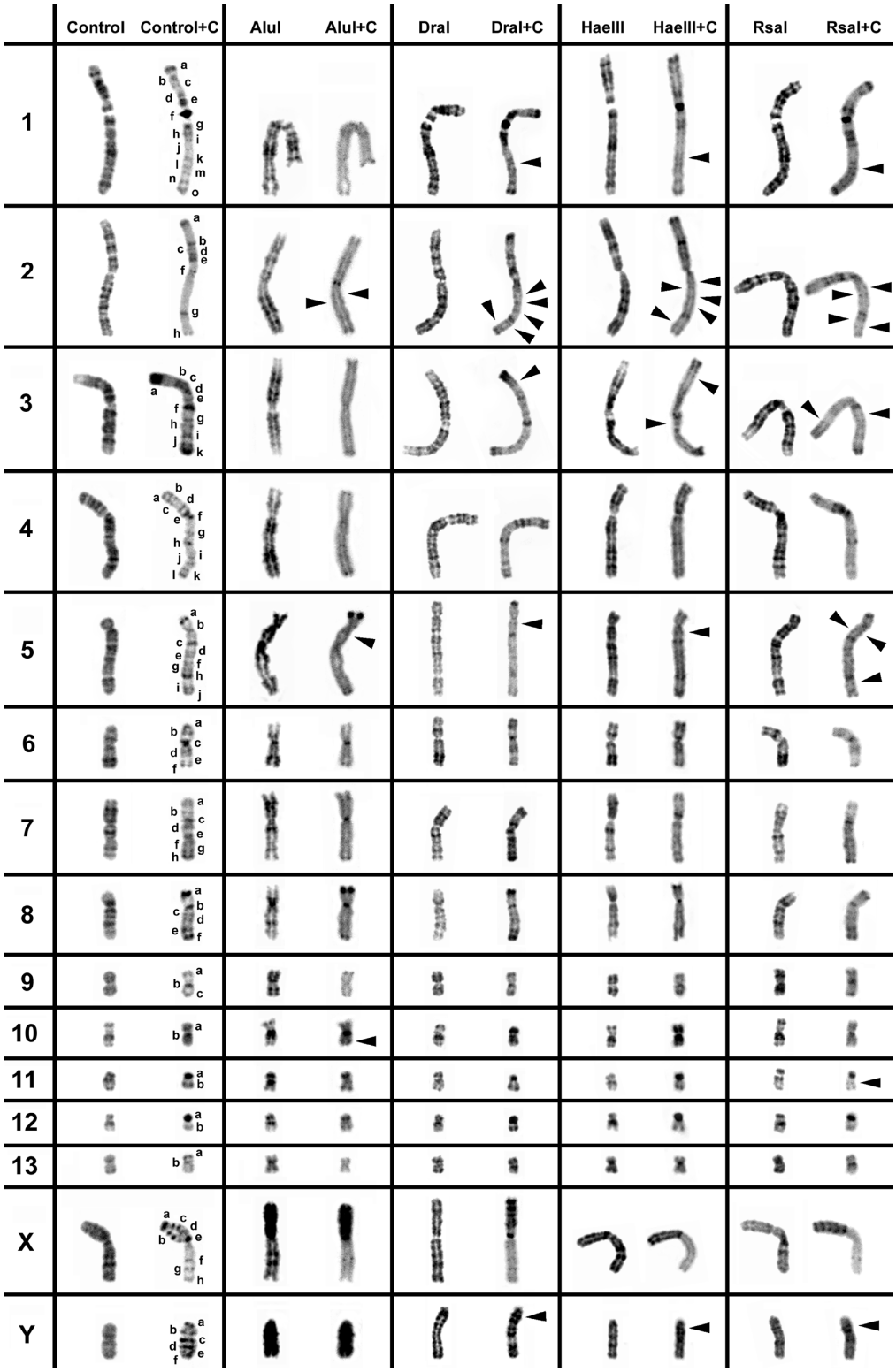


Fig. A.2. *In situ* restriction of *Phodopus sungorus* chromosomes and sequential C-banding. Control G- and C-banding of *Phodopus sungorus* chromosomes (2n = 28) are shown on the left column. The other columns show the bands produced by each of the four restriction endonucleases (REs) used, before and after C-banding. The letters (a-g) represent the C-bands according to their order of appearance in each chromosome. Arrowheads indicate C-positive heterochromatin bands only revealed by previous RE treatment (cryptic C-bands).

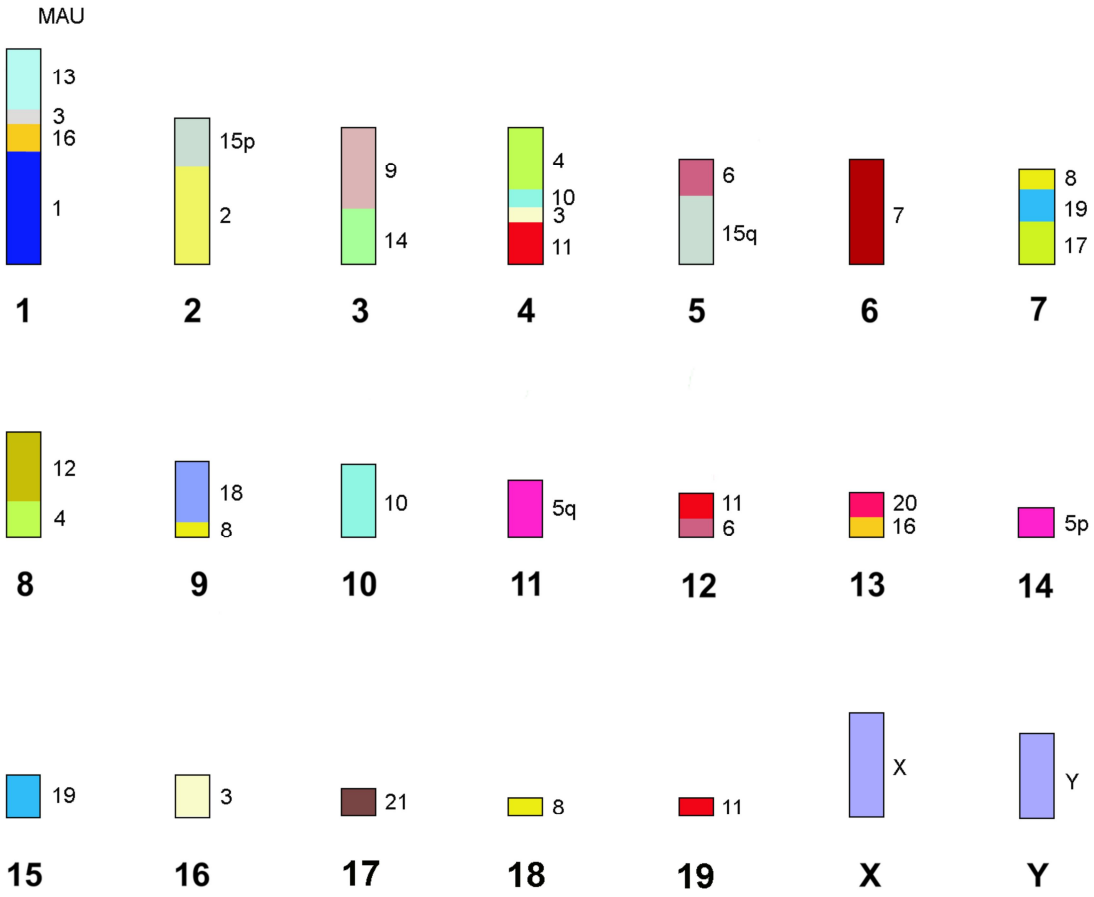
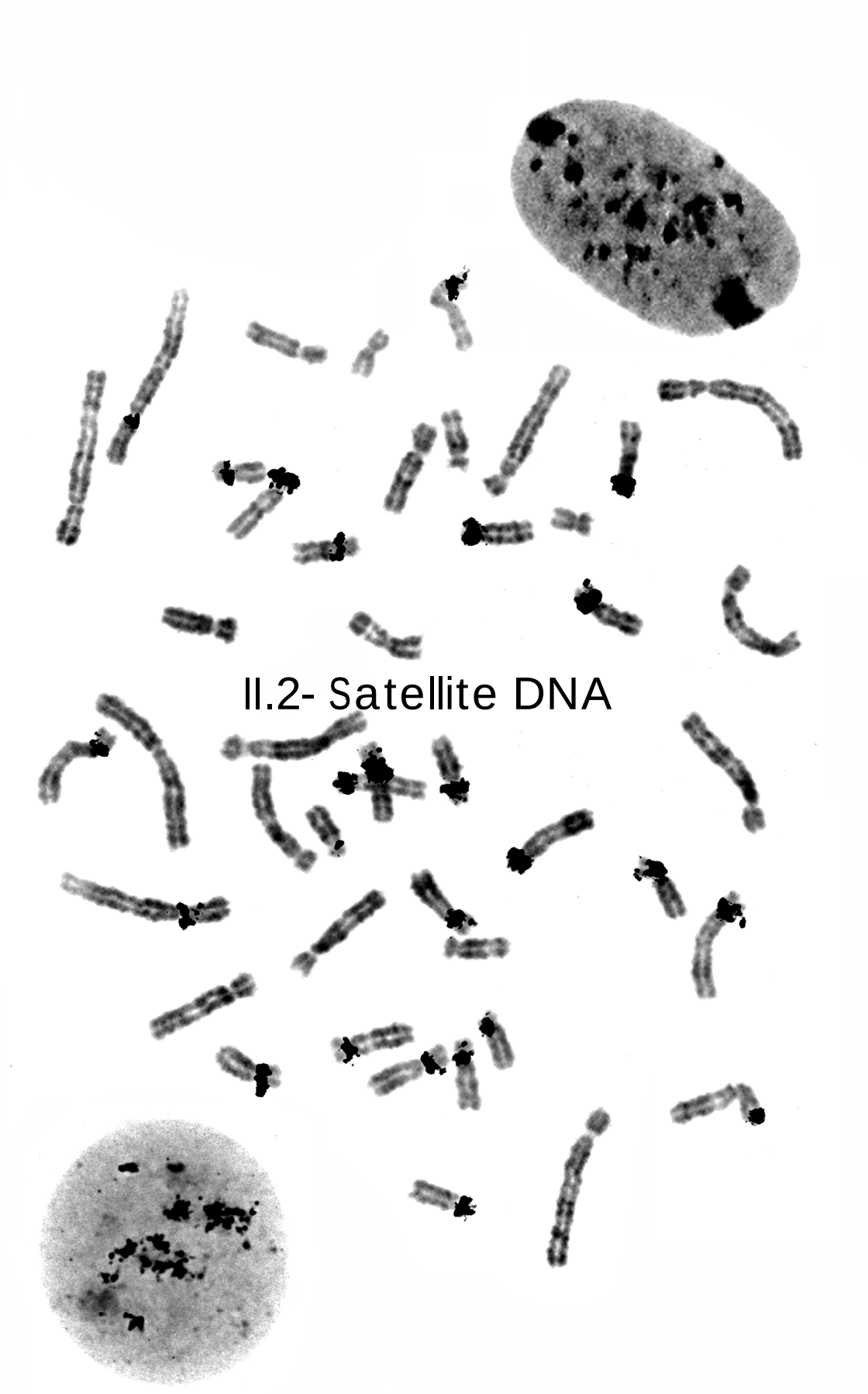


Fig. B.1. Hypothetical ancestral *Phodopus* karyotype 2n = 40. The coloured blocks correspond to chromosomes of *Mesocricetus auratus* (MAU). The hypothetical ancestral *Phodopus* karyotype (APK) was constructed based on Romanenko et al. (2007) data.



II.2- Satellite DNA

With the objective to investigate the genomic importance of satellite DNA (satDNAs) sequences, particularly their involvement in karyotype restructuring, in this work were isolated six novel satDNAs from four rodent species studied, *Cricetus cricetus, Peromyscus eremicus, Phodopus roborovskii* and *Phodopus sungorus*. For *C. cricetus* and both *Phodopus* species, these sequences correspond indeed to the first satDNAs isolated in these genomes. This investigation resulted in four papers that increase the evidences regarding the responsibility of satDNAs in genome plasticity, due to their high molecular dynamics. The chromosomal location and copy number analyses of these sequences in different species genomes allow to propose models for the evolution of these sequences.

Despite the results obtained here that implicate these sequences in chromosomal restructurings, the data also supports other functions for satDNAs. An analysis focused in the transcriptional activity of these sequences, particularly for PSUcentSat and PERcentSat, increase the evidences about the role of satDNAs in chromatin remodelation, control of gene expression, centromeric function and in the response to stress. This work also shows that not all the copies of a satDNA transcriptionally active may be available for transcription. In this way, satDNAs may present simultaneously more than one function in a genome, a function played by their transcripts and probably a more structural function assumed, perhaps, by the copies that are not transcribed.

II.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 Tra´s-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 characterized 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.

The molecular analysis of repetitive sequences and their physical mapping in chromosomes of different

* 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

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 Gosá lvez 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, Muse´um 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 ºC 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 ºC 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 ºC. Control slides were also prepared according to the aforementioned procedures but only with buffer. The slides were washed in distilled water and airdried. 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. CBPbanding [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 ºC for 40 min. The slides where then counterstained with PI (1.5 mI/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 hybridization*

Chromosome preparations from *C. cricetus* and *P. eremicus* were aged at 65 8C 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 8C (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 ºC, allowing sequences with more than 77% similarity to remain hybridised. Digoxigenin-labelled probes were detected with antidigoxigenin- 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 lack hybridisation signal, suggesting that

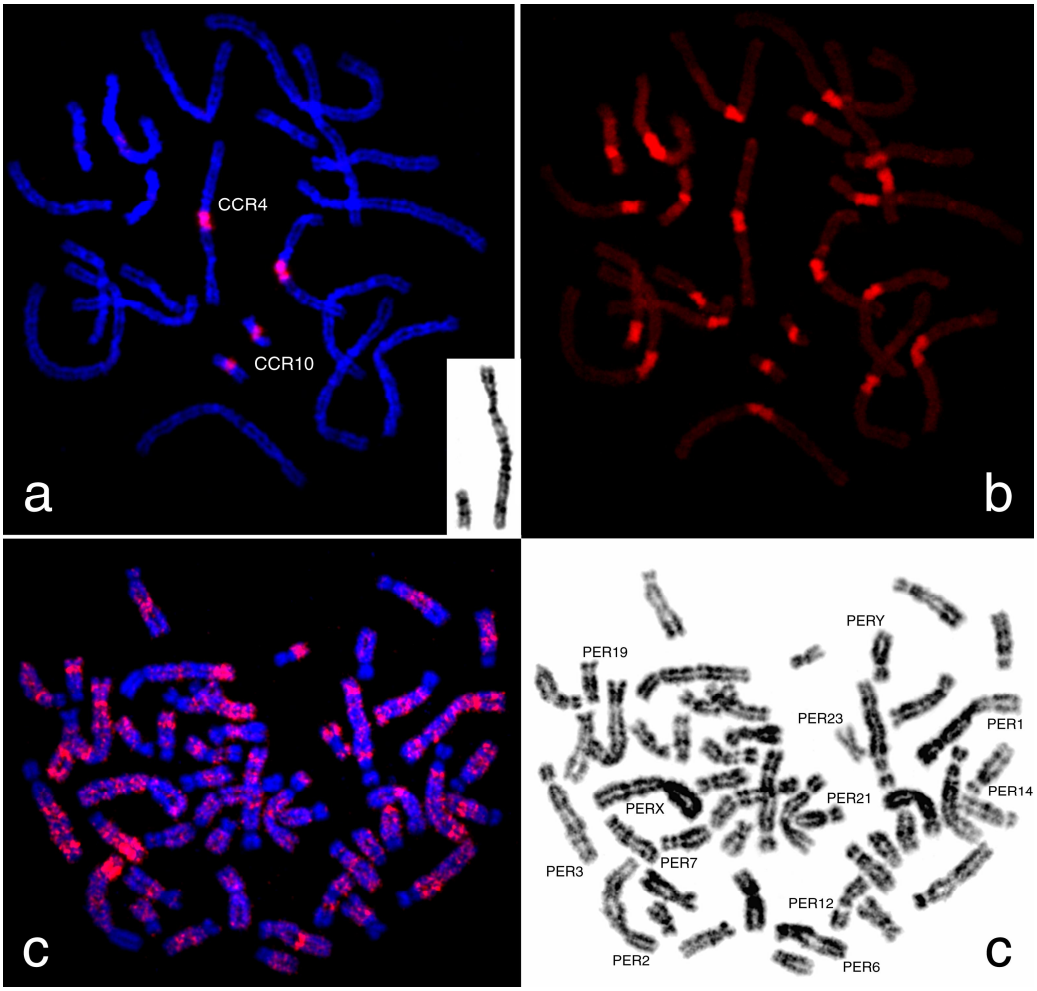


Fig. 1. Representative *in situ* hybridisation of CCR4/10sat sequences from *Cricetus cricetus* (CCR) onto *C. cricetus* chromosomes. CCR4 and CCR10 were DAPI inverted for chromosome identification (a). The same metaphase after C-banding (b). *In situ* hybridisation of CCR4/10sat onto *Peromyscus eremicus*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

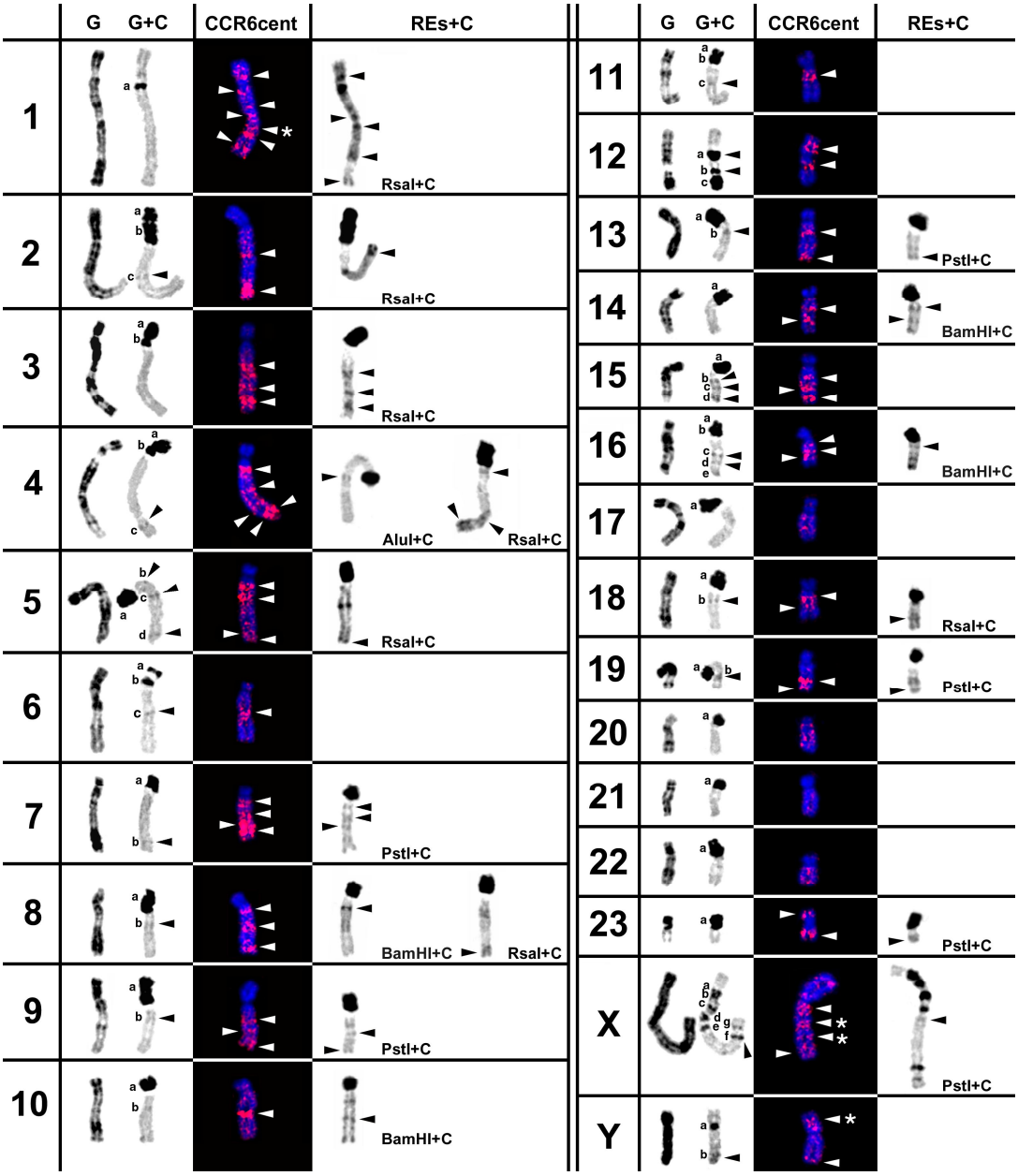


Fig. 2. Table showing the *in situ* hybridisation pattern of CCR4/10sat in *Peromyscus eremicus* chromosomes. G- and C-banding of each *P. eremicus* chromosomes are shown in the left column. The letters (a–g) represent the C-bands according to order of appearance in each chromosome. In the other columns it is possible to observe the hybridization 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.

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 colocalizes 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 hybridization 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 hybridization pattern of the CCR4/10sat probe (white arrowheads) and the action of the different REs used (AluI, BamHI, PstI, RsaI) with sequential Cbanding 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 hybridization 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 hybridization 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 scatteredpattern, 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 severalmechanisms (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

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 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., Fro¨ nicke, 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., Cserpa´ n, I., Praznovszky, T., Kereso¨ , 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., Lo´ pez-Ferna´ ndez, 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.

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.

II.2.2

Evolutionary story of a satellite DNA sequence from *Phodopus sungorus* (Rodentia, Cricetidae)

Evolutionary story of a satellite DNA sequence from *Phodopus sungorus* (Rodentia, Cricetidae)

A. Paço[1], F. Adega[1], N. Meštrović[3], M. Plohl[3], R. Chaves[1,2*]

[1]Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGBUTAD), Apdo 1013, 5001-801 Vila Real, Portugal

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

[3] Department of Molecular Genetics, Ruđer Bošković Institute, Bijenička 54, HR-10002 Zagreb, Croatia

*Author for Correspondence: Raquel Chaves, Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGB-UTAD), Vila Real, Portugal, +351259350841, rchaves@utad.pt

Type of manuscript: research article

Keywords: Rodentia, Satellite DNA, *Phodopus sungorus*, Chromosomal distribution, Copy number, Transcriptional activity

Running head: Evolutionary story of a satellite DNA sequence

Abstract

We isolated and characterized for the first time a satellite DNA sequence (PSUcentSat) from *Phodopus sungorus* (Cricetidae) genome, presenting a monomeric unit of ~ 330 bp. Physical mapping of PSUcentSat in *Phodopus sungorus* chromosomes showed large arrays of this sequence located at the heterochromatic (peri)centromere of five autosomal pairs and the Y-chromosome. FISH and southern blot hybridizations indicate the presence of orthologous PSUcentSat sequences in the genomes of four other rodent species belonging to Muridae and Cricetidae families, presenting a dispersed chromosomal distribution. Contrarily to what would be expected by parsimony rules, this chromosomal location indicates a scattered PSUcentSat distribution in the ancestor of Muridae/Cricetidae families, assuming afterwards, in a descendant genome, a restricted location to few chromosomes in a specific chromosomal region, the (peri)centromeric region. This peculiar PSUcentSat chromosomal distribution allowed us to propose a model to explain the evolution of this sequence in *Phodopus sungorus* genome. After divergence of the studied species, PSUcentSat was most probably highly amplified in the (peri)centromeric region of some chromosome pairs of this hamster. This amplification could be mediated through different recombinational mechanisms, as unequal crossing-over. The bouquet chromosome configuration (during the early prophase I) possibly has a role in this PSUcentSat selective amplification, providing physical proximity of centromeric regions between chromosomes with similar size and/or morphology. This seems particularly important for the acrocentric chromosomes of *Phodopus sungorus* (including the Y-chromosome), presenting all large arrays of PSUcentSat repeats in its (peri)centromeric region. The preservation of this sequence in the studied genomes and its (peri)centromeric amplification in *Phodopus sungorus*, strongly suggests functional significance. This is supported by the verification of PSUcentSat transcriptional activity in normal proliferative cells of the studied species, suggesting that this satellite transcription is not limited, for

example, to a developmental stage, as described for other satellites. Moreover, our results also show that the role of PSUcentSat cannot be restricted to its transcripts function(s), since not all PSUcentSat copies are transcriptionally active but are maintained in the genomes.

Introduction

The genomes of higher eukaryotes harbor large amounts of repeated sequences. According to their organization, two major classes can be distinguished, interspersed and tandem repeats. Satellite DNAs (satDNAs) are classified as highly tandem repeated sequences, located in heterochromatic regions preferentially around centromeres, but also at chromosome interstitial and terminal positions (reviewed in Adega et al. 2009). Structurally these sequences are commonly formed by long tandem arrays up to 100 Mb, presenting the monomers (or repeat units) a sequential arrangement positioned one after the other (e.g. Plohl et al. 2008).

It is generally accepted that satDNAs follow the principles of concerted evolution (e.g. Palomeque and Lorite 2008; Plohl et al. 2008), a non-independent evolution mode of satellite monomers within a genome. According to this evolution model, mutated satellite units could be spread or eliminated in the satellite arrays leading to homogenization of repeats, which occurs through mechanisms of non-reciprocal transfer within and between chromosomes, as gene conversion, unequal crossing-over, rolling circle replication/reinsertion and transposon-mediated exchange (Walsh 1987; Elder and Turner 1995; Dover 2002). The chromosome configuration during the early prophase I (bouquet configuration) may facilitate the homogenization process on non-homologous chromosomes, by the physical proximity between centromeres of chromosomes with a similar size and morphology (Brannan et al. 2001; Cazaux et al. 2011). As consequence of the independent action of these homogenization mechanisms in different genomes, orthologous satDNAs could present high differences in its monomer size, nucleotide sequence, copy number or chromosome organization and location (reviewed in Plohl et al. 2008).

To date, the knowledge about the genomic importance of satDNAs is limited, nevertheless several functions have been proposed to this eukaryotic genome fraction. It has been

suggested the involvement of satDNAs in functions as diverse as, centromeric activity (e.g. Marshall and Clarke 1995), tridimensional organization of the interphase nucleus (Manuelidis 1982) and a driver of genome reorganization during evolution (e.g. Wichman et al. 1991; Garagna et al. 1997). This last role of satDNAs is mainly justified by the high molecular dynamics of these repeats, which is a consequence of its evolution mode. Recent works however, also show that the overexpression of satDNAs is directly associated to the occurrence of chromosomal rearrangements. Centromeric and pericentromeric regions have long been regarded as transcriptionally inert portions of chromosomes. Nevertheless, an increasing number of studies in the past 10 years refute this idea and provide credible evidences that these regions are transcriptionally active in several biological contexts (e.g. Vourc'h and Biamonti 2011; Enukashvily and Ponomartsev 2013). In fact, the transcription of satDNAs seems to be a general phenomenon (reviewed in Ugarković 2005). In accordance to what has been described, satDNAs transcripts could act as long non-coding RNAs or as precursors of small interfering RNAs, which have an important role in epigenetic processes of chromatin remodeling/heterochromatin formation and in control of gene expression (reviewed in Vourc'h and Biamonti 2011; Bierhoff et al. 2013). The organismal developmental stage and the tissue-specific expression observed in some satDNAs unequivocally point to a regulatory role for these transcripts, although the real function for most satellite transcripts is still unclear (Vourc'h and Biamonti 2011).

The species studied in this work belong to the Cricetidae and Muridae families, the most specious rodent families (Musser and Carleton 2005). Based on molecular data, the divergence time between Muridae/Cricetidae can be estimated at 17 Million years (My) (Robinson et al. 1997). From the studied species, *Rattus norvegicus* is the only that belongs to the Muridae family. *Cricetus cricetus, Peromyscus eremicus, Phodopus roborovskii* and *Phodopus sungorus* all belong to Cricetidae family, according to the classification of Tree of

Life web project (http://www.tolweb.org/tree/). In this work we report the isolation and molecular characterization of a *Phodopus sungorus* satDNA, PSUcentSat. This novel satellite is located exclusively at the heterochromatic (peri)centromeric region of five *Phodopus sungorus* autosomal pairs and in the Y-chromosome. *In situ* and southern blot hybridizations suggest the presence of PSUcentSat orthologous sequences in the other four rodent studied species. The transcriptional activity of this sequence was verified in normal proliferative fibroblast cell. Our data strongly suggest a functional significance of PSUcentSat in the studied genomes.

Material and Methods

Chromosome preparations and Genomic DNA extraction

Fixed chromosome preparations from *Cricetus cricetus* (CCR), *Peromyscus eremicus* (PER), *Phodopus roborovskii* (PRO), *Phodopus sungorus* (PSU) and *Rattus norvegicus* (RNO) were obtained from fibroblast cell cultures, using standard procedures described elsewhere (Chaves et al. 2004). Genomic DNA of the different species was obtained from these fibroblast cell cultures using the JETQUICK DNA kit (Genomed).

Isolation, cloning and sequencing of PSUcentSat sequence

PSU genomic DNA was digested with the restriction endonuclease (RE) MboI, according to the manufacturers' instructions (Invitrogen Life Technologies), resulting in a smear with DNA fragments ranging between 3 kb to 100 bp. The restriction products were later inserted into the plasmid vector pUC19 (FERMENTAS Life Science) and used to transform competent *Escherichia coli* DH5α cells (Invitrogen Life Technologies). A part of the obtained colonies were transferred onto a nylon membrane Hybond^{TM}-N^{+} (Amersham, GE Healthcare) and the

DNA in the membrane probed to MboI restriction products labeled with digoxigenin-11-d'UTP, using DIG DNA labeling Kit (Roche Diagnostics). Hybridization was performed at 68ºC as described by Bruvo et al. (2003). The positive signals were visualized using chemiluminiscent CDP-Star system (Roche Diagnostics). The plasmid DNA of the positive clones, which included PSUcentSat sequence, was isolated using the High Pure Plasmid Isolation kit (Roche Diagnostics) and sequenced in both directions using M13 primers.

Sequence analysis of PSUcentSat sequence

PSUcentSat was analyzed with different sequence database tools and bioinformatic softwares: NCBI Blast (http://www.ncbi.nlm.gov/Blast/), RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker), (EMBOSS CpG plot (http://www.ebi.ac.uk/Tools/emboss/cpgplot/), EMBOSS einverted (http://emboss.bioinformatics.nl/cgi-bin/emboss/einverted), Tandem repeats Finder (Benson 1999, version 4.00, free download in http://tandem.bu.edu/trf/trf.html) and vector NTI advance 11 (Invitrogen Life Technologies). A BLAST search of PSUcentSat sequence against nucleotide sequences existent in Genbank and RepBase was accomplished using NCBI blast and RepeatMasker tools. Sequence alignments were performed with the software vector NTI advance 11 that apply the Clustal W algorithm (Thompson et al. 1994) to determine sequence similarities. The search for direct or inverted repeats within the PSUcentSat sequence was done using the software Tandem Repeats Finder and the EMBOSS einverted tool, respectively. EMBOSS einverted tool was used with a minimum score threshold of 20%. The EMBOSS CpG plot allowed the identification of CpG islands. Sequence data from the PSUcentSat clone were deposited in the NCBI Nucleotide database with the follow accession number: KJ649148.

Physical mapping of PSUcentSat sequence

Physical mapping of PSUcentSat in the chromosomes of the five studied species (CCR, PER, PRO, PSU and RNO) was carried out following Fluorescent *in situ* Hybridization (FISH) procedures, described by Schwarzacher and Heslop-Harrison (2000). The PSUcentSat sequence was labeled with digoxigenin-11-d-UTP (Roche, Molecular Biochemicals) by PCR amplification. The most stringent post-hybridization wash was in 50% formamide/2×SSC at 42 ºC. Digoxigenin-labeled probes were detected with antidigoxigenin-5'TAMRA (Roche Molecular Biochemicals).

CBP-banding sequential to physical mapping of PSUcentSat sequence

After distaining the slides, CBP-banding (C-bands by Barium Hydroxide with Propidium Iodide) was performed according to the standard procedure of Sumner (1972) with slight modifications. Shortly, the slides were submitted to hydrochloric acid (0.1 M) during 20 minutes, barium hydroxide (5% solution) during 7 minutes and 2× saline sodium citrate at 60 ºC for 40 minutes.

Capture and preparation of images

Chromosomes were observed in a Zeiss Axioplan 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); contrast and color optimization were the functions used and affected the whole image equally. The chromosomes of PSU were identified according to Romanenko et al. (2007), and RNO chromosomes according to Levan (1974).

Southern hybridization analysis

Genomic DNA from PSU was digested with the endonucleases AluI, HhaI and MboI. Genomic DNA of the other studied rodent species (CCR, PER, PRO and RNO) was digested with AluI and MboI. The resulting fragments were separated in a 0,8% agarose gel and blotted onto a Nylon membrane Hybond^{TM}-N^{+} (Amersham, GE Healthcare) by alkali capillary blotting, using 0.4 M NaOH. The membranes were then probed with the cloned PSUcentSat sequence, previously labeled by PCR with digoxigenin-11-d'UTP (Roche Diagnostics). Hybridization was performed at 42ºC in hybridization solution (Roche Diagnostics). The positive signals were visualized using chemiluminiscent CDP-Star system (Roche Diagnostics). Selection of REs was performed using the CLC Sequence Viewer software (version 6.2, http://www.clcbio.com/index.php?id=28).

Satellite DNA copy number quantification (absolute and relative) by TaqMan assay

For PSUcentSat quantification it was performed a quantitative real-time PCR approach as previous described in Louzada et al. (submitted for publication). TaqMan specific assay mix (primers/probe) was designed using Primer Express® Software v3.0 (Life Technologies Applied Biosystems) based in PSUcentSat sequence. PCR primers PSUcentSat F (5'^{145}GCTACACTGCGCAAGAGAGATAAG3') and PSUcentSat R (5'^{209} GAGACGCTTT TCGCGAATGCTGTC3') locate between the positions 146 and 210 bp of PSUcentSat sequence, allowing the amplification of a 64 bp product. The probe [5'^{170}(FAM) CACTGTGAGAGTAAAGAG3' (NFQ)] had the fluorescent reporter dye, 6-carboxy-fluorescein (FAM), located at the 5' end and the non-fluorescent quencher (NFQ) located at the 3' end.

For PSUcentSat absolute quantification in PSU genome, the standard curve method was performed. A 10-fold serial dilution series of the plasmid DNA standard, 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 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 3014 bp (pUC 19 vector 2686 bp and the insert 328 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 PSUcentSat. Briefly, the standard curve includes a plot of the C_T values *versus* the log concentration of the plasmid DNA standard. For PSU genomic DNA, the unknown total DNA sample was obtained by interpolating its C_T value against the standard curve. We used 1 ng and 5 ng of PSU genomic DNA in the PCR reactions. 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. This experiment was carried out in StepOne real-time PCR system (Life Technologies Applied Biosystems), where the samples were subjected to an initial denaturation at 95 ºC (10 minutes), and then to 40 cycles at 95 ºC for 15 seconds followed by 60 ºC for 1 minute. 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 analyze the data. 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 PSUcentSat allowed determining the copy number of this sequence in PSU genome to 1 and 5 ng, which comprises 333 and 1667 haploid genomes, respectively.

For PSUcentSat quantification within the other species genomes (CCR, PER, PRO and RNO) a relative quantification real-time PCR approach was used, being the PSU genome the control sample. We used the same PSUcentSat TaqMan assay described for the absolute quantification and the 18S gene (HS99999901_s1, Life Technologies Applied Biosystems) as the reference assay. For this comparative analysis, PCR reactions were performed with 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 18S gene. The $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) was used to calculate fold changes in the amount of PSUcentSat in the different species. Results are shown as the $\log_{10}$ of $2^{-\Delta\Delta CT}$ PSUcentSat copy number in CCR, PER, PRO and RNO relatively to PSU (control 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.

As it is not yet available information about the genome size (bp) and mass (pg) of PSU genome, we considered that the haploid PSU genome presents approximately 3.10^9 bp and weights 3 pg, according to the size and mass of other Cricetidae genomes in the Animal Genome Size database (http://www.genomesize.com/). The same was considered for PRO haploid genome. The genome mass of *Cricetus cricetus*, *Peromyscus eremicus* and *Rattus norvegicus* is approximately 3.44 pg, 3.3 pg and 3.2 pg (respectively).

RNA isolation and reverse transcription quantitative real-time PCR

Total and small RNA from a PER, PSU and RNO fibroblast cell line was isolated using the mirVana Isolation Kit (Ambion, Invitrogen Life technologies), following the manufacturer's recommendations. Expression experiments were performed using the TaqMan® RNA-to-C_T™ 1-Step Kit (Life Technologies Applied Biosystems). We used the same PSUcentSat TaqMan assay described previously as target and as reference assay the Glyceraldehyde-3-phosphate dehydrogenate (GAPDH, Rn01749022_g1, Life Technologies Applied Biosystems). The 20 µl reactions included 5 µl of RNA sample (50 ng/µl from total or small RNA), 1 µl of the primer/probe assay mixture, 10 µl of PCR Master Mix, 0.5 µl of RT enzyme mix (Life Technologies Applied Biosystems) and 3.5 µl of DEPC-treated water. This experiment was carried out in the StepOne real-time PCR system (Life Technologies Applied Biosystems), where the samples were subjected to 48 ºC for 15 minute and 95 ºC for 10 minute, followed by 40 cycles of 95 ºC for 15 seconds and 60 ºC for 1 minute. 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 relative expression level was normalized with GAPDH gene expression. The $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001) was used to calculate fold changes in the expression levels of the PSUcentSat sequence in different genomes, using the expression in PSU as control. Besides, the fold changes in the expression levels of total and small RNA in each species were calculated using total RNA as control.

Results

Molecular analysis of PSUcentSat

In this work it was isolated, sequenced and molecularly characterized a novel satellite DNA (satDNA) in the genome of *Phodopus sungorus* (PSUcentSat). BLAST search revealed no significant similarity between this sequence and any other described and deposited in Genbank or in RepBase databases. As can be observed in figure 1, direct and inverted short internal repeats were detected within PSUcentSat sequence (PSUcentSat monomer whose length was determined by southern blot analysis as described below). Namely, two different short direct repeats with more than 11 bp: a repeat with 11 bp (two copies with 94% similarity) and a repeat with 17 bp (two copies with 100% similarity), a GT rich region presenting 19 tandem GT repeats and an inverted short repeat with 13 bp (two copies with 84% similarity). A CpG analysis also allowed the identification of a CpG island with 107 bp between the positions 135 to 241 bp (cf. figure 1).

Physical distribution of the PSUcentSat in chromosomes of five rodent species

Physical mapping of PSUcentSat in *Cricetus cricetus*, *Peromyscus eremicus*, *Phodopus roborovskii*, *Phodopus sungorus* and *Rattus norvegicus* genomes was performed by FISH. *Rattus norvegicus* was used as outgroup for this analysis since it is the only species outside Cricetidae (belongs to Muridae). In *Phodopus sungorus* genome PSUcentSat presents a chromosome distribution characteristic of a tandem repeat sequence, being organized as large blocks at the (peri)centromeric region in five autosomal pairs and in the Y-chromosome, PSU6, PSU8, PSU10, PSU11, PSU12 and PSUY (cf. figure 2a). C-banding sequentially to FISH (Figure 2b) evidenced a co-localization of this sequence with constitutive heterochromatin (CH), as can be seen in Figure 2c. In the other four species, PSUcentSat

presents a scattered distribution along all the chromosomes of the complement (as can be observed in figure 2d for *Rattus novergicus* chromosomes). Besides, the majority of the (peri)centromeric regions in these four species presents a depletion of the sequence (some of these regions are evidenced by arrowheads in figure 2d), being PSUcentSat only located in a few chromosome pairs, at the (peri)centromeric regions.

Genomic organization of PSUcentSat

In order to investigate the genomic organization of PSUcentSat in the five studied rodent species, southern blot analyses were carried out. The ladder hybridization pattern obtained for PSUcentSat in *Phodopus sungorus* using AluI, HhaI and MboI enzymes (Figure 3a), indicates the tandem organization of this sequence, characteristic of a satellite sequence. A common band with approximately ~ 330 bp was obtained with all the enzymes used (monomer), and other bands were also observed showing a 330 bp periodicity: 660 bp (dimmer) and 990 bp (trimmer). The enzymes used in these analyses cut only once the PSUcentSat clone, allowing the determination of PSUcentSat monomer length (bp), since these enzymes in *Phodopus sungorus* genome cut the PSUcentsat arrays with a periodicity that corresponds to the monomer length, or multiples of it. According to similarities in size (bp), we assumed that PSUcentSat monomers present a very similar sequence with the PSUcentSat clone isolated here, presenting a length of 328 bp and an AT content of ~56%. As can be seen in the figure 3b, the southern hybridization pattern obtained for *Cricetus cricetus, Peromyscus eremicus, Phodopus roborovskii* and *Rattus norvegicus* species is not indicative of a tandem organization for the PSUcentSat in these genomes. Contrary to what occurs in the *Phodopus sungorus*, a scattered pattern of hybridization was observed for these four species (cf. figure 3b). *Phodopus sungorus* genomic DNA digested with AluI and MboI was used as the experiment control.

PSUcentSat DNA copy number analysis

A satellite copy number quantification, performed using a new methodology based in real time quantitative PCR allied to Taqman chemistry (as described in Louzada et al. submitted for publication), shows significant differences in the copy number of PSUcentSat in the five studied genomes. Absolute quantification, using a standard curve (cf. figure 4a), revealed that at least 0.2% of *Phodopus sungorus* haploid genome is comprised by PSUcentSat, corresponding to at least 17895 copies per haploid genome. Considering that PSUcentSat can present different monomer variants and here we only analyzed one, the copy number estimated by this approach for *Phodopus sungorus* is considered the minimal number of copies that this satDNA can present in this genome. Relative quantification showed that the amount of PSUcentSat in the other species genome is lower than in *Phodopus sungorus* (940 to 7000 times lower) (Figure 4b), presenting all the results statistically significant values ($p<0.05$). From the other analyzed species, it is the *Rattus norvegicus* genome that presents the lower number of copies of PSUcentSat (~7000 times lower) in comparison with the genome of *Phodopus sungorus*.

Transcription analysis of PSUcentSat satellite sequence

In this work we verified PSUcentSat transcription in total and small RNA isolated from normal proliferative fibroblast cells of *Peromyscus eremicus*, *Phodopus sungorus* and *Rattus norvegicus*. Figure 5 resumes the results of the relative reverse transcription quantitative real time PCR (RT-qPCR) quantification, in terms of the fold change in PSUcentSat RNA expression, normalized using GAPDH gene expression and calculated relatively to *Phodopus sungorus* PSUcentSat expression (expression in different genomes) or relatively to total RNA PSUcentSat expression (expression in each genome). The levels of PSUcentSat transcription in both total and small RNA is higher in *Peromyscus eremicus* and lower in *Rattus*

norvegicus, relatively to what happens in *Phodopus sungorus* (cf. figures 5a and 5b). In *Peromyscus eremicus*, PSUcentSat transcription in small RNA is higher relatively to the transcription in total RNA (cf. figure 5d). The expression values presented in figures 5a, 5b and 5d were considered statistically significant following analyses using Student's t-test with a p value <0.05. In *Phodopus sungorus* and *Rattus norvegicus* the differences in the transcription level for both RNA fractions were considered statistically non significant (cf. figures 5c and 5e).

Discussion

As far as we know, this report corresponds to the first study describing a satellite DNA (satDNA) sequence (PSUcentSat) from *Phodopus sungorus* genome. BLAST search revealed no significant similarity between PSUcentSat and any other described DNA sequence, both in Genbank or in RepBase databases, indicating that this sequence correspond to a novel satellite described.

The study of PSUcentSat genomic organization in *Phodopus sungorus* shows that this sequence presents a monomer length with ~ 330 bp. As revealed by sequence analysis, different short direct and inverted repeat submotifs were identified within PSUcentSat monomer. SatDNAs from different organisms, as primates, cattle, rodents, nematodes and insects, also present internal short repeats (e.g. Miklos and Gill 1982; Singer 1982; Modi 1992; Modi 1993; Castagnone-Sereno et al. 2000; Modi et al. 2003; Lorite et al. 2004; Mravinać et al. 2004). The functional significance of this internal repeats is unclear, however it has been assumed that these short repeats are associated with chromatin conformation (Modi 1993; Plohl 2010). Secondary and tertiary structures of the DNA molecule can be induced by particular distribution of nucleotides (reviewed by Plohl 2010). According to this, the identified repeats' submotifs may play a role in PSUcentSat chromatin conformation,

which probably has an effect in the homogenization of these satellite repeats by recombinational mechanisms (e.g. unequal crossing-over), and also in its accessibility for transcription.

Physical mapping of PSUcentSat in *Phodopus sungorus* chromosomes showed large arrays of this sequence on heterochromatic (peri)centromere of five autosomal pairs and in the Y-chromosome (cf. figure 2a). FISH and southern blot hybridizations indicated the presence of orthologous PSUcentSat sequences in the genomes of four other rodent species, belonging to Cricetidae and Muridae families, *Cricetus cricetus, Peromyscus eremicus, Phodopus roborovskii* and *Rattus norvegicus*. In fact, a new methodology based in real time quantification also confirmed the presence of PSUcentSat in these genomes, but estimates a lower PSUcentSat copy number in these species comparing with *Phodopus sungorus* (940 to 7000 fold lower), presenting the sequence, in these genomes, a dispersed hybridization pattern (cf. figure 2d). Therefore, regarding PSUcentSat chromosomal distribution in the considered outgroup species (*Rattus norvegicus*) and contrarily to what would be expectable by parsimony rules, we can conclude that in an ancestor of Muridae/Cricetidae families (diverged at ~17 My according Robinson et al. 1987), PSUcentSat presented a scattered distribution, assuming after a restricted location to few chromosomes in a specific chromosomal region ((peri)centromeric region) in a descendant genome. Furthermore, during its evolution, PSUcentSat also altered its genomic organization, from an initial interspersed organization turned out to be highly amplified as a tandem repeat (satDNA), in *Phodopus sungorus*. The reasons for this different PSUcentSat amount, chromosomal location and genomic organization in the studied genomes is enigmatic, however its presence evidences a functional significance for this sequence. The preservation of this sequence during at least ~17My corroborates its importance in the studied genomes. Moreover, the amplification and maintenance of PSUcentSat as large arrays located in the (peri)centromeric regions of some

Phodopus sungorus chromosomes brings, most probably, an adaptive advantage to this species, possibly in the centromeric function.

The peculiar chromosomal distribution of PSUcentSat allows us to propose a model to explain the evolution of this sequence in the genome of *Phodopus sungorus*. We believe that initially, in this genome, PSUcentSat presented a lower copy number and a dispersed distribution, as it is observed in the other studied species. Later, PSUcentSat was highly amplified in the (peri)centromeric region of some chromosome pairs, originating large PSUcentSat repeat arrays. The high level of amplification of PSUcentSat found in these chromosomes may have been mediated through different recombinational mechanisms, as unequal crossing-over and rolling circle amplification/reinsertion. Most probably, not all *Phodopus sungorus* chromosomes displayed this sequence in its (peri)centromeric region, as it was observed in the other studied species (cf. figure 2d), thus restricting PSUcentSat amplification to only a few chromosomes. The bouquet chromosome configuration (during early prophase I) possibly also has played an important role in PSUcentSat selective amplification, as it provides physical proximity of centromeric regions of chromosomes with similar size and/or morphology. In this stage, all chromosomes migrate to one area of the nucleus and adopt an orientation in which all telomeres attach to the nuclear membrane (Scherthan et al. 1996). All the acrocentric chromosomes, independently of their size, will have more proximity between the (peri)centromeric regions during the bouquet stage, favoring the occurrence of recombinational events in these regions (if homology exists). In *Phodopus sungorus* genome, all acrocentric chromosomes (PSU11, PSU12 and PSUY) present large arrays of PSUcentSat repeats at the (peri)centromeric region. This explains why PSUcentSat was amplified in the Y but not in the X-chromosome. Moreover, the analysis of synaptonemal complexes between the sex-chromosomes of *Phodopus sungorus* (Spyropoulos et al. 1982) shows that the (peri)centromeric regions of the sex chromosomes do not pair,

supporting our theory and justifying the apparently absence of this satellite in the X-chromosome. Simultaneously with PSUcentSat amplification in some (peri)centromeric regions, the dispersed PSUcentSat sequences initially present in *Phodopus sungorus* chromosomes were probably reduced in its copy number or eliminated, as these were not detected in the FISH analysis (cf. figure 2a).

In this work we also report transcriptional activity of PSUcentSat in normal proliferative fibroblast cells of *Peromyscus eremicus*, *Phodopus sungorus* and *Rattus norvegicus*. This transcription was demonstrated in total and small RNA of these cells, being the level of transcription significantly higher in *Peromyscus eremicus* in comparison to *Phodopus sungorus*. These results are very interesting as they show that not all PSUcentSat copies are transcriptionally active. For instance, *Peromyscus eremicus* exhibits a lower number of copies of PSUcentSat than *Phodopus sungorus*, but it displays, however, a higher transcription level of this sequence. In the literature it is frequent the description of satDNAs temporally transcribed at particular developmental stages or at different cell types, tissues or organs (reviewed by Ugarković 2005). As an example, the major satellite in mouse is differently expressed during development of the central nervous system, as well as in the adult liver and testis (Rudert et al. 1995). In contrast, PSUcentSat transcripts were detected in normal proliferative cells, which point to an important functionality of these transcripts. The transcription analysis here conducted also showed that in *Peromyscus eremicus* the transcription level is significantly higher in small RNA relatively to total RNA (cf. figure 5d), which leads us to assume the possibility that the transcripts of this sequence can result in small interfering RNAs. Interestingly, in this satellite sequence, a CpG island was identified, which can be associated with its transcription regulation. In human and mouse, approximately 60% of all promoters colocalizes with CpG islands (Antequera 2003).

Future works focused in a complete characterization of PSUcentSat RNAs (transcripts length, occurrence of single or both strands transcripts and sub-cellular localization) will certainly enlighten the functional significance of this repeated sequence. Nevertheless, in this context it is important to emphasize that the role of PSUcentSat in the studied genomes could not to be restricted to the function of the transcripts, since not all PSUcentSat copies are transcriptionally active but are maintained in the genomes. These copies that are not being transcribed (in our analysis) possibly have a structural role. However, we cannot ignore the possibility of these PSUcentSat copies are active in other cells or during different development stages of *Phodopus sungorus*. In any case, PSUcentSat most probably exhibit different functions or "ways of action" in the studied genomes.

Acknowledgements

This work was supported by the project POCI/BIA-BCM/58541/2004, a PhD grant (SFRH/BD/41574/2007) 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.

References

Acosta MJ, Marchal JA, Fernández-Espartero C, Romero-Fernández I, Rovatsos MT, Giagia-Athanasopoulou EB, Gornung E, Castiglia R, Sánchez A. 2010. Characterization of the satellite DNA Msat-160 from species of Terricola (*Microtus*) and Arvicola (Rodentia, Arvicolinae). Genetica 138: 1085–1098.

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.

Antequera F. 2003. Structure, function and evolution of CpG island promoters. Cell. Mol. Life Sci. 60: 1647–1658.

Benson G. 1999. “Tandem repeats finder: a program to analyze DNA sequences". Nucleic Acids Res 27 (2): 573-580.

Bierhoff H, Postepska-Igielska A, Grummt I. 2013. Noisy silence: Non-coding RNA and heterochromatin formation at repetitive elements. Epigenetics 9 (1).

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.

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.

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. Genetica 138: 1201–1210.

Castagnone-Sereno P, Leroy F, Abad P. 2000. Cloning and characterization of an extremely conserved satellite DNA family from the root-knot nematode *Meloidogyne arenaria.* Genome 43: 346–353.

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

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

Dover G. 2002. Molecular drive. Trends Genet 18: 587–589.

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: 283–290.

Enukashvily NI, Ponomartsev NV. 2013. Mammalian Satellite DNA: A Speaking Dumb. Advances in Protein Chemistry and Structural Biology 90: 31-65.

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.

Garagna S, Pérez-Zapata A, Zuccotti M, Mascheretti S, Marziliano N, Redi CA, Aguilera M, Capanna E. 1997. Genome composition in Venezuelan spiny-rats of the genus *Proechimys* (Rodentia, Echimyidae). I. Genome size, C-heterochromatin and repetitive DNAs *in situ* hybridization patterns. Cytogenet Cell Genet 78: 36-43.

Horz W, Zachau HG. 1977. Characterization of distinct segments in mouse satellite DNA by restriction nucleases. Eur J Biochem 73: 383-392.

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

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.

Lorite P, Carrillo JA, Tinaut A, Palomeque T. 2004. Evolutionary dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera, Formicidae). Gene 332:159–168.

Louzada S, Vieira-da-Silva A, Kubickova S, Rubes J, Adega F, Raquel Chaves R. 2014. A novel satellite DNA sequence in *Peromyscus* genome that evolves by copy number fluctuation: PMSat Is it an "ancient" or a "nascent" satellite DNA?. Submitted for publication to Heredity.

Manuelidis L. 1982. Repeated DNA sequences and nuclear structure. In: Dover GA, Flavell RB, editors. Genome evolution. Academic Press, London. p. 263–285.

Marshall LG, Clarke L. 1995. A novel cis-acting centromeric DNA elements affects *S. pombe* chromatin structure at a distance. J Cell Biol 128: 445–454.

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: 1062–1068.

Miklos G, Gill. 1982. Nucleotide sequences of highly repeated DNAs, compilation and comments. Genet Res 39:1-30.

Modi WS. 1992. Nucleotide sequence and genomic organization of a tandem satellite array from the rock vole *Mierotus ehrotorrhinus* (Rodentia). Mamm Genome 3: 226-232.

Modi WS. 1993. Rapid, localized amplification of a unique satellite DNA family in the rodent *Microtus chrotorrhinus.* Chromosoma 102: 484-490.

Modi WS, Serdyukova NA, Vorobieva NV, Graphodatsky AS. 2003. Chromosomal localization of six repeated DNA sequences among species of *Microtus* (Rodentia). Chromosome Res 11: 705-713.

Mravinać B, Plohl M, Ugarković Đ. 2004. Conserved patterns in the evolution of *Tribolum* satellite DNAs. Gene 332: 169–177.

Musser G, Carleton M. 2005. Superfamily Muroidea. In: Wilson DE, Reeder DM, editors. Mammal Species of the World. Washington DC: Smithsonian Institution Press. p. 894-1522.

Palomeque T, Lorite P. 2008. Satellite DNA in insects: a review. Heredity 100: 564–573.

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. 2010. Those mysterious sequences of satellite DNAs. Period Biol 112 (4): 403–410.

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.

Romanenko SA, Volobouev VT, Perelman PL, et al. (12 co-authors). 2007. Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroideia, Rodentia) inferred from chromosomal painting and banding comparison. Chromosome Res. 15: 283-297.

Rouleux-Bonnin F, Renault S, Bigot Y, Periquet G. 1996. Transcription of four satellite DNA subfamilies in *Diprion pini* (Hymenoptera, Symphyta, Diprionidae). Eur J Biochem 238: 752–759.

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

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.

Schwarzacher T, Heslop-Harrison JS. 2000. Practical *in situ* hybridization. Oxford: Bios Scientific Publishers.

Singer MF. 1982. Highly repeated sequences in mammalian genomes. Int Rev Cytol 76: 67-112.

Spyropoulos B, Ross PD, Moens PB, Cameron DM. 1982. The Synaptonemal Complex Karyotypes of Palearctic Hamsters, *Phodopus roborovskii* Satunin and *P. sungorus* Pallas. Chromosoma 86: 397-408.

Sumner AT. 1972. A simple technique for demonstrating centromeric heterochromatin. Exp Cell Res 75:304-306.

Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22): 4673-4680.

Ugarković Đ. 2005. Functional elements residing within satellite DNAs. EMBO rep 6(11): 1035-1039.

Vourc'h C, Biamonti G. 2011. Long Non-Coding RNAs. Progress in Molecular and Subcellular Biology. In: Ugarkovic D, editor. Transcription of Satellite DNAs in Mammals. Springer-Verlag Berlin Heidelberg. p. 95-118.

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

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: 369-377.

Figures legends

Figure 1- Organization of PSUcentSat. Schematic representation of PSUcentSat molecular features and monomer length. Colored lines indicate the region for which TaqMan specific assay mix (primers/probe) was designed, and was used for copy number quantification and transcription analysis. Blue line corresponds to the PSUcentSat forward primer and the green line corresponds to the PSUcentSat reverse primer.

Figure 2- Physical mapping of PSUcentSat on chromosomes of *Phodopus sungorus* and *Rattus norvegicus*. (a) Representative *in situ* hybridization presenting the chromosomal localization of PSUcentSat on chromosomes of *Phodopus sungorus* (PSU). The sequence was labeled with digoxigenin-11-dUTP and detected with 5'TAMRA (red), but here it is presented in the pseudo color green. Chromosomes were counterstained with DAPI (blue). (b) Same metaphase after sequential C-banding. Chromosomes were counterstained with Propidium Iodide (red). (c) Overlapping of PSUcentSat hybridization signals with C-bands. (d) Representative *in situ* hybridization presenting the chromosomal localization of PSUcentSat on chromosomes of *Rattus norvegicus* (RNO). Arrowheads evidence a depletion of PSUcentSat at the (peri)centromeric regions of some RNO chromosomes.

Figure 3- Southern blot analysis. (a) Electrophoresis separation of *Phodopus sungorus* genomic DNA after digestion with AluI, HhaI and MboI (shown on the left). The corresponding southern blot obtained after hybridization with PSUcentsat is shown on the right. (b) Electrophoresis separation of *Cricetus cricetus* (CCR), *Peromyscus eremicus* (PER), *Phodopus roborovskii* (PRO) *and Rattus norvegicus* (RNO) genomic DNA after digestion with AluI and MboI (shown on the left). The corresponding southern blot obtained after hybridization with PSUcentsat is shown on the right.

Figure 4- PSUcentsat copy number quantification. (a) Standard calibration curve used in the absolute quantification of PSUcentSat copy number in the genome of *Phodopus sungorus* (PSU). For this analysis were used 1 ng and 5 ng of genomic DNA (gDNA PSU). The two groups of blue cubes indicate the copy number estimated for 1 ng and 5 ng of gDNA. (b) Relative quantification (represented as $\log_{10}$) of PSUcentSat in *Cricetus cricetus* (CCR),

Phodopus roborovskii (PRO), *Peromyscus eremicus* (PER) and *Rattus norvegicus* (RNO) using PSU as control. Error bars represent ± SD.

Figure 5- Relative expression analysis of PSUcentSat in fibroblast cells of *Phodopus sungorus, Peromyscus eremicus* and *Rattus norvegicus*. (a) Relative expression analysis of PSUcentSat in total RNA from fibroblast cells of *Phodopus sungorus* (PSU), *Peromyscus eremicus* (PER) and *Rattus norvegicus* (RNO). (b) Relative expression analysis of PSUcentSat in small RNA from fibroblast cells of PSU, PER and RNO. Expression results were obtained by reverse transcription quantitative real time PCR (RT-qPCR), normalized with the expression of the reference gene glyceraldehyde-3-phosphate dehydrogenate (GAPDH) and the PSUcentSat expression in PER and RNO genomes compared with the expression in PSU genome (control). (c) Relative expression analysis of PSUcentSat in total and small RNA from a fibroblast cells of PSU; (d) PER; (e) and RNO. Expression results were obtained by RT-qPCR, normalized with the expression of GAPDH gene and PSUcentSat expression in small RNA compared with the expression in the total RNA (control). Data is presented as mean corresponding to fold change relative to the control sample ($p < 0.05$). Error bars represent ± SD.

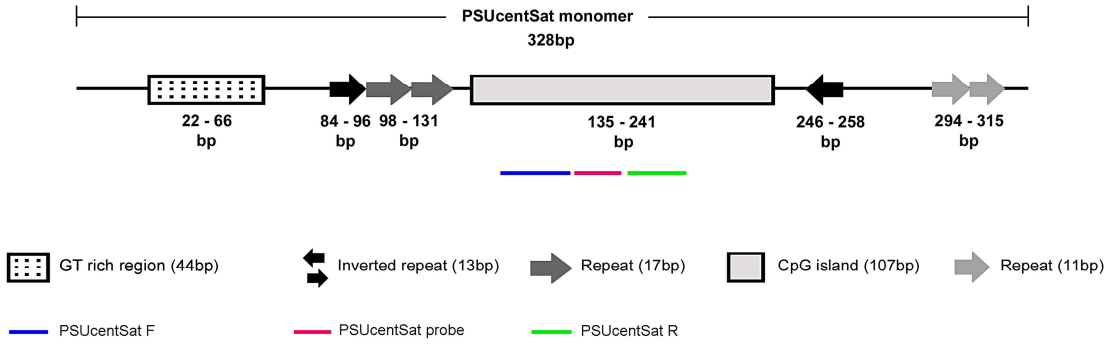


Figure 1

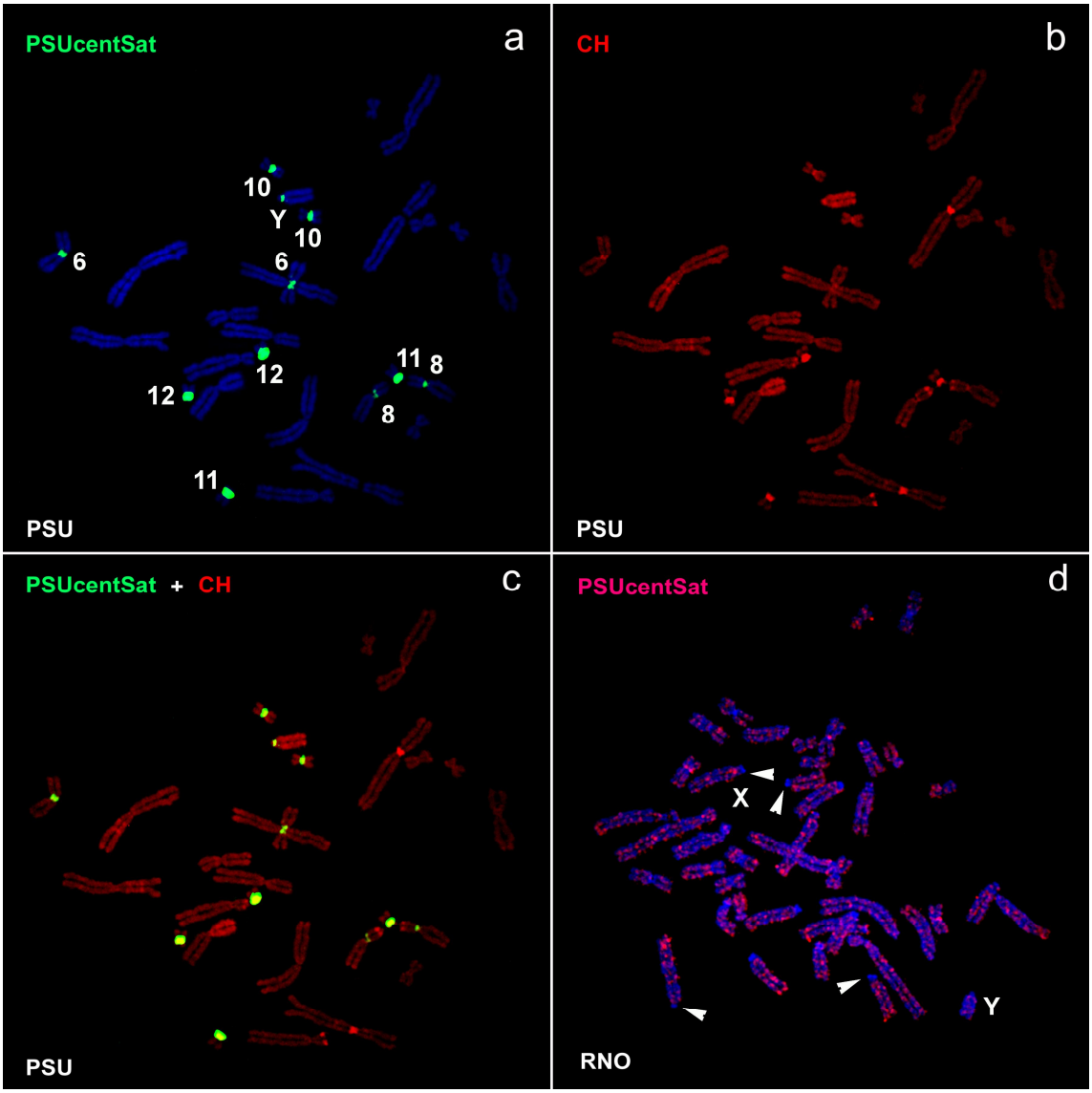

PSUcentSat
a
10
Y
10
6
6
12
12
11
8
8
11
PSU
CH
b
PSU
PSUcentSat + CH
c
PSU
PSUcentSat
d
X
Y
RNO

Figure 2

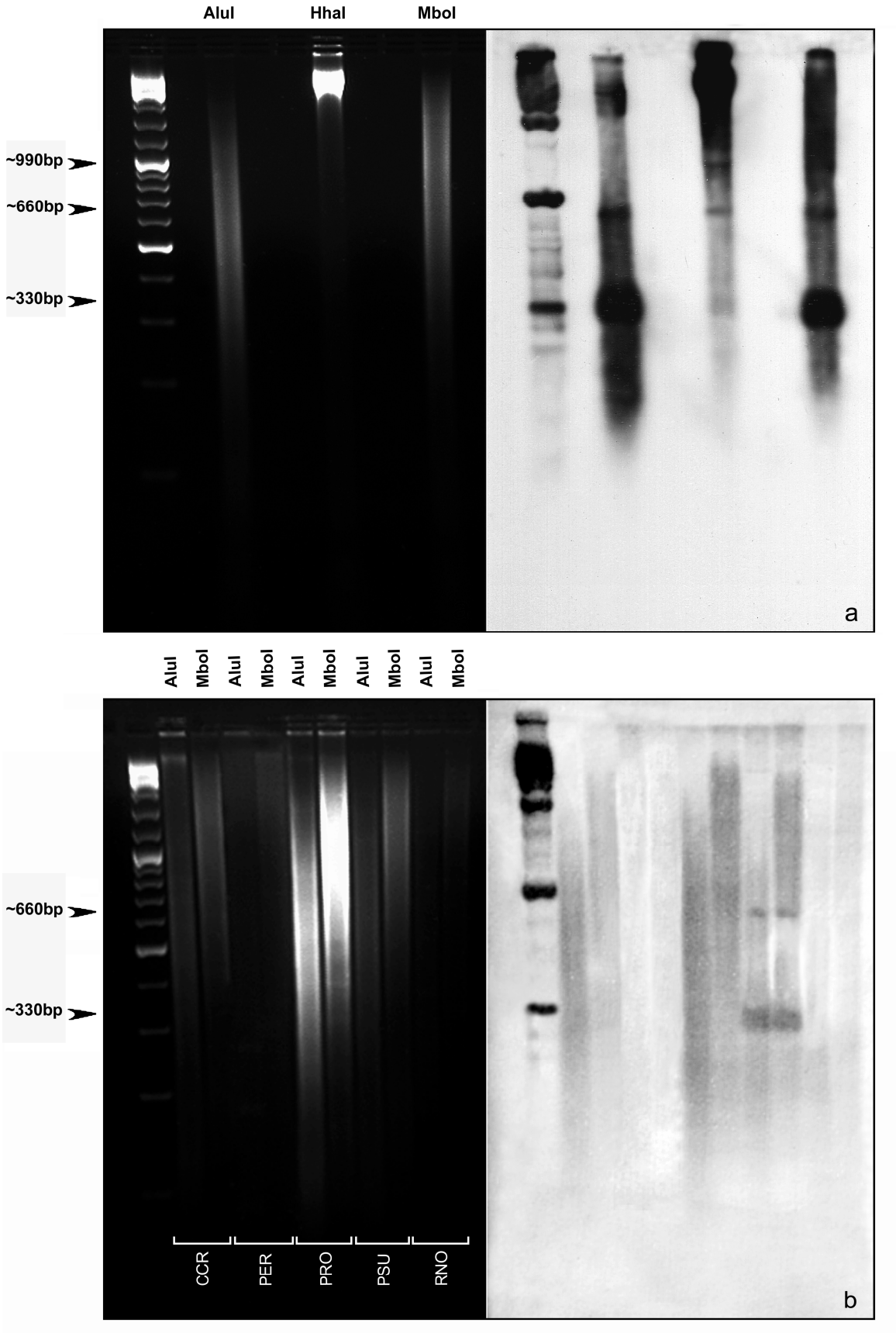

AluI
HhaI
MboI
~990bp
~660bp
~330bp
a
AluI
MboI
AluI
MboI
AluI
MboI
AluI
MboI
AluI
MboI
~660bp
~330bp
CCR
PER
PRO
PSU
RNO
b

Figure 3

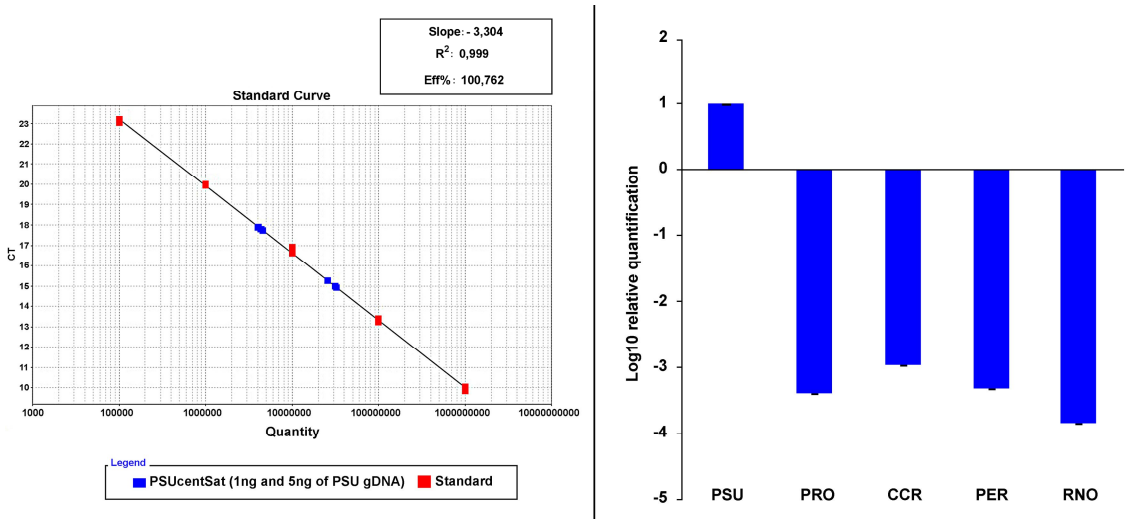


Figure 4

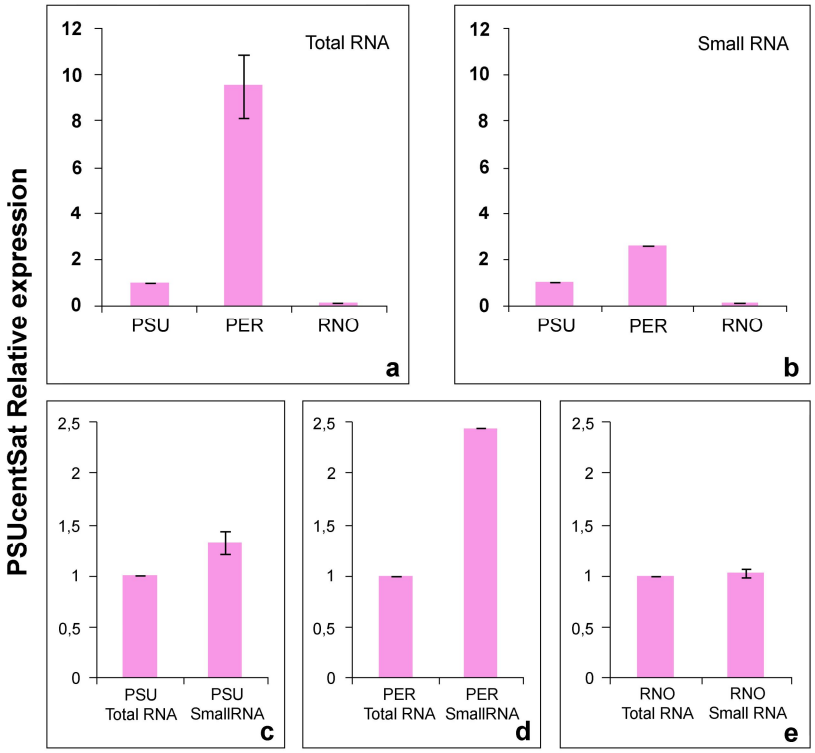


Figure 5

II.2.3

High-resolution organization of repetitive DNA sequences in *Phodopus roborovskii* and *Phodopus sungorus* genomes (Cricetidae, Rodentia)

High-resolution organization of repetitive DNA sequences in *Phodopus roborovskii* and *Phodopus sungorus* genomes (Cricetidae, Rodentia)

A. Paço[*], F. Adega[*], N. Meštrović[†], M. Plohl[†], R. Chaves[*,§,1]

[*]Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGBUTAD), Apdo 1013, 5001-801 Vila Real, Portugal

[§] Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

[†] Department of Molecular Genetics, Ruđer Bošković Institute, Bijenička 54, HR-10002 Zagreb, Croatia

Sequence data from this article have been deposited with the GenBank

Data Library under accession numbers. KJ649144, KJ649145, KJ649146, KJ649147.

Running title: Phodopus repetitive DNA evolution

Keywords: Rodentia, Phodopus, Repetitive Sequences, chromosomal evolution

[1] Corresponding author: Raquel Chaves, Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGB-UTAD), Vila Real, Portugal. Phone number: +351259350841. E-mail: rchaves@utad.pt

ABSTRACT

In this work we describe three novel repetitive DNA sequences presenting a similar heterochromatic chromosomal location in two hamster species: *Phodopus roborovskii* and *Phodopus sungorus*. Namely, two species-specific repeats (PROsat from *Phodopus roborovskii* and PSUchr1sat from *P. sungorus*) presenting each a neighboring chromosomal location to a third repetitive sequence shared by both hamster genomes (PsatDNA). Fiber-FISH analyses revealed that PROsat intermingles with PsatDNA in *P. roborovskii* and PSUchr1sat with PsatDNA in *P. sungorus*. This intricate repetitive sequences organization and chromosomal distribution allowed us to propose a model for the evolution of these sequences in the two genomes. A reconstruction of the chromosomal evolutionary events elapsed was possible, where the ancestral chromosomal forms were inferred. Evidences for an association between these repeats and the process of chromosome evolution were observed, mainly for PROsat, where it was verified that the evolutionary breakpoints that modulated the ancestral chromosomes occurred within and in boundaries of PROsat blocks. Besides, according to the sequences' similarities, it is also proposed that the first repetitions of satellite PROsat were originated from LINE-1 retrotransposons. The results obtained here increase the evidence about the possible origin of satellite DNAs from transposable elements. The high diversity of repeats at the heterochromatic regions of Phodopus chromosomes, together with its complex organization, lead us to propose these species as important models for evolutionary studies and the functional significance of repetitive sequences in eukaryotic genomes.

INTRODUCTION

A significant portion of eukaryotic genomes is comprised by repetitive DNA sequences that are mainly located at the heterochromatic regions of chromosomes. The more commonly repetitive sequences found within the heterochromatic genome blocks are the highly tandemly repeated sequences, satellite DNAs (satDNAs) (Enukashvily and Ponomartsev 2013). These can present several millions of copies in a genome, organized into long arrays (e.g. Adega *et al.* 2009, Plohl 2010), being preferentially located in or around centromeres, could however also assume an interstitial and/or terminal location (e.g. Santos *et al.* 2004, Louzada *et al.* 2008). Despite the abundance of satDNAs in heterochromatic regions, these regions are also frequently inhabited by transposable elements (Grewal and Jia 2007, Pezer and Ugarković 2008), presenting these elements a dispersed organization, whose copies are not clustered, but showing a widely distribution throughout the genomes (Slamovits and Rossi 2002, Wong and Choo 2004, Richard *et al*. 2008). Different families of repetitive sequences may exist in the same heterochromatic block, organized in a juxtapose form or intercalated (e.g. Mayorov *et al.* 1996, Chaves *et al.* 2005, Marchal *et al.* 2006). *In situ* hybridization techniques can be used to establish the organization of repetitive sequences in these dense heterochromatic blocks. Particularly Fiber-FISH experiments (*in situ* hybridization on stretched chromatin), allows determining the fine organization of highly repeated sequences located in the same region (e.g. Garagna *et al.* 2002, Kuznetsova *et al.* 2006).

Despite initially considered useless genomic elements (Ohno 1972), the repetitive sequences are presently seen as an important fraction of eukaryotic genomes to which several functions have been ascribed (e.g. Richard *et al.* 2008, Hall *et al.* 2012, Rebollo *et al.* 2012, Enukashvily and Ponomartsev 2013), thus justifying its study. The centromeric activity is one of the main roles accepted for (peri)centromeric satDNAs, which is associated with kinetochore formation, spindle microtubule attachment and sister chromatid cohesion (e.g.

Guenatri *et al.* 2004, Enukashvily and Ponomartsev 2013). Besides, it was also attributed to satDNAs an important responsibility in the origin of chromosomal rearrangements, which is mainly justified by its high molecular dynamics (reviewed by Adega *et al.* 2009). Studies in the last decades shed light on a previous "dark" area where transcription was not accepted as a trait of satDNAs. We are all now aware of the fact that these repeats are transcribed, having some of these non coding RNAs crucial roles in the maintenance of proper cell division and genome stability in the eukaryotic genome (Wong *et al.* 2007, Zhu and Pao *et al.* 2011, Hall *et al.* 2012). Regarding the transposable elements, these sequences are generally associated with the regulation of gene expression, mostly because they present a noticeable ability to produce mutations when integrating at new genomic sites (reviewed by Rebollo *et al.* 2012).

Nevertheless, and despite the progressive accumulation of data regarding the repetitive fraction of eukaryotic genomes, the great variety of repetitive classes and families, as well as its genomic abundance, blurs the exact functions that these sequences might play in the genomes. There also many unanswered questions concerning the origin and evolution mode of these sequences. The origin of the first repetitions from which satellites evolved is one of these questions (reviewed by Slamovits and Rossi 2002). Molecular mechanisms as unequal crossing-over and rolling circle replication/reinsertion are commonly proposed to explain the amplification and homogenization of satDNA monomers within genomes, in the process of concerted evolution (Hamilton et al. 1992, Ugarković and Plohl 2002, Plohl et al. 2008, Plohl 2010). However, it is not completely clear how the first tandem repetitions are originated, culminating in the millions of copies of a satDNA sequence (Slamovits and Rossi 2002). It is accepted that new satellites can evolve from the preexisting pool of satDNAs in the genome, or can in turn, be originated *de novo* from non satDNA sequences (Kapitonov *et al.* 1998). Theoretical models and computer simulations suggest that satellite units can be generated from a large variety of non satDNA and propagated into an array by unequal crossing-over

(Smith 1976, Slamovits and Rossi 2002). The best known example comes from several reports indicating a role of transposable elements in the origin/expansion of certain tandem repeats (Rossi *et al.* 1993, Batistoni *et al.* 1995, Heikkinen *et al.* 1995, Kapitonov *et al.* 1998, Kapitonov and Jurka 1999, Cheng and Murata 2003, López-Flores *et al.* 2004, Macas *et al.* 2009). In this way, the characterization of new repeats, particularly regarding its molecular features and chromosomal distribution, as performed in this work, will certainly contribute to understand the origin and evolution mode of these sequences, increasing our knowledge about its involvement in genome evolution.

Here we report the isolation and molecular characterization of three repetitive sequences from two Phodopus genomes: a species-specific repeat from *Phodopus roborovskii*, a species-specific repeat from *P. sungorus* and a third repetitive sequence common to both Phodopus. These species belong to the Cricetidae family, one of the most specious rodent families (Carleton and Musser 2005). The data assembled in this work allowed us to establish considerations about the origin and evolution of the studied repeats, being also possible to infer ancestral chromosomal forms and to reconstruct chromosomal rearrangements during these species evolution. A connection between these repeats and the process of chromosome evolution was observed, mainly for PROsat, since breakpoints for chromosomal rearrangements occurred within and in borders of PROsat blocks.

MATERIAL AND METHODS

Chromosome and nuclei preparations, released chromatin preparations and genomic DNA extraction

Fixed chromosome preparations and interphase nuclei from *Cricetus cricetus* (CCR), *Peromyscus eremicus* (PER), *Phodopus roborovskii* (PRO), *Phodopus sungorus* (PSU) and *Rattus norvegicus* (RNO) were obtained from fibroblast cell lines, which belong to 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 the five species was performed according to the method described by Chaves *et al*. (2004). Extended chromatin fibers were prepared from the same PRO and PSU cell cultures, as described in Verma and Babu (1995). Genomic DNA of the different species was obtained from fibroblast cell cultures using the Jequick DNA kit (Genomed), according to the manufacture instructions.

Isolation, cloning and sequencing of PROsat, PSUchr1sat and PsatDNA

PROsat and PsatDNA were isolated from PRO genome and PSUchr1sat from PSU genome. PRO genomic DNA was digested with the restriction endonucleases (REs) *Mbo*I or *Hin*fI (for the isolation of PROsat and PsatDNA, respectively) and the PSU genomic DNA with the endonuclease *Mbo*I (allowing the isolation of PSUchr1sat). These digestion reactions were performed by overnight incubation with 20 U of enzyme/µg of DNA, resulting in a smear that ranged between 3 kb to 100 bp. Subsequently, the restriction products were inserted into the plasmid vector pUC19 (Thermo scientific) and used to transform competent *Escherichia coli* DH5α cells (Invitrogen Life Technologies). Part of the obtained colonies were transferred onto a nylon membrane Hybond^{TM}-N^{+} (Amersham, GE Healthcare), following colony-lift hybridization procedures. For the isolation of PROsat, the DNA in the membrane was probed to *Eco*RV restriction prominent band (~2200 bp) obtained from PRO genome, labeled with digoxigenin-11-dUTP using the DIG DNA labeling Kit (Roche Molecular Biochemicals). For the isolation of PsatDNA and PSUchr1sat, the DNA in the membranes was probed to *Hin*fI or *Mbo*I restriction products (respectively), labeled as referred previously. Hybridization was performed at 68 º as described by Bruvo *et al.* (2003). Positive signals were visualized using the chemiluminescent CDP-Star system (Roche

Molecular Biochemicals). Plasmid DNA of the positive clones was isolated using the High Pure Plasmid Isolation kit (Roche Molecular Biochemicals) and sequenced in both directions using universal M13 primers.

Sequence analysis of PROsat, PSUchr1sat and PsatDNA

The isolated and sequenced clones of PROsat, PSUchr1sat and PsatDNA were analyzed with different sequence database tools and bioinformatic softwares: NCBI Blast (http://www.ncbi.nlm.gov/Blast/), RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker), Genomatix MatInspector (http://www.hearne.com.au/products/genomatix/edition/matinspector/, Cartharius *et al.* 2005), EMBOSS einverted (http://emboss.bioinformatics.nl/cgi-bin/emboss/einverted), Tandem repeats Finder (Benson 1999, version 4.00, free download in http://tandem.bu.edu/trf/trf.html) and vector NTI advance 11 (Invitrogen Life Technologies). A BLAST search for PROsat, PSUchr1sat and PsatDNA sequences against nucleotide sequences present in Genbank and Repbase was accomplished using NCBI blast and RepeatMasker tools. Sequence alignments were performed with the software vector NTI advance 11 that apply the Clustal W algorithm (Thompson *et al.* 1994) to verify sequence similarities. The search for direct or inverted repeats within PROsat, PSUchr1sat and PsatDNA was done using the Tandem Repeats Finder software and the EMBOSS einverted tool, respectively. EMBOSS einverted tool was used with a minimum score threshold of 35%. MatInspector tool allowed the analysis for the presence of transcription factor binding sites in these sequences. Sequence data from the three repeats isolated, PROsat, PsatDNA and PSUchr1sat, were deposited in the NCBI Nucleotide database, with the following accession numbers: KJ649145, KJ649146, KJ649147 (respectively).

Physical mapping of PROsat, PSUchr1sat and PsatDNA

Physical mapping of PROsat, PSUchr1sat and PsatDNA was carried out in metaphasic chromosomes, interphase nuclei and extended chromatin fibers of different rodent species (CCR, PER, PRO, PSU and RNO) by Fluorescent *in situ* Hybridization (FISH) procedures, as described by Schwarzacher and Heslop-Harrison (2000). The three sequences were labeled with digoxigenin-11-dUTP or biotin-16-dUTP (Roche, Molecular Biochemicals) by PCR. The most stringent post-hybridization wash was 50% formamide/2×SSC at 42 º. Biotin-labeled probes were detected by FITC conjugated with avidin (Roche Molecular Biochemicals), digoxigenin-labeled probes were detected with anti-digoxigenin-5'TAMRA (Roche Molecular Biochemicals).

CBP-banding sequential to Fluorescent in situ hybridization

After distaining the slides, sequential CBP-banding (C-bands by Barium Hydroxide with Propidium Iodide) was performed according to the standard procedure of Sumner (1972) with slight modifications. Shortly, the slides were submitted to hydrochloric acid (0.1 M) during 20 min, barium hydroxide (5% solution) during 7 min and 2× saline sodium citrate (2×SSC: 0.3 mol/L NaCl, 0.03 mol/L sodium citrate) at 60 º for 40 min.

Capture and preparation of images

Chromosomes, interphase nuclei and chromatin fibers were analyzed in a Zeiss Axioplan 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); contrast and color optimization were the functions used and affected the whole image equally. Chromosomes of PRO and PSU were identified according to Romanenko *et al.* (2007).

Southern hybridization analysis

Genomic DNA from the five rodent species (CCR, PER, PRO, PSU and RNO) was digested with the endonucleases *Eco*RV and *Eco*RI, for PROsat analysis. For the analysis of the other two sequences, the genomic DNA of these species was digested with *Alu*I and *Pvu*II (PSUchr1sat analysis) and *Eco*RV, *Hae*III and *Hin*fI (PsatDNA analysis). These enzymes cut only once or not cut the sequences under study. The resulting fragments were separated in a 0.8% agarose gel and blotted onto a Nylon membrane Hybond^{TM}-N^{+} (Amersham, GE Healthcare) by alkali capillary blotting, using 0.4 M NaOH. The membranes were then probed respectively with PROsat, PSUchr1sat and PsatDNA sequences, previously labeled by PCR with digoxigenin-11-d'UTP (Roche, Molecular Biochemicals). Hybridization was performed at 42º in hybridization solution (Roche Molecular Biochemicals). The positive signals were visualized using the chemiluminescent CDP-Star system (Roche Molecular Biochemicals). Selection of REs was done using the CLC sequence Viewer software (version 6.2, http://www.clcbio.com/index.php?id=28).

Inverted PCR analysis

Inverted PCR was performed using primers with a reverse orientation for the amplification of PROsat (PROsat1 5'-GTGTCGAGTATGGCAGACCATTTC-3' and PROsat2 5'-CATGTAATCAGCCCGACAGAGTTC-3'), allowing the determination of its flanking sequences, as described by Ochman *et al.* (1988). One of the strongest obtained bands, with approximately 470 bp, was inserted into the plasmid vector pUC19 (Thermo scientific) and used to transform competent *Escherichia coli* DH5α cells (Invitrogen Life Technologies). The inserts of the obtained clones were sequenced in both directions using universal M13 primers.

RESULTS

Molecular analysis of PROsat, PSUchr1sat and PsatDNA

In this work three novel repetitive DNA sequences were isolated and molecularly characterized. One of these sequences seems to be exclusive of *Phodopus roborovskii* (PROsat) genome, other specific of *P. sungorus* (PSUchr1sat) and the last one is common and shared by both genomes (PsatDNA). BLAST search revealed no significant similarity between these sequences and any other described or reported repetitive sequence; however, a small region of PROsat presents high similarity with part of a LINE-1 retrotransposon, within the ORF2 region (88% of similarity in an extension of 122 bp, Figure 1). In this sequence (PROsat), an inverted short repeat with 12 bp was also identified (100% similarity between inverted copies). A direct repeat of 46 bp (2.1 copies with 84% of similarity) was observed in PSUchr1sat, as can be seen in figure 1. No direct or inverted short internal repeats were identified within the PsatDNA. The analysis of transcription factor binding sites in the three sequences showed that PROsat presents two binding sites for a vertebrate TATA-binding factor (Figure 1, Table 1).

Genomic distribution of PROsat, PSUchr1sat and PsatDNA

PROsat, PSUchr1sat and PsatDNA were physical mapped by FISH in the chromosomes, interphase nuclei and chromatin fibers (Fiber-FISH) of *Phodopus roborovskii* and *P. sungorus*, which allow a high resolution analysis of their genomic distribution. These sequences were also hybridized onto the chromosomes of other rodent species, *Cricetus cricetus*, *Peromyscus eremicus* (Cricetidae) and *Rattus norvegicus* (Muridae), however no hybridization signals were detected (data not shown).

In *P. roborovskii* chromosomes, PROsat hybridizes at the (peri)centromeric region and at an interstitial band in the long arm of two autosomal pairs (PRO12 and PRO13), the short arm of the X-chromosome and the long arm of the Y-chromosome (Figure 2A). The simultaneous hybridization of PROsat and PsatDNA on *P. roborovskii* chromosomes revealed a similar location for these two sequences, except for the long arm of the PRO12 and PRO13, where it is only possible to observe PROsat signals (Figures 2B and 2C, arrowheads within the circles). An intricate pattern of PROsat and PsatDNA distribution is also detected in *P. roborovskii* interphase nuclei (Figures 3A – 3D). These sequences seem to be mainly colocalized (yellow signals, Figure 3C); nevertheless it is also possible to observe distinguishable signals at some nuclear sites (PROsat green signals and PsatDNA red signals, Figure 3C). C-banding performed sequentially to FISH evidenced a colocalization of these two sequences with constitutive heterochromatin (Figure 2D and 3D). In *P. sungorus,* PSUchr1sat hybridizes only at the (peri)centromeric region of chromosome 1 (Figure 2E). In this species, PSUchr1sat and PsatDNA also present a very similar chromosome distribution (Figures 2E– 2H), being both chromosome-specific. A similar location was also observed for these two sequences in interphase nuclei of *P. sungorus* (Figures 3E - 3H).

The clarification of this complex repetitive sequences distribution pattern in Phodopus genomes was further achieved by Fiber-FISH. The fine organization analysis of PROsat and PsatDNA performed on released chromatin of *P. roborovskii* showed the distinct location presented by these two sequences (Figure 4A). The same could be observed for PSUch1satDNA and PsatDNA on released chromatin of *P. sungorus* (Figure 4B). PROsat intermingles with PsatDNA in *P. roborovskii* genome and PSUchr1sat with PsatDNA in *P. sungorus* (Figures 4A and 4B). Chromatin regions which do not present hybridization signals were also observed within these heterochromatic blocks (evidenced by white arrowheads, Figure 4).

Genomic organization of PROsat, PSUch1sat and PsatDNA

In order to investigate the genomic organization of PROsat, PSUchr1sat and PsatDNA on both Phodopus genomes studied, southern blot analyses was performed (Figure 5). Nevertheless, the hybridization pattern obtained for the three sequences did not allowed the confirmation of its tandem organization, characteristic of satDNAs. The southern hybridization results obtained for PROsat in *P. roborovskii* using *Eco*RV (Figure 5A), show a pattern of bands with a periodicity of ~ 1250 bp, intercalated with other bands of varying lengths (Figure 5A), what can be justified by the sequence variability of PROsat repeat units in *P. roborovskii* genome. Differently, the southern hybridization pattern obtained for this sequence using *Eco*RI does not exhibit a defined band length periodicity (Figure 5A). Inverted PCR results, using primers to amplify PROsat flanking regions, point to a tandem organization of this sequence, corresponding to a satDNA with a monomer unit of 1587 bp (available in NCBI Nucleotide, accession number KJ649144). This PCR analysis, as the southern blot results, shows that PROsat presents several monomer variants, since a variety of bands were obtained (Figure 6).

The southern blot results of PSUchr1sat sequence in *P. sungorus* genome show a ladder hybridization pattern with a periodicity of ~ 50 bp (Figure 5B), for both enzymes used (*Alu*I and *Pvu*II). However, there were only observed two bands, which are not sufficient to confirm the tandem organization of the sequence. Cross-hybridization of both sequences indicates the presence of PROsat only in *P. roborovskii* and PSUchr1sat only in *P. sungorus* genomes, at least according with this technique resolution (Figure 5B, showing the absence of PSUchr1sat southern signals in *P. roborovskii*). Regarding PsatDNA, our southern results indicate the presence of this sequence in both studied Phodopus genomes. The hybridization results obtained for PsatDNA in *P. roborovskii*, using *Hin*fI enzyme, show a pattern of bands with a periodicity of ~150 bp (Figure 5C). Nevertheless, we only identified multiples up to 4 times

the size of 150 bp, which is not satisfactory to validate the tandem organization of this sequence. Other bands not included in this ladder pattern (with variable lengths), were also observed. The restriction enzymes *Eco*RV and *Hae*III did not produce a defined band length periodicity (Figure 5C). At the same time, the hybridization pattern obtained with *Hae*III for *P. sungorus* is comparable to the pattern obtained for *P. roborovskii*. In this case, the enzyme *Hae*III produced two bands with ~210 and 520 bp. The action of *Hinf*I in *P. sungorus* resulted in very weak bands, presenting the smallest one ~200 bp. PsatDNA in this genome does not seem to have recognition sites for *Eco*RV (Figure 5C).

According to the southern analyses resolution (data not shown), PROsat, PSUchr1sat and PsatDNA seem to be absent on *Cricetus cricetus, Peromyscus eremicus* (Cricetidae) and *Rattus norvegicus* (Muridae) genomes.

DISCUSSION

Molecular features and genomic organization of PROsat, PSUchr1sat and PsatDNA

In this work we report the isolation and molecular characterization of three novel repetitive sequences from two Phodopus genomes, a repeat from *Phodopus roborovskii* (PROsat), other from *P. sungorus* (PSUchr1sat) and a repeat shared by both hamster genomes (PsatDNA). BLAST analyses revealed no significant similarity between these sequences and any other repetitive sequence already described. Nevertheless, a region of PROsat presents a high similarity with a fraction of a LINE-1 retrotransposon, within the ORF2 region (88% of similarity in an extension of 122 bp, Figure 1). There are growing evidences that transposable elements can be involved in the origin and evolution of satellite DNAs, based on sequence similarities found between these two classes of sequences (e.g. Rossi *et al.* 1993, Batistoni *et al.* 1995, Heikkinen *et al.* 1995, Kapitonov *et al.* 1998, Kapitonov and Jurka 1999, Cheng and Murata 2003, López-Flores *et al.* 2004, Macas *et al.* 2009). Kapitonov *et al.* (1998)

reported, in cetaceans, a satDNA displaying similarity to part of the 3'UTR and ORF2 regions from LINE-1 retrotransposons, suggesting that the entire unit of this satDNA was derived from a LINE-1 fragment which underwent extensive internal deletions and other mutational events. In accordance, we can also suppose that PROsat originated from a LINE-1 retrotransposon, conserving a part of the ORF2 region in their repeat unit sequences. Non-homologous recombinational mechanisms (e.g. unequal crossing-over), involving homologous retrotransposons, can be responsible for the initial tandem duplications in the satDNAs culminating in arrays of repeats. Subsequent mutational changes and sequence homogenization (concerted evolution) events can justify the fact that only a small region is conserved between these satDNAs and transposable elements (as proposed by Wong and Choo (2004)).

Many satellite DNA families in both plants and animals present repetitions of 150-180 bp and 300-360 bp, which seems to be associated with requirements in the DNA length for wrapping around one or two nucleosomes (Schmidt and Heslop-Harrison 1998, Henikoff *et al.* 2001). However, in the literature we can also find descriptions of satDNAs exhibiting larger repeat units. The largest satDNA repeat unit reported in mammals presents ~2570 bp (Modi 1993). Our data showed as well that PROsat presents a large monomer unit with 1587 bp, nevertheless some degree of repeat unit variability was also observed. Furthermore, the transcription factors binding sites analysis performed in this work suggest a transcriptional activity of this satellite, as vertebrate TATA-binding factor sites were found in PROsat (Figure 1). The hypothesis of transcriptional activity gains strength due to the fact that the TATA box corresponds to an element presented in the core promoter of most protein-coding genes (Persengiev *et al.* 2003). The detection of short inverted repeat motifs in PROsat (Figure 1) could have an important role in the accessibility of the transcription factors to this sequence, since these short motifs have been associated with the conformation of chromatin

(Modi 1993, Plohl 2010). Future studies about the transcription profile of PROsat are mandatory to better understand how the transcription of this satellite is regulated.

Physical mapping and evolution model

The physical mapping of PROsat and PSUchr1sat on metaphasic chromosomes of five Cricetidae and Muridae species (only positive results are shown, Figures 2A and 2E), suggests the presence of PROsat only in *P. roborovskii* genome and the PSUchr1sat only in *P. sungorus*, considering at least the FISH technique resolution. This was also supported by southern blot analyses (Figure 5B, showing the absence of PSUchr1sat southern signals in *P. roborovskii*). PsatDNA is found in both Phodopus species (Figures 2B and 2F). Interestingly, FISH analyses also show that the location of PROsat is very similar to the location of PsatDNA in *P. roborovskii*, happening the same with PSUch1sat and PsatDNA in *P. sungorus* (Figures 2C and 2G, respectively). Despite this intricate pattern of distribution in each Phodopus genome, the analyses of the location of these repeats in interphase nuclei (Figure 3) and particularly on released chromatin (Figure 4), showed different locations for these sequences. PROsat intermingles with PsatDNA in *P. roborovskii* genome and PSUchr1sat intermingles with PsatDNA in *P. sungorus* (Figure 4). The chromatin regions within the analyzed heterochromatic blocks that do not present hybridization signals (evidenced by white arrowheads, Figure 4) suggest the presence of other repetitive sequences, which increase the complexity of the repetitive sequences organization in these genomes.

The chromosomal distribution of the three repeats in both Phodopus genomes allowed us to propose a model for its evolution, and also to establish evolutionary considerations on the chromosomes inhabited by these repeats (Figure 7). PROsat and PsatDNA sequences were most probably located in a single ancestral Phodopus chromosome (APK14, considering the ancestral karyotype presented by Paço *et al.* 2012), which in turn originated PRO12 in *P.*

roborovskii and a part of PSU1 in *P. sungorus*, including the (peri)centromeric region (according to Romanenko *et al.* 2007 data, Figure 7). Later, in each Phodopus genome, these repeats experienced different pathways. In *P. roborovskii* a pericentric inversion dividing the PROsat block seems to have occurred, originating a smaller PROsat block that occupies the pericentromeric/short arm proximal region and a second block in the long arm of PRO12 (Figure 7). PsatDNA repeat seem to have been maintained at the (peri)centromere in PRO12, being afterwards expanded to PRO12 short arm, allowing some degree of intermingling with PROsat (Figure 7). PsatDNA amplification could have occurred through recombinational mechanisms within PRO12 short arm, as gene conversion or unequal crossing-over (Walsh 1987, Elder and Turner 1995, Dover 2002). As the PROsat presents similarity with LINE-1 retrotransposons, non-homologous recombination events could explain the spread of PROsat to the heterochromatic region of the sex-chromosomes and to PRO13, both rich in LINE-1 sequences (unpublished data). PsatDNA probably also assumed these locations by association with PROsat, caused by the intermingling between these two sequences. Considering Romanenko *et al.* (2007) data, PSU1 results from the fusion of three ancestral Phodopus chromosomes that present homology to PRO1 (short arm), PRO12 and PRO4 (Figure 7). In accordance, we propose that by the occurrence of one of these fusion events, the PROsat block present in the ancestral chromosome was eliminated, thus justifying the absence of this sequence in *P. sungorus* (Figure 7). PsatDNA was maintained in *P. sungorus* and was afterwards amplified, resulting in the large PsatDNA block observed in PSU1 (Figure 7). The presence of PSUchr1sat only in *P. sungorus* genome and also its chromosome-specific location may suggest a recent origin for this repeat, certainly after the divergence of *P. roborovskii* and *P. sungorus* species. The subsequent expansion of this repeat within the PsatDNA block can explain the intermingling organization of these two sequences.

A previous work (Paço *et al.* 2012), also focused in the evolution of the two Phodopus species chromosomes, describe a high CH molecular heterogeneity in both Phodopus species, which suggest the presence of a higher number of repeats in these genomes, as well as a high molecular dynamics of the Phodopus repetitive fraction. In this regard, we believe in a recurrent origin of new repeats mostly from other preexisting repeats in these genomes, continuing the older sequences to be maintained, which results in an accumulation of a large variety of repeats during Phodopus evolution. This is supported by the results obtained here, where the described repeats are only restricted to a few chromosomes pairs (even chromosome-specific), with an intermingling organization. The reason for this growing diversity of repetitive sequences during Phodopus evolution is not yet understood, but indicates an important functionality of this high heterogeneous genomic fraction. In accordance, we propose these species as important models to study the evolutionary mode and the functional significance of repetitive sequences in the eukaryotic genome.

ACKNOWLEDGEMENTS

This work was supported by the project POCI/BIA-BCM/58541/2004, a PhD grant (SFRH/BD/41574/2007) 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. We are deeply grateful to Dr. Vitaly Volobouev for providing the Rodentia species cell cultures.

LITERATURE CITED

Adega, F., H. Guedes-Pinto, R. Chaves, 2009 Satellite DNA in the Karyotype Evolution of Domestic Animals – Clinical Considerations. Cytogenet Genome Res 126: 12–20.

Batistoni, R., G. Pesole, S. Marracci, I. Nardi, 1995 A Tandemly Repeated DNA Family Originated from SINE-Related Elements in the European Plethodontid Salamanders (Amphibia, Urodela). J Mol Evol 40: 608-615.

Benson, G., 1999 “Tandem repeats finder: a program to analyze DNA sequences". Nucleic Acids Res 27 (2): 573-580.

Bruvo, B., J. Pons, Đ. Ugarković, C. Juan, E. Petitpierre *et al*., 2003 Evolution of low-copy number and major satellite DNA sequences coexisting in two Pimelia species-groups (Coleoptera). Gene 312: 85–94.

Carleton, M. D., and G. G Musser, 2005 Order Rodentia, pp 745-1601 in *Mammal Species of the World: A Taxonomic and Geographic Reference*, edited by D. E. Wilson and D. M. Reeder. Johns Hopkins University Press, Baltimore.

Cartharius, K., K. Frech, K. Grote, B. Klocke, M. Haltmeier *et al*., 2005 MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21: 2933-2942.

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

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

Cheng, Z. J., and M. Murata, 2003 A centromeric tandem repeat family originating from a part of Ty3/gypsy –retroelement in wheat and its relatives. Genetics 164: 665-672.

Dover, G., 2002 Molecular drive. Trends Genet 18: 587–589.

Elder, J. F., B. J. Turner, 1995 Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev Biol 70:297-320.

Enukashvily, N. I., and N. V. Ponomartsev, 2013 Mammalian Satellite DNA: A Speaking Dumb. Adv Protein Chem Struct Biol 90: 31-65.

Garagna, S., M. Zuccotti, E. Capanna, C. A. Redi, 2002 High resolution organization of mouse telomeric and pericentromeric DNA. Cytogenet Genome Res 96: 125–129.

Grewal, S. I. S., and S. Jia, 2007 Heterochromatin revisited. Nature Rev 8: 35-46.

Guenatri, M., D. Bailly, C. Maison, G. Almouzni, 2004 Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J cell Biol 166 (4): 493-504.

Hall, L. E., S. E. Mitchell, R. J. O'Neill, 2012 Pericentric and centromeric transcription: a perfect balance required. Chromosome Res (2012) 20:535–546.

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

Heikkinen, E., V. Launonen, E. Muller, L. Bachmann, 1995 The pvB370 BamHI satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic pDv elements. J Mol Evol 41: 604-614.

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

Kapitonov, V. V., G. P. Holmquist, J. Jurka, 1998 L1 Repeat Is a Basic Unit of Heterochromatin Satellites in Cetaceans. Mol Biol Evol 15(5): 611–612.

Kapitonov, V. V., and J. Jurka, 1999 Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica 107: 27-37.

Kuznetsova, I., O. Podgornaya, M. A. Ferguson-Smith, 2006 High-resolution organization of mouse centromeric and pericentromeric DNA. Cytogenet Genome Res 112: 248–255.

López-Flores, I., R. de la Herrán, M. A. Garrido-Ramos, P. Boudry, C. Ruiz-Rejón *et al*., 2004 The molecular phylogeny of oysters based on a satellite DNA related to transposons. Gene 339: 181–188.

Louzada, S., A. Paço, S. Kubickova, F. Adega, H. Guedes-Pinto *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.

Macas, J., A. Koblížková, A. Navrátilová, P. Neumann, 2009 Hypervariable 3′ UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. Gene 448: 198–206.

Marchal, J. A., M. J. Acosta, M. Bullejos, E. Puerma, R. Díaz de la Guardia R, A. Sánchez, 2006 Distribution of L1-retroposons on giant sex chromosomes of *Microtus cabrerae* (Arvicolidae, Rodentia): functional and evolutionary implications. Chromosome Res 14: 177–186.

Mayorov, V. I., L. R. Adkison, N. V. Vorobyeva, E. A. Khrapov, N. G. Kholodhov, I. B. Rogozin, T. B. Nesterova, A. I. Protopopov, O. V. Sablina, A. S. Graphodatsky, S. M. Zakian, 1996 Organization and chromosomal localization of a B1-like containing repeat of *Microtus subarvalis*. Mamm Genome 7: 593-597.

Modi, W. S., 1993 Rapid, localized amplification of a unique satellite DNA family in the rodent *Microtus chrotorrhinus*. Chromosoma 102: 484-490.

Ochman, H., A. S. Gerber, D. L. Hart, 1988 Genetic Applications of an Inverse Polymerase Chain Reaction. Genetics 120: 621-623.

Ohno, S., 1972 So much “junk” DNA in our genome. In: Evolution of genetic systems. Brookhaven Symp Biol 23: 366-370.

Paço, A., R. Chaves, A. Vieira-da-Silva, F. Adega, 2012 The involvement of repetitive sequences in the remodelling of karyotypes: The Phodopus genomes (Rodentia, Cricetidae). Micron 46: 27-34.

Persengiev, S. P., X. Zhu, B. L. Dixit, G. A. Maston, E. L. W. Kittler *et al*., 2003 TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. Proc Natl Acad Sci USA 100(25): 14887–14891.

Pezer, Ž., and Đ. Ugarković (2008). Role of non-coding RNA and heterochromatin in aneuploidy and cancer. Sem Cancer Biol 18: 123-130.

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

Plohl, M., 2010 Those mysterious sequences of satellite DNAs. Period Biol 112 (4): 403–410.

Rebollo, R., M. T. Romanish, D. L. Mager, 2012 Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. Annual Review of Genetics 46: 21-42.

Richard, G. F., A. Kerrest, B. Dujon, 2008 Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes. Microbiol Mol Biol R: 686-727.

Romanenko, S. A., V. T. Volobouev, P. L. Perelman, V. S. Lebedev, N. A. Serdukova *et al*., 2007 Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroideia, Rodentia) inferred from chromosomal painting and banding comparison. Chromosome Res 15: 283-297.

Rossi, M. S., C. G. Pesce, O. A. Reig, A. R. Kornblihtt, J. Zorzópulos, 1993 Retroviral-like features in the repetitive unit of the major satellite DNA from the South American rodents of the genus *Ctenomys*. DNA Seq 3: 379–381.

Santos, S., R. Chaves, H. Guedes-Pinto, 2004 Chromosomal localization of the major satellite DNA family (FA-SAT) in the domestic cat. Cytogenet Genome Res 107:119–122.

Schmidt, T., and J. S. Heslop-Harrison, 1998 Genomes, genes and junk: the large scale organization of plant chromosomes. Trends Plant Sci 3: 195–199.

Schwarzacher, T., and J. S. Heslop-Harrison, 2000 Practical *in situ* hybridization. Bios Scientific Publishers, Oxford.

Slamovits, C. H., and M. S. Rossi, 2002 Satellite DNA: agent of chromosomal evolution in mammals. A review. Mastozoologia Neotropical/J. Neotrop. Mammal 9 (2):297–308.

Smith, G. P., 1976 Evolution of repeated DNA sequences by unequal crossover. Science 191: 528–535.

Sumner, A. T., 1972 A simple technique for demonstrating centromeric heterochromatin. Exp Cell Res 75:304-306.

Thompson, J. D., D. G. Higgins, T. J. Gibson, 1994 Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22): 4673-4680.

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

Verma, R. S., and A. Babu, 1995 *Human chromosomes – Principles and Tecniques.* 2nd edition. McGraw-Hills Book Company, New York.

Walsh, J. B., 1987 Persistence of tandem arrays: implications for satellite DNA and simple-sequence DNAs. Genetics 115: 553-567.

Wong, L. H., and K. H. A. Choo, 2004 Evolutionary dynamics of transposable elements at the centromere. Trends Genet 20(12): 611-616.

Wong, L. H., K. H. Brettingham-Moore, L. Chan, J. M. Quach, M. A. Anderson *et al.*, 2007 Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res 17(8): 1146–1160.

Zhu, Q., G. M. Pao, A. M. Huynh, H. Suh, N. Tonnu *et al.*, 2011 BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature 477(7363): 179–184.

TABLE 1

Summary of the analysis for transcription factor binding sites in PROsat

Satellite	Transcription factor binding site	Position	Sequence
PROsat	Vertebrate TATA-binding factor	1080 to 1096 bp	atgttTAAAtcctcaaa
		1110 to 1126 bp	ccttattTAATcgcagg

[a]Core sequence of transcription factor binding sites in capitals.

FIGURES LEGENDS

Figure 1.- PROsat, PSUchr1sat and PsatDNA sequences internal characteristics. Schematic representation of the PROsat, PSUchr1sat and PsatDNA molecular features.

Figure 2.- Physical mapping of PROsat, PsatDNA and PSUchr1sat on chromosomes of *Phodopus roborovskii* and *Phodopus sungorus*. Representative *in situ* hybridization (FISH) presenting the chromosomal localization of PROsat (A) and PsatDNA (B) on *P. roborovskii* chromosomes. Overlapping of PROsat and PsatDNA hybridization signals (C). Same metaphase after sequential C-banding (D). Representative FISH presenting the chromosomal localization of PSUchr1sat (E) and PsatDNA (F) on *P. sungorus* chromosomes. Overlapping of PSUchr1sat and PsatDNA hybridization signals (G). Same metaphase after sequential C-banding (H). The satellite sequences were labeled with biotin-16-dUTP detected by FITC conjugated with avidin (green) or digoxigenin-11-dUTP detected with 5'TAMRA (red).

Chromosomes were counterstained with DAPI (blue) in FISH experiments or Propidium Iodide (red) in C-banding.

Figure 3.- Physical mapping of the PROsat, PsatDNA and PSUchr1sat on interphase nuclei of *Phodopus roborovskii* and *Phodopus sungorus*. Representative *in situ* hybridization (FISH) presenting the localization of PROsat (A) and PsatDNA (B) on interphase nuclei of *P. roborovskii*. Overlapping of PROsat and PsatDNA hybridization signals (C). Same interphase nuclei after sequential C-banding (D). Representative FISH presenting the chromosomal localization of PSUchr1sat (E) and PsatDNA (F) on interphase nuclei of *P. sungorus.* (F) Overlapping of PSUchr1sat and PsatDNA hybridization signals (G). Same interphase nuclei after sequential C-banding (H). The satellite sequences were labeled with biotin-16-dUTP detected by FITC conjugated with avidin (green) or digoxigenin-11-dUTP detected with 5'TAMRA (red). Chromatin was counterstained with DAPI (blue) in FISH experiments or Propidium Iodide (red) in C-banding.

Figure 4.- Representative dual color *in situ* hybridization of PROsat, PSUchr1sat and PsatDNA on extended chromatin fibers from *Phodopus roborovskii* and *P. sungorus*. (A) Dual color FISH with PROsat (green signals) and PsatDNA satellite (red signals) on released chromatin of *P. roborovskii*. DAPI was used as counterstaining (blue). (B) Dual color FISH with PSUchr1sat (green signals) and PsatDNA satellite (red signals) on released chromatin of *P. sungorus.* DAPI was used as counterstaining (blue). White arrowheads point to chromatin regions without satellite signals.

Figure 5.- Restriction enzyme digestions and Southern blot hybridizations. (A) Electrophoresis separation of *Phodopus roborovskii* (PRO) genomic DNA after digestion with

*Eco*RV and *Eco*RI (shown on the left). The corresponding southern blot obtained after hybridization with PROsat is shown on the right. (B) Electrophoresis separation of *P. roborovskii* and *P. sungorus* (PSU) genomic DNA after digestion with *Alu*I and *Pvu*II (shown on the left). The corresponding southern blot obtained after hybridization with PSUchr1sat is shown on the right. (C) Electrophoresis separation of *Phodopus roborovskii* and *P. sungorus* genomic DNA after digestion with *Eco*RV, *Hae*III and *Hin*fI (shown on the left). The corresponding southern blot obtained after hybridization with PsatDNA is shown on the right. The black arrows point to southern blot resultant bands that present evident multiple lengths (periodicity of ~1250bp, ~50bp or ~150bp). The ladder use to evaluate the bands length corresponds to O'GeneRuler DNA Ladder Mix (Fermentas).

Figure 6.- Inverted PCR analysis of PROsat repeat. Pink bars represent the known PROsat sequence fragments amplified by PROsat1 and PROsat2 primers. Amplification sense of primers is shown by black arrows. The green bar represents the PROsat flanking regions amplified by PROsat1 and PROsat2 primers. The PROsat sequence plus its flanking sequences compose the PROsat satellite complete monomer.

Figure 7.- Hypothetical model explaining the evolution of PROsat, PsatDNA and PSUchr1sat in the *Phodopus roborovskii* and *P. sungorus* genomes. In this figure are schematized the most probable evolutionary pathways that occurred in the Phodopus chromosomes explaining the observed physical distribution of the PROsat, PsatDNA and PSUchr1sat repeats. The chromosomes of the ancestral Phodopus karyotype (APK) are identified according to Paço *et al.* (2012, see supplementary data). In the ancestral chromosome APK14 is represented the most probable centromere position (constriction).

Green blocks correspond to the PROsat, pink blocks to PsatDNA and blue blocks to PSUchr1sat.

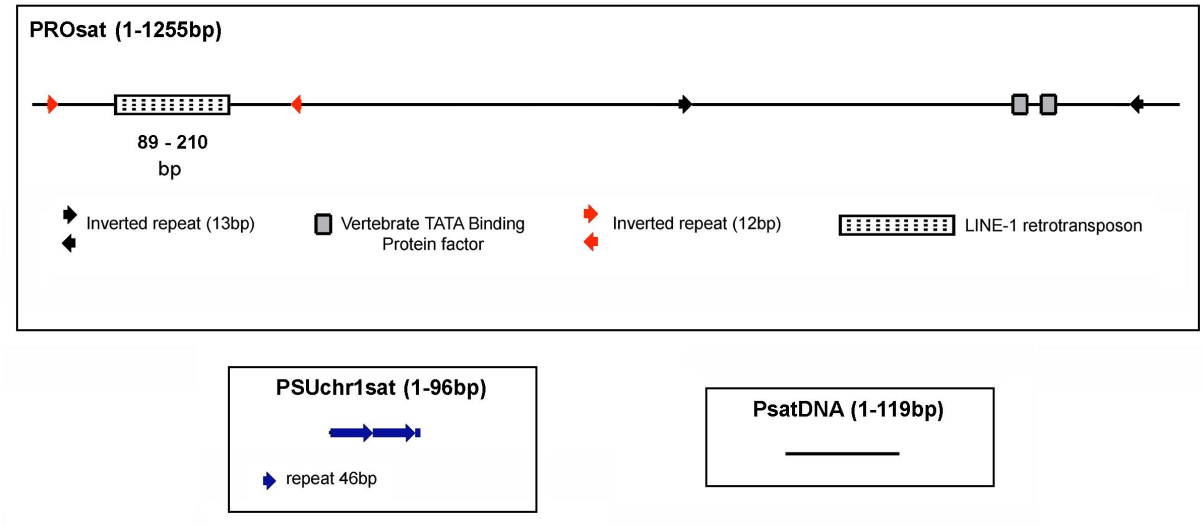


Figure 1

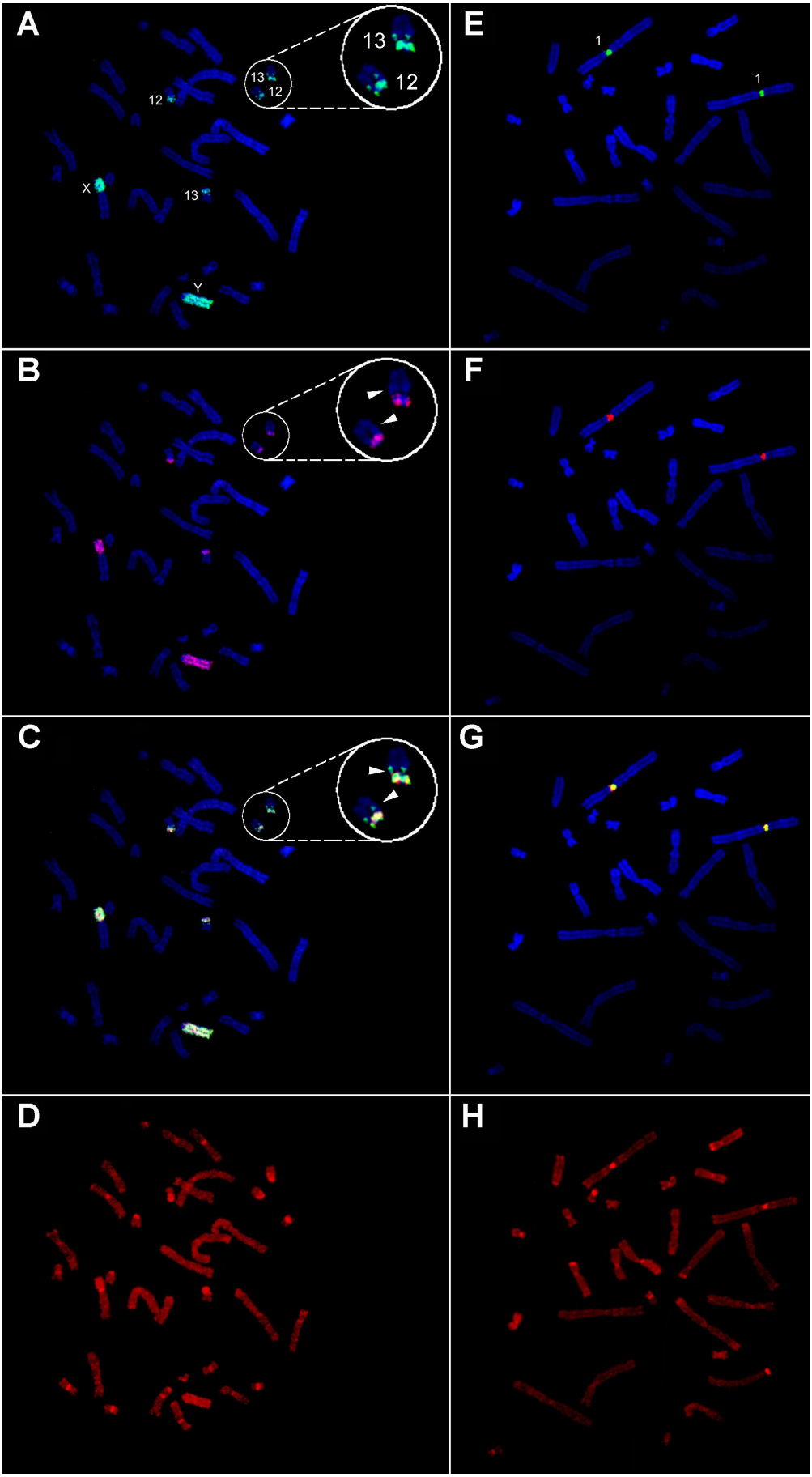


Figure 2

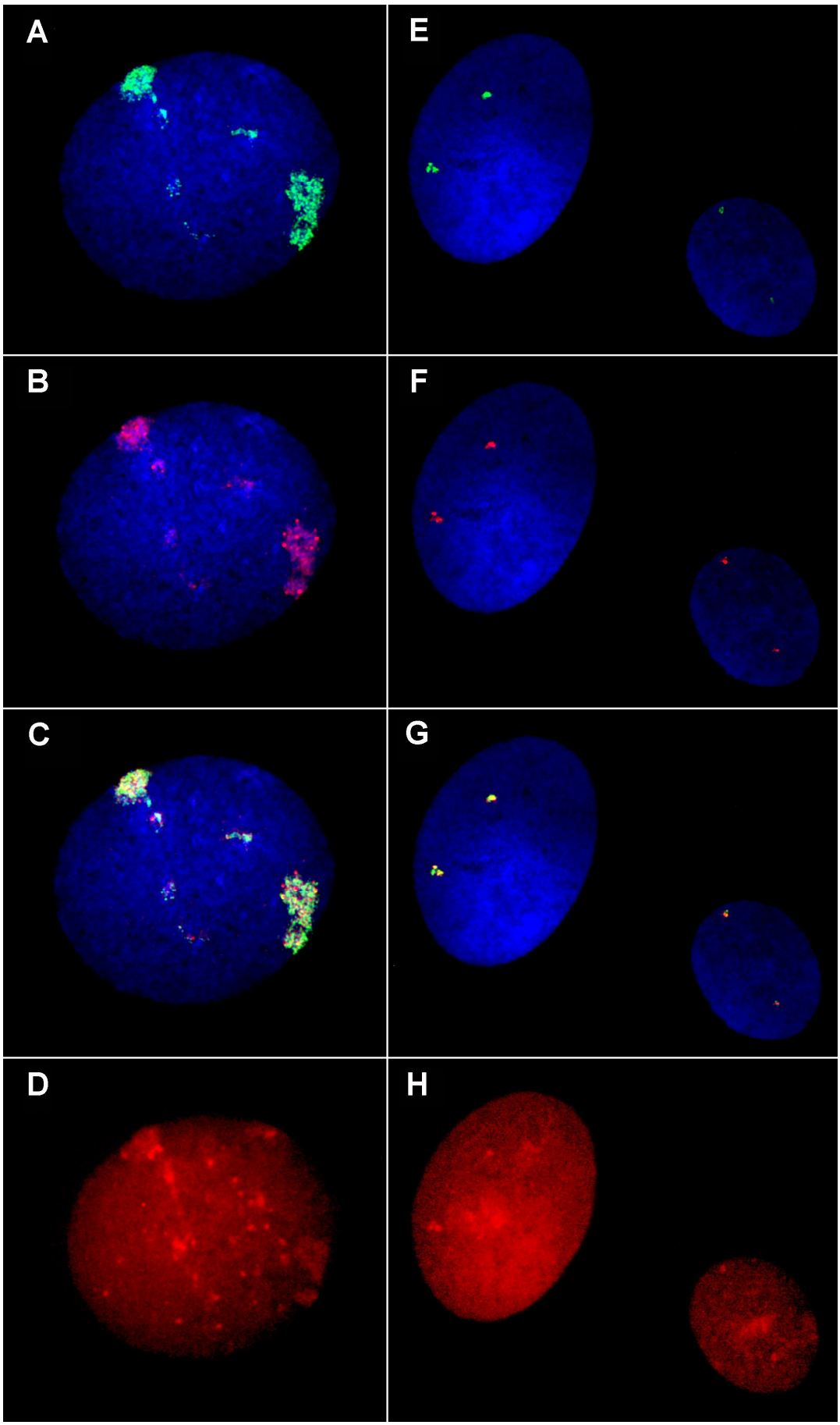


Figure 3

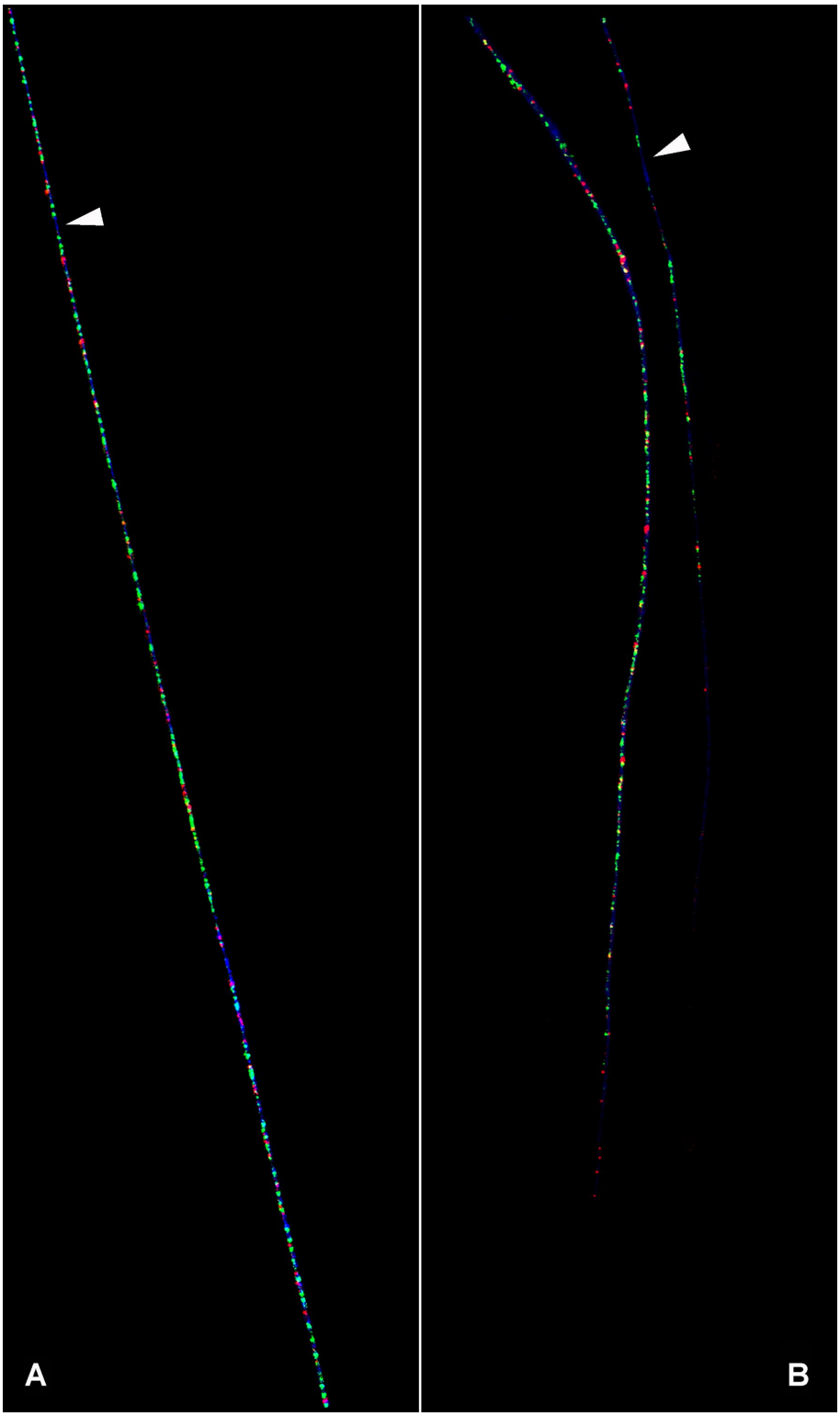


Figure 4

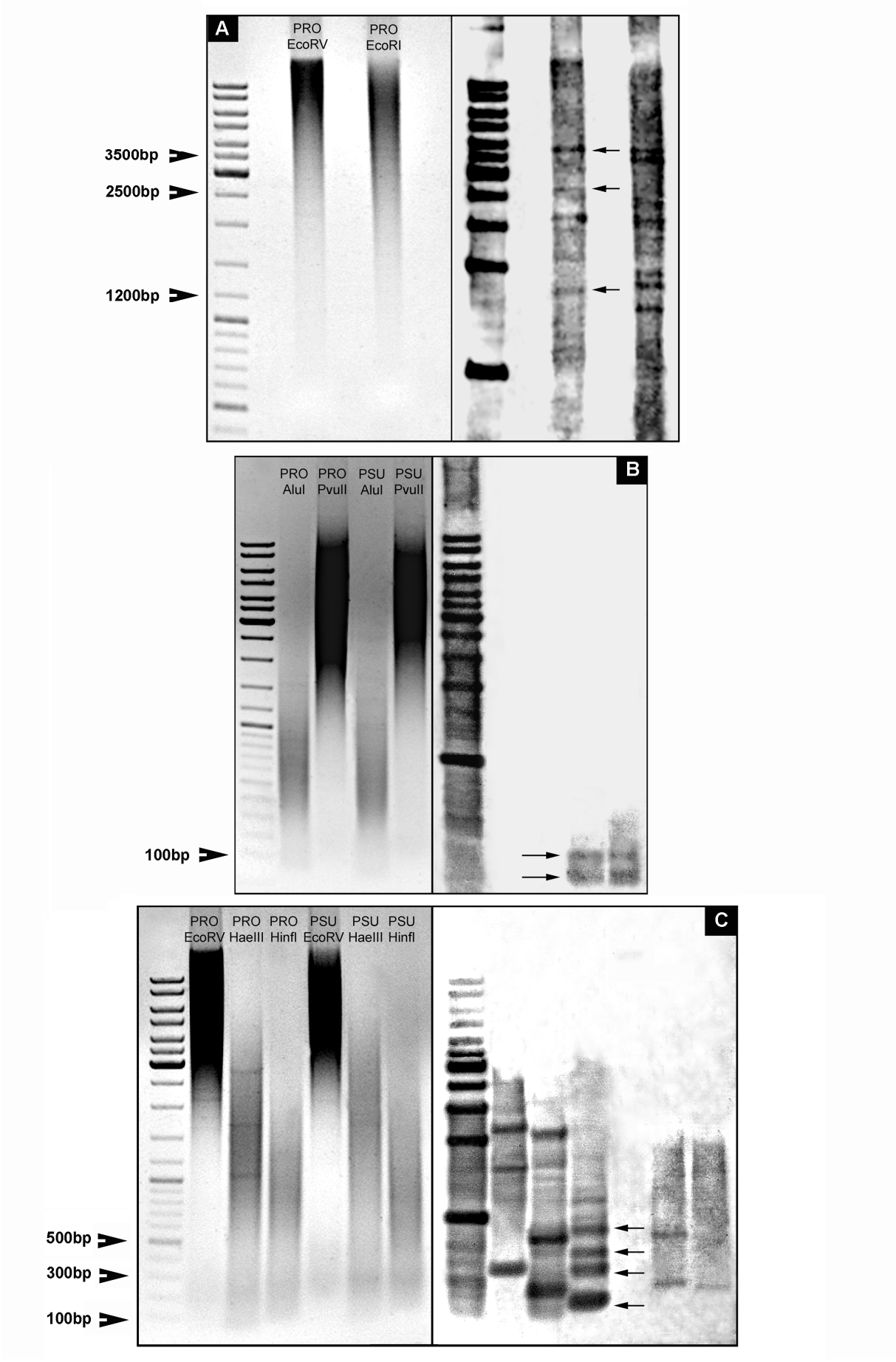

A
PRO EcoRV
PRO EcoRI
3500bp
2500bp
1200bp
B
PRO AluI
PRO PvuII
PSU AluI
PSU PvuII
100bp
C
PRO EcoRV
PRO HaeIII
PRO HinfI
PSU EcoRV
PSU HaeIII
PSU HinfI
500bp
300bp
100bp

Figure 5

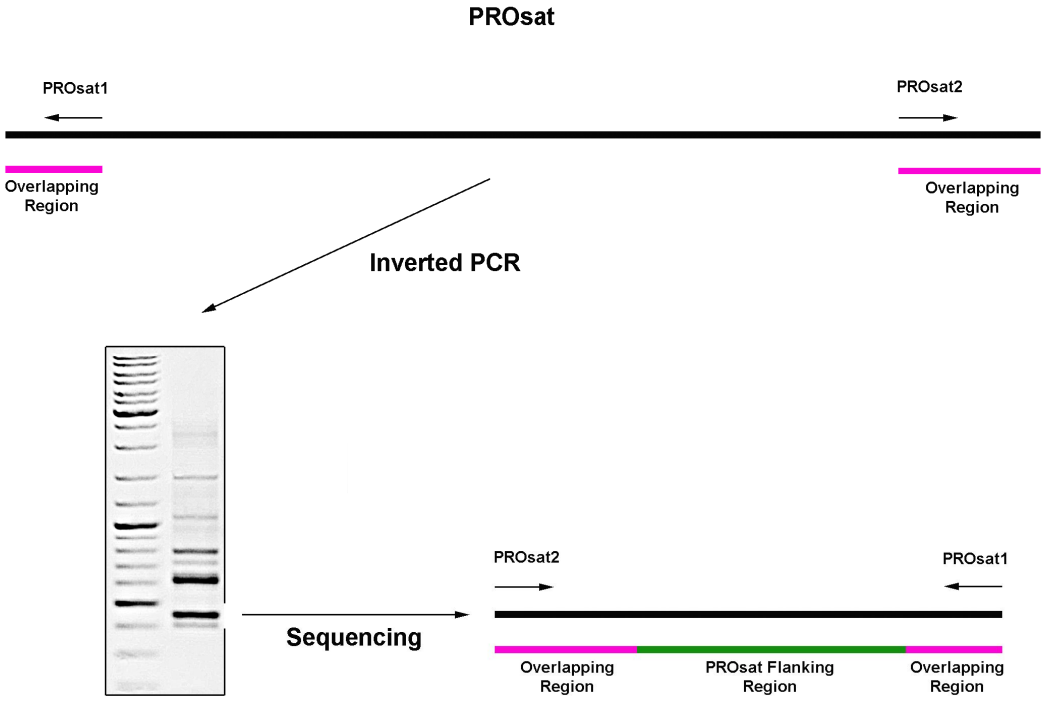

PROsat
PROsat1
PROsat2
Overlapping Region
Overlapping Region
Inverted PCR
PROsat2
PROsat1
Sequencing
Overlapping Region
PROsat Flanking Region
Overlapping Region

Figure 6

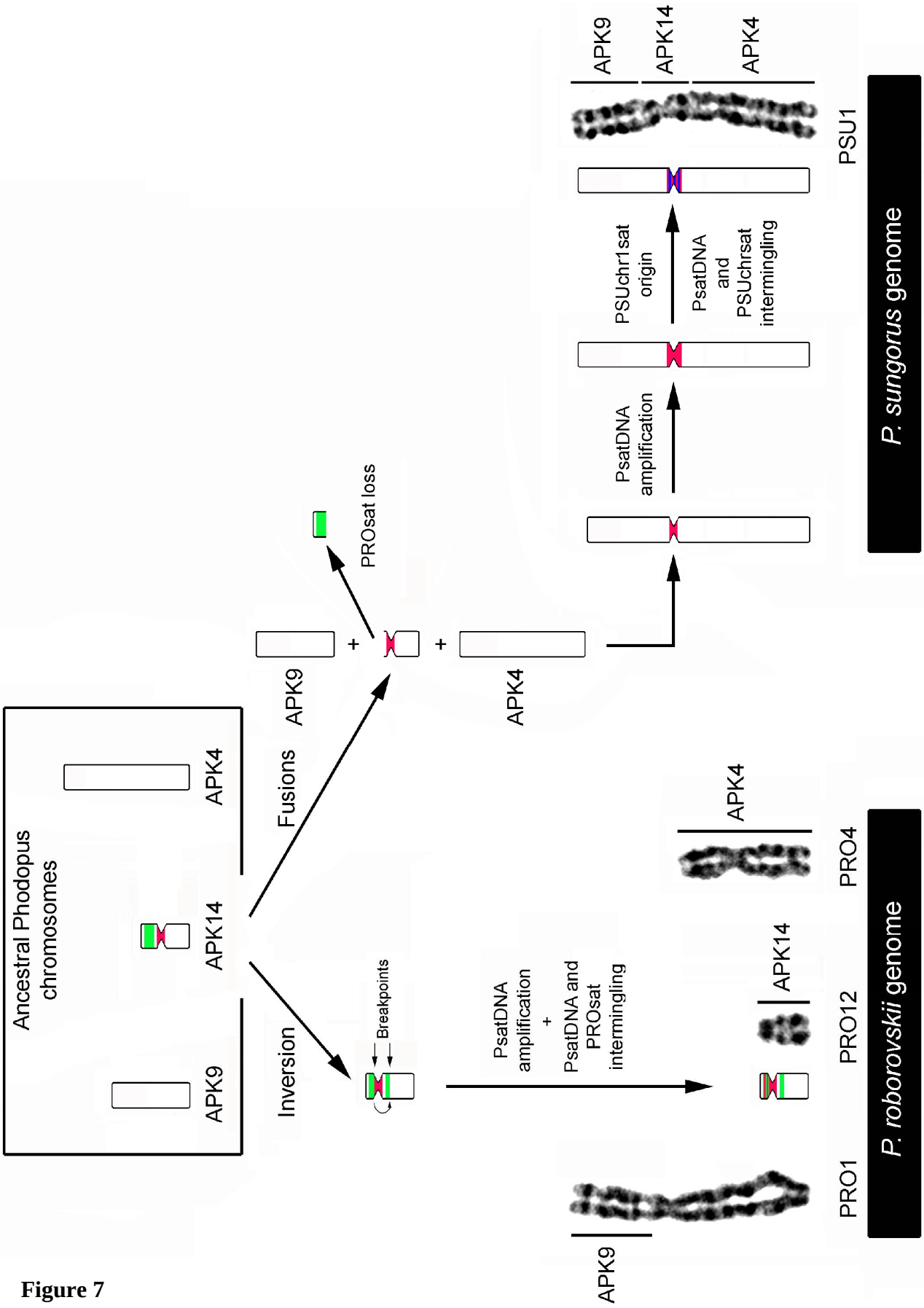

Ancestral Phodopus chromosomes
APK4
APK14
APK9
Fusions
Inversion
Breakpoints
PROsat loss
APK9
APK4
PsatDNA amplification
PSUchr1sat origin
PsatDNA and PSUchrsat intermingling
APK9
APK14
APK4
PSU1
P. sungorus genome
PsatDNA amplification + PsatDNA and PROsat intermingling
APK4
APK14
APK9
PRO4
PRO12
PRO1
P. roborovskii genome

Figure 7

II.2.4

Quest for the functional significance of a satellite DNA sequence from *Peromyscus eremicus* (Cricetidae, Rodentia)

Quest for the functional significance of a satellite DNA sequence from *Peromyscus eremicus* (Cricetidae, Rodentia)

[1]Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGBUTAD), Apdo 1013, 5001-801 Vila Real, Portugal

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

*Author for Correspondence: Raquel Chaves, Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGB-UTAD), Vila Real, Portugal, +351259350841, rchaves@utad.pt

Abstract

Despite the large collection of works describing the isolation and molecular characterization of satellite DNAs (satDNAs) in eukaryotic genomes, much more is needed until the exact role(s) played by these repetitive sequences be completely elucidated. Here we isolated a novel (peri)centromeric satDNA from the genome of *Peromyscus eremicus* (Cricetidae), named PERcentSat, presenting a monomeric unit of 21 bp. The molecular analysis and the similarity observed between PERcentSat and two other satDNAs, MSAT21 from *Microtus* and the human centromeric satellite HSAT6, suggest that these three satellites were probably originated from an ancestral satDNA with an 18 bp repeat unit. The origin of this ancestral sequence dates back to the divergence time between Primates and Rodents, at ~91.9My ago. The identification of transcription factor binding sites in PERcentSat, Heat

Shock Factor 1 (HSF1) and CRE-Binding protein 1/c-Jun heterodimer (CRE-BP1/c-Jun) supports its transcription induced by stress stimulus. Moreover, the similarity presented by PERcentSat repeat units and a small region of two previously described coding mRNAs, allowed us to suggest that this satellite can control gene expression by RNAi post-transcriptional silencing. Interestingly, this expression silencing may act as feedback control for PERcentSat transcription. Contrary to the idea of satDNAs as non-coding sequences, here three Open Reading Frames (ORFs) were also identified for PERcentSat. This allows us to conjecture about the possibility of this satellite is evolving towards a coding sequence. The similarity observed between a putative PERcentSat polypeptide and a Microtubule-associated protein 1A points to a role of PERcentSat in centromeric function, which is also supported by the conservation of this satDNA centromeric location during ~91.9 My. In accordance with all these evidences, we can attribute several functions to PERcentSat, as an involvement in the cellular response to stress (by its transcripts), control of gene expression and/or centromeric function, acting probably in the chromosome segregation by affecting microtubule assembly.

Keywords: satellite DNA, *Peromyscus eremicus*, Transcription, Stress stimulus, Centromeric Activity.

Introduction

Satellite DNAs (satDNAs) are tandemly repeated sequences, organized into long uninterrupted arrays (usually megabase-sized arrays) in the heterochromatic regions of the chromosomes (Charlesworth et al. 1994), assuming like this a preferential centromeric/pericentromeric location. However, the presence of these sequences in (sub)telomeric or interstitial regions were also reported (e.g. Santos et al. 2004, Louzada et al. 2008). In eukaryotes, a large fraction of its genomes is composed by this type of repeats. A good example of that is the kangaroo rat (*Dipodomys ordii*), where satDNAs comprises up to 50% of the genome (Singer 1982).

It has been postulated that satDNAs evolve in a concerted way within a genome, resulting in a homogenization of changes among monomeric units (repeat units) and their subsequent fixation in the members of a population (Dover 1986, Dover 2002). Different molecular mechanisms, such as gene conversion, unequal crossing-over, rolling circle replication/reinsertion and transposon-mediated exchange, are responsible for this homogenization of satDNA units (Walsh 1987, Elder and Turner 1995, Dover 2002). During the homogenization of repeats, these mechanisms can enable the drastic change in copy number of repeat units and also the intragenomic movements of satDNAs. Therefore, satDNAs show significant sequence divergence, as well as variation in copy number and/or chromosomal location, even between closely related species (Slamovits and Rossi 2002, Adega et al. 2009).

The described satDNA sequence structure based on tandem repeats as well as its heterochromatic localization, led the initial belief that these sequences were not transcribed (Skinner et al. 1977). Presently, it is becoming increasingly accepted that the transcription of satDNAs is a general event, having been reported satellite transcripts in several organisms including vertebrates, invertebrates and plants (Wong et al. 2007, Vourc'h and Biamonti

2011, Hall et al. 2012, Pezer and Ugarković 2012, Enukashvily and Ponomartsev 2013). But so far little is known about basic mechanisms of satDNA expression and regulation (Pezer and Ugarković 2008, Vourc'h and Biamonti 2011, Enukashvily and Ponomartsev 2013). In light of the current knowledge, satellite transcripts are usually heterogeneous in size and the transcription can proceeds in both DNA strands or to be strand-specific (Rudert et al. 1995, Rouleux-Bonnin et al. 1996). Most of these transcripts are present as polyadenylated RNA in the cytoplasm but some are found exclusively in the nucleus (Trapitz et al. 1988, Bonaccorsi et al. 1990). The occurrence of developmental stage and tissue-specific differential expression of some satellites provide clues that their transcripts have a regulatory role (Pezer and Ugarković 2012, Enukashvily and Ponomartsev 2013).

Different works also report that satDNA transcripts can act as long RNAs or precursors of small interfering RNAs (siRNAs). Long satellite RNAs have been considered as functional components of kinetochore, participating in recruitment of kinetochore proteins (Wong et al. 2007, Hall et al. 2012, Enukashvily and Ponomartsev 2013). Besides, was also referred the involvement of long satellite RNAs in the regulation of splicing during stress or its action as ribozymes with self-cleavage activity (reviewed by Ugarković 2005). Considering the satellite derived siRNAs, it is recognized that these transcripts have an important role in chromatin remodelation (leading to the heterochromatin formation and maintenance) and in the control of gene expression (Vourc'h and Biamonti 2011, Enukashvily and Ponomartsev 2013). For that, satellite siRNAs need to be associated with two different complexes of argonaute proteins, resulting in the RNA-induced transcriptional silencing complex (RITS) or in the RNA-induced silencing complex (RISC). The siRNA-loaded RIST complex regulates chromatin modifications, as DNA and histone methylations characteristic of heterochromatin, via interactions with nascent transcripts (reviewed by Pezer and Ugarković 2008, Hall et al. 2012). The siRNA-loaded RISC complex controls the gene expression by the pairing between

satellite siRNAs and complementary gene mRNAs, targeting these mRNAs for destruction (reviewed by Buckingham 2003, Ugarković 2005).

Works describing the presence of a major satDNA (PMsat) in *Peromyscus eremicus* pericentromeric regions and in some chromosome short arms (Louzada et al. 2014 submitted for publication), as well as, a satDNA with an interspersed location (CCR4/10sat, Louzada et al. 2008), correspond to the most recent works on the repetitive genome fraction of this Cricetidae species. In this paper we report the isolation and molecular characterization of a novel *Peromyscus eremicus* satDNA, PERcentSat. The evidences collected here point to a role of this satellite in the response to stress, control of gene expression and centromeric activity.

Material and Methods

Chromosome preparations and Genomic DNA extraction

Fixed chromosome preparations from *Peromyscus eremicus* were obtained from fibroblast cell cultures, using standard procedures described elsewhere (Chaves et al. 2004). Genomic DNA was obtained from these fibroblast cell cultures using the JETQUICK DNA kit (Genomed).

Isolation, cloning and sequencing of PERcentSat sequence

PER genomic DNA was digested with the restriction endonuclease (RE) HaeIII, according to manufacturers' instructions (Invitrogen Life Technologies), resulting in a smear with DNA fragments that range between 3kb to 100bp. The restriction products were later inserted into the plasmid vector pUC19 (Thermo scientific) and used to transform competent *Escherichia coli* DH5α cells (Invitrogen Life Technologies). A part of the obtained colonies was lift onto a nylon membrane Hybond^{TM}-N^{+} (Amersham, GE Healthcare) and the DNA in the membrane

was probed to HaeIII restriction products labeled with digoxigenin-11-d'UTP, using DIG DNA labeling Kit (Roche Molecular Biochemicals). Hybridization was performed at 68 ºC as described by Bruvo et al. (2003). Positive signals were visualized using chemiluminiscent CDP-Star system (Roche Molecular Biochemicals). The plasmid DNA of the positive clones was isolated using the High Pure Plasmid Isolation kit (Roche Molecular Biochemicals) and sequenced in both directions using universal M13 primers.

Sequence analysis of PERcentSat sequence

PERcentSat was analyzed with different sequence databases tools and bioinformatic software's: NCBI Blast (http://www.ncbi.nlm.gov/Blast/), RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker), NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), Vector NTI advance 11 (Invitrogen Life Technologies), Tandem repeats Finder (Benson 1999, version 4.00, free download in http://tandem.bu.edu/trf/trf.html) and TFsearch (http://www.cbrc.jp/research/db/TFSEARCH.html). A BLAST search for PERcentSat sequence, against nucleotide sequences present in Genbank and RepBase was accomplished using NCBI blast and RepeatMasker tools. The identification of Open Reading Frames was performed using NCBI ORF Finder database tool. Sequence alignments were performed with the software vector NTI advance 11 that apply the Clustal W algorithm (Thompson et al. 1994) to verify sequence similarities. The search for direct repeats within the PERcentSat sequence was performed using the Tandem Repeats Finder software. TFsearch tool allowed the analysis for the presence of transcription factor binding sites in this sequence.

Physical mapping of PERcentSat and PMsat sequence

The physical mapping of PERcentSat and PMsat sequences in the chromosomes of *Peromyscus eremicus* was carried out by Fluorescent *in situ* Hybridization (FISH) procedures, described by Schwarzacher and Heslop-Harrison (2000). A clone of PMsat was provided by Louzada et al. (2014 submitted for publication). The two sequences were labeled with digoxigenin-11-dUTP or biotin-16-dUTP (Roche, Molecular Biochemicals) by PCR. The probe was denatured at 85°C for 15 min and then snap-cooled on ice. The most stringent post-hybridization wash was in 50% formamide/2×SSC at 42 ºC. Biotin-labeled probes were detected by FITC conjugated with avidin (Roche Molecular Biochemicals), digoxigenin-labeled probes were detected with anti-digoxigenin-5'TAMRA (Roche Molecular Biochemicals).

CBP-banding sequential to physical mapping of PERcentSat sequence

After distaining the slides, sequential CBP-banding [C-bands by Barium Hydroxide with Propidium Iodide] was performed according to the standard procedure of Sumner (1972) with slight modifications. Shortly, the slides were submitted to hydrochloric acid (0.1 M) during 20 min, barium hydroxide (5% solution) during 7 min and 2× saline solution citrate (2×SSC: 0.3 mol/L NaCl, 0.03 mol/L sodium citrate) at 60 º for 40 min.

Capture and preparation of images

Chromosomes were analyzed in a Zeiss Axioplan 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); contrast and color optimization were the functions used and affected the whole image equally. The

chromosomes of *Peromyscus eremicus* were identified according to The Committee for Standardization of Chromosomes of *Peromyscus* (1977).

Results

Molecular features of PERcentSat

In this work it was isolated and molecularly characterized a satellite DNA (satDNA) sequence from the genome of *Peromyscus eremicus*, named PERcentSat, presenting a monomeric unit with 21 bp. The determination of this repeat unit length was achieved analyzing PERcentSat clones with the Tandem repeats Finder software, allowing the identification of several tandem short repeats with a length of 21 bp (Figure 1A). Similarity values between these repeats varies within 76,2% to 95,2% (Figure 1B).

BLAST search revealed significant similarity between PERcentSat and other repetitive sequences deposited in NCBI nucleotide or in Repbase databases, a repeated DNA sequence of *Microtus* species (MSAT21, NCBI L43073.1, Modi et al. 2003) and a human centromeric satellite (HSAT6), as can be observed in figure 1A. The alignment between the 21 bp PERcentSat repeat unit and the 21 bp MSAT21 repeat unit (*consensus* sequences) is present in figure 1C. These two sequences present a similarity of 95.2%. The 21bp PERcentSat repeat unit (*consensus* sequence) and the 18 bp subunit of HSAT6 present a similarity of 66.7%. Moreover, PERcentSat also presents similarity with mRNA sequences, *Nannospalax galili* isotig18792.Nagamubr mRNA [NCBI JL986735.1, belonging to the transcriptome sequencing library of *Nannospalax galili* (Malik et al. 2011)], *Ciona intestinalis* E3 ubiquitin-protein ligase HECTD2 mRNA (NCBI XM_002125758.1, Gene ID: 100179398, Protein NCBI XP_002125794.1) and *Ciona intestinalis* fibroblast growth factor receptor mRNA (NCBI NM_001044355.1, Gene ID: 445706, Protein NCBI NP_001037820.1). Despite the

high similarity (80 to 81%) verified between PERcentSat and the last two referred mRNAs, the region in question is very short. About 2,2% and 5% of the *Ciona intestinalis* E3 ubiquitin-protein ligase HECTD2 mRNA and fibroblast growth factor receptor mRNA (respectively) align with this satellite, suggesting a possible involvement of the satDNA in the regulation of these genes by the siRNA mechanism. Meanwhile, PERcentSat also displays a high similarity (82%) with a large region of *Nannospalax galili* isotig18792.Nagamubr mRNA (78% of this mRNA) (82% of similarity). This mRNA sequence was isolated from *Nannospalax galili* cells submitted to stress stimulus.

Analysis of PERcentSat with the NCBI ORF Finder allows the identification of three Open Reading Frames (ORFs) (Figure 1A), whose putative polypeptides present similarity with hypothetical, predicted or described proteins, *Branchiostoma floridae* hypothetical protein BRAFLDRAFT (NCBI XP_002589899.1), *Saccoglossus kowalevskii* predicted protein-like (NCBI 002732031.1), *Helicobacter pylori* cell surface protein (NCBI YP_005774454.1), *Cricetulus griseus* Microtubule-associated protein 1A (NCBI EGW00944.1) and *Cricetulus griseus* WD repeat-containing protein (NCBI EGW03489.1). These similarity values vary from 48% to 69%, presenting the *Cricetulus griseus* Microtubule-associated protein 1A the higher similarity with a putative PERcentSat polypeptide. However, it is important to refer that only 6.2% of this protein align with a putative PERcentSat polypeptide.

A search for transcription factor binding sites in PERcentSat clones, using the TFsearch tool, allowed the recognition of a binding site for the Heat Shock Factor 1 (HSF1) and four sites for CRE-Binding protein 1/c-Jun heterodimer (CRE-BP1/c-Jun), as can be seen in figure 1D.

Chromosomal distribution of PERcentSat

Physical mapping of PERcentSat onto *Peromyscus eremicus* chromosomes revealed that this sequence presents a chromosome distribution characteristic of a tandem repeat, hybridizing with the (peri)centromeric region of 14 autosomal pairs, PER2, PER7, PER8, PER10, PER11, PER13, PER14, PER15, PER16, PER17, PER18, PER20, PER21 and PER22 (Figure 2A). In this work, the chromosomal distribution of PERcentSat was also analyzed in comparison with the location of PMsat, previously described on this species (Louzada et al. 2014 submitted for publication). In figure 2B is possible to observe the characteristic hybridization pattern of PMsat on the chromosomes of *Peromyscus eremicus*, being possible to observe in the figure 2C the simultaneous hybridization of PERcentSat (red signals) and PMsat sequences (green signals) with the chromosomes of this species. C-banding performed sequentially to FISH evidenced a co-localization of PERcentSat with constitutive heterochromatin (CH), as can be seen in figure 2D. In figure 2E was compared in more detail PERcentSat location with PMsat and CH distribution. PERcentSat presents a more centromeric location comparing with PMsat, presenting this last satellite a more pericentromeric position (seems to occupy all the entire short arm of these chromosomes). In some chromosome pairs, two PMsat repetitive blocks “sandwiched” a PERcentSat block in the centromeric region, as it can be seen in chromosomes PER7, PER11, PER13, PER15, PER16, PER20 and PER21 (Figure 2E).

Discussion

In this work we describe a novel (peri)centromeric satellite DNA (satDNA) sequence from the genome *Peromyscus eremicus* (PERcentSat). This sequence presents a short repeat unit of 21 bp and orthology to other satDNAs already reported, MSAT21 from *Microtus* species (Modi et al. 2003) and the human centromeric satellite HSAT6 (submitted directly by

Pavlicek and Jurka to Repbase database). MSAT21 was first described in the sex-chromosomes of *Microtus chrotorrhinus,* composed by tandem repetitions of 21 bp (Ivanov and Modi 1996). Later, this sequence was also identified in the sex-chromosomes and/or mainly in the autosomal centromeric regions of seven more *Microtus* species (Modi et al. 2003). HSAT6 was characterized as a human centromeric satellite, combined by tandem repetitions of 42 bp with two subunits of 18 bp. The alignment region and the similarity values between these three satellites (PERcentSat, MSAT21 and HSAT6), suggests that they were originated from an ancestral satDNA with a repeat unit of 18 bp, present in the common ancestral genome of the species exhibiting it. The origin of this ancestral sequence dates back to the divergence time between Primates and Rodents, at ~ 91.9 My (according to OneZoom Tree of Life Explorer database, http://www.onezoom.org/index.htm, updated December 2013).

In a genome, the homogenization of satDNA repeat units (concerted evolution) seems to occur more efficiently within a localized subset of repeat motifs, decreasing the efficiency when changes are homogenized between different arrays (reviewed by Plohl et al. 2008). Likewise homogenization mechanisms can generate satellite subfamilies in a genome, or induce mutations and sequence rearrangements that could culminate in the generation of novel satellite repeats, often from simple sequence motifs (Ugarković and Plohl 2002). This kind of evolution could also explain the origin of PERcentSat, MSAT21 and HSAT6 repeat units from a short motif of 18 bp. Examples of this satDNA repeat unit evolution were reported for mouse Major satellite, presenting a 234 bp repeat unit based on a 9 bp motif (Horz and Altenburger 1981), or for the a bovine 1.715 satDNA, originated from a basic 31 bp subrepeat (Jobse et al. 1995). The long evolutionary preservation of these repeats sharing an ancestral sequence with at least 91.9 My, is indicative of functional significance.

The similarity observed between PERcentSat and a *Nannospalax galili* isotig18792.Nagamubr mRNA (Malik et al. 2011), as also the extension of the alignment region between these two sequences (~78% of the *Nannospalax galili* mRNA), suggest that PERcentSat is transcribed. The identification of transcription factor binding sites in PERcentSat, Heat Shock Factor 1 (HSF1) and CRE-Binding protein 1/c-Jun heterodimer (CRE-BP1/c-Jun) binding sites, corroborates PERcentSat transcription activity, allowing also to speculate how the transcription of this satellite can be regulated. Like PERcentSat, the pericentromeric human satellite III has a binding motif for HSF1, being specifically expressed under stress, resulting in long single-stranded polyadenylated transcripts (Jolly et al. 2004, Metz et al. 2004). This heat shock factor drives RNA pol II transcription (Metz et al. 2004), pointing to an involvement of this enzyme in PERcentSat transcription when a stress heat shock stimulus occurs. The identification of CRE-BP1/c-Jun heterodimer binding sites in PERcentSat also supports a transcription induced by stress stimulus of this satellite. The transcriptional activity of CRE-BP1/c-Jun heterodimer (or ATF2/ c-Jun heterodimer, Hai and Curran 1991) is stimulated by cellular stress following the JNK signal transduction pathway (Hayakawa et al. 2003). This signalling pathway is activated primarily by exposure to environmental stress, thereby controlling the response of cells to these stimuli. According to that, the high stress environment conditions existent in the habitat of *Peromyscus eremicus*, living in desert regions throughout the southwestern United States, North/Central Mexico and Baja California (Veal and Caire 1979), might control the transcription of PERcentSat that in turn certainly plays a role in the cellular response to these stress conditions. The similarity identified between PERcentSat and a *Nannospalax galili* mRNA (mentioned previously), belonging to cDNA pools from muscle and brain tissues isolated from animals exposed to stress conditions (Malik et al. 2011), is also in agreement with PERcentSat transcription induction by response to environmental stress stimulus.

PERcentSat may also have a role in the control of gene expression. The similarity observed between PERcentSat repeat units and a small region of two genes mRNA, gene for E3 ubiquitin-protein ligase HECTD2 and for the fibroblast growth factor receptor of *Ciona intestinalis*, may suggest an involvement of PERcentSat transcripts in the control of these two genes' expression. Complementary PERcentSat siRNAs could target these genes mRNA for destruction, in a process named post-transcriptional silencing by RNAi, inhibiting the function of these genes (codification for an E3 ubiquitin ligase that targets proteins to degradation by the proteosome (Zhao et al. 2000) and codification for growth factor receptors (Herbst 2004)). Similar assumptions were also proposed in other species. For example, the presence of several coding mRNAs in human and chick embryos that contain α-like satellite repeat as part of their 5' or 3' untranslated region indicates that their expression could be controlled by siRNAs derived from α-satellite repeats (Li and Kirby 2003). The small size of PERcentSat repeat units (21 bp) is also in accordance with siRNAs originated from this satellite. Interestingly, the putative capacity of PERcentSat siRNAs for the control of these two genes expression, may allow a feedback regulation of the satellite transcription. The ubiquitin-proteasome pathway plays a central role in the regulation of essential cellular processes such as transcription/signal transduction (Zhao et al. 2000), having been described that transcription factors like C-Jun are regulated by ubiquitination (Treier et al. 1994). Like this, the silencing of E3 ubiquitin-protein ligase gene, could results into the non degradation of CRE-BP1/c-Jun heterodimer (likely involved in PERcentSat transcription) and consequently in the maintenance of PERcentSat transcription. Regarding the gene coding for growth factor receptors, its silencing by PERcentSat siRNAs may results in the absence of cell surface receptors and consequently in the incapacity of these cells to "sense" environmental stimuli. If these stimuli do not arrive to the nucleus, PERcentSat may not be transcribed.

The identification of three Open Reading Frames (ORFs) in PERcentSat, whose putative polypeptides present similarity with hypothetical, predicted or described proteins (values ranging from 48% to 69% comprising short alignment regions), reinforce the functional significance of this satellite. In this context, it could be also speculated that the evolutionary pathway of this satellite, which began at least 91.9 My ago, could result in the origin of a coding sequence. This is supported by works reporting the origin of new protein-coding genes from non-coding DNA (Wu et al. 2011, Murphy and McLysaght 2012, Xie et al. 2012). The similarity identified between a putative PERcentSat polypeptide and a cell surface protein, is in agreement with the probable induction of PERcentSat transcription in response to environmental stress stimulus. Moreover, the similarity of a putative PERcentSat polypeptide and a Microtubule-associated protein 1A points to a role of PERcentSat in centromeric function, since this protein bind tubulin subunits that make up microtubules regulating their assembly (Maccioni and Cambiazo 1995). The exclusive (peri)centromeric chromosomal location of this satellite in *Peromyscus eremicus* supports its role in centromeric function. From all satDNAs described up to now in *Peromyscus eremicus* genome (PMsat and CCR4/10sat, Louzada et al. 2008 and Louzada et al. 2014 submitted for publication), PERcentSat is the one presenting an exclusive and more precise centromeric location. The importance of PERcentSat in the centromeric function is also supported by the similarity between this satellite and MSAT21 and HSAT6, two satellites that present a centromeric location.

Taken together, the results obtained in this work sustain without any doubts that satDNAs definitely present important functions in the genomes. Regarding particularly the satDNA described here (PERcentSat) we can attribute three major roles to this sequence: involvement in the cellular response to stress, control of gene expression and centromeric function, operating possibly in the chromosome segregation by affecting microtubule assembly.

Acknowledgements

This work was supported by the project POCI/BIA-BCM/58541/2004, a PhD grant (SFRH/BD/41574/2007) 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. We are deeply grateful to Dr. Vitaly Volobouev for providing the *Peromyscus eremicus* 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.

Bonaccorsi S, Gatti M, Pisano C, Lohe A (1990). Transcription of a satellite DNA on two Y chromosome loops of *Drosophila*. Chromosoma 99: 260–266.

Benson G (1999). "Tandem repeats finder: a program to analyze DNA sequences". Nucleic Acids Res 27 (2): 573-580.

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.

Buckingham S (2003). The major world of microRNAs. Horizon Symposia 1-3.

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

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

Committee for standartization of chromosomes of *Peromyscus* (1977). Standard katyotype of the deer mice, Permyscus (Rodentia). *Cytogenet. Cell Genet*. 19: 38-43.

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

Dover G (2002). Molecular drive. Trends Genet. 18: 587–589.

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

Enukashvily NI, Ponomartsev NV (2013). Mammalian Satellite DNA: A Speaking Dumb. Advances in *Protein Chemistry* and Structural Biology 90: 31-65.

Hai T and Currant T (1991). Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proc. Nati. Acad. Sci. USA 88: 3720-3724.

Hall LE, Mitchell SE, O'Neill RJ (2012). Pericentric and centromeric transcription: a perfect balance required. Chromosome Research (2012) 20:535–546.

Hayakawa J, Depatie C, Ohmichi M, Mercola D (2003). The Activation of c-Jun NH2-terminal Kinase (JNK) by DNA-damaging Agents Serves to Promote Drug Resistance via Activating Transcription Factor 2 (ATF2)-dependent Enhanced DNA Repair. The Journal of Biological Chemistry 278(23): 20582–20592.

Herbst RS (2004). Review of Epidermal Growth Factor Receptor Biology. Int. J. Radiation Oncology Biol. Phys. (59)2: 21–26.

Horz W, Altenburger W (1981). Nucleotide sequence of mouse satellite DNA. Nucleic Acids Res. 9: 683-696.

Ivanov SV, Modi WS (1996). Molecular characterization of the complex sex-chromosome heterochromatin in the rodent *Microtus chrotorrhinus*. Cytogenet Cell Genet. 75(1): 49-56.

Jobse C, Buntjer JB, Haagsma N, Breukelman HJ, Beintema JJ, Lenstra JA (1995) Evolution and recombination of bovine DNA repeats. J. Mol. Evol. 41: 277-283.

Jolly C, Metz A, Govin J, Vigneron M, Turner BM, Khochbin S, Vourc'h C (2004). Stress-induced transcription of satellite III repeats. J Cell Biol 164: 25–33.

Li YX, Kirby ML (2003). Coordinated and conserved expression of alphoid repeat and alphoid repeat-tagged coding sequences. Dev Dynamics 228: 72–81.

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.

Louzada S, Vieira-da-Silva A, Kubickova S, Rubes J, Adega F, Raquel Chaves R (2014). A novel satellite DNA sequence in *Peromyscus* genome that evolves by copy number fluctuation: PMSat Is it an "ancient" or a "nascent" satellite DNA?. Submitted for publication to Heredity.

Maccioni RB, Cambiazo V (1995). Role of Microtubule-Associated Proteins in the Control of Microtubule Assembly. Physiological Reviews 75 (4): 835-864.

Malik A, Korol A, Hübner S, Hernandez AG, Thimmapuram J, Ali S, Glaser F, Paz A, Avivi A, Band M (2011). Transcriptome Sequencing of the Blind Subterranean Mole Rat, Spalax galili: Utility and Potential for the Discovery of Novel Evolutionary Patterns. PLoS ONE 6(8): e21227.

Metz A, Soret J, Vourc'h C, Tazi J, Jolly C (2004). A key role for stress-induced satellite III transcripts in the relocalization of splicing factors into nuclear stress granules. J Cell Sci 117: 4551–4558.

Modi WS, Serdyukova NA, Vorobieva NV, Graphodatsky AS (2003). Chromosomal localization of six repeated DNA sequences among species of *Microtus* (Rodentia). Chromosome Research 11: 705-713.

Murphy DN, McLysaght A (2012). De Novo Origin of Protein-Coding Genes in Murine Rodents. PLOS ONE 7(11): e48650.

Pezer Ž and Ugarković Đ (2012). Satellite DNA-associated siRNAs as mediators of heat shock response in insects. RNA Biology 9(5): 587-595.

Pezer Ž, Ugarković Đ (2008). Role of non-coding RNA and heterochromatin in aneuploidy and cancer. Seminars in Cancer Biology 18: 123–130.

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.

Rouleux-Bonnin F, Renault S, Bigot Y, Periquet G (1996). Transcription of four satellite DNA subfamilies in *Diprion pini* (Hymenoptera, Symphyta, Diprionidae). Eur J Biochem 238: 752–759.

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

Santos S, Chaves R, Guedes-Pinto H (2004). Chromosomal localization of the major satellite DNA family (FA-SAT) in the domestic cat. Cytogenet. Genome Res. 107: 119–122.

Schwarzacher T, Heslop-Harrison JS (2000). Practical *in situ* hybridization. Bios Scientific Publishers, Oxford.

Singer MF (1982) Highly repeated sequences in mammalian genomes. Int Rev Cytol 76:67-112.

Skinner DM (1977). Satellite DNAs. BioSciences 27: 790–796.

Slamovits CH, Rossi MS (2002). Satellite DNA: Agent of chromosomal Evolution in mammals. A review. Mastozoología Neotropical / J. Neotrop. Mammal 9(2): 297-308.

Sumner AT (1972). A simple technique for demonstrating centromeric heterochromatin. Exp Cell Res 75:304-306.

Thompson JD, Higgins DG, Gibson TJ (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22): 4673-4680.

Trapitz P, Wlaschek M, Bunemann H (1988). Structure and function of Y chromosomal DNA. II. Analysis of lampbrush loop associated transcripts in nuclei of primary spermatocytes of *Drosophila hydei* by *in situ* hybridization using asymetric RNA probes of four different families of repetitive DNA. Chromosoma 96: 159–170.

Treier M, Staszewski LM, Bohmann D (1994). Ubiquitin-dependent c-Jun degradation in vivo is mediated by the delta domain. Cell 78: 787-798.

Ugarković Đ, Plohl M (2002). Variation in satellite DNA profiles-causes and effects. The EMBO Journal 21(22): 5955-5959.

Ugarković Đ (2005). Functional elements residing within satellite DNAs. EMBO reports 6(11): 1035-1039.

Veal R, Caire W (1979). *Peromyscus eremicus*. Mammalian species 118: 1-6.

Vourc'h C, Biamonti G (2011). Long Non-Coding RNAs. Progress in Molecular and Subcellular Biology. In Ugarkovic D (Ed) Transcription of Satellite DNAs in Mammals. Springer-Verlag Berlin Heidelberg, pp. 95-118.

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, et al. (2007). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res 17: 1146–1160.

Wu D-D, Irwin DM, Zhang Y-P (2011). De Novo Origin of Human Protein-Coding Genes. PLoS Genetics 7(11): e1002379.

Xie C, Zhang YE, Chen J-Y, Liu C-J, Zhou W-Z, Li Y, Zhang M, Zhang R, Wei L, Li C-Y (2012). Hominoid-Specific De Novo Protein-Coding Genes Originating from Long Non-Coding RNAs. PLOS Genetics 8(9): e1002942.

Zhao J, Tenev T, Martins LM, Downward J, Lemoine NR (2000). The ubiquitin-proteasome pathway regulates survivin degradation in a cell cycle-dependent manner. Journal of Cell Science 113: 4363-4371.

Legends of figures

Figure 1- Sequence analysis of a PERcentsat clone. (A) Schematic representation of PERcentSat molecular features. The sequence analysis similarity performed for this sequence is also shown in this figure. (B) Alignment of PERcentSat repeat units. (C) Alignment between the 21 bp repeat units of PERcentSat and MSAT21 (*consensus* sequences) and alignment between the 21 bp repeat unit of PERcentSat (*consensus* sequence) and the 18 bp HSAT6 subunit. (D) Identification of transcription factor binding sites in a PERcentSat clone.

Figure 2- Organization of two satellite DNA sequences in (peri)centromeric regions of *Peromyscus eremicus* chromosomes. (A) Representative *in situ* hybridization presenting the chromosomal location of PERcentSat on chromosomes of *Peromyscus eremicus.* The satellite sequence was labelled with digoxigenin-11-dUTP and detected with 5'TAMRA (red). (B) Representative *in situ* hybridization presenting the chromosomal localization of PMsat on chromosomes of *Peromyscus eremicus.* The satellite sequence was labelled with biotin-16-dUTP and detected by FITC conjugated with avidin (green). (C) Overlapping of PERcentSat and PMsat hybridization signals. (D) Same metaphase after sequential C-banding. (E) Resume table presenting the chromosomes that display PERcentSat sequence. All chromosomes were counterstained with Propidium Iodide (red). The comparison between the location of this sequence with PMsat and the constitutive heterochromatin can also be observed in this table.

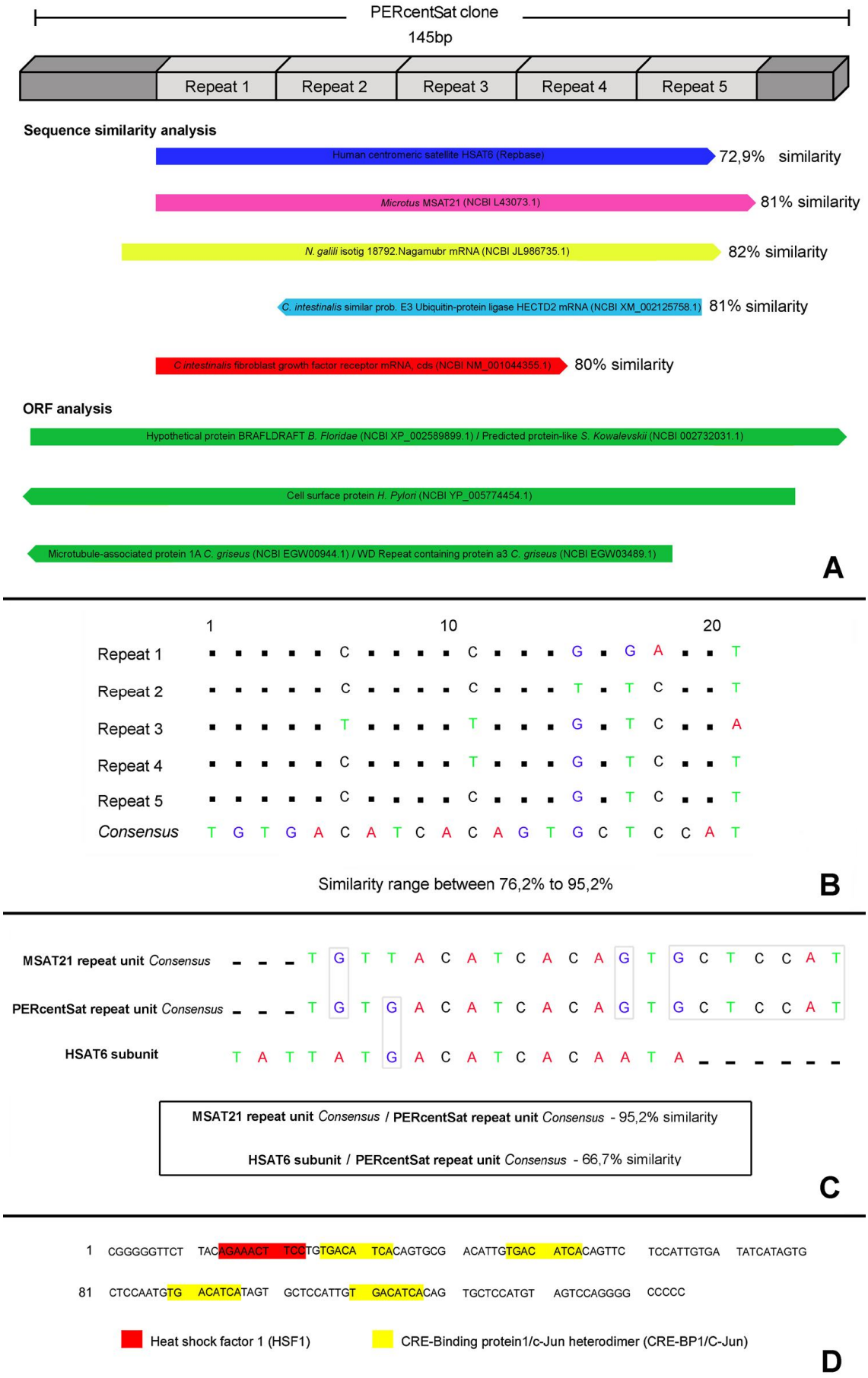

PERcentSat clone
145bp
Repeat 1
Repeat 2
Repeat 3
Repeat 4
Repeat 5
Sequence similarity analysis
Human centromeric satellite HSAT6 (Repbase)
72,9% similarity
Microtus MSAT21 (NCBI L43073.1)
81% similarity
N. galili isotig 18792.Nagamubr mRNA (NCBI JL986735.1)
82% similarity
C. intestinalis similar prob. E3 Ubiquitin-protein ligase HECTD2 mRNA (NCBI XM_002125758.1)
81% similarity
C intestinalis fibroblast growth factor receptor mRNA, cds (NCBI NM_001044355.1)
80% similarity
ORF analysis
Hypothetical protein BRAFLDRAFT B. Floridae (NCBI XP_002589899.1) / Predicted protein-like S. Kowalevskii (NCBI 002732031.1)
Cell surface protein H. Pylori (NCBI YP_005774454.1)
Microtubule-associated protein 1A C. griseus (NCBI EGW00944.1) / WD Repeat containing protein a3 C. griseus (NCBI EGW03489.1)
A
1
10
20
Repeat 1 ▪ ▪ ▪ ▪ ▪ C ▪ ▪ ▪ ▪ C ▪ ▪ ▪ G ▪ G A ▪ ▪ T
Repeat 2 ▪ ▪ ▪ ▪ ▪ C ▪ ▪ ▪ ▪ C ▪ ▪ ▪ T ▪ T C ▪ ▪ T
Repeat 3 ▪ ▪ ▪ ▪ ▪ T ▪ ▪ ▪ ▪ T ▪ ▪ ▪ G ▪ T C ▪ ▪ A
Repeat 4 ▪ ▪ ▪ ▪ ▪ C ▪ ▪ ▪ ▪ T ▪ ▪ ▪ G ▪ T C ▪ ▪ T
Repeat 5 ▪ ▪ ▪ ▪ ▪ C ▪ ▪ ▪ ▪ C ▪ ▪ ▪ G ▪ T C ▪ ▪ T
Consensus T G T G A C A T C A C A G T G C T C C A T
Similarity range between 76,2% to 95,2%
B
MSAT21 repeat unit Consensus - - - T G T T A C A T C A C A G T G C T C C A T
PERcentSat repeat unit Consensus - - - T G T G A C A T C A C A G T G C T C C A T
HSAT6 subunit T A T T A T G A C A T C A C A A T A - - - - - -
MSAT21 repeat unit Consensus / PERcentSat repeat unit Consensus - 95,2% similarity
HSAT6 subunit / PERcentSat repeat unit Consensus - 66,7% similarity
C
1 CGGGGGTTCT TACAGAAACT TCCTGTGACA TCACAGTGCG ACATTGTGAC ATCACAGTTC TCCATTGTGA TATCATAGTG
81 CTCCAATGTG ACATCATAGT GCTCCATTGT GACATCACAG TGCTCCATGT AGTCCAGGGG CCCCC
Heat shock factor 1 (HSF1)
CRE-Binding protein1/c-Jun heterodimer (CRE-BP1/C-Jun)
D

Figure 1

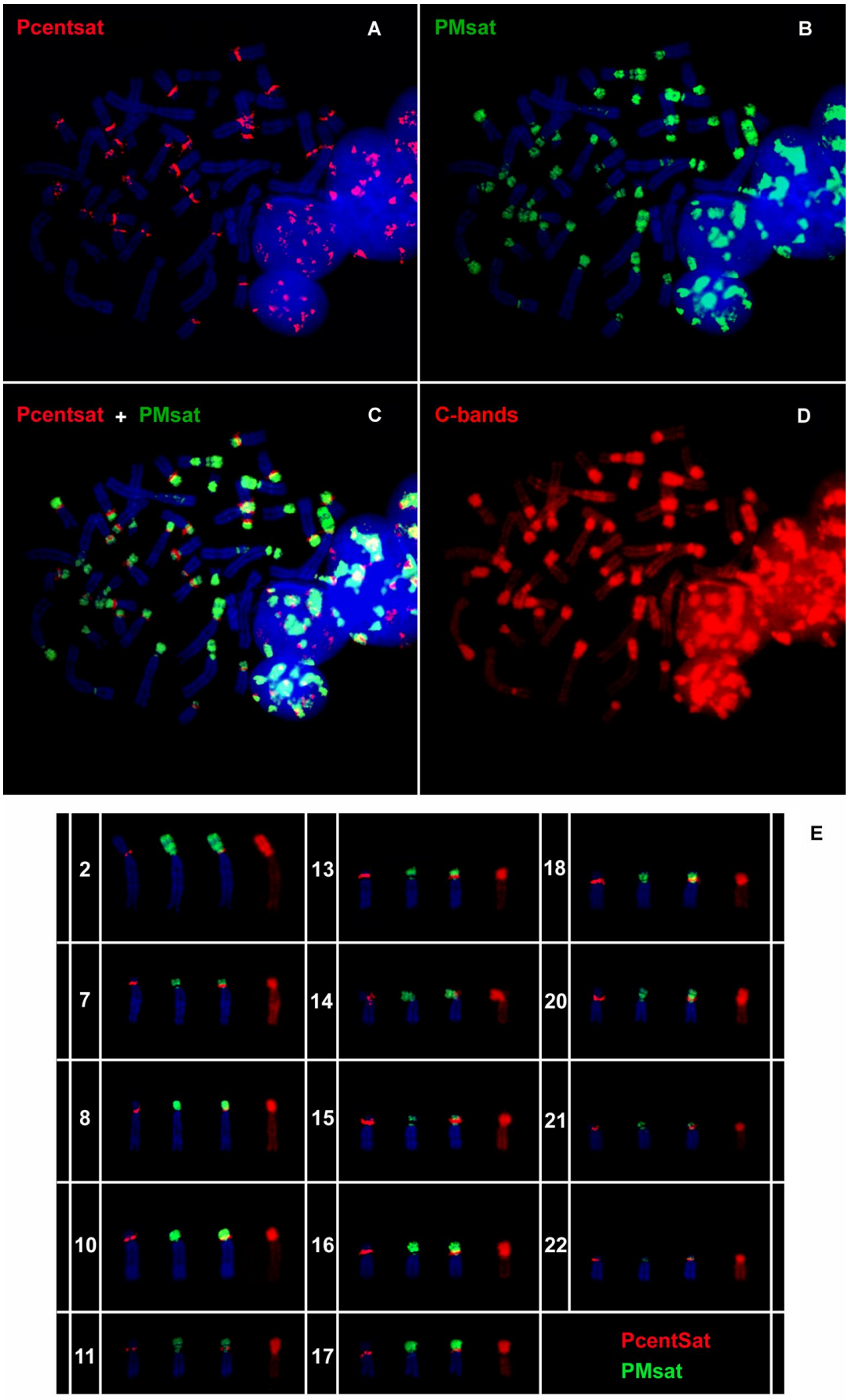
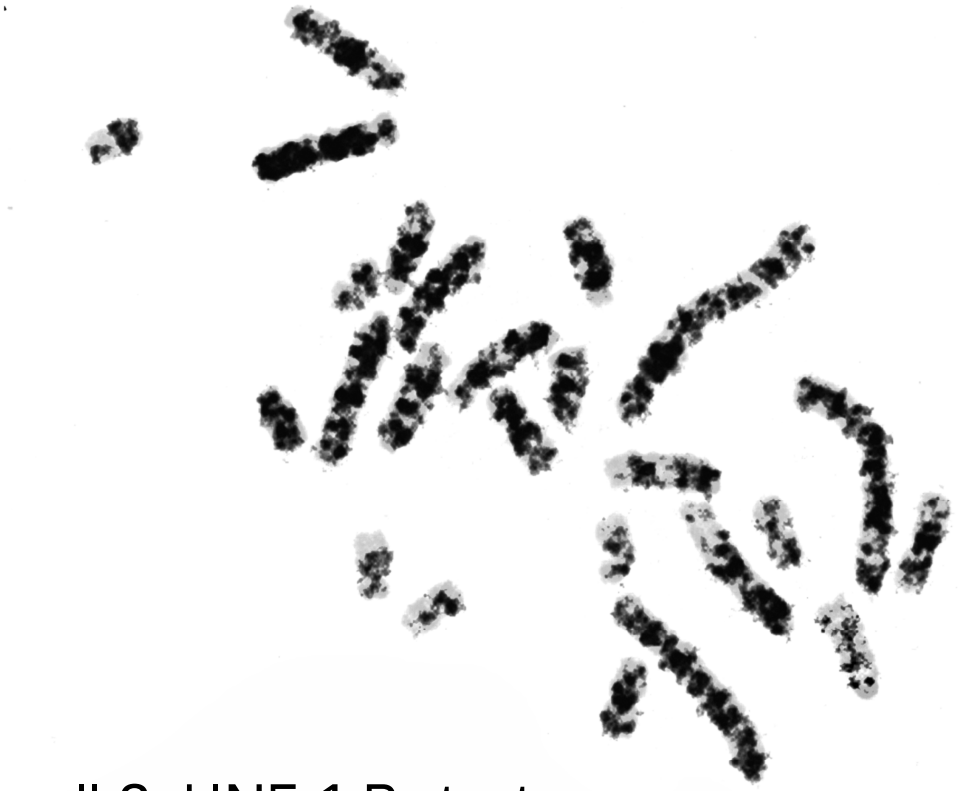

Pcentsat
A
PMsat
B
Pcentsat + PMsat
C
C-bands
D
E
2
13
18
7
14
20
8
15
21
10
16
22
11
17
PcentSat
PMsat

Figure 2



II.3- LINE-1 Retrotransposons

Despite that the occurrence of LINE-1 sequences in mammalian genomes has been traditionally explained by the selfish DNA hypothesis, recently it was argued that these retrotransposons could play important roles in genomes, being implicated in regulation of gene expression, X-chromosome inactivation, progression of early embryogenesis and in genome reorganization (e.g. Lyon 1998, Song and Boissinot 2007, Akagi et al. 2008). To understand better the importance of these highly abundant sequences in mammalian genomes, it was isolated and analyzed, in this work, a fraction of the ORF2 LINE-1 sequence from three rodent species, *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi*. This study resulted in a manuscript, presenting clear evidences about the functions of these sequences in the studied genomes. According to the results obtained, it seems very probable that LINE-1 retrotransposons are involved in control of gene expression, specifically in imprinting of genes and X-chromosome inactivation. Moreover, according to its chromosomal location, these sequences seem also to present a significant role in satDNA evolution.

II.3.1

Line-1 Retrotransposons: from "parasite" sequences to Functional Elements

Journal of Applied Genetics

Line-1 Retrotransposons: from "parasite" sequences to Functional Elements

--Manuscript Draft—

Manuscript Number:	
Full Title:	Line-1 Retrotransposons: from "parasite" sequences to Functional Elements
Article Type:	Original Paper
Section/category:	Animal Genetics
Corresponding Author:	Raquel Chaves Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGB–UTAD) Vila Real, PORTUGAL
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGB–UTAD)
Corresponding Author's Secondary Institution:	
First Author:	Ana Paço, Master
First Author Secondary Information:	
Order of Authors:	Ana Paço, Master Filomena Adega, PhD Raquel Chaves, PhD
Order of Authors Secondary Information:	
Abstract:	Long interspersed nuclear elements-1 (LINE-1) are the most abundant and active retrotransposons in the mammalian genomes. Traditionally, the occurrence of LINE-1 sequences in the genome of mammals has been explained by the selfish DNA hypothesis. Nevertheless, recently it has been also argued that these sequences could play important roles in these genomes, as in the regulation of gene expression, genome modelling and X-chromosome inactivation. The non-random chromosomal distribution is a striking feature of these retroelements that somehow reflects its functionality. In the present study we have isolated and analyzed a fraction of the ORF2 LINE-1 sequence from three rodent species, *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi*. Physical mapping of the isolated sequences revealed an interspersed longitudinal AT pattern of distribution along all the chromosomes of the complement in the three genomes. A detailed analysis show that these sequences are preferentially located in the euchromatic regions, although some signals could be detected in the heterochromatin. In addition, a coincidence between the location of imprinted gene regions (as Xist and Tsix gene regions) and the LINE-1 retroelements was also observed. According to these results, we propose an involvement of LINE-1 sequences in different genomic events as gene imprinting, X-chromosome inactivation and evolution of repetitive sequences located at the heterochromatic regions (e.g. satellite DNA sequences) of the rodents' genomes in analysis.
Suggested Reviewers:	Juan Alberto Marchal Ortega jamaor@ujaen.es Work cited in the manuscript.
	Jiri Rubes rubes@vri.cz Work cited in the manuscript.
	Leopoldo Iannuzzi leopoldo.iannuzzi@ispaam.cnr.it Member of Editorial Board with work in Physical mapping.

Line-1 Retrotransposons: from "parasite" sequences to Functional Elements

Ana Paço, Filomena Adega, Raquel Chaves*

Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro (IBB/CGB–UTAD), 5000-801 Vila Real, Portugal

Telephone: +351 259 350 936, E-mail: rchaves@utad.pt

Abstract

Long interspersed nuclear elements-1 (LINE-1) are the most abundant and active retrotransposons in the mammalian genomes. Traditionally, the occurrence of LINE-1 sequences in the genome of mammals has been explained by the selfish DNA hypothesis. Nevertheless, recently it has been also argued that these sequences could play important roles in these genomes, as in the regulation of gene expression, genome modelling and X-chromosome inactivation. The non-random chromosomal distribution is a striking feature of these retroelements that somehow reflects its functionality. In the present study we have isolated and analyzed a fraction of the ORF2 LINE-1 sequence from three rodent species, *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi*. Physical mapping of the isolated sequences revealed an interspersed longitudinal AT pattern of distribution along all the chromosomes of the complement in the three genomes. A detailed analysis show that these sequences are preferentially located in the euchromatic regions, although some signals could be detected in the heterochromatin. In addition, a coincidence between the location of imprinted gene regions (as Xist and Tsix gene regions) and the LINE-1 retroelements was also observed. According to these results, we propose an involvement of LINE-1 sequences in different genomic events as gene imprinting, X-chromosome inactivation and evolution of repetitive sequences located at the heterochromatic regions (e.g. satellite DNA sequences) of the rodents' genomes in analysis.

Keywords LINE-1 Retrotransposons; Rodentia; Genomic distribution; Repeat sequences functional significance; Imprinting.

Introduction

Long interspersed nuclear elements-1 (LINE-1) are autonomous non-long-terminal repeat retrotransposons that constitute an important component of mammalian genomes, composing about 17 and 19% of the human and mouse genomes, respectively (Lander et al. 2001; Waterston et al. 2002). A complete and transpositionally active LINE-1 sequence is approximately 6 to 7 kb long and contains a 5'-untranslated region (UTR) with an internal promoter, two Open Reading Frames (ORF1, ORF2), a 3'-UTR that presents a G-rich polypurine tract, ending in a A-rich region (e.g. Furano and Usdin 1995; Kazazian 2000; Deininger and Batzer 2002; Moran and Gilbert 2002; Weiner 2002; Lee et al. 2010). It is assumed that both Open Reading Frames encode proteins required for transposition. The role of ORF1 protein is not completely clear as the amino acid sequence predicted for its polypeptide (40 kDa) lacks homology with any known functional protein (see Hohjoh and Singer 1996; Martin 2006), however, it is presently regarded as a non-specific nucleic acid binding protein with nucleic acid chaperone activity (e.g. Martin 2010; An et al. 2011; Dai et al. 2011). By contrast, the role of ORF2 protein is best known, having its 146 kDa multifunctional polypeptide simultaneously endonuclease and reverse transcriptase activities, crucial for retrotransposition (e.g. Dewannieux and Heidmann 2005; Doucet et al. 2010), presenting although a cysteine-rich domain with a still unknown function (Fanning and Singer 1987; Dai et al. 2011).

LINE-1 sequences in mammalian genome have traditionally been referred as selfish elements, persisting over time due to their replicative advantage above the host genome. Nowadays this hypothesis has been argued and a functional meaning, or a more symbiotic significance, has been addressed to LINE-1 sequences, as they play an important role in regulation of gene expression (e.g. Yang et al. 1998; Han and Boeke 2004; Muotri et al. 2007; Akagi et al. 2008), genome modelling (e.g. Boissinot et al. 2006; Song and Boissinot 2007; Kolb et al. 2009; Longo et al. 2009), progression of early embryogenesis (see Vitullo et al. 2011) and in X-chromosome inactivation (see Lyon 1998; Bailey et al. 2000; Lyon 2006). The majority of the proposed functions for LINE-1 are directly related with its non-random distribution in the mammalian genomes. It has been shown for some species (e.g. human and mouse) that these retrotransposons are preferentially located at AT-rich regions, corresponding to chromosome G-bands (see Korenberg and Rykowski 1988; Lander et al. 2001; Waterston et al. 2002), in regions with monoallelically expressed genes (see Allen et al.

2003; Walter et al. 2005) and low levels of recombination (e.g. Boissinot et al. 2001; Graham and Boissinot 2006). A controversial issue is the special accumulation that they seem to exhibit on the X-chromosome when compared with the autosomes of some species, as human and mouse (see Lyon 1998; Bailey et al. 2000; Lyon 2006).

In a general way, these sequences are arguably one of the most significant dynamic forces operating on the mammalian genome, seeming to have great impact on the genesis of genetic diseases as cancer (e.g. Rodić and Burns 2013) and also, on genome evolution (e.g. Martin et al. 2005; Wallace et al. 2008; Lee et al. 2010; Lupski 2010; Martin 2010). LINE-1 retrotransposition can cause shuffling of gene promoters, enhancers and even exons, by the transcription of flanking non-LINE-1 sequences (e.g. Moran et al. 1999; Han and Boeke 2005; Medstrand et al. 2005; Akagi et al. 2008). Gene function can be affected by LINE-1 retrotransposition when the gene coding region is the target site of insertion. However, the ability of these retrotransposons to alter gene expression without interfering with the coding region was also documented (see Landry et al. 2001; Waterston et al. 2002). LINE-1 insertion in the UTR regions of a gene can affect the regulation of its transcription and translation. It has been estimated that more then 18,4% of all known genes in mouse contain transposable elements inside its UTR and/or promoter regions (see Van de Lagemaat et al. 2003). Besides this, transposition events generate homologous sequences in non-homologous regions, providing opportunities for recombination on misaligned chromosomes, what may result in chromosomal rearrangements (e.g. Moran and Gilbert 2002; Boissinot et al. 2006; Song and Boissinot 2007).

The rodents studied in this work belong to the two most specious Muroidea families: the Cricetidae *Cricetus cricetus* (CCR) and *Peromyscus eremicus* (PER) and the Muridae *Praomys tullbergi* (PTU) [NCBI Taxonomy (http://www.ncbi.nlm.nih.gov/sites/entrez) and Tree of Life Web Project (http://www.tolweb.org/tree)]. In the present work, a fraction of the LINE-1 retroelements from these three genomes was isolated and sequenced for the first time. The molecular analysis of these sequences was integrated with the available sequences in NCBI Nucleotide database, allowing its comparison. The genomic distribution analysis of the isolated sequences reinforces the functional significance of LINE-1 in mammal's genomes.

Materials and Methods

Genomic DNA extraction

Genomic DNA from the species under analysis was isolated from fibroblast cell lines which belong to the cell and tissue collection housed at the Department of Systematics and Evolution, of Muséum National d'Histoire Naturelle (MNHN) from Paris (France), using the JETQUICK DNA kit (Genomed).

Isolation, cloning and sequencing of LINE-1 sequences

LINE-1 sequences from *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi* were obtained from genomic DNA by PCR, using specific primers designed from regions conserved between mouse, rat, rabbit and human ORF2 LINE-1 (Dobigny 2002): ORF2 L1F 5'-CCATGCTCATSGATTGG-3' and ORF2 L1R 5'-ATTCTRTTCCATTGGTCTA-3'. The cycling parameters of these PCR reactions were 30 cycles of 94°C for 45s; 52°C for 45s; 72°C for 45s, after a 10 min denaturation at 94°C. Products were visualised on a 1% agarose gel.

The obtained PCR products were then isolated from the ethidium bromide-stained gel and purified using the kit GeneClean® II (QBioGene MP Biomedicals). Resulting DNA fragments were cloned in pCR® 4-TOPO vector (Invitrogen Life Technologies) according to the manufacturer's instructions. The plasmid DNA was isolated using the High Pure Plasmid Isolation kit (Roche Molecular Biochemicals). The selected positive clones were sequenced in both directions using universal primers (M13R 5'- AGCGGATAACAATTTCACACA GG-3' and M13F 5'-CCC AGT CAC GAC GTT GTA AAA CG-3').

Molecular and Phylogenetic analysis of LINE-1 sequences

The BLAST search analysis of the isolated sequences was done using the NCBI Blast database tool (http://www.ncbi.nlm.gov/Blast/). These sequences were also analysed using the software vector NTI advance 11 (Invitrogen Life Technologies) that allowed sequence alignments using the Clustal W algorithm (Thompson et al. 1994). Sequence data from the LINE-1 clones was deposited in the NCBI Nucleotide database with the following accession numbers: HQ386006, HQ386007 and HQ386008. Based on the similarity between the obtained sequences and other available LINE-1 sequences in NCBI database, a dendrogram

was built with the software Mega 4.1 (Beta 3) (Tamura et al. 2007), that uses a Neighbor Joining algorithm (Saitou and Nei 1987).

Physical mapping of LINE-1 sequences

Physical mapping of LINE-1 sequences of the three studied species was carried out by Fluorescent *in situ* Hybridization (FISH). Fixed chromosome preparations from each species were obtained from fibroblast cell lines, by standard cell culture procedures described elsewhere (Chaves et al. 2004). The chromosomes were aged at 65ºC overnight, dehydrated with 100% chilled ethanol, air dried and then denatured in 70% formamide/2×SSC for 1 minute at 70°C. LINE-1 sequences' probes were obtained from LINE-1 clones and labelled with digoxigenin-11-dUTP (Roche, Molecular Biochemicals) by PCR, using the universal primers M13. The probe was denatured at 80°C for 10 min and then snap-cooled on ice. Hybridization was carried out overnight in a moist chamber at 37ºC and the most stringent post-hybridization wash was 50% formamide/2×SSC at 37ºC. Digoxigenin-labelled probes were detected with anti-digoxigenin-5'TAMRA (Roche Molecular Biochemicals). Chromosomes were counterstained with DAPI (Vector Laboratories), producing the DAPI banding specific of AT-rich regions.

CBP-banding sequential to physical mapping of LINE-1 sequences

After distaining the slides, CBP-banding [C-bands by Barium Hydroxide using Propidium Iodide (PI)] was performed according to the standard procedure of Sumner (1972) with slight modifications. Shortly, the slides were submitted to hydrochloric acid (0.1 M) during 20 min, barium hydroxide (5% solution) during 7 min and 2×saline solution citrate (2×SSC: 0.3 mol/L NaCl, 0.03 mol/L sodium citrate) at 60 ºC for 40 min.

Capture and preparation of images

Chromosomes were observed in a Zeiss Axioplan 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); contrast and color optimization were the functions used and affected the whole image equally.

In Silico* Analysis of the imprinted gene regions in chromosomes of *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi

For the imprinted gene regions analysis in the chromosomes of the three studied species, these were first located in the chromosomes of *Mus musculus* accessing the information available in Geneimprint database (http://www.geneimprint.com/site/genes-by-species). Afterwards, imprinted gene regions were annotated in the chromosomes of *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi* using comparative chromosome maps performed by our group (Chaves et al. 2012 and unpublished data), which allowed to identify the syntenic regions between the chromosomes of *Mus musculus* and the three studied species. G-banding was used to increase the resolution of this mapping (Stanyon 2006) allowing a more precise location of the imprinted gene regions.

Results

Sequence analysis data

In this work a fraction of the ORF2 LINE-1 sequence (ORF2 L1) with approximately 300 bp from three rodent species, *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi*, was isolated and sequenced; namely, a part of reverse transcriptase region, considering the specificity of the primers used (Waters et al. 2007). A BLAST search of the isolated sequences allowed its identity confirmation based on the similarity with all available sequences in NCBI Nucleotide database. The *Cricetus cricetus* ORF2 L1 Partial sequence (CCR L1) presents high similarity with a fraction of the ORF2 L1 from the rodents *Microtus arvalis* (above 83%; AY041352), *Nectomys squamipes* (above 80%; AY041425), *Rhipidomys nitela* and *Sigmodon hispidus* (above 78%; AY041598 and AY041608), all belonging to the Cricetidae family. It was also observed great similarity among the CCRL1 and a partial LINE-1 sequence from *Microtus cabrerae* (above 82%; AM041146), also belonging to Cricetidae. The *Peromyscus eremicus* ORF2 L1 Partial sequence (PERL1) shows a high similarity with a ORF2 L1 fraction from *Peromyscus californicus* (above 92% specifically with a part of reverse transcriptase region; U70833), *Peromyscus nudipes*, *Peromyscus maniculatus*, *Peromyscus leucopus* and *Reithrodontomys fulvescens* (above 89%; AY041575, AY041527, AY041520 and AY041635, respectively), all belonging to Cricetidae, and most

of the species belonging to the genus *Peromyscus*. Also, *Praomys tullbergi* ORF2 L1 Partial sequence (PTUL1) revealed a high similarity with sequences present in all the *Mus musculus* chromosomes, with similarity values between 87 and 89%.

Alignments between the three isolated sequences (CCRL1, PERL1 and PTUL1) and LINE-1 sequences of two rodent index species, *Mus musculus* [complete LINE-1 sequence of *Mus musculus* (MMU L1), D84391] and *Rattus norvegicus* [complete ORF2 L1 sequence of *Rattus norvegicus* (RNO ORF2 L1), U83119] were also conducted. The similarity values between each pair of sequences (pairwise alignment) are resumed in table 1. In a general analysis, from the three isolated sequences, the one that presents higher similarity with MMU L1 is PTUL1 (85% similarity). The CCRL1 is the sequence presenting the less similarity value (75%). Identical results were obtained when aligning the three isolated sequences and RNO ORF2 L1 (Table 1). A schematization of the alignments among CCRL1, PERL1 and PTUL1 with the MMU L1 and RNO ORF2 L1 is shown in figure 1. In this figure we can also observe the alignments among LINE-1 sequences of other ten Cricetidae species, available in the NCBI Nucleotide database (chosen according our BLAST search results), with the LINE-1 sequences of the two index genomes. The similarity values among all the analysed sequences are above 75%. Furthermore, these alignments reveal the occurrence of insertions/deletions in the ORF2 L1 sequences of *Reithrodontomys fulvescens* (RFU), *Peromyscus californicus* (PCA) and *Rhipidomys nitela* (RNI), marked in figure 1 with a (*).

Analysis of sequence similarity considering phylogenetic relationships

The alignments between CCRL1, PERL1 and PTUL1 isolated in this work and the other rodent LINE-1 sequences that we obtained from NCBI Nucleotide database (previously referred), allowed to build a dendrogram that correlates the different sequences, and consequently infer the phylogenetic relationships of the respective rodent species. In this dendrogram (Fig. 2) the higher similarity is observed among sequences from species belonging to the Muridae family (PTU, MMU and RNO), and between LINE-1 sequences of Cricetidae species. Besides this, a high similarity is also observed among sequences from species belonging to the same subfamily and to the same genus (phylogenetic classification according to NCBI Taxonomy and Tree of Life Web Project databases).

Physical distribution of the isolated LINE-1 sequences in chromosomes of *Cricetus cricetus, Peromyscus eremicus* and *Praomys tullbergi*

The physical mapping of CCRL1, PERL1 and PTUL1 in the chromosomes of the respective species demonstrated that these sequences are scattered along all the chromosomes of the complement (Fig. 3). In a more detailed analysis, it is also observed that these sequences are non-randomly distributed in the genome of the studied species, displaying a longitudinal differentiation pattern similar to an AT chromosome banding (DAPI like banding). Interestingly, two of these sequences (CCRL1 and PERL1) present a preferential location in some chromosome regions. Notice the accumulation of CCRL1 in both arms of chromosome CCR10 (Fig. 3a, magnified within the circle) and PERL1 in a short arm region of chromosome PER1 (Fig. 3d, within the circle). It is also important to note the greatest amount of CCRL1 and PERL1 in a specific region of *Cricetus cricetus* and *Peromyscus eremicus* X chromosomes, respectively (Figs. 3a and 3d). Differently, PTUL1 seems to present a uniform pattern of distribution in *Praomys tullbergi* chromosomes, without a prevalent accumulation in any chromosome region (Fig. 3g).

When C-banding was performed sequentially to LINE-1 *in situ* hybridization, the distribution euchromatin/heterochromatin of the isolated sequences could be accessed. As follows, PERL1 and PTUL1 seem to present an exclusive euchromatic location (Figs. 3f and 3i). By contrast, CCRL1 is located both in euchromatin as in (peri)centromeric heterochromatic regions. Specifically, in chromosomes CCR1, CCR2, CCR4, CCR5, CCR6 and CCR10, CCRL1 seems to be located only at the boundaries of the (peri)centromeric heterochromatic blocks (Fig. 3c, see in more detail chromosome CCR10 within the circle). Figures 4, 5 and 6 resume the chromosome´s distribution of the isolated LINE-1 sequences relatively to the AT-rich regions and euchromatin/heterochromatin for each of the species in analysis. In a previous work (Paço et al. 2009) the constitutive heterochromatin (CH) of the three studied genomes was detailed characterized using a panel of seven restriction endonucleases followed by C-banding. Furthermore, this methodology allowed the identification of several additional C-bands not evidenced by classical C-banding, cryptic C-bands. The comparison between the location of the cryptic C-bands (brown blocks in Figs. 4, 5 and 6) and the distribution of CCRL1, PERL1 and PTUL1 (pink blocks in the Figs. 4, 5 and 6), highlighted a clear co-localization of some of these C-bands with LINE-1 sequences. See, for instance, chromosomes CCR2, CCR8 (Fig. 4), PER1, PER14 (Fig. 5), PTU3 and PTU10 (Fig. 6). In these same figures (Figs. 4, 5 and 6), imprinted gene regions (blue blocks) were

also identified in the chromosomes of each studied species and interestingly, almost all these imprinted gene regions are co-localized with LINE-1 sequences.

Discussion

Molecular and Comparative analysis of LINE-1 sequences

According to BLAST search and sequence similarity analysis, a fraction (with approximately 300 bp) of the ORF2 LINE-1 sequence (ORF2 L1) from three rodent species, *Cricetus cricetus*, *Peromyscus eremicus* (Cricetidae) and *Praomys tullbergi* (Muridae) was for the first time isolated in the current work. As an example, CCRL1 presents similarity values above 83% with an ORF2 L1 fraction from *Microtus arvalis* and PERL1 similarity values above 92% with an ORF2 L1 fraction from *Peromyscus californicus. Praomys tullbergi* LINE-1 (PTUL1) presents a high similarity with the ORF2 L1 from the Muridae species *Mus musculus* and *Rattus norvegicus* (~ 85% of similarity).

The simultaneous alignment (schematized in Fig. 1) of CCRL1, PERL1, PTUL1 with LINE-1 sequences from 12 other Cricetidae/Muridae species (available in the NCBI Nucleotide database), demonstrated ~75% of sequence conservation for the analysed ORF2 L1 region among these species that diverged at least 17 Million Years ago (Robinson et al. 1997). This is in accordance with the literature, being suggested that this fraction of LINE-1 sequence, namely the LINE-1 reverse transcriptase region, corresponds to a more conserved part of this retrotransposon (Xiong and Eickbush 1990; Casavant et al. 1996), what can be justified by its important role in the retrotransposition (Feng et al. 1996; Casavant et al. 2000; Martin 2006). The evolution of this ORF2 L1 fraction in the different genomes is marked by the occurrence of insertions/deletions, as in *Reithrodontomys fulvescens, Peromyscus californicus* and *Rhipidomys nitela*, evidenced in figure 1 with a (*). The alignment of all these sequences also allowed to build a dendrogram correlating the different sequences and suggesting the phylogenetic relationships of the respective rodent species. In this dendrogram is possible to discern more similarity amongst LINE-1 sequences from rodents belonging to the same family, subfamily and genus (Fig. 2), according to the taxonomic classification presented at the NCBI Taxonomy and Tree of Life Web Project databases.

It is recognized that LINE-1 insertions often fail to include the 5' end of the retrotransposon (promoter region), generating mostly defective copies which remain in the genome without autonomous mobile capacity (transposition inability). However, novel

replication competent LINE-1 sequence variants could also be produced, displaying the ability to pass the new mutations acquired to all subsequent sequence copies, generating a new family of retrotransposons (e.g. Mayorov et al. 1999; Casavant et al. 2000; Boissinot and Furano 2001; Furano et al. 2004). Due to these features, LINE-1 sequences are increasingly being used as phylogenetic characters (see Pascale et al. 1990; Hayward et al. 1997; Kriegs et al. 2006; Waters et al. 2007; Waters et al. 2008), often regarded as homoplasy-free phylogenetic markers (see Batzer et al. 1994; Furano and Usdin 1995; Verneau et al. 1997; Serdobova and Kramerov 1998; Verneau et al. 1998). Nevertheless, some authors mention the need to construct a LINE-1 phylogenetic tree involving a high range of mammal species to legitimate these sequences as valid phylogenetic characters, testing if LINE-1 insertions are indeed unique and not convergent characters (e.g. Waters et al. 2007). The phylogenetic relationships performed in this work, inferred from the comparative analysis of LINE-1, provide additional evidence supporting the use of these sequences as phylogenetic markers.

Genomic distribution of LINE-1 sequences

Physical mapping of CCRL1, PERL1 and PTUL1 demonstrates that these sequences are scattered along all the chromosomes of the complement in each of the three studied species (Fig. 3), presenting a longitudinal pattern similar to a DAPI banding (specific for AT-rich regions). This distribution pattern mimics the observed in other eutherian mammal species, like human (Korenberg and Rykowski 1988; Lander et al. 2001), rabbit (Waters et al. 2004), mouse (Boyle et al. 1990; Waterston et al. 2002), rodents from the genus *Taterillus* (Muridae) (Dobigny et al. 2002) or *Cryptomys* (Bathyergidae) (Deuve et al. 2006), all belonging to the Superorder Euarchontoglires (Murphy et al. 2001; Delsuc et al. 2002). Waters et al. (2004) proposed that this banding pattern observed for LINE-1 sequences is a shared feature only for mammals belonging to the Euarchontoglires Superorder (which includes Rodents, Primates and Lagomorpha), since species belonging to other mammalian Superorders as Laurasiatheria, Xenarthra and Afrotheria do not exhibit such banding pattern (Thomsen and Miller 1996; Parish et al. 2002; Waters et al. 2004). In fact, all species here analysed belong to Euarchontoglires and present that LINE-1 specific distribution. However, a preceding work in Arvicolinae rodents (Acosta et al. 2008) does not support Waters et al. (2004) theory.

In this work we also observed a preferential location for two of the isolated sequences, CCRL1 and PERL1, in different autosomes regions (CCR10 and PER1) and X-chromosome (Figs. 3a and 3d). The PTUL1 presents however a uniform pattern of distribution in *Praomys*

tullbergi chromosomes (regarding our FISH resolution), without a preferential location at any chromosome region (Fig. 3g). Diverse theories have been proposed to explain the preferential location of LINE-1 sequences in some chromosome regions. Several authors suggest a LINE-1 accumulation in regions presenting low levels of recombination (e.g. Boissinot et al. 2001; Graham and Boissinot 2006), being also proposed that regions with monoallelically expressed genes present more density of these retroelements (see Allen et al. 2003; Walter et al. 2005). According to our results, we cannot support the idea of LINE-1 accumulation in regions with low levels of recombination, as we didn't observe a preferential location of these sequences at the Y chromosome of *Peromyscus eremicus* or *Praomys tullbergi* (Figs. 3d and 3g), the chromosome of the complement with the lowest level of recombination. Nevertheless, a closer look to regions with monoallelically expressed genes, particularly imprinted genes, revealed that almost all the imprinted regions are co-localized with LINE-1 sequences in the three studied genomes (Figs. 4, 5 and 6). This finding opens the idea for a possible involvement of LINE-1 retrotransposons in the imprinting of genes that must be, however, further explored in a higher number of species.

A controversial issue is the likely involvement of LINE-1 sequences in X-chromosome inactivation, hypothesized for the first time by Lyon (1998), based in the high accumulations of LINE-1 in human and mouse X-chromosomes, in comparison with the autosomes (Korenberg and Rykowski 1988; Boyle et al. 1990). Yet, in this work, we have observed identical amounts of LINE-1 sequences in autosomes and X-chromosome of *Cricetus cricetus*, *Peromyscus eremicus* (e.g. CCR10 vs CCRX or PER1 vs PERX) and *Praomys tullbergi*. Nevertheless, it is worth mentioning that we identified LINE-1 sequences in the imprinted Xist and Tsix gene regions (X-chromosome) in the three rodents analysed (Figs. 3-5), that have in turn, a primordial role in X-chromosome inactivation. Different studies point to a parallel mechanistic between X-chromosome inactivation and autosomal genes imprinting (reviewed by Reik and Lewis 2005). In accordance and based on our data, it is worth considering the involvement of these retroelements in the imprinting of genes, and also in X-chromosome inactivation, as proposed before by Lyon in 1998, but through another mechanism(s). Whatever the reason(s) behind the accumulation of LINE-1 sequences in some chromosome regions, the fact is that certainly provides some kind of selective advantage to genomes.

The analysis of LINE-1 location/distribution reveals that the three isolated sequences, CCRL1, PERL1 and PTUL1, are preferentially located in the euchromatin (Figs. 4, 5 and 6),

but a heterochromatic coincidence was also observed. The location in euchromatin is in accordance with the involvement of these retrotransposons in control of gene expression (e.g. imprinting of genes as here suggested), being this genomic fraction enriched in unique coding sequences (e.g. Pezer and Ugarković 2008). Regarding the LINE-1 heterochromatic location, diverse works report the location of LINE-1 in heterochromatic blocks of different mammalian genomes (e.g. Mayorov et al. 1996; Waters et al. 2004; Marchal et al. 2006; Acosta et al. 2008), suggesting that these retroelements are probably intermingled with other repeated sequences, forming a complex structure (e.g. Mayorov et al. 1996; Marchal et al. 2006). For some chromosomes from the analysed species this was also observed, what might be related with the evolution of the other repeats, as satellite DNA. Sequences of satellite DNA, the mainly DNA component of heterochromatin, regularly evolve in a concerted way, leading to rapid change between repeats in different genomes, throughout sequence modification, amplification of new variants and intragenomic movements (Ugarković and Plohl 2002; Louzada et al. 2008; Adega et al. 2009; Plohl 2010). Therefore, the heterochromatic location of the three LINE-1 sequences isolated leads us to propose that, when in association with satellite DNA, LINE-1 retrotransposons could enable the amplification and the intragenomic movements of satellite sequences.

Conclusions

In this work, and as far as we know, a fraction of ORF2 LINE-1 sequence from the rodent species, *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi* was isolated and sequenced for the first time. Integrated comparative analysis with other rodent LINE-1 sequences available in the NCBI database, allowed building a dendrogram that phylogenetically relates the considered rodent species. The phylogenetic relationships inferred here are in agreement with the taxonomic classification proposed for Muroidea Rodents, thus supporting the use of LINE-1 sequences as phylogenetic markers. Physical mapping demonstrated that the isolated LINE-1 sequences are scattered along all the chromosomes of the complement in the three genomes, presenting a longitudinal pattern specific to AT-rich regions. After detailed analysis, a coincidence between the location of these three LINE-1 sequences and regions of imprinted genes was observed, presenting also these sequences a location both in euchromatic and heterochromatic regions. According to our results it seems reasonable to consider that LINE-1 sequences are involved in several genomic events, as

imprinting of genes, X-chromosome inactivation and evolution of other repetitive sequences, as satellite DNA sequences.

Acknowledgements

This work was supported by the project POCI/BIA-BCM/58541/2004, and a PhD grant (SFRH/BD/41574/2007) from the Science and Technology Foundation (FCT) from Portugal. We are deeply grateful to Dr. Vitaly Volobouev for providing the cell cultures Rodentia species.

References

Acosta MJ, Marchal JA, Fernández-Espartero CH, Bullejos M, Sánchez A (2008) Retroelements (LINEs and SINEs) in vole genomes: Differential distribution in the constitutive heterochromatin. Chromosome Res (8):1253-1263

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

Akagi K, Li J, Stephens RM, Volfovsky N, Symer DE (2008) Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. Genome Res 18:869–880

Allen E, Horvath S, Tong F, Kraft P, Spiteri E, Riggs AD, Marahrens Y (2003) High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. Proc Natl Acad Sci USA 100 (17):9940–9945

An W, Dai L, Niewiadomska AM, Yetil A, O'Donnell KA, Han JS, Boeke JD (2011) Characterization of a synthetic human LINE-1 retrotransposon ORFeus-Hs. Mobile DNA 2(2):1-12

Bailey JA, Carrel L, Chakravarti A, Eivhler EE (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proc Natl Acad Sci USA 97 (12):6634-6639

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL (1994) African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 91(25):12288–12292

Boissinot S, Furano AV (2001) Adaptative Evolution in LINE-1 retrotransposons. Mol Biol Evol 18(12):2186–2194

Boissinot S, Entezam A, Furano AV (2001) Selection against deleterious LINE-1-containing loci in the human lineage. Mol Biol Evol 18:926–935

Boissinot S, Davis J, Entezam A, Petrov D, Furano AV (2006) Fitness cost of LINE-1 (L1) activity in humans. Proc Natl Acad Sci USA 103:9590–9594

Boyle AL, Ballard SG, Ward DC (1990) Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence *in situ* hybridization. Proc Natl Acad Sci USA 87:7757-7761

Casavant NC, Sherman AN, Wichman HA (1996) Two Persistent LINE-1 in *Peromyscus* Have Unequal Rates of Evolution. Genetics 142:1289-1298

Casavant NC, Scott L, Cantrell MA, Wiggins LE, Baker RJ, Wichman HA (2000) The end of the LINE?: lack of recent L1 activity in a group of South American rodents. Genetics 154:1809-1817

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

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

Dai L, Huang Q, Boeke JD (2011) Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. BMC Biochem 12(18):1-11

Deininger PL, Batzer MA (2002) Mammalian retroelements. Genome Res 12:1455–1465

Delsuc F, Scally M, Madsen O, Stanhope MJ, De Jong WW, Catzeflis FM, Springer MS, Douzery EJ (2002) Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Mol Biol Evol 19:1656–1671

Deuve JL, Bennett NC, O'Brien PCM, Ferguson-Smith MA, Faulkes CG, Britton-Davidian J, and Robinson TJ (2006) Complex evolution of X and Y autosomal translocations in the giant mole-rat, *Cryptomys mechowi* (Bathyergidae). Chromosome Res 14:681–691

Dewannieux M, Heidmann T (2005) LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modelling. Cytogenet Genome Res 110:35-48

Dobigny G (2002) Speciation chromosomique chez les especes ouest-africaines de *Taterillus* (Rodentia, Gerbillinae): implication systematiques et biogeographiques, hypotheses genomiques. Dissertation, Museum National d'Histoire Naturelle

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

Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV, Gilbert N (2010) Characterization of LINE-1 ribonucleoprotein particles. PLoS Genet 6:1-19

Fanning T, Singer M (1987) The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. Nucleic Acids Res 15(5):2251-2260

Feng Q, Moran JV, Kazazian HH, Boeke JD Jr, Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonucleases required for retrotransposition. Cell 86:905–916

Furano AV, Usdin K (1995) DNA Fossils and Phylogenetic Analysis using L1 (LINE-1, Long Interspersed Repeated) DNA to determine the Evolutionary History of Mammals. J Biol Chem 270:25301–24304

Furano AV, Duvernell DD, Boissinot S (2004) L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. Trends Genet 20:9–14

Graham T, Boissinot S (2006) The Genomic Distribution of L1 Elements: the role of Insertions Bias and Natural Selection. J Biomed Biotechnol Volume 2006, Article ID 75327:1-5

Han JS, Boeke JD (2004) A highly active synthetic mammalian retrotransposon. Nature 429:314–318

Han JS, Boeke JD (2005) LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression?. BioEssays 27:775–784

Hayward BE, Zavanelli M, Furano AV (1997) Recombination Creates Novel L1 (LINE-1) Elements in *Rattus norvegicus*. Genetics 146:641-654

Hohjoh H, Singer MF (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J 15:630–639

Kazazian HH Jr (2000) Genetics. L1 retrotransposons shape the mammalian genome. Science 289:1152–1153

Kolb J, Chuzhanova N A, Högel J, Vasquez K M, Cooper D N, Bacolla A, Kehrer-Sawatzki H (2009) Cruciform-forming inverted repeats appear to have mediated many of the microinversions that distinguish the human and chimpanzee genomes. Chromosome Res 17:469–483

Korenberg JR, Rykowski MC (1988) Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. Cell 53:391–400

Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J (2006) Retroposed elements as archives for the evolutionary history of placental mammals. PLoS Biol 4:e91

Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921

Landry JR, Medstrand P, Mager DL (2001) Repetitive elements in the 5' untranslated region of a human zinc-finger gene modulate transcription and translation efficiency. Genomics 76:110–116

Lee SH, Cho SY, Shannon MF, Fan J, Rangasamy D (2010) The Impact of CpG Island on Defining Transcriptional Activation of the Mouse L1 Retrotransposable Elements. PLoS ONE 5(6):1-14

Longo M S, Carone D M, Comparative Sequencing Program NISC, Green E D, O'Neill M J, O'Neill R J (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, 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

Lupski JR (2010) Retrotransposition and Strutural Variation in the Human Genome. Cell 141:1110-1112

Lyon MF (1998) X-Chromosome Inactivation Spreads Itself: Effects in Autosomes. Am J Hum Genet 63:17–19

Lyon MF (2006) Do LINEs Have a Role in X-chromosome Inactication?. J Biomed Biotechnol Volume 2006, Article ID 59746:1-6

Marchal JA, Acosta MJ, Bullejos M, Puerma E, Díaz de la Guardia R, Sánchez A (2006) Distribution of L1-retroposons on giant sex chromosomes of *Microtus cabrerae*

(Arvicolidae, Rodentia): functional and evolutionary implications. Chromosome Res 14:177-186

Martin SL, Li WLP, Furano AV, Boissinot S (2005) The structure of mouse and human L1 elements reflect their insertion mechanism. Cytogenet Genome Res 110:223-228

Martin SL (2006) The ORF1 protein encoded by LINE-1: structure and function during L1 Retrotransposition. J Biomed Biotechnol Volume 2006, Article ID 45621:1-6

Martin SL (2010) Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. RNA Biol 7:706–711

Mayorov VI, Adkison LR, Vorobyeva NV, Khrapov EA, Kholodhov NG, Rogozin IB, Nesterova TB, Protopopov AI, Sablina OV, Graphodatsky AS, Zakian SM (1996) Organization and chromosomal localization of a B1-like containing repeat of *Microtus subarvalis*. Mamm Genome 7:593-597

Mayorov VI, Rogozin IB, Adkison LR (1999) Characterization of several LINE-1 elements in *Microtus kirgisorum*. Mamm Genome 10:724-729

Medstrand P, Van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL (2005) Impact of transposable elements on the evolution of mammalian gene regulation. Cytogenet Genome Res 110:342-352

Moran JV, DeBerardinis RJ, Kazazian HH Jr. (1999) Exon shuffling by L1 retrotransposition. Science 283:1530–1534

Moran JV, Gilbert N (2002) Mammalian LINE-1 retrotransposons and related elements. In:, Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. American Society of Microbiology Press, Washington DC, pp 836-869

Muotri AR, Marchetto MC, Coufal NG, Gage FH (2007) The necessary junk: new functions for transposable elements. Hum Mol Gen 16 (2):R159–R167

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ (2001) Molecular phylogenetics and the origins of placental mammals. Nature 409:614–618

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

Parish DA, Vise P, Wichman HA, Bull JJ, Baker RJ (2002) Distribution of LINEs and other repetitive elements in the karyotype of the bat *Carollia*: implications for X-chromosome inactivation. Cytogenet Genome Res 96:191–197

Pascale E, Valle E, Furano AV (1990) Amplification of an ancestral mammalian LI family of long interspersed repeated DNA occurred just before the murine radiation. Proc Natl Acad Sci USA 87:9481-9485

Pezer Ž, Ugarković Đ (2008) Role of non-coding RNA and heterochromatin in aneuploidy and cancer. Seminars in Cancer Biology 18:123–130

Plohl M (2010) Those mysterious sequences of satellite DNAs. Period Biol 112 (4):403–410

Reik W, Lewis A (2005) Co-evolution of X-chromosome inactivation and imprinting in mammals. Nature Rev Genet (6):403-410

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

Rodić N, Burns KH (2013) Long Interspersed Element-1 (LINE-1): Passenger or Driver in Human Neoplasms?. PLoS Genetics 9 (3):e1003402

Saitou N, Nei M (1987) The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. Mol Biol Evol 4(4):406-425

Serdobova IM, Kramerov DA (1998) Short Retroposons of the B2 Superfamily: Evolution and Application for the Study of Rodent Phylogeny. J Mol Evol 46:202–214

Song M, Boissinot S (2007) Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. Gene 390:206–213

Stanyon (2006) Selected karyotypes. In: O'Brien SJ, Nash WG, Menninger JC (eds) ATLAS of Mammalian chromosomes. John Wiley, Chichester, pp 307

Sumner AT (1972) A simple technique for demonstrating centromeric heterochromatin. Exp Cell Res 75:304-306

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol Biol Evol 24(8):1596–1599

Thomsen PD, Miller JR (1996) Pig genome analysis: differential distribution of SINE and LINE sequences is less pronounced than in the human and mouse genomes. Mamm Genome 7:42– 46

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673-4680

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

Van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003) Transposable elements in mammals promotes regulatory variation and diversification of genes with specialized functions. Trends Genet 19:530–536

Verneau O, Catzeflis F, Furano AV (1997) Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. J Mol Evol 45:424–436

Verneau O, Catzeflis F, Furano AV (1998) Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. Proc Natl Acad Sci USA 95:11284-11289

Vitullo P, Sciamanna L, Baiocchi M, Sinibaldi-Vallebona P, Spadafora C (2011) LINE-1 Retrotransposon Copies Are Amplified During Murine Early Embryo Development. Mol Reprod Dev 2011:1-10

Wallace NA, Belancio VP, Deininger PL (2008) L1 mobile element expression causes multiple types of toxicity. Gene 419:75 – 81

Walter J, Hutter B, Khare T Paulsen M (2005) Repetitive elements in imprinted genes. Cytogenet Genome Res 113:109-115

Waters PD, Dobigny G, Pardini AT, Robinson TJ (2004) LINE-1 distribution in Afrotheria and Xenarthra: implications for understanding the evolution of LINE-1 in eutherian genomes. Chromosoma 113:137–144

Waters PD, Dobigny G, Waddell PJ, Robinson TJ (2007) Evolutionary History of LINE-1 in the Major Clades of Placental Mammals. PLoS ONE 2(1):1-7

Waters PD, Dobigny G, Waddell PJ, Robinson TJ (2008) LINE-1 Elements: Analysis by Fluorescence *In-Situ* Hybridization and Nucleotide Sequences. In: Murphy WJ (ed) Methods in Molecular Biology: Phylogenomics. Humana Press Inc, Totowa, pp 227-237

Waterston RH, Lindblad-Toh K, Birney E et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520-562

Weiner AM (2002) SINEs and LINEs: the art of biting the hand that feeds you. Curr Opin Cell Biol 14:343–350

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362

Yang Z, Boffelli D, Boonmark N, Schwartz K, Lawn R (1998) Apolipoprotein(a) gene enhancer resides within a LINE element. J Biol Chem 273 (2):891–897

Table 1 Evaluation of the similarity values among different LINE-1 sequences

	CCRL1	PERL1	PTUL1	MMUL1	RNO ORF2 L1
CCRL1	100%				
PERL1	83%	100%			
PTUL1	75%	78%	100%		
MMUL1	75%	80%	85%	100%	
RNO ORF2 L1	78%	81%	85%	87%	100%

Figures legends

Fig. 1 Schematic representation of all the alignments performed tends as comparison the complete LINE-1 sequence of *Mus musculus* (MMU L1) and the complete ORF2 L1 sequence of *Rattus norvegicus* (RNO ORF2 L1). The analysis of these alignments allowed to observe the occurrence of insertions/deletions in the alignment region of different LINE-1 sequences (marked with *). The CCRL1, PERL1 and PTUL1 sequences correspond to the *Cricetus cricetus* ORF2 L1 Partial sequence, *Peromyscus eremicus* ORF2 L1 Partial sequence and *Praomys tullbergi* ORF2 L1 Partial sequence isolated in this work. (F) ORF2 LINE-1 foward primer. (R) ORF2 LINE-1 reverse primer. This figure was not made to scale, but on top of all sequences is presented their alignment position (in bp) relative to the sequence MMUL1

Fig. 2 Dendrogram that correlates several LINE-1 sequences belonging to Cricetidae and Muridae species. This dendrogram was built in agreement with the similarity values among the different LINE-1 sequences using a Neighbor Joining algorithm (Saitou and Nei 1987). The CCRL1, PERL1 and PTUL1 sequences correspond to the *Cricetus cricetus* ORF2 L1 Partial sequence, *Peromyscus eremicus* ORF2 L1 Partial sequence and *Praomys tullbergi* ORF2 L1 Partial sequence isolated in this work. The numbers at each node indicate bootstraps values as a percentage of 1000 replicates

Fig. 3 Representative *in situ* hybridization of *Cricetus cricetus* ORF2 L1 Partial sequence (CCRL1), *Peromyscus eremicus* ORF2 L1 Partial sequence (PERL1) and *Praomys tullbergi* ORF2 L1 Partial sequence (PTUL1) in chromosomes of *Cricetus cricetus* (a), *Peromyscus eremicus* (d) and *Praomys tullbergi* (g), respectively (green signals). The same metaphases after C-banding are also present in this figure (b, e and h, respectively). Propidium Iodide was use as counterstaining. The overlapped images of the LINE-1 *in situ* hybridization and the sequential C-banding are present in the third column of the figure (c, f and i, respectively). The most interesting chromosomes are identified

Fig. 4 Genomic distribution of *Cricetus cricetus* ORF2 L1 Partial sequence (CCRL1) in *Cricetus cricetus* chromosome's. The chromosomes of *Cricetus cricetus* present a DAPI banding specific of AT-rich regions. The CCRL1 chromosome distribution was represented by pink and red blocks, correspond the red blocks to the preferential location of this sequence. The heterochromatic bands identified in these chromosomes were represented by black or brown blocks and the imprinted gene regions were represented by blue blocks. Heterochromatic cryptic C-bands (brown blocks) present here were evidenced in a previous work (Paço et al. 2009)

Fig. 5 Genomic distribution of *Peromyscus eremicus* ORF2 L1 Partial sequence (PERL1) in *Peromyscus eremicus* chromosome's. The chromosomes of *Peromyscus eremicus* present a DAPI banding specific of AT-rich regions. The PERL1 chromosome distribution was represented by pink and red blocks, correspond the red blocks to the preferential location of this sequence. The heterochromatic bands identified in these chromosomes were represented by black or brown blocks and the imprinted gene regions were represented by blue blocks. Heterochromatic cryptic C-bands (brown blocks) present here were evidenced in a previous work (Paço et al. 2009)

Fig. 6 Genomic distribution of *Praomys tullbergi* ORF2 L1 Partial sequence (PTUL1) in *Praomys tullbergi* chromosome's. The chromosomes of *Praomys tullbergi* present a DAPI banding specific of AT-rich regions. The PTUL1 chromosome distribution was represented by pink blocks. The heterochromatic bands identified in these chromosomes were represented by black or brown blocks and the imprinted gene regions were represented by blue blocks.

Heterochromatic cryptic C-bands (brown blocks) present here were evidenced in a previous work (Paço et al. 2009).

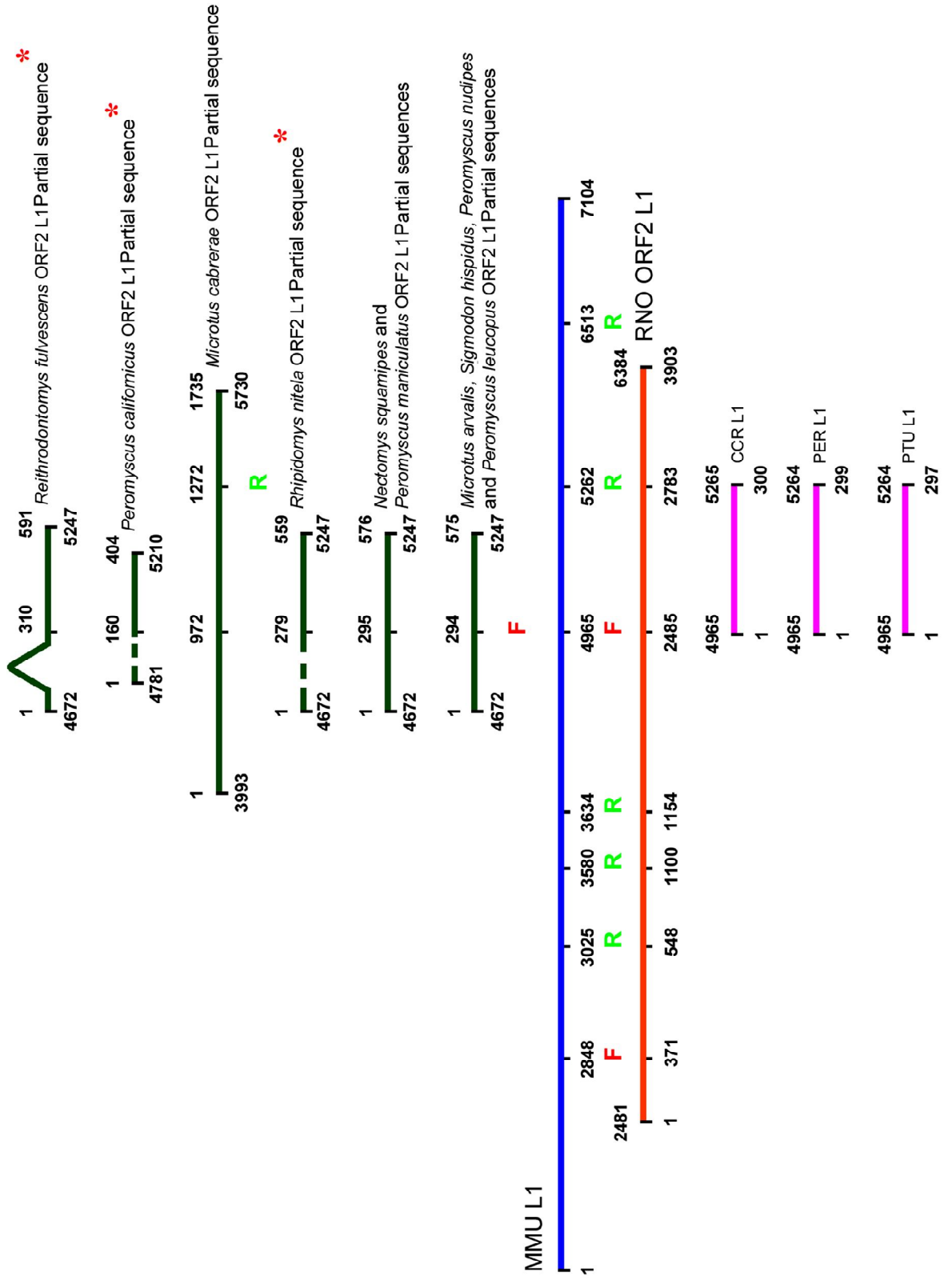


Figure 1

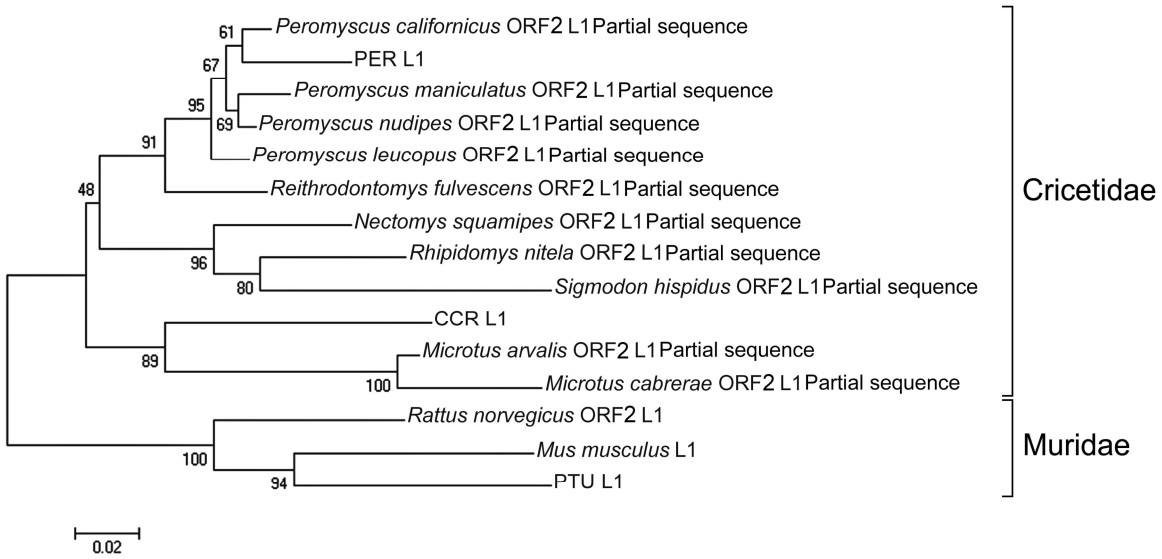

Peromyscus californicus ORF2 L1 Partial sequence
PER L1
Peromyscus maniculatus ORF2 L1 Partial sequence
Peromyscus nudipes ORF2 L1 Partial sequence
Peromyscus leucopus ORF2 L1 Partial sequence
Reithrodontomys fulvescens ORF2 L1 Partial sequence
Nectomys squamipes ORF2 L1 Partial sequence
Rhipidomys nitela ORF2 L1 Partial sequence
Sigmodon hispidus ORF2 L1 Partial sequence
CCR L1
Microtus arvalis ORF2 L1 Partial sequence
Microtus cabrerae ORF2 L1 Partial sequence
Rattus norvegicus ORF2 L1
Mus musculus L1
PTU L1
61
67
95
69
91
48
96
80
89
100
100
94
0.02
Cricetidae
Muridae

Figure 2

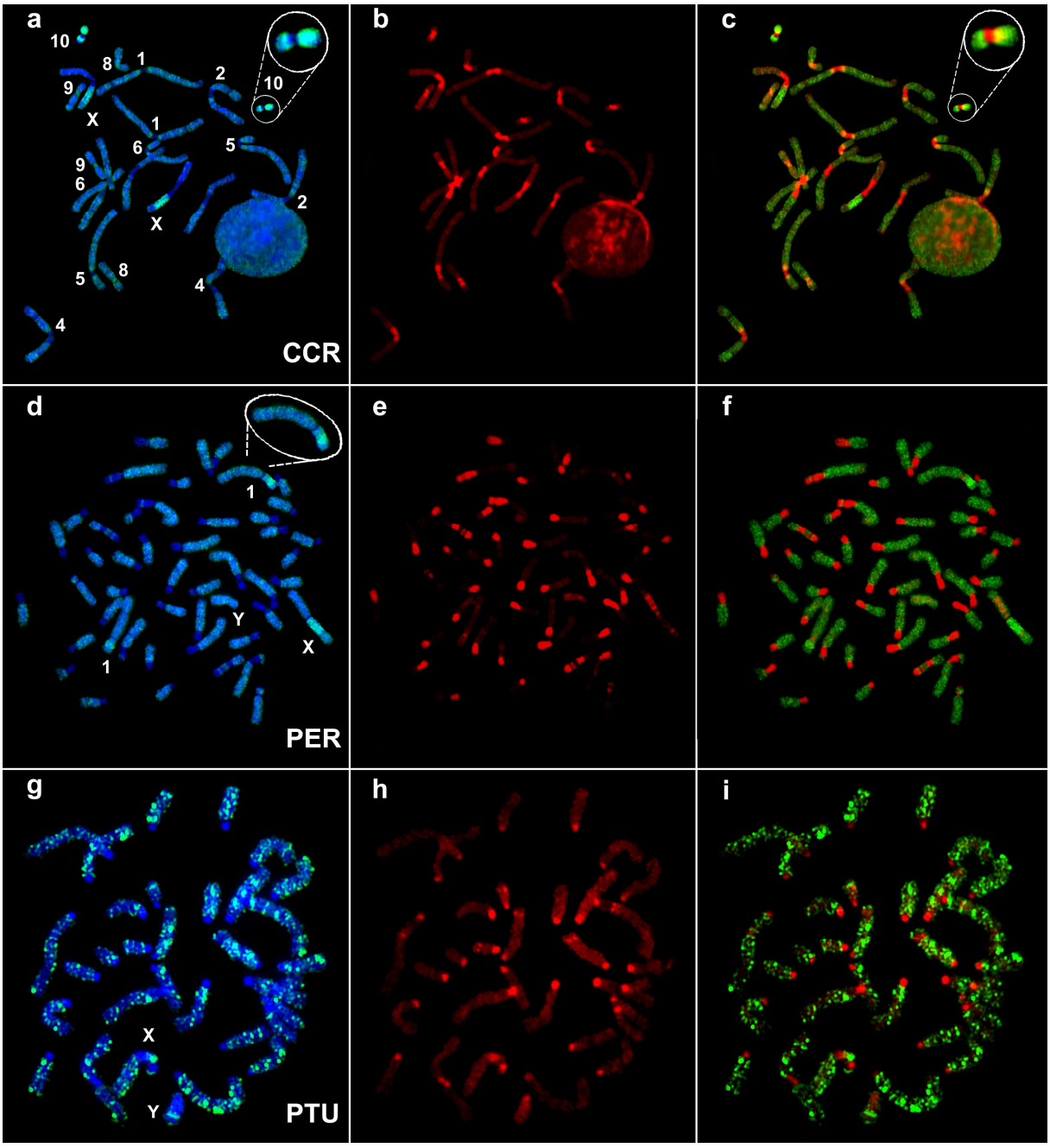


Figure 3

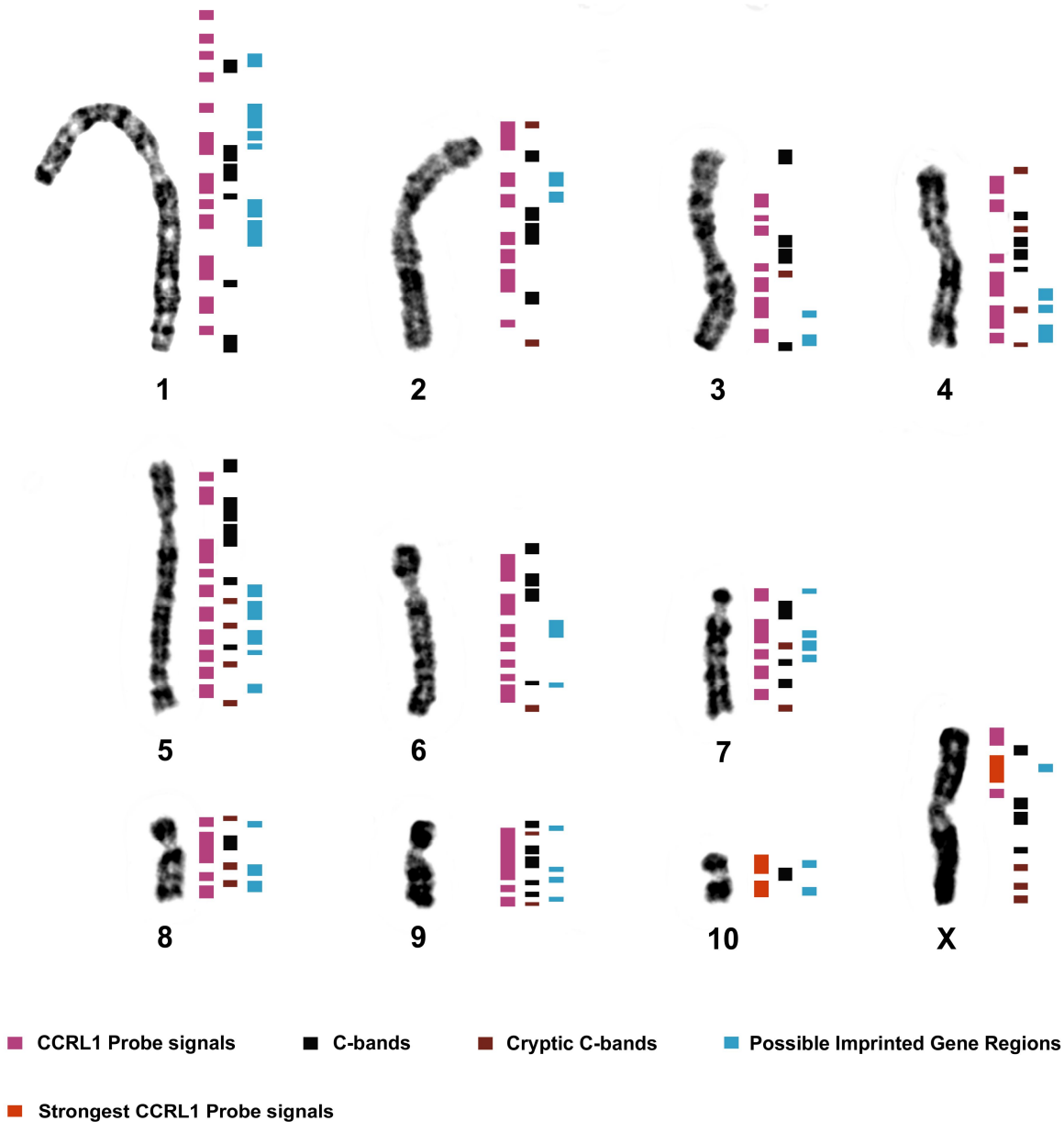

1
2
3
4
5
6
7
8
9
10
X
CCRL1 Probe signals
C-bands
Cryptic C-bands
Possible Imprinted Gene Regions
Strongest CCRL1 Probe signals

Figure 4

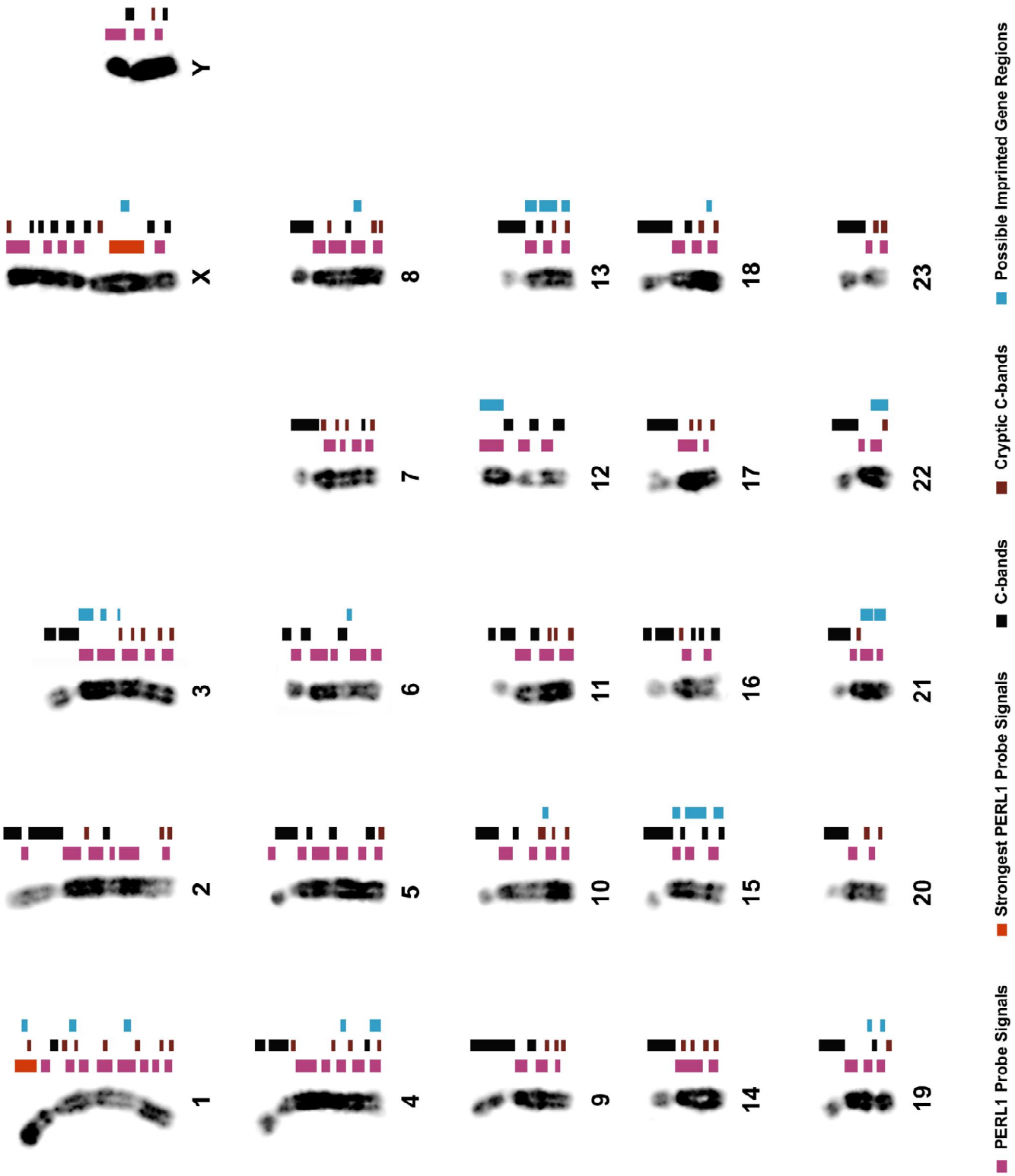

Y
X
8
13
18
23
7
12
17
22
3
6
11
16
21
2
5
10
15
20
1
4
9
14
19
PERL1 Probe Signals
Strongest PERL1 Probe Signals
C-bands
Cryptic C-bands
Possible Imprinted Gene Regions

Figure 5

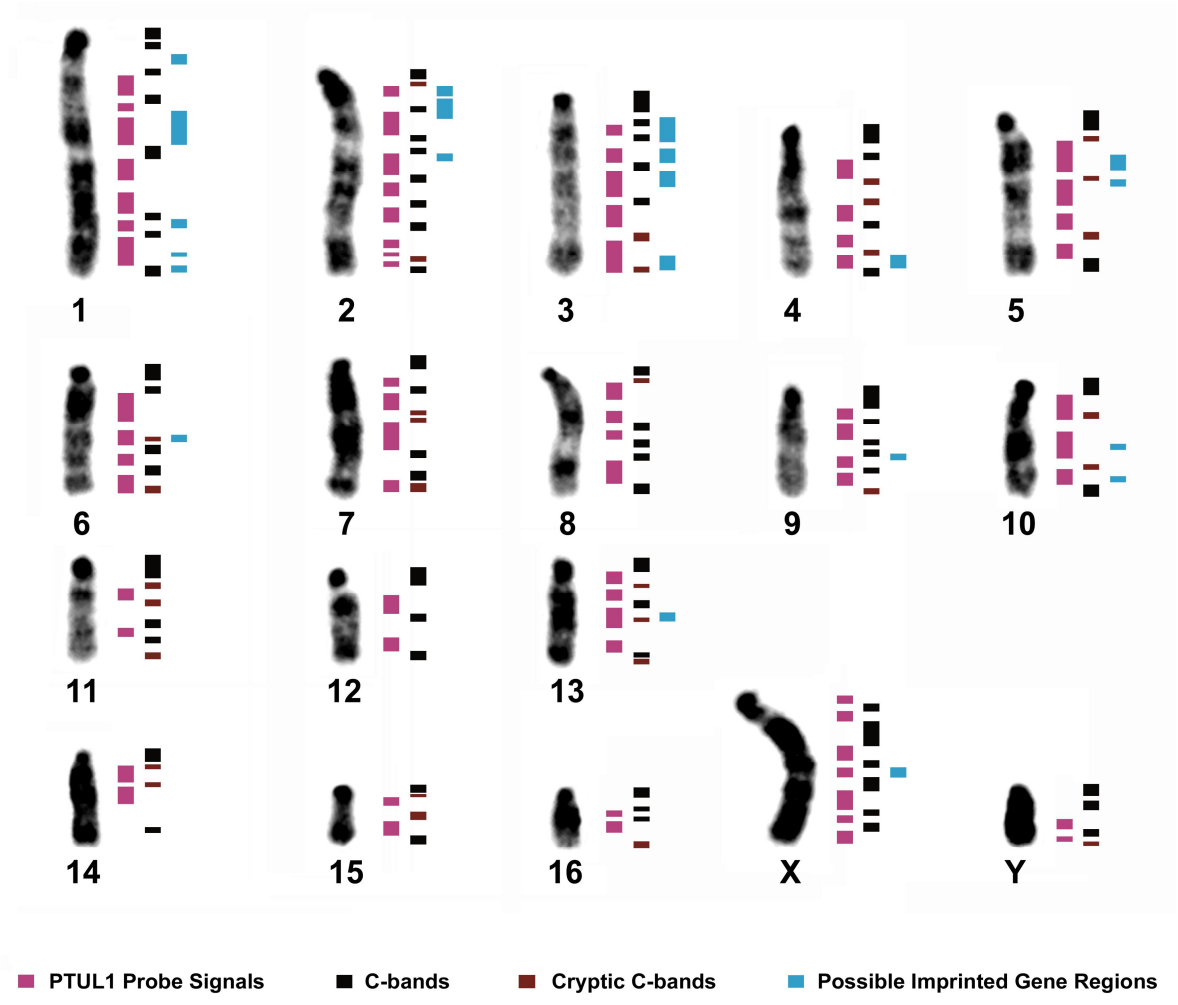

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
X
Y
PTUL1 Probe Signals
C-bands
Cryptic C-bands
Possible Imprinted Gene Regions

Figure 6

Chapter III

General Discussion and Future Perspectives

CHAPTER III. GENERAL DISCUSSION AND FUTURE PERSPECTIVES

III.1- General discussion

With the purpose of contributing to a better knowledge of the repetitive sequences significance in the eukaryotic genomes, particularly its role in genome evolution, here was analysed the repetitive genomic fraction of five Cricetidae and Muridae Rodentia species, regarding tandem and dispersed repeats: Satellite DNAs, Interstitial Telomeric Sequences and LINE-1 retrotransposons. A detailed analysis about the distribution and molecular nature of the Constitutive Heterochromatin of these rodent genomes was also performed

The obtained data allowed to understand how these genomes evolved and to reconstruct the chromosomal evolutionary events elapsed. Evidences for an association between repetitive sequences and the process of chromosome evolution and genome's remodelling were observed. Besides, here was also present data that supports the role of repetitive sequences in many other functions, as in centromeric activity, regulation of gene expression, chromatin remodelation and response to stress.

In this chapter it will be given an overall discussion of all the results obtained, integrating the data about the genomic organization of the different repetitive sequences classes studied in the five genomes. Furthermore, this analysis will correlate the diverse issues in an evolutionary perspective. This thesis section is thus divided in five parts: III.1.1) Chromosomal location and molecular nature of Constitutive Heterochromatin; III.1.2) Characterization and evolution of Satellite DNAs; III.1.3) Telomeric Repeats genomic distribution; III.1.4) LINE-1 retrotransposons genomic distribution and III.1.5) Concluding Remarks.

III.1.1- Chromosomal location and molecular nature of the Constitutive Heterochromatin

In the current work is presented a detailed Constitutive Heterochromatin (CH) description for the five studied species, *Cricetus cricetus*, *Phodopus roborovskii*, *Phodopus sungorus*, *Peromyscus eremicus* and *Praomys tullbergi*, using *in situ* restriction endonuclease (RE) digestion followed by C-banding. This analysis allowed characterizing the CH in terms of its

chromosome location and molecular heterogeneity, as well as the identification of several C-bands impossible to identify by classical C-banding (cryptic C-bands).

In a general overview, according to CH location, three major classes of CH were identified in the five studied species: (peri)centromeric, interstitial and (sub)telomeric. Considering the CH molecular nature, in each of these three CH classes it was recognized a high number of CH subclasses, distinguished by the different restriction patterns when submitted to the same panel of REs. This high molecular heterogeneity suggests the existence of different types of repetitive DNA in these genomes, namely satellite DNA families or variants, possibly originated by a rapid turnover of the repeats.

Comparing the results obtained for the five studied species, it is notorious a distinct distribution of CH as well as a different level of CH heterogeneity in these genomes, what seems to be deeply related with the evolution of their karyotypes. In *Cricetus cricetus* and *Peromyscus eremicus* the CH is preferentially located in the (peri)centromeric regions of the chromosomes (Figures 10a and 10b). *Cricetus cricetus* has an almost entirely meta/submetacentric karyotype, with the majority of these chromosomes exhibiting two very large (peri)centromeric CH blocks, suggesting the occurrence of chromosome fusions and/or heterochromatin additions during the course of this karyotype evolution. This is supported by comparative chromosome painting (Vieira-da-Silva and Louzada et al. *unpublished data*). *Peromyscus eremicus* has a very distinct karyotype that comprises only submetacentric chromosomes, the majority exhibiting large (peri)centromeric C-bands almost extending to the whole short arms (Figure 10b). This suggests the occurrence of heterochromatin amplifications/additions in these short arms, as already proposed by Robbins and Baker (1981) and also recently by Vieira-da-Silva and Louzada et al. (*unpublished data*). Additionally, here was also noticed a great CH heterogeneity in these short arms, which is certainly indicative for the coexistence of different satellite DNA (satDNA) families or variants at these chromosome regions. This is confirmed by recent works, having been identified three satDNA families in these highly heterochromatic chromosomes' short arms (Louzada et al. 2014 *submitted for publication*, Chaves and Adega (*unpublished data*), Vieira-da-Silva et al. (*unpublished data* a)). Also in this work other (peri)centromeric satDNA is described in *Peromyscus eremicus* genome (next section of this chapter). *Phodopus roborovskii* and *Phodopus sungorus* present karyotypes where the majority of the chromosomes are meta/submetacentric, like *Cricetus cricetus*, however the C-bands identified in these karyotypes display (peri)centromeric, interstitial or (sub)telomeric locations, without

a preferential representation (Figure 10c and 10d). The same is observed for the almost completely acrocentric karyotype of *Praomys tulbergi* (Figure 10e). There are now some works suggesting an involvement of the repeats located in CH regions in the chromosomal rearrangements, and consequently in the evolutionary reshaping of karyotypes (e.g. Yunis and Yasmineh 1971, Peacock et al. 1982, John 1988, Chaves et al. 2004), either by promoting chromosome reorganizations (e.g. Wichman et al. 1991, Adega et al. 2009, Zhu and Pao et al. 2011) and/or as fragile regions prone to chromosome breakage (e.g. Ruiz-Herrera et al. 2006). The particular scattered CH distribution presented in *Phodopus roborovskii*, *Phodopus sungorus* and *Praomys tullbergi* chromosomes, points to an association of this genomic fraction with the chromosomal restructurings that modulated these karyotypes, being these repeats the remnants of these events. Also the high CH molecular variability found in these genomes, indicative of an elevated molecular dynamics of the repeats composing these CH

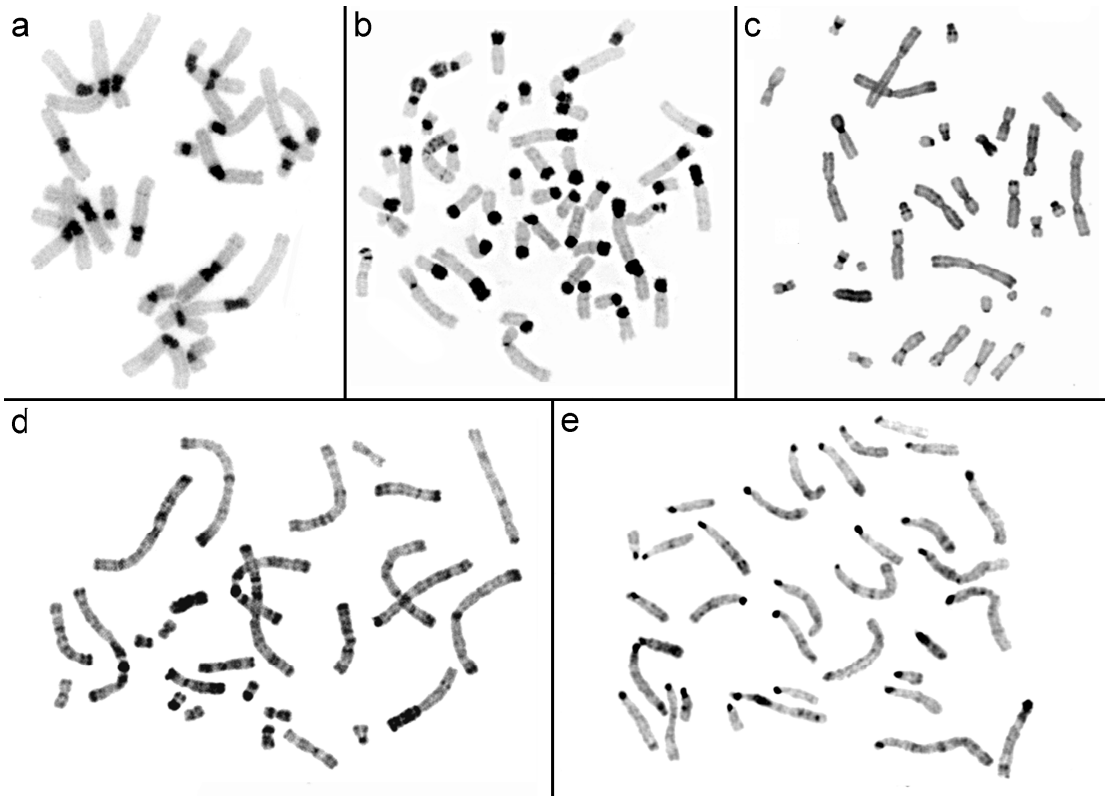


Figure 10- C-banding pattern in chromosomes of *Cricetus cricetus, Peromyscus eremicus, Phodopus roborovskii, Phodopus sungorus* and *Praomys tullbergi*. (a) C-banding pattern in chromosomes of *C. cricetus* (C-banding performed after HaeIII digestion). (b) C-banding pattern in chromosomes of *P. eremicus* (C-banding performed after RsaI digestion). (c) C-banding pattern in chromosomes of *P. roborovskii* (C-banding performed after G-banding - Classical C-banding). (d) C-banding pattern in chromosomes of *P. sungorus* (C-banding performed after DraI digestion). (e) C-banding pattern in chromosomes of *P. tullbergi* (C-banding performed after RsaI digestion).

regions supports the hypothesis of an association CH/karyotype evolution in these species. To assess this probable relationship the CH distribution was compared with the location of evolutionary breakpoint regions in *Phodopus roborovskii* and *Phodopus sungorus* chromosomes. This analysis revealed approximately 95% of coincidence in *Phodopus roborovskii* and 100% in *Phodopus sungorus*. Regarding *Praomys tullbergi* genome, Chaves et al. (2012) also reported a high colocalization of CH with evolutionary breakpoints regions identified in its chromosomes.

Collecting all these evidences, it is possible to point a clear involvement of the repeats located in the CH regions in the chromosomal restructurings that marked the karyotype evolution of the five species, probably representing hotspots for the occurrence of these events.

III.1.2- Characterization and evolution of Satellite DNAs

The important role attributed to satDNAs in the occurrence of chromosomal reorganizations is mainly justified by the dynamic molecular behaviour of these repeats, based in an evolutionary mode mediated through the occurrence of non-homologous chromosomal recombination events, thereby promoting chromosomal reorganization (e.g. Froenicke and Lyons 2008). More recently, however, this role was also suggested by the direct association observed between the increase level of satDNA transcription and genomic instability (Ting et al. 2011, Zhu and Pao et al. 2011). With the objective of investigating these repetitive sequences and their involvement in karyotypes' restructuring, six satDNAs were isolated in this work from the genome of four Cricetidae species, *Cricetus cricetus*, *Peromyscus eremicus, Phodopus roborovskii* and *Phodopus sungorus*.

The first satellite sequence isolated in this work was a (peri)centromeric and almost chromosome-specific satDNA in *Cricetus cricetus* genome (CCR4/10sat), which presents orthologous variants in the *Peromyscus eremicus* genome exhibiting a scattered pattern of distribution in almost all chromosomes of the complement (Figure 11). The most parsimonious scenario for the evolution of these sequences seems to be its "exclusive" presence at the (peri)centromeric regions of an ancestral Cricetidae karyotype, as the observed condition in *Cricetus cricetus* chromosomes, being afterwards these sequences moved throughout the genome, assuming a scattered pattern in the genome of *Peromyscus eremicus*. This CCR4/10sat repositioning and amplification in *Peromyscus eremicus* chromosomes may have resulted from intragenomic movements of these sequences, which may occur by several recombinational mechanisms of unequal transfer (e.g. unequal crossing-over and rolling circle amplification), that can also be accompanied by transposon mediated exchanges. These events certainly facilitated the occurrence of chromosome rearrangements, since it allows that homologous sequences in non-homologous regions recombine. Therefore, we strongly believe that some of the chromosomal rearrangements that occurred during *Peromyscus eremicus* karyotype restructuring might have been a consequence of the CCR4/10sat molecular dynamics.

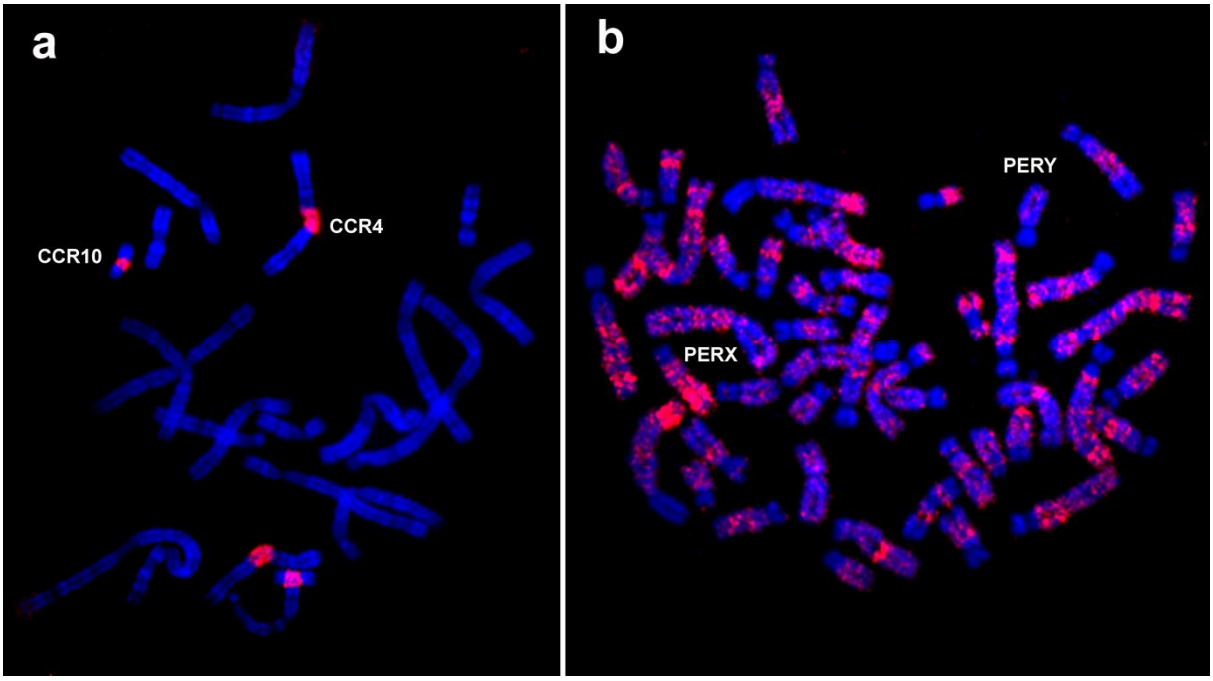


Figure 11- Representative *in situ* hybridization of CCR4/10sat sequences on *Cricetus cricetus* and *Peromyscus eremicus* chromosomes. (a) FISH with CCR4/10sat sequences (red signals) onto chromosomes of *C. cricetus.* CCR4 - *C. cricetus* chromosome 4. CCR10 - *C. cricetus* chromosome 10. The chromosomes were counterstained with DAPI (blue). (b) FISH with CCR4/10sat sequences (red signals) onto chromosomes of *P. eremicus.*

In addition to CCR4/10sat, from the satDNAs isolated here, two other were identified in the *Peromyscus eremicus* genome, PSUcentSat and PERcentSat, however revealing very distinct evolutionary pathways in comparison to CCR4/10sat. Indeed, PSUcentSat seems to follow a completely opposite evolutionary course, since this sequence certainly presented a dispersed chromosomal distribution in the chromosomes of an ancestral Muridae/Cricetidae genome, as observed in *Peromyscus eremicus* and in the Muridae species *Rattus norvegicus*, but a restricted (peri)centromeric location in only a few *Phodopus sungorus* chromosomes. This PSUcentSat distribution in *Phodopus sungorus* genome occurred most probably by selective amplification of this satellite in some chromosome (peri)centromeric regions after the divergence of these species, together with the decrease and elimination of dispersed PSUcentSat copies that most probably were also initially present in the chromosomes of *Phodopus sungorus*. Regarding PERcentSat, orthologous sequences of this satellite could be found in phylogenetically distant species, which are included (at least) in the Orders of Primates and Rodents, presenting almost exclusively a centromeric location. Contrary to what

was described for CCR4/10sat and PSUcentSat, the centromeric location of PERcentSat is highly conserved during evolution, with at least ~ 91,9 My, according to OneZoom Tree of Life Explorer database (http://www.onezoom.org/index.htm, updated December 2013), what points to an important role of this satellite in the different genomes, most probably presenting a centromeric function, due to its conserved location.

From the genome of the two *Phodopus* species studied in this work (*Phodopus roborovskii* and *Phodopus sungorus*) were also isolated more three repeats with a narrower phylogenetic location: a species-specific satDNA in *Phodopus roborovskii* (PROsat), a repeat shared by both *Phodopus* genomes (PsatDNA) and a species-specific and also chromosome-specific repeat of *Phodopus sungorus* (PSUchr1sat). These three repeats present an intricate genome organization with PROsat intermingling with PsatDNA in *Phodopus roborovskii* chromosomes, happening the same with PSUchr1sat and PsatDNA in *Phodopus sungorus* genome. The origin of these sequences seems to be recent, because they are only present within the *Phodopus* genus, being most probably PROsat and PsatDNA originally chromosome-specific repeats that were spread afterwards to other chromosomes. In fact, PsatDNA is currently still chromosome-specific in *Phodopus sungorus* genome. The amplification and distribution/spread of PROsat and PsaDNA from one ancestral chromosome to some *Phodopus roborovskii* chromosome pairs, can be explained by non-homologous recombination events between homologous regions in different chromosomes, as unequal crossing-over. The study of the evolutionary path of these three repeats allowed inferring the ancestral chromosomal forms of *Phodopus*, having been verified that the breakpoints for the rearrangements that modulate these ancestral chromosomes occurred in PROsat blocks.

Despite the dynamic molecular evolution of these sequences might result in the occurrence of chromosomal rearrangements, many other roles are assigned to satDNAs (e.g. Richard et al. 2008, Plohl 2010, Zhu and Pao et al. 2011, Hall et al. 2012, Enukashvily and Ponomartsev 2013). In this work, an analysis more associated with the transcriptional activity of these sequences, particularly for PSUcentSat and PERcentSat, increased the evidences presented in different works about the important role of satDNAs in the control of gene expression, chromatin remodelation, centromeric activity and in the response to stress (reviewed in Enukashvily and Ponomartsev 2013). Briefly, the significantly higher amount of PSUcentSat in small RNA relatively to total RNA in *Peromyscus eremicus* genome, allows suggesting the possibility that this satellite transcripts may result in siRNAs. These RNAs are recognized for having an important role in chromatin remodelation, leading to heterochromatin formation

and maintenance, and in the control of gene expression (reviewed in Vourc'h and Biamonti 2011, Enukashvily and Ponomartsev 2013). The PERcentSat possibly may also have a role in control of gene expression. The similarity identified by NCBI blast analysis between this satellite repeat units (with 21 bp) and a small region of two genes mRNAs, gene for probable E3 ubiquitin-protein ligase HECTD2 and for the fibroblast growth factor receptor of *Ciona intestinalis*, suggests an involvement of the PERcentSat transcripts in the control of these genes expression, by siRNAs. Similar idea was presented by Li and Kirby (2003) for human and chicken alpha-satellite transcripts.

The similarity found between PERcentSat and a *Nannospalax galili* mRNA (NCBI blast analysis), belonging to cDNA pools from muscle and brain tissues isolated from animals exposed to hypoxic and normoxic conditions (Malik et al. 2011), is in agreement with PERcentSat transcription by response to stress environment stimulus. The identification in PERcentSat of binding sites for transcription factors that promote transcription in response to stress stimulus, as the Heat Shock Factor 1 (HSF1) and the CRE-Binding protein 1/c-Jun heterodimer (CRE-BP1/c-Jun), as well corroborate the involvement of PERcentSat in the cellular response to stress. Interestingly it was also identified different Open Reading Frames in PERcentSat, whose putative polypeptides present similarity with hypothetical, predicted or already characterized, proteins. One of these proteins is a cell surface receptor and also this result is in agreement with the PERcentSat transcription in response to stress. Other protein is the microtubule-associated protein 1A that regulates the assembly of microtubules by binding tubulin subunits (Maccioni and Cambiazo 1995). This sustains a centromeric function for PERcentSat, already suggested by the conservation of its (peri)centromeric location during at least ~91.9My, probably operating in chromosome segregation by affecting microtubule assembly.

Making a general analysis of the results obtained for the different satDNAs isolated, the most obvious conclusion that can be drawn is that they present different evolutionary stories and a distinct phylogenetic dispersion. Some are shared by various species (orthologous sequence) that belongs to the same or to a very different taxonomic group (within or out Rodentia Order), but in contrast others are restricted to a single genome or even a pair of chromosomes. Therefore, the results obtained here give a good picture of the complex molecular behavior of satDNAs, which is certainly associated with the role(s) played by these sequences in the genomes.

III.1.3- Telomeric repeats genomic distribution

In addition to its characteristic terminal position, blocks of telomeric repeats can also be found at interstitial and (peri)centromeric regions of the chromosomes, known as interstitial telomeric sequences (ITSs) (e.g. Meyne et al. 1990, Liu and Fredga 1999). Commonly, the origin of ITSs is explained by the occurrence of chromosomal rearrangements as fusions (e.g. Slijepcevic 1998, Li et al. 2000) or pericentric inversions (Rovatsos et al. 2011), corresponding these repeat blocks to "scars" or remnants of these reorganizations. Therefore, to get a better look on the chromosomal evolution of the five studied species, we have also investigated the telomeric repeats genomic distribution in these karyotypes.

Assembling the results obtained here and in other works (Multani et al. 2001, Meles et al. 2007), it is observed that three of the five studied species present ITSs, namely *Phodopus roborovskii*, *Phodopus sungorus* and *Praomys tullbergi*. Regarding specifically the ITS pattern of *Phodopus* species (obtained in this work, Figure 12), and the chromosomal syntenies of *Phodopus* chromosomes with hypothetical ancestral karyotypes, it is possible to suggest that some of these ITS blocks have its origin in robertsonian-like fusions and pericentric inversions. In *Praomys tullbergi* genome, most of the ITSs origin was attributed to tandem fusions (Meles et al. 2007). However in these three genomes, not all the ITSs can be

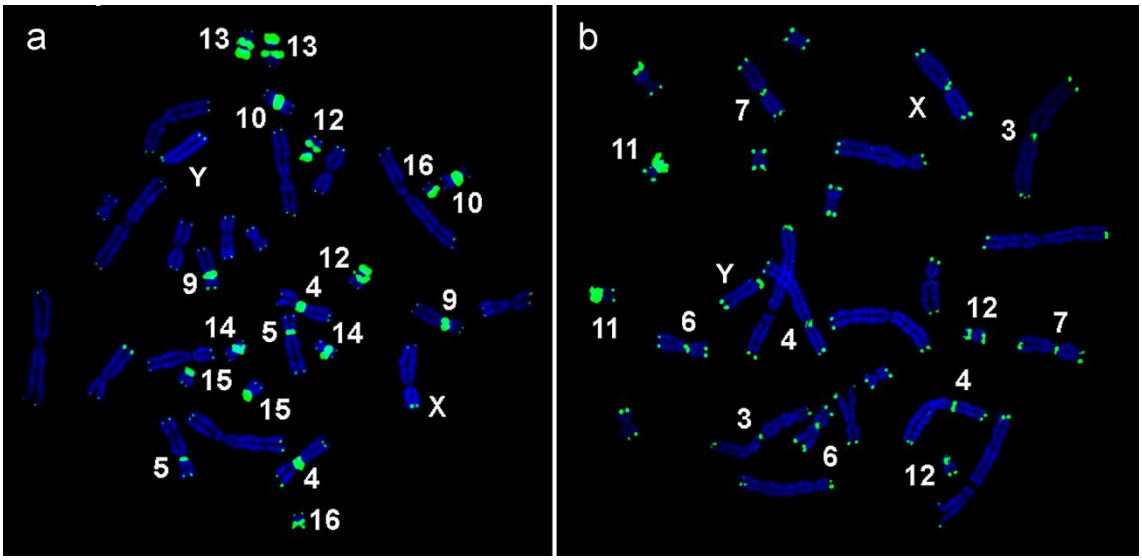


Figure 12- Representative *in situ* hybridization of telomeric repeats (TTAGGG)n on *Phodopus roborovskii,* and *P. sungorus* chromosomes. (a) FISH with telomeric sequences (green signals) on chromosomes of *P. roborovskii*. The chromosomes were counterstained with DAPI (blue). (b) FISH with telomeric sequences (green signals) on chromosomes of *P. sungorus.*

justified by chromosomal rearrangements. Relocation of telomeric repeats is the most parsimonious scenario for the location of some of these ITSs. As it happens for satDNAs, recombinational mechanisms, such unequal crossing-over can be involved in the amplification and relocation of ITS blocks. These molecular events certainly occur independently in the different genomes, which can explain the different ITS patterns observed even in close related species, as *Phodopus roborovskii* and *Phodopus sungorus* (Figure 12). Integration of telomeric repeats within the copies of a satellite sequence, as it was observed for the major satDNA family of *Felis catus*, FA-SAT (Fanning 1987, Santos et al. 2004), was also proposed to explain the location of some ITSs whose origin could not be ascribed to chromosomal reorganizations. The amplification and relocation of this kind of satDNAs may also have led to the amplification and relocation of the telomeric sequences.

Curiously, ITSs are not simply related with chromosomal evolution as a consequence of chromosomal rearrangements, but also as a cause of these reorganizations. Several studies point these repeats as unstable regions, representing hotspots for chromosome fusions/fissions, tandem fusions or inversions (e.g. Ashley and Ward 1993, Nanda et al. 2002, Farré et al. 2009). This was particularly proposed for large ITSs blocks, due to their propensity to form secondary structures, creating the opportunity for recombination events that may induce breakage (Balajee et al. 1994, Fernández et al. 1995, Nergadze et al. 2007), and consequently chromosome rearrangements. In accordance, the large ITS blocks observed in the two *Phodopus* species studied here, particularly in *Phodopus roborovskii* chromosomes, might represent unstable regions that were involved in some of the rearrangements that modulate these karyotypes during evolution. For example, the large ITS blocks that occupy the entire short arm of some small *Phodopus* chromosomes (namely PRO12, PRO13, PRO15, PRO16 and PSU11), from the (peri)centromeric region to the telomere border, could easily be the result of telomeric repeats amplification from the telomeric domain (e.g. by unequal crossing-over). However, we cannot discard the hypothesis of these large ITS blocks have a more internal location in ancestral chromosomal forms, being its origin attributed to chromosomal rearrangements and in turn, due to its high molecular dynamics, be the cause of subsequent chromosomal fissions. As a consequence of that, these repeats blocks are now present in some of the more small acrocentric chromosomes of these karyotypes.

III.1.4- LINE-1 retrotransposons genomic distribution

Currently, it is generally accepted that LINE-1 retrotransposons have a great impact in genome evolution, acting directly as a source of chromosomal rearrangements (e.g. Song and Boissinot 2007); presenting as well an ability to alter gene expression (e.g. Muotri et al. 2007, Akagi et al. 2008); an participation in the mechanism of mammalian X-chromosome inactivation (e.g. Lyon 2006) and a probable involvement in the progression of early embryogenesis (e.g. Vitullo et al. 2011). Here it was analysed the LINE-1 genomic distribution in the genomes of the rodents *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi*. Specifically, it was isolated, sequenced and mapped physically in the chromosomes a fraction of the ORF2 LINE-1 sequence from the referred species.

As stated previously, different are the works implicating transposable elements at different stages of satellite DNA evolution, namely in its origin, homogenization and genome dispersion (e.g. Kapitonov et al. 1998, Cheng and Murata 2003, López-Flores et al. 2004, Macas et al. 2009). Examples of a LINE-1 role in satDNA evolution can also be suggested for two of the satellites isolated and analysed in this work, CCR4/10sat and PROsat. We believe that before the acquisition of a dispersed location, CCR4/10sat probably existed in the genome of *Peromyscus eremicus* intermingled or in proximity with LINE-1 sequences, as currently observed in the genome of *Cricetus cricetus* (the pericentromeric region of chromosome CCR10 presents a co-localization of CCR4/10sat sequences and a high accumulation of LINE-1 retrotransposons). This distribution scenario can easily explain the scattered CCR4/10sat pattern observed now in *Peromyscus eremicus* genome, where the initial intragenomic movements of these sequences may have been promoted by LINE-1 retrotransposition, allowing the transduction of non-LINE-1 DNA to new genomic locations (e.g. Cordaux and Batzer 2009). Following recombinational events enabled the amplification and additional dispersion of CCR4/10sat in the *Peromyscus eremicus* genome. Other example of the LINE-1 implication in satDNA evolution is observed in PROsat. The high similarity observed between a region of PROsat repeat unit and a part of the LINE-1 retrotransposon, suggests a possible origin of PROsat from LINE-1 elements. Molecular mechanisms as unequal crossing-over involving homologous dispersed LINE-1 sequences, could have initiated the tandem duplications that later originated the arrays of PROsat repeats. Subsequent mutational changes and sequence homogenization (concerted evolution) can justify the finding of only a small conserved region between PROsat and LINE-1,

corresponding to a small part of the ORF2 region. Nevertheless, looking for the participation of LINE-1 in PROsat evolution, it is also alternatively possible that LINE-1 could have been simply inserted within a PROsat repeat unit by retrotransposition (perhaps being inserted a shorter truncated LINE-1 element), and during the evolution of this satellite, the LINE-1 fragments were maintained and spread to all the repeats in the arrays by sequence homogenization mechanisms.

Additionally, an analysis about the preferential location of LINE-1 sequences in chromosome regions with monoallelically expressed genes, namely imprinted genes, was performed in this work. Interestingly, almost all the imprinted regions identified in *Cricetus cricetus*, *Peromyscus eremicus* and *Praomys tullbergi* chromosomes present LINE-1 sequences. According to these results, it seems probable the involvement of LINE-1 elements in the imprinting of these genes. Two of these imprinted gene regions correspond to the *Xist* and *Tsix* gene regions, having these a primordial role in X-chromosome inactivation. Therefore, as LINE-1 retroelements have a likely role in the imprinting of *Xist* and *Tsix* genes of the analysed species, consequently they are also implicated in the complex mechanism of X-chromosome inactivation, as initially proposed by Lyon (1998), but through another mechanism.

LINE-1 sequences are increasingly used as phylogenetic characters (see Pascale et al. 1990, Hayward et al. 1997, Kriegs et al. 2006, Waters et al. 2007, 2008), being regarded by diverse authors as homoplasy-free phylogenetic markers (see Batzer et al. 1994, Furano and Usdin 1995, Verneau et al. 1998, Serdobova and Kramerov 1998). Nevertheless, this is not yet a general idea and some reservations persist about this issue, requiring the construction of a LINE-1 phylogenetic tree involving a high range of mammal species to legitimate these sequences as valid phylogenetic characters (e.g. Waters et al. 2008). Also during this work, the LINE-1 sequences obtained were aligned with several others LINE-1 elements from Cricetidae and Muridae species (available in the NCBI Nucleotide Database), what allowed to build a dendrogram that correlates the different sequences and consequently infer the phylogenetic relationships of the respective rodent species. This phylogenetic analysis is in agreement with the taxonomic classification proposed for these species (Tree of Life Web Project), which increases the evidences that support the use of LINE-1 sequences as phylogenetic markers.

As conclusion of this point of the work, it can be mentioned that the results obtained here strongly support some of the previously suggested roles for LINE-1 retrotransposons, being

these sequences certainly involved in the imprinting of genes, X-chromosome inactivation and satDNA evolution of the analysed species. The data achieved in this work also support the use of LINE-1 sequences as valid phylogenetic markers.

III.1.5- Concluding remarks

The integration of all the data discussed here clearly points to a role of repetitive sequences in the evolution of the five studied genomes, being unquestionable that these sequences are implicated in the occurrence of chromosomal rearrangements. The repetitive nature *per se* of every class of repeats studied, satellite DNAs, ITSs and LINE-1, favours recombinational events between homologous sequences in non-homologous regions, which may culminate in chromosomal rearrangements. However, this role in the origin of chromosomal reorganizations is particularly proposed for satDNAs, because of its characteristic evolutionary mode (concerted evolution) generally marked by rapid sequence variations, copy number alterations and/or intragenomic movements (high molecular dynamic behaviour), driven by different recombinational events as unequal crossing-over or rolling circle replication/reinsertion.

In figure 13 are summarized all the results obtained in this work, regarding Constitutive Heterochromatin (CH) location and molecular heterogeneity; Satellite DNA (satDNA) characterization and distribution; Interstitial Telomeric Sequences (ITSs) chromosomal location and LINE-1 Retrotransposons genomic distribution in the five studied species, *Cricetus cricetus, Peromyscus eremicus, Phodopus roborovskii, Phodopus sungorus* and *Praomys tullbergi*. These data is also complemented with information from other works. Considering only the CH analysis performed here, it could be suggested that *Phodopus roborovskii, Phodopus sungorus* and *Praomys tullbergi* karyotypes are more derivative compared to the ones of *Cricetus cricetus* and *Peromyscus eremicus,* being originated by a higher number of chromosomal rearrangements. This is supported by the scattered distribution and the higher molecular dynamics of the repeats located in Constitutive Heterochromatin regions (higher CH molecular heterogeneity) of *Phodopus roborovskii, Phodopus sungorus* and *Praomys tullbergi* chromosomes (Figure 13). The availability of Comparative Chromosome Painting (CCP) data for all the five studied species (Romanenko et al. 2007, Chaves et al. 2012, Vieira-da-Silva and Louzada et al. *unpublished data*), increased the knowledge about the evolutionary paths that these karyotypes followed. CCP data indicate, in conformity with the CH analysis, that *Phodopus roborovskii* and *Phodopus sungorus* have more derivative karyotypes, in comparison with *Cricetus cricetus* and *Peromyscus eremicus*. The *Phodopus roborovskii* karyotype evolved from Muroidea ancestral by at least 29 rearrangements, the *Phodopus sungorus* karyotype by 32 (Romanenko et al.

2007), the *Cricetus cricetus* by 28 and *Peromyscus eremicus* by six rearrangements (Vieira-da-Silva and Louzada et al. *unpublished data*). But, contrarily to what was expected from the CH analysis (Figure 13), *Cricetus cricetus* seems to display a more derivative karyotype in comparison to the one of *Praomys tullbergi*, that evolved from the Muroidea ancestral by 23 rearrangements (Chaves et al. 2012). This discordance between the results of the CH analysis and the CCP data can easily be explained if during the fusions and fissions events that typify *Cricetus cricetus* karyotype evolution, loss of CH blocks have occurred, followed by amplification of specific remaining repeats, becoming these regions more homogeneous. A more intense rate of satDNA repeat units homogenization (concerted evolution) in *Cricetus cricetus* could also have resulted in the lowest CH heterogeneity observed in this karyotype comparing with the other studied species. As it is accepted that chromosomal fusions are associated with loss of centromeric and telomeric sequences at the fusion breakpoints (e.g. Elder 1980, Iannuzzi et al 1987, Garagna et al. 1995), the loss of repetitive sequences in *Cricetus cricetus* is also supported by the absence of ITSs in this karyotype (Figure 13), being on the contrary these telomeric repeats particularly abundant in *Phodopus roborovskii*, *Phodopus sungorus* and *Praomys tullbergi* chromosomes (present work, Figure 13 and Meles et al. 2007). Regarding *Peromyscus eremicus* karyotype evolution, CCP data suggested that this karyotype is the more conserved relatively to the Muroidea ancestral (Vieira-da-Silva and Louzada et al. *unpublished data*), and this is in accordance with our CH analysis and also with the absence of ITSs in *Peromyscus eremicus* chromosomes (Multani et al. 2001, Figure 13). If it is accepted that the majority of ITSs can result from chromosome rearrangements during evolution (reviewed by Ruiz-Herrera et al. 2008), the non occurrence of rearrangements should not allow the origination of ITSs.

When we compare only the two *Phodopus* species, the higher CH heterogeneity identified in *Phodopus sungorus* chromosomes, which is synonymous of a higher molecular dynamics of the repeats located in these regions, points that this species have an even more derivative karyotype than *Phodopus roborovskii* (Figure 13). This is also supported by different cytogenetic studies (Schmid et al. 1986, Romanenko et al. 2007), which suggest that *Phodopus roborovskii* retains a more primitive karyotype than *Phodopus sungorus,* differing from the *Phodopus* ancestor by three and six rearrangements, respectively.

Figure 13- Table resuming all the results obtained in this work for the five studied species. The results obtained here are integrated with data obtained in other works. The divergence degree from the Muroidea ancestral is present according to the syntenic associations detected by Romanenko et al. (2007), Chaves et al. (2012) and Vieira-da-Silva and Louzada et al. (*unpublished data*). The CH heterogeneity (CH subclasses) is identified using a panel of four REs (AluI, DraI, HaeIII, and RsaI) and is referent to autosomes and to the X-chromosome in the five species.

According to what has been referred, the molecular dynamics of the repeats located in CH are implicated in the chromosomal rearrangements that occurred in the five studied genomes. However, this dynamics may not only concern the nucleotide sequence alterations of repeats, focused in the origin of new sequences, but also implicate the variation of copy number and/or intragenomic movements of repeats, which do not reflect CH heterogeneity. This can fit with what was previously described for *Cricetus cricetus*, having this species the lowest CH heterogeneity, but showing one of the more derivative karyotypes of the studied species. Contrarily, in the specific case of *Phodopus* species, it seems that the molecular dynamics of the repeats located in CH falls mainly on changes in the nucleotide sequence, because of its very high CH heterogeneity (Figure 13). An illustration of this high heterogeneity is the identification of two satDNAs (PROsat and PsatDNA) colocalized in the same heterochromatic blocks of a few *Phodopus roborovskii* chromosomes and the presence of two chromosome-specific satDNAs (PSUchr1sat and PsatDNA) located in the same

heterochromatic (peri)centromeric region of *Phodopus sungorus* chromosome 1 (Figure 13). Accordingly, in these genomes, it seems that new satDNAs are recurrently originated mostly from other preexisting satDNAs, being the older sequences maintained. This culminates in an accumulation of a large variety of satDNAs in *Phodopus* genomes during evolution, presenting these sequences the capacity to originate new satellites and at the same time to conserve its identity. The reason for this increasing diversity of satDNAs during chromosome evolution in *Phodopus* is not understood, mainly because it is accepted that satDNAs evolve in such a concerted way that it leads to homogenization of repeats within a genome, ultimately resulting in the elimination of old satellites with the simultaneous origin of new sequences. Thus, the finding of a CH fraction quite heterogeneous in the *Phodopus* genomes analyzed in the present work seems to have a biological meaning associated. For this reason, here we propose that these genomes can be used as models to study the evolution and functions of satDNAs in eukaryotic genomes.

Additionally, beyond its important function in genome restructuring, the data obtained in this work also suggests a satDNA role in the centromeric function, control of gene expression, chromatin remodelation and cellular response to stress. LINE-1 sequences as well have important functions in control of gene expression, acting in gene imprinting and in X-chromosome inactivation. Thereby, despite being initially considered useless genomic elements, presently it is impossible to deny that the repetitive sequences are crucial for proper functioning and evolution of the eukaryotic genomes, dethroning to our view the importance given in the past only to the protein-coding sequences. Looking for instance at long term, by creating chromosomal instability, these DNA elements are responsible for mutations, which result in genetic diversity, increasing the possibilities for species survival in more inhospitable environmental conditions. The importance scale of these DNA elements for eukaryotic genomes is even more evident when we consider that protein-coding regions account only for a tiny part of these genomes, as for example only 1,5% of the human genome. This could not better justify the goal of the work done in this thesis.

III.2- Future perspectives

Despite the diversity of the reports available in the literature describing the isolation and molecular characterization of repetitive sequences in eukaryotic genomes, much more work is needed until be completely elucidated the role(s) played by this large genomic fraction. The study presented in this thesis is a small contribution to that, and to continue it different approaches can be suggested.

Here, satellite sequences were isolated from almost all studied species, excluding the *Praomys tullbergi*. Therefore, future analyses need to focus in the isolation and physical mapping of satDNAs in this genome. The description of other satDNAs in *Cricetus cricetus* genome is as well interesting, mainly because this species presents very large heterochromatic (peri)centromeric regions. Many more satDNAs could also be isolated in the remaining studied species, particularly in *Phodopus roborovskii* and *Phodopus sungorus*, as they present a high CH molecular heterogeneity. Nevertheless, probably more important than isolate novel satDNAs, is the analysis of their transcription profile. Future works focused in a complete characterization of satDNA transcripts, namely transcripts length, occurrence of single or both strands transcripts and transcripts sub-cellular localization, will certainly highlight our knowledge regarding the genomic importance of these repetitive sequences. The enlargement of satDNA analysis to other rodent genome's, beyond the five species studied here, can also surely lead to a better understanding about the evolution of these repeats and how this is in turn related with complex rodent's genome evolution.

Regarding LINE-1 sequences, it is also intend to test more high range of genomes for the co-localization of LINE-1 and imprinted genes regions, so as to reinforce the results obtained here. Presently, is possible to extrapolate the imprinted regions in all species chromosomes where were applied mouse and human painting probes, being precisely known for these two species the location of imprinted genes. According to that and concerning rodent species, the analysis about LINE-1 and imprinted regions location can be accomplished in rodent species belonging to all Rodentia Superorders (Romanenko et al. 2012).

To obtain a complete picture about the repetitive fraction of the five genomes studied here, the analysis of other dispersed repeats, like LTR-retrotransposons and DNA transposons, seems also a subsequent step to follow.

Chapter IV

References

CHAPTER IV. REFERENCES

Abuín M, Martínez P, Sánchez L (1996). Localization of the repetitive telomeric sequence (TTAGGG)n in four salmonid species. Genome 39: 1035-1038.

Acosta MJ, Marchal JA, Fernández-Espartero CH, Bullejos M, Sánchez A (2008). Retroelements (LINEs and SINEs) in vole genomes: Differential distribution in the constitutive heterochromatin. Chromosome Research 8: 1253-1263.

Acosta MJ, Marchal JA, Fernández-Espartero CH, Romero-Fernández I, Rovatsos MT, Giagia-Athanasopoulou EB, Gornung E, Castiglia R, Sánchez A (2010). Characterization of the satellite DNA Msat-160 from species of Terricola (*Microtus*) and Arvicola (Rodentia, Arvicolinae). Genetica 138: 1085-1098.

Adega F, Chaves R, Guedes-Pinto H (2005). Chromosome restriction enzyme digestion in domestic pig (*Sus scrofa*). Constitutive heterochromatin arrangement. Genes & Genetic Systems 80: 49-56.

Adega F, Chaves R, Guedes-Pinto H (2007) Constitutive heterochromatin characterization of white-lipped and collared peccaries (Tayassuidae). Journal of Genetics 86: 19-26.

Adega F, Guedes-Pinto H, Chaves R (2009). Satellite DNA in the Karyotype Evolution of Domestic Animals – Clinical Considerations. Cytogenetic and Genome Research 126: 12-20.

Adey NB, Comer MB, Edgell MH, Hutchison CA (1991). Nucleotide sequence of a mouse full-length F-type L1 element. Nucleic Acids Research 19(9): 2497.

Adey NB, Schichman SA, Graham DK, Peterson SN, Edgell MH, Hutchison CA (1994). Rodent Ll Evolution Has Been Driven by a Single Dominant Lineage That Has Repeatedly Acquired New Transcriptional Regulatory Sequences. Molecular Biology Evolution 11(5): 778-789.

Akagi K, Li J, Stephens RM, Volfovsky N, Symer DE (2008). Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. Genome Research 18: 869-880.

Allen E, Horvath S, Tong F, Kraft P, Spiteri E, Riggs AD, Marahrens Y (2003). High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. Proceedings of the National Academy of Sciences 100 (17): 9940-9945.

Ashley T, Ward DC (1993). A “hot spot” of recombination coincides with an interstitial telomeric sequence in the Armenian hamster. Cytogenetics and Cell Genetics 62: 169-171.

Babushok DV, Ohshima K, Ostertag EM, Chen X, Wang Y, Mandal PK, Okada N, Abrams CS, Kazazian HH Jr (2007). A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. Genome Research 17: 1129-1138.

Bailey JA, Carrel L, Chakravarti A, Eivhler EE (2000). Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proceedings of the National Academy of Sciences 97 (12): 6634-6639.

Balajee AS, Oh HJ, Natarajan AT (1994). Analysis of restriction enzyme-induced chromosome aberrations in the interstitial telomeric repeat sequences of CHO and CHE cells by FISH. Mutation Research 307: 307-313.

Baskin JA (1986). The late Miocene radiation of Neotropical sigmodontine rodents in North America. Contributions to Geology, University of Wyoming 3: 287-303.

Batistoni R, Pesole G, Marracci S, Nardi I (1995). A Tandemly Repeated DNA Family Originated from SINE-Related Elements in the European Plethodontid Salamanders (Amphibia, Urodela). Journal of Molecular Evolution 40: 608-615.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL (1994). African origin of human-specific polymorphic Alu insertions. Proceedings of the National Academy of Sciences 91(25): 12288–12292.

Beklemisheva VR, Romanenko SA, Biltueva LS, Trifonov VA, Vorobieva NV, Serdukova NA, Rubtsova NV, Brandler OV, O’Brien PCM, Yang F, Stanyon R, Ferguson-Smith MA, Graphodatsky AS (2011). Reconstruction of karyotype evolution in core Glires. I. The genome homology revealed by comparative chromosome painting. Chromosome Research 19: 549-565.

Belancio VP, Hedges DJ, Deininger P (2008). Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. Genome Research 18: 343-358.

Benton MJ, Donoghue PC (2007). Paleontological evidence to date the tree of life. Molecular Biology and Evolution 24: 26-53.

Beraldi R, Pittoggi C, Sciamanna I, Mattei E, Spadafora C (2006). Expression of LINE-1 retroposons is essential for murine preimplantation development. Molecular Reproduction and Development 73: 279-287.

Bergmann JH, Jakubsche JN, Martins NM, Kagansky A, Nakano M, Kimura H, Kelly DA, Turner BM, Masumoto H, Larionov V, Earnshaw WC (2012). Epigenetic engineering: histone H3K9 acetylation is compatible with kinetochore structure and function. Journal of Cell Science 125(Pt2): 411-421.

Bernard P, Maure JF, Partridge JF, Genier S, Javerzat JP, Allshire RC (2001). Requirement of heterochromatin for cohesion at centromeres. Science 21: 2539-2542.

Bertoni L, Attolini C, Faravelli M, Simi S, Giulotto E (1996). Intrachromosomal telomere-like DNA sequences in Chinese hamster. Mammalian Genome 7: 853-855.

Blackburn EH (2001). Switching and signaling at the telomere. Cell 106: 661-673.

Boissinot S, Furano AV (2001). Adaptive Evolution in LINE-1 Retrotransposons. Molecular Biology and Evolution 18(12): 2186-2194.

Boissinot S, Entezam A, Furano AV (2001). Selection against deleterious LINE-1-containing *loci* in the human lineage. Molecular Biology and Evolution 18: 926-935.

Boissinot S, Davis J, Entezam A, Petrov D, Furano AV (2006). Fitness cost of LINE-1 (L1) activity in humans. Proceedings of the National Academy of Sciences 103: 9590-9594.

Bolzán AD, Bianchi MS (2006). Telomeres, interstitial telomeric repeat sequences, and chromosomal aberrations. Mutation Research 612: 189-214.

Bonaccorsi S, Gatti M, Pisano C, Lohe A (1990). Transcription of a satellite DNA on two Y chromosome loops of *Drosophila*. Chromosoma 99: 260-266.

Borstnik B, Pumpernik D, Lukman D, Ugarković Đ, Plohl M (1994). Tandemly repeated pentanucleotides in DNA sequences of eucaryotes. Nucleic Acids Research 22: 3412–3417.

Boulesteix M, Weiss M, Biémont C (2006). Differences in the Genome Size between closely related species: the *Drosophila melanogaster* species subgroup. Molecular Biology and Evolution 23(1): 162-167.

Boyle AL, Ballard SG, Ward DC (1990). Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence *in situ* hybridization. Proceedings of the National Academy of Sciences 87: 7757-7761.

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

Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr (2003). Hot L1s account for the bulk of retrotransposition in the human populations. Proceedings of the National Academy of Sciences 100: 5280-5285.

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.

Buckingham S (2003). The major world of microRNAs. Horizon Symposia 1-3.

Canapa A, Barucca M, Cerioni PN, Olmo E (2000). A satellite DNA containing CENP-B box-like motifs is present in the Antarctic scallop *Adamussium colbecki*. Gene 247: 175-180.

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

Carleton MD, Musser GG (2005). Order Rodentia. In: Wilson DE, Reeder DM (Eds), Mammal Species of the World: A Taxonomic and Geographic Reference. Baltimore: Johns Hopkins University Press, pp 745-1601.

Casavant NC, Sherman AN, Wichman HA (1996). Two Persistent LINE-1 Lineages in *Peromyscus* Have Unequal Rates of Evolution. Genetics 142: 1289-1298.

Casavant NC, Lee RN, Sherman AN, Wichman HA (1998). Molecular Evolution of Two Lineages of L1 (LINE-1) Retrotransposons in the California Mouse, *Peromyscus californicus.* Genetics 150: 345-357.

Casavant NC, Scott L, Cantrell MA, Wiggins LE, Baker RJ, Wichman HA (2000). The end of the LINE?: lack of recent L1 activity in a group of South American rodents. Genetics 154: 1809-1817.

Cavalier-Smith T (1985). Selfish DNA and the origin of introns. Nature 315: 283-284.

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

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

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

Cheng ZJ, Murata M (2003). A centromeric tandem repeat family originating from a part of Ty3/gypsy – retroelement in wheat and its relatives. Genetics 164: 665-672.

Cordaux R, Batzer MA (2009). The impact of retrotransposons on human genome evolution. Nature reviews 10: 691-703.

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.

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

Dai L, Huang Q, Boeke JD (2011). Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. BMC Biochemistry 12(18): 1-11.

Dawe RK, Henikoff S (2006). Centromeres puts epigenetics in the driver's seat. Trends in Biochemical Sciences 31: 662-669.

DeBerardinis RJ, Kazazian HH Jr (1999). Analysis of the Promoter from an Expanding Mouse Retrotransposons Subfamily. Genomics 56: 317-323.

Deininger PL, Batzer MA (1999). Alu repeats and human disease. Molecular Genetics and Metabolism 67: 183-193.

Deininger PL, Batzer MA (2002). Mammalian retroelements. Genome Research 12: 1455-1465.

Delsuc F, Scally M, Madsen O, Stanhope MJ, De Jong WW, Catzeflis FM, Springer MS, Douzery EJ (2002). Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Molecular Biology and Evolution 19: 1656-1671.

Deuve JL, Bennett NC, O'Brien PCM, Ferguson-Smith MA, Faulkes CG, Britton-Davidian J, Robinson TJ (2006). Complex evolution of X and Y autosomal translocations in the giant mole-rat, *Cryptomys mechowi* (Bathyergidae). Chromosome Research 14: 681-691.

Devos KM (2010). Grass genome organization and evolution. Current Opinion in Plant Biology 13: 139-145.

Dewannieux M, Heidmann T (2005). LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modelling. Cytogenetic and Genome Research 110: 35-48.

Ding W, Lin L, Chen B, Dai J (2006). L1 Elements, Processed Pseudogenes and Retrogenes in Mammalian Genomes. International Union of Biochemistry and Molecular Biology Life 58(12): 677-685.

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

Doolittle WF, Sapienza C (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature 284: 601-603.

Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV, Gilbert N (2010). Characterization of LINE-1 ribonucleoprotein particles. PLoS Genetics 6: 1-19.

Dover GA (2002). Molecular drive. Back to basics. Trends in Genetics 18: 587–589.

Dover GA, Flavell RB (1984). Molecular coevolution: DNA divergence and the maintenance of function. Cell 38: 623-624.

Elder FF (1980). Tandem fusions, centric fusions, and chromosomal evolution in the cotton rats, genus *Sigmodon*. Cytogenetics and Cell Genetics 26: 199-210.

Elder JF Jr, Turner BJ (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. The Quarterly Review of Biology 70: 297-320.

Enukashvily NI, Ponomartsev NV (2013). Mammalian Satellite DNA: A Speaking Dumb. Advances in Protein Chemistry and Structural Biology 90: 31-65.

Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H, Chalajour F, Kilic N, Strätling WH, Schumann GG (2004). Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. The Journal of Biological Chemistry 279: 27753-27763.

Eymery A, Callanan M, Vourc'h C (2009). The secret message of heterochromatin: new insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription. The International Journal of Developmental Biology 53: 259-268.

Fagundes V, Christoff AU, Yonenaga-Yassuda Y (1998). Extraordinary Chromosomal Polymorphism with 28 different Karyotypes in the Neotropical Species *Akodon cursor* (Muridae, Sigmodontinae), one of the Smallest Diploid Number in Rodents (2n = 16, 15 and 14). Hereditas 129: 263-274.

Fahlbusch V (1969). Pliozäne und pleistozäne Cricetinae (Rodentia, Mammalia) aus Polen. Acta Zoologica Cracovienisa 14: 99-137.

Fanning T (1987). Origin and evolution of a major feline satellite DNA. Journal of Molecular Biology 197: 627-634.

Farré M, Ponsà M, Bosch M (2009). Interstitial telomeric sequences (ITSs) are not located at the exact evolutionary breakpoints in primates. Cytogenetic and Genome Research 124: 128-131.

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. Nucleic Acids Research 22 (18): 3728-3736.

Feldhamer GA, Drickamer LC, Vessey SH, Merritt JF (2003). Mammalogy. Adaptation, Diversity and Ecology. Second ed. Mc Graw Hill, Boston.

Fernández JL, Gosálvez J, Goyanes V (1995). High frequency of mutagen-induced chromatid exchanges at interstitial telomere-like DNA sequence blocks of Chinese hamster cells. Chromosome Research 3: 281-284.

Feschotte C, Pritham EJ (2007). DNA transposons and evolution of eukaryotic genomes. Annual Review of Genetics 41: 331-368.

Finato AO, Varella-Garcia M, Tajara EH, Taddei VA, Morielle-Versute E (2000). Intrachromosomal distribution of telomeric repeats in *Eumops glaucinus* and *Euntops perotis* (Molossidae, Chiroptera). Chromosome Research 8: 563-569.

Frausto SD, Lee E, Tang H (2013). Cyclophilins as Modulators of Viral Replication. Viruses 5: 1684-1701.

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

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.

Furano AV, Usdin K (1995). DNA "Fossils" and Phylogenetic Analysis. The Journal of Biology Chemistry 270(43): 25301-25304.

Furano AV, Duvernell DD, Boissinot S (2004). L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. Trends in Genetics 20: 9-14.

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, Broccoli D, Redi CA, Searle JB, Cooke HJ, Capanna E (1995). Robertsonian metacentrics of the house mouse lose telomeric sequences but retain some minor satellite DNA in the pericentromeric area. Chromosoma 103: 685-692.

Garagna S, Pérez-Zapata A, Zuccotti M, Mascheretti S, Marziliano N, Redi CA, Aguilera M, Capanna E (1997). Genome composition in Venezuelan spiny-rats of the genus *Proechimys* (Rodentia, Echimyidae). I. Genome size, C-heterochromatin and repetitive DNAs *in situ* hybridization patterns. Cytogenetics and Cell Genetics 78: 36-43.

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

Goodier JL, Ostertag EM, Du K, Kazazian HH Jr (2001). A Novel Active L1 Retrotransposon Subfamily in the Mouse. Genome Research 11: 1677-1685.

Gosálvez J, López-Fernández C, Goyanes R, Mezzanotte V (1997). Chromosome differentiation using nucleases: An overview. In: Henriques-Gil N, Parker JS and Puertas MJ (eds), Chromosomes Today. London: Chapman & Hall, pp 23-49.

Graham T, Boissinot S (2006). The Genomic Distribution of L1 Elements: The Role of Insertion Bias and Natural Selection. Journal of Biomedicine and Biotechnology 75327: 1-5.

Graphodatsky AS (1989). Conserved and variable elements of mammalian chromosomes. In: Halnan CRE (Eds), Cytogenetics of Animals. United Kingdown, CAB International Press, pp 95-123.

Graphodatsky AS, Yang F, Dobigny G, Romanenko SA, Biltueva LS, Perelman PL, Beklemisheva VR, Alkalaeva EZ, Serdukova NA, Ferguson-Smith MA, Murphy WJ, Robinson TJ (2008). Tracking genome organization in rodents by Zoo-FISH. Chromosome Research 16: 261-274.

Gregory TR (2011). Animal genome size database. Available from: http://www.genomesize.com.

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

Grewal SIS, Jia S (2007). Heterochromatin revisited. Nature Reviews 8: 35-46.

Hall LE, Mitchell SE, O'Neill RJ (2012). Pericentric and centromeric transcription: a perfect balance required. Chromosome Research 20:535-546.

Hamilton MJ, Honeycutt RL, Baker RJ (1990). Intragenomic movement, sequence amplification and concerted evolution in satellite DNA in harvest mice, *Reithrodontomys:* Evidence from *in situ* hybridization. Chromosoma 99: 321 329.

Han JS (2010). Non-Long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mobile DNA 1: 15.

Han JS, Boeke JD (2004). A highly active synthetic mammalian retrotransposon. Nature 429: 314-318.

Han JS, Boeke JD (2005). LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression?. BioEssays 27: 775-784.

Hayward BE, Zavanelli M, Furano AV (1997). Recombination Creates Novel Ll (LINE1) Elements in *Rattus norvegicus.* Genetics 146: 641-654.

Hedges DJ, Deininger PL (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutation Research 616: 46-59.

Heikkinen E, Launonen V, Muller E, Bachmann L (1995). The pvB370 BamHI satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic pDv elements. Journal of Molecular Evolution 41: 604-614.

Heitz E (1928). Das Heterochromatin der Moose. I. Jahrb. Wiss Bot 69: 762-818.

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

Hohjoh H, Singer MF (1996). Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. The EMBO Journal 15: 630-639.

Hsu TC, Arrighi FE (1966). Chromosomal evolution in the genus *Peromyscus* (Cricetidae, Rodentia). Cytogenetics 5: 355-359.

Huchon D, Madsen O, Sibbald MJ, Ament K, Stanhope MJ, Catzeflis F, de Jong WW, Douzery EJP (2002). Rodent Phylogeny and a Timescale for the Evolution of Glires: Evidence from an Extensive Taxon Sampling Using Three Nuclear Genes. Molecular Biology and Evolution 19(7): 1053-1065.

Hughes AL, Friedman R (2000). Evolutionary diversification of protein-coding genes of hantaviruses. Molecular Biology and Evolution 17: 1558–1568.

Iannuzzi L, Di Berardino D, Gustavsson I, Ferrara L, Di Meo GP (1987). Centromeric loss in translocation of centric fusion type in cattle and watter buffalo. Hereditas 106: 73-81.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature 409: 860-921.

Inukai T (2004). Role of transposable elements in the propagation of minisatellites in the rice genome. Molecular Genetics and Genomics 271: 220-227.

Jacobs LL (1978). Fossil rodents (Rhizomyidae and Muridae) from Neogene Siwalik deposits, Pakistan. *Mus*. North Arizona Press Bull 52: 1-104.

Jansa SA, Weksler M (2004). Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. Molecular Phylogenetics and Evolution 31: 256-276.

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

Jurka J, Kapitonov VV, Kohany O, Jurka M V (2007). Repetitive Sequences in Complex Genomes: Structure and Evolution. Annual Review of Genomics and Human Genetics 8: 241-59.

Kapitonov VV, Jurka J (1999). Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica 107: 27-37.

Kapitonov VV, Jurka J (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. Nature Reviews Genetics 9: 411-412.

Kapitonov VV, Holmquist GP, Jurka J (1998). L1 Repeat Is a Basic Unit of Heterochromatin Satellites in Cetaceans. Molecular Biology and Evolution 15(5): 611-612.

Kapitonov VV, Tempel S, Jurka J (2009). Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domais protein sequence. Gene 448: 207-213.

Kazazian HH Jr (2000). L1 retrotransposons shape the mammalian genome. Science 289: 1152-1153.

Kazazian HH Jr (2004). Mobile elements: drivers of genome evolution. Science 303: 1626-1632.

Kolb J, Chuzhanova N A, Högel J, Vasquez K M, Cooper D N, Bacolla A, Kehrer-Sawatzki H (2009). Cruciform-forming inverted repeats appear to have mediated many of the microinversions that distinguish the human and chimpanzee genomes. Chromosome Research 17: 469-483.

Koop BF, Baker RJ, Genoways HH (1983). Numerous chromosomal Polymorphisms in a natural population of rice rats (*Oryzomys*, Cricetidae). Cytogenetics and Cell Genetics 35: 131-135.

Korenberg JR, Rykowski MC (1988). Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. Cell 53: 391-400.

Kowalski K (2001). Pleistocen rodents of Europe. Folia Quarternaria 72.

Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J (2006). Retroposed elements as archives for the evolutionary history of placental mammals. PLoS Biology 4: e91.

Lecompte E, Denys C, Granjon L (2005). Confrontation of morphological and molecular data: The *Praomys* group (Rodentia, Muridae) as a case of adaptative convergences and morphological stasis. Molecular Phylogenetics and Evolution 37: 899-919.

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 Evolutionary Biology 8: 199.

Lee SH, Cho SY, Shannon MF, Fan J, Rangasamy D (2010). The Impact of CpG Island on Defining Transcriptional Activation of the Mouse L1 Retrotransposable Elements. PLoS ONE 5 (6): 1-14.

Li WH (1997). Molecular Evolution. Sunderland, Sinauer.

Li YC, Lee C, Sanoudou D, Hsu TH, Li SY, Lin CC (2000). Interstitial colocalization of two cervid satellite DNAs involved in the genesis of the Indian muntjac karyotype. Chromosome Research 8: 363–373.

Li YX, Kirby ML (2003). Coordinated and conserved expression of alphoid repeat and alphoid repeat-tagged coding sequences. Developmental Dynamics 228: 72–81.

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

Longo M S, Carone D M, Comparative Sequencing Program NISC, Green E D, O'Neill M J, O'Neill R J (2009). Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. BMC Genomics 10: 334.

López-Flores I, de la Herrán R, Garrido-Ramos MA, Boudry P, Ruiz-Rejón C, Ruiz-Rejón M (2004). The molecular phylogeny of oysters based on a satellite DNA related to transposons. Gene 339: 181-188.

Lorite P, Carrillo JA, Tinaut A, Palomeque T (2004). Evolutionary dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera, Formicidae). Gene 332: 159-168.

Louzada S, Vieira-da-Silva A, Kubickova S, Rubes J, Adega F, Raquel Chaves R (2014). A novel satellite DNA sequence in *Peromyscus* genome that evolves by copy number

fluctuation: PMSat Is it an "ancient" or a "nascent" satellite DNA?. Submitted for publication to Heredity.

Lucchinetti E, Feng J, Silva R, Tolstonog GV, Schaub MC, Schumann GG, Zaugg M (2006). Inhibition of LINE-1 expression in the heart decreases ischemic damage by activation of Akt/PKB signaling. Physiological Genomics 25: 314-324.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389: 251-260.

Lupski JR (2010). Retrotransposition and Strutural Variation in the Human Genome. Cell 141: 1110-1112.

Lyon MF (1998). X-Chromosome Inactivation Spreads Itself: Effects in Autosomes. The American Journal of Human Genetics 63: 17-19.

Lyon MF (2006). Do LINEs Have a Role in X-chromosome Inactication?. Journal of Biomedicine and Biotechnology, Article ID 59746: 1-6.

Macas J, Koblížková A, Navrátilová A, Neumann P (2009). Hypervariable 3′ UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. Gene 448: 198–206.

Maccioni RB, Cambiazo V (1995). Role of Microtubule-Associated Proteins in the Control of Microtubule Assembly. Physiological Reviews 75(4): 835-864.

Malik A, Korol A, Hübner S, Hernandez AG, Thimmapuram J, Ali S, Glaser F, Paz A, Avivi A, Band M (2011). Transcriptome Sequencing of the Blind Subterranean Mole Rat, *Spalax galili*: Utility and Potential for the Discovery of Novel Evolutionary Patterns. PLoS ONE 6(8): e21227.

Marchal JA, Acosta MJ, Bullejos M, Puerma E, Díaz de la Guardia R, Sánchez A (2006). Distribution of L1-retroposons on giant sex chromosomes of *Microtus cabrerae* (Arvicolidae, Rodentia): functional and evolutionary implications. Chromosome Research 14: 177-186.

Martin SL (2006). The ORF1 protein encoded by LINE-1: structure and function during L1 Retrotransposition. Journal of Biomedicine and Biotechnology 2006 (ID 45621): 1-6.

Martin SL (2010). Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. RNA Biology 7: 706-711.

Martin SL, Li WLP, Furano AV, Boissinot S (2005). The structure of mouse and human L1 elements reflect their insertion mechanism. Cytogenetic and Genome Research 110: 223-228.

Martinez-Lage A, Rodriguez-Farina F, Gonzalez-Tizon A, Mendez J (2005). Origin and evolution of *Mytilus* mussel satellite DNAs. Genome 48: 247-256.

Mashkova T, Oparina N, Alexandrov I, Zinovieva O, Marusina A, Yurov Y, Lacroix MH, Kisselev L (1998). Unequal crossing-over is involved in human alpha satellite DNA rearrangements on border of the satellite domain. FEBS Letters 441: 451-457.

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

Matthey R (1958). Les chromosomes et la position systématique de quelques Murinae africains (Mammalia–Rodentia). Acta Tropica 15: 97–117.

Matthey R (1960). Chromosomes, hétêrochromosomes, et cytology comparée des Cricetinar Paléarctiques (Rodentia). Caryologia 13: 199-223.

Mayorov VI, Adkison LR, Vorobyeva NV, Khrapov EA, Kholodhov NG, Rogozin IB, Nesterova TB, Protopopov AI, Sablina OV, Graphodatsky AS, Zakian SM (1996). Organization and chromosomal localization of a B1-like containing repeat of *Microtus subarvalis*. Mammalian Genome 7: 593-597.

Mayorov VI, Rogozin IB, Adkison LR (1999). Characterization of several LINE-1 elements in *Microtus kirgisorum*. Mammalian Genome 10: 724-729.

McClintock B (1950). The origin and behavior of mutable *loci* in maize. Proceedings of the National Academy of Sciences 36: 344–355.

McDonald JF (1995). Transposable elements: possible catalysis of organismic evolution. Trends in Ecology and Evolution 10: 123-126.

McKenna MC, Bell SK (1997). Classification of Mammals above the Species Level. Columbia University Press, New York.

Meles S, Adega F, Guedes-Pinto H, Chaves R (2007). 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. Molecular Biology and Evolution 15: 1062-1068.

Meštrović N, Castagnone-Sereno P, Plohl M (2006). High conservation of the differentially amplified MPA2 satellite DNA family in parthenogenetic root-knot nematodes. Gene 376: 260-267.

Meštrović N, Pavlek M, Car A, Castagnone-Sereno P, Abad P, Plohl M (2013). Conserved DNA Motifs, Including the CENP-B Box-like, Are Possible Promoters of Satellite DNA Array Rearrangements in Nematodes. PLoS ONE 8(6): e67328.

Metcalfe CJ, Eldridge MD, Johnston PG (2007). Mapping the distribution of the telomeric sequence (T2AG3) n in the Macropodoidea (Marsupialia) by fluorescence *in situ* hybridization. II. The ancestral 2n=22 macropodid karyotype. Cytogenetic and Genome Research 116: 212-217.

Meyne J, Baker RJ, Hobart HH, Hsu TC, Ryder OA, Ward OG, Wiley JE, Wurster-Hill DH, Yates TL, Moyzis RK (1990). Distribution of non-telomeric sites of the (TTAGGG)n telomeric sequence in vertebrate chromosomes. Chromosoma 99: 3-10.

Michaux J, Catzeflis F (2000). The bushlike radiation of muroid rodents is exemplified by the molecular phylogeny of the LCAT nuclear gene. Molecular Phylogenetics and Evolution 17: 280-293.

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

Modi WS (1993). Rapid, localized amplification of a unique satellite DNA family in the rodent *Microtus chrotorrhinus.* Chromosoma 102: 484-490.

Moran JV, Gilbert N (2002). Mammalian LINE-1 retrotransposons and related elements. In Craig NL, Craigie R, Gellert M, Lambowitz AM (Eds), Mobile DNA II. Washington DC: American Society of Microbiology Press, pp. 836-869.

Mouse ENCODE Consortium (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biology 13(418): 2-5.

Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562.

Mravinac B, Plohl M, Meštrović N, Ugarković Đ (2002). Sequence of PRAT satellite DNA "frozen" in some Coleopteran species. Journal of Molecular Evolution 54: 774-783.

Mravinac B, Ugarković Đ, Franjević D, Plohl M (2005). Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. Journal of Molecular Evolution 60: 513-525.

Multani AS, Ozen M, Furlong CL, Zhao YJ, Hsu TC, Pathak S (2001). Heterochromatin and interstitial telomeric DNA homology. Chromosoma 110: 214-220.

Muotri AR, Marchetto MC, Coufal NG, Gage FH (2007). The necessary junk: new functions for transposable elements. Human Molecular Genetics 16 (2): R159-R167.

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ (2001). Molecular phylogenetics and the origins of placental mammals. Nature 409: 614-618.

Nachman MW, Searle JB (1995). Why is the house mouse karyotype so variable?. Trends in Ecology and Evolution 10: 397-402.

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292: 110-113.

Nanda I, Schrama D, Feichtinger W, Haaf T, Schartl M, Schmid M (2002). Distribution of telomeric (TTAGGG)(n) sequences in avian chromosomes. Chromosoma 111: 215-227.

Nechay G (2000). Status of Hamsters *Cricetus cricetus, Cricetus migratorius, Mesocricetus newtoni* and other hamster species in Europe. Nature and Environment Series No. 106.

Nergadze SG, Santagostino MA, Salzano A, Mondello C, Giulotto E (2007). Contribution of telomerase RNA retrotranscription to DNA double-strand break repair during mammalian genome evolution. Genome Biology 8: R260.

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. Molecular Phylogenetics and Evolution 39: 135-148.

Nicolas V, Verheyen E, Verheyen W, Hulselmans J, Dillens M, Akpatou B, Dudu A, Wendelen W, Colyn M (2005). Systematics of African lowland rainforest *Praomys* (Rodentia, Muridae) based on molecular and craniometrical data. Zoological Journal of the Linnean Society 145: 539-553.

Nonaka N, Kitajima T, Yokobayashi S, Xiao G, Yamamoto M, Grewal SIS, Watan Y (2002). Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. Nature Cell Biology 4: 89-93.

Norris RW, Zhou K, Zhou C, Yang G, Kilpatrick CW, Honeycutt RL (2004). The Phylogenetic position of the zokors (Myospalacinae) and comments on the families of muriods (Rodentia). Molecular Phylogenetics and Evolution 31: 972-978.

Nowak RN (1991). Walker's Mammals of the World. The Johns Hopkins University Press, Baltimore.

O'Brien SJ, Menninger JS, Nash WG (2006). Atlas of Mammalian Chromosomes. John Wiley & Sons Inc, United States of America.

Obuse C, Iwasaki O, Kiyomitsu T, Goshima G, Toyoda Y, Yanagida M (2004). A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. Nature Cell Biology 6: 1135-1141.

Ohta T, Dover GA (1984). The cohesive population genetics of molecular drive. Genetics 108: 501-521.

Okada T, Ohzeki J, Nakano M, Yoda K, Brinkley WR, Larionov V, Masumoto H (2007). CENP-B Controls Centromere Formation Depending on the Chromatin Context. Cell 131: 1287-1300.

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

Ostertag EM, Kazazian HH Jr. (2001). Twin priming: A proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Research 11: 2059-2065.

Palomeque T, Lorite P (2008). Satellite DNA in insects: a review. Heredity 100: 564-573.

Parish DA, Vise P, Wichman HA, Bull JJ, Baker RJ (2002). Distribution of LINEs and other repetitive elements in the karyotype of the bat *Carollia*: implications for X-chromosome inactivation. Cytogenetic and Genome Research 96: 191-197.

Pascale E, Valle E, Furano AV (1990). Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before themurine radiation. Proceedings of the National Academy of Sciences 87: 9481-9485.

Patton JL, Sherwood SW (1982). Genome evolution in pocket gophers (genus *Thomomys*). I. Heterochromatin variation and speciation potential. Chromosoma 85: 149-162.

Peacock WJ, Dennis ES, Gerlach WL (1982). DNA sequence changes and speciation. Progress in Clinical Biological Research 96: 123-142.

Pellegrino KC, Rodrigues MT, Yonenaga-Yassuda Y (1999). Chromosomal evolution in the Brazilian lizards of genus *Leposoma* (Squamata, Gymnophthalmidae) from Amazon and Atlantic rain forests: banding patterns and FISH of telomeric sequences. Hereditas 131: 15-21.

Petrov DA (2001). Evolution of genome size: new approaches to an old problem. Trends in Genetics 17: 23-28.

Pezer Ž, Ugarković Đ (2008a). Role of non-coding RNA and heterochromatin in aneuploidy and cancer. Seminars in Cancer Biology 18: 123-130.

Pezer Ž, Ugarković Đ (2008b). RNA Pol II Promotes Transcription of Centromeric Satellite DNA in Beetles. PLoS ONE 3(2): e1594.

Pezer Ž and Ugarković Đ (2012). Satellite DNA-associated siRNAs as mediators of heat shock response in insects. RNA Biology 9(5): 587-595.

Pieczarka JC, Nagamachi CY, Muniz JAPC, Barros RMS, Mattevi MS (1998). Analysis of constitutive heterochromatin of *Aotus* (Cebidae, Primates) by restriction enzyme and fluorochrome bands. Chromosome Research 6: 77-83.

Pittoggi C, Sciamanna I, Mattei E, Beraldi R, Lobascio AM, Mai A, Quaglia MG, Lorenzini R, Spadafora C (2003). A role of endogenous reverse transcriptase in murine early embryo development. Molecular Reproduction and Development 66:225-236.

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

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, Šatović 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: 543-551.

Pritham EJ (2009). Transposable elements and factors influencing their success in eukaryotes. Journal of Heredity 100: 648.

Probst AV, Almouzni G (2008). Pericentric heterochromatin: Dynamic organization during early development in mammals. Differentiation 76: 15-23.

Rat Genome Sequencing Project Consortium (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428: 493-521.

Rebollo R, Romanish MT, Mager DL (2012). Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. Annual Review of Genetics 46: 21-42.

Richard G.F, Kerrest A, Dujon B (2008). Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes. Microbiology and Molecular Biology Reviews: 686-727.

Riddle BR, Hafner DJ, Alexander LF (2000). Phylogeography and Systematics of the *Peromyscus eremicus* Species Group and Historical Biogeography of North American Warm Regional Deserts. Molecular Phylogenetics and Evolution. 17(2): 145-160.

Robbins LW, Baker RJ (1981). An assessment of the nature of chromosomal rearrangements in 18 species of *Peromyscus* (Rodentia: Cricetidae). Cytogenetics and Cell Genetics 31: 194-202.

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

Robinson TJ, Ruiz-Herrera A, Avise JC (2008). Hemiplasy and homoplasy in the karyotypic phylogenies of mammals. PNAS 105(38): 14477–14481.

Robles F, Herrán R, Ludwig A, Ruiz RC, Ruiz RM, Garrido-Ramos MA (2004). Evolution of ancient satellite DNAs in sturgeon genomes. Gene 338: 133–142.

Rocco L, Morescalchi MA, Costagliola D, Stingo V (2002). Karyotype and genome characterization in four cartilaginous fishes. Gene 295: 289-298.

Rogers DS, Greenbaum IF, Gunn SJ, Engstrom MD (1984). Cytosystematic value of chromosomal inversion data in the genus *Peromyscus* (Rodentia, Cricetidae). Journal of Mammalogy 65: 457-465.

Roizes G (2006). Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. Nucleic Acids Research 34: 1912-1924.

Romanenko SA, Volobouev VT, Perelman PL, Lebedev VS, Serdukova NA, Trifonov VA, Biltueva LS, Nie W, O'Brien PCM, Bulatova NS, 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 Research 15: 283-297.

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

Rossi MS, Pesce CG, Reig OA, Kornblihtt AR, Zorzópulos J (1993). Retroviral-like features in the repetitive unit of the major satellite DNA from the South American rodents of the genus *Ctenomys*. DNA Sequence 3: 379-381.

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*. Cytogenetics and Cell Genetics 69: 179-184.

Rouleux-Bonnin F, Renault S, Bigot Y, Periquet G (1996). Transcription of four satellite DNA subfamilies in *Diprion pini* (Hymenoptera, Symphyta, Diprionidae). European Journal of Biochemistry 238: 752-759.

Rovatsos MTh, Marchal JA, Romero-Fernández I, Fernández FJ, Giagia-Athanosopoulou EB, Sánchez A (2011). Rapid, independent, and extensive amplification of telomeric repeats in pericentromeric regions in karyotypes of arvicoline rodents. Chromosome Research 19: 869-882.

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

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

Ruiz-Herrera A, Nergadze SG, Santagostino M, Giulotto E (2008). Telomeric repeats far from the ends: mechanisms of origin and role in evolution. Cytogenetics and Genome Research 122: 219-228.

Santani A, Raudsepp T, Chowdhary BP (2002). Interstitial telomeric sites and NORs in Hartmann's zebra (*Equus zebra hartmannae*) chromosomes. Chromosome Research 10: 527-534.

Santos S, Chaves R, Guedes-Pinto H (2004). Chromosomal localization of the major satellite DNA family (FA-SAT) in the domestic cat. Cytogenetics. Genome Research 107: 119-122.

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. Journal of Heredity 97(2): 114-118.

Sayah DM, Sokolskaja E, Berthoux L, Luban J (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature 430: 569-573.

Schlötterer C, Harr B (2001). Microsatellite Instability. Encyclopedia of Life Sciences Nature Publishing Group: 1-4.

Schmid MT, Haaf T, Weis H, Schempp W (1986). Chromosomal homologies in hamster species of the genus *Phodopus* (Rodentia, Cricetinae). Cytogenetics and Cell Genetics 43: 168-173.

Schmid MT, Haaf T, Steinlein C, Nanda I, Mahony M (2002). Chromosome banding in Amphibia: XXV. Karyotype evolution and heterochromatin characterization in

Australian *Mixophyes* (Anura, Myobatrachidae). Cytogenetic Genome Research 97: 239-253.

Schmidt T, Heslop-Harrison JS (1998). Genomes, genes and junk: the largescale organization of plant chromosomes. Trends in Plant Science 3: 195-199.

Schueler MG, Dunn JM, Bird CP, Ross MT, Viggiano L, NISC Comparative Sequencing Program, Rocchi M, Willard HF, Green ED (2005). Progressive proximal expansion of the primate X chromosome centromere. Proceedings of the National Academy of Sciences 102: 10563-10568.

Serdobova IM, Kramerov DA (1998). Short Retroposons of the B2 Superfamily: Evolution and Application for the Study of Rodent Phylogeny. Journal of Molecular Evolution 46: 202-214.

Severynse DM, Hutchison CA, Edgell MH (1992). Identification of transcriptional regulatory activity within the 5' A-type monomer sequence of the mouse LINE-1 retroposon. Mammalian Genome 2: 41-50.

Silva MJJ, Yonenaga-Yassuda Y (1998). Karyotype and chromosomal polymorphism of an undescribed *Akodon* from Central Brazil, a species with the lowest known diploid chromosome number in rodents. Cytogenetics and Cell Genetics 81(1): 46-50.

Sitnikova NA, Romanenko SA, O'Brien PCM, 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 Research 15: 447-456.

Slamovits C.H, Rossi MS (2002). Satellite DNA: Agent of chromosomal Evolution in mammals. A review. Mastozoología Neotropical / Journal of Neotropical Mammalogy 9(2): 297-308.

Slijepcevic P (1998). Telomeres and mechanisms of Robertsonian fusion. Chromosoma 107: 136-140.

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

Smýkal P, Kalendar R, Ford R, Macas J, Griga M (2009). Evolutionary conserved lineage of Angela-family retrotransposons as a genome-wide microsatellite repeat dispersal agent. Heredity 103: 157-167.

Song M, Boissinot S (2007). Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. Gene 390: 206-213.

Spadafora C (2008). A reverse transcriptase-dependent mechanism plays central roles in fundamental biological processes. Systems Biology in Reproductive Medicine 54: 11-21.

Stephan W (1989). Tandem-repetitive noncoding DNA: forms and forces. Molecular Biology and Evolution 6: 198-212.

Steppan SJ, Adkins RM, Anderson J (2004). Phylogeny and Divergence-Date Estimates of Rapid Radiations in Muroid Rodents Based on Multiple Nuclear Genes. Systematic Biology 53(4): 533-553.

Steppan SJ, Adkins RM, Spinks PQ, 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.

Stitou S, Díaz de la Guardia R, Jiménez R, Burgos M (1999). Isolation of a species-specific satellite DNA with a novel CENP-B-like box from the North African rodent *Lemniscomys barbarus*. Experimental Cell Research. 250(2):381-386.

Strachan T, Read AP (2004). Human Molecular Genetics 3. Third ed. Garland Science, London.

Strachan T, Webb D, Dover GA (1985). Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. The EMBO Journal 4: 1701-1708.

Sullivan BA, Blower MD, Karpen GH (2001). Determining centromere identity: cyclical stories and forking paths. Nature Reviews Genetics 2: 584-596.

Sumner AT (2003). Chromosomes Organization and Function. Blackwell Publishing Company, United Kingdom.

Sun C, Shepard DB, Chong RA, Arriaza JL, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL (2011). LTR Retrotransposons Contribute to Genomic Gigantism in *Plethodontid* Salamanders. Genome Biology and Evolution 4(2): 168-183.

Sunkar R, Chinnusamy V, Zhu J, Zhu JK (2007). Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends in Plant Science 12: 301-309.

Svartman M, Stone G, Stanyon R (2005). Molecular cytogenetics discards polyploidy in mammals. Genomics 85(4): 425-430.

Szybalski W (1968). Use of cesium sulfate for equilibrium density gradient centrifugation. Methods in Enzymology 12B: 330–360.

The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74.

Thompson-Stewart D, Karpen GH, Spradling AC (1994). A transposable element can drive the concerted evolution of tandemly repetitious DNA. Proceedings of the National Academy of Sciences 91: 9042-9046.

Thomsen PD, Miller JR (1996). Pig genome analysis: differential distribution of SINE and LINE sequences is less pronounced than in the human and mouse genomes. Mammalian Genome 7: 42-46.

Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfard S, Coffman EJ, Contino G, Deshpande V, Lafrate AJ, Letovsky S, Rivera MN, Bardeesy N, Maheswaran S, Haber DA (2011). Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. Science 331: 593-595.

Trapitz P, Wlaschek M, Bunemann H (1988). Structure and function of Y chromosomal DNA. II. Analysis of lampbrush loop associated transcripts in nuclei of primary spermatocytes of *Drosophila hydei* by *in situ* hybridization using asymetric RNA probes of four different families of repetitive DNA. Chromosoma 96: 159-170.

Trifonov VA, Perelman PL, Kawada SI, Iwasa MA, Oda SI, Graphodatsky AS (2002). Complex structure of B-chromosmes in two mammalian species. *Apodemus peninsulae* (Rodentia) and *Nyctereutes procyonoides* (Carnivora). Chromosome Research 10: 109-116.

Ugarković Đ (2005). Functional elements residing within satellite DNAs. EMBO reports 6(11): 1035-1039.

Ugarković Đ, Plohl M (2002). Variation in satellite DNA profiles-causes and effects. The EMBO Journal 21(22): 5955-5959.

Usakin L, Abad J, Vagin VV, de Pablos B, Villasante A, Gvozdev VA (2007). Transcription of the 1.688 satellite DNA Family Is Under the Control of RNA Interference Machinery in *Drosophila melanogaster* ovaries. Genetics 176: 1343-1349.

Valgardsdottir R, Chiodi I, Giordano M, Cobianchi F, Riva S, Biamonti G (2005). Structural and functional characterization of noncoding repetitive RNAs transcribed in stressed human cells. Molecular Biology of the Cell 16: 2597-2604.

Venner S, Feschotte C, Biémont C (2009). Dynamics of transposable elements: towards a community ecology of the genome. Trends in Genetics 25: 317-323.

Ventura K, Silva MJ, Fagundes V, Christoff AU, Yonenaga-Yassuda Y (2006). Non-telomeric sites as evidence of chromosomal rearrangement and repetitive (TTAGGG)n arrays in heterochromatic and euchromatic regions in four species of *Akodon* (Rodentia, Muridae). Cytogenetic and Genome Research 115: 169-175.

Vermeesch JR, De Meurichy W, Van Den Berghe H, Marynen P, Petit P (1996). Differences in the distribution and nature of the interstitial telomeric (TTAGGG)n sequences in the chromosomes of the Giraffidae, okapi (*Okapia johnstoni*), and giraffe (*Giraffa camelopardalis*): evidence for ancestral telomeres at the okapi polymorphic rob(5; 26) fusion site. Cytogenetics and Cell Genetics 72: 310-315.

Verneau O, Catzeflis F, Furano AV (1998). Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. Proceedings of the National Academy of Sciences 95: 11284-11289.

Veyrunes F, Dobigny G, Yang F, O'Brien PCM, 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. Proceedings of the Royal Society 273: 2925-2934.

Veyrunes F, Watson J, Robinson TJ, Britton-Davidian J (2007). Accumulation of rare sex chromosome rearrangements in the in the African pygmy mouse, *Mus* (*Nannomys*) *minutoides*: a whole-arm reciprocal translocation (WART) involving an X-autosome fusion. Chromosome Research 15: 223-230.

Vitullo P, Sciamanna L, Baiocchi M, Sinibaldi-Vallebona P, Spadafora C (2011). LINE-1 Retrotransposon Copies Are Amplified During Murine Early Embryo Development. Molecular Reproduction and Development 2011: 1-10.

Vogel W, Jainta S, Rau W, Geerkens C, Baumstark A, Correa-Cerro LS, Ebenhoch C, Just W (1998). Sex determination in *Ellobius lutescens*: the story of an enigma**.** Cytogenetics and Cell Genetics 80(1-4): 214-221.

Vorontsov NN, Krjukova EP (1969). *Phodopus przhewalskii* species nova, a new species of desert hamsters (Cricetinae, Cricetidae, Rodentia) from the Zaissan Basin. In: Vorontsov NN (Ed), The mammals. Novosibirsk: Academy of Sciences of the USSR, pp 102-104.

Vourc'h C, Biamonti G (2011). Long Non-Coding RNAs. Progress in Molecular and Subcellular Biology. In: Ugarković Đ (Ed), Transcription of Satellite DNAs in Mammals. Springer-Verlag Berlin Heidelberg, pp 95-118.

Wallace NA, Belancio VP, Deininger PL (2008). L1 mobile element expression causes multiple types of toxicity. Gene 419: 75-81.

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

Walter J, Hutter B, Khare T, Paulsen M (2005). Repetitive elements in imprinted genes. Cytogenetic and Genome Research 113: 109-115.

Warburton PE, Willard HF (1990). Genomic analysis of sequence variation in tandemly repeated DNA. Evidence for localized homogeneous sequence domains within arrays of alpha-satellite DNA. Journal of Molecular Biology 216: 3-16.

Waters PD, Dobigny G, Pardini AT, Robinson TJ (2004). LINE-1 distribution in Afrotheria and Xenarthra: implications for understanding the evolution of LINE-1 in eutherian genomes. Chromosoma 113: 137-144.

Waters PD, Dobigny G, Waddell PJ, Robinson TJ (2007). Evolutionary History of LINE-1 in the Major Clades of Placental Mammals. PLoS ONE 2(1): 1-7.

Waters PD, Dobigny G, Waddell PJ, Robinson TJ (2008). LINE-1 Elements: Analysis by Fluorescence *In-Situ* Hybridization and Nucleotide Sequences. In: Murphy WJ (Ed), Methods in Molecular Biology: Phylogenomics. Totowa: Humana Press Inc, pp. 227-237.

Watts CHS, Baverstock PR (1995). Evolution in the Murinae (Rodentia) assessed by microcomplement fixation of albumin. Australian Journal of Zoology 43: 105-118.

Weiner AM (2002). SINEs and LINEs: the art of biting the hand that feeds you. Current Opinion in Cell Biology 14: 343-350.

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. Journal of Heredity 82: 369-377.

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 A.H (2007). A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8: 973-982.

Wiley JE, Meyne J, Little ML, Stout JC (1992). Interstitial hybridization sites of the (TTAGGG)n telomeric sequence on the chromosomes of some North American hylid frogs. Cytogenetics and Cell Genetics 61: 55-57.

Willard HF, Waye JS (1987). Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. Journal of Molecular Evolution 25: 207-214.

Wong LH, Choo KHA (2004). Evolutionary dynamics of transposable elements at the centromere. Trends in Genetics 20(12): 611-616.

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 Research 17: 1146-1160.

Wurster-Hill DH, Ward OG, Davis BH, Park JP, Moyzis RK, Meyne J (1988). Fragile sites, telomeric DNA sequences, B chromosomes, and DNA content in raccoon dogs, *Nyctereutes procyonoides*, with comparative notes on foxes, coyote, wolf, and raccoon. Cytogenetics and Cell Genetics 49: 278-281.

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 Research 8: 219-227.

Yang Z, Boffelli D, Boonmark N, Schwartz K, Lawn R (1998). Apolipoprotein(a) gene enhancer resides within a LINE element. The Journal of Biological Chemestry 273 (2): 891-897.

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

Zacharias H (2001). Key word: chromosome. Chromosome Research 9: 345-355.

Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH, Verma IM (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature 477(7363): 179–184.

WEBSITES

NCBI genome - http://www.ncbi.nlm.nih.gov/genome/browse/

OneZoom Tree of Life Explorer database - http://www.onezoom.org/index.htm

Tree of Life Web Project - http://www.tolweb.org/tree/