Universitat Autònoma de Barcelona

On the role of chromosomal rearrangements in evolution:
Reconstruction of genome reshuffling in rodents and analysis of Robertsonian fusions in a house mouse chromosomal polymorphism zone

by
Laia Capilla Pérez

A thesis submitted for the degree of Doctor of Philosophy in Animal Biology

Supervisors:
Dra. Aurora Ruiz-Herrera Moreno and Dr. Jacint Ventura Queija

Institut de Biotecnologia i Biomedicina (IBB)
Departament de Biologia Cel·lular, Fisiologia i Immunologia
Departament de Biologia Animal, Biologia Vegetal i Ecologia
Universitat Autònoma de Barcelona

Supervisor	Supervisor	PhD candidate
Aurora Ruiz-Herrera Moreno	Jacint Ventura Queija	Laia Capilla Pérez

Bellaterra, 2015

A la mare

Al pare

Al mano

"Visto a la luz de la evolución, la biología es, quizás, la ciencia más satisfactoria e inspiradora. Sin esa luz, se convierte en un montón de hechos varios, algunos de ellos interesantes o curiosos, pero sin formar ninguna visión conjunta."

Theodosius Dobzhansky

"La evolución es tan creativa. Por eso tenemos jirafas."

Kurt Vonnegut

This thesis was supported by grants from:

- Ministerio de Economía y Competitividad (CGL2010-15243 and CGL2010-20170).
- Generalitat de Catalunya, GRQ 1057.
- Ministerio de Economía y Competitividad. Beca de Formación de Personal Investigador (FPI) (BES-2011-047722).
- Ministerio de Economía y Competitividad. Beca para la realización de estancias breves (EEBB-2011-07350).

Covers designed by cintamontserrat.blogspot.com

INDEX

Abstract 15-17
Acronyms 19-20

1. GENERAL INTRODUCTION 21-60
1.1 Chromosomal rearrangements 23-46
1.1.1 Types of chromosomal rearrangements 23-25
1.1.2 Origin of chromosomal rearrangements: Mechanisms of DNA repair 26-27
1.1.3 DNA sequences associated with CRs 27-32
1.1.3.1 Segmental duplications 28
1.1.3.2 Transposable elements 28-30
1.1.3.3 Tandem repeats 30-31
1.1.3.4 Telomeric repeats 31-32
1.1.4 How to detect CRs among taxa 33-36
1.1.4.1 Comparative cytogenetics 33-34
1.1.4.2 Genetic mapping 34-35
1.1.4.3 Comparative genomics 35-36
1.1.5 Models of ditribution of CRs within genomes 37-42
1.1.5.1 The *random breakage model* 37
1.1.5.2 The *fragile brakage model* 37-38
1.1.5.3 The *intergenic breakage model* 38-39
1.1.5.4 The *integrative breakage model* 39-42
1.1.6 The role of CRs in speciation 42-46
1.1.6.1 The *hybrid dysfunction model* *42-43*
1.1.6.2 The *supressed recombination model* *43-44*
1.1.6.3 *Bateson-Dobzhansky-Muller Incompatibilities* *45-46*
1.2 Meiotic recombination and speciation 48-56
1.2.1 Meiosis 48-50
1.2.2 Variation of recombination rates within genomes 52-53
1.2.3 Variation of recombination rates within and among species 53-56
1.2.4 The role of *Prdm9* gene in meiotic recombination and speciation 56-60

1.3 The western house mouse as a model for chromosomal speciation 56-60
1.3.1 The western house mouse chromosomal variability 57
1.3.2 The Barcelona Rb system 58-59
1.3.3 Previous studies in house mouse Rb populations 59-60

2. OBJECTIVES 61-63

3. MATERIALS AND METHODS 65-90
3.1 Bioinformatic analysis 67-72
3.1.1 Alignment of mammalian genomes 67
3.1.2 Detection of regions of synteny (HSBs and SFs) 68
3.1.3 Detection and classification of EBRs 68-69
3.1.4 Analysis of genomic features 70-71
3.1.4.1 Gene content and ontology 70
3.1.4.2 Recombination rates 71
3.1.4.3 Cosntitutive lamina assocaited domains (cLADs) 71
3.1.5 Permutation tests 72
3.2 Biological samples 72-73
3.3 Molecular biology techinques 73-79
3.3.1 Bacterial artificial chromosomes (BACs) selection and purification 73-75
3.3.1.1 Vector plasmid culture and DNA extraction 73-74
3.3.1.2 DNA labelling by nick translation 75
3.3.2 Genomic DNA purification 75-77
3.3.3 *Prdm9* amplification 77-78
3.3.4 DNA purification from gel bands and Sanger sequencing 79
3.4 Cell biology techniques 80-90
3.4.1 Cell cultures and chromosome harvest 80-81
3.4.2 Spermatocyte spreads 82
3.4.3 Immunofluorescence (IF) 83-84
3.4.4 Fluorescence *in situ* hybridization (FISH) 84-89
3.4.4.1 FISH with chromosome paintings on metaphase chromosomes 85-86
3.4.4.2 FISH with BACs on spermatocytes 86-88
3.4.4.3 Quantitative-fluorescence *in situ* hybridization (Q-FISH) on metaphase chromosomes 88-89
3.4.5 Image processing and analysis 89-90

4. RESULTS 91-163

4.1. RECONSTRUCTION OF GENOME RESHUFFLING IN RODENTS 93-125

4.1.1 Comparative analysis of rodent genomes reveals evolutionary signatures of genome reshuffling 95-125

4.1.1.1 Introduction 95-97

4.1.1.2 Materials and methods 97-101

4.1.1.2.1 Whole genome comparisons 97-98

4.1.1.2.2 Gene content and ontology 99-100

4.1.1.2.3 Recombination rates 100

4.1.1.2.4 Constitutive lamina associated domains (cLADs) 100-101

4.1.2.5 Statistical analysis 101

4.1.1.3 Results 101-109

4.1.1.3.1 Genome reshuffling in Rodentia 101-104

4.1.1.3.2 Rodentia EBRs are gene-rich regions 104-107

4.1.1.3.3 Rodentia EBRs correspond to regions of low recombination rates 108

4.1.1.3.4 Rodentia EBRs are depleted in cLADs 108-109

4.1.1.4 Dicussion 109-112

4.1.1.5 Acknowledgments 112-113

4.1.1.6 Bibliography 113-117

4.1.2 Supplementary information 118-125

4.2. STUDY OF THE ROLE OF TELOMERES IN THE FORMATION OF RB FUSIONS IN THE BARCELONA RB POLYMORPHISM SYSTEM 127-137

4.2.1 On the origin of Robertsonian fusions in nature: evidence of telomere shortening in wild house mice 129-137

4.3. ANALYZING THE ROLE OF RB FUSIONS AND PRDM9 SEQUENCE ON THE MEIOTIC DYNAMICS OF THE BARCELONA RB POLYMORPHISM SYSTEM 139-163

4.3.1 Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and *Prdm9* allelic background 141-149

4.3.2 Supplementary information 150-163

4.3.2.1 Supplementary materials and methods 150-153

4.3.2.1.1 Animals and chromosomal

characterization 150
4.3.2.1.2 Immunofluorescence 150
4.3.2.1.3 Fluorescence *in situ* hybridization 151
4.3.2.1.4 Image processing and data analysis 151-152
4.3.2.1.5 *Prdm9* genotyping 152-153
4.3.2.2 Supplementary figures and Tables 153-162
4.3.2.3 Supplementary references 163

5.GENERAL DISCUSSION 165-184
5.1 Genome reshuffling in Rodentia: causes and consequences of the genomic distribution of EBRs 166-173
5.1.1 Functional constrains 169-172
5.1.2 Chromatin structure: a new player in evolutionary genome reshuffling? 172-173
5.2 The Barcelona Rb polymorphism system as a model for the study of CRs 173-184
5.2.1 The role of telomere shortening on the origin of Rb fusions 174-176
5.2.2 The effect of Rb fusions on fertility 177-184
5.2.2.1 Rb fusions and meiotic recombination 178-181
5.2.2.2 *Prdm9* and the study of chromosomal evolution 181-184

6. CONCLUSIONS 185-188

7.BIBLIOGRAPHY 189-216

Abstract

In order to understand the evolutionary dynamics of mammalian genomes, is necessary to analyze chromosome configuration as well as the genomic changes that have occurred at a large-scale (in the form of chromosomal rearrangements) and at a micro-scale (in the form of nucleotide changes) within species. Chromosomal rearrangements (i.e., inversions, translocations, fusions or fissions) have played a crucial role during evolution as they have led to genomic changes with consequences for the species differentiation. Within mammals, rodents represent the most specious taxon with a wide spectrum of karyotypes.

In this thesis, we have first analyzed the chromosomal reorganizations along rodents evolution together with the factors that have been involved in the distribution of chromosomal rearrangements. Taking advantage of the increasing number of available whole-genomes sequenced, we have compared the genomes of six rodent species (including the mouse genome as a reference) and six outgroup species corresponding to different mammalian taxa (Primates, Artiodactyla, Carnivora and Perissodactyla). We have identified genomic regions of homology (or homologous synteny blocks, HSBs) and the regions of synteny disruption (or Evolutionary Breakpoint regions, EBRs) among rodents. Moreover, the localization of EBRs has permitted us to analyze the genomic features that could be involved in the origin of chromosomal rearrangements. Our results showed that EBRs present a non-homogeneus distribution across the mouse genome. Additionally, EBRs are characterized by specific genomic features such as higher gene content, lower recombination rates and low proportion of lamina associated domains (cLADs) compared with the rest of the mouse genome.

Secondly, it is known that the western house mouse (*Mus musculus domesticus*) natural populations present a wide variety of diploid numbers due to the presence of Robertsonian (Rb) fusions. Within all these populations analyzed, one of them, localized in the Barcelona, Lleida and Girona provinces, presents a specific structure, where no metacentric race has been described, being the Rb fusions found in a polymorphic state. This chromosomal polymorphism zone is known as The Barcelona Rb system. Giving the specific characteristics of this population, we have: (i) analyzed the role of telomeres in the occurrence of the Rb fusions and (ii) studied the effect of the Rb fusions and *Prdm9* gene on meiotic recombination. We have detected that telomere shortening in acrocentric p-arms can be one of the factors that could explain the occurrence of Rb fusions by promoting the interaction between chromosomal ends and thus, to the fusion events. Moreover, we have

observed that the presence of Rb fusions leads to a decrease in recombination rates due to a re-distribution of crossovers towards the telomeres in metacentric chromosomes. Furthermore, we have detected that this phenomenon is due to an interference effect of the centromere in metacentric chromosomes, which acts suppressing recombination within the pericentromeric regions. Additionally, we have also characterized the *Prdm9* allelic distribution within the Barcelona Rb polymorphism system, as well as an effect of the *Prdm9* sequence on recombination rates.

Therefore, and in the light of our results, we propose that the effect of suppression of recombination on individuals with Rb fusions is due to a mechanicstic (by the centromeric interference effect) and genetic (the *Prdm9* allelic sequence) factors. These results, together with the characterization of the genomic features that have been involved in the occurrence of evolutionary chromosomal rearrangements in rodents, would help us to understand the dynamics of chromosomal speciation along evolution and how chromosomal rearrangements occur in natural populations.

Resum

Per tal de poder entendre la dinàmica evolutiva dels genomes de mamífers és necessari estudiar l'estructura dels cromosomes i quins han estat els canvis que s'han donat tant a gran escala (en forma de reorganitzacions cromosòmiques) com a petita escala (en forma de mutacions d'un sol nucleòtid). Les reorganitzacions cromosòmiques han jugat un paper crucial en el procés evolutiu ja que els canvis a què han donat lloc (inversions, translocacions, fusions o fissions) han provocat canvis genòmics amb conseqüències per a la diferenciació de les espècies. Dins dels mamífers, a més, els rosegadors són el grup que presenta més diversitat d'espècies amb un ampli espectre de cariotips.

En aquesta tesi, en primer lloc hem estudiat les reorganitzacions cromosòmiques al llarg de l'evolució dels rosegadors així com els factors que han condicionat la distribució genòmica de les reorganitzacions. Gràcies a la creixent disponibilitat de genomes seqüenciats, hem pogut comparar els genomes de sis espècies de rosegadors (incloent com a referència el del ratolí) i de sis *outgrups* corresponents a diferents taxons dins dels mamífers (Primats, Carnivora, Artiodactyla i Perissodactyla). Amb aquest anàlisi hem identificat les regions d'homologia (o *Homologous Synteny Blocks*; HSBs) i les regions de disrupció o trencament de l'homologia (o *Evolutionary Breakpoint Regions*; EBRs) que s'han donat al llarg de l'evolució dels rosegadors. La localització de les EBRs ens ha permès, doncs, estudiar les característiques

genòmiques de les regions d'inestabilitat que han donat lloc a les reorganitzacions cromosòmiques en el genoma de ratolí i que s'han donat al llarg de l'evolució dels rosegadors. Els nostres resultats mostren que les EBRs presenten una distribució no-homogènea i es caracteritzen per presentar un alt contingut gènic, taxes de recombinació meiòtica més baixes i una proporció més baixa de *lamina associated domains* (cLADs), comparant amb la resta del genoma de ratolí.

D'altra banda, està descrita la tendència de poblacions naturals de ratolí domèstic occidental (*Mus musculus domesticus)* a presentar una gran diversitat de nombres diploides a causa de l'aparició de fusions Robertsonianes (Rb). De totes les poblacions estudiades fins al moment, n'hi ha una, localitzada a les províncies de Barcelona, Lleida i Girona, que presenta una estructura diferent a la resta ja que no existeix en ella una raça metacèntrica i les fusions Rb detectades es troben en un estat de polimorfisme, previ a la seva fixació: Aquesta zona de polimorfisme cromosòmic es coneix amb el nom de systema Robertsonià de Barcelona.

En segon lloc, donades les característiques úniques d'aquest sistema Robertsonià hem analitzat: (i) el paper que podien jugar els telòmers en l'aparició de les fusions Rb i (ii) l'efecte d'aquestes fusions i del gen *Prdm9* sobre la recombinació meiòtica. En el nostre treball hem detectat que l'escurçament telomèric dels braços p dels cromosomes acrocèntrics podria ser un dels factors que explicarien la aparició de les fusions Rb ja que afavoririen la seva interacció i la posterior fusió entre ells.

D'altra banda, em vist que la presència de les fusions Rb provoca una baixada en la taxa de recombinació que es deu a una redistribució dels *crossovers* (o punts de recombinació homòloga) cap al telòmers en els cromosomes fusionats. Hem detectat, a més, que aquest fenomen es deu a un efecte d'interferència del centròmer que suprimeix la recombinació a la zona pericentromèrica. A més, també hem analitzat la distribució allèlica del gen *Prdm9* en el sistema Robertsonià de Barcelona, així com un possible efecte de la seqüència d'aquest gen sobre la taxa de recombinació.

Per tant, donats els nostres resultats, proposem que l'efecte de supressió de la recombinació que hem detectat en els individus amb fusions Rb és a causa d'un factor mecànic (com es l'efecte d'interferència centromèrica) i d'un factor genètic (com es la seqüencia al·lèlica del gen *Prdm9*). Aquests resultats aporten una informació esencial per entendre la dinàmica dels processos d'especiació cromosòmica en poblacions naturals.

Acronyms

ACs - Astrocytes
Alt-EJ - Alternative End Joining
ATM - Ataxia Telangiectasia Mutated
ATR - ATM-Rad3-Related
BAC - Bacterial Artificial Chromosome
BDMIs – Bateson-Dobzhansky-Muller incompatibilities
BSA - Bovine Serum Albumin
CHORI - Children's Hospital Oakland Research Institute
cLAD - Constitutive Lamina Associated Domain
cM - Centimorgan
CO - Crossover
CR - Chromosomal Rearrangements
CREST – Complication of Raynaud's phenomenon, Esophageal dysfunction, Sclerodactily and Telangiectasia
CTCF - CCCCTC-binding factor
CT - Chromosomal Territory
Cy3 - Cyanine 3
Cy5 - Cyanine 5
DAPI - 4 ',6-diamino-2-fenilindol
DAVID - Database for Annotation, Visualization and Integrated Discovery
dHJ - Double holliday Junction
DIG - Digoxigenin
DIRS - Dictyostelium Intermediate Repeat Sequence
DMC1 - Disrupted Meiotic cDNA 1
DMEM - Dulbecco's Modified Eagle Medium
DMSO - Dimethyl Sulfoxide
dNTP - Deoxynucleotide Triphosphate
DSB - Double Strand Break
dUTP - Deoxyuridine Triphosphate
dUTP-DIG – dUTP Digoxigenin
dUTP-Cy3 – dUTP Cyanine 3
EASE - Expression Analysis Systematic Explorer
EBA - Evolutionary Breakpoint Analyzer
EBR - Evolutionary Breakpoint Region
EDTA - Ethylenediamine-Tetraacetic Acid
EH – Evolution Highway Comparative Chromosome Browser
EtBr – Ethidium Bromide
FBS - Foetal Bovine Serum
FISH - Fluorescence *in situ* Hybridization
FITC - Fluorescein Isothiocianate
FoSTeS - Fork Stalling and Template Switching
GCD – Germ Cell Death
GEO - Gene Expression Omnibus
GO - Gene Ontology
H3K4 – Histone 3 methylated at lysine 4
H3K9me3 - Histone 3 trimethylated at lysine 9
HSB - Homologous Synteny Block
Hst1 - Hybrid sterility 1
IF- Immunofluorescence
ITS - Interstitial Telomeric Sequences
Kbp - Kilobase pairs
KRAB – Krueppel-Associated Box
LB - Luria-Bertani
Lcn - Lipocalin
LCRs - Low Copy Repeats
LD - Linkage Disequilibrium

LINEs - Long Interspersed Nuclear Elements
LTR - Long Terminal repeat
LTR-ERV 1 - Long Terminal Repeat Endogenous Retrovirus 1
MajSat - Major Satellite DNA
Mbp – Megabase pairs
MEFs - Mouse Embryonic Fibroblasts
Meisetz - Meiosis induced factor containing a PR/SET domain and zinc finger motif
MinSat - Minisatellite DNA
MLH1 - MuTL Homolog 1
MLH3 - MuTL Homolog 3
MMBIR - Microhomology-Mediated Break-Induced Repair
MMEJ - Microhomology-Mediated End Joining
MSH4 - MutS Homologs 4
MSH5 - MutS Homologs 5
Myr - Million Years
NAHR - Non-Allelic Homologous Recombination
NCO - Non-Crossover
NHEJ - Non-Homologous End Joining
NL - Nuclear Lamina
NPCs - Neural Precursor Cells
Ods - Odysseus
Okt1Ko MEFs - Oct1 Knockout Mouse Embrionic Fibroblasts
Ovd - Overdirve
PBS - Phosphate Serum Saline
PLE - Penelope-Like Element
PNA – Peptide Nucleic Acid
Prdm9 - PR domain containing 9
Q-FISH - Quantitative Fluorescence *in situ* Hybridization
QTL - Quantitative Trait Locus
Rad51 - Radiation sensitive 51
Rb - Robertsonian
RPA - Replication Protein A
Rpm- Revolutions per minute
SC - Synaptonemal Complex
SD - Segmental Duplications
SDSA - Synthesis-Dependent Strand Annealing
SF - Syntenic Fragment
SINEs - Short interspersed Nuclear Elements
SNP - Single Nucleotide Polymorphism
SS - Satsuma Synteny
SSA - Single Strand Annealing
SSC - Saline Sodium Citrate
ssDNA - single-stranded DNA
ST - Synteny Tracker
SYCP1 - Synaptonemal Complex Protein 1
SYCP2 - Synaptonemal Complex Protein 2
SYCP3 - Synaptonemal Complex Protein 3
TAD - Topological Associated Domain
TE - Transposable Element
TFU – Telomere Fluorescence Units
TIR - Terminal Inverted Repeats
tL1 – Minisatellite L1
TLCSat - Telocentric satellite
TR - Tandem Repeats
WART - Whole-Arm Reciprocal Translocation
ZnF - Zinc finger
γH2AX - Phosphorylation of Histone H2AX on serine 139

1.

GENERAL INTRODUCTION

Understanding the mechanisms by which speciation takes place has been a topic of discussion since Charles Darwin proposed its revolutionary theory of the origin of the species in 1859 to explain the diversity of the living world ("*On the Origin of Species*", 1859) and setting the bases for the modern evolutionary biology. In these early days, species were seen as arbitrary constructs made by humans (Darwin 1859). However, this discussion seemed to be unproductive until (Mayr 1942) proposed the "Biological Species Concept", defining species as "groups of actually or potentially inbreeding natural populations, which are reproductively isolated from other such groups". Mayr, along with T. Dobzhansky (Dobzhansky 1937; 1951), defined the modern synthesis of evolutionary theory, integrating Mendel's laws of inheritance with Darwin's view by placing the emphasis on the genetic relationships among populations. Thus, what reproductive isolation consists of and which are its genetic consequences? T. Dobzhansky (1937, 1951) greatly contributed to our understanding of the mechanisms that are involved in reproductive isolation (see BOX 1) and, with the "Biological Species Concept" as the center of discussion, the first models of speciation were focused on elucidating the mechanisms underlying speciation barriers.

1.1 Chromosomal rearrangements

Early observations on the correlation between karyotype diversity and the presence of Chromosomal Rearrangements (CRs) in different taxa (White 1973), triggered the development of different speciation models involving CRs in order to explain the evolution of species. CRs constitute the basis of genome variability, since they contribute to genome reshuffling, providing new chromosomal forms on which natural selection can act. As Peng and collaborators (2006) posted it, CRs "comprise evolutionary "earthquakes" that dramatically change de landscape". This is well exemplified by mammals, a taxonomic group characterized by a high diversity of karyotypes, with diploid numbers ranging from 2n=6 in the Indian Muntjak female (*Muntiacus muntjak*) (Wurster and Benirschke 1970) to 2n=102 in the case of the Red vizcacha rat (*Octomys mimax*) (Contreras et al. 1994).

1.1.1 Types of chromosomal rearrangements

Chromosomal rearrangements consist of large-scale reorganizations that can occur within or among chromosomes. They encompass different types of events

affecting from a few Kilobase pairs (Kbp) to whole chromosomes, being generally classified as balanced or unbalanced CRs (Griffiths et al. 1999).

Balanced CRs include those rearrangements that do not change the final gene dosage and include inversions, reciprocal translocations, fusions and fissions (Figure 1.1). Inversions occur when an internal region of a chromosome changes its orientation with respect to flanking regions and can involve the centromere (pericentric inversion, Figure 1.1A) or not (paracentric inversion). Reciprocal translocations, on the other hand, consist of the exchange of two genomic regions between two different chromosomes (Figure 1.1B). This rearrangement can also be unbalanced, when one chromosomal region is inserted into a different chromosome (non-reciprocal translocation) (Griffiths et al. 1999).

BOX 1: Barriers that contribute to reproductive isolation

Reproductive barriers can be initiated by divergent selection (ecological or sexual selection) or by genetic drift (as a direct consequence of evolution or through the accumulation of genetic incompatibilities) (Dobzhansky 1937; Dobzhansky 1951; Coyne and Orr 2004) and include:

Prezygotic isolation barriers: They act before fertilization of the oocyte.

Premating isolation: Barriers act before the transference of the sperm to the oocyte and comprise:

- Behavioral isolation: The mechanism by which there is a lack of cross-interaction between individuals.
- Ecological isolation: Based on the ecological differences between species that are products of the adaptation to the local environment, and this can be:
  - Habitat isolation: The species occupy different habitats during the mating period.
  - Temporal isolation: The species mating period are developed at different times.
- Mechanical isolation: Incompatibility between reproductive structures that avoid the mating process.

Postmating isolation: Mechanisms that act after the sperm transfer but before fertilization.

- Copulatory behavioral isolation: The behavior of an individual during copulation avoids fertilization.
- Gametic isolation: Sperm cannot fertilize the oocyte.

Postzygotic isolation barriers: They act after the fertilization of the oocyte.

Extrinsic postzygotic isolation: This mechanism depends on the environment conditions (either biotic or abiotic).

- Ecological inviability: Hybrids develop normally but they suffer low viability, as they cannot find an appropriate ecological niche to develop.
- Behavioral sterility: Hybrids have normal gametogenesis but their reproductive efficiency is low, as they cannot find appropriate mates.

Intrinsic postzygotic isolation

- Hybrid inviability: Hybrids suffer developmental difficulties causing full or partial lethality.
- Hybrid sterility: Hybrids present meiotic problems that affect to the proper development of gametes.

Chromosomal fusions occur when two non-homologous chromosomes are combined to form a new chromosome. One of the most common type of chromosomal fusion found in nature is the so-called centric fusion or Robertsonian (Rb) fusion (Robertson 1916). In this later case, two acrocentric chromosomes fuse by its centromeric region following the breakage within the microsatellite sequences, leading to the formation of single metacentric chromosome (Figure 1.1C). This type of rearrangement does not affect gene dosage although they can induce variation in diploid numbers. Finally, chromosomal fissions also lead to changes in chromosome numbers by transforming one initial chromosome into two (see the example of centric fission, Figure 1.1D).

Unbalanced CRs, on the other hand, encompass the type of rearrangement where gene dosage is altered, either as a gain (duplication) or loss (deletion) of genetic material. Duplicated regions can be localized adjacent to each other (i.e., tandem duplication), in a novel location within the same chromosome, or in another chromosome (i.e., insertional duplication) (Griffiths et al. 1999). Deletions, on the contrary, occur when chromosomal fragments or whole chromosomes are lost (Griffiths et al. 1999).

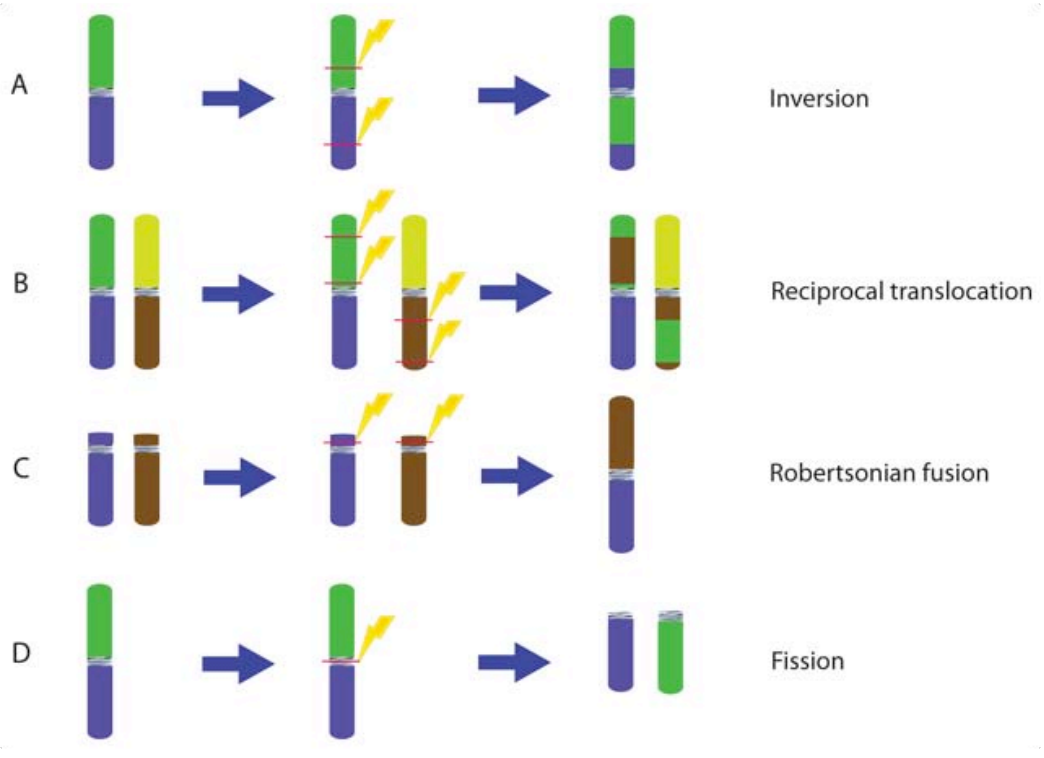


Figure 1.1. Representation of the different types of CRs: For each type of rearrangement, the initial state of the chromosome, the localization of chromosomal breakages and the resulting rearranged chromosome are represented for an (A) pericentric inversion, (B) reciprocal translocation, (C) Robertsonian fusion and (D) fission.

1.1.2 Origin of chromosomal rearrangements: mechanisms of DNA repair

Chromosomal rearrangements that are shaping mammalian genomes are originated by Double Strand Breaks (DSBs). This type of lesion is generated either by exogenous factors (ionizing radiation and/or chemical agents), endogenous agents (free radicals or stall of the replication fork) or by highly specialized cellular processes such as meiosis. In all cases, DSBs are repaired through different mechanisms, as it will be explained below. When any of these mechanisms fail, DSBs are ill repaired, which can lead to cell death or large-scale CRs that enhance genome instability. If these new chromosomal forms are produced in the germ line, they will have the probability to be fixed in the population providing new variability and eventually contributing to the formation of new species.

DNA repair mechanisms are classified into two main groups, depending on the type of sequences (homologous or non-homologous) used as template (see Onishi-Seebacher and Korbel 2011, and references therein). DNA repair mediated by homologous sequences include two mechanisms: (i) Single Strand Annealing (SSA) (Lee 2014) and (ii) Non-Allelic Homologous Recombination (NAHR) (Sasaki et al. 2010). The mechanisms mediated by non-homologous regions or regions of microhomology are more diverse and comprise: (i) Non Homologous End Joining (NHEJ) (Lieber 2010), (ii) Alternative End Joining (Alt-EJ) (Frit et al. 2014), (iii) Microhomology-Mediated End Joining (MMEJ) (McVey and Lee 2008), (iv) Microhomology-Mediated Break-Induced Repair (MMBIR) (Yu and McVey 2010), (v) Fork Stalling and Template Switching (FoSTeS) (Hastings et al. 2009) and (vi) chromothripsis (Stephens et al. 2011; Kloosterman et al. 2012).

Within the DNA repair mechanisms that are mediated by homologous sequences, SSA occurs when two homologous sequences are incorrectly annealed (normally repetitive sequences) without involving DNA strand invasions, and normally, leads to deletions (Heyer et al. 2010). Regarding NAHR or ectopic recombination, this takes place by three different pathways, depending on the relative position of the homologous sequences. In the first case, when the sequences that recombine are adjacent with the same orientation, the result is either a duplication or deletion. Conversely, in the case where homologous regions are found in non-homologous chromosomes, it results in a translocation. Finally, if the homologous sequences are found in the same chromosome but with inverted orientations, the result is an inversion (reviewed in Bailey and Eichler 2006).

Regarding non-homology or microhomology-mediated mechanisms, NEHJ constitutes a DSBs repair pathway that involves the joining of two free ends through a process that is independent of sequence homology and, therefore, produce junctions that can vary in their sequence composition (between 1bp and 4bp in size) (Takata et al. 1998; Hefferin and Tomkinson 2005). Alt-EJ and MMEJ constitute alternative repair pathways of DSBs free ends by microhomology that differ from NEHJ by the size of the region involved (between 5bp to 25bp in this case) (Hastings et al. 2009). Additionally MMBIR/FoSTeS mechanisms have been proposed more recently in order to explain the occurrence of some types of rearrangements that could not be explained by any of the previously mentioned pathways and would be involved in genomic instability related with specific human diseases (Lee et al. 2007; Zhang et al. 2009). This mechanism takes place during DNA replication. In this case, the replication fork can stall and switch different templates using complementary template microhomology to anneal and prime DNA replication. This mechanism involves the insertion of small portions of non-homologous regions within the replication fork (Hastings et al. 2009).

Finally, chromothripsis (also known as chromosome shattering) refers to the large amount of genomic breakages observed in several cancers and human genomic disorders (Stephens et al. 2011). This phenomenon is considered as a chromosomal "disaster" and was defined on the basis of three main characteristics: (i) the occurrence of remarkable numbers of rearrangements in localized chromosomal regions, (ii) a low number of copy number states across the rearranged region; and (iii) alternation in the chromothriptic areas of regions where heterozygosity is preserved with regions presenting loss of heterozygosity (Stephens et al. 2011; Forment et al. 2012).

1.1.3 DNA sequences associated with CRs

Specific DNA features have been associated with the occurrence of CRs, either as the mechanism of origin or as the result of the reorganization. This is the case of repetitive elements and telomeric repeats. Three types of repetitive sequences have been described to play a role in genome reshuffling: Segmental Duplications (SDs), Transposable Elements (TEs) and Tandem Repeats (TRs).

1.1.3.1 Segmental Duplications

Segmental Duplications (SDs), also known as Low Copy Repeats (LCRs), consist of duplicated DNA fragments larger than 1Kbp in size that spread across the genome. SDs present a non-random genome distribution, clustering at subtelomeric and pericentromeric regions (Bailey and Eichler 2006; She et al. 2008), although they can also be present in euchromatic regions of specific chromosomes (She et al. 2004). They show >90% of sequence homology (Bailey and Eichler 2006), which explain the propensity of these regions to promote NAHR. However, there is evidence that support the involvement of SDs in homology-independent DNA repair mechanisms, such as NHEJ (Koszul and Fischer 2009; Quinlan et al. 2010).

Given their repetitive nature, SDs are an important source of genomic instability by birthing new genes, favoring functional diversification and expansion of gene families (Newman et al. 2005; Wilson et al. 2006). SDs can encode protein products that, although are not necessarily essential for the viability of the organism, can be relevant for species adaptation (Duda and Palumbi 1999; Chang and Duda 2012). In fact, commonly duplicated genes in mammals include those associated with innate immunity, digestion, drug detoxification, olfaction, and sperm competition (Beckmann et al. 2007; Liu et al. 2009).

Moreover, SDs are also involved in the occurrence of CRs implicated in human disorders (Lupski 1998; Pollack et al. 2002; Sharp et al. 2006; Uddin et al. 2011) and chromosomal evolution (Murphy et al. 2005b; Elsik et al. 2009). The correlation between SDs and evolutionary CRs has been described in some mammalian species (Murphy et al. 2005b), such as mouse (Armengol et al. 2003; Bailey et al. 2004; Armengol et al. 2005), rat (Zhao and Bourque 2009), dog (Zhao and Bourque 2009) and primates (Carbone et al. 2006; Kehrer-Sawatzki and Cooper 2008b; Girirajan et al. 2009; Marques-Bonet et al. 2009; Capozzi et al. 2012; Carbone et al. 2014). Overall, nearly 40% of evolutionary CRs have been found to be associated with SDs in mammals (Bailey and Eichler 2006).

1.1.3.2 Transposable Elements

Transposable Elements (TEs) were discovered by Barbara McClintock (1984) in maize and since then, they have been found in almost all organisms, from prokaryotes to eukaryotes (Capy 1998). They are classified into two large families:

(i) retrotransposons (class I elements) and (ii) DNA transposons (class II elements) (Richard et al. 2008).

Retrotransposons, also known as class I elements, form a group of TEs integrated by a variety of different repeats (Wicker et al. 2007). They are characterized by the absence of introns, an adenine-rich tail and direct repeats localized in the adjacent regions, that are produced by transposition, and include: Long Terminal Repeats (LTRs), Long Interspersed Nuclear Elements (LINEs), Short Interspersed Nuclear Elements (SINEs) (that includes *Alu* elements described in primates), Dictyostelium Intermediate Repeat Sequence (DIRS-like elements) and Penelope-Like Elements (PLE) (Wicker et al. 2007). Their mechanism of transposition consists of the retro-transcription of its sequences to RNA that is subsequently transformed to DNA by an inverse transcriptase, a process that can be auto regulated by the retrotransposon itself. On the other hand, class II elements, or DNA transposons, consist of mobile DNA repeats that codify for a transposase that split from the DNA sequence and use single strand breaks or DSBs to transpose itself in a new genomic region by repairing the breaks, using the sister chromatid or homologous chromosomal region as template (Wicker et al. 2007). This group includes several types of transposons such as Terminal Inverted Repeats (TIR), Crypton, Elitron and Maverick repeats (Wicker et al. 2007).

Due to their interspersed sequence homology, TEs would favor genome instability involved both in genomic disorders and species adaptation, mirroring what has been described for SDs (Deininger and Batzer 1999; Schibler et al. 2006; Lee et al. 2008; Kehrer-Sawatzki and Cooper 2008b; Longo et al. 2009; Farré et al. 2011). Additionally, they also play an important role in environmental adaptation by facilitating the acquisition of new genetic material (reviewed in Casacuberta and González 2013). The formation of CRs triggered by TEs can be due to the result of either indirect (homologous recombination) or direct DNA repair mechanisms (alternative transposition process) (Wicker et al. 2007; Richard et al. 2008). In the first case, CRs can result from recombination between non-allelic sequences between TEs interspersed across the genome, either by NAHR or SSA (Sen et al. 2006). NAHR is thought to be the source of many TE-related instability events, particularly those resulting in duplications and translocations (Elliott et al. 2005). In humans, most of the *Alu*-mediated deletion events are consistent with either NAHR or SSA mechanisms (Sen et al. 2006). This has been the case also for *Saccharomyces*, where recombination events between transposons have been described probably due to their high sequence homology (Fischer et al. 2000).

Transposable elements, as with SDs, have the capacity to influence genome

plasticity. This can be done, for example, by (i) the alteration of the gene function and regulation, (ii) contributing to the creation of new genes, and (iii) inducing CRs (see Feschotte and Pritham 2007; Cordaux and Batzer 2009, for reviews). In fact, genome reshuffling triggered by TEs have been extensively recorded both in plants and animals (Walker et al. 1995; Delprat et al. 2009). In the case of mammals, this is exemplified by primates (Kehrer-Sawatzki and Cooper 2008a; Lee et al. 2008; Farré et al. 2011; Carbone et al. 2014), Cetartyodactyla (Groenen et al. 2012) and marsupials (Longo et al. 2009), where lineage-specific CRs have been found to be enriched in TEs.

1.1.3.3 Tandem Repeats

Tandem Repeats (TRs) consist of a large series of repeated sequences sequentially distributed or clustered in specific regions of the genome. They follow a head-to-tail-fashion distribution, being also called "direct repeats" (Richard et al. 2008). TRs can be classified into two groups: (i) minisatellites, which contain repeat units larger than 7bp, and (ii) microsatellites, with repeat units from 1bp to 6bp (Näslund et al. 2005).

Along with TEs, TRs have been reported as an important source of DNA variation and mutation (Armour 2006). TRs can form different secondary DNA structures from the Watson-Crick classical conformation such as hairpins, cruciform or triplex conformations that would promote genomic instability (Bacolla et al. 2008; Kolb et al. 2009). These DNA structures are associated with DNA replication malfunctions and meiotic recombination (Usdin and Grabczyk 2000; Shaw and Lupski 2004; Kelkar et al. 2008). Expansions of the repeat array occur when an unusual secondary structure is formed in the lagging daughter strand during DNA replication. Deletions, on the other hand, occur when an unusual configuration develops in the template for lagging-strand DNA synthesis (Usdin and Grabczyk 2000). Thus, TRs are relevant facilitators of CRs involved in many diseases (Campuzano et al. 1996; Usdin and Grabczyk 2000). From an evolutionary perspective, TRs were initially shown to be concentrated in evolutionary chromosomal regions in the human genome (Ruiz-Herrera et al. 2006) as well as in primate-specific chromosomal breakpoint regions (Farré et al. 2011). In primates, certain TRs (that presents AAAT motif) are enriched in *Alu* elements, suggesting the involvement of these elements in primate evolution (Farré et al. 2011). However, given the AAAT motif is similar to the insertion motif 5'-TTAAA-3' (Jurka 1997) and that *Alu* elements are capable of insertion at target

sites that are slightly different (although always AT-rich) (Levy et al. 2010), the AAAT motif could be site-specific for *Alu* insertion in primate evolutionary breakpoints. This is consistent with the observations that young *Alu* families were located in AT-rich regions in the human genome (Kvikstad and Makova 2010), or the presence of LTR Endogenous Retrovirus 1 (LTR-ERV1), satellite repeats and tRNAGlu-derived SINEs in certartiodactyla-specific EBRs (Groenen et al. 2012).

1.1.3.4 Telomeric repeats

Telomeric repeats have been also described to be involved in genome reshuffling (Nergadze et al. 2004; Ruiz-Herrera et al. 2008). Telomeres are specialized ribonucleoprotein structures localized at the end of chromosomes, composed of tandem TTAGGG repeats, bound to specific proteins forming the shelterin complex (Moyzis et al. 1988; de Lange 2005) and associated with specific non-coding telomeric RNA molecules called TERRA (Azzalin et al. 2007; Schoeftner and Blasco 2008; Reig-Viader et al. 2014). Telomeres protect chromosomal ends to be recognized as DSBs or "free ends" and thus prevent the formation of chromosomal fusions (Zakian 1997), recombination between telomeres and DNA erosion (Li et al. 1998). Telomeres balance the end-replication problem, which consists of telomeric DNA shortening in each round of replication caused by the inability of polymerases to replicate DNA. This process can be reverted by the action of either a recombination-based mechanism or an enzymatic strategy based on the action of telomerase, a ribonucleoprotein DNA polymerase that adds telomeric repeats at chromosomal ends (O'Sullivan and Karlseder 2010).

Given the importance of telomeres in maintaining genome integrity, alterations of their structure are often related to the occurrence of CRs, such as Rb fusions (Blasco et al. 1997; Slijepcevic et al. 1997; Slijepcevic 1998). In this context, the formation of Rb fusions would require either the elimination or inactivation of telomeres prior to the reorganization (Ruiz-Herrera et al. 2008). Moreover, provided that Rb fusions are extended in natural populations from several mammalian species representative of bovids, cervids and rodents (Piálek et al. 2005; Pagacova et al. 2009; Gauthier et al. 2010; Robinson and Ropiquet 2011; Aquino et al. 2013), the study of telomere dynamics represents an important challenge for the understanding of the role of Rb fusions in speciation. In this context, three different mechanisms have been proposed to explain the occurrence and fixation of Rb fusions in natural populations in relation to the presence/absence of telomeric sequences in the resultant metacentric

chromosome (Slijepcevic 1998): (i) telomere shortening, (ii) chromosome breakage within minor satellite sequences and (iii) telomere inactivation (Figure 1.2).

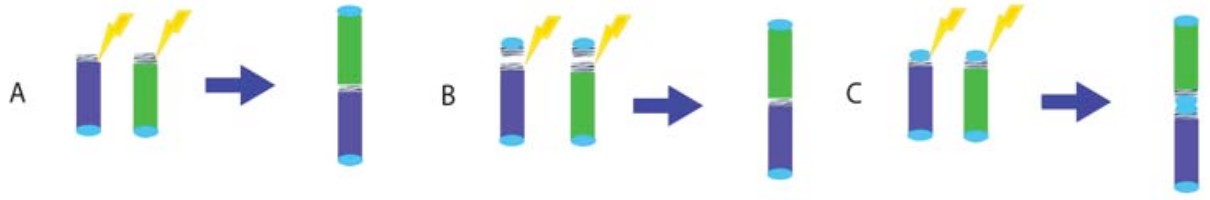


Figure 1.2. Types of Rb fusions and their mechanisms of origin: Telomeres are depicted in blue. (A) Rb fusion by telomere shortening (B) Rb fusion by chromosome breakage within minor satellite sequences and (C) Rb fusion by telomere inactivation. Adapted from Slijepcevic (1998).

Initial studies regarding the role of telomere shortening in the origin of Rb fusions originated from laboratory observations using telomerase knockout mice as models (Blasco et al. 1997). Blasco and collaborators (1997) observed a high degree of chromosome fusions in tumor cell cultures that presented telomerase inactivation (Blasco et al. 1997), a process also observed in humans (Harley et al. 1990; Allsopp et al. 1992) and *Mus spretus* primary cells (Prowse and Greider 1995). These observations suggest that short telomeres would lose the capacity to protect chromosomes from fusions (Figure 1.2A). Despite such evidence, whether this phenomenon occurs in natural populations remains to be tested.

Robertsonian fusions can also occur when telomeric sequences are lost due to chromosomal breakage within minor satellite sequences at the centromere (Figure 1.2B). This phenomenon was proposed to occur in wild-derived house mouse (Garagna et al. 1995; Nanda et al. 1995; Garagna et al. 2001). In this case, breakage leads to the loss of both telomeres and satellite repeats and, in both cases, the absence of telomeric repeats could promote ill DNA repair that could lead to the joining of two acrocentric or telocentric chromosomes, leading to the formation of a metacentric (Figure 1.2B).

And finally, it has been reported that Rb fusions can occur without the loss of telomeric repeats (Figure 1.2C). The observation of chromosomes of many vertebrates with the presence of Interstitial Telomeric Sequences (ITSs) supports this view (Meyne et al. 1990; Ruiz-Herrera et al. 2008; Bruschi et al. 2014). In this case, the presence of telomeric repeats at fusion regions would be inconsistent with the presence of functional telomeres, as Slipjecevic suggested (Slijepcevic 1998).

1.1.4 How to detect CRs among taxa

The study of genome organization and the mechanisms by which CRs are involved in speciation and adaptation are both key points to better understand the evolutionary dynamics of genomes. In order to investigate the causes and evolutionary consequences of CRs, different approaches have been developed to study genome reshuffling among taxa.

1.1.4.1 Comparative cytogenetics

Since the first human karyotype was described contemporarily by two research groups (Ford and Hamerton 1956; Tjio and Levan 2010), relevant technological advances have been achieved in the study of CRs and karyotype diversity. Initial cytogenetic approaches in the late 1960s were focused on banding patterns comparisons (G- and R- banding) that permitted to identify and differentiate chromosomes between species (Caspersson et al. 1970; de Grouchy et al. 1972; Egozcue et al. 1973a; Egozcue et al. 1973b; Lentzios et al. 1980). This was the case of early karyotype comparisons in great apes (Egozcue et al. 1973a; Egozcue et al. 1973b; Dutrillaux et al. 1975; Yunis and Prakash 1982; Clemente et al. 1990). The subsequently development of molecular cytogenetics techniques, such as Fluorescence *in situ* Hybridization (FISH), allowed researchers to refine the resolution at which CRs can be identified among different species. This approach relies on the use of DNA probes corresponding to whole chromosomes (chromosome painting probes) that can be hybridized on chromosome preparations of target species (Scherthan et al. 1994). In this way, it is possible to establish conserved chromosomal regions between two given species, based on DNA sequence homology. Such information, framed in an evolutionary context (i.e., by using the appropriate outgroups), has allowed the identification of the evolutionary direction of CRs. In fact, using whole-chromosomal paintings, Zoo-FISH has been used to compare a large amount of mammalian species representative of the major phylogenetic groups (Chowdhary et al. 1998; Froenicke et al. 2005). These species included primates (e.g., Stanyon et al. 2000; 2004; García et al. 2000; Ruiz-Herrera et al. 2002; Ventura et al. 2007; Kehrer-Sawatzki and Cooper 2008b; Girirajan et al. 2009), rodents (Veyrunes et al. 2006; Graphodatsky et al. 2008; Trifonov et al. 2010; Di-Nizo et al. 2015), afrotherian (Yang et al. 2003; Frönicke et al. 2003; Pardini et al. 2007), equids (Pawlina and Bugno-Poniewierska 2012; Musilova et al. 2013), carnivores (Perelman et al. 2012)

and birds (Guttenbach et al. 2003; Griffin et al. 2008; de Oliveira et al. 2010). In this way, three different types of chromosome conserved regions can be identified: (i) whole chromosomes that remain intact in the ancestral karyotype as a single block, (ii) fragments of chromosomes corresponding to a single chromosome ancestor, and (iii) chromosomal associations (i.e., synteny), that is, chromosomal fragments or whole chromosomes that are now present in independent chromosomes, but were represented as a single chromosomal form in a recent common ancestor.

But, despite these methodological advances, molecular cytogenetic methodologies present some limitations, as FISH probes can only hybridize between closely related species and the lack of resolution that it presents does not allow for the detailed characterization of chromosomal breakpoints, at a sequence level, in distantly related species. Thus, additional comparative approaches have been developed to fill this gap.

1.1.4.2 Genetic mapping

Genetic mapping has been broadly used to identify chromosomal regions of homology among species. It is based on the experimental obtaining of the gene order between two species assuming that its conservation corresponds to regions of homology. Genetic maps can follow two different strategies: (i) high-density molecular linkage maps or (ii) radiation hybrid maps.

On one hand, the characterization of the relative position of two *loci* based on the frequency they are observed together in gametes is the approximation used in linkage maps. If this frequency is high, this indicates that loci are closely located. Conversely, low frequencies indicate distant positions between *loci*. This approach was initially used to define regions of homology between human and mouse genomes (Lalley et al. 1978), but was rapidly extended to other species such as cat and dog (O'Brien et al. 1995; O'Brien et al. 1997), horse (Caetano 1999), chimpanzee (Crouau-Roy et al. 1996), baboon (Rogers et al. 1995), vole (McGraw et al. 2011), flycatcher and zebra finch (Kawakami et al. 2014), among others.

On the other hand, radiation hybrid maps, are based on the ionizing irradiation of genomes, followed by cell hybridization (Cox et al. 1990). By this approach, chromosomes are broken into fragments that can be studied in hybrid cells. The estimation of the distance between two loci is calculated based on the frequency these two *loci* are found in the same cell. Radiation hybrid maps have been constructed in human (Gyapay et al. 1996), rat (Watanabe et al. 1999), dog (Mellersh et al. 2000), cow (Band et al. 2000), pig (Hawken et al. 1999), cat (Murphy et al. 2000), rhesus

macaque (Murphy et al. 2005a) and zebrafish (Geisler et al. 1999) among others.

1.1.4.3 Comparative genomics

Since the publication of the first draft of the human genome (Lander et al. 2001), new advances on genome sequencing projects and the availability of mathematical algorithms have transformed the study of genome structure and evolution. In this first approach, >20,000 large Bacterial Artificial Chromosome (BAC) clones were used to obtain genomic data (Lander et al. 2001). Now, with the development of next-generation sequencing methods such as 454/FLX or Solexa (reviewed in Mardis 2008) both the time and cost of sequencing have been gradually reduced. In fact, and thanks to new international collaborative efforts, the genomes of 41 mammalian species have been sequenced to differing degrees of completion (Ensembl database, 26th release). This includes 18 Euarchontoglires species (guinea pig, rat, mouse, rabbit, kangaroo rat, squirrel, pika, tarsier, vervet monkey, olive baboon, mouse lemur, bushbaby, marmoset, macaque, chimpanzee, orangutan, gorilla and human), 14 laurasiatherian representatives (megabat, microbat, shrew, dolphin, pig, cow, alpaca, horse, dog, cat, hedgehog, ferret, panda and sheep), 3 afrotherian species (elephant, hyrax and lesser hedgehog tenrec), 2 xenathrans (armadillo and sloth), 2 metatherian (opossum and wallaby) and the platypus, as a prototherian representative. This has been possible through the implementation of two large sequencing projects: (i) the 1,000 Genomes Project (Siva 2008) whose purpose is to obtain whole-genome sequences from 1,000 humans in order to discover and understand the catalog of human variation, and (ii) the Genome 10K Project (Genome 10K Community of Scientists 2009; Koepfli et al. 2015), which has as a main objective to obtain whole-genome sequences of 10,000 vertebrate species. Up to now, about 205 vertebrate genomes are already published or in progress, yielding to a new genomics era that is changing the comparative genomics landscape.

Therefore, the availability of whole-genome data of a wide range of species has fueled the development of algorithms that allowed for the detection of genomic regions of homology among taxa (the so-called Homologous Synteny Blocks, HSBs), and regions of disruption of homology (Evolutionary Breakpoint Regions, EBRs) (BOX 2). Both the identification of HSBs and whole-genome alignments rely on identifying specific markers (i.e., genes or short sequences that are highly conserved between genomes and long enough to make their conservation statistically significant) within genomes. If a sufficiently dense homology region is identified, then those regions

are considered as HSBs. Algorithms developed for genome alignments in order to detect HSBs and EBRs generally use markers as anchors (i.e., fixed references in the alignment) and these markers are usually orthologous genes, k-mers or even nucleotide sequences. Some examples of these algorithms include *MGR* (Bourque and Pevzner 2002), *GRIMM* (Tesler 2002), *GRIMM-synteny* (Pevzner and Tesler 2003a), *CHAINNET* (Kent et al. 2003), *CASSIS* (Baudet et al. 2010), *Cyntenator* (Rödelsperger and Dieterich 2010), *Synteny Tracker* (Donthu et al. 2009), *Satsuma Synteny* (Grabherr et al. 2010) and *i-ADHoRe* (Proost et al. 2012), among others. These algorithms have been used in a wide range of comparative studies, mainly in mammals. In this way, a large number of syntenic regions have been identified at different degrees of resolution (Nadeau and Taylor 1984; Pevzner and Tesler 2003a; Bourque et al. 2004; Larkin et al. 2009; Skinner and Griffin 2012; Zhang et al. 2014).

However, although all these methods have been deeply improved in recent years, such as the approaches to narrow EBRs coordinates (Lemaitre et al. 2008; Larkin et al. 2009) there is still an inaccurate identification of CRs depending on the detecting resolutions used in each approach (Lemaitre et al. 2008; Attie et al. 2011). This lack of consensus between different methods makes that still nowadays new approaches are being developed in order to identify CRs, HSBs and EBRs with high resolution and maximum confidence.

BOX 2. Graphical representation of Homologous Synteny Blocks (HSBs) and Evolutionary Breakpoint Regions (EBRs)

Homologous synteny blocks consist of conserved genomic regions between two or more species. They are defined as a minimun of two adjacent markers without homology interruption on the same chromosome or in different chromosomes of two species. Conversely, EBRs are considered as the interval between two contigous HSBs or regions of disruption of homology that result from a reorganization (Figure 1.3). Their detection is useful for evolutionary studies as they allow for the reconstruction of the origin of mammalian genomes and the evolutionary forces that have shaped them (Murphy et al. 2005b; Ruiz-Herrera et al. 2006; Larkin et al. 2009).

Figure 1.3: HSBs (in green) and EBRs detected between two chromosomes from two species (A and B).

1.1.5 Models of distribution of CRs within genomes

By the use of all the methodological approaches described above, the study of chromosomal variability among taxa as a result of genome reshuffling has been an exciting issue in evolutionary biology and different models have been proposed in order to explain the distribution of EBRs across genomes.

1.1.5.1 The *random breakage model*

The random breakage model has been the dominant paradigm of genome reshuffling since it was first proposed in the late 80's by Nadeau and Taylor (1984), based on a previous work by Ohno, years before (Ohno 1973). Their thesis relied on two main assumptions: (i) that many chromosomal segments are expected to be conserved among species, and (ii) that CRs are randomly distributed within genomes. In their work, Nadeau and Taylor (1984) compared the mouse and human linkage maps based on the correspondence of 83 orthologous genes and described the presence of approximately 180 HSBs between both genomes. Additionally, they hypothesized that the distribution of the conserved regions sizes followed an exponential function as a result of a random process. Therefore, they proposed that CRs were distributed uniformly across genomes. This view was generally accepted for many years, supported by additional data of new described orthologous genes between mouse and human genomes (Sankoff et al. 1997).

1.1.5.2 The *fragile breakage model*

Several years after Nadeau and Taylor proposed their model, genome sequencing efforts and the development of new mathematical algorithms for whole-genome analysis, proved that the first assumption was true: there is a conservation of large genomic regions among mammalian species (Froenicke et al. 2006; Robinson et al. 2006). Regarding the second postulate, it was questioned by Pevzner and Tesler (2003b) who proposed the *fragile breakage model*, arguing against the random distribution of CRs within genomes. In their work, the *GRIMM-Synteny* algorithm was used to characterize regions of homology between the human and mouse genomes, identifying 281 HSBs. This analysis also revealed an unexpectedly large number of small and closely localized HSBs that were not described before and whose lengths followed an exponential distribution, a circumstance that could

not be explained by the *random breakage model*. This new evidence yielded to an intense discussion on the suitability of both models (Sankoff and Trinh 2005; Peng et al. 2006; Sankoff 2006; Becker and Lenhard 2007) although previous experimental data in mammals supported the non-randomness of EBRs distribution (Froenicke 2005; Ruiz-Herrera et al. 2005). Subsequently, comparative studies have given support to the *fragile breakage model*; that is, there are regions throughout the mammalian genome that are prone to break and reorganize (Pevzner and Tesler 2003b; Bourque et al. 2004; Zhao et al. 2004; Murphy et al. 2005; Ruiz-Herrera et al. 2006; Peng et al. 2006; Sankoff 2006; Becker and Lenhard 2007; Kemkemer et al. 2009; Lemaitre et al. 2009; Larkin et al. 2009; Gordon et al. 2009; Alekseyev and Pevzner 2010; Mlynarski et al. 2010). These studies also reported that some EBRs that occurred during mammalian evolution were reused (i.e., that were involved in CRs independently in two different lineages without a recent common ancestor) (e.g., Pevzner and Tesler 2003b; Murphy et al. 2005b; Sankoff 2006; Larkin et al. 2009; Alekseyev and Pevzner 2010; Farré et al. 2011) a circumstance that also supports the *Fragile breakage model*. This view is also supported by the presence of repetitive sequences associated to CRs (see section 1.1.3).

1.1.5.3 The *intergenic breakage model*

Following the efforts made by evolutionary biologists to understand why specific genomic regions are prone to break, Peng and collaborators (2006) provided a new approximation to the field by proposing the *intergenic breakage model*. By simulating random CRs in the human and mouse genomes and by analyzing the genomic distribution of both intergenic and gene regulatory regions, they proposed that the combination of long regulatory regions and the distribution of intergenic regions sizes can be considered a conditional factor for the genomic distribution of EBRs. This model also suggests that selection would avoid the presence of EBRs within genes that could affect their expression causing deleterious effects. Consequently, under the *intergenic breakage model*, EBRs are considered regions of genomic instability not because of their intrinsic DNA properties but due to their gene composition as they are localized in regions where negative selection is minimal. Thus, this model predicts that DSBs, as the origin of EBRs, would appear to have a random distribution, but only those DSBs that do not have deleterious effects would become fixed within populations. Data supporting this hypothesis has been provided in different models. For example, Mongin and collaborators (2009) analyzed CRs that occurred along the human lineage when compared with the opossum genome,

showing that long-range transcriptional regulation played a role in the fixation of chromosome breaks.

In this context, it can be hypothesized that CRs tend to localize in intergenic regions in order to escape from negative selection (Peng et al. 2006; Farré et al. 2015). Which means that if a CR disrupts an important gene involved in cell maintenance or development, the deleterious effect of that affection would have as a result its elimination from the population. Notwithstanding the existence of evidence supporting this view, the presence of EBRs in high-density genic regions has been described in mammals (Murphy et al. 2005b; Larkin et al. 2009; Zhang et al. 2014). Genes found within EBRs in species as pig and macaca have been found to be associated with adaptive responses to external factors such as inflammatory and immune response as well as muscle contraction (Groenen et al. 2012; Ullastres et al. 2014). Conversely, genes localized within highly conserved regions of genomes present an enrichment of genes involved in cell cycle and nervous system development (Larkin et al. 2009). This is also supported by the case described in mouse, were the presence of housekeeping genes near the centromeres avoid the fixation of Rb fusions in natural populations (Ruiz-Herrera et al. 2010). Accordingly, these findings suggests that EBRs would be under strong positive selection meaning that the fitness cost of a deleterious rearrangements would be high and thus, this would only be transmitted to the offspring in the case any important developmental gene is disrupted.

1.1.5.4 The *integrative breakage model*

Despite their apparent discrepancies, the *fragile breakage model* and the *intergenic breakage model* are not mutually exclusive, as several works published in the recent years have presented supporting data for different aspects of both (Murphy et al. 2001; Ruiz-Herrera et al. 2006; Larkin et al. 2009; Larkin 2010; Farré et al. 2011). This includes the role of repetitive sequences and functional constrains in the genomic localization of EBRs (Armengol et al. 2005; Ruiz-Herrera et al. 2006; Elsik et al. 2009; Larkin et al. 2009; Farré et al. 2011). In addition to these factors, several lines of evidence suggest that chromatin structure is probably also affecting genome plasticity as the permissiveness of some regions of the genome to undergo CRs could be determined by changes in chromatin conformation (Lemaitre et al. 2009; Carbone et al. 2009). Based on these findings, there is evidence to support that certain properties of local DNA sequences (i.e., repetitive sequences, section 1.1.3),

together with the epigenetic state of the chromatin, could promote the change of chromatin contributing to the origin of CRs. In this context, a new model has recently been proposed to integrate this view, the *integrative breakage model* (Farré et al. 2015). This model proposes that, genome reshuffling would occur in regions that (i) physically interact inside the 3D nuclear space (the "nucleome"), (ii) have an accessible chromatin state that promotes genomic instability and (iii) do not disrupt essential genes and/or their association with long-range *cis*-regulatory elements (Farré et al. 2015).

Chromatin organization is basically composed by several superimposed layers that include: (i) chemical modifications of the DNA or epigenetic signatures such as acetylation and methylation, (ii) the presence of nucleosomes that wrap the DNA around an octomer of eight histones (two copies of each four histone proteins: H2A, H2B, H3 and H4) joined to each other by a linker histone (H1) and an intervening stretch of DNA (Woodcock 2006), (iii) the high-order organization of the chromatin compartments inside the nucleus (such as chromosomal territories or open/close chromatin conformation), and (iv) gene expression during cell cycle and development. Taking this into account, Farré and collaborators (2015) focused their attention on four different levels of hierarchical genome organization in order to explain the role of chromatin architecture in genome plasticity: (i) Chromosomal Territories (CTs), (ii) open/closed compartments, (iii) Topological Associated Domains (TADs), and (iv) looping interactions (Dekker et al. 2013; Nora et al. 2013; Phillips-Cremins 2014) (Figure 1.4).

Chromosomal territories were initially described by FISH with whole-chromosome painting probes that showed that chromosomes localize within the interphase nucleus as discrete globular domains (Mora et al. 2006; Cremer et al. 2008). This localization is non-random, as big chromosomes tend to localize in the nuclear periphery whereas small chromosomes are positioned towards the center (Cremer and Cremer 2001; Parada and Misteli 2002; Kozubek et al. 2002; Mora et al. 2006) a conformation that is evolutionary conserved in mammals (Tanabe et al. 2002; Mora et al. 2006). Inside CTs, chromatin can present open or closed conformation states. Open chromatin (termed "A" conformation) provides access to a range of DNA binding proteins, necessary for genetic regulation processes such as transcription, DNA repair and recombination, being associated with regulatory elements (i.e., promoters and enhancers). On the other hand, closed chromatin (termed "B" conformation) does not facilitate access to protein binding (Lieberman-Aiden et al. 2009). These conformation states are regulated by specific histone modifications such as acetylations and methylations in specific residues of the

histones, as well as similar type of changes at DNA cytosine residues (Barski et al. 2007). Markers for open chromatin states include H3K9ac, H3K27ac, H3K4me3 and monomethylations of H3K27, H3K9, H4K20, H3K79 and H2BK5, whereas closed chromatin states mainly include trimethylations such as H3K27me3, H3K9me3 and H3K79 (Barski et al. 2007; Chai et al. 2013). Additional genomic signatures have also been co-localized with open chromatin regions such as high gene density, CpG islands (Terrenoire et al. 2010), DNase I hypersensitive sites (Birney et al. 2007) and nuclear Lamina-Associated Domains (LADs) (Peric-Hupkes et al. 2010). The last of these is strongly associated with the nuclear periphery and contains around half of the human genome with sizes ranging between 40Kbp and 15Mbp (Megabase pairs) (Guelen et al. 2008) that are highly conserved between species and have been shown to be depleted of EBRs (Meuleman et al. 2013).

The development of high-throughput methodologies to study chromatin interactions such as ChIP-seq and, more recently, the chromosome conformation capture (3C) method and its derivatives (such as Hi-C), permitted to determine histone interactions as well as the description of genome-wide chromatin interaction maps (Lieberman-Aiden et al. 2009; Dekker et al. 2013; Dekker 2014). 3C-based techniques revealed the presence of sub-Mbp structures referred as TADs, which constitute discrete, contiguous regions (of 800Kbp medium length) that contain *loci* with a higher tendency to interact among themselves than with *loci* outside the region (Dixon et al. 2012; Sexton et al. 2012). Interestingly, some genomic features, such as the CCCCTC-binding Factor (CTCF) and cohesins have been found to be enriched at TADs boundaries (Schmidt et al. 2012; Rao et al. 2014) (Figure 1.4).

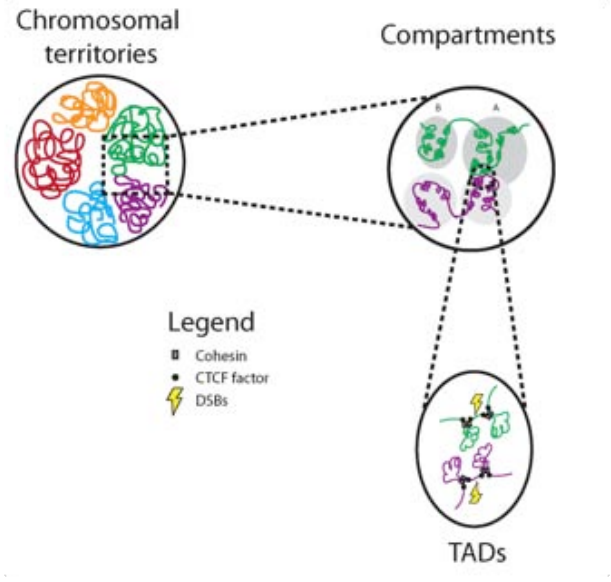


Figure 1.4. Model of chromatin structure within interphase nucleus in eukaryotic cells: Chromosomal territories (CTs, depicted in different colors) occupy specific regions within the nucleus and are compartmentalized into open (''A'') or closed (''B'') chromatin states. Both compartments contain Topological Associated Domains (TADs) delimited by cohesins and CTCF factors. DSBs can occur between two TADs and can recombine leading to CRs.

In this context, and according to the *Intergenic Breakage Model*, DSBs that lead to CRs are expected to occur in specific open-chromatin regions where DNA sequences are accessible to the protein DNA repair machinery (Farré et al. 2015) (Figure 1.4) and several lines of evidence support this interpretation. Initial studies showed that EBRs tend to localize in regions of high transcriptional activity due to the open chromatin conformations (Lemaitre et al. 2009). Also, it has been shown that regions involved in chromosomal translocations had a close localization within the nucleus in mouse (Véron et al. 2011; Zhang et al. 2012) and, more recently, a study comparing high-resolution EBRs detection between five mammalian species and chromatin states profiles, revealed that distribution of CRs could be accurately explained as mis-repaired breaks between open chromatin regions that were bought into contact (Berthelot et al. 2015). Taking together, all this evidence point out to the importance of the study of epigenetic features and chromatin state in order to understand the mechanism underlying the formation of CRs.

1.1.6 The role of CRs in speciation

T. Dobzhansky, in his "*Genetics and the Origin of Species*" (Dobzhansky 1937; 1951), identified the existence of two main components involved in the origin of reproductive isolation between species: CRs (chromosomal speciation) and genetic divergence (genetic speciation) (Dobzhansky 1937; 1951; Dobzhansky and Sturtevant 1938). Whether chromosomal speciation and genetic speciation act solely or in combination by promoting divergence between populations is a hot debate still under discussion (Faria et al. 2011). In this context, two main models have been proposed to explain the role of CRs in the speciation process: the *hybrid dysfunction model* and the *suppressed recombination model*.

1.1.6.1 *The hybrid dysfunction model*

This model was initially proposed by White (1969, 1978), who hypothesized that speciation takes place when structural CRs become fixed in a population. This process is likely to occur in small populations with high consanguinity where the presence of CRs in the hybrid would reduce fertility by the generation of unbalanced gametes due to meiosis impairment. According to this view, CRs are divided into three functional categories: (i) deleterious rearrangements that would be eliminated from the population by natural selection, (ii) CRs that are capable of giving rise to balanced polymorphisms, and (iii) CRs that reduce heterozygous fertility and,

therefore, would be playing a role in generating reproductive isolation. This later effect in hybrid sterility has been classically known as the underdominance effect (Figure 1.5).

The *hybrid dysfunction model* was initially considered to explain the presence of CRs in natural populations. This was the case of species that presented monobranchial fusions (that is, Rb fusions with homology in one arm) such as *Sorex*, *Castor* and *Mus* (Baker and Bickham 1986), or Australian grasshoppers (*Vandiemenella*), which are characterized by an extensive chromosomal variation due to fusions, fissions, translocations and inversions (King 1993). However, the *hybrid dysfunction model* received several criticisms due to the fact that the spread and subsequent fixation of an underdominant CR is problematic from a population genetics perspective, representing an unresolved paradox: If CRs are underdominant, it is unlikely that they would be fixed in the population as they would be eliminated of the population by natural selection (Wright 1941; Lande 1985).

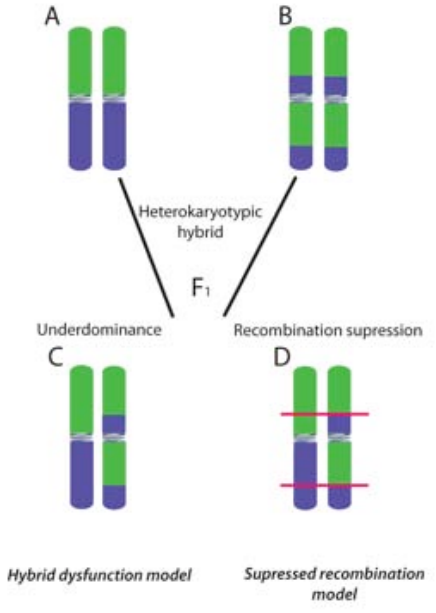


Figure 1.5. Chromosomal speciation models: In both models the CRs represented is an inversion. Following hybridization between the parental form (A) and the reorganized chromosome (B), a heterokaryotype hybrid results (F1). (C) The *hybrid dysfunction model* postulates that the hybrid will experience underdominance, whereas the *suppressed recombination model* (D) postulates inverted regions will experience suppressed recombination thus facilitating the accumulation of genic incompatibilities. Adapted from Brown and O'Neill (2010).

1.1.6.2 The *suppressed recombination model*

In order to overcome the initial limitations of the *hybrid disfunction model*, new studies have proposed that CRs could potentially contribute to speciation not through the underdominance of the hybrid but by the suppression of recombination in the presence of gene flow (either in parapatry or sympatry) (Noor et al. 2001; Rieseberg 2001). In this way, CRs do not necessarily affect fertility, but they contribute to gene flow reduction within populations by suppression of meiotic recombination within rearranged regions (Rieseberg 2001) (Figure 1.5). This effect would increase

divergence in the genomic regions affected by the CRs, favoring the accumulation of genic incompatibilities that could lead, eventually, to the fixation of locally adapted genes and produce, in the long term, a partial reproductive isolation. In this context, three main conditions need to be met: (i) CRs must suppress recombination, (ii) gene flow suppression within rearranged regions must play a major role in reproductive isolation and (iii) there must be different CRs in sister taxa (Faria and Navarro 2010). Although this model was initially proposed to explain the presence of inversions (Rieseberg 2001; Noor et al. 2001), new evidence is suggesting its applicability to Rb fusions (Dumas and Britton-Davidian 2002; Franchini et al. 2010).

Evidence for the suppression of recombination can be extracted from highly diverse model systems and methodologies. Cytogenetic approaches in mammals described this phenomenon within pericentric inversions in different rodent species, such as the sand rat (Ashley et al. 1981), and the deer mice (Greenbaum and Reed 1984; Hale 1986). This reduction in meiotic recombination was described later in the case of Rb fusions occurring in natural populations of common shrews (Borodin et al. 2008), house mice (Castiglia and Capanna 2002; Dumas and Britton-Davidian 2002; Dumas et al. 2015) and tuco-tucos (Basheva et al. 2014). Similar results have been obtained when studying sequence divergence (patterns of nucleotide differentiation) between species as an indirect estimation of recombination (see BOX 3). Rieseberg and collaborators (1995, 1999) observed in isolated hybrid populations of the sunflower genera *Helianthus* that the rate of introgression was lower within rearranged chromosomes than in collinear chromosomes. In the same vein, Besansky and collaborators (2003) found high sequence divergence rates within an inversion in the X chromosome of *Anopheles* mosquitoes, mirroring what has been described in *Drosophila*, in genomic regions close to inversions (Navarro et al. 1997; Brown et al. 2004; Noor et al. 2007). Studies of nucleotide differentiation due to inversions in mammals have resulted in more heterogeneous results. In the case of primates, Navarro and Barton (2003a) found a higher rate of protein evolution (measured by dN/dS ratios) within rearranged genomic regions between chimpanzee and human. More recently, Farré and collaborators (2013) performed a comparison between human and chimpanzee recombination maps, detecting recombination suppression in inverted regions compared to non-inverted regions. Mirroring this pattern, several studies using microsatellites in small mammals have shown restricted gene flow near the centromeres of chromosomes resulting from Rb fusions among different chromosomal races in the common shrew (Basset et al. 2006; Yannic et al. 2009) and the house mouse (Franchini et al. 2010; Giménez et al. 2013; Janoušek et al. 2015).

1.1.6.3 *Bateson-Dobzhansky-Muller Incompatibilities*

As a result of recombination suppression, genomic regions involved in CRs (either inversions or fusions) can accumulate high rates of genetic differentiation, also referred as "islands of speciation" (Turner et al. 2005; Harr 2006; Feder and Nosil 2009). These regions are attributed to have limited gene flow, due to reduced recombination and/or diversifying selection in sympatry (reviewed in Butlin 2005; Feder et al. 2012). Therefore, restricted recombination preserves and generalizes the effects of diminished gene flow and, thus, the accumulation of genomic incompatibilities within reorganized regions (Navarro and Barton 2003b). This phenomenon was already proposed by Bateson (1909) and later by Dobzhansky (1937) and Muller (1942) in the form of the *Bateson-Dobzhansky-Muller Incompatibilities* (BDMIs) *model* in order to explain sequence divergence between isolated populations, where the appearance of point mutations at specific sites over generations can generate incompatible alleles. This model is based on the epistatic interactions between genes that act when a gene can allow (epistasis) or disallow (negative-epistasis) the phenotypic expression of another gene (Figure 1.6).

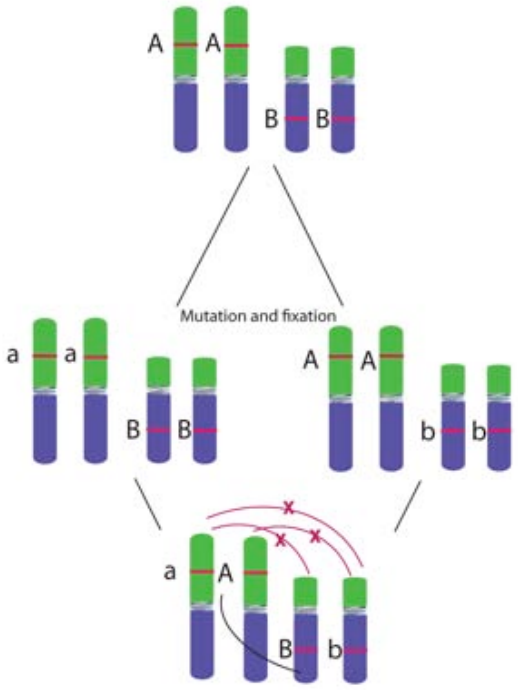


Figure 1.6. The *Bateson-Dobzhansky–Muller incompatibilities* model: In an ancestral population (top) two epistatic genes, AA and BB, are homozygous. Following mutation and fixation in two distinct populations, each population now carries a variation of these, aaBB and AAbb. In an inter/intraspecific hybrid, the diverged alleles are now present in a heterozygote as AaBb. Whereas alleles A and B are still capable of epistatic interaction, as they were in the common ancestor, Ab, aB, and ab are no longer in epistasis and hybrid incompatibility results. Adapted from Brown and O'Neill (2010).

A simple theoretical example to explain this process is represented in Figure 1.6. One population presents alleles AA and BB (considered the ancestral alleles). After a process of mutation and fixation, two populations in allopatry (derived from the ancestral population that evolved independently) would present the

combination of the ancestral and derived genes, aaBB and AAbb. Later on, in a hypothetic sympatry scenario, these two populations would produce heterozygotes with the combination AaBb. In this case, the AB combination still presents epistatic interactions whereas Ab, aB and ab combinations will show incompatibilities. These genetic incompatibilities would then produce hybrid sterility caused by negative epistatic interactions between *loci*, thus acting as an intrinsic postzygotic isolation barrier.

Different *loci* causing hybrid incompatibilities have been identified in the literature and genes responsible for this effect have been defined as "speciation genes". Several specific speciation genes have been described in different *Drosophila* species: *Odisseus* (*Ods*), mapped in chromosome X in *D. simulans* (Perez et al. 1993; Ting 1998), *JYAlpha*, mapped in chromosome 4 in *D. Melanogaster* (Masly et al. 2006) and *Overdrive* (*Ovd*), mapped in X chromosome in D. pseudoobscura subspecies (Phadnis and Orr 2009). All three genes have been suggested to play an important role in segregation distortion (meiotic drive) and hybrid sterility (Phadnis and Orr 2009).

In the case of mouse subspecies, Forejt (1996) initially described one of the *loci* responsible for hybrid sterility between *Mus m. musculus* and *Mus m. domesticus*, the so-called Hybrid Sterility 1 (*Hst1*). This *loci* was later identified as the PR domain containing 9 (*Prdm9)* gene (Mihola et al. 2009), and was subsequently described in mammalian species such as primates (Myers et al. 2010) and rodents (Oliver et al. 2009), among others. These findings are in agreement with the studies performed in house mouse hybrid natural populations (i.e., regions of contact between two distinct chromosomal races) (Hauffe et al. 2012) in which researchers examined the exchange of genetic variation between the two subspecies and detected the presence of several autosomal regions exhibiting epistasis (Payseur et al. 2004; Teeter et al. 2008; Janoušek et al. 2012), thus suggesting the role of specific loci in reproductive isolation. Subsequent genetic mapping studies using F1 and F2 hybrids from laboratory crosses have identified additional *loci* and genic interactions that are most probably contributing to sterility phentoypes (Storchová et al. 2004; Good et al. 2008; White et al. 2011; Dzur-Gejdosova et al. 2012). In fact, recent evidence suggest the existence of Quantitative Trait Locus (QTL) interactions that can induce variations in gene expression affecting fertility in *M. musculus* and *M. domesticus* F1 hybrids (Bhattacharyya et al. 2013; Turner et al. 2014; Turner and Harr 2014; Janoušek et al. 2015). A similar effect has been observed in *Drosophila* subspecies hybrids (Morán and Fontdevila 2014). Therefore, reproductive isolation is revealing itself as a multigenic trait (e.g., White et al. 2011).

BOX 3: Methodological approaches used for the study of recombination

Cytological approaches

Giemsa staining has been traditionally used to identify meiotic chromosomes for the study of bivalents and chiasmata in metaphase I, especially in mammals (e.g., Castiglia and Cappana 2002) (Figure 1.7A). However, if the objective is to analyze the recombination process as it occurs, alternative methods, such as the *in situ* immunolocalization of recombination proteins on spermatocyte spreads can be applied (Froenicke et al. 2002; Lynn et al. 2002). This allows for the analysis of the distribution of MLH1 *foci* in each cell individually at pachynema, as a proxy of recombination events (Figure 1.7B). This approach provide with recombination maps at large-scale resolutions (Mbp).

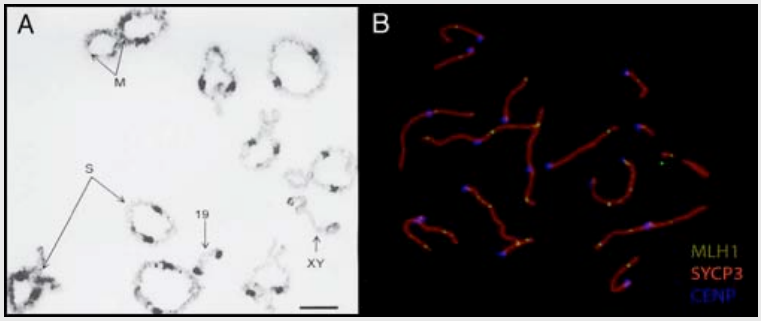


Figure 1.7. (A) Chiasmata detection by uniform staining (from Dumas and Britton Davidian, 2005). (B) Immunedetection of MLH1 (in green), SYCP3 (in red) and centromeres (in blue; Capilla et al., unpublished). Both images correspond to spermatocyte preparations from mice with Rb fusions.

High-resolution and population genetic approaches

Pedigree analyses estimate recombination rates by examining parent-to-offspring transmission of alleles at specific *loci*. This method provides estimates of current recombination rates per physical distance unit (cM, centimorgans per Mbp) in the population under study and can serve to determine recombination hotspots. Estimates of Linkage Disequilibrium (LD) through population genetic analyses, estimate historical recombination events expressed as $4N_e r/Kbp$ (Brunschwig et al. 2012). This approach represents an integration of population-level processes over several generations from Single Nucleotide Polymorphisms (SNPs) data (Clark et al. 2010) (Figure 1.8). This method presents the difficulty to detect low-frequency events and the variable results can be obtained depending on the density of markers used (Arnheim et al. 2003). Sperm typing analysis, on the other hand, directly detects recombinant DNA molecules and consists of allele-specific PCR amplification from sperm DNA by using SNPs as anchors. The advantage of this approach is that it allows the screening of large amounts of recombinant DNA from single individuals (Wu et al. 2010), although the resolution can be a limitation depending on marker density, the number of individuals analyzed, and the size of the amplifiable target (<15Kbp). Finally, the analysis of genetic differentiation (expressed as F_{st} values) using markers such as microsatellites or SNPs between populations can be considered as a proxy of recombination. If the recombination is reduced or suppressed, then, higher sequence divergences are expected (e.g., Giménez et al. 2013).

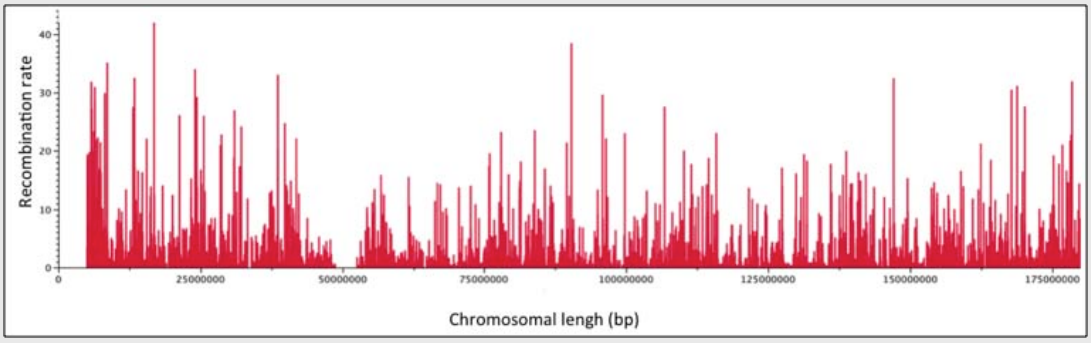


Figure 1.8. Standardized recombination rates across human chromosome 4 estimated by LD analysis. Recombination rates data extracted from Kong and collaborators (2010).

1.2 Meiotic recombination and speciation

All speciation genes characterized so far (*Ovd, Ods, JYAlpha* and *Prdm9*) present common functional features such as being involved in DNA binding, epigenetic modifications and gene expression regulation, specially in meiosis (Presgraves 2010). Therefore, the understanding of the recombination process, that takes place during meiosis, is of relevance for the study of the role of CRs in the speciation process.

1.2.1 Meiosis

Meiosis is the reductional division that takes place in the gonads of the sexually reproducing organisms, producing gametes that own an haploid chromosome dotation. This is achieved by two consecutive cell divisions (meiosis I and meiosis II), preceded by a single round of DNA replication (S-phase) (reviewed in Handel and Schimenti, 2010). Especially relevant for the role of speciation genes, is the first meiotic division (meiosis I) where synapsis of homologous chromosomes and recombination (i.e., DNA exchange between homologous chromosomes) takes place. Meiosis I comprises four stages: prophase I, metaphase I, anaphase I and telophase I (Figure 1.9). Meiotic recombination between homologous chromosomes occurs during prophase I, which is divided into four differentiated stages: leptonema, zygonema, pachynema and diplonema (Handel and Schimenti 2010). This process involves the activation of complex and highly regulated mechanisms.

At leptonema, chromosomes are aggregated by their telomeres to the nuclear envelope in a configuration known as the *bouquet* (Scherthan et al. 1996). This structure promotes the pairing of homologous chromosomes by the formation of proteinaceus structures along chromosomes formed by cohesins (REC8 and SMC1) and proteins of the Synaptonemal Complex (SC). The SC consists of a tripartite structure with the axial elements, Synaptonemal Complex Protein 2 and 3 (SYCP2 and SYCP3) connected by perpendicular filament proteins with the overlapping central element, Synaptonemal Complex Protein 1 (SYCP1) (Heyting 1996; Page and Hawley 2004; Henderson and Keeney 2005) (Figure 1.9) that permits the alignment and pairing of homologous chromosomes. Contemporarily, meiotic recombination starts with the formation of DSBs by the endonuclease protein Spo11 (Keeney et al. 1997; Romanienko and Camerini-Otero 2000; Longhese et al. 2009), a highly conserved protein with any or little DNA sequence specificity (Keeney 2008). The DSBs formation induce the phosphorylation of histone H2AX on serine 139 (H2AX)

by the proteins Ataxia Telangiectasia Mutated (ATM) and ATM-Rad3-Related (ATR) (Rogakou et al. 1998; Burma et al. 2001; Kuo and Yang 2008) that constitutes the first step in recruiting proteins involved in DSBs repair.

At zygonema, DSBs are repaired leading to synapsis between homologous chromosomes producing, either Crossovers (COs) or Non-Crossovers (NCOs) (Allers and Lichten 2001; Hunter and Kleckner 2001; Baudat and de Massy 2007). In both cases, after H2AX phosphorylation at leptonema, the Replication Protein A (RPA) binds to the 3' strand overhangs of DSBs (He et al. 1995). RPA is then displaced by the protein Radiation sensitive 51 (Rad51) and/or Disrupted Meiotic cDNA 1 (DMC1) (Pittman et al. 1998; Yoshida et al. 1998) that form nucleoprotein filaments that catalyze strand invasion. In the case of COs formation, a D-loop is created capturing the second 3' end of the homologous chromosome, and after the DNA synthesis and ligation, a double Holliday Junction (dHJ), with heteroduplex DNA flanking the DSBs site, is produced (Collins and Newlon 1994; Schwacha and Kleckner 1995). This structure is then resolved by the cleavage and ligation of strands of same polarities at identical positions, which generates COs. In this case, COs formation involves the invasion of the homologous chromosomes by the single end created by DSB, leading to the formation of chiasmata connecting both homologous chromosomes. Mismatch repair of heteroduplexes can lead to either, gene conversion or restoration depending on the choice of the corrected strand. However, in most organisms, the number of DSBs largely exceeds the number of resulting COs, indicating that the majority of DSBs is resolved as NCOs. For example, in mice about 250-300 DSBs are initiated, but only about one-tenth of them are processed into COs (Koehler et al. 2002a; Moens et al. 2007; Murakami and Keeney 2008). NCO products are thought to be generated by either a dHJ or by a mechanism called Synthesis-Dependent Strand Annealing (SDSA) (McMahill et al. 2007). In SDSA, the D-loop is disassembled by displacement of the newly synthesized strand, which anneals with the other DSB end. DSBs repair is completed with DNA synthesis by a DNA polymerase (using the homologous sequence as template) followed by a ligation to the original strand. In this case, mismatch repair of heteroduplex generates gene conversion without resulting in a CO (Figure 1.9).

Is at pachynema when recombination is resolved. The SCs are completely established creating bivalent structures (Oud et al. 1979; Speed 1982) and COs are resolved by the repair pathway directed by proteins MutS Homologs 4 and 5 (MSH4 and MSH5) (that appear earlier at zygonema) and MutL Homologs 1 and 3 (MLH1 and MLH3) (Baker et al. 1996; Kneitz et al. 2000; Santucci-Darmanin 2000; Lipkin et al. 2002; Snowden et al. 2004). Finally, at diplonema, homologous chromosomes

start to segregate by the disegregation of SCs. Homologous chromosomes will keep contact only in the CO sites until anaphase by the chiasmata structures (Oud et al. 1979; Speed 1982). With this process, prophase I ends and starts metaphase I, where the nuclear envelope disappears and the chromosomes, which are totally condensed, migrate to the equatorial plane of the cell. At telophase I, daughter cells are generated with haploid chromosome dotation and genetically different to each other due to genetic recombination process.

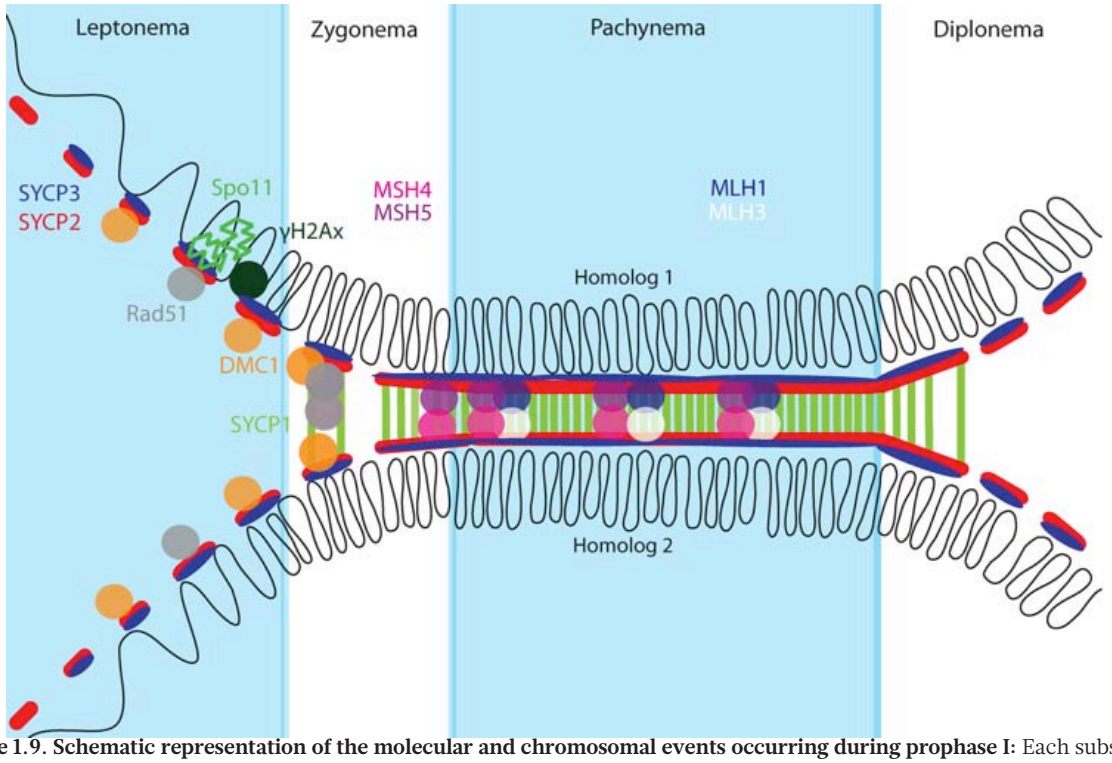


Figure 1.9. Schematic representation of the molecular and chromosomal events occurring during prophase I: Each substages of meiosis I prophase is represented. During leptonema, pairing of homologous chromosomes starts, although they are not completely paired until pachynema. The SC is being formed by the load of lateral elements (SYCP3 and SYCP2) and chromatids experience the formation of DSBs, induced by Spo11. DSBs are recognized by the homologous recombination repair machinery including phosphorylation of histone H2AX by ATM, which triggers the binding of proteins DMC1 and RAD51, among others. At zygonema, homologous chromosomes are completely paired by the central element of SC (SYCP1) and synapsis is initiated. Pachynema is defined by completion of synapsis and includes the repair of DSBs resulting in the maturation of a subset that are resolved into COs, marked by the mismatch repair proteins MLH1 and MLH3. After recombination is completed, chromosomes undergo desynapsis and condense during the final diplonema substage. Image extracted and modified from Cohen and Holloway (2010).

1.2.2 Variation of recombination rates within genomes

It is well known that COs do not distribute randomly across the genome, but rather at preferred locations known as recombination hotspots (reviewed in Petes 2001). Hotspots span across small regions of the genome (1-2Kbp length) and are surrounded by areas where recombination is normally suppressed (reviewed in Kauppi et al. 2004). Evidence from studies in many organisms using both cytological and genetic approaches (see BOX 3), have described this non-random distribution of recombination events (de Massy 2003). However, the mechanisms responsible of

the formation and genomic distribution of DSBs and thus, of recombination hotspots, are not fully understood.

The distribution of COs across the genome is carefully regulated by meiotic homeostasis, which is conditioned by different mechanistic and genetic factors depending on the organism studied (Cole et al. 2012). However, it is generally accepted that COs distribution presents three main characteristics: (i) COs take place in discrete regions of the genome, (ii) there is at least one CO per bivalent, the so-called "obligatory chiasma" as its presence ensures the proper segregation of homologous chromosomes (Bishop and Zickler 2004; Kauppi et al. 2004), and (iii) if there are two or more COs along a bivalent they tend not to appear close to each other, a phenomenon known as "CO interference" (Kauppi et al. 2004; Berchowitz and Copenhaver 2010).

Observations in mammalian species suggest that the total number of COs per cell is generally correlated with the number of chromosomal arms (Pardo-Manuel de Villena and Sapienza 2001; Li and Freudenberg 2009; Segura et al. 2013). Moreover, the total number of COs on each chromosome is restricted by CO interference, which prevents COs from occurring near each other, operating over distances spanning tens of megabases (Broman et al. 2002; Petkov et al. 2007). Therefore, the requirement of having a minimum of one CO per chromosome, but limited by the density interference, results in a strong tendency for shorter chromosomes to have more COs per unit length than larger chromosomes (Kaback et al. 1992; Lynn et al. 2002). This interference has not only been observed between COs but also near the centromeres, that produces an increasing pattern of recombination rates towards the telomeres (reviewed in Nishant and Rao 2006).

Additionally, other specific genomic features have been described to play a role in COs positioning along bivalents. For example, recombination hotspots are associated with the GC content being high recombination rates associated with GC-rich regions (Kauppi et al. 2004; Coop and Przeworski 2007; Buard and de Massy 2007). It is known that recombination influence GC-content evolution by favoring the fixation of G and C alleles through the increasing rate of A/T to G/C substitutions by gene conversion (Clément and Arndt 2013). This phenomenon has been described in humans (Eisenbarth 2001) as well as in *Saccharomyces* (Gerton et al. 2000) and pig (Tortereau et al. 2012). These GC-rich regions are organized in chromosome domains or isochores (Bernardi 2000) also corresponding to gene-rich regions that present different chromatin features inside the nucleus such as high transcriptional activity (Jabbari and Bernardi 1998). Moreover, and depending on the species analyzed, recombination hotspots have been found to be associated with specific consensus

sequence motifs which have been found to be species-specific and, a percentage of them, associated with the function of a meiosis-specific gene: *Prdm9*. In humans, a 13-mer motif has been observed associated with 40% of the recombination hotspots (Myers et al. 2008). Similar predictions have been observed in yeast, although the motif presented was different (being in this case 18-mer) (Steiner and Smith 2005), and mice (Buard et al. 2009; Smagulova et al. 2011) where the motif recognized is 12-mer (although variability has been described). Therefore, these observations involve the role of specific genes on the localization of COs in mammals (see section 1.2.4).

1.2.3 Variation of recombination rates within and among species

Variability in recombination rates has been observed among species and among individuals of the same species. Within the same species, recombination presents a sex-bias (males *vs.* females) (Lenormand and Dutheil 2005; Coop and Przeworski 2007). In humans, for example, the genetic map length in females is 1.6 times longer than in males (Broman et al. 1998; Kong et al. 2010). Differences between sexes are also found when analyzing chiasmata distributions, being in males localized near the telomeres (Tease and Hultén 2004) whereas female recombination is more evenly distributed across the genome (Froenicke et al. 2002; Lynn et al. 2005). This pattern has also been observed not only in mouse, where female recombination maps are 1.09 times longer than in males (Lawrie et al. 1995; Cox et al. 2009), but also in pigs (Tortereau et al. 2012) and *Arabidopsis* (Giraut et al. 2011). Hypothesis to explain these differences were first proposed by Haldane (1922) who suggested that the heterogametic sex had low COs rates as a consequence of selection against recombination between the sex chromosomes. However, this rule has not been supported by empirical data. Strikingly, other factors have been claimed to involve sex recombination differences such as COs interference, which operates over shorter genomic distances in females than in males (de Boer et al. 2006; Petkov et al. 2007; Giraut et al. 2011) and are related to the SC length being also different between sexes (Tease and Hultén 2004). However, meiotic dimorphism between sexes is still under discussion as different other genetic factors could be involved, such as GC content.

Recombination maps have been constructed for many species by different approaches (reviewed in Smukowski and Noor 2011) (see BOX 3). This has been the case of yeast (de Castro et al. 2012; Liu et al. 2012), plants (Harushima et al. 1998; van Os et al. 2006), invertebrates such as *Drosophila* (Santos-Colares et al. 2004;

Kulathinal et al. 2008), grasshoppers (Taffarel et al. 2015) and honeybees (Wilfert et al. 2007; Meznar et al. 2010) and higher vertebrates, including birds such as zebra finch (Backström et al. 2010) or chicken (Groenen et al. 2009; Elferink et al. 2010). Regarding mammals, meiotic studies have included human and non-human primates (Sun et al. 2005; Hassold et al. 2009; Garcia-Cruz et al. 2011; Auton et al. 2012; Munch et al. 2014), mouse (Froenicke et al. 2002; Koehler et al. 2002b; Paigen et al. 2008; Cox et al. 2009; Dumont et al. 2011; Brunschwig et al. 2012), common shrews (Borodin et al. 2008), horse (Lindgren et al. 1998; Al-Jaru et al. 2014) and dog (Wong et al. 2010), among others.

Although the use of different methodological approaches can result in different recombination resolutions (Kbp *vs.* Mbp) (reviewed in Butlin, 2005), a degree of correspondence have been described in orthologous chromosomes from different species (Garcia-Cruz et al. 2011). Conservation in both location and the intensity of recombination hotspots has been reported in yeast (*S. paradaous vs. S. cerevisiae*) (Tsai et al. 2010). In mammals, however, the conventional argument has been that although recombination rates may vary considerably between species when comparing high-resolution (Kbp), these differences disappear at a broader scale (Mbp). In fact, recent studies in mice (Dumont and Payseur 2008), primates (Garcia-Cruz et al. 2011) and other mammalian species representative of Afrotheria, Euachontoglires and Laurasiatheria (Segura et al. 2013), have suggested that closely related species tend to have similar average rates of recombination. These observations suggested that phylogenetic dimension might also play an important role on meiotic recombination (Segura et al. 2013).

1.2.4 The role of the *Prdm9* gene in meiotic recombination and speciation

As explained in section 1.1.6.3, mouse *Hst1* was first defined as a hybrid sterility *loci* responsible for male sterility in hybrids between *M. musculus* subspecies (Forejt 1996). It was initially mapped in mouse chromosome 17 at a genetic distance of 8.4 cM from the centromere (between 15,545,119 and 15,564,354bp, extracted from Mouse Genome Informatics – www.informatics.jax.org) (Forejt et al. 1991; Gregorová and Forejt 2000). Later on, the gene corresponding to this *loci*, also known as *Meisetz* (Meiosis induced factor containing a PR/SET domain and zinc finger motif), was found to codify for a PR domain 9 protein (PRDM9) (Mihola et al. 2009).
The PRDM9 protein presents a Krueppel-Associated Box (KRAB) domain in the

N-terminal region which behaves as a transcriptional repressor domain (see Urrutia 2003, and references therein), a SET domain with histone 3 lysine 4 (H3K4) methyltransferase activity and an array of C2H2 Zinc Finger (ZnF) repeats in the C-terminal domain that recognizes a DNA-specific repeat sequence (Figure 1.10A). The *Prdm9* structure has been described in several mammalian species, such as humans (Baudat et al. 2010; Jeffreys et al. 2013), non-human primates (Oliver et al. 2009; Schwartz et al. 2014), mouse (Baudat et al. 2010), other rodent species such as *Mus musculus castaneus, Mus macedonicus, Mus spicilegus, Coelomys pahari, Apodemus sylvaticus, Meriones unguiculateus, Peromyscus leucopus, Peromyscus maniculatus, Peromyscus polionotus, Microtus agrestis* and *Arvicola terrestris* (Oliver et al. 2009) and equids (Steiner and Ryder 2013), among others. Moreover, *Prdm9* expression has also been detected in the germ line of additonal metazoan species such as trout, cattle, pig, sea urchin, and the gastropod snail (Oliver et al. 2009). Exceptionally, in dogs, the gene was early inactivated during its evolution (Axelsson et al. 2012).

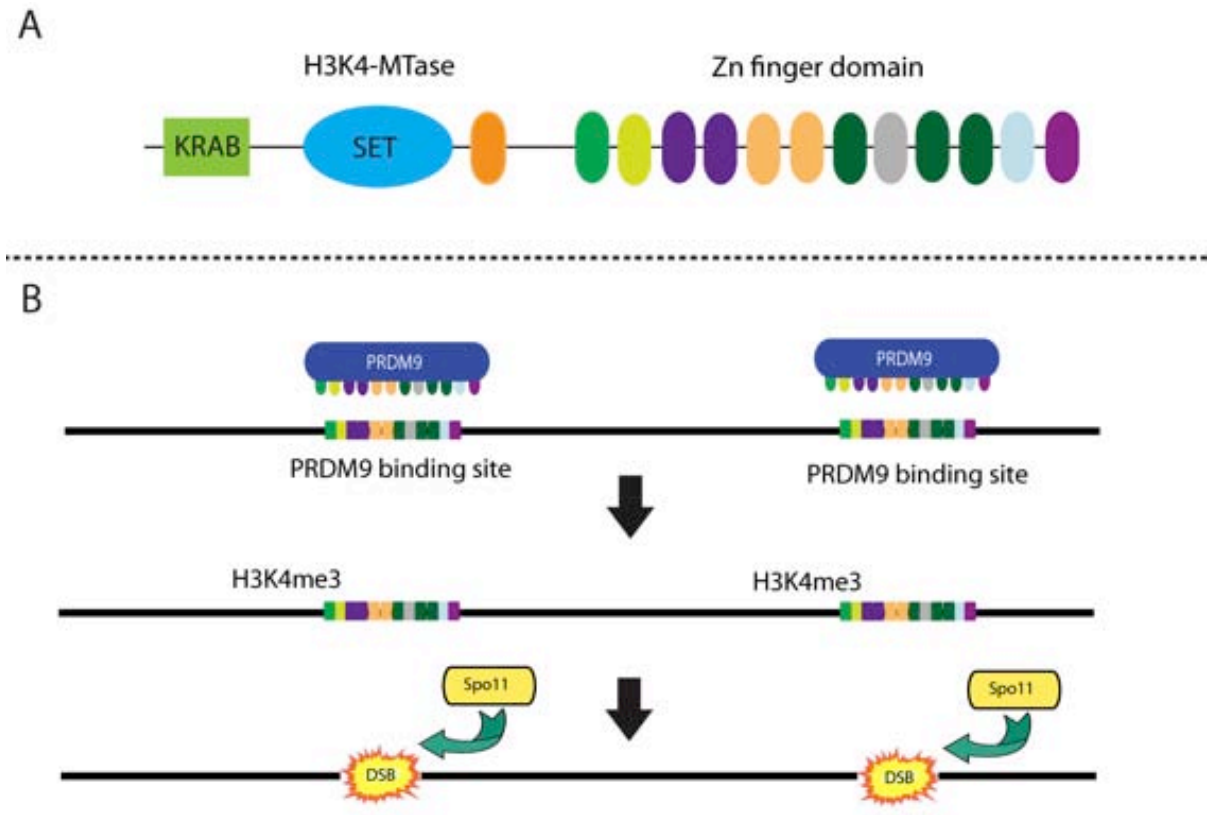


Figure 1.10. *Prdm9* structure and function: (A) The *Prdm9* gene codifies for a protein with different domains, such as a KRAB domain in the N-terminal region, followed by a SET domain with methyltransferase activity and a ZnF array of 12 repeats in the C-terminal region (in the case of the mouse reference allele). (B) The PRDM9 protein recognizes a specific DNA sequence in the genome through its ZnF domain and methylates lysine 4 of histone 3. This epigenetic mark recruits Spo11 that will induce the formation of DSBs in early meiosis.

Prdm9 is expressed only during early stages of meiosis in testis and ovaries (Hayashi et al. 2005) and it was first described as a marker of transcription recognition sites (Brick et al. 2012). The mechanism by which *Prdm9* controls the distribution pattern of meiotic recombination is still largely unknown, but it has been postulated that it occurs by the recognition of a degenerated DNA motif by its C2H2 ZnF domain (Figure 1.10A), which is, as it as been explained above, species-specific (Baudat et al. 2010). Once the DNA motif is recognized, the SET domain leaves an epigenetic signature (H3K4me3) and recruits the protein Spo11 responsible of the formation of DSBs (Keeney et al. 1997) (see section 1.2.1, Figure 1.10B). In mouse, Parvanov and collaborators (2010) also identified *Prdm9* as the gene that would control the distribution of recombination hotspots by mapping this gene region in a inter-subspecies mouse cross, data also supported by the enrichment of H3K4me3 observed within mouse hotspots (Baudat et al. 2010).

What is relevant about this gene from an evolutionary perspective is that the ZnF array is characterized by a high variability in both the number and the sequence of the ZnF repeats (Oliver et al. 2009; Baudat et al. 2010). These differences in size and sequence have been described between closely related species (Oliver et al. 2009; Schwartz et al. 2014) and even within the same species (Oliver et al. 2009; Buard et al. 2014; Kono et al. 2014; Pratto et al. 2014). Interestingly, modifications of the ZnF domain can be translated into a re-distribution of the recombination sites (Baudat et al. 2010). This effect was initially described in mouse by Brick and collaborators (2012) by comparing two close related *Prdm9* alleles (with 9 and 13 repeats, 9R and 13R), which shared only 1.1% of recombination hotspots positions due to the differences of *Prdm9* sequence. The authors, proposed that a single amino-acid change within the ZnF domain could lead to new *Prdm9* variants with novel DNA binding specificity that, in turn, could potentially create a new genome-wide distribution of hotspots (Brick et al. 2012). In fact, recent studies in humans (Pratto et al. 2014) determined that three different alleles resulted in different distribution of DSBs, although, the implication of additional factors other than *Prdm9* in modulating the frequencies of recombination initiation cannot be ruled out.

In fact, recombination hotspots distribution have been proposed to be affected by a complex interplay between the *cis*-acting DNA sequence at hotspots (or the gene conversion bias of COs repairing mechanisms) and the *trans*-acting factor that binds to that DNA sequence (being in this case the PRDM9 function). Different studies have identified that one of the two chromatids present a higher probability of undergoing gene conversion, where the donor chromatid is used as the template to repair the DNA sequence lost from the active partner in the course of creating

the DSB (Baudat and de Massy 2007; Paigen et al. 2008; Berg et al. 2011). Therefore, this biased gene conversion predicts that hotspots should undergo evolutionary erosion, a phenomenon known as the "hotspot paradox" (Coop and Myers 2007; Baker et al. 2015) as if hotspots drive themselves to extinction, its seems difficult that recombination persists. However, PRDM9 function brings a solution to this paradox, as its rapid evolutionary change can overcome hotspots loss by undergoing mutations altering its ZnF array, and thus, changing the genome wide distribution of hotpots. However, the question of why *Prdm9* is evolving so fast it is still under discussion. One of the current explanations is that *Prdm9* is under positive selection, associated with the rapid evolution of its binding sites (Coop and Myers 2007).

1.3 The western house mouse as a model for chromosomal speciation

The western house mouse, *M. musculus domesticus* (Schwarz and Schwarz 1943) taxonomically belongs to the Subfamily Murinae, one of the taxa of the Family Muridae. As other *Mus* representatives, western house mice are ecologically highly opportunistic animals and, in general, strictly commensal, although some populations have undergone a secondary feralization and colonized several outdoor habitats. They are small-sized rodents, showing body mass, head and body length variability, ranging between 12.5-29g and 73-101.5mm, respectively (Sans-Fuentes 2004). Commensal individuals show a relatively darker coloration than feral mice, with the back, tail and legs dark grey-brown, the ventral region slightly clearer than the back, and without a conspicuous delimitation between the coloration of the upper and lower parts. They are mainly crepuscular or nocturnal animals, however they alternate periods of activity during the day related with the search for food (Ballenger 1999).

The western house mouse has been localized in Western Europe, Africa and Middle East (e.g., Guénet and Bonhomme 2003), and due to its commensal association with humans it arrived to America, Australia (Auffray et al. 1990) and Antartica (Jansen van Vuuren and Chown 2006). It constitutes a significant pest in many areas, producing economic wastage by distorting and contaminating stored foods and crops and other kinds of damage. Nevertheless, house mice (*sensu lato*) also constitute one of the most important model organisms in biology and medicine. A striking characteristic of *M. m. domesticus* is its exceptionally variable karyotype throughout its range of distribution compared to other house mouse subspecies.

1.3.1 The western house mouse chromosomal variability

The genus *Mus* is characterized by having a standard karyotype of 40 acrocentric chromosomes (19 pairs of autosomal and one pair of sex chromosomes) (Gropp et al. 1972; Zima et al. 1990; Boursot et al. 1993) with a highly conserved banding pattern between the 14 species that comprises (Veyrunes et al. 2006). Additionally to the standard karyotype, many *M. m.domesticus* populations show metacentric chromosomes derived from Rb fusions (Piálek et al. 2005). Additionally, the metacentrics could evolve into new combinations of chromosomes by Whole-Arm Reciprocal Translocations (WARTs), which consist of the exchange of one arm of a metacentric chromosome by other arm of a metacentric or acrocentric chromosome (Searle 1993; Capanna and Redi 1995). As a result, a high variety of diploid numbers have been described in natural populations, ranging from 2n=22 (that is, all the autosomal chromosomes fused) to 2n=40 (Piálek et al. 2005; Hauffe et al. 2012). Rb fusions can be fixed in homozygous state in populations of a well limited geographic area forming a chromosomal race, also called metacentric race, which differs from other such races by its specific set of metacentrics (Hausser et al. 1994; Nachman and Searle 1995). In some cases, one metacentric race can contact with the standard or other metacentric races resulting in the presence of hybrids with an intermediate karyotype. These contact regions have been well studied and are called hybrid zones (Searle 1993; Hauffe et al. 2012). Almost 100 different metacentric populations have been described in western Europe and north-Africa area (Piálek et al. 2005; Hauffe et al. 2012), presenting overall 101 of the 171 possible combinations of metacentrics (Hauffe et al. 2012) and distributed in several highly localized systems identified in Scotland, Denmark, northern Switzerland, southern Switzerland, northern Italy, Croatia, Spain, central-southern Italy, Peloponnesus, mainland Greece and Madeira Island (Figure 1.11).

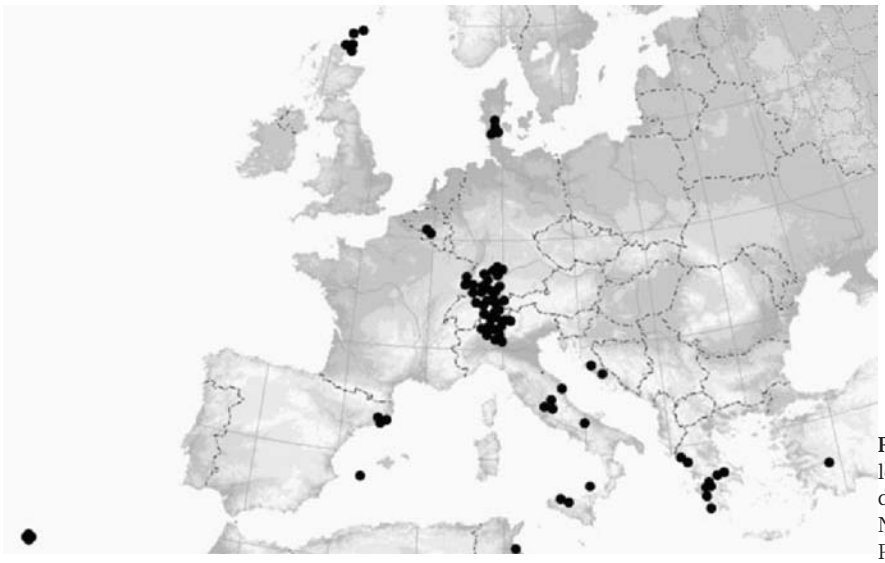


Figure 1.11. Map showing the location of metacentric populations described until 2005 in Europe and North Africa. Image adapted from Piálek and collaborators (2005)

M. m. domesticus is a primary model system for studies of chromosomal speciation, (i.e., reproductive isolation promoted by the presence of CRs, section 1.1.6) (King 1993). When two chromosomal races that differ in only few metacentrics contact, gene flow can occur without significant restrictions (Wallace et al. 2002). Nevertheless, the accumulation of a relatively high number of Rb fusions is related to hybrid subfertility or even sterility as a result of malsegregation during meiosis (Wallace et al. 1992; Castiglia and Capanna 2002). Additionally, high heterozygosity for CRs is expected to reduce fitness of hybrid mice, thus limiting the gene flow between races (Hauffe and Searle 1998; Nunes et al. 2011).

1.3.2 The Barcelona Rb system

The Barcelona Rb system is located within the provinces of Barcelona, Tarragona and Lleida (northeastern Iberian Peninsula) where house mice presenting diploid numbers ranging between 2n=27 and 2n=40 have been described (see Medarde et al. 2012 and references therein). In this population, seven different metacentric chromosomes, [Rb (3.8), (4.14), (5.15), (6.10), (7.17), (9.11) and (12.13)] have been found in an area of approximately 5,000 km^2. One exciting characteristic of this zone is that it constitutes a unique example within all house mouse Rb zones reported to date lacking a Rb race (Hauffe et al. 2012). Their metacentrics are geographically distributed with a staggered fashion leading to a progressive reduction in the diploid numbers towards the center of the Rb polymorphism zone (Gündüz et al. 2001; Sans-Fuentes et al. 2007; Medarde et al. 2012). Other particular characteristics of

this zone are that it shows a high grade of structural heterozygosity with up to 7 fusions in heterozygous state (Sans-Fuentes 2004) and the clinal distribution of the metacentric chromosomes has not varied in a decade (Gündüz et al. 2001; Gündüz et al. 2010; Medarde et al. 2012). For all this, this area is considered to be a Rb polymorphism zone rather than a typical hybrid zone which is a product of the contact of two metacentric races (Sans-Fuentes et al. 2007), taking into account the definition of Rb system given by Piálek and collaborators (2005) (i.e., "a group of Rb populations from a restricted geographical region and sharing a set of metacentrics with an apparently common evolutionary origin").

Therefore, it has been considered that the Barcelona Rb system is a particular Rb scenario that has been originated by primary intergradation, although secondary contact cannot be excluded in principle (Gündüz et al. 2001). Consequently, it may represent an example of raciation process eventually leading to the formation of a Rb race without geographic isolation (Sans-Fuentes et al. 2009). In concordance with this hypotheses, previous studies have showed that the number of Rb fusions and/or the level of structural heterozygosity alter some sensitive systems, such as skeletal morphology (Muñoz-Muñoz et al. 2006; 2011; Sans-Fuentes et al. 2009; Martínez-Vargas et al. 2014), circadian rhythm of motor activity (Sans-Fuentes et al. 2005), spermatogenesis (Sans-Fuentes et al. 2010; Medarde et al. 2015), and sperm size and shape (Medarde 2013; Medarde et al. 2013). A common result of all these studies is that standard mice (2n=40) show significant differences against Rb mice, particularly against those with a high number of fusions and a relatively high level of heterozygosity.

1.3.3 Previous studies in house mouse Rb populations

According to Hewitt (1988), hybrid zones are "natural laboratories for evolutionary studies" as they permit the investigation of a speciation process in progress and, in this case, a previous step that permits to understand which are the forces that lead to the fixation of CRs in nature. Many works have been performed in order to test both models of chromosomal speciation (*hybrid dysfunction model vs. suppressed recombination mode*l; see section 1.1.6) by using house mice from different metacentric races (Gropp et al. 1982; Tichy and Vucak 1987; Hübner and Koulischer 1990; Searle 1991; Capanna and Redi 1995; Castiglia and Capanna 2000; Britton-Davidian et al. 2002; Britton-Davidian et al. 2005; Solano et al. 2008; Sans-Fuentes et al. 2010; Medarde et al. 2015). Some of these studies deal with the effects of Rb fusions on the subfertility or sterility phenotypes in hybrids between chromosomal

races. Although there is some controversy on the effect of these mutations when its number is relatively moderate (Searle 1993; Hauffe and Searle 1998; Castiglia and Capanna 2000; Piálek et al. 2001; Sans-Fuentes et al. 2007), Moreover, significantly reduced fertility has been detected in individuals with a high number of Rb fusions (Searle 1993; Hauffe and Searle 1998; Castiglia and Capanna 2000; Piálek et al. 2001), suggesting the role of this mutations as postzygotic barrier. In the light of these results, the recombination dynamics has started to be tested in Rb populations of *M. m. domesticus*. Initial studies of chiasmata frequencies revealed the existence of a significant decrease in the recombination rates in Rb hybrids (Bidau et al. 2001; Castiglia and Capanna 2002; Dumas and Britton-Davidian 2002) and a possible higher mechanical interference of the metacentric centromeres has been proposed as the possible cause of this reduction (Dumas and Britton-Davidian 2002). Subsequent studies have been focused on the incidence of asynapsis during meiosis in mice carrying multiple Rb fusions (Manterola et al. 2009; Vasco et al. 2012), however, the meiotic and recombination dynamics in a polymorphic Rb system without the presence of a metacentric race, together with the possible influence of specific genetic factors such as *Prdm9* in these systems remains to be elucidated.

2.

OBJECTIVES

The research activity of our group is focused on the study of genome evolution and, more specifically, on the mechanisms responsible for genome reshuffling in mammals. By investigating the plasticity of mammalian genomes, much can be learned about its significance in speciation.

Despite long-standing discussions on the role of CRs in speciation, empirical evidence contrasting chromosomal speciation models (*hybrid dysfunction model vs. suppressed recombination model*) is still scarce in nature. Moreover, the recent description of mammalian hybrid sterility genes is balancing the focus on the importance of meiosis and, more specifically, recombination in evolutionary biology. In this context, the existence of house mouse wild populations with the presence of Rb fusions, offers an inestimable opportunity for the study of the role of CRs in speciation and the mechanisms that promote their appearance and fixation in nature.

Given this background, the main aim of this work was to study the mechanisms that are involved in genome reshuffling using rodents, and more specifically the house mouse as model species. This has been achieved taking advantage of the availability of whole-genome sequences of several mammalian species, together with the analysis of a house mouse Rb population, the Barcelona Rb system.

In order to reach the main goal, three specific objectives have been delineated:

1. To reconstruct genome reshuffling in rodents, paying special attention to the specific genomic signatures that characterize the distribution of evolutionary breakpoint regions (EBRs).

2. To elucidate the role of telomeres in the occurrence of Rb fusions in the Barcelona Rb system.

3. To analyze the role of Rb fusions and *Prdm9* sequence on meiotic dynamics and recombination rates in the Barcelona Rb system.

3.

MATERIALS AND METHODS

3.1 Bioinformatic analysis

3.1.1 Alignment of mammalian genomes

Eleven sequenced genomes representative of different mammalian orders such as Rodentia (*Spalax galilii, Jaculus jaculus, Heterocephalus glaber, Microtus ochrogaster, Rattus norvegicus* and *Mus musculus*), Primates (*Homo sapiens, Macaca mulatta*, and *Pongo pygmaeus*), Artiodactyla (*Equus caballus*), Carnivora (*Felis catus*) and Perissodactyla (*Bos taurus*) were included in our study (section 4.1). These species were selected on the grounds of the quality of their genomes; that is, they presented N50 values >2Mbp, being the N50 value a statistic that represents an approximate estimation of the mean scaffold lengths of a given genome.

Out of all genomes analyzed, eight (*M. ochrogaster, R. norvegicus, M. musculus, H. sapiens, M. mulatta, P. pygmaeus, E. caballus, F. catus* and *B. taurus*) were assembled into chromosomes whereas the remaining (*S. galilii, J. jaculus* and *H. glaber*) consisted of scaffold-based genomes (section 4.1, Table S1). All genomes were downloaded from the GenBank FTP site (ftp://ftp.ncbi.nih.gov/) except for *S. galilii* that was obtained from the original paper (Fang et al. 2014).

Whole-genome alignments were performed using *Satsuma Synteny* (SS) (Grabherr et al. 2010) created by the Vertebrate Genome Biology Program from the Broad Institute of MIT and Harvard (https://www.broadinstitute.org/). This algorithm performs high-sensitivity whole-genome alignments between a reference genome and a target through cross-correlation (Figure 3.1A). Thus, it establishes homologies between DNA regions that have diverged over time by a measure of similarity between sequences. In the present work, the analyses were performed considering the mouse as the reference genome (NCBIm37 assembly). In all cases, the mouse genome was compared with each of the rest of the genomes (i.e., target genome) in a pair-wise manner (Figure 3.1A). For each pair-wise genome comparison, an output with the aligned sequences was obtained for each chromosome that was then cleaned for overlapping matches using an in-house perl script.

3.1.2 Detection of regions of synteny (HSBs and SFs)

In order to characterize regions of homology between pair-wise alignments *Synteny Tracker* (ST) (Donthu et al. 2009) was used. This algorithm identifies and unifies sequence alignment homologies between two genomes establishing blocks of homology (i.e., syntenic regions) (see BOX 2 and Figure 3.1B). Based on the type of the target genome, two types of syntenic regions were obtained: (i) HSBs when the target genome was assembled into chromosomes, and (ii) Syntenic Fragments (SFs) when the target genome was only available into scaffolds. In each case, the sizes defined for each syntenic region were established depending on a minimum size threshold.

In the present work, the analysis for each pair-wise comparison was performed by triplicate, setting three different minimum sizes for syntenic regions: 100Kbp, 300Kbp and 500Kbp. This analysis at different resolutions also allowed us to infer the quality of the target genomes. In the case the number of regions of homology detected was proportional between the three resolutions (100Kbp, 300Kbp and 500Kbp), then the target genome was considered to present high sequencing quality (section 4.1, Figure S1).

3.1.3 Detection and classification of EBRs

Once HSBs and SFs were detected, the program *Evolutionary Breakpoint Analyzer* (EBA) (Farré et al., under revision) was used in order to characterize and classify EBRs in a phylogenetic context. EBA is a perl-based algorithm that identifies the coordinates of all EBRs (defined as the interval delimited by two adjacent HSB boundaries) that have occurred during the evolution of the species analyzed and classifies them in a phylogenetic context. In the case of the genomes consisting of scaffolds (*S. galilii, J. jaculus* and *H. glaber*), EBA only considers the presence of an EBR when SFs boundaries are located within the same scaffold.

In the present work, analyses were performed using the HSBs dataset obtained with ST for all pair-wise comparisons, based on the more recent mammalian phylogenetic relationships (section 4.1, Figure 1) (Meredith et al. 2011; dos Reis et al. 2012). EBA was used to detect and classify EBRs by their time of appearance in each phylogenetic lineage (Figure 3.1C), providing a score of reliability for each EBR detected as a ratio of classification confidence as well as the number of species that present the same EBR respect to the ones that present a genomic gap (regions

without sequencing data). Both values provide an estimation of reliability for each EBR, avoiding false positives.

EBA was run by using syntenic regions data (HSBs and SFs) at 300Kbp as the reference block size resolution chosen (given that it represents the mean threshold value) and a phylogenetic tree input. In order to avoid false positives, the final set of EBRs was obtained by fixing the reliability score threshold at ≥34 and the percentage of species with the EBR at ≥60% (taking into account the total number of species included in the study).

Once EBRs coordinates were established, the *Evolution Highway Comparative Chromosome Browser* (EH) (http://evolutionhighway.ncsa.uiuc.edu/) was used in order to visualize all EBRs detected in each phylogenetic lineage. EH is a tool that creates graphic representations for comparative genomic data for each reference chromosomes, in our case the mouse genome, compared with the target species in a pair-wise manner (Figure 3.1D). Additionally, this representation allows the confirmation of each EBR classification characterized by EBA.

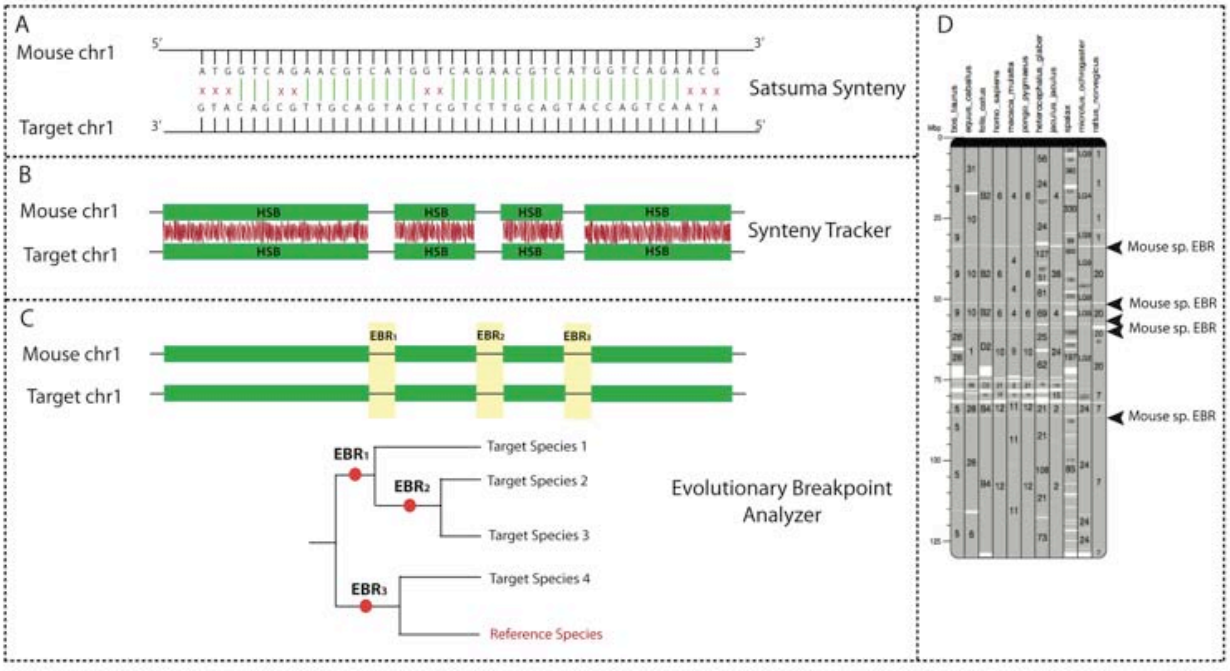


Figure 3.1. Schematic representation of the analysis performed by each bioinformatic tool used in this study: (A) *Satsuma Synteny* identifies sequence homologies (depicted with a green line) between a reference genome (mouse) and a target genome. (B) *Synteny Tracke*r unifies genome sequence homologies in syntenic blocks: Homologous Synteny Blocks (HSBs) or Syntenic Fragments (SFs) depending on how the target genome is assembled (into chromosomes or scaffold-based, respectively). In this case, HSBs are represented in green. (C) *Evolutionary Breakpoint Analyzer* (EBA) identifies the regions of disruption of homology as EBRs (depicted in yellow) and classifies them in a phylogenetic context based on a species tree provided. (D) E*volution Highway* chromosome representation with of homology regions (HSBs and SFs, in grey) and EBRs (in white) of mouse chromosome 10 with each of the target genomes studied in this analysis

3.1.4 Analysis of genomic features

By using an in-house Perl script, the mouse genome was divided into 10Kbp non-overlapping windows and merged with the EBRs and HSBs coordinates resulted from the analysis using ST and EBA. With this approach, the mouse genome was classified into four different genomic regions: (i) EBRs, (ii) HSBs, (iii) interphase regions (genomic 10Kbp non-overlapping windows overlapping with the start or the end coordinates of any EBR) and (iv) 100Kbp regions upstream or downstream the EBRs coordinates (section 4.1, Figure 3A). This file was merged with the data obtained for different genomic features (gene content, recombination rates and constitutive LADs), as explained below, in order to further analyze their distribution across the mouse genome.

3.1.4.1 Gene content and ontology

Data used for the gene content analysis was obtained from the Biomart portal (http://www.ensembl.org/biomart/martview/) and consisted of two main files: (i) all mouse RefSeq genes and (ii) all mouse protein-coding genes datasets (NCBIm37 release). Gene coordinates were then merged across the mouse genome in 10Kbp non-overlapping windows previously classified as EBRs, HSBs, interphases and 100Kbp upstream or downstream EBRs. Kruskal-Wallis non-parametric test was performed with JMP statistical package (release 7.1) in order to compare mean gene number for each genomic region.

Subsequently, gene ontology analyses were performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID, Huang et al. 2009). DAVID is an integrated database and analytic tool that permits the extraction of biological features or enrichments associated with a specific gene list query. Thus, it is a useful tool for the identification of overrepresented biological terms contained in any given region compared to the rest of the genome. Additionally, DAVID provides Benjamini's statistical test to control false positives and sort biological terms by its degree of enrichment and the Expression Analysis Systematic Explorer score (EASE score) a modified Fisher's exact p-value, as the threshold to consider a term significantly overrepresented (fixed on <0.05).

3.1.4.2 Recombination rates

Genetic maps of recombination rates for the house mouse was extracted from Brunschwig and collaborators (2012) who provided high-resolution recombination rates estimates across the mouse genome by SNPs mapping that were generated as part of the Mouse Genome Project (http://www.sanger.ac.uk/resources/mouse/genomes/). Two different datasets were used to detect recombination hotspots. The first of them included SNPs data obtained from 12 mouse-inbred strains (129S5/SvEvBrd, AKR/J, A/J, BALB/cJ, C3H/HeJ, C57BL/6NJ, CBA/J, DBA/2J, LP/J, NOD/ShiLtJ, NZO/HILtJ, and WSB/EiJ). The second dataset consisted of 100 classical strains genotyped with the Mouse Diversity Array (Yang et al. 2011).

Mean recombination rates (expressed as $4N_e r$/Kbp) (see BOX 3) were distributed into 10Kbp non-overlapping windows and merged with the four genomic regions described above (section 3.1.4). Kruskal-Wallis non-parametric test was performed with JMP statistical package (release 7.1) in order to compare mean recombination rates between EBRs, interphases, 100Kbp adjacent regions and HSBs.

3.1.4.3 Constitutive lamina associated domains (cLADs)

Lamina associated domains are genomic regions associated to the Nuclear Lamina (NL), a protein lining the nuclear envelope of eukaryotic cells. These regions provide information regarding the spatial architecture of the DNA inside the nucleus given that genomic regions in contact with the NL are normally positioned in the periphery of the cell nucleus. Meuleman and collaborators (2013) published a refined map of the genomic regions associated to two types of NL (A and B) in four different mouse cell types (Lamina A in astrocytes and neural precursor cells and B1 in wild type and Oct1 knockout embryonic fibroblasts). Those lamina regions that were common in all cell types were considered as constitutive LADs (cLADs). In our study, cLADs genomic coordinates were extracted from Meuleman and collaborators (2013) and distributed in 10Kbp non-overlapping windows using in-house Perl scripts. This dataset was merged with the four genomic regions described above and statistically analyzed as it was done with gene density and recombination rates.

3.1.5 Permutation tests

Genome-wide association analysis of each genomic signatures included in our study (gene density, recombination rates and cLADS) were performed with *RegioneR* (Diez-Villanueva et al. 2015), a R-implemented algorithm. RegioneR compares, through randomization tests, the number of overlaps between a query and a reference region set to the distribution of the number of overlaps obtained by randomizing the regions of interest for each chromosome (http://gattaca.imppc.org/regioner/). In this work, 10,000 permutation tests were run per chromosome for the following genomic signatures coordinates: (i) all RefSeq genes, (ii) protein coding genes (iii) recombination rates and (iv) cLADs by comparing them with EBRs and HSBs coordinates. In all cases, significance was considered when $p<0.05$.

3.2 Biological samples

A total number of 31 specimens of *M. m. domesticus* were live-trapped in 10 different populations belonging to the Barcelona Rb polimorphism system (Figure 3.2). The sampled area consisted of four populations (Castellfollit del Boix, Vacarisses, Arbeca and Santa Perpètua de la Mogoda) with standard karyotype (2n=40) and six populations (L'Ametlla de Segarra, Sant Sadurní d'Anoia, El Papiol, Viladecans, Prat de Llobregat and Castelldefels) with Rb mice with fusions ranging from 1 to 12 (2n=28-39). Specimens were captured using Sherman animal traps that were placed at the evening in different locations inside the farms and collected the next morning. Legal permission for animal capture was granted by the "Departament de Medi Ambient" of the "Generalitat de Catalunya".

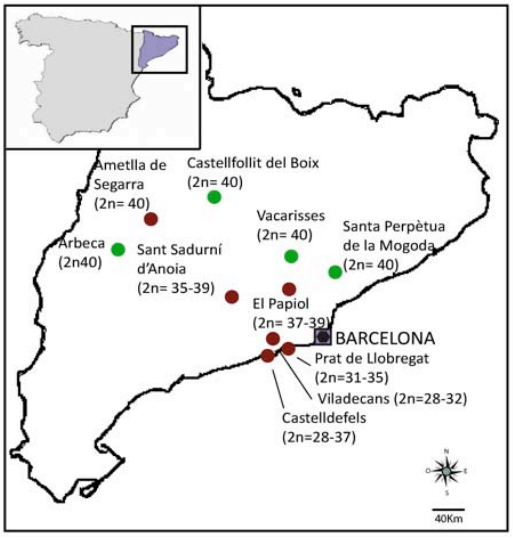


Figure 3.2. Map of the Barcelona Rb system: All sampled populations in this study are represented differentiating standard populations (in green) and Rb populations (in red). For each population, the diploid number range is indicated.

All captured mice were immediately transferred to the laboratory in order to house them under controlled conditions for three days. The following samples were obtained for each specimen: testicular tissue, muscle from posterior leg and conjunctive tissue from intercostal muscle. Animals were handled by Dr. Medarde (see Medarde 2013) in compliance with the guidelines and ethical approval of the "Comissió d'Ètica en L'Experimentació Animal i Humana" of the Universitat Autònoma de Barcelona and by the "Departament d'Agricultura, Ramaderia, Pesca, Alimentació i Medi Natural" of the "Generalitat de Catalunya" (reference of the experimental procedure authorization: DAAM 6328).

3.3 Molecular biology techniques

3.3.1 Bacterial Artificial Chromosome (BAC) selection and purification

Bacterial Artificial Chromosomes (BACs) are constructs that allow cloning DNA fragments of interest, ranging in size from 150Kbp to 350Kbp, in a bacterial vector (normally *E. coli*). By using BACs as a probe in FISH experiments, the region of interest can be identified on chromosomal preparations. In the present work, this method was used to identify meiotic chromosomes on mouse spermatocyte spreads previously treated with Immunofluorescence (IF) (see section 3.4.3). BACs corresponding to mouse chromosomes 4, 9, 11, 12, 13 and 14 were selected from the UCSC genome Browser (http://genome.ucsc.edu). For each mouse chromosome, a set of overlapping BAC clones were purchased from the BACPAC resources center in Children's Hospital Oakland Research Institute(CHORI, https://bacpac.chori.org/).

3.3.1.1 Vector plasmid culture and DNA extraction

Materials

- Plasmid vector supplied by CHORI in Luria-Bertani (LB) broth agar stab
- LB medium powder (Liofilchem)
- Agar powder (Sigma Aldrich)
- Cloramphenicol solution (34mg/ml, Sigma Aldrich)
- Glycerol solution (60%, Sigma Aldrich)

- Isopropanol (Sigma Aldrich)
- Absolute ethanol (Merck)
- MilliQ water
- Plasmid midi Kit (Quiagen)
- Petri dishes
- 50ml ultracentrifuge tubes
- 50ml tubes
- 500ml erlenmeyer flasks
- 2ml cryotubes
- Nanodrop H-1000 (Thermo Scientific)

Protocol

- Prepare LB agar petri dishes with cloramphenicol solution at a final concentration of 12.5µg/ml.
- Inoculate the bacterial vector on a LB agar petri dish and incubate overnight at 37ºC.
- Pick up a single colony from the petri dish and inoculate into 10ml of liquid LB culture with cloramphenicol (12.5µg/ml).
- Incubate overday at 37ºC in agitation (300rpm).
- Add 300µl of the previous culture to fresh 100ml of liquid LB medium with cloramphenicol (12.5µg/ml).
- Incubate overnight at 37ºC in agitation (300rpm).
- Prepare a glycerinate of the bacterial culture for future uses. Mix 670µl of the liquid LB culture with 330µl of glycerol (60%) in a cryotube. Store at -80ºC.
- Obtain the cell pellet by centrifuging the bacterial culture for 30 minutes at 6000*xg*.
- Purify DNA using the Quiagen plasmid midi kit following manufacturer instructions.
- Dilute the DNA pellet in milliQ water.
- Estimate DNA concentration with Nanodrop.
- Store at -20ºC until use.

3.3.1.2 DNA labeling by nick translation

Materials

DNA solution (>50ng/µl)

- Buffer 10X: tris-HCl (0.5M), MgCl2 (50mM), Bovine Serum Albumin (BSA, 0.5mg/ml)
- Deoxynucleotide Triphosphate (dNTP) mix: dATP (0.5mM), dCTP (0.5mM), dGTP (0.5mM) and dTTP (0.1mM)
- ß-mercaptoethanol (10mM, Sigma Aldrich)
- Klenow enzyme (Roche Diagnostics)
- DNase (2000U/ml, Roche Diagnostics)
- Deoxyuridine Triphosphate-Digoxigenin (dUTP-DIG, Roche Diagnostics) or dUTP-Cyanine 3 (dUTP-Cy3, GE Healthcare)
- MilliQ water
- Ethylenediamine-tetraacetic acid (EDTA 0.5M, Sigma Aldrich)
- 1.5ml eppendorf tubes
- Thermoblock (Eppendorf)

Protocol

Prepare nick translation mix:

- 1µg of DNA
- 5µl of buffer 10X
- 5µl of ß-mercaptoethanol
- 5µl of dNTP mix
- 2.5µl of dUTP-DIG or dUTP-Cy3
- 1µl of DNase
- 1µl of Klenow enzyme
- MilliQ water up to a total volume of 50µl
- Incubate overnight at 15ºC in a thermoblock.
- Stop the reaction with 1µl of EDTA (0.5M).
- Store at -20ºC until use.

3.3.2 Genomic DNA purification

Mouse genomic DNA was obtained from either cell cultures (see section 3.4.1) or muscle tissue (stored in ethanol at -20ºC) using the standard phenol-chloroform

protocol as follows.

Materials

- Extraction buffer: Tris-HCl (10mM, pH 8.0), EDTA (10mM, pH 8.0), NaCl (150mM), SDS (0.5%)
- RNase cocktail (Ambion)
- Proteinase K (Ambion)
- Phenol solution: equilibrated with Tris HCl (10mM, pH 8.0), EDTA (1mM, Sigma Aldrich)
- Phenol-Chloroform:IAA MB Grade (Ambion)
- Chloroform:IAA MB grade (Sigma Aldrich)
- Sodium acetate buffer solution (NaAc, 3M, pH 5.2, Sigma Aldrich)
- Absolute ethanol (Merck)
- Ethanol 70%
- MilliQ water
- 2ml eppendorf tubes
- 1.5ml eppendorf tubes
- Centrifuge 15ml tubes
- Eppendorf centrifuge
- Thermoblock (Eppendorf)
- Nanodrop H-1000 (Thermo Scientific)

Protocol

- Add 0.75ml of extraction buffer, 2.5µl of RNase cocktail and 3.75µl of proteinase K to 2ml eppendorf tube containing the sample.
- Incubate in a thermoblock overnight at 37ºC in agitation (300rpm).
- Add 0.75ml of phenol solution.
- Mix by hand until the two phases are mixed.
- Centrifuge 15 minutes at 5000*xg*.
- Recover the aqueous phase (the upper one) and transfer to a new 2ml tube.
- Add 0.75ml of phenol solution.
- Mix by hand until the two phases are mixed.
- Centrifuge 15 minutes at 5000*xg*.
- Recover the aqueous phase (the upper one) and transfer to a new 2ml tube.
- Add 0.75ml of phenol-chloroform and swirl gently by hand until the two phases are mixed.
- Centrifuge 10 minutes at 5000*xg*.

- Recover the aqueous phase (the upper one) and transfer to a new 2ml tube.
- Add 0.75ml of chloroform:IAA MB grade and swirl gently by hand until the complete emulsion of the two phases.
- Centrifuge 10 minutes at 5000*xg*.
- Recover the aqueous phase and transfer to a 15ml tube.
- Add 0.1 volumes of NaAc (3M) and 2 volumes of absolute ethanol and transfer the solution to a new 1.5ml eppendorf.
- Mix slowly by inversion until DNA precipitates.
- Centrifuge 1 minute at 5000*xg*.
- Eliminate supernatant and add 0.75ml of 70% ethanol.
- Centrifuge 1 minute at 5000*xg*.
- Eliminate the supernatant and leave to air dry for a few minutes.
- Dilute the DNA pellet in sterile milliQ water.
- Estimate DNA concentration with Nanodrop.
- Store at -20ºC until use.

3.3.3 *Prdm9* amplification

In this work, exon 12 of the *Prdm9* gene, from repeat #2 to the C-terminal domain, was amplified and sequenced by Sanger sequencing in all the mice specimens included in the study. Exon 12 contains the ZnF domain array that recognizes and methylates the specific DNA sequences (Parvanov et al. 2010).

Materials

- DNA solution (>50ng/µl)
- MilliQ water
- Buffer 10X (20mM, Takara Bio Inc.)
- dNTP mix (2.5mM each dNTP, Invitrogen)
- Primer 5', fl1822U24 (10µM, Invitrogen)
- Primer 3', 2848L23 (10µM, Invitrogen)
- Dimethyl Sulfoxide (DMSO, Sigma Aldrich)
- ExTaq (TaKaRa ExtaqTM,Takara Bio inc.)
- Agarose D1 low EEO (Condalab)
- Ethidium Bromide (EtBr, Sigma Aldrich)
- Buffer 1XTAE for electrophoresis
- DNA ladder 100bp (Invitrogen)

- Loading Buffer
- 1.5ml eppendorf tubes
- 0.2 eppendorf tubes (Fisher Scientific)
- PCR thermocycler (Bio-Rad)

PCR amplification protocol

- Mix the following reagents in a 1.5ml eppendorf tube (volumes are indicated for a single sample):
  - 4µl of buffer 10X
  - 3.2µl of dNTP mix
  - 1µl of each primer (5' and 3')
  - 1.6µl of DMSO
  - 0.25µl of ExTaq
- MilliQ water up to a total volume of 36µl
- In a 96 well plate, add 36µl of PCR mix to each well, which already contains 3µl of DNA sample previously added.
- Run the following program in the thermocycler:
  - Initial denaturation: 95ºC, 3 minutes
  - Denaturation: 95ºC, 30 seconds
  - Annealing: 56ºC, 30 seconds
  - Elongation: 72ºC, 90 seconds

  (repeat 30 cycles)
  - Final elongation: 72ºC, 10 minutes
  - Infinite hold: 10ºC
- Store at 4ºC until use.

Genotyping protocol

- Prepare a 2X agarose gel for electrophoresis: 100ml TAE buffer 1X, 2g agarose, 2µl of EtBr.
- Prepare samples: 5µl of loading buffer, 5µl of DNA sample.
- Prepare DNA ladder solution: 2µl of loading buffer, 1µl of DNA ladder (100bp).
- Run the gel for 2.5 hours at 75V.
- Capture gel image.
- Identify different alleles by size.

3.3.4 DNA purification from gel bands and Sanger sequencing

Depending on whether the specimen was homozygote or heterozygote for the *Prdm9* allele, different procedures were followed to obtain DNA for subsequent sequencing. In the case that mice were heterozygous for the *Prdm9* allele, amplified bands were cut from the agarose gel and DNA was purified using a DNA extraction kit (IllustraTM GFXTM). In the case of samples homozygote for the *Prdm9* gene, Sanger sequencing was performed directly from PCR products.

Materials

- Surgical blade
- Transilluminator UV
- DNA extraction kit (IllustraTM GFXTM PCR DNA and gel band purification kit, GE Healthcare)
- 1.5ml eppendorf tubes
- 0.2 eppendorf tubes (Fisher Scientific)
- Nanodrop H-1000 (Thermo Scientific)

DNA purification from gel bands protocol

- Localize the bands corresponding to heterozygote samples with a UV transilluminator.
- Cut the bands with a surgical blade and transfer to a 1.5ml eppendorf tube.
- Store the bands for 1 week maximum at -20ºC until DNA extraction.
- Purify DNA by using gel band DNA extraction kit (IllustraTM GFXTM PCR DNA and gel band purification kit, GE Healthcare) following manufacturer instructions.
- Calculate DNA concentration with Nanodrop.
- Store DNA at -20ºC until sequencing.

Sequencing

Samples were sent to Macrogen Europe (https://dna.macrogen.com/) for Sanger sequencing procedure. All reads (forward and reverse reads) and their corresponding chromatograms were analyzed with Bioedit sequence alignment editor (version 7.2.5).

3.4 Cell biology techniques

3.4.1 Cell cultures and chromosome harvest

Fibroblast primary cell lines were established from mice from the Barcelona Rb system using conjunctive tissue biopsy obtained from the intercostal muscle as described below. For those animals from which cell lines were not successfully established, chromosomes were obtained from bone marrow as previously described (Medarde 2013).

Materials

- Transport medium: 100ml Dulbecco's Modified Eagle Medium (DMEM, Life technologies), 10ml Foetal Bovine Serum (FBS, Life technologies), 1ml antibiotic-antimycotic (Life technologies), 0.7ml gentamicin (Life technologies)
- Washing medium: 100ml of Phosphate Serum Saline 1X (PBS), 0.3g BSA (Sigma Aldrich), 0.1g glucose (Sigma Aldrich), 0.7g gentamicin, 1ml antibiotic-antimycotic
- Cell culture medium (DMEM)
- Sterile 1XPBS (Life technologies)
- Trypsin-EDTA (Life technologies)
- DMSO (Sigma Aldrich)
- Colcemid (KarioMax®, Invitrogen)
- Hypotonic solution (KCl, 0.075M)
- Fixative solution: Methanol (Merck), Acetic Acid (Merck) at 3:1 concentration (freshly prepared)
- Absolute ethanol (Merck)
- 1.5ml eppendorf tubes
- Superfrost slides (Waldemar Knittel)
- Surgical blade
- T25 cell culture flasks (Orange Scientific)
- Centrifuge 15ml tubes
- Laminar flow cabin for cell culture
- Centrifuge

Cell culture protocol

- Immediately after the specimen was sacrificed, a conjunctive tissue sample from the intercostal muscle was collected and transported in ice to the cell culture laboratory in a 15ml tube with 5ml of transport medium.
- Working in a laminar flow cabin for cell culture, wash the tissue three times in washing medium.
- Cut the tissue into small explants with a surgical blade and place them inside a cell culture flask.
- Add 10ml of DMEM supplemented with 15% of FBS and incubate at 37ºC (CO_2 10%).
- Once fibroblasts start to grow in monolayer from the tissue, remove the explants.
- When cell culture reaches confluence, trypsinize and expand (1:2) in two T25 flasks.
- Explants and cells were cryo-preserved in liquid nitrogen with DMSO:FBS (concentration 1:9) at early passages.

Chromosome harvest protocol

- In order to arrest cell cycle at metaphase stage, add 80µl of colcemid (10µg/ml) to 10ml medium when cultures are at 80% of confluence.
- Incubate for 2 hours at 37ºC.
- Trypsinize the cell culture and centrifuge in a 15ml tube (5 minutes at 1400rpm).
- Break the cell pellet by vortexing and add 10ml of hypotonic solution (pre-warmed at 37ºC) very slowly.
- Incubate the cells for 25 minutes at 37ºC inverting every 5 minutes.
- Centrifuge 5 minutes at 1400rpm.
- Add 5ml of fixative solution.
- Centrifuge 5 minutes at 1400rpm.
- Wash the pellet twice by adding 5ml of fixative solution and centrifuge (5 minutes at 1400rpm).
- Dilute the cells in 1ml of fixative solution and store at -20ºC in an eppendorf tube until use.
- Prepare chromosomal spreads by placing a drop (approximately 20µl) of chromosome suspension on a superfrost slide.
- Leave the fixative solution to air dry and store the slides at -20ºC until use.

3.4.2 Spermatocyte spreads

Mice testicular biopsies were obtained immediately after animal dissection, and maintained at -80º with isopentane (Sigma Aldrich) for the first 24 hours in order to avoid temperature fluctuations.

Materials

- Buffer 1XPBS (pH 7.2-7.4)
- Fixative solution: Paraformaldehyd (4%, pH 9.8, Sigma Aldrich), Triton X-100 (0.15%) in milliQ water
- Lypsol 1% freshly prepared in milliQ water
- Washing solution: 1% PhotoFlo (Kodak) in milliQ water
- 4',6-diamidino-2-phenylindole (DAPI) solution: DAPI (125ng/ml) in antifade (Vectashield, Vector Laboratories)
- Humidified chamber
- Superfrost adhesive slides (Waldemar Knittel)
- 22X60 coverslips
- Surgical blade
- Petri dishes

Protocol

- Obtain a cell suspension from the testicular biopsy by scattering the tissue in a petri dish in 20µl of cold 1XPBS.
- Transfer the cell suspension on cold superfrost slides.
- Add 90µl of lypsol 1% and leave it for 14 minutes in a humidified chamber.
- Add 100µl of paraformaldehyd 4% and leave it for 20 minutes in a humidified chamber.
- Let the slides to air dry for 30 minutes. It is important not to let all the solution to dry completely as the proteins could be damaged.
- Wash the slides 3 times with washing solution and let them to air dry.
- Add 20µl of DAPI and place a coverslip.
- Verify with an epifluorescence microscope the quality of the spreads (number and morphology of cells).
- Store at -20ºC until use.

3.4.3 Immunofluorescence (IF)

The IF technique allows the identification of proteins of interest by using different sets of primary antibodies that recognize protein epitopes. Such epitopes are subsequently revealed with fluorescent secondary antibodies. In this work, different proteins involved in meiotic division process were analyzed on mouse spermatocytes. These proteins included:

- MLH1: protein involved in recombination COs at pachynema stage (Anderson et al. 1999).
- SYCP3: this protein forms the lateral element of the SC (Henderson and Keeney 2005).
- Histone 3 trimethylated at lysine 9 (H3K9me3): is considered a marker of constitutive heterochromatin (Hublitz et al. 2009).
- RPA: recognizes single stranded DNA (ssDNA) during DSBs formation and contributes to the replication process with the homologous chromosome during meiotic recombination (Plug et al. 1997).
- γH2AX: this protein is phosphorylated during the formation of DSBs and it is involved in gene inactivation of asynapsed regions (Baarends et al. 2005).
- Serum CREST: Human serum of Complication of Raynaud's phenomenon, Esophaegal dysfunction, Sclerodactily and Telangiectasia (CREST) usually used for the identification of centromeres. In this case we identificate CENP-B protein, used to identify the centromere (Fachinetti et al. 2013)

Materials

- Blocking solution (PTBG): 1XPBS, Tween-20 (0.05%, Sigma Aldrich)
- Fixative solution: formaldehyd 1% (Scharlau) in 1XPBS
- MilliQ water
- Buffer 1XPBS (pH 7.2-7.4)
- Buffer Saline Sodium Citrate 2X (2XSSC, pH 7)
- DAPI solution: DAPI (125ng/ml) in antifade (Vectashield, Vector Laboratories)
- Set of primary antibodies. All antibodies were purchased from Abcam, except the γH2AX that was obtained from Millipore. The serum CREST was kindly donated by Dr. M. Fritzler (Calgary University, Canada):
  - Anti-SYCP3 (mouse or rabbit, concentration 1:400 and 1:600,

respectively)
  - Anti-H3K9me3 (rabbit, concentration 1:50)
  - Anti-MLH1 (mouse, concentration 1:100)
  - Anti-RPA (mouse, concentration 1:200)
  - Anti-γH2AX (mouse, concentration 1:200)
  - Serum CREST (human, concentration 1:200)
- Set of secondary antibodies conjugated with Cyanine 3 (Cy3), Cyanine 5 (Cy5) or Fluorescein Isothiocyanate (FITC). All secondary antibodies were purchased from Jackson Immunoresearch and were applied at a concentration of 1:200.
  - Anti-Rabbit Cy3
  - Anti-Rabbit FITC
  - Anti-Mouse FITC
  - Anti-Human Cy5
- Humidified chamber
- Parafilm

Protocol

- Remove the coverslip by washing the slides in milliQ water for 10 minutes in agitation (100rpm).
- Wash the slide for 10 minutes with PTBG in agitation (100rpm).
- Apply 100µl per slide of the primary antibodies diluted in PTBG, place a parafilm coverslip and incubate overnight in a humid chamber (4ºC).
- Wash the slides twice for 5 minutes in PTBG at 37ºC in agitation (100rpm).
- Apply 100µl per slide of the secondary antibodies mix with PTBG, cover it with a piece of parafilm and incubate 1 hour in a humidified chamber at 37ºC.
- Wash the slides twice for 5 minutes in PTBG in agitation (100rpm).
- Wash the slides 10 minutes in fixative solution.
- Wash the slides in 1XPBS for 5 minutes twice.
- Add 20µl per slide of DAPI solution.

3.4.4 Fluorescence *in situ* Hybridization (FISH)

Different FISH protocols were used in this study, depending on the type of sample used (mitotic or meiotic chromosomes) and the type of probe hybridized (commercial whole-chromosome paintings, BAC probes or telomeric probes).

3.4.4.1 FISH with chromosome painting on metaphase chromosomes

In this protocol, commercial whole-chromosome painting probes were hybridized on mouse metaphase chromosomes in order to identify specific chromosomes involved in Rb fusions.

Materials

- Pepsin solution: pepsin (0.5%, Sigma Aldrich) in HCl (0.01M)
- Mouse painting probes for chromosomes 4, 9, 11, 12, 13 and 14 (XCyting, MetaSystems)
- Buffer 1XPBS (pH 7.2-7.4)
- $MgCl_2$ solution: $MgCl_2$ 10% in 1XPBS
- Fixative solution: Paraformaldehyd 50%, $MgCl_2$ 10% in 1XPBS
- Dehydration ethanols (70%, 90% and 100%)
- Washing solution 1: 0.4XSSC (pH 7-7.5)
- Washing solution 2: 2XSSC, Tween-20 (0.05%, pH 7.0)
- MilliQ water
- DAPI solution: DAPI (125ng/ml) in antifade (Vectashield, Vector Laboratories)
- 22X22 coverslips
- Thermoblock (Eppendorf)
- Humidified chamber

Slides pre-treatment protocol

- Incubate slides at 65ºC for 2 hours.
- Incubate slides for 30 minutes in pepsin solution pre-warmed at 37ºC.
- Wash in 1XPBS for 2 minutes.
- Wash in $MgCl_2$ solution for 5 minutes.
- Wash in fixative solution for 5 minutes.
- Wash in 1XPBS for 5 minutes.
- Dehydrate the slides in a gradient of ethanol (70%, 90% and 100%) for 2 minutes each.
- Store in cold absolute ethanol (4ºC) until use.

Hybridization protocol

- Let slides to air dry.
- Add 10µl of the commercial probe mixture (5µl probe, 5µl buffer, all supplied by the manufacturer) on the metaphase slide and cover with a small coverslip (22X22mm).
- Denature on a thermoblock at 75ºC for 3 minutes.
- Incubate in a humidified chamber at 37ºC overnight.
- Remove the coverslip with washing solution 1 at 72ºC for 2 minutes.
- Rinse the slide for 30 seconds with washing solution 2.
- Apply 20µl of DAPI solution and add a coverslip.

3.4.4.2 FISH with BACs on spermatocytes

This protocol was applied to hybridize mouse BAC probes previously selected (section 3.3.1) and fluorescently labeled (see section 3.3.2) on mouse spermatocyte spreads. This allowed for the identification of specific meiotic chromosomes that were previously analyzed by IF (see section 3.4.3).

Materials

- Buffer 1XPBS (pH 7.2-7.4)
- Buffer 2XSSC (pH 7)
- Mouse DNA COT-1 (Invitrogen)
- Salmon sperm DNA (Invitrogen)
- NaAc (3M)
- Absolute ethanol (Merck)
- Dehydration ethanols (70%, 90% and 100%)
- Ethanol 70% at 4ºC
- Denaturation solution: Formamide (70%) in 1XPBS (pH=7.4).
- Washing solution: Formamide (50%) in 1XPBS (pH=7.4).
- Hybridization buffer: Formamide (50%), dextran sulfate (1%) in 20XSSC.
- PBD buffer: 1l of milliQ water, 1g Na_2CO_3, 5ml Igepal (CA-630®, Sigma Aldrich).
- DAPI solution: DAPI (125ng/ml) in antifade (Vectashield, Vector Laboratories).
- DNA probe labeled with Digoxigenin (DIG) or Cy3 (see section 3.3.1.2)
- Anti-DIG-FITC in 1XPBS with BSA (concentration 1:150).
- Parafilm
- 1.5ml eppendorf tubes

- 20x60 coverslips
- Humidified chamber (37ºC)
- Thermoblock (Eppendorf)

DNA probe denaturation protocol

- DNA precipitation mix:
  - 1µg of the labeled DNA
  - 10µl of mouse DNA COT-1
  - 10µl of salmon sperm DNA
  - 0.1 volumes of NaAc 3M
  - 3 volumes of EtOH 100%
- Incubate at -20ºC overnight.
- Centrifuge at 13,300rpm for 30 minutes.
- Wash the DNA pellet twice with ethanol 70% and let to air dry.
- Add 14µl of hybridization buffer.
- Denature for 8 minutes at 74ºC in the thermoblock.

Spermatocyte slides treatment and hybridization

- Rinse the slides in 1XPBS in agitation (100rpm).
- Wash in 1XPBS for 5 minutes in agitation (100rpm).
- Rinse in the dehydration ethanols (70%, 90% and 100%) for 2 minutes in each solution.
- Denature the slides in denaturation solution at 74ºC for 3 minutes.
- Immediately after denaturation, wash the slides in cold ethanol (70%) for 2 minutes.
- Dehydrate the slides in a gradient of ethanol (70%, 90% and 100%) for 2 minutes each.
- Add 14µl of DNA probe mix on the slide and place a coverslip. Incubate overnight in a humidified chamber at 37ºC.

Post-hybridization washes protocol

- Remove the coverslip carefully and wash the slides 3 times for 5 minutes in washing solution at 45ºC.
- Repeat the 3 washes in 2XSSC at 45ºC.
- If the probe is directly labeled with Cy3:
  - Let slides to air dry and add 20µl of DAPI solution and add a coverslip.
- If the probe is conjugated with DIG:

- Add 100µl of anti-DIG-FITC and place a parafilm coverslip.
- Incubate in a dark humidified chamber at 37ºC for 30 minutes.
- Wash the slides in PBD buffer twice for 3 minutes.
- Let slides to air dry.
- Add 20µl of DAPI solution per slide and add a coverslip.

3.4.4.3 Quantitative-Fluorescence *in situ* Hybridization (Q-FISH) on metaphase chromosomes

The Q-FISH is a technique that allows the analysis of DNA telomeric length directly on metaphase chromosomes. This technique is based on fluorescence *in situ* hybridization of labeled synthetic DNA mimics called Peptide Nucleic Acid (PNA) oligonucleotides. The measuring of fluorescence intensity signal using an specific software is used as an estimation of telomeric length (TFL-TeloV2, BC Cancer Research Center, Canada) (Poon et al. 1999).

Materials

- Chromosome spreads previously fixed and stored at -20ºC (see section 3.4.1)
- Pepsin solution: pepsin (0.5%, Sigma-Aldrich) in HCl (0.01M)
- Formaldehyd 37% (Scharlau)
- Buffer 1XPBS (pH 7.2-7.4)
- Buffer 2XSSC (pH 7)
- Dehydration ethanols (70%, 85% and 100%, at 4ºC)
- PNA probe $(CCCTAA)_3$(TelC)(Panagene)
- Hybridization buffer: $NaHPO_4$ (10mM, pH 7.4), NaCl (10mM), Tris (20mM, pH 7.5), formamide (70%, Sigma Aldrich)
- Washing solution 1: Tween-20 (0.1%) in 1XPBS
- Washing solution 2: Tween-20 (0.1%) in 2XSSC
- DAPI solution: DAPI (125 ng/ml) in antifade (Vectashield, Vector Laboratories)
- 22X22mm coverslips
- Humidified chamber
- Thermoblock (Epperndorf)

Protocol

- Incubate the slides for 10 minutes at 67ºC.
- Rehydrate the slides in 1XPBS for 15 minutes.
- Fix the cells with formaldehyde 4% in 1XPBS for 4 minutes.
- Wash the slides 5 minutes in 1XPBS (twice).
- Incubate the slides in pepsin solution for 4 minutes at 37ºC.
- Wash in 1XPBS for 3 minutes twice.
- Dehydrate the slides in a gradient of ethanol (70%, 90% and 100%) for 1 minute in each solution.
- Let the slides to air dry.
- Add 15µl of PNA probe at 800ng/ml in hybridization buffer and add a coverslip.
- Denature in a thermoblock for 5 minutes at 80ºC.
- Incubate for 90 minutes in a humidified chamber.
- Rinse in washing buffer 1 and remove the coverslip.
- Incubate in washing buffer 1 for 20 minutes at 57ºC.
- Rinse in washing buffer 2 for 1 minute.
- Add 20µl of DAPI solution per slide and add a coverslip.

3.4.5 Image processing and analysis

Microscopy image capturing and processing

All chromosomal preparations obtained from both IF and FISH protocols were visualized with an epifluorescence microscope (model Zeiss Axioskop) equipped with the appropriate filters for FITC, Cy3, Cy5 and DAPI detection together with a charged coupled device camera (ProgResR CS10Plus, Jenoptik). Images were captured and produced with the ProgResR software (2.7.7). All analyses were performed blindly (a numerical code was assigned to each specimen in order to avoid unintentional scores bias) with Adobe Photoshop CS (version 8.0).

Recombination analysis

The analysis of meiotic DSBs and COs distribution in mouse meiotic chromosomes was performed using *Micromeasure* (version 3.3) (Reeves 2001), a tool designed to provide quantitative distance estimations from cytogenetic microscopic images. *Micromeasure* was used to obtain recombination maps of the mouse chromosomes identified by FISH (4, 9, 11, 12, 13 and 14) (see section 3.4.4.2). Thus, it permitted to calculate the medium relative distance (as a percentage of the SC length) to the centromere for each MLH1 or RPA *foci* analyzed. These data were used

to construct cumulative sequence plots in order to obtain chromosomal distributions of MLH1 and RPA *foci* along SCs. Distribution plots were then compared by using Kolmogorov-Smirnov Test. *Micromeasure* was also used for the analysis of the signal area of H3K9me3 overlapping the SC for each chromosomal arm (measured in micrometers).

Telomere length analysis

The TFL-TeloV2 software (BC, Cancer Research Center, Canada) (Poon et al. 1999) is designed to calculate telomere length (expressed as Telomere Fluorescence Units, TFUs) from Q-FISH images. It compares the intensity of each telomere respect to the intensity of the background and gives the telomere length estimation for each cell analyzed. This program allowed comparing q- and p- arms telomeres length for each metaphase giving the estimation of telomere length in TFUs.

4.

RESULTS

4.1 Reconstruction of genome reshuffling in rodents

The reconstruction of genome reshuffling during the evolution of species is of special interest for understanding the mechanisms that promote genome evolution. Although several theoretical models have tried to explain the role of different factors in genome reshuffling, evidence is still scarce in mammals, especially in rodents. In this work, we analyzed the genomic distribution of genome reshuffling across rodents evolution, taking advantage of whole-genome sequences available in the databases. We identified Rodentia specific EBRs together with the genomic features that could be involved in their origin, including gene content, recombination rates and chromatin structure.

Comparative genomics data of this work was obtained during a research stage granted by travel fellowship by "Ministerio de Economía y Competitividad" (ref. EEBB-I-13-07350) in the laboratory of Dr. Denis M. Larkin in Aberystwyth University (Wales, UK). This manuscript is currently under preparation.

4.1.1 Comparative analysis of rodent genomes reveals evolutionary signatures of genome reshuffling

Capilla L[1,2], Sánchez-Guillén RA[1], Farré M[3], Alföldi J[4], Lindblad-Toh K[4], Malinverni R[5], Ventura J[2], Larkin DM[3], Ruiz-Herrera A[1,6]*

[1]Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. [2]Departament de Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. [3]Department of Comparative Biomedical Sciences, The Royal Veterinary College, London, UK. [4]Vertebrate Genome Biology. Broad Institute of MIT and Harvard, USA. [5]Institute of Predictive and Personalized Medicine of Cancer (IMPPC), Barcelona, Spain. [6]Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

*Corresponding author: aurora.ruizherrera@uab.cat

4.1.1.1 Introduction

Unlocking the genetic basis of speciation is of crucial importance to explain species diversity and adaptation to a changing environment. Similarly, understanding the role that large-scale chromosomal changes play in reproductive isolation has been a focus of evolutionary biologists (White 1978; Ayala and Coluzzi 2005). Particularly, whether these act as barriers to gene flow (Rieseberg 2001; Navarro and Barton 2003; Faria and Navarro 2010) or by modifying both the structure and regulation of genes located at, or near, the affected regions (Murphy et al. 2001; Larkin et al. 2009; Ullastres et al. 2014). The main motivation behind these studies has been to find evidence of the adaptive value of genome reshuffling and of the mechanisms of its formation during mammalian diversification. Compelling evidence has shed light on genomic features that characterize evolutionary breakpoint regions (EBRs) (i.e., regions of disruption of genome homologies) and their genomic distribution. Repetitive elements including segmental duplications (SDs) (Bailey and Eichler 2006; Kehrer-Sawatzky and Cooper 2007; Zhao and Bourque 2009), tandem repeats (TRs) (Kehrer-Sawatzki et al. 2005; Ruiz-Herrera et al. 2006; Farré et al. 2011) and transposable elements (TEs) (Longo et al. 2009; Carbone et al. 2009; Farré et al. 2011) have all been associated with their presence. However, given the diversity of repetitive elements found within EBRs it is likely that sequence composition is not alone in influencing genome instability. In fact, it was initially reported that EBRs are located in gene-rich regions (Murphy et al. 2005; Lemaitre et al. 2009), those

containing gene functional process networks, such as genes related to the immune system (Groenen et al. 2012; Ullastres et al. 2014). This suggests that changes in gene expression caused by genome reshuffling could reflect a selective advantage through the development of new adaptive characters (Larkin et al. 2009; Groenen et al. 2012; Ullastres et al. 2014). But how universal this pattern is among other mammals, needs further validation.

Rodentia is the most diverse and species rich mammalian order with more than 2,000 defined species (Carleton and Musser 2005) that occupy a wide range of habitats and adaptive features. Although the rodent phylogeny has been hotly contested due to its complexity, it is widely accepted that Rodentia can be classified into three major clades: the mouse-related clade (Anomaluromorpha, Castorimorpha and Myomorpha), the squirrel-related clade (Sciuromorpha) and Ctenohystrica (Hystricomorpha) (Huchon et al. 2002; Montgelard et al. 2008; Blanga-Kanfi et al. 2009; Churakov et al. 2010) Rodentia are generally considered to present specific features such as higher rates of nucleotide substitution (Wu and Li 1985), lower recombination rates and higher genome reshuffling rates [although this is mainly based on Mus (Stanyon et al. 1999; Veyrunes et al. 2006)] than when compared to other Laurasiatheria (Dumont and Payseur 2011; Segura et al. 2013). In fact, one of the most intriguing features that characterize rodents is the high chromosomal variability. This is exemplified by a wide range of diploid numbers ranging from 2n=10 in *Akodon spp.* (Myodonta clade) to 2n=102 in *Tympanoctomys barerae* (Ctenohystrica clade) (Silva and Yonenaga-Yassuda, 1998; Gallardo et al. 2004). Previous comparative studies have provided relevant information on both ancestral karyotype reconstructions for the group (Bourque et al. 2004; Froenicke et al. 2006; Ma et al. 2006; Graphodatsky et al. 2008; Mlynarski et al. 2010; Romanenko et al. 2012) and specific large-scale rearrangements (Pevzner and Tesler 2003; Zhao et al. 2004; Froenicke et al. 2006; Mlynarski et al. 2010). In particular, it has been proposed that mouse-specific rearrangements are mostly inter-chromosomal (such as fissions or fusions) while rat specific rearrangements are intra-chromosomal (such as inversions) (Zhao et al. 2004). However, the reason(s) behind the extremely high rate of genome reshuffling has, and continues, to puzzle evolutionary biologists. Therefore, a more comprehensive picture of rodent genome evolution at a finer scale remains to be uncovered.

With the availability of fully sequenced genomes from several different rodent species, we can now delineate the evolutionary history of genomic reshuffling in rodents in order to better understand both the adaptive value of chromosomal rearrangements within the group and the mechanisms underlying this pattern.

Here we present a refined analysis of the Rodentia EBRs as an estimation of genome reshuffling in the mouse by comparing six rodent genomes and six mammalian outgroup species. This has permitted the examination of EBRs across Rodentia phylogeny, that were analyzed for gene content, recombination rates and cLADs. Our results provide evidence for the presence of rodent-specific genomic signatures, reinforcing the adaptive role of genome reshuffling and highlighting the influence of chromatin spatial organization in the formation of large-scale evolutionary chromosomal changes.

4.1.1.2 Materials and methods

4.1.1.2.1 Whole-genome comparisons

Pair-wise alignments were established between the genomes of the mouse (NCBIm37 assembly) and 11 representative species of mammalian phylogeny by *Satsuma Synteny* (Grabherr et al. 2010) (Table S1). Based on the sequence homologies provided by *Satsuma Synteny*, the *Synteny Tracker* algorithm (Donthu et al. 2009) was used to establish regions of homology (syntenic regions) between the mouse genome (reference genome) and each of the mammalian species included in the analysis based on a minimum block size threshold. We differentiated two types of syntenic regions: (i) Homologous Synteny Blocks (HSBs) when pair-wise comparisons were established between genomes assembled into chromosomes, and (ii) Synteny Fragments (SFs), for pair-wise comparisons between genomes only available into scaffolds (Table S2). For each pair-wise alignment, three different syntenic block sizes (including both HSBs and SFs) were compared (100Kbp, 300Kbp and 500Kbp) (Table S3; Figure S1). This allowed us to evaluate genome assembly reliability. When the number of HSBs or SFs was not proportional between the three resolutions, it was assumed that the genome contained structural assembly errors.

Once syntenic regions were established for all species, EBRs were defined and classified by the *Evolutionary Breakpoint Analyzer* (EBA) algorithm (Farré et al 2015, submitted) using 300Kbp as the reference block size resolution. EBA provides all EBRs detected in each pair-wise comparison and gives a reliability score for each classification. The main values are determined by the ratio of the scores and the percentage of species with breakpoints with respect to genomic gaps. By taking the total number of species used in our analysis into account and the percentage of species

that presented the genome in scaffolds, the threshold was fixed at a ratio ≥34, and a percentage >60%. Then, two different groups of EBRs were established: (i) EBRs corresponding to any of the 11 species studied (hereafter, lineage-specific EBRs) and (ii) EBRs that appeared in any of the differentiation nodes of the phylogenetic tree (hereafter, clade-specific EBRs, Figure 1; Table S4). In fact, and based on the phylogenetic relationships among the species included in our analysis, ten different nodes/clades were considered (Figure 1): Clade 1 - Boreoeutheria, which included all mammalian species compared in our analysis; Clade 2 - Euarchontoglires, including all rodent and primate species; Clade 3 - Catarrhini, which included *Homo sapiens, Macaca mulatta,* and *Pongo pygmaeus*; Clade 4 - Hominoidea, with only *H. sapiens* and *P. pygmaeus*; Clade 5 - Rodentia, which included all rodent species compared in our study; Clade 6 - Myodonta, all rodents species compared, except *Heterocephalus glaber*; Clade 7 - Muroidea, with *Spalax galilii*, *Microtus ochrogaster, Rattus norvegicus* and *Mus musculus*; Clade 8 – Cricetidae+Muridae, including *M. ochrogaster, R. norvegicus* and *M. musculus*; Clade 9 - Muridae, with *R. norvegicus* and *M. musculus*; and Clade 10 - Laurasiatheria, with Bos taurus, Equus caballus and Felis catus. In order to estimate the average rate of EBRs occurring for each phylogenetic branch (number of EBRs per million years - Myr), divergence times (autocorrelated rates and hard-bounded constraints) were extracted from (Meredith et al. 2011) for each lineage and clade phylogenetic branches, with the exception of Muridae. In this latter instance, data provided by (dos Reis et al. 2012) was used (Table S5).

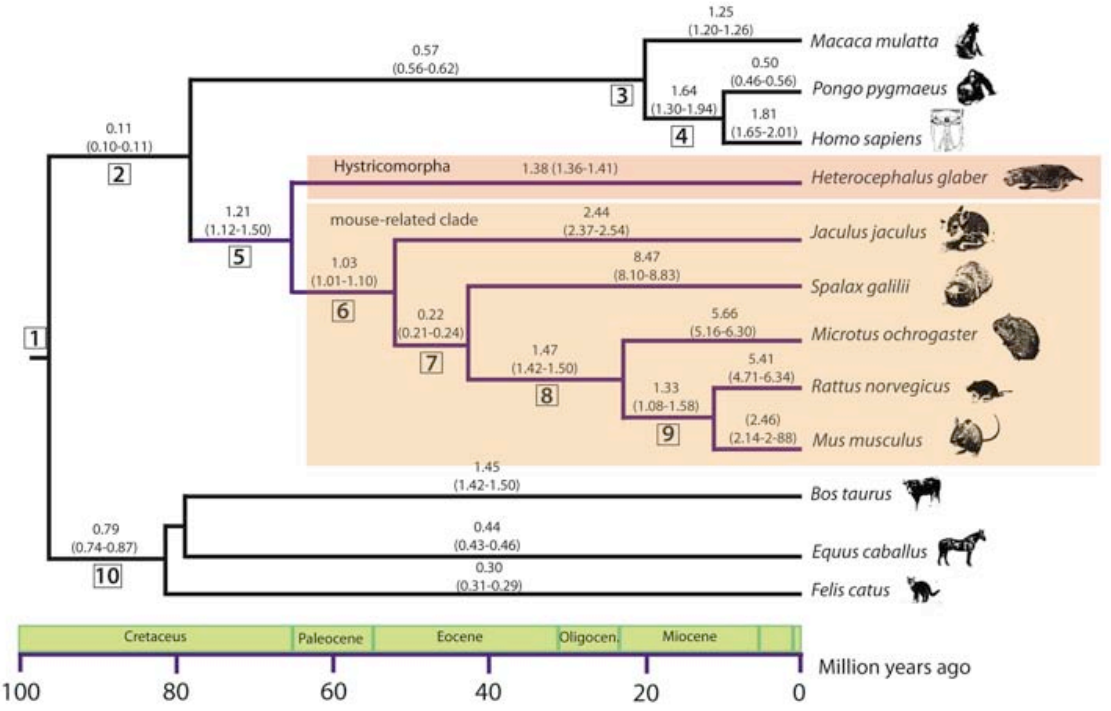


Figure 1. EBRs mapped in the time tree of the mammalian species included in the study: Time tree was based on divergence times (autocorrelated rates and hard-bounded constraints) described by Meredith et al. (2011), with the exception of two species (*M. musculus* and *R. norvegicus*) and one clade (Muridae) which were estimated by using dos Reis et al. (2012) time tree. In the upper section of each branch, the mean rate of EBRs per Myr and the range (in brackets) is shown. Numbers framed in squares represent mammalian phylogenetic nodes: 1-Boreoeutheria; 2-Euarchontoglires; 3-Catarrhini; 4-Hominoidea; 5-Rodentia; 6-Myodonta; 7-Muroidea; 8- Cricetidae+Muridae; 9-Muridae; 10-Laurasiatheria.

4.1.1.2.2 Gene content and ontology

Sequence coordinates of all mouse genes were obtained from BioMart (RefSeq genes, NCBIm37). Genes were clustered into two groups: (i) total genes, which included protein-coding genes, novel genes with unknown function, pseudogenes and RNA genes and (ii) protein coding genes, which included only genes with known function. Genes were assigned either to HSBs or EBRs when coordinates fell within these regions. EBRs used in this and subsequent analysis were specifically all rodents clade-specific and mouse lineage-specific. Gene density was analyzed by calculating the mean number of genes contained in non-overlapping windows of 10Kbp across the mouse genome as previously described (Ullastres et al. 2014). Four different genomic regions were taken into account: (i) HSBs, (ii) EBRs, (iii) interphase regions (regions of overlapping within the start or the end coordinates of any given EBRs) and (iv) 100Kbp regions upstream or downstream from the EBRs coordinates. Given the high incidence of assembly errors at the telomeres/subtelomeres and the centromeric/pericentromeric areas, we excluded 3Mbp section of each region from the analysis.

The functional annotation and clustering tool DAVID (Database for Annotation, Visualization, and Integrated Discovery, v6.7) (Huang et al. 2009) was used to identify overrepresented biological terms contained in EBRs. Functional annotation clustering allows for the biological interpretation and functional annotation charts identifying the most relevant (overrepresented) biological terms associated with a given gene list (Huang et al. 2009). We used the Benjamini's test to control false positives. This compares the proportion of genes in the analyzed regions (i.e., EBRs) to the proportion of the genes of the rest of the genome (i.e., HSBs), and produces an EASE score. EASE scores ≤ 0.05 and containing a minimum of two Gene Ontology (GO) terms were considered significantly overrepresented.

4.1.1.2.3 Recombination rates

The mouse genetic map was extracted from (Brunschwig et al. 2012). This contains high-resolution recombination rate estimates across the mouse genome (the autosomic chromosomes) based on 12 classically sequenced mouse strains (129S5/SvEvBrd, AKR/J, A/J, BALB/cJ, C3H/HeJ, C57BL/6NJ, CBA/J, DBA/2J, LP/J, NOD/ShiLtJ, NZO/HILtJ, and WSB/EiJ). From this map, we estimated recombination rates for non-overlapping windows of 10Kbp across the mouse genome as previously described (Farré et al. 2013). For each 10Kbp window, the recombination rate was calculated as the average of all recombination rates. These values were subsequently merged with the genomic positions from the four different genomic regions included in the gene density analysis using in-house Perl scripts. Centromeric and telomeric regions were not included in the analysis.

4.1.1.2.4 Constitutive lamina associated domains (cLADs)

Genomic data for mouse Lamina Associated Domains (LADs) was extracted from (Meuleman et al. 2013) available at the NCBI Gene Expression Omnibus (GEO) (accession number GSE36132). LADs were obtained using DamID maps (Peric-Hupkes and van Steensel 2010) of lamina A in mouse astrocytes (ACs) and neural precursor cells (NPCs) and Lamina B1 in wild type and Oct1 knockout mouse embryonic fibroblasts (MEFs and Oct1koMEFs respectively). Constitutive LADs (cLADs) resulted from selecting lamina regions that were identified in all cell types of cells analyzed. Once cLADs positions were obtained, their genomic distribution was analyzed in non-overlapping windows of 10Kbp as described above. Each 10Kbp

window was subsequently classified into different genomic regions as was done in the gene content and recombination analyses (EBRs, HSBs, interphases and 100Kbp adjacent regions) described above.

4.1.1.2.5 Statistical analysis

The genome-wide distribution of EBRs was estimated using an average frequency across the mouse genome and by assuming a homogeneous distribution of all detected EBRs. We used a $\chi 2$ test with a Bonferroni correction to assess any possible deviation from the homogeneous distribution.
Mean comparison of gene density, recombination rates and cLADs with the genome wide division of 10Kbp windows was performed with Kruskal-Wallis non-parametric test using JMP statistical package (release 7.1). Genome-wide association analysis between EBRs, gene content and cLADs were obtained using RegioneR— an R package based on permutation tests (Phipson and Smyth 2010; Diez-Villanueba et al. 2015) (http://gattaca.imppc.org/regioner/). RegioneR compares the number of observed overlaps between a query and a reference region-set to the distribution of the number of overlaps obtained by randomizing the regions-set over the genome for each chromosome. We performed 10,000 permutations with randomization for each analysis. The threshold was fixed in the 5%, therefore p-values<0.05 indicate significant association. Gene content, recombination rate and cLADs were analyzed using Spearman correlation test using the JMP statistical package (release 7.1).

4.1.1.3 Results

4.1.1.3.1 Genome reshuffling in Rodentia

Defining syntenic regions and EBRs in Rodentia. In order to determine the evolutionary genomic landscape in Rodentia, we compared the mouse genome to those of five rodent species: one representative of the Hystricomorpha clade (*H. glaber*) and four species belonging to the Myodonta clade (*J. jaculus, S. galilii, M. ochrogaster and R. norvegicus*). In addition, the inclusion of six mammalian species from Primates (*H. sapiens, M. mulatta,* and *P. pygmaeus*), Artiodactyla (*B. taurus*), Carnivora (*F. catus*) and Perissodactyla (*E. caballus*) allowed us to refine the characterization of EBRs in a phylogenetic context.

We first determined the syntenic regions (HSBs and SFs) in all species (Table S2), identifying a total of 3,392 HSBs with a mean size ranging from 13.22 Mbp in R. norvegicus, to 5.56 Mbp in B. taurus (Table S2). We detected a total of 3,142 SFs, with a mean size ranging from 1.14Mbp in *S. galilii*, to 5.14Mbp in *H. glaber* (Table S2). The number of estimated HSBs differed depending on species and ranged from 280 HSBs (representing the 95.60% of the mouse genome) between mouse and rat, to 521 HSBs (representing 91.11% of the mouse genome) between mouse and cow (Table S2). In the case of scaffold-based genomic comparisons, the number of SFs was slightly higher in *J. jaculus* (559) and *H. glaber* (598) and especially pronounced in *S. galilii* (1,985). The syntenic regions detected represented >80% of the mouse genome, reaching 95.6% in the mouse/rat comparison, and 93.5% for the mouse/horse comparison (Table S2). This is a reflection of the high conservation of their genomes.

Once the syntenic regions were determined for all species, we estimated the number and genomic distribution of EBRs in the mouse genome and classifed them based on their appearance during Rodentia evolution. We detected a total of 1,333 EBRs, the majority of which (1,179) were classified as unique EBRs (i.e., EBRs that appeared only once during Rodentia evolution, in a specie lineage or clade) (Figure 1 and Table S3). The rest, representing 154 EBRs, were classified as reused (i.e., EBRs that are shared by a subset of species from the same clade). Of the unique EBRs detected, 1,024 were lineage-specific (i.e., specific for each of the species when compared to the mouse genome), and the remaining 130 EBRs were classified as clade-specific (Primate, Hominoidea, Laurasiatheria, Euarchontoglires, Rodentia, Myodonta, Muroidea, Cricetidae+Muridae and Muridae) (Table S3). The number of lineage-specific EBRs was variable and ranged from 8 EBRs in *P. pygmaeus* to 360 EBRs in *S. galilii*. In the case of the clade-specific EBRs, the number of evolutionary regions ranged from 2 EBRs in Euarchontoglires to 33 EBRs in Catharrini (Table S3). Likewise, EBRs mean size depended on the pair-wise species comparison (Table S3). In order to corroborate the EBR estimations, we analyzed the number of syntenic blocks obtained at 100Kbp, 300Kbp and 500Kbp resolutions for all pair-wise comparisons. With the exception of *R. norvegicus*, the number of syntenic blocks was proportional between the three levels of resolution (Figure S1 and Table S4) supporting the reliability of genome assemblies and EBR estimations.

To provide an estimation of the genome reshuffling rate (expressed as the number of EBRs detected in each phylogenetic branch per Myr) that occurred in Rodentia, we placed the total estimated EBRs in a phylogenetic context considering the species included in the study (Figure 1). We detected that the presence of EBRs in

Rodentia was higher (1.21 EBRs/Myr) than in the rest of major mammalian clades (i.e., 0.79 EBRs/Myr for Laurasiatheria or 0.11 EBRs/Myr for Euarchontoglires) (Figure 1). This observation corroborates the long-standing view that points rodents as one of the mammalian orders with the highest genome reshuffling rates. There is, however, variability among Rodentia clades—the highest rate of the genome reshuffling was detected in the mouse-like group (Muridae, 1.47 EBRs/Myr) while a lower rate was detected in Muroidea (0.22 EBRs/Myr). In terms of the species-specific genome reshuffling rates, rodents in general showed higher rates that any other mammalian species included in the study (Figure 1). That was the case, for example, of *J. jaculus* (2.44 EBRs/Myr), *M. ochrogaster* (5.66 EBRs/Myr), *R. norvegicus* (6.41 EBRs/Myr) and *S. galilii* (8.47 EBRs/Myr). However, we need to be conservative in defining genome reshuffling rates in *R. norvegicus* since the number of HSBs detected was not proportional in the three different resolutions of ST (100Kbp, 300Kbp and 500Kbp, Figure S1).

Genome-wide distribution of Rodentia EBRs. Given that our main goal was to define genome reshuffling in Rodentia, and more specifically, to determine the presence of genomic signatures that occurred during mouse evolution, we focused our efforts on analyzing the distribution of both Rodentia clade-specific EBRs and mouse-specific EBRs across the mouse genome. Of the 655 EBRs detected in the rodent species analyzed, 105 (covering 0.31% of the mouse genome) appeared in the lineage leading to the Mus. These included 75 clade-specific EBRs: 15 EBRs defined Rodentia, 14 Myodonta, 3 Muroidea, 28 Cricetidae+Muridae, 15 Muridae and 30 EBRs were specific to *Mus musculus* (Figure 1 and Table S4). Assuming a homogeneous distribution across the genome, we observed that EBRs were not randomly distributed throughout the mouse genome (Figure 2 and Figure S2). In fact, three chromosomes (MMU8, MMU17 and MMU18) appeared to contain significantly more EBRs than expected under a random distribution (MMU17: $\chi2 = 13.57$, p-value < 0.001 and MMU18:$\chi2= 14.96$, p-value < 0.001; Figure S2). Additionally, three other chromosomes (MMU4, MMU16 and MMUX) contained less EBRs than expected under a random distribution (MMU4: $\chi2 = 4.54$, p-value < 0.05; MMU16: $\chi2= 3.93$, p-value <0.05; and MMUX: $\chi2 = 4.81$, p-value <0.05; Figure S2). Moreover, EBRs appeared to be localized in clusters (i.e., genomic regions with a higher density of EBRs per Mbp), for example in MMU8 and MMU17 (Figure 2).

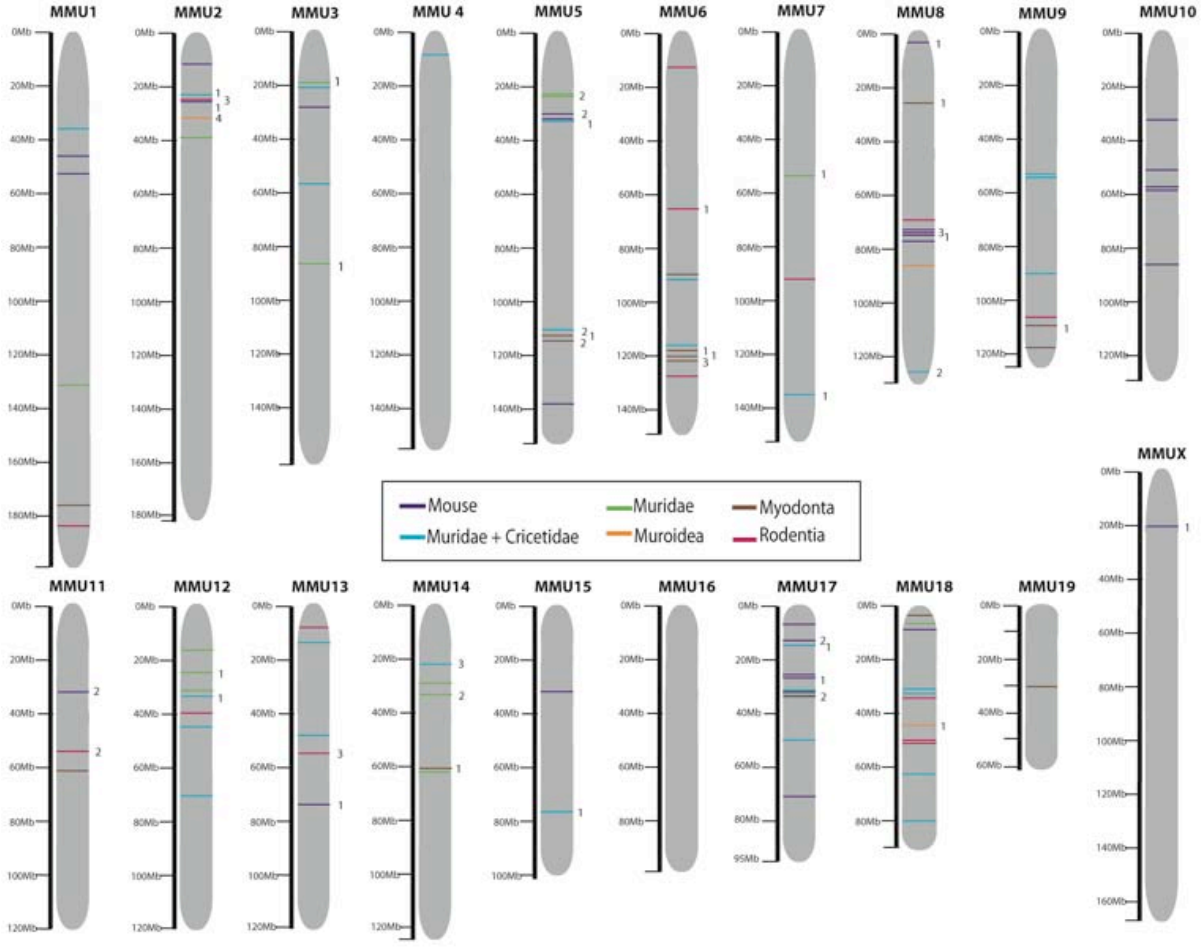


Figure 2. EBRs mapped in the mouse genome: The positions of each EBRs (lineage and clade-specific) are color-coded (see inset legend). Number of protein-coding genes detected within each EBRs are depicted on the right of each mouse EBR.

4.1.1.3.2 Rodentia EBRs are gene-rich regions

We further examined the genomic characteristics of EBRs that lead to the mouse differentiation, searching for the presence of specific evolutionary signatures. To this end, we first analyzed the genome-wide distribution of genes, paying special attention to GO. A total of 36,381 genes were identified and included in the analysis. These were divided into two groups: (i) all genes (n=36,381) and (ii) protein coding genes (n=22,352). The mean distribution of genes (including protein-coding genes, non-coding RNA genes and pseudogenes) found in the mouse genome was 0.09 genes per 10Kbp, although these were non-homogeneously distributed across chromosomes (Kruskal-Wallis test, p-value<0.001). Mouse chromosomes 7, and 11 are gene-rich (0.14 genes per 10Kbp in both cases) whereas chromosomes 12, 18 and X (0.06 genes per 10Kbp in all cases) are low on genes. We then analyzed gene density for all Rodentia EBRs detected (including clade-specific and those that are

mouse lineage-specific). Our results showed that EBRs are gene-rich regions with an average density of 0.18 genes per 10Kbp compared to the rest of the genome (0.09 genes per 10Kbp, Kruskal-Wallis test, $p<0.001$). This enrichment was confirmed using a genome-wide permutation test (based on 10,000 permutations with randomization, $p<0.05$) (Table 1; Figure 3B). When analyzing the gene density at the vicinity of EBRs (Figure 3A), we observed that these flanking regions have a high concentration of genes when compared to the rest of the genome (HSBs) (Kruskal-Wallis test, p-value<0.001, Figure 3A), especially so in regions that are up-stream EBRs. Additionally, we studied the presence of protein coding genes (n=22,352) overlapping either the start or the end coordinates of the analyzed EBRs (both clade- and mouse-specific). This allowed us to detect whether gene sequences were disrupted by the presence of the estimated EBRs coordinates. In total, we detected 63 protein-coding genes that were disrupted (35 genes at the start and 28 at the end of EBRs) representing all types of clade-specific and in mouse-specific EBRs (Table S6). Of these, 55 genes were disrupted in intronic regions (87.5%). In only 8 instances EBR coordinates found to be positioned inside an exon (Table S6).

	Protein-coding genes	
EBR type	**p-value**	**z-score**
Mouse specific	0.029*	2.53
Muridae specific	0.009**	1.43
Cricetidae+Muridae specific	0.049*	2.95
Muroidea specific	0.004**	3.81
Myodonta specific	0.009**	2.93
Rodentia specific	0.003**	3.21
All EBRs	0.001**	6.25

Table 1. Gene clusters found enriched within EBRs: For each EBR we have specified the mouse chromosome (chr), the start and end position (in bp), the corresponding gene enrichment cluster or gene family name, the ID and GO terms and the distance of the gene start from the up-stream region of the EBR (in Kbp).

Since chromosomal rearrangements can potentially affect the structure and regulation of genes in or nearby the affected regions, we focused on the adaptive role of EBRs by analyzing gene ontology of the 107 protein coding genes detected within Rodentia EBRs in the mouse genome. We found two gene families localized within specific EBRs. Moreover, there was one enrichment cluster in EBRs that presented the highest statistical support when compared to the rest of the genome (n=3; EASE≤0.05)(Table S7). The first gene family included the Calycin superfamily and more specifically the Lipocalins (Lcn) that were localized within two nearby EBRs (one Rodentia-specific and one mouse-specific EBR) in mouse chromosome

2. In particular, we detected Lcn genes that were involved in the transportation of lipophilic molecules (Lcn4, with the expression limited to the vomeronasal organ), male fertility (Lcn13), sperm maturation and retinoid carrier proteins within the epididymis (Lcn5) and odorant binding proteins (Lcn14). The second gene family found was localized in mouse chromosome 11 and included five genes belonging to the haemoglobin (Hb) family (involved in binding and/or transporting oxygen). All four genes were Hb subunits and localized in a mouse-specific EBR which included Hb X, Hb alfa (Hb-alfa , chains 1 and 2), and haemoglobin theta A and B (Hb-Theta, 1B and 1A).

Lastly, and most intriguing, the only statistically significant enrichment cluster found in our analysis (Benjamini test, p-value=0.02; Table S7) included five genes clustered as a Krueppel-associated box (KRAB) that were localized in three EBRs (classified as mouse- and Muridae-specific) and distributed in three different mouse chromosomes. KRAB proteins are transcription factors with ZnF binding domains (Knight and Shimeld 2001) that are mainly expressed during meiotic process (Baudat et al. 2010; Parvanov et al. 2010) and include, among others, *Prdm9*, the only known speciation-associated gene described for mammals, initially described in mice (Mihola et al. 2009; Capilla et al. 2014). Moreover, our analysis revealed genes from the Lcn family in the oldest Rodentia EBRs (Rodentia-specific), whereas, both the haemoglobin family and the transcription regulation gene enrichment cluster were localized in the EBRs leading to the mouse lineage (transcription regulation gene cluster; n=8 genes, enrichment score=2.39; Benjamini test, p-value=0.18).

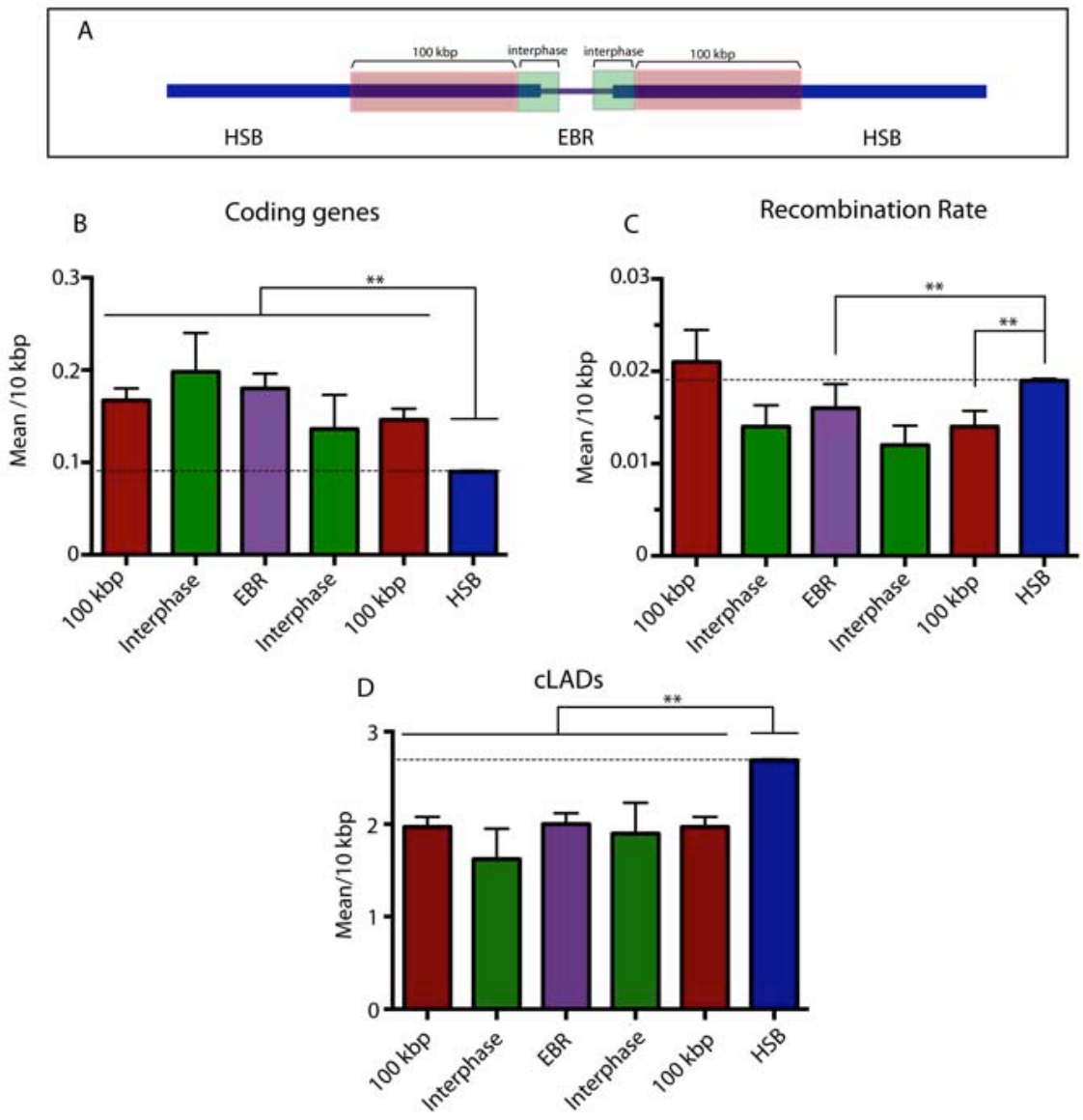


Figure 3. Gene content and recombination rates around EBRs: (A) Schematic representation of the genomic regions considered for the analysis (see material and methods for details). (B) Distribution of protein coding genes. The X-axis represents the genomic regions analyzed, whereas the Y-axis display the mean number of genes detected per 10Kbp. (C) Distribution of recombination rates. The X-axis represents the genomic regions analyzed, whereas the Y-axis display the mean recombination rate detected per 10Kbp. (D) Distribution of constitutive Lamina Associated Domains (cLADs). The X-axis represents de genomic regions analyzed, whereas the y-axis display the mean number of cLADs identified per each 10Kbp windows. Standard error bars are represented. Punctuated lines represent genome-wide means. Asterisk indicates statistical significance (Kruskal-Wallis test, **p-value<0.001).

4.1.1.3.3 Rodentia EBRs correspond to regions of low recombination rates

It is known that genome reshuffling affects recombination (Rieseberg 2001; Navarro and Barton 2003; Noor and Bennett 2009), but data on the interplay between EBRs and recombination is restricted to few studies (Navarro et al. 1997; Farré et al. 2013). To address this, we analyzed the genome-wide distribution of recombination rates in the mouse genome and tested whether there was a correlation with EBRs. We found that recombination rates were not homogeneously distributed across the mouse genome. Chromosomes 17 and 19 had the highest recombination rates (0.015 $4N_e r$/Kbp in both cases) while the chromosome 8 showed the lowest rate (0.003 $4N_e r$/Kbp). The mean genome-wide recombination rate was 0.019 $4N_e r$/Kbp. These observations corroborate previous data in mammals that showed smaller chromosomes to have higher recombination rates than large chromosomes, thereby ensuring their correct segregation during meiosis (Sun et al. 2005; Farré et al. 2013). Moreover, our analysis indicated that Rodentia EBRs presented a significantly lower mean recombination rate (0.016 $4N_e r$/Kbp) compared to the rest of the genome (0.019 $4N_e r$/Kbp, Kruskal-Wallis test, $p<0.001$). To further explore these observations we estimated the mean recombination rates for clade-specific and mouse-specific EBRs and found a significantly lower recombination rate in the mouse-specific and Muridae-specific EBRs (0.013 and 0.006 $4N_e r$/Kbp respectively, Kruskal-Wallis test, $p<0.001$). Moreover we also analyzed mean recombination rates around the EBRs (Figure 3C). The analysis suggested a tendency for the presence of low recombination rates in EBRs flanking regions (0.014 and 0.012 $4N_e r$/Kbp) and then an increment in the following 100Kbp surrounding EBRs (0.021 and 0.019 $4N_e r$/Kbp) that tend to reach the values observed for HSBs (Figure 3C).

4.1.1.3.4 Rodentia EBRs are depleted in LADs

We further investigated whether the distribution of EBRs in the mouse genome was influenced by the spatial organization of the chromatin. We analyzed the distribution of cLADs and found that the total 715,804 cLADs described in the mouse, were not homogenously distributed across the genome, but were inversely correlated with gene distribution (Figure S3A) thus mirroring similar studies on human cells (Guelen et al. 2008). The X chromosome had the highest cLADs density (3.75 cLADs/10Kbp), whereas chromosomes 11 and 19 had the lowest (1.80

and 1.72 cLADs/10Kbp, respectively) (Kurskal-Wallis test, $p<0.001$). Gene density was inversely correlated to cLADs density per chromosome, the only exceptions being chromosomes 4, 15 and 16 (Figure S3B). When looking at the genome-wide distribution of cLADs in each chromosome, the same pattern was observed (Figure S3B). We subsequently analyzed the relationship between EBRs (both Rodentia and mouse lineage specific EBRs) and cLADs. Our results indicated a significant decrease in cLADs density in all EBRs (2 cLADs/10Kbp) as well as in interphase regions (1.62 and 1.90 cLADs/10Kbp) when compared to the rest of the genome (2.68 cLADs/10Kbp; Kruskal-Wallis test, $p<0.001$; Figure 3D). This pattern was corroborated by permutation test (based on 10,000 permutations with randomization, z-score= -2.46; $p<0.05$).

Finally, the relationships between the three genomic characteristics studied in this work (gene content, recombination rate and cLADs) were examined using pair-wise correlations between all three variables. This indicated a significant negative correlation between the number of cLADs and the number of coding genes (Spearman correlation test, $p=-0.093$; p-value<0.001) and less but also significant between cLADs and the recombination rates (Spearman correlation test, $p=-0.015$; p-value<0.001).

4.1.1.4 Discussion

The genome comparative analysis of six rodent species representative of two of the three major clades that include all rodents (Hystricomorpha and Myodonta clades) together with eleven mammalian representative species has allowed us to reconstruct the most detailed comprehensive picture of the evolutionary rodent genome reshuffling. We have been able to identify lineage and clade-specific EBRs among the Rodentia species analyzed and to compare their rate of chromosome breakage (number of EBRs/Myr) as an estimate of genome reshuffling, with respect to other mammalian outgroups such as Primates, Perissodactyla, Artyodactila and Carnivora. Our results are in agreement with previous studies that reflected a high genome reshuffling rate within Rodentia differentiation (either in the clades and species differentiation) (Murphy et al. 2005; Larkin et al. 2009). In fact, when considering the main mammalian diversification nodes, Rodentia presented approximately two orders of magnitude increase in EBRs per Myr, than either Euarchontoglires or Laurasiathera. But, more intriguingly, this rate increased when analyzing lineage-specific EBRs. Previous cytogenetic studies indicated

that the myomorph rodents showed more highly reorganized patterns (reviewed in Romanenko et al. 2012), whereas the comparative genome analysis performed here, showed the Muroidea species (*S. galilii, M. ochrogaster, R. norvegicus* and *M. musculus*) with the highest rates of genome reshuffling (a 2- to 5-fold increase when compared to other eutherian mammals).

In searching for signatures that characterize evolutionary genome reshuffling we detected a significantly higher gene density in EBRs when compared to the rest of the mouse genome. Although others before us have detected this trend in other mammalian species (Murphy et al. 2005; Lemaitre et al. 2009; Larkin et al. 2009; Groenen et al. 2012), the reasons behind this pattern have remained unclear. Our results offer a material advance in that they suggest that both the state of the chromatin and the adaptive role of EBRs are most probably affecting its genomic distribution in the mouse genome and it seems likely that this will hold for other mammalian orders.

First, we detected that rodent EBRs were depleted in cLADs and that these structural genomic regions negatively correlated with gene content. Nuclear lamina (NL) anchor chromosomal domains in mammalian chromatin by interacting with cLADs. Previously is was thought that cLADs interact with the NL independently of cell type and are conserved in human and mouse (Meuleman et al. 2013). The pattern that we observed is most probably related with the fact that the chromatin status in cLADs is mostly transcriptionally inactive and silenced (Reddy et al. 2008; Peric-Hupkes et al. 2010; Kind and van Steensel 2010; Kohwi et al. 2013). Therefore, genomic regions outside cLADs are expected to be more exposed to the transcription machinery. As a consequence of this spatial chromatin organization and according to the new integrative breakage model for genome evolution (Farré et al. 2015) gene-rich regions would be more susceptible to the occurrence of large-scale chromosomal reorganizations, due to their accessibility. Our observation of a depletion of cLADs in rodent EBRs, in conjunction with a high-density of protein-coding genes, supports this view. That is, "open" chromatin configurations in regions with high transcriptional activity are gene-rich and may drive genome reshuffling.

Additionally, and despite the possibility that genome reshuffling would disrupt genes essential for survival, and therefore be subject to purifying selection, EBRs can represent opportunities for the development of novel functions that may promote the adaptation of species. This is consistent with the idea that there is a connection between mammalian EBRs and the development of new adaptive gene functions, such as in the immune system or olfactory receptors (Larkin et al. 2009; Groenen et al. 2012; Ullastres et al. 2014). In this context, rodents are a particularly

useful model since they are the largest mammalian order, whose species show an enormous array of evolutionary adaptations. We detected the presence of two gene families in our rodent data (Lcn and Hb) and one functional enrichment cluster (KRAB genes) within clade- and lineage-specific EBRs in the Rodentia phylogeny that might support the adaptive hypothesis of genome reshuffling. Lipocalins are globular secreted proteins involved in many biological processes, especially in rodents (Stopková et al. 2009). The lipocalins found within rodent EBRs belong to two main functional groups: (i) odor-binding proteins (OBPs) involved in chemical communication (Snyder et al. 1989), and (ii) epididymal retinoic acid binding proteins (ERABP), which are specifically expressed in the epididymis and, therefore, relevant for assuring fertility through sperm maturation acquire (Suzuki et al. 2007). Given that chemical communication in rodents is extremely important for sexual reproduction driving mate choice between individuals (Hurst and Beynon 2004), the original function of lipocalins may have been favored by natural selection during the evolution of the chemical communication in mice (Stopková et al. 2009). In fact, although lipocalins are poorly conserved at the sequence level, their folding pattern and structure is very well conserved across species (Flower 1996; Beynon and Hurst 2004). In addition to this observation, the impairment of antioxidative mechanisms in rodents have been also described to be adaptive under uncertain conditions, such as altitude or extreme thermal conditions, among others (Storz et al. 2007; Storz et al. 2009). In this context, developing new variants of haemoglobin can provide selective advantage, exemplified by the high levels of haemoglobin polymorphisms that have been described in rodent species (Natarajan et al. 2013; Kotlík et al. 2014). But perhaps the most relevant result was the presence of an enrichment cluster in rodent EBRs that included KRAB-ZnF genes, a group of transcription factors with ZnF domains. Most of the KRAB-ZnF proteins, with the exception of *Prdm9*, are not functionally fully characterized, but are known to be organized in clusters (Huntley et al. 2006; Ding et al. 2009) and are thought to play a role in speciation given their role in reproductive isolation (Nowick et al. 2013; Turner et al. 2014). In fact, studies in mice have shown that the PRDM9 protein, a meiotic-specific histone methyltransferase, determines the position where recombination occurs (Brick et al. 2012) as well as determining recombination rates in mice natural populations (Capilla et al. 2014). KRAB-ZnF genes are, indeed, fast evolving (for a review see Nowick et al. 2013) and, in the case of *Prdm9*, a large diversity in the number and sequence of ZnF have been reported (Oliver et al. 2009; Steiner and Ryder 2013; Buard et al. 2014; Kono et al. 2014; Capilla et al. 2014). Strikingly, we found *Prdm9* together with poorly characterized KRAB genes, such as Zfp169, Zfp182 and Zfp300 in different

Rodentia EBRs. It may be possible that the rapid evolution characterizing this gene family might be related to the instability created by genome reshuffling within these regions and occurs by altering both sequence composition and expression patterns of the genes located within EBRs.

As a consequence, can evolutionary breakpoint regions be considered "genomic islands of speciation" (as referred by Turner et al. 2005)? While initial studies have found that EBRs tend to show higher divergence rates than other regions in the genome (Navarro et al. 1997; Marques-Bonet and Navarro 2005), others have detected lower recombination rates (Farré et al. 2013). Mirroring these results, we detected a significant reduction on recombination rates within EBRs when compared to the rest of the mouse genome. This reduction was only maintained in EBRs corresponding to the mouse lineage and the Muridae clade, in consonance with the short effect of chromosomal rearrangements on recombination rates along the species evolution (Coop and Myers 2007). But, one may ask whether the presence of speciation genes within EBRs (here exemplified by *Prdm9*) with low recombination rates, give rise to linkage disequilibrium that facilitate selection. Genes involved in reproductive isolation are expected to be found in regions of low recombination (Rieseberg 2001; Noor 2002; Navarro and Barton 2003). In fact, gene incompatibilities, reduced introgression and higher differentiation are associated with genomic regions with reduced recombination (Geraldes et al. 2011; Seehausen et al. 2014; Janoušek et al. 2015). Therefore, we propose that low recombination rates in EBRs could lead to a high genomic differentiation and the fixation of new mutations in genes related to the species-specific phenotypes (such as genes involved in mating and individual recognition, reproductive isolation and oxidative stress), thereby reinforcing the adaptive value of genome reshuffling.

In conclusion, our results support the new multidisciplinary Integrative Breakage Model for the study of genome evolution (Farré et al. 2015). That is, to fully understand the mechanism(s) shaping mammalian genomes and driving speciation, it is necessary to take not only the functional constrains that would accompany genome reshuffling, but also the analysis of the high-level structural organization of genomes into consideration.

4.1.1.5 Acknowledgments

LC is the beneficiary of a FPI predoctoral fellowship (BES-2011-047722). RASG has a postdoctoral grant from "Alianza 4 Universidades". This study was supported

by a research grant from the Ministerio de Economia y Competitividad (MINECO, CGL-2010-20170) to ARH. Genome sequencing and assembly of the *J. jaculus*, and *M. ochrogaster* genomes by the Broad Institute of MIT and Harvard were supported by grants from the National Human Genome Research Institute (NHGRI). T.J. Robinson is acknowledged for comments on the draft of this manuscript.

4.1.1.6 BIBLIOGRAPHY

Ayala FJ, Coluzzi M (2005) Chromosome speciation: humans, *Drosophila*, and mosquitoes. Proc Natl Acad Sci 102 Suppl :6535–42.

Bailey JA, Eichler EE (2006) Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet 7:552–64.

Baudat F, Buard J, Grey C, et al (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327:836–40.

Beynon RJ, Hurst JL (2004) Urinary proteins and the modulation of chemical scents in mice and rats. Peptides 25:1553–63.

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

Bourque G, Pevzner PA, Tesler G (2004) Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. Genome Res 14:507–16.

Brick K, Smagulova F, Khil P, et al (2012) Genetic recombination is directed away from functional genomic elements in mice. Nature 485:642–5.

Brunschwig H, Levi L, Ben-David E, et al (2012) Fine-scale maps of recombination rates and hotspots in the mouse genome. Genetics 191:757–64.

Buard J, Rivals E, Dunoyer de Segonzac D, et al (2014) Diversity of *Prdm9* zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite. PLoS One 9:e85021.

Capilla L, Medarde N, Alemany-Schmidt A, et al (2014) Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and *Prdm9* allelic background. Proc Biol Sci. 281:20140297.
Carbone L, Harris RA, Vessere GM, et al (2009) Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. PLoS Genet 5:e1000538.

Carleton MD, Musser GG (2005) Order Rodentia. In: Mammal Species of the World. The Johns Hopkins University Press, pp 745–752.

Churakov G, Sadasivuni MK, Rosenbloom KR, et al (2010) Rodent evolution: back to the root. Mol Biol Evol 27:1315–26.

Coop G, Myers SR (2007) Live hot, die young: transmission distortion in recombination hotspots. PLoS Genet 3:e35.

Diez-Villanueva, A Malinverni R, Gel B (2015) regioneR: Association analysis of genomic regions based on permutation tests. R package version 1.0.3.

Ding G, Lorenz P, Kreutzer M, et al (2009) SysZNF: the C2H2 zinc finger gene database. Nucleic Acids Res 37:D267–73.

Donthu R, Lewin H a, Larkin DM (2009) SyntenyTracker: a tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. BMC Res Notes 2:148.

Dos Reis M, Inoue J, Hasegawa M, et al (2012) Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc Biol Sci 279:3491–500.

Dumont BL, Payseur B a (2011) Genetic analysis of genome-scale recombination rate evolution in house mice. PLoS Genet 7:e1002116.

Faria R, Navarro A (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends Ecol Evol 25:660–9.

Farré M, Bosch M, López-Giráldez F, et al (2011) Assessing the role of tandem repeats in shaping the genomic architecture of great apes. PLoS One 6:e27239.

Farré M, Micheletti D, Ruiz-Herrera A (2013) Recombination rates and genomic shuffling in human and chimpanzee-a new twist in the chromosomal speciation theory. Mol Biol Evol 30:853–64.

Farré M, Robinson TJ, Ruiz-Herrera A (2015) An Integrative Breakage Model of genome architecture, reshuffling and evolution: The Integrative Breakage Model of genome evolution, a novel multidisciplinary hypothesis for the study of genome plasticity. Bioessays 37:479–88.

Finlan LE, Sproul D, Thomson I, et al (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. PLoS Genet 4:e1000039.

Flower DR (1996) The lipocalin protein family: structure and function. Biochem J 318:1–14.

Froenicke L, Caldés MG, Graphodatsky A, et al (2006) Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? Genome Res 16:306–10.

Gallardo M, Garrido O, Bahamonde R, et al (1994) Gametogenesis and nucleotypic effects in the tetraploid red vizcacha rat, *Tympanoctomys barrerae* (Rodentia, Octodontidae) Biological Res 37:767-75.

Garcia-Cruz R, Pacheco S, Brieño MA, et al (2011) A comparative study of the recombination pattern in three species of Platyrrhini monkeys (primates). Chromosoma 120:521–30.

Geraldes A, Basset P, Smith KL, Nachman MW (2011) Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. Mol Ecol 20:4722–36.

Grabherr MG, Russell P, Meyer M, et al (2010) Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics 26:1145–51.

Graphodatsky AS, Yang F, Dobigny G, et al (2008) Tracking genome organization in rodents by Zoo-FISH. Chromosome Res 16:261–74.

Groenen MAM, Archibald AL, Uenishi H, et al (2012) Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491:393–8.

Guelen L, Pagie L, Brasset E, et al (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453:948–51.

Huang DW, Sherman BT, Lempicki R a (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57.

Huchon D, Madsen O, Sibbald MJJB, et al (2002) Rodent Phylogeny and a Timescale for the Evolution of Glires: Evidence from an Extensive Taxon Sampling Using Three Nuclear Genes. Mol Biol Evol 19:1053–1065.

Huntley S, Baggott DM, Hamilton AT, et al (2006) A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. Genome Res 16:669–77.

Hurst JL, Beynon RJ (2004) Scent wars: the chemobiology of competitive signalling in mice. Bioessays 26:1288–98.

Janoušek V, Munclinger P, Wang L, et al (2015) Functional organization of the genome may shape the species boundary in the house mouse. Mol Biol Evol 32:1208–20.

Kehrer-Sawatzki H, Sandig C, Chuzhanova N, et al (2005) Breakpoint analysis of the pericentric inversion distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (*Pan troglodytes*). Hum Mutat 25:45–55.

Kehrer-Sawatzky HK, Cooper DN (2007) Understanding the Recent Evolution of the Human Genome : Insights from Human – Chimpanzee Genome Comparisons. Human Mutation 28(2):99-130
Kind J, van Steensel B (2010) Genome-nuclear lamina interactions and gene regulation. Curr Opin Cell Biol 22:320–5.

Knight RD, Shimeld SM (2001) Identification of conserved C2H2 zinc-finger gene families in the Bilateria. Genome Biol 2:RESEARCH0016.

Kohwi M, Lupton JR, Lai S-L, et al (2013) Developmentally regulated subnuclear genome reorganization restricts neural progenitor competence in *Drosophila*. Cell 152:97–108.

Kono H, Tamura M, Osada N, et al (2014) *Prdm9* polymorphism unveils mouse evolutionary tracks. DNA Res 21:315–26.

Kotlík P, Marková S, Vojtek L, et al (2014) Adaptive phylogeography: functional divergence between haemoglobins derived from different glacial refugia in the bank vole. Proc Biol Sci. 281:1786

Larkin DM, Pape G, Donthu R, et al (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res 19:770–7.

Lemaitre C, Zaghloul L, Sagot M-F, et al (2009) Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. BMC Genomics 10:335.

Longo MS, Carone DM, Green ED, et al (2009) Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. BMC Genomics 10:334.

Ma J, Zhang L, Suh BB, et al (2006) Reconstructing contiguous regions of an ancestral genome. Genome Res 16:1557–65.

Marques-Bonet T, Navarro A (2005) Chromosomal rearrangements are associated with higher rates of molecular evolution in mammals. Gene 353:147–54.

Meredith RW, Janečka JE, Gatesy J, et al (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science 334:521–4.

Meuleman W, Peric-hupkes D, Kind J, et al (2013) Constitutive nuclear lamina – genome interactions are highly conserved and associated with A/T-rich sequence. 23:270–280.

Mihola O, Trachtulec Z, Vlcek C, et al (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science 323:373–5.

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

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

Murphy WJ, Stanyon R, O'Brien SJ (2001) Evolution of mammalian genome organization inferred from comparative gene mapping. Genome Biol 2:REVIEWS0005.

Murphy WJ, Larkin DM, Everts-van der Wind A,

et al (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309:613–7.

Natarajan C, Inoguchi N, Weber RE, et al (2013) Epistasis among adaptive mutations in deer mouse hemoglobin. Science 340:1324–7.

Navarro A, Betrán E, Barbadilla A, Ruiz A (1997) Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics 146:695–709.

Navarro A, Barton NH (2003) Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. Science 300:321–4.

Noor M a, Grams KL, Bertucci L a, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. Proc Natl Acad Sci 98:12084–8.

Noor M a F, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103:439–44.

Nowick K, Carneiro M, Faria R (2013) A prominent role of KRAB-ZNF transcription factors in mammalian speciation? Trends Genet 29:130–9.

Oliver PL, Goodstadt L, Bayes JJ, et al (2009) Accelerated evolution of the *Prdm9* speciation gene across diverse metazoan taxa. PLoS Genet 5:e1000753.

Parvanov ED, Petkov PM, Paigen K (2010) *Prdm9* controls activation of mammalian recombination hotspots. Science 327:835.

Peric-Hupkes D, Meuleman W, Pagie L, et al (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell 38:603–13.

Peric-Hupkes D, van Steensel B (2010) Role of the nuclear lamina in genome organization and gene expression. Cold Spring Harb Symp Quant Biol 75:517–24.

Pevzner P, Tesler G (2003) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc Natl Acad Sci 100:7672–7.

Phipson B, Smyth GK (2010) Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Stat Appl Genet Mol Biol 9:Article39.

Reddy KL, Zullo JM, Bertolino E, Singh H (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. Nature 452:243–7.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. Trends Ecol Evol 16:351–358.

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

Ruiz-Herrera A, Garcia F, Frönicke L, et al (2004) Conservation of aphidicolin-induced fragile sites in *Papionini* (Primates) species and humans. Chromosome Res 12:683–90.

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

Scally A, Dutheil JY, Hillier LW, et al (2012) Insights into hominid evolution from the gorilla genome sequence. Nature 483:169–75.

Seehausen O, Butlin RK, Keller I, et al (2014) Genomics and the origin of species. Nat Rev Genet 15:176–92.

Segura J, Ferretti L, Ramos-onsins S, et al (2013) Evolution of recombination in eutherian mammals : insights into mechanisms that affect recombination rates and crossover interference. Proc R Soc B 280:20131945.

Silva M, 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.

Cytogenet Cell Genet 81:46–50.

Snyder SH, Sklar PB, Hwang PM, Pevsner J (1989) Molecular mechanisms of olfaction. Trends Neurosci 12:35–8.

Stanyon R, Yang F, Cavagna P, et al (1999) Reciprocal chromosome painting shows that genomic rearrangement between rat and mouse proceeds ten times faster than between humans and cats. Cytogenet Cell Genet 84:150–5.

Steiner CC, Ryder OA (2013) Characterization of *Prdm9* in equids and sterility in mules. PLoS One 8:e61746.

Stopková R, Hladovcová D, Kokavec J, Vyoral D (2009) Multiple roles of secretory lipocalins (Mup, Obp) in mice. 58:29–40.

Storz JF, Sabatino SJ, Hoffmann FG, et al (2007) The molecular basis of high-altitude adaptation in deer mice. PLoS Genet 3:e45.
Storz JF, Runck AM, Sabatino SJ, et al (2009) Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. Proc Natl Acad Sci 106:14450–5.

Sun F, Oliver-Bonet M, Liehr T, et al (2005) Discontinuities and unsynapsed regions in meiotic chromosomes have a cis effect on meiotic recombination patterns in normal human males. Hum Mol Genet 14:3013–8.

Suzuki K, Yu X, Chaurand P, et al (2007) Epididymis-specific lipocalin promoters. Asian J Androl 9:515–21.

Turner LM, White MA, Tautz D, Payseur BA (2014) Genomic networks of hybrid sterility. PLoS Genet 10:e1004162.

Turner TL, Hahn MW, Nuzhdin S V (2005) Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol 3:e285.

Ullastres A, Farré M, Capilla L, Ruiz-Herrera A (2014) Unraveling the effect of genomic structural changes in the rhesus macaque - implications for the adaptive role of inversions. BMC Genomics 15:530.

Veyrunes F, Dobigny G, Yang F, et al (2006) Phylogenomics of the genus Mus (Rodentia; Muridae): extensive genome repatterning is not restricted to the house mouse. Proc Biol Sci 273:2925–34.

White MJD (1978) Modes of Speciation. W.H.Freeman & Co.

Wu CI, Li WH (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. Proc Natl Acad Sci 82:1741–1745.

Zhao H, Bourque G (2009) Recovering genome rearrangements in the mammalian phylogeny. Genome Res 19:934–42.

Zhao S, Shetty J, Hou L, et al (2004) Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. Genome Res 14:1851–60.

Reeves A (2001) Micromeasure: a new computer program for the collection and analysis of cytogenetic data. Genome 44: 439– 443.

4.1.2 Supplementary information

Table S1. Species included in the analysis: Data regarding taxonomy classification, genome version, N50 and diploid number (2n) are included. The majority of the species presented their genomes assembled in chromosomes with the exception of *H. glaber, J. jaculus* and *S. galilii*, whose genomes were only available into scaffolds. In the case of *M. ochrogaster* we considered all data available (assembled chromosomes and linkage groups). All genomes, except for *S. galilii*, were downloaded from Genbank FTP site (ftp://ftp.ncbi.nlm.nih.gov).

Order	Family	Species	Genome Version	N50	2n
Artiodactyla	Bovidae	*Bos taurus*	Bos_taurus_UMD_3.1	Chromosomes	60
Perissodactyla	Equidae	*Equus caballus*	EquCab2.0	Chromosomes	66
Carnivora	Felidae	*Felis catus*	Felis_catus_6.4	Chromosomes	68
Primates	Hominidae	*Homo sapiens*	GRCh37.p10	Chromosomes	46
Primates	Hominidae	*Pongo pygmaeus*	P_pygmaeus_2.0.2	Chromosomes	48
Primates	Cercopithecidae	*Macaca mulatta*	Mmul_p51212	Chromosomes	42
Rodentia	Bathyergidae	*Heterocephalus glaber*	HetGla_female_1.0	20.53Mb	60
Rodentia	Dipodidae	*Jaculus jaculus*	JacJac1.0	22.08Mb	48
Rodentia	Spalacidae	*Spalax galilii*	Fang et al. 2014	3.6Mbp	60
Rodentia	Cricetidae	*Microtus ochrogaster*	MicOch1.0	Chromosomes	54
Rodentia	Muridae	*Mus musculus*	NCBIm37	Chromosomes	40
Rodentia	Muridae	*Rattus norvegicus*	Rnor5.0	Chromosomes	42

Table S2. List of HSBs and SFs obtained for each pair-wise comparison (300Kbp resolution): In all cases, the mouse genome was used as reference (version NCBIm37). "N" denotes the number of HSBs and SFs detected and "type" refers to the type of syntenic region. Total, mean, maximum and minimum lengths are expressed in Mbp.

Species Compared	N	Type	Length (Mbp)				% Genome
			Total	Mean	Minimum	Maximum	
R. norvegicus	280	HSB	2,512.37	13.22	0.30	127.98	95.60
M. ochrogaster	459	HSB	2,370.98	6.43	0.31	60.48	90.22
S. galilii	1985	SF	2,153.00	1.14	0.30	10.91	81.93
J. jaculus	598	SF	2,385.86	4.70	0.30	43.68	90.79
H. glaber	559	SF	2,410.46	5.14	0.31	30.23	91.72
P. pygmaeus	420	HSB	2,446.09	7.41	0.30	52.69	93.08
H. sapiens	459	HSB	2,421.46	6.56	0.31	66.22	92.14
M. mulatta	437	HSB	2,454.65	7.07	0.30	52.06	93.41
F. catus	391	HSB	2,428.35	8.07	0.30	59.58	92.41
E. caballus	425	HSB	2,457.01	7.33	0.30	71.36	93.49
B. taurus	521	HSB	2,394.30	5.56	0.31	58.78	91.11

Table S3. EBRs identified: Twelve lineage-specific (*R. norvegicus, M. ochrogaster, S. galilii, J. jaculus, H. glaber, P. pygmaeus, H. sapiens, M. mulatta, F. catus, E. caballus* and *B. taurus*) and eight clade-specific (*Muridae, Cricetidae+Muridae, Muroidea, Myodonta, Rodentia, Hominoidea, Catarrhini, Laurasiatheria,* and *Euarchontoglires*) pair-wise comparisons were established using *M. musculus* as the reference genome. Reused EBRs shared by any of the 11 species used in the study are also shown. N denotes the number of EBRs detected. Total, mean, minimum and maximum lengths are expressed in Kbp.

			Length (Kbp)				% Genome
Type	**Species/Clade**	**N**	**Total**	**Mean**	**Minimum**	**Maximum**	**Coverage**
Lineage-specific	*M. musculus*	30	2.388	79.62	8.64	336.15	0.09
	R. norvegicus	66	8.766	132.82	28.88	1,129.23	0.33
	M. ochrogaster	133	12.313	92.58	4.44	344.28	0.47
	S. galilii	360	51.682	143.56	10.72	1,306.54	1.97
	J. jaculus	128	13.372	104.47	2.49	557.17	0.51
	H. glaber	91	10.186	111.94	1.36	666.99	0.39
	P. pygmaeus	8	1.137	142.20	42.03	288.52	0.04
	H. sapiens	29	4.216	145.39	39.52	1,150.26	0.16
	M. mulatta	26	3.948	151.87	30.72	776.72	0.15
	F. catus	24	2.49	103.79	24.46	498.64	0.09
	E. caballus	36	3.145	87.37	4.31	415.19	0.12
	B. taurus	118	13.484	114.28	5.39	780.16	0.51
Clade-specific	Murinae	15	833	55.58	5.08	102.59	0.03
	Muridae	28	2.48	88.6	29.31	303.16	0.09
	Muroidea	3	230	76.70	2.16	149.04	0.01
	Myodonta	14	1.372	98.00	22.37	394.80	0.05
	Rodentia	15	1.053	70.21	35.65	132.51	0.04
	Hominoidea	8	710	88.76	22.46	173.28	0.03
	Catarrhini	33	2.036	61.71	9.27	148.24	0.08
	Laurasiatheria	12	1.552	129.41	15.00	466.36	0.06
	Euarchontoglires	2	270	135.32	34.21	236.44	0.01
Reused		154	17.462	113.39	0.154	3,056.91	0.66

Table S4. HSBs and SFs at different resolutions: Comparison of the number of HSBs and SFs for each *Synteny Tracker* pair-wise comparison and for each resolution (100Kbp, 300Kbp and 500Kbp).

Species	**100Kbp**	**300Kbp**	**500Kbp**
Felis catus	371	281	245
Bos taurus	560	411	344
Equus caballus	437	315	270
Pongo pygmaeus	439	310	277
Macaca mulatta	472	327	283
Homo sapiens	496	349	292
Heterocephalus glaber	535	449	416
Jaculus jaculus	600	488	437
Spalax galilii	2417	1876	1477
Microtus ochorgaster	431	350	327
Rattus norvegicus	577	170	109

Table S5: Divergence times. Phylogenetic distances described by Meredith et al. (2011) (autocorrelated rates and hard-bounded constraints) and by dos Reis et al. (2012) (marginal prior divergence times) "na" denotes data not available. Values are mean and 95% CI (in brackets).

	Divergence Times (Myr)			
Clade	**Meredith et al. (2011)**		**dos Reis et al. (2012)**	
Placentalia	96.4	(92.12-99.29)	124.1	(102.1-139.2)
Laurasiatheria	81.32	(78.46-83.28)	114.6	(87.4-133.3)
Euarchontoglires	78.45	(74.08-80.12)	114	(87.4-133.3)
Catarrhini	20.85	(20.56-21.65)	29.5	(23.7-34.1)
Hominoidea	15.98	(14.41-17.54)	na	na
Rodentia	66.03	(64.12-66.70)	61.1	(56.0-65.8)
Myodonta	52.42	(50.34-54.06)	na	na
Muroidea	42.48	(40.73-44.42)	na	na
Muridae	23.48	(21.09-25.74)	na	na
Murinae	na	na	12.2	(10.4-14.0)

Table S6: List of disrupted genes

chr	gene (ncbim37)	gene start	gene end	EBR start (bp)	EBR end (bp)	phylogenetic group	disruption site	disruption site	Name
1	ENSMUSG00000026095	53382351	53409596	53383643	53477238	mouse	3' region	3' region	Asnsd1-001
1	ENSMUSG00000026401	132285103	132319586	132241026	132313250	muridae	INTRON	INTRON 5-6	Daf2-001
1	ENSMUSG00000038633	184205903	184212935	184079303	184211812	rodent	INTRON	INTRON 1-2	Degs1-001
2	ENSMUSG00000036449	25508640	25511737	25510722	25615814	rodent	EXON	EXON 4	Lcn8-001
2	ENSMUSG00000026753	39049874	39081971	39055148	39115672	muridae	EXON	EXON 7	Ppp6c-001
2	ENSMUSG00000026825	32163991	32208849	32085580	32164488	muroidea	EXON	EXON 21	Dnm1-005
3	ENSMUSG00000074655	28791539	28825646	28741302	28797838	mouse	INTRON	INTRON 3-4	m1527-201
5	ENSMUSG00000028999	23293529	23326187	23295729	23375080	muridae	INTRON	INTRON 2-3	Rint1-001
5	ENSMUSG00000014956	32761347	32819806	32765594	33101739	mouse	INTRON	INTRON 1-2	Ppp1cb-001
5	ENSMUSG00000023452	33078949	33128295	32765594	33101739	mouse	INTRON	INTRON 3-4	Pisd-001
5	ENSMUSG00000029507	111202686	111209678	111208811	111283234	g1	EXON	EXON 2	Pus1-002
5	ENSMUSG00000029521	111269036	111303152	111208811	111283234	g1	INTRON	INTRON 5-6	Chek2-201
5	ENSMUSG00000042216	113672240	113739806	113709212	113828290	g2	INTRON	INTRON 10-11	Sgsm1-001
5	ENSMUSG00000072694	115258705	115273477	115271581	115377612	g2	discrepancies in gene position v37	???	1500011B03Ri 001
5	ENSMUSG00000041827	115373249	115387924	115271581	115377612	g2	INTRON	INTRON 1-2	Oasl1-002
5	ENSMUSG00000029727	138334160	138362847	138302234	138340445	mouse	INTRON	INTRON 10-11	Cyp3a13-001
6	ENSMUSG00000030089	90554719	90596406	90565760	90639578	g2	INTRON	INTRON 2-3	Slc41a3-001
6	ENSMUSG00000072623	118411968	118429338	118413942	118546832	g2	EXON	EXON 6	Zfp9-001
6	ENSMUSG00000030180	120314117	120394592	120363590	120443311	g2	INTRON	INTRON 19-20	Kdm5a-201
6	ENSMUSG00000030107	121195924	121220934	121207056	121601857	g2	INTRON	INTRON 4-5	Usp18-201
6	ENSMUSG00000024104	116158051	116212686	116114690	116188257	g1	INTRON	INTRON 19	Fam21-201
6	ENSMUSG00000030111	121586191	121629255	121207056	121601857	g2	INTRON	INTRON 14-15	A2m-201
6	ENSMUSG00000030359	128433585	128476738	128399721	128464831	rodent	INTRON	INTRON 10-11	Pzp-201
7	ENSMUSG00000030835	53289066	53339582	53308744	53374272	muridae	INTRON	INTRON 11-12	Nomo1-001
7	ENSMUSG00000030781	135409171	135415944	135411640	135524011	g1	INTRON	INTRON 5-6	Slc5a2-001

7	ENSMUSG00000030844	135517135	135562272	135411640	135524011	g1	INTRON	INTRON 3-4	Rgs10-001
8	ENSMUSG00000031574	26918946	26925782	26920431	26998236	g2	INTRON	INTRON 3-4	Star-001
8	ENSMUSG00000031805	74200195	74214474	74213382	74365250	mouse	INTRON	INTRON 24-25	Jak3-001
8	ENSMUSG00000079019	74213113	74214474	74213382	74365250	mouse	INTRON	INTRON 1-2	Insl3-001
8	ENSMUSG00000031622	75247187	75282102	75281769	75321990	mouse	INTRON	INTRON 19	Sin3b-201
8	ENSMUSG00000031967	126001803	126027816	126027276	126184642	g1	INTRON	INTRON 17	Afg3l1-001
8	ENSMUSG00000039960	126177829	126187784	126027276	126184642	g1	INTRON	INTRON 2-3	Rhou-001
9	ENSMUSG00000041986	53759267	53823108	53801032	53902951	g1	INTRON	INTRON 1-2	Elmod1-201
9	ENSMUSG00000074345	53873413	53916218	53801032	53902951	g1	INTRON	INTRON 1-2	Tnfaip8l3-201
9	ENSMUSG00000037493	54392601	54408025	54342642	54407473	g1	INTRON	INTRON 1-2	Cib2-001
9	ENSMUSG00000037410	90096885	90165642	90140251	90222237	g1	INTRON	INTRON 2-3	Tbc1d2b-001
9	ENSMUSG00000020258	106055188	106060469	105997280	106058783	rodent	INTRON	INTRON 3	Glyctk-002
9	ENSMUSG00000032496	110921796	110945270	110937469	111018429	g2	INTRON	INTRON 13-14	Ltf-201
9	ENSMUSG00000032446	118387330	118395250	118392005	118440224	g2	INTRON	INTRON 5-6	Eomes-001
10	ENSMUSG00000062593	51200446	51216419	51207626	51285212	mouse	INTRON	INTRON 4-5	Lilrb4-201
11	ENSMUSG00000020268	54640368	54674418	54655891	54724903	rodent	INTRON	INTRON 4-5	Lyrm7-005
12	ENSMUSG00000001627	40928154	40975091	40859948	40930665	rodent	INTRON	INTRON 10-11	Ifrd1-001
12	ENSMUSG00000021076	72038844	72065705	71998160	72057812	g1	INTRON	INTRON 10-11	Actr10-201
13	ENSMUSG00000021500	55736388	55782617	55680231	55754460	rodent	INTRON	INTRON 8-9	Ddx46-202
14	ENSMUSG00000021982	60179733	60216673	60185855	60253528	g2	INTRON	INTRON 9-10	Cdadc1-203
14	ENSMUSG00000021810	21139081	21167343	20839441	21142605	g1	INTRON	INTRON 13-14	Ecd-201
15	ENSMUSG00000018893	76845919	76881100	76768807	76848361	g1	EXON	EXON 2	Mb-203
17	ENSMUSG00000073471	8138518	8172689	8146616	8220308	mouse	INTRON	INTRON 2-3	Rsph3a-201
17	ENSMUSG00000060475	13159662	13185412	13179122	13269924	mouse	INTRON	INTRON 1-2	Wtap-002
17	ENSMUSG00000055602	13253977	13275347	13179122	13269924	mouse	INTRON	INTRON 9-10	Tcp10b-003
17	ENSMUSG00000024188	26389855	26422449	26394039	26490733	mouse	INTRON	INTRON 2-3	Luc7l-003
17	ENSMUSG00000024193	27070072	27074835	27006215	27074393	mouse	EXON	EXON 15	Phf1-201

17	ENSMUSG00000024027	31038785	31073455	31045859	31121158	g1	INTRON	INTRON 2-3	Glp1r-201
17	ENSMUSG00000054134	31091624	31147653	31045859	31121158	g1	INTRON	INTRON 9-10	Umodl1-203
17	ENSMUSG00000038146	32257765	32303825	32209345	32287588	mouse	EXON	EXON 15	Notch3-001
17	ENSMUSG00000041881	33961559	33975258	33961996	34041618	mouse	INTRON	INTRON 1-2	Ndufa7-001
17	ENSMUSG00000024056	71845440	71876197	71865642	71932256	mouse	INTRON	INTRON 7-8	Ndc80-201
17	ENSMUSG00000052525	71901401	71938873	71865642	71932256	mouse	INTRON	INTRON 6-7	Spdya-201
18	ENSMUSG00000024231	3382986	3436698	3409154	3481072	g2	INTRON	INTRON 2-3	Cul2-001
18	ENSMUSG00000005873	34504543	34533069	34532486	34599406	rodent	INTRON	INTRON 1-2	Reep5-201
18	ENSMUSG00000024259	31739822	31770183	31690467	31763255	g1	INTRON	INTRON 4-5	Slc25a46-201
18	ENSMUSG00000014503	34569077	34602445	34532486	34599406	rodent	INTRON	INTRON 14-15	Pkd2l2-201
19	ENSMUSG00000024863	30307418	30314175	30310005	30332371	g2	INTRON	INTRON 4-5	Mbl2-201

Table S7. Enriched functional annotation charts in all Rodentia EBRs:

Category	Term
	Cluster 1: enrichment Score: 4.51
INTERPRO	Calycin
INTERPRO	Lipocalin
INTERPRO	Lipocalin-related protein and Bos/Can/Equ allergen
INTERPRO	Von Ebner's gland protein/Bos/Can allergen
	Cluster 2: enrichment Score: 3.6
INTERPRO	Hemoglobin, alpha
GOTERM_CC_FAT	Hemoglobin complex
GOTERM_BP_FAT	Oxygen transport
INTERPRO	Globin, subset
INTERPRO	Globin
GOTERM_MF_FAT	Oxygen transporter activity
GOTERM_MF_FAT	Oxygen transport
SP_PIR_KEYWORDS	Gas transport
GOTERM_BP_FAT	*Oxygen binding*
GOTERM_MF_FAT	*Cytosolic part*
GOTERM_CC_FAT	*Iron*
SP_PIR_KEYWORDS	*Iron ion binding*
GOTERM_MF_FAT	*Heme binding*
GOTERM_MF_FAT	*Heme*
SP_PIR_KEYWORDS	*Tetrapyrrole binding*
GOTERM_MF_FAT	*cytosol*
	Cluster 3: enrichment score: 1.69
SMART	*KRAB*
GOTERM_BP_FAT	*Regulation of transcription, DNA dependent*
GOTERM_BP_FAT	*Regulation of RNA metabolic process*
INTERPRO	*Krueppel-associated box*

Figure S1. HSBs and SFs: Number of HSBs and SFs detected by *Synteny Tracker* for each of the pair-wise comparisons and for each resolution (100Kbp, 300Kbp and 500Kbp).

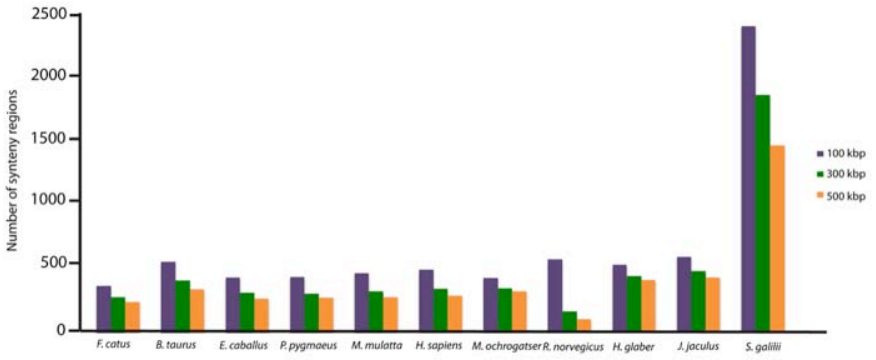


Figure S2. Distribution of unique EBRs across the mouse genome: Frequency of EBRs in the mouse genome (lineage and clade-specific) (n=105) detected for each chromosome. Dotted line represents the estimated frequency of EBRs in the mouse genome assuming a homogeneous distribution.(χ^2 test, ** p-value<0.001).

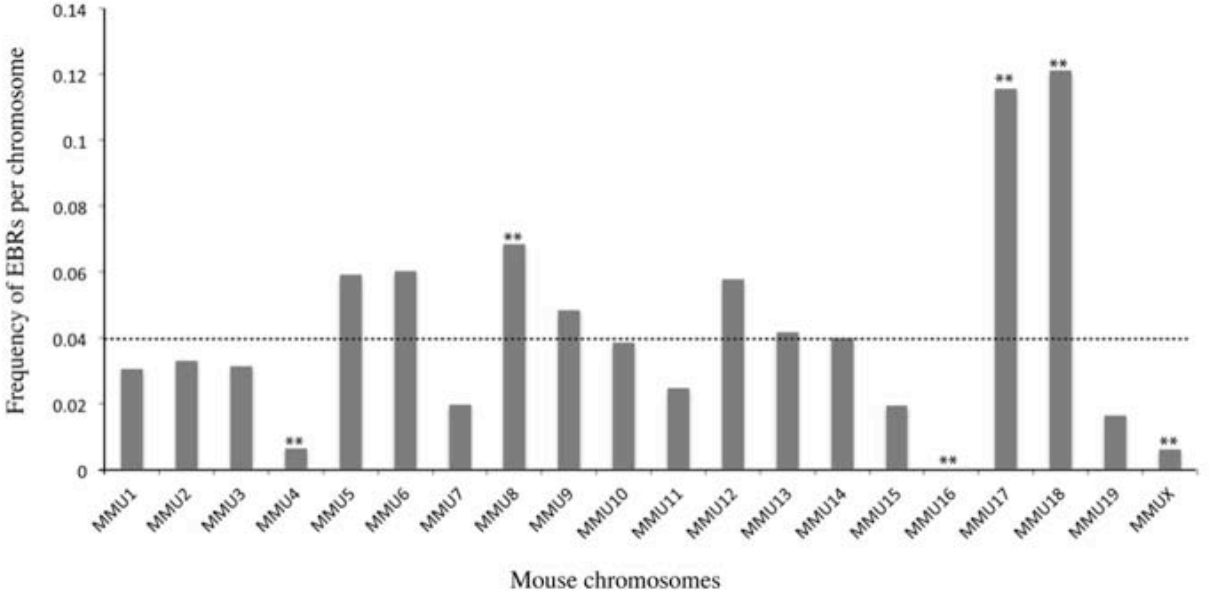
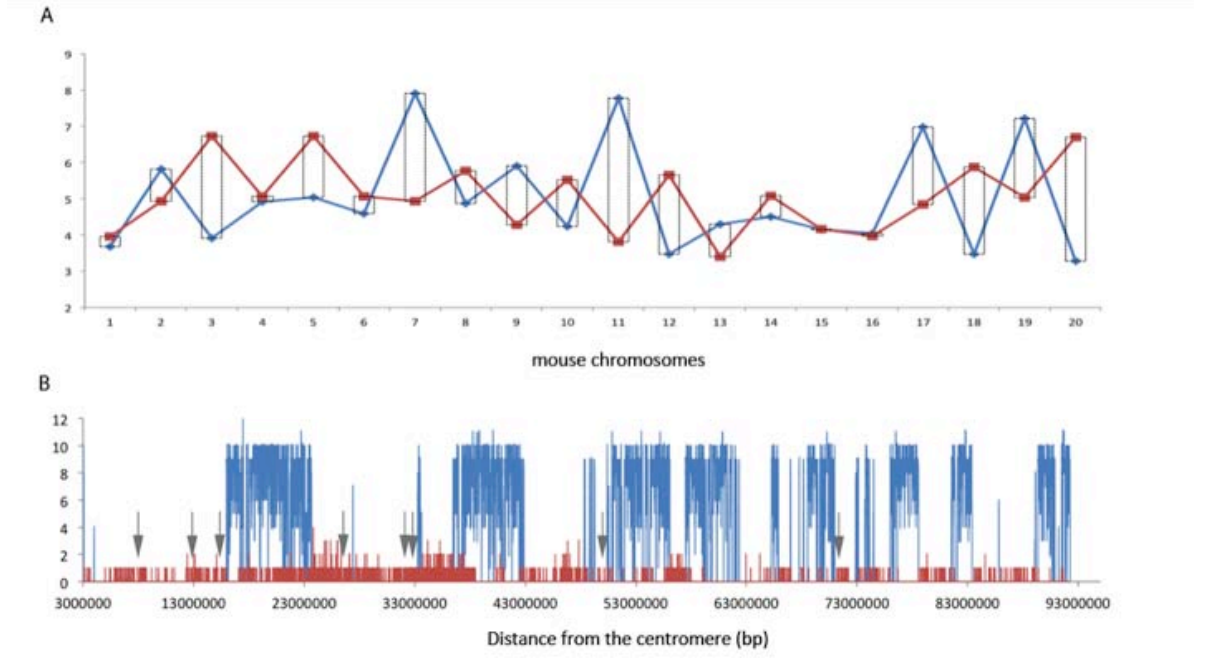


Figure S3. Genome-wide distribution of cLADs and genes in the mouse genome: (A) Number of protein coding genes (blue) and cLADs (red) per each mouse chromosome. Mean values of genes (blue line) and cLADs (red line) per 10Kbp windows are represented in the y-axis. (B) Genome distribution of protein coding genes (red) and cLADs (blue) along mouse chromosome 17. Number of genes (blue line) and cLADs (red line) per 10Kpb windows are represented in the y-axis. Arrows indicate the position of estimated EBRs in this work.



4.2 Study of the role of telomeres in the formation of Rb fusions in the Barcelona Rb system

Despite the role of chromosomal rearrangements (such as Rb fusions) in speciation has been long discussed in evolutionary biology, the mechanisms involved in their origin and fixation in nature remain largely unexplored. The wide distribution of Rb fusions in mice natural populations represents an interesting model to analyze the mechanisms that are driving its origin and fixation. Here, we have analyzed the role of telomere shortening in the origin of Rb fusions by using as a model the Barcelona Rb polymorphism zone.

4.2.1 On the origin of Robertsonian fusions in nature: evidence of telomere shortening in wild house mouse

Sánchez-Guillén RA, Capilla L, Reig-Viader R, Martínez-Plana M, Pardo-Camacho C, Andrés-Nieto M, Ventura J, Ruiz-Herrera A (2015) **On the origin of Robertsonian fusions in nature: evidence of telomere shortening in wild house mouse**. Journal of Evolutionary Biology. 28:241-249.

Impact factor: 3.48
Q1 in Ecology
Q2 in Evolutionary Biology
Q2 in Genetics and Heredity

JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12568

On the origin of Robertsonian fusions in nature: evidence of telomere shortening in wild house mice

R. A. SÁNCHEZ-GUILLÉN*, L. CAPILLA*†, R. REIG-VIADER*‡, M. MARTÍNEZ-PLANA*‡, C. PARDO-CAMACHO*‡, M. ANDRÉS-NIETO*‡, J. VENTURA† & A. RUIZ-HERRERA*‡

*Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona, Barcelona, Spain
†Departament de Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona, Barcelona, Spain
‡Departament de Biologia Cel lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Barcelona, Spain

Keywords:

chromosomal evolution;
Robertsonian fusions;
telomere shortening;
western house mouse.

Abstract

The role of telomere shortening to explain the occurrence of Robertsonian (Rb) fusions, as well as the importance of the average telomere length vs. the proportion of short telomeres, especially in nature populations, is largely unexplored. In this study, we have analysed telomere shortening in nine wild house mice from the Barcelona Rb system with diploid numbers ranging from 29 to 40 chromosomes. We also included two standard (2n = 40) laboratory mice for comparison. Our data showed that the average telomere length (considering all chromosomal arms) is influenced by both the diploid number and the origin of the mice (wild vs. laboratory). In detail, we detected that wild mice from the Rb Barcelona system (fused and standard) present shorter telomeres than standard laboratory mice. However, only wild mice with Rb fusions showed a high proportion of short telomeres (only in p-arms), thus revealing the importance of telomere shortening in the origin of the Rb fusions in the Barcelona system. Overall, our study confirms that the number of critically short telomeres, and not a simple reduction in the average telomere length, is more likely to lead to the origin of Rb fusions in the Barcelona system and ultimately in nature.

Introduction

Chromosomal rearrangements in the form of inversions, translocations and chromosome fusions/fissions are key factors for evolution. This evolutionary role is especially true for Robertsonian (Rb) fusions, whose importance is revealed by its occurrence in taxa as diverse as mammals, reptiles, insects or mollusks (White, 1973; King, 1995). For instance, Rb fusions, together with inversions, contribute to chromosomal speciation through underdominance and/or by suppression of recombination when they are present in heterozygous form (Rieseberg, 2001; Dumas & Britton-Davidian, 2002; Faria & Navarro, 2010; Farré *et al.*, 2013; Capilla *et al.*, 2014). Although most of the recent literature has been focused on the diversity of factors influencing the fixation of chromosomal rearrangements within populations (Faria & Navarro, 2010 and references therein), the mechanism(s) responsible for the origin of Rb fusions have received less attention.

Robertsonian fusions occur when two telocentric or acrocentric chromosomes fuse resulting in one metacentric chromosome (Robertson, 1916). Several factors have been invoked for the appearance of Rb fusions in mammalian species; these include chromosome size, GC and DNA content, gene density, illegitimate recombination between repetitive sequences and telomere loss or inactivation (Slijepcevic, 1998; Garagna *et al.*, 2001; Ruiz-Herrera *et al.*, 2010; Wesche & Robinson, 2012). Telomeres are ribonucleoprotein structures composed of TTAGGG tandem repeats, a protein complex (shelterin) and noncoding telomeric RNA (TERRA) that cap the ends of eukaryotic chromosomes to protect them from being recognized as double-strand breaks (DSBs) (Azzalin *et al.*, 2007; O'Sullivan & Karlseder, 2010; Reig-Viader *et al.*, 2014). Therefore, Rb fusions

Correspondence: Rosa A. Sanchéz-Guillén and A. Ruiz-Herrera, Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona, Campus UAB, 08193 Barcelona, Spain.
Tel.: +34 93 5811379, +34 93 5812051; fax: +34 93 581 3357; e-mails: rguillenuvigo@hotmail.com and aurora.ruizherrera@uab.es

would require either the elimination or the inactivation of telomeres prior to the chromosomal rearrangement. In this context, three different mechanisms have been formally proposed to explain the origin of Rb fusions in relation to the presence/absence of telomeric sequences in the resultant fused chromosome: (i) telomeric inactivation, (ii) chromosomal breakage within centromeric satellite sequences and (iii) telomeric shortening (Slijepcevic, 1998).

The first mechanism assumes the occurrence of chromosomal fusions by telomere inactivation without loss of telomeric repeats (Slijepcevic, 1998), and occasionally, the fused chromosomes would exhibit large blocks of telomeric repeats that would be retained and/or amplified in the pericentromeric regions (see Ruiz-Herrera *et al.*, 2008 for a review). In fact, the presence of telomeres with an altered structure is common in many vertebrates (see Slijepcevic, 1998), with numerous examples of end-to-end telomere fusions in the absence of significant telomere shortening (e.g. Bailey *et al.*, 1999; Espejel & Blasco, 2002). According to the second mechanism (chromosomal breakage in small satellite sequences), chromosomal breaks within centromeric repetitive sequences would result in the total loss of telomeres of the p-arms, promoting fusion of the chromosomal ends. This mechanism has been proposed to explain chromosomal evolution in several rodent species (Nanda *et al.*, 1995; Garagna *et al.*, 2001). The last of the suggested mechanisms (telomeric shortening) proposes that Rb fusions could result from progressive telomere shortening due to the 'end of replication problem' (Allsopp *et al.*, 1995). This states that due to the inefficiency of DNA polymerase to replicate the whole telomeric sequences of the lagging strand during the S-phase of mitotic division, telomeres progressively shorten, resulting in loss of the telomeric DNA and reaching critical length in the absence of telomere elongation mechanisms. This would precipitate the fusion of chromosomal ends (Slijepcevic, 1998). Knockout mice lacking telomerase in successive generations represent a good example of chromosomal fusion due to telomere shortening (Blasco *et al.*, 1997). In fact, Hemann *et al.* (2001) detected a preference for the fusion of chromosomes with short telomeres in telomerase-null mice. They predict that as telomeres shorten, chromosomes undergo fusion. Nevertheless, and despite the experimental evidence, telomere length has never been formally studied in natural Rb populations.

The western house mouse (*Mus musculus domesticus*) is arguably the best studied and understood model of Rb variation in nature. The accumulation of Rb fusions is thought to encompass the past 10 000 years (Auffray, 1993). The standard karyotype of the house mouse consists of all acrocentric chromosomes (2n = 40); however, a wide range of diploid numbers (2n) (from 22 to 39) has been described in the literature (Piálek *et al.*, 2005). In fact, 106 different Rb fusion combinations are distributed across Europe and the Mediterranean basin. One of these Rb systems, the Barcelona Rb system, has diploid numbers ranging from 27 to 39 chromosomes (Medarde *et al.*, 2012 and references therein). This system occurs in a 5000 km^2 area of Barcelona province (Spain) and is characterized by a high level of chromosomal polymorphisms [i.e. seven different Rb chromosomes including Rb (3.8), (4.14), (5.15), (6.10), (7.17), (9.11) and (12.13) showing nongeographically coincident clines] (Medarde *et al.*, 2012) and low recombination rates (Capilla *et al.*, 2014). To disentangle the mechanism(s) responsible for the formation of Rb fusions in natural populations, we have investigated telomere shortening in the Barcelona Rb system. To this end, telomere length and the proportion of short telomeres were measured in standard and Rb mice from the system, and compared to standard laboratory mice (C57BL/6; B6). Our results indicate that telomere shortening can explain the origin of the Rb fusions in the Barcelona system and thus confirm the role of telomere shortening in shaping the occurrence of Rb fusions in nature.

Materials and methods

Cell cultures and chromosome extraction

Primary fibroblast cell lines were previously established [from nine wild male mice livetrapped in commensal habitats from five different localities representative of the Barcelona Rb system] (Fig. 1a) and karyotyped by Capilla *et al.* (2014). Cell lines were characterized by diploid numbers that ranged from 29 to 40 (2n = 40, three mice; 2n = 38, one mouse; 2n = 37, one mouse; 2n = 31, one mouse; 2n = 30, two mice; and 2n = 29, one mouse) (Table 1). In addition, two primary fibroblast cell lines were established from two laboratory mice (2n = 40, B6), which together with HeLa cells (human cells with known telomere length) were included in the analysis as internal controls for telomere length measurements. Tissue culture protocols followed Capilla *et al.* (2014). Chromosome preparations were obtained at early passages in culture (from 2th to 7th).

Quantitative-fluorescence *in situ* hybridization (Q-FISH)

Q-FISH method on metaphases was used based on its high sensitivity, which allows the quantification of individual telomeres in single cells, as well as the identification of short telomeres (telomeric repeats < 0.15 kbp, the so-called 'signal-free ends'; Poon *et al.*, 1999; Aubert *et al.*, 2012), which appear as telomeres without FISH signal. Metaphase chromosomes were hybridized using the peptide–nucleic acid-FISH (PNA-FISH) preparation method described by the manufacturer

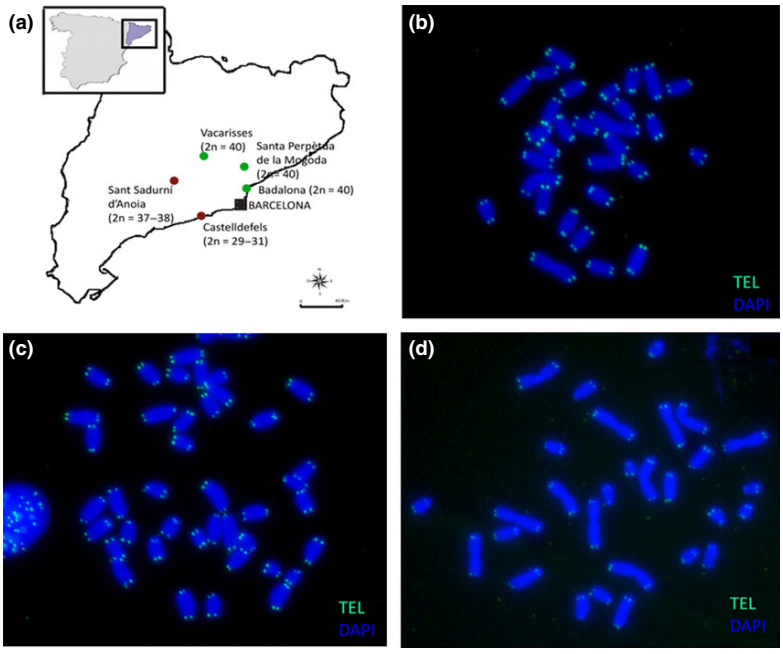


Fig. 1 Distribution and type of Rb fusions. (a) Sample distribution in the Barcelona chromosomal polymorphic zone. Diploid numbers are shown for standard (green) and Rb (dark red) animals of the Rb system. (b–d) Examples of Q-FISH experiments performed on metaphases of specimens with (b) three fusions, specimen SS18, (c) standard karyotype from a B6 mouse and (d) ten fusions, specimen 968. Telomeres are shown in green (TEL) and chromosomes are stained in blue (DAPI).

(www.panagene.com). The FAM-conjugated $(CCCTAA)_3$ PNA probe (telo C) (Panagen) was used to label telomeres; DNA counterstaining was performed using DAPI (40,6-diamidino-2-phenylindole). For each Q-FISH experiment, an internal control consisting of HeLa cells derived from the same chromosome extraction (i.e. same passage) was included.

Telomere length measurements

Chromosomal preparations were visualized with a Zeiss Axioskop epifluorescence microscope equipped with the appropriate filters for FAM/DAPI detection and a charged coupled device camera (ProgResR CS10plus, Jenoptik, Jena, Germany). Images were captured and produced by the PROGRESR software (2.7.7), maintaining the same exposure time. Fluorescence intensities of telomeres were analysed with the TFLTELO-v2 software package (Terry Fox Laboratory, BC Cancer Research Centre, Vancouver, BC, Canada) (Poon *et al.*, 1999). Analyses of telomere fluorescence were carried out maintaining the same tolerance threshold values. At least 28 metaphases were analysed per cell line (Table 1) given that this sample size gives comparable results to conventional southern analysis (Poon *et al.*, 1999; Vera & Blasco, 2012). Telomere lengths were obtained in terms of arbitrary telomere fluorescence units (TFUs) following previous studies (Reig-Viader *et al.*, 2014).

We investigated the phenomenon of telomere shortening in terms of telomere length and proportion of short telomeres, which normally participate in end-to-end fusions. To this end, q- and p-arms of each acrocentric and metacentric (fused) chromosome were identified in each metaphase and their telomere lengths and proportion of short telomeres were measured. In this study, telomeres of metacentric chromosomes were always considered as q-arm telomeres. Analyses were performed following three approaches. Firstly, we considered only the type of chromosome arm (q-arms vs. p-arms). Secondly, we considered only diploid numbers, and thus we included six groups: (i) 2n = 40, (ii) 2n = 38, (ii) 2n = 37, (iv) 2n = 31, (v) 2n = 30 and (vi) 2n = 29. Thirdly, analyses were performed considering both origin and type of mice: (i) Laboratory-standard, which included B6 laboratory mice (2n = 40, B6, two mice); (ii) Rb-standard, which included standard wild mice from the Rb system (2n = 40, three mice) and (iii) Rb-fused, which included wild mice from the Rb system with Rb fusions (2n = 29 to 2n = 38, six mice).

Correlation between q- and p-arm telomeres

Telomere length data were normally distributed, and thus we used an ANOVA (as implemented in JMP 11th version) to compare q- and p-arm telomere lengths and also to test whether p-arms are more prone to loss telomere FISH signals (Almeida & Ferreira, 2013). Additionally, we investigated whether telomere shortening is similarly affecting q- and p-arms. To this end, length ratio (q-arm/p-arm) in the three types of mice was investigated using an ANOVA.

Table 1 Data characteristics of the specimens analysed: locality, group of animals, diploid number (2n), number of metaphases and telomeres analysed, mean telomere lengths (expressed as TFUs), proportion of short telomeres (S.T.) and proportion of metacentric (Meta) chromosomes.

						Acrocentric chromosomes					Metacentric chromosomes		
						Q-arm		P-arm			Q-arm		
Specimen	Locality	Group	2n	Metaphases (telomeres scored)	Mean TFUs per cell	Mean ± SE (TFUs)	S.T. (%)	Mean ± SE (TFUs)	S.T. (%)	Meta (%)	Mean ± SE (TFUs)	S.T. (%)
B6-1	Lab strain	Laboratory-standard	40	30 (3944)	1852.63	2311.80 ± 19.62	0.05	1569.76 ± 16.06	2.23	0.00	0.00	0.00
B6-3	Lab strain	Laboratory-standard	40	31 (4036)	1944.87	2041.34 ± 15.46	0.64	1662.98 ± 13.80	1.14	0.00	0.00	0.00
900	Santa Perpètua de la Mogoda	Rb-standard	40	30 (3944)	1746.66	1932.05 ± 15.28	0.48	1557.75 ± 13.16	2.34	0.00	0.00	0.00
VA4	Vacarisses	Rb-standard	40	31 (4036)	1795.77	2007.40 ± 16.86	0.47	1580.29 ± 14.25	2.24	0.00	0.00	0.00
927	Badalona	Rb-standard	40	33 (4624)	1441.24	1647.99 ± 12.35	0.45	1232.35 ± 9.83	1.48	0.00	0.00	0.00
SS7	Sant Sadurní d'Anoia	Rb-fused	38	32 (4724)	1121.74	1228.69 ± 10.05	0.82	1017.37 ± 9.19	1.89	6.06	1093.29 ± 25.90	2.69
SS18	Sant Sadurní d'Anoia	Rb-fused	37	30 (4868)	1316.49	1508.59 ± 14.46	0.33	1114.33 ± 11.05	2.02	8.56	1351.74 ± 27.87	0.35
CS10	Castelldefels	Rb-fused	31	30 (4176)	827.59	922.02 ± 13.31	2.46	735.29 ± 9.73	9.29	31.12	818.01 ± 11.37	2.71
955	Castelldefels	Rb-fused	30	33 (3352)	1393.75	1466.19 ± 19.17	2.22	1180.71 ± 16.57	8.10	28.00	1588.20 ± 31.31	0.00
968	Castelldefels	Rb-fused	30	28 (3668)	1386.60	1446.44 ± 18.15	0.42	1153.02 ± 17.10	5.53	33.61	1452.32 ± 15.63	1.54
967	Castelldefels	Rb-fused	29	32 (3923)	1366.19	1475.92 ± 19.09	1.38	1044.35 ± 15.79	4.45	38.02	1335.59 ± 17.47	2.04

Telomere shortening: average telomere length and proportion of short telomeres

The relationship between telomere length and: (i) type of chromosomal arm (q- and p-arms), (ii) diploid number (from 2n = 29 to 2n = 40) and (iii) type of mice (Laboratory-standard, Rb-standard and Rb-fused) was examined using general linear models (GLMs) as implemented in JMP 11th version. Telomere length (expressed as TFUs) was analysed with a GLM using a normal distribution and an identity-link function, with telomere length (TFUs) as the response variable and the type of arm, diploid number and the group of animals as predictor variables. The significance of parameters was determined from the Wald chi-square statistic. Additionally, the relationship between the proportion of short telomeres and: (i) type of chromosomal arm, (ii) diploid number and (iii) the type of mice was analysed using a GLM with a binomial distribution (assigning 1 for telomeres without signal telomeres and 0 for telomeres with signal) and a logit-link function with the number of short telomeres as the response variable, and the type of chromosomal arm, diploid number and group of animals as predictors.

ANOVA test (telomere length) and Kruskal–Wallis (short telomeres) and subsequent post hoc analysis were applied to investigate response variables (type of chromosomal arm and group of animals) with a significant effect over telomere length and/or over the proportion of short telomeres. Additionally, the correlation (Pearson's) between diploid number and telomere length and/or proportion of short telomeres was tested by a post hoc linear regression with telomere length or with the proportion of short telomeres as the response variable and the diploid number as the predictor variable.

Results

Data description

We studied the effect of telomere shortening in the occurrence of Rb fusions in nature by analysing a total of 44 295 telomeres including 11 mice (28–33 metaphases of each cell line; Fig. 1, Table 1). Telomere lengths followed a normal distribution in all cell lines, including HeLa control cell line (Fig. 2).

Telomere length (TFUs) of the laboratory mice (Laboratory-standard) (mean ± SE per population = 1898 ± 8.83 TFUs; n = 7899 telomeres) ranged from 91.81 to 5806.23 TFUs, of the standard mice from the Rb system (Rb-standard) (mean ± SE per population = 1658 ± 6.02 TFUs; n = 14040 telomeres) ranged from 104.4 to 12383 TFUs and of the Rb mice from the Rb system (Rb-fused) (mean ± SE per population = 1216 ± 3.9 TFUs; n = 21800 telomeres) ranged from 18.0 to 9383.23 TFUs. Table 1 summarizes all data

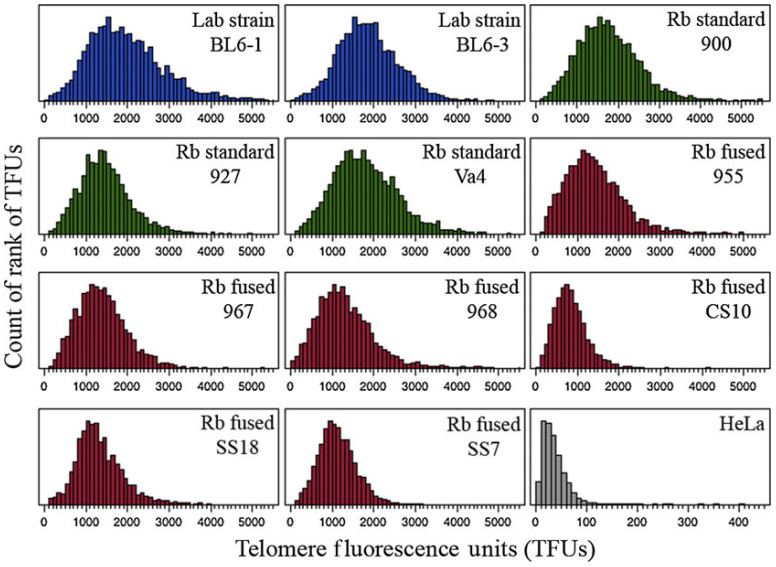


Fig. 2 Telomere length distribution. Distribution of telomeres for each measured TFUs in the specimens analysed. Lab strain (individuals from laboratory strain) (dark blue), Rb-standard (individuals from the Rb population with standard karyotype) (green), Rb-fused (individuals from the Rb population with Rb fusions) (dark red) and HeLa (grey) lines.

(telomeres scored, mean and SE (TFUs), percentage of short arms and percentage of metacentric chromosomes) per specimen and type of chromosomal arm. We found that mouse cells showed wider telomere length ranges and higher mean telomeric lengths than human (HeLa) cells confirming previous observations (Zijlmans *et al.*, 1997; Reig-Viader *et al.*, 2014).

Correlation between q- and p-arm telomeres

It is known that, in wild-type mice, p-arm telomeres are significantly shorter than q-arm telomeres (see Zijlmans *et al.*, 1997). Our analysis confirmed this trend given that telomere lengths were significantly different between q- and p-arms (ANOVA: $F = 1697$; $P < 0.0001$). Our analysis confirmed that acrocentric p-arm telomeres are significantly shorter (mean ± SD: 1310.93 ± 627.85) than their q-arm counterparts (1699.27 ± 775.05) ($t = 54.46$, $P < 0.0001$; Fig. 3a) and significantly similar to the telomere length of q-arms of metacentric chromosomes (1312.93 ± 747.24) ($t = 1.02$, $P = 0.84$; Fig. 3a). However, acrocentric q-arms have significantly higher telomere lengths than metacentric q-arms ($t = 37.58$, $P < 0.0001$; Fig. 3a).

Thus, if p-arm telomeres are significantly shorter than q-arm telomeres, they may be more sensitive to lose telomeric repeats than q-arm telomeres (Blasco *et al.*, 1997). According to this view, the proportion of short telomeres was significantly different between q- and p-arms of acrocentric and metacentric chromosomes (χ^2=286.64, $P < 0.0001$; Fig. 3b). P-arms of acrocentric chromosomes have a higher proportion of short telomeres (3.16%) than q-arms of the metacentrics (2.07%) (Fig. 3b), and both, have higher proportion of short telomeres than q-arms of the acrocentric chromosomes (0.77%) (Fig. 3b).

Q-arm/p-arm telomere length ratio among type of mice (Laboratory-standard, Rb-standard and Rb-fused) was also significantly different (ANOVA: $F = 14.31$, $P < 0.0001$). This ratio was significantly lower in the standard mice from the Rb system (mean ± SD: 1.50 ± 0.94) ($t = 4.74$, $P < 0.0001$) than both laboratory mice (1.62 ± 1.30) and Rb mice from the Rb system (1.59 ± 01.61) ($t = 4.29$, $P < 0.0001$), but not significantly different between laboratory and Rb mice ($t = 1.24$, $P = 0.21$). The decrease of the q-arm/p-arm telomere length ratio in the standard mice from the Rb system is consistent with a significant correlation ($r = 0.156$, $P < 0.001$) between q- and p-arms, in both laboratory mice ($r = 0.114$, $P < 0.0001$) and Rb mice from the Rb system ($r = 0.108$, $P < 0.0001$).

Telomere shortening: average telomere length and proportion of short telomeres

To examine the role of telomeric shortening on the origin of the Rb fusions in the Barcelona Rb system, telomere length and proportion of short telomeres were analysed with GLMs. Table 2 includes Wald chi-square values, degrees of freedom and *P* values.

Overall, we observed that average telomere length was influenced by the type of chromosomal arm (q and p), the group of mice (Laboratory-standard, Rb-standard and Rb-fused) and the diploid number (from 2n = 29 to 2n = 40) (Table 2). In Rb-fused mice, the average telomere length (mean ± SD) (q-arm = 1333.63 ± 604.24 and p-arm = 1036.32 ± 488.40) was significantly shorter than both, the average telomere

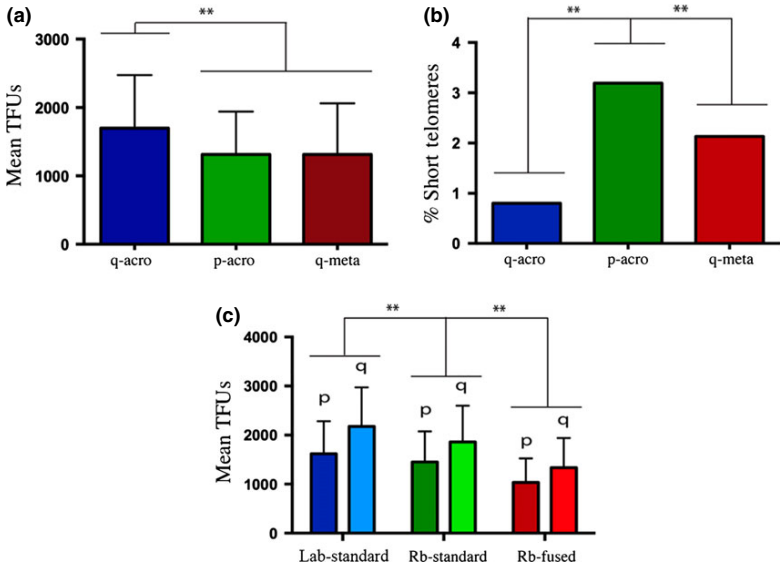


Fig. 3 Telomere shortening analysis. (a) Mean telomere lengths (in TFUs) for each type of chromosomal arms: p-arms of acrocentric chromosomes (p-acro), q-arms of acrocentric chromosomes (q-acro) and q-arms of metacentric chromosomes (q-meta) (Student's *t*-test, ** *P*-value < 0.001). (b) Percentage of short telomeres for each type of chromosomal arms: p-arms of acrocentric chromosomes (p-acro), q-arms of acrocentric chromosomes (q-acro) and q-arms of metacentric chromosomes (q-meta) (Student's *t*-test, ** *P*-value < 0.001). (c) Mean telomere lengths (in TFUs) for each type of chromosomal arms (p- and q-arms) for the three different animal groups compared: laboratory mice (Lab-standard), standard mice from the Rb system (Rb-standard) and Rb mice from the Rb system (Rb-fused) (Student's *t*-test, ** *P*-value < 0.001).

Table 2 Generalized linear model for telomere length and proportion of short telomeres according to diploid number (2n), group of animals (Laboratory-standard, Rb-Standard and Rb-Fused) and chromosome arm (q- and p-arms).

	Telomere length			Short telomeres		
Fixed effects	χ^2	d.f.	*P*	χ^2	d.f.	*P*
2n	11.97	1	0.0005*	113.17	1	< 0.0001*
Group	3390.67	2	< 0.0001*	2.44	2	0.2953
Arm	3351.69	2	< 0.0001*	353.00	2	< 0.0001*

P refers to type II deviance values. * indicates statistical significance.

length of the Rb-standard mice (q-arm = 1859.80 ± 739.83 and p-arm = 1453.19 ± 622.64) (q-arm: $t = 46.55$, $P < 0.0001$; p-arm: $t = 43.90$, $P < 0.0001$; Fig. 3c) and the Laboratory-standard mice (q-arm = 2175.40 ± 797.26 and p-arm = 1617.17 ± 662.99) (q-arm: $t = 62.49$, $P < 0.0001$; p-arm: $t = 51.44$, $P < 0.0001$; Fig. 3c). Moreover, average telomere length of Rb-standard mice was significantly shorter than average telomere length of the Laboratory-standard mice (q-arm: $t = 22.80$, $P < 0.0001$; p-arm: $t = 14.18$, $P < 0.0001$; Fig. 3c). Finally, and only in the case of Rb-fused mice, due to their unique composition of metacentric and acrocentric q-arms, when comparing average telomere length of acrocentric (1333.63 ± 604.24) and metacentric chromosomes in Rb mice (1300.05 ± 747.44), we detected that metacentric q-arms were significantly shorter ($t = 3.38$, $P = 0.0007$).

Moreover, diploid numbers correlated significantly with telomere length of both type of chromosomal arms (q-arm: $r = 0.103$, $P < 0.0001$; p-arm: $r = 0.092$, $P < 0.0001$), that is, shorter average telomeres appeared in mice with lower diploid numbers (with more Rb fusions) (Fig. 4a,b). However, only telomere length of q-arms correlated with number of fusions (q-arm: $r = 0.0006$, $P = 0.02$; p-arm: $r = 0.0001$, $P = 0.35$).

Conversely, the proportion of short telomeres was influenced by both, the type of chromosomal arm and the diploid number but not by the type of mice (see Table 2). The percentage of short telomeres in both chromosomal arms (q- and p-arms) was significantly correlated with the diploid number (q-arm: $r = 0.003$, $P < 0.0001$; p-arm: $r = 0.013$, $P < 0.0001$; Fig. 4c,d) and also with the number of fusions (q-arm: $r = 0.002$, $P < 0.0001$; p-arm: $r = 0.011$, $P < 0.0001$).

Discussion

The role of telomere shortening to explain Rb fusions in natural populations, as well as the importance of the average telomere length vs. the proportion of short

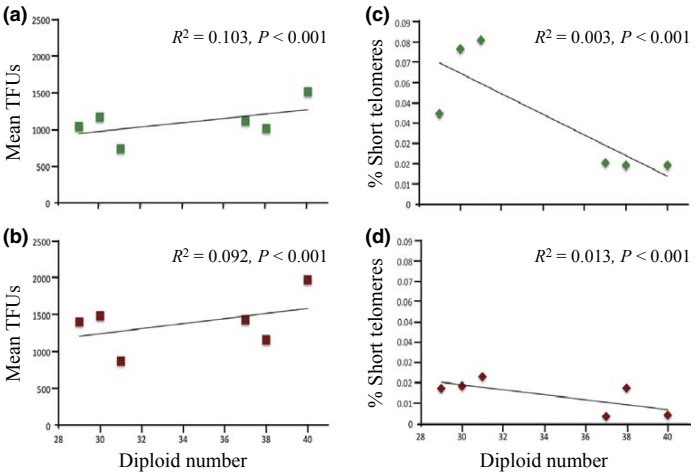


Fig. 4 Karyotype and telomere length. Correlation analysis between diploid number and mean telomere length (expressed as TFUs) for each type of chromosomal arms: (a) q-arms and (b) p-arms. Correlation analysis between diploid numbers expressed percentage of short telomeres of (c) q-arms and (d) p-arms.

telomeres is yet under discussion. It is known that telomere length can vary among species, individuals and even among telomeres within the same cell (Zijlmans *et al.*, 1997; Canela *et al.*, 2007). As the generation of Rb fusions due to telomere shortening has been observed exclusively *in vitro* (i.e. in tissue cultures established from telomerase-null mice) (Blasco *et al.*, 1997), the occurrence of such mechanisms in nature, together with the stabilization of the resultant chromosomes through subsequent generations, needs further validation. Here, we provide the first evidence of telomere shortening, in terms of telomere length and percentage of short telomeres, in a wild house mouse Rb system. We detected that both types of wild mice from the Barcelona Rb system (with Rb fusions and standards) presented, on average, significantly lower telomere lengths than laboratory standard mice, and this correlation held when considering p- and q-arms. Additionally, and more interestingly, the percentage of short telomeres was (in this case, only for the q-arms) negatively correlated to the diploid number, but was influenced by the type of mice, because only mice with Rb fusions showed high proportions of short telomeres.

It is well established that mouse p-arm telomeres are shorter than their q-arm counterparts (Zijlmans *et al.*, 1997). We found that the length of telomeres on q- and p-arms differed among the three groups of mice analysed and that this was influenced by the diploid number. In fact, in wild mice from the Barcelona Rb system, p-arm telomeres were significantly shorter than their q-arm counterparts and presented a higher proportion of short telomeres than their q-arm counterparts. Additionally, and consistently with Nanda *et al.* (1995) study on laboratory mice strains with Rb fusions, we did not detect telomeric signals at the centromeres of the metacentric chromosomes in the analysed specimens. Moreover, we observed that the average telomere length of the metacentric chromosomes was more similar (although significantly higher) to the average length of the q-arms telomeres of the acrocentric chromosomes, than to the average length of their p-arm telomeres. Therefore, these results indicate that Rb fusion formation between p-arms would most probably be triggered by the loss of whole p-arm telomeres.

Previous studies on telomerase-null mice (Blasco *et al.*, 1997; Hemann *et al.*, 2001) showed that chromosomes with short telomeres have a preference for fusion, and based on this finding, they predict that as telomeres shorten, chromosomes would undergo fusions as a way to stabilize chromosomal ends. Consistent with this prediction, we have similarly detected that wild mice with Rb fusions have shorter telomeres (q- and p-arms) than standard mice (from the Rb system and laboratory mice), and the diploid number was negatively correlated with the proportion of short telomeres. Therefore, considering the three different mechanisms that have been proposed to explain the origin of Rb fusions in relation to the presence/absence of telomeric repeats in the resultant chromosome, we suggest that telomere shortening is involved in the formation of Rb fusions in natural populations. Known as the widespread distribution of Rb fusions in the Barcelona polymorphic zone (Medarde *et al.*, 2012 and references therein), Rb fusions would represent an effective way to stabilize the progressive loss of telomeric sequences within the population.

The fact that mice from the Barcelona Rb system, and especially those with Rb fusions, present shorter telomeres when compared to standard laboratory mice is an intriguing result and highlights the importance of focusing on natural populations. Some studies have suggested that telomere length is under genetic control, for example humans (Slagboom *et al.*, 1994; Jeanclos *et al.*, 2000; Vasa-Nicotera *et al.*, 2005) and mice (Zhu *et al.*, 1998; Ding *et al.*, 2004). Zhu *et al.* (1998) showed, based on the telomere length polymorphism

between mice of the interfertile species *Mus spretus* and *M. musculus*, that an unidentified gene located in a region on distal chromosome 2 regulates telomere length in mice. More recently, Ding *et al.* (2004) identified Rtel (a 'regulator of the telomere gene') that is thought to codify a helicase which is implicated in telomere length regulation and prevents genetic instability. We detected that standard mice from the Rb system had shorter telomere lengths in both q- and p-arms but similar proportions of short telomeres than the standard mice from the laboratory strain. The low average length of both types of wild mice from the Barcelona Rb systems (standard and fused) could be explained by the presence and/or dominance of an allelic variant of gene (s) coding for protein(s) involved in telomere length maintenance. In this way, chromosome fusions would take place when telomere length reached the threshold recognized by the DSBs repair machinery.

Apart from telomerase activity, the so-called alternative lengthening of telomeres (ALT) has been reported as an additional telomerase-independent mechanism involved in telomere lengthening and maintenance through homologous recombination (HR) between telomeres (Cesare & Reddel, 2010). Although this mechanism has been mainly described in tumour cell lines (Bryan *et al.*, 1995; Dunham *et al.*, 2000), it has been reported that short telomeres have an increased level of telomere recombination in primary cells (i.e. in the absence of telomerase), suggesting that the ALT mechanism can be present as a way to stabilize telomeric-free ends (Neumann *et al.*, 2013). Recently, we have reported a displacement of recombination events to telomeres in fused metacentric chromosomes of spermatocytes of Rb wild mice from the Barcelona Rb system (Capilla *et al.*, 2014). Such observation, together with the fact that telomeres are significantly shorter in wild mice (standard and fused) from the Barcelona Rb system compared to standard laboratory mice, suggests that the HR could be contributing to telomere length maintenance in the Barcelona Rb system, maybe by the ALT mechanism. Therefore, telomeres would be maintained to avoid telomeric lengths below 15 kbp (which would be close to the critic length that leads telomeres to be recognized as DSBs by the repair machinery). In this way, cells avoid fusions between all acrocentric chromosomes in Rb mice and would delay the generation of fused metacentric chromosomes in standard mice.

In conclusion, we detected that Rb-fused mice present shorter q- and p-arms telomeres and higher percentage of short telomeres (only in p-arms) than the standard mice. However, standard mice from the Barcelona Rb system showed, on average, shorter telomeres, but similar proportion of short telomeres than the standard mice from the B6 laboratory strain. This fact suggests that the number of critically short telomeres, and not the reduction in the telomere length *per se*, is probably the responsible of the occurrence of Rb fusions in wild mice from the Barcelona Rb system.

Acknowledgments

The authors acknowledge T. J. Robinson for insightful comments on earlier versions of the manuscript and to Francisca Garcia and Francisco Cortés from the SCAC (Servei de Cultius Cel.lulars, Universitat Autónoma de Barcelona) for their assistance with the cell lines. We are also grateful to Nuria Medarde for her contribution in the fieldwork. RAS-G was supported by a postdoctoral grant from 'Alianza 4 Universidades', LC is the beneficiary of a FPI predoctoral fellowship (BES-2011-047722), and RRV was supported by a Personal Investigador en Formació (PIF) fellowship of the Universitat Autònoma de Barcelona. This study was partially supported by a grant from the Ministerio de Economia y Competitividad (MINECO, CGL-2010-20170) to ARH. The capture of mice in the wild was supported by a grant from the Ministerio de Economia y Competitividad (MINECO, CGL2010-15243) to JV.

References

Allsopp, R.C., Chang, E., Kashefi-Aazam, M., Rogaev, E., Piatyszek, M.A., Shay, J.W. *et al.* 1995. Telomere shortening is associated with cell division *in vitro* and *in vivo*. *Exp. Cell Res.* **220**: 194–200.

Almeida, H. & Ferreira, M.G. 2013. Spontaneous telomere to telomere fusions occur in unperturbed fission yeast cells. *Nucleic Acids Res.* **41**: 3056–3067.

Aubert, G., Hills, M. & Lansdorp, P.M. 2012. Telomere length measurement-caveats and a critical assessment of the available technologies and tools. *Mutat. Res.* **730**: 59–67.

Auffray, J.-C. 1993. Chromosomal divergence in house mice in the light of palaeontology: a colonization-related event? *Quatern. Int.* **19**: 21–25.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**: 798–801.

Bailey, S.M., Meyne, J., Chen, D.J., Kurimasa, A., Li, G.C., Lehnert, B.E. *et al.* 1999. DNA double-strand break repair proteins are required to cap the ends of mammalian chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 14899–14904.

Blasco, M.A., Lee, H.W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R.A. *et al.* 1997. Telomere shortening and tumor formation by mouse cells lacking telomerase. *Cell* **91**: 25–34.

Bryan, T.M., Englezou, A., Gupta, J.Y., Bacchetti, S. & Reddel, R.R. 1995. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.* **14**: 4240.

Canela, A., Vera, E., Klatt, P. & Blasco, M.A. 2007. High-throughput telomere length quantification by FISH and its application to human population studies. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 5300–5305.

Capilla, L., Medarde, N., Alemany-Schmidt, A., Oliver-Bonet, M., Ventura, J. & Ruiz-Herrera, A. 2014. Genetic

recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and Prdm9 allelic background. *Proc. R. Soc. B* **281**: 20140297.

Cesare, A.J. & Reddel, R.R. 2010. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat. Rev. Genet.* **11**: 319–330.

Ding, H., Schertzer, M., Wu, X., Gertsenstein, M., Selig, S., Kammori, M. *et al.* 2004. Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. *Cell* **117**: 873–886.

Dumas, D. & Britton-Davidian, J. 2002. Chromosomal rearrangements and evolution of recombination: comparison of chiasma distribution patterns in standard and Robertsonian populations of the house mouse. *Genetics* **162**: 1355–1366.

Dunham, M.A., Neumann, A.A., Fasching, C.L. & Reddel, R.R. 2000. Telomere maintenance by recombination in human cells. *Nat. Genet.* **26**: 447–450.

Espejel, S. & Blasco, M.A. 2002. Identification of telomere-dependent "senescence-like" arrest in mouse embryonic fibroblasts. *Exp. Cell Res.* **276**: 242–248.

Faria, R. & Navarro, A. 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**: 660–669.

Farré, M., Micheletti, D. & Ruiz-Herrera, A. 2013. Recombination rates and genomic shuffling in human and chimpanzee–a new twist in the chromosomal speciation theory. *Mol. Biol. Evol.* **30**: 853–864.

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.

Hemann, M.T., Strong, M.A., Hao, L.Y. & Greider, C.W. 2001. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* **107**: 67–77.

Jeanclos, E., Schork, N.J., Kyvik, K.O., Kimura, M., Skurnick, J.H. & Aviv, A. 2000. Telomere length inversely correlates with pulse pressure and is highly familial. *Hypertension* **36**: 195–200.

King, M. 1995. Species evolution: the role of chromosome change. *Syst. Biol.* **44**: 578.

Medarde, N., López-Fuster, M.J., Muñoz-Muñoz, F. & Ventura, J. 2012. Spatio-temporal variation in the structure of a chromosomal polymorphism zone in the house mouse. *Heredity* **109**: 78–89.

Nanda, I., Schneider-Rasp, S., Winking, H. & Schmid, M. 1995. Loss of telomeric sites in the chromosomes of *Mus musculus domesticus* (Rodentia: Muridae) during Robertsonian rearrangements. *Chromosome Res.* **3**: 399–409.

Neumann, A.A., Watson, C.M., Noble, J.R., Pickett, H.A., Tam, P.P. & Reddel, R.R. 2013. Alternative lengthening of telomeres in normal mammalian somatic cells. *Genes Dev.* **27**: 18–23.

O'Sullivan, R.J. & Karlseder, J. 2010. Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **11**: 171–181.

Piálek, J., Hauffe, H.C. & Searle, J.B. 2005. Chromosomal variation in the house mouse. *Biol. J. Linn. Soc.* **84**: 535–563.

Poon, S.S., Martens, U.M., Ward, R.K. & Lansdorp, P.M. 1999. Telomere length measurements using digital fluorescence microscopy. *Cytometry* **36**: 267–278.

Reig-Viader, R., Vila-Cejudo, M., Vitelli, V., Buscà, R., Sabaté, M., Giulotto, E. *et al.* 2014. Telomeric repeat-containing RNA (TERRA) and telomerase are components of telomeres during mammalian gametogenesis. *Biol. Reprod.* **90**: 103.

Rieseberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**: 351–358.

Robertson, W. 1916. Chromosome studies. I. Taxonomic relationships shown in the chromosomes of Tettigidae and Acrididae. V-shaped chromosomes and their significance in Acrididae, Locustidae and Gryllidae: chromosome and variation. *J. Morphol.* **27**: 179–331.

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

Ruiz-Herrera, A., Farré, M., Ponsà, M. & Robinson, T.J. 2010. Selection against Robertsonian fusions involving housekeeping genes in the house mouse: integrating data from gene expression arrays and chromosome evolution. *Chromosome Res.* **18**: 801–808.

Slagboom, P.E., Droog, S. & Boomsma, D.I. 1994. Genetic determination of telomere size in humans: a twin study of three age groups. *Am. J. Hum. Genet.* **55**: 876–882.

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

Vasa-Nicotera, M., Brouilette, S., Mangino, M., Thompson, J.R., Braund, P., Clemitson, J.R. *et al.* 2005. Mapping of a major locus that determines telomere length in humans. *Am. J. Hum. Genet.* **76**: 147–151.

Vera, E. & Blasco, M. 2012. Beyond average: potential for measurement of short telomeres. *Aging* **4**: 379–392.

Wesche, P.L. & Robinson, T.J. 2012. Different patterns of Robertsonian fusion pairing in Bovidae and the house mouse: the relationship between chromosome size and nuclear territories. *Genet. Res.* **94**: 97–111.

White, M.J.D. (1973). *Animal Cytology and Evolution*, 3rd edn (N.). Cambridge University Press, Cambridge.

Zhu, L., Hathcock, K.S., Hande, P., Lansdorp, P.M., Seldin, M.F. & Hodes, R.J. 1998. Telomere length regulation in mice is linked to a novel chromosome locus. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 8648–8653.

Zijlmans, J.M., Martens, U.M., Poon, S.S., Raap, A.K., Tanke, H.J., Ward, R.K. *et al.* 1997. Telomeres in the mouse have large inter-chromosomal variations in the number of T2AG3 repeats. *Proc. Natl. Acad. Sci. U.S.A.* **94**: 7423–7428.

Received 6 October 2014; revised 21 November 2014; accepted 23 November 2014

4.3 Analyzing the role of Rb fusions and *Prdm9* sequence on the meiotic dynamics of the Barcelona Rb system

Since it was initially proposed (Rieseberg 2001; Noor 2001), the suppressed recombination model has been considered as a relevant framework to explain the role of chromosomal rearrangements in speciation. However empirical evidence that support this model is scarce in mammals. Here, we have tested the effect of chromosomal rearrangements (Rb fusions) on meiotic recombination by using as a model the Barcelona Rb system, a house mouse population characterized by a high rate of karyotype diversity due to the presence of Rb fusions in a polymorphic state. Additionally we also have analyzed the gene *Prdm9* in the same population, a genetic factor that has been seen to be involved in the distribution of meiotic recombination.

4.3.1 Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and Prdm9 allelic backgorund

Capilla L, Medarde N, Alemany-Schmidt A, Oliver-Bonet M, Ventura J, Ruiz-Herrera A (2014). **Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and *Prdm9* allelic background**. Proceedings of the Royal Society of Biological Sciences 281:20140297.

Impact factor: 5.29
Q1 in Biology
Q1 in Ecology
Q1 in Evolutionary Biology

PROCEEDINGS OF THE ROYAL SOCIETY B

rspb.royalsocietypublishing.org

Research

Cite this article: Capilla L, Medarde N, Alemany-Schmidt A, Oliver-Bonet M, Ventura J, Ruiz-Herrera A. 2014 Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and *Prdm9* allelic background. *Proc. R. Soc. B* **281**: 20140297. http://dx.doi.org/10.1098/rspb.2014.0297

Received: 5 February 2014
Accepted: 24 April 2014

Subject Areas:
evolution, genetics

Keywords:
house mouse, Robertsonian fusion, speciation, heterochromatinization, recombination, *Prdm9*

Authors for correspondence:
Jacint Ventura
e-mail: jacint.ventura.queija@uab.cat
Aurora Ruiz-Herrera
e-mail: aurora.ruizherrera@uab.cat

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2014.0297 or via http://rspb.royalsocietypublishing.org.

Royal Society Publishing

Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and *Prdm9* allelic background

Laia Capilla[1,2], Nuria Medarde[2], Alexandra Alemany-Schmidt[4], Maria Oliver-Bonet[4], Jacint Ventura[2] and Aurora Ruiz-Herrera[1,3]

[1]Institut de Biotecnologia i Biomedicina (IBB), [2]Departament de Biologia Animal, Biologia Vegetal i Ecologia, and [3]Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Campus UAB 08193, Barcelona, Spain
[4]Unitat d'Investigació, Hospital Universitari Son Espases, Ctra Valldemossa 79, Palma 07010, Spain

Despite the existence of formal models to explain how chromosomal rearrangements can be fixed in a population in the presence of gene flow, few empirical data are available regarding the mechanisms by which genome shuffling contributes to speciation, especially in mammals. In order to shed light on this intriguing evolutionary process, here we present a detailed empirical study that shows how Robertsonian (Rb) fusions alter the chromosomal distribution of recombination events during the formation of the germline in a Rb system of the western house mouse (*Mus musculus domesticus*). Our results indicate that both the total number of meiotic crossovers and the chromosomal distribution of recombination events are reduced in mice with Rb fusions and that this can be related to alterations in epigenetic signatures for heterochromatinization. Furthermore, we detected novel house mouse *Prdm9* allelic variants in the Rb system. Remarkably, mean recombination rates were positively correlated with a decrease in the number of ZnF domains in the *Prdm9* gene. The suggestion that recombination can be modulated by both chromosomal reorganizations and genetic determinants that control the formation of double-stranded breaks during meiosis opens new avenues for understanding the role of recombination in chromosomal speciation.

1. Introduction

The role of chromosomal reorganizations in speciation has been a long-standing question in biology. Understanding the genetic and mechanistic basis of these processes will provide insights into how biodiversity originates and is maintained. Compelling evidence supports that chromosomal rearrangements may reduce gene flow and therefore potentially contribute to speciation by the suppression of recombination [1–3]. Under this 'suppressed recombination' model, chromosome reorganizations in heterokaryotypes have a minimal influence on fitness, but rather affect recombination thus contributing to a reduction of gene flow within these genomic regions and the consequent accumulation of genetic incompatibilities [3]. Data supporting recombination suppression by inversions have been reported in different model organisms (see [4] and references therein), whereas studies regarding the effects of fusions have been restricted to the western house mouse, *Mus musculus domesticus* [5–14] and the common shrew, *Sorex araneus* [15,16]. The general view is that chromosomal reorganizations disturb the chromosomal distribution of recombinational events altering, in the long term, the final outcome of crossovers (COs). If this phenomenon was perpetuated within the population for long periods of time, gene flow within the reorganized regions would be reduced, thus

eventually contributing to the establishment of genetic incompatibilities and subsequent progress towards complete reproductive isolation [3]. Despite recent progress in the field, the mechanisms and the genetic basis underlying the process remain elusive. In this regard, recent attention has focused on the *Prdm9* gene, the only known speciation-associated gene described for mammals. The protein encoded by this gene, the PR domain zinc finger 9 (PRDM9) [17], is a meiotic-specific histone (H3) methyltransferase with a C-terminal tandem repeat zinc finger (ZnF) domain that accumulates at recombination sites through its recognition of a species-specific and highly mutagenic repetitive DNA motif [17–19]. In humans, the high nucleotide variation detected at this locus suggests that the protein (together with the repetitive DNA motif that it recognizes) is highly mutable enabling binding to new motifs as soon as they are formed [20,21]. Studies in laboratory strains of mice, moreover, have revealed that the PRDM9 protein could be directly involved in the recruitment of the recombination initiation machinery during meiosis [22]. It has also been reported that variations in the *Prdm9* sequence affect the positioning of meiotic double-stranded breaks (DSBs), in the same way that the number and sequence of ZnF could determine the strength and specificity of DNA binding [19,22].

Despite the existence of formal models to explain how chromosomal rearrangements could be fixed in two parapatrically distributed populations under divergent selection, and by which mechanisms these may contribute to speciation [23], few empirical data are available, especially in mammals (see [4] and references therein). Therefore, the appreciation of how meiotic recombination is controlled during meiosis and how chromosomal reorganization affects the process is of critical importance for our understanding of speciation mechanisms. In this context, natural populations represent an excellent scenario for determining the genetic and mechanistic factors underlying the origin of meiotic disturbances triggered by chromosomal reorganizations. The standard karyotype of *M. musculus domesticus* consists of all-acrocentric chromosomes ($2n = 40$). However, a wide range of diploid numbers (from 22 to 40) have been described in natural populations resulting from the occurrence of Robertsonian (Rb) fusions and/or whole arm reciprocal translocations [24,25]. According to Piálek *et al.* [25], a Rb system is defined as a series of house mouse populations from a restricted geographical area that share a set of metacentrics with an apparently common evolutionary origin. In this context, a total of 97 Rb systems distributed across Europe, and the Mediterranean basin has been formally recognized [25]. Among them, the Barcelona Rb system is found in the provinces of Barcelona, Tarragona and Lleida (Spain), extending over approximately 5000 km^2 (figure 1*a*). Diploid numbers ($2n$) range from 27 to 39 chromosomes, with the lowest $2n$ values observed at sites located approximately 30 km west of the city of Barcelona ([26] and references therein). This Rb system is characterized by a high level of chromosomal polymorphism, with seven different Rb chromosomes (Rb (3.8), (4.14), (5.15), (6.10), (7.17), (9.11) and (12.13)) showing non-geographically coincident (staggered) clines ([26] and references therein). Owing to the absence of a Rb race, as defined by Hausser *et al.* [27], this system has been considered to be a Rb polymorphism zone rather than be a typical hybrid zone ([26] and references therein). Although the chromosomal composition and structure of the Barcelona Rb system are known [26,28], the meiotic dynamics and genetic recombination of the mice that comprise it remain elusive. Here, we have studied the recombination features of this Rb system with two specific aims: (i) to analyse the effect of Rb fusions on both recombination rate and its chromosomal distribution, and (ii) to determine the genetic and mechanistic factors that are shaping the process.

2. Results

(a) Recombination events are reduced in Robertsonian mice

We studied the effect of Rb fusions on meiotic recombination by analysing the variation in the number of autosomal MLH1 *foci* (marker of COs) detected at pachynema in males (figure 1*a–c* and the electronic supplementary material, table S1). The average of MLH1 *foci* per cell observed ranged from 20.16 ($\pm$1.18) to 21.65 ($\pm$2.43) in wild mice with standard (St) karyotype ($2n = 40$), and from 18.13 ($\pm$1.78) to 20.57 ($\pm$1.90) in Rb mice (figure 1*d* and the electronic supplementary material, table S1). Wild St mice presented, on average, significantly less COs (20.88 ± 1.83) per cell than did mice derived from laboratory strains (21.32 ± 2.12; Mann–Whitney *U*-test, $p \leq 0.05$; figure 1*d* and the electronic supplementary material, table S1). More importantly, mice with Rb fusions showed significantly smaller mean number of COs than wild St mice (Kruskal–Wallis test, $p \leq 0.001$; figure 1*d* and the electronic supplementary material, table S1). Among Rb mice, we could distinguish two different groups (Mann–Whitney *U*-test, $p \leq 0.001$): mice with high diploid numbers ($2n = 39–37$, 19.93 ± 1.77 COs per cell) and mice with low diploid numbers ($2n = 32–28$, 19.39 ± 1.68 COs per cell). These results indicate that Rb mice have a significant decrease in the overall number of recombination events per bivalent as a result of the presence of fused chromosomes.

In order to disentangle the effect of different factors on the decreased number of COs observed in Rb mice, we evaluated correlations between diploid number and the degree of heterozygosity in terms of rearrangements (heterokaryotypes). Owing to the absence of a fixed metacentric race and the high level of chromosomal polymorphisms that characterizes this Rb system, we pooled Rb compositions in three categories: one, two and three Rb fusions in heterozygote state. When considering both St and Rb animals, we found a positive correlation between the mean number of COs and the diploid number (Spearman correlation test $\rho = 0.757$, $p \leq 0.001$; electronic supplementary material, figure S1); however, when applying the same correlation only among the Rb specimens, this was not significant (Spearman correlation tests $\rho = 0.236$, $p = 0.315$). Additionally, the comparison of the mean number of COs in Rb mice revealed that recombination was not affected by the degree of heterozygosity (Kruskal–Wallis test, $p = 0.345$). These results suggest that additional factors (other than diploid number reduction and the degree of heterozygosity) play a role in the final outcome of COs in this Rb system.

Considering the low mean values of COs observed in Rb mice, we investigated whether a similar pattern was reflected by the proteins implicated in the repair of DSBs in the early stages of prophase I (early pachynema). Meiotic recombination is initiated by DSBs generated by the protein Spo11

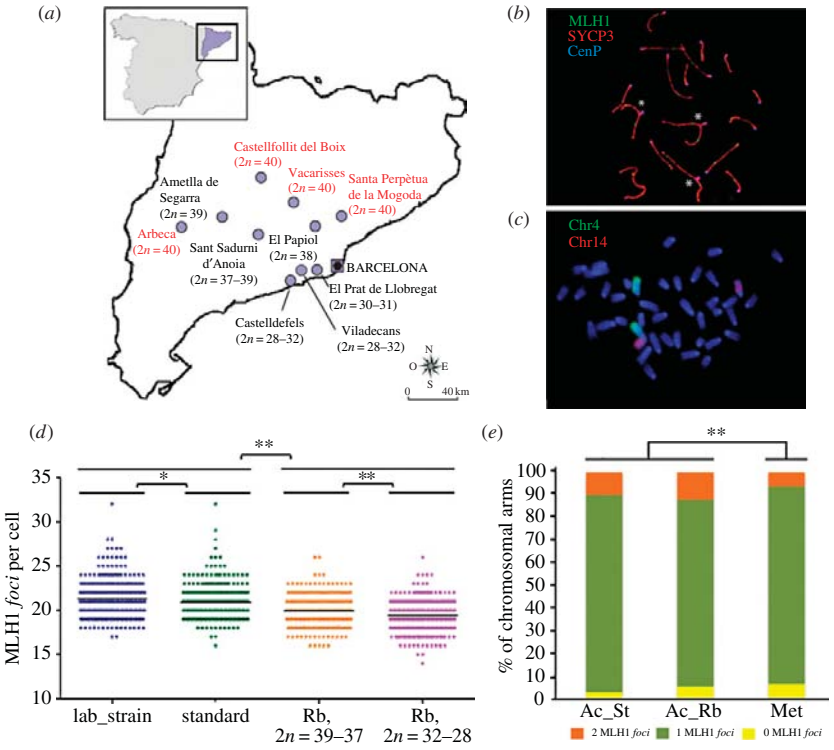


Figure 1. Genetic recombination diversity in the Barcelona chromosomal polymorphism zone. (*a*) Map showing the sampling distribution. Diploid numbers are shown for each locality (see the electronic supplementary material, table S1). Locations with standard and Rb individuals are indicated in red and black, respectively. (*b*) Example of immunolocalization of meiotic recombination events in mouse spermatocytes at pachynema from Rb5 ($2n = 37$). MLH1 *foci* are depicted in green, centromeres in blue and the synaptomenal complexes in red. Asterisks indicate trivalent structures. (*c*) Fluorescence *in situ* hybridization on a metaphase spread from Rb3 ($2n = 38$) with the chromosomes involved in the Rb fusion painted in green (chromosome 4) and red (chromosome 14). (*d*) Distribution of the mean numbers of MLH1 *foci* per cell observed in the specimens analysed ($n = 34$): lab_strain, standard mice from laboratory strain ($n = 310$ cells); standard, wild standard mice ($n = 349$ cells); Rb, $2n = 39-37$, Rb mice with diploid numbers between $2n = 37-39$ ($n = 270$ cells); and Rb, $2n = 32-28$, Rb mice with diploid numbers between $2n = 28-32$ ($n = 333$ cells). Asterisks indicate statistical significance (Mann–Whitney *U*-test or Kruskal–Wallis test; **p*-value ≤ 0.05, ***p*-value ≤ 0.001). (*e*) Percentage of chromosomal arms showing 0, 1 or 2 MLH1 *foci*. Three groups are differentiated: Ac_St, all acrocentric chromosomes belonging to standard wild mice ($n = 1197$); Ac_Rb, all acrocentric chromosomes from Rb mice ($n = 2211$); and Met, all chromosomal arms involved in Rb fusions ($n = 2056$). Asterisk indicates statistical significance (Kruskal–Wallis test, ***p*-value ≤ 0.001).

[29]. The replication protein A (RPA protein) associates with single-stranded DNA following DSBs formation and subsequently accumulates at the DSB sites [30]. Therefore, by analysing the number of RPA *foci* in early pachynema, we can determine the progression of DSBs as early recombinational nodules. If the final outcome of CO events observed in Rb mice is directly related to the initial DSBs formed in early stages of prophase I, then we would expect to find a reduced number of RPA *foci* when compared with St mice. We found that the mean number of RPA *foci* per cell at early pachynema was similar in wild St (67.49 ± 18.55) and Rb mice (64.83 ± 17.40; Mann–Whitney test, $p = 0.536$). However, wild St and Rb mice differed significantly with respect to laboratory mice (83.32 ± 19.92; Kruskal–Wallis test, $p \leq 0.001$; electronic supplementary material, figure S2), suggesting that the difference between wild and laboratory mice is probably related with a reduction in the initial number of DSBs at early meiosis (exemplified here as RPA *foci*, representative of early nodules).

(b) Robertsonian fusions alter the chromosomal distribution of recombination events

Moved by this striking pattern, we tested whether the overall decrease in recombination events observed was as a result of a reduction of COs in reorganized chromosomes. We analysed the percentage of chromosomal arms showing 0, 1 or 2 MLH1 *foci* considering three different groups based on the state of the chromosome (acrocentric, Ac, or metacentric, Met) and the type of specimen (St or Rb): (i) chromosomal arms involved in Rb fusions in Rb mice (Met), (ii) chromosomal arms not involved in Rb fusions belonging to wild St mice (Ac_St), and (iii) chromosomal arms not involved in Rb fusions in Rb mice (Ac_Rb; figure 1*e* and the electronic supplementary material, table S2). Our results show that when chromosomes are involved in Rb fusions (Met) the frequency of chromosomal arms with two COs decreases significantly when compared with acrocentric chromosomes (Kruskal–Wallis test, $p \leq 0.001$; figure 1*e* and the electronic supplementary material,

table S2). More importantly, the frequency of chromosomal arms with the absence of COs is not altered in the three groups (Kruskal–Wallis test, $p = 0.071$; figure 1*e* and the electronic supplementary material, table S2).

In order to analyse the redistribution of COs in Rb chromosomes in greater detail, we studied the chromosomal-specific distribution of MLH1 *foci* by applying sequential immunostaining and fluorescence *in situ* hybridization (FISH) with chromosome-specific bacterial artificial chromosomes (BAC) clones (electronic supplementary material, table S3 and figure 2*a*). This allowed for the identification of each of the chromosomes implicated in the Rb fusions in different mice, enabling us to test whether the overall reduced number of COs detected in Rb mice was due to the chromosomes involved in Rb fusions. We focused our efforts on the chromosomes most frequently involved in Rb fusions described in the area [26] by analysing the relative positions of MLH1 *foci* along each synaptonemal complex (SC; figure 2*a–h*). These were chromosomes 4, 9, 11, 12, 13 and 14. First, we established chromosome-specific recombination maps in acrocentric forms; these data serve as controls (figure 2*b–h* and the electronic supplementary material, figure S3). The COs distribution in acrocentrics was in accordance with the pattern previously reported for mice [31]; that is, a bimodal distribution of COs in long- and medium-sized chromosomal arms (i.e. chromosomes 4, 9, 11 and 12) and a telomeric distribution in short chromosomes (i.e. chromosomes 13 and 14; figure 2*b–h* and the electronic supplementary material, figure S3). However, and more importantly, we observed a trend in Rb chromosomes, that is, the distribution of CO events was displaced towards the telomeric regions (figure 2*b–h* and the electronic supplementary material, figure S3). In fact, we observed a significantly reduction of COs in the proximal area of the SCs (from the centromere to 30% of SC length) in Rb chromosomes, either in homozygosis (0.99 ± 2.10) or in heterozygosis (1.51 ± 2.67) when compared with acrocentric (3.70 ± 3.33; Kruskal–Wallis test, $p \leq 0.001$). These results clearly show that when chromosomal arms are implicated in the Rb fusions, there is a reduction in the number of COs, displacing the recombination event towards the telomeric regions of the chromosomes, mirroring previous observations in mice [5,12,13].

(c) Robertsonian fusions alter the chromosomal distribution of H3K9me3 and γH2AX

We detected that the number of early recombinational nodules (RPA *foci*) was not altered in Rb mice when compared with wild St mice (electronic supplementary material, figure S2*a–c*). Moreover, the chromosomal distribution of RPA *foci* was not significantly different between acrocentric and metacentric chromosomes (electronic supplementary material, figure S2*d*), indicating that disturbances in the final outcome of COs occur during the resolution of early recombination nodules into COs, and not during the formation of DSBs. Therefore, we tested whether additional mechanistic factors (i.e. synapsis alterations and heterochromatinization) were altering the chromosomal distribution of COs in reorganized chromosomes.

We analysed the morphology of trivalents and sex chromosomes in order to detect pairing disturbances, because early studies in mice already noted the presence of chromosome unpairing in trivalents [32]. The γH2AX protein recognizes and localizes at DSBs, working as a marker for gene inactivation of asynapsed regions [33,34]. The sex body showed positive signal for γH2AX in all specimens analysed (electronic supplementary material, figure S4*a*,*b*) indicating a normal progression of the meiotic silencing of asynapsed chromatin that characterizes mammalian sex chromosomes [35]. However, we detected significantly higher numbers of cells with sex chromosomes totally asynapsed in Rb mice when compared with St animals (Fisher's test, $p \leq 0.05$; electronic supplementary material, figure S4*e–g*). Regarding autosomes, when analysing the pairing dynamics of trivalents, we identified two different configurations: (i) closed trivalents and (ii) open trivalents (the electronic supplementary material, figure S4*e*). The frequency of spermatocytes detected with open (asynapsed) trivalents varied depending on the specimen analysed and these ranged from 4% to 21% of cells. As a general trend, open trivalents showed positive γH2AX labelling (electronic supplementary material, figure S4*d*,*e*). However, we also detected an abnormal pattern for γH2AX along chromosomes in specimens with Rb fusions, with γH2AX localizing along the SC in synapsed regions in both bivalents and trivalents in 48.6–52% of pachynema analysed (electronic supplementary material, figure S4*f*). This pattern was significantly less frequent in wild St mice (Fisher's test, $p \leq 0.001$), and especially rare in the laboratory strain (only 10.9% of the cells analysed; Fisher's test $p \leq 0.001$; electronic supplementary material, figure S4*f*). These results indicate that the persistence of regions with non-repaired DSBs through pachynema in animals with Rb fusions does not compromise the correct formation of the sex body, allowing the progression through prophase I and thus producing germ cells.

We also tested whether CO reduction in Rb mice was owing to a displacement towards telomeric regions triggered by an expansion centromeric interference effect as it has been previously postulated, although not experimentally tested [5]. In doing so, we analysed the distribution of the histone H3 lysine 9 tri-methylated (H3K9me3), an epigenetic signal for constitutive heterochromatin [36], along the SCs in several mice with different chromosomal configurations: $2n = 29$ (11 Rb fusions; two specimens), $2n = 38$ (two Rb fusions; one specimen), wild $2n = 40$ (two specimens) and a laboratory mouse $2n = 40$. In all mice analysed, H3K9me3 signals were located at autosomal centromeric regions in both acrocentric and metacentric chromosomes (figure 3*a*,*b*). However, the distribution pattern of H3K9me3 signals around the centromere differed between acrocentric and metacentric chromosomes (figure 3*c*). In fact, we identified two different configurations: (i) signals projected outside the SCs and only restricted to the centromeres, and (ii) H3K9me3 signals overlapping both the centromere and the pericentromeric area of the SCs. We detected that in the majority (90%) of the acrocentric chromosomes analysed, H3K9me3 signals did not overlap the SC. However, this proportion decreased significantly (Fisher's test, p-value ≤ 0.001) in chromosomes implicated in Rb fusions (70–60%; figure 3*d*). Moreover, the area (expressed in μm of SC length) occupied by the H3K9me3 signals from the centromere overlapping the SC in trivalents was, on average, significantly larger than in acrocentric chromosomes (Kruskal–Wallis test, p-value ≤ 0.001; figure 3*e*).

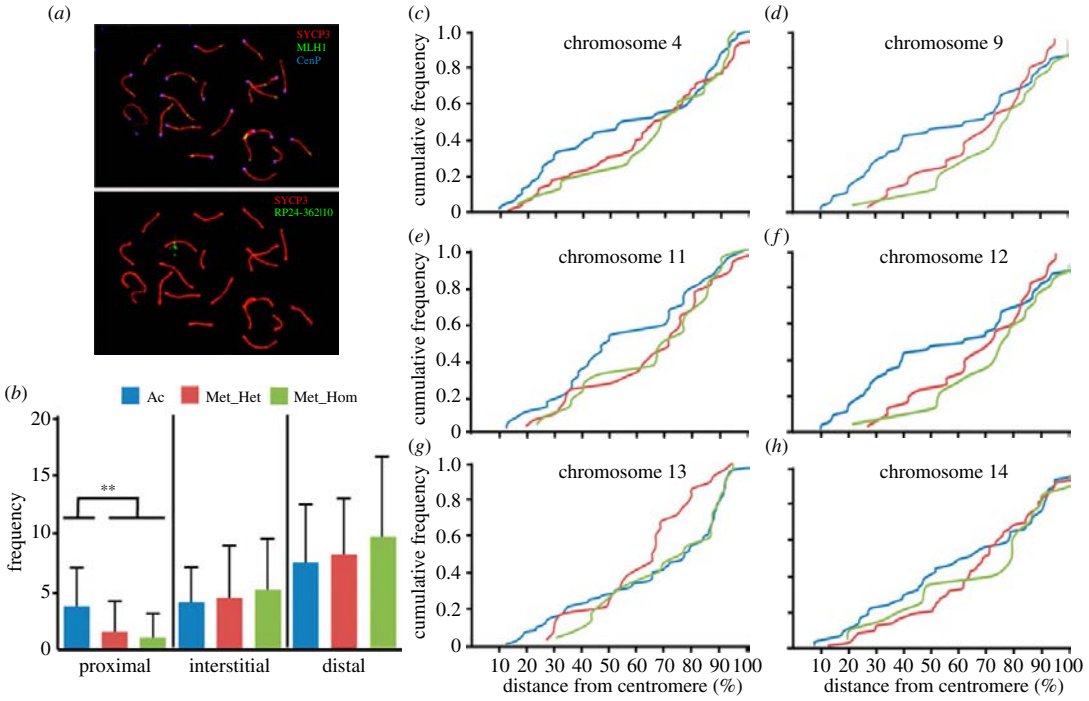


Figure 2. Cumulative frequencies of MLH1 *foci* in chromosomes implicated in Rb fusions. (*a*) Sequential image of a mouse spermatocyte at pachynema depicting a triple immunostaining with SYCP3 (red), MLH1 (green) and centromeres (blue) (upper panel) and fluorescent *in situ* hybridization (FISH; bottom panel) with a BAC probe corresponding to mouse chromosome 11 (RP23–362I10) in green. (*b*) Representation of the MLH1 *foci* frequencies obtained after merging data for all chromosomes analysed (no. 4, no. 9, no. 11, no. 12, no. 13 and no. 14). MLH1 *foci* frequencies along the SCs are shown for three regions depending on the relative distance from the centromere: (i) proximal, from the centromere to 30% of the SC; (ii) interstitial, between 30% and 70% of SC; and (iii) distal, from 70% to telomeric region (Kruskal–Wallis test; $^{**}p$-value ≤ 0.001). Three different groups were considered: Ac, chromosomes in acrocentric form; Met_Hom, chromosomes involved in Rb fusions in homozygosis; and Met_Het, chromosomes involved in Rb fusions in heterozygosis. (*c–h*) Cumulative frequency plots representing the CO distribution of specific mouse chromosomes. In all instances, blue lines indicate chromosomes in the acrocentric form ($n = 38$ cells analysed for no. 4, $n = 25$ for no. 9, $n = 36$ for no. 11, $n = 31$ for no. 12, $n = 71$ for no. 13 and $n = 46$ for no. 14). Green lines represent the distribution when chromosomes are involved in Rb fusions in homozygosis ($n = 21$ cells analysed for no. 4, $n = 22$ for no. 9 $n = 30$ for no. 11, $n = 21$ for no. 12, $n = 20$ for no. 13 and $n = 36$ for no. 14) and red lines when in heterozygosis ($n = 86$ cells analysed for no. 4, $n = 22$ for no. 9, $n = 20$ for no. 11, $n = 28$ for no. 12, $n = 26$ for no. 13, $n = 73$ for no. 14).

(d) *Prdm9* allelic background and recombination rates

To directly assay the genetic basis underlying the alterations in the number and distribution of COs observed in Rb mice, we screened our sample for new *Prdm9* allelic variants that might account for such diversity. *Prdm9* plays a key role in determining the patterning of recombination events [22]; therefore, our working hypothesis was to consider that the observed diversity in recombination rates (figure 1 and the electronic supplementary material, table S1) was related to the *Prdm9* genetic background. With this in mind, we sequenced the *Prmd9* exon 12 containing the Zn finger domain from Zn repeat +3 towards the C-terminal domain in 27 mice from our sample (electronic supplementary material, table S1). We detected that 20 specimens had alleles of the same length, whereas the remaining specimens had alleles of two different lengths (electronic supplementary material, table S1 and figure S5). Our genetic screening detected five *Prdm9* allelic variants present in the study area, which differed both in nucleotide sequences and the number of Zn finger repeats from previous studies [17,37]. These were referred to as *10A*, *10B*, *11B*, *12B* and *12C* (electronic supplementary material, figures S6 and S7). The alleles *12B* and *12C* shared the same number of ZnF domains as the alleles previously described in *M. musculus domesticus* [17,37], but differed in the DNA sequence of the ZnF domains (electronic supplementary material, figures S7 and S8). The allele *11B* shared the same number of ZnF domains as previously described in *Mus musculus castaneus* [37], but, as with the above, differed in the DNA sequence of ZnF domains (electronic supplementary material, figure S7). More importantly, we detected two alleles with 10 ZnF repeats (allele *10A* and *10B*), which are specific for the Barcelona chromosomal polymorphism zone (electronic supplementary material, figures S6–S8). The analysis of the nucleotide and amino acid sequence revealed that the highest replacement rates for all alleles were detected at positions −1, +3 and +6, all regions that were previously described as being highly polymorphic (electronic supplementary material, figures S7 and S8). These positions correspond to the amino acids that recognize the DNA repeat sequence-specific for the Prdm9 protein in the mouse.

Allele frequency and distribution varied among specimens and among localities (electronic supplementary material,

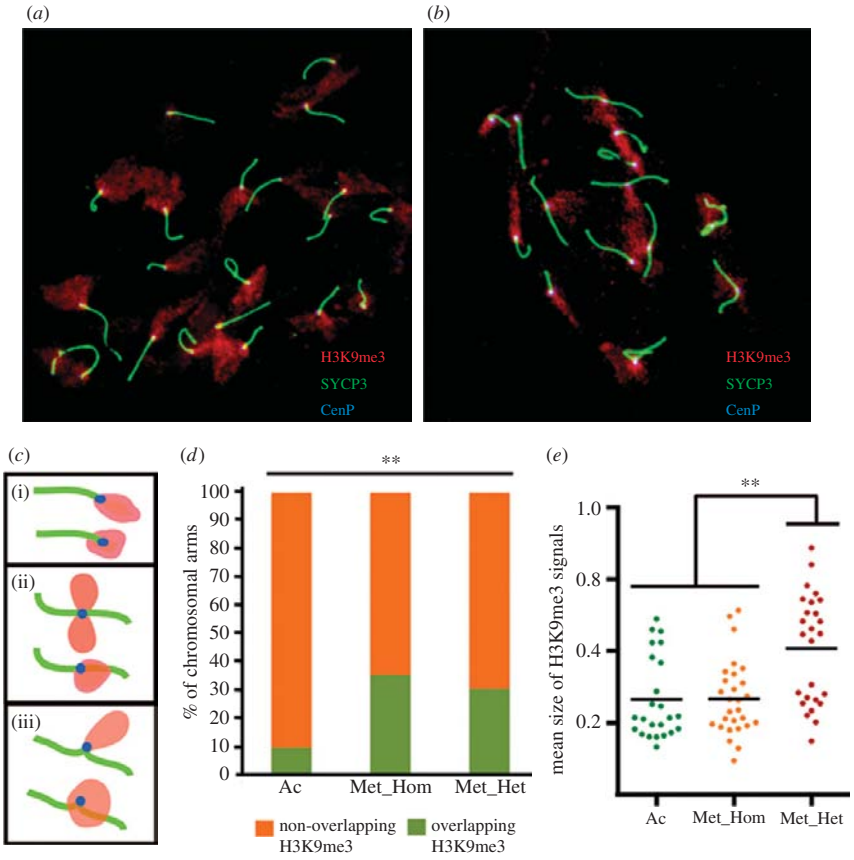


Figure 3. Chromosomal distribution of H3K9me3 signals. (*a*,*b*) Examples of mouse spermatocytes from standard (*a*) and Rb (*b*) mice displaying the H3K9me3 distribution (red) with SC (green) and centromeres (blue). (*c*) Representation of the different H3K9me3 signal patterns observed in the sample: (i) acrocentric chromosomes, (ii) heterozygote metacentric, and (iii) homozygote metacentric. In all cases, H3K9me3 signals (red) were found in two different situations: overlapping and not-overlapping the SC. (*d*) Frequency of chromosomal arms with H3K9me3 signal overlapping (green) or non-overlapping (orange) the SC in acrocentric (Ac, $n = 392$) homozygote metacentric (Met_Hom, $n = 172$) and heterozygote metacentric chromosomal arms (Met_Het, $n = 71$). (*e*) Distribution of the mean size (expressed in μm of SC length) of the H3K9me3 signals overlapping the SC measured in the centromeric area ($n = 78$) in the three types of chromosomal arms described in (*c*,*d*). Asterisks indicate statistical significance (Fisher's test, $^{**}p$-value ≤ 0.001; Kruskal–Wallis test, $^{**}p$-value ≤ 0.001).

figure S6*a*,*b*). Allele *10A* was the most frequently observed (77.77%) in our sample followed by *12B* (14.81%), *12C* (3.70%) and to a lesser extent *11B* (1.85%) and *10B* (1.85%; electronic supplementary material, figure S6*a*). The highest number of different alleles was observed within St wild mice that surrounded the localities containing Rb mice (electronic supplementary material, figure S6*b* and table S1). Remarkably, these Rb mice were genetically more homogeneous, being mostly homozygous for allele *10A* (electronic supplementary material, figure S6*b* and table S1).

We additionally investigated whether *Prdm9* genetic background influences recombination, as has been previously shown in humans [38]. As it has been suggested that low numbers of ZnF repeats are correlated with lower recombination rates [38], three different groups were considered: mice carrying two common 10 alleles (10/10), mice carrying one non-10 allele (10/N) and mice carrying two non-10 alleles (N/N). The comparison of CO frequency (mean number of MLH1 *foci*/cell) in specimens with 10/10, 10/N and N/N genotypes revealed that those carrying the 10 alleles in homozygosis state showed, on average, a significantly lower number of COs than N/10 and N/N mice (Kruskal–Wallis test, p-value ≤ 0.001; electronic supplementary material, figure S6*c*). Moreover, we applied different correlation analysis between *Prdm9* allelic diversity, diploid number and mean number of MLH1 *foci* per cell. We detected a highly similar correlation value between these three factors: $\rho = 0.56$ between total number of PRDM9 ZnF domains and recombination rate (p-value < 0.05), $\rho = 0.58$ between total number of PRDM9 ZnF domains and diploid number (p-value < 0.05), and $\rho = 0.75$ between diploid number and recombination rate (p-value < 0.001), suggesting that the three (mechanistic factors, genetic factors and recombination rates) are somehow related.

3. Discussion

In an evolutionary context, our study represents a detailed empirical demonstration that Rb fusions affect meiotic recombination and that this can be related to alterations in epigenetic signatures for heterochromatinization. Although some analyses performed were based on a limited number of individuals, this is, to our knowledge, the first meiotic study on

a Rb polymorphism zone. This has allowed us to detect important trends. First, we found that Rb males have significantly lower recombination rates that wild St mice, despite the variability observed in diploid numbers. Second, our results suggest that Rb fusions indeed have an effect on the chromosomal distribution of COs, altering their chromosomal distribution. The average number of recombination events is reduced in chromosomal arms involved in Rb fusions when compared with acrocentrics, reflecting a reduction in the percentage of chromosomal arms with two COs in Rb animals. These data add to preliminary observations that have reported a reduction in chiasmata number (in latter stages of meiosis; i.e. metaphase II) and MLH1 *foci* in different house mice populations with Rb fusions [5,6,11–13]. Our approach, however, provided a more detailed and chromosomal-specific analysis of CO redistribution. In fact, we observed that the distribution of COs along chromosomal arms that occur in acrocentrics is altered in Rb chromosomes; recombinational events closer to the centromeric region are frequently lost and COs tend to be more terminal in metacentrics. It is well known that meiotic recombination is repressed close to the centromere, a pattern that has been conserved in all eukaryote species. This is because COs occurring close to centromere interfere with normal chromatid segregation, inducing aneuploidy [39]. Although the molecular mechanisms underlying centromere interference are still largely unknown, pericentric heterochromatin formation (here exemplified by H3K9me3 signals) is considered crucial for the process [40]. In fact, the epigenetic status of the chromatin is, in general, important for recombination. Meiotic DSBs tend to occur in open and highly transcribed euchromatic regions [19], whereas DNA methylation suppresses CO formation. Recent studies in plants have suggested that DNA methylation can affect the distribution of recombination events *in cis*, this alteration being chromosomal-dependent [41]. This fact has important implications for our observations given that chromosomes involved in the Rb fusions locally affect the overall reduction of COs detected in Rb mice. We observed larger and more expanded H3K9me3 signals over the SC in trivalents, which are characterized by the presence of three centromeres non-completely aligned. These results, together with the presence of altered γH2AX pattern along the SC in synapsed regions in both bivalents and trivalents in Rb mice, suggest that DSBs are not properly repaired, especially in trivalent structures, affecting the final outcome of COs.

We also detected that the formation of the sex body at pachynema is not compromised in the Rb males by either the presence of asynapsed regions or asynapsed sex chromosomes. But more importantly, the number of chromosomal arms with no CO is not significantly altered when comparing acrocentric and metacentric chromosomes in Rb animals. COs are highly regulated in mammals to ensure the proper disjunction of homologous chromosomes during meiosis [42] and mammalian species normally present (on average) one CO per chromosomal arm [43,44]. Our observations have important implications for CO homeostasis [42], indicating that cells carrying Rb fusions modulate the final CO outcome without losing the obligatory one chiasmata per arm necessary to allow even chromosomal segregation, thus not compromising the viability of germ cells. Therefore, and despite the presence of unrepaired DNA regions in both open trivalents and synapsed autosomes at pachynema in Rb mice, the sex body is probably established and cells escape the pachytene checkpoint [34]. Given the widespread distribution of Rb animals in the Barcelona chromosomal polymorphic zone, the low recombination rates observed and the presence of asynapsed regions in the trivalents probably would have a mild negative effect on fertility, mirroring previous observations [34]. This interpretation is in line with analyses performed on mice from the Barcelona Rb system [26,45,46], which suggest that Rb fusions have a reduced effect on mice fertility, thus permitting the de novo occurrence of chromosomal rearrangements within this zone.

Together with the CO mechanistic disturbances observed, the *Prdm9* allelic diversity has been revealed as an additional factor modulating genetic recombination. The PRDM9 protein is directly involved in the recruitment of the recombination initiation machinery during meiosis [22]. Previous studies have described the presence of four different haplotypes in mice with differences in the number of Zn fingers: 9-Zn fingers, 11-Zn fingers, 12-Zn fingers, 13-Zn fingers and 14-Zn fingers [17,19,37], and our data add new *Prdm9* allelic variants to the picture, confirming the high *Prdm9* sequence variability in nature. More importantly, our results suggest that the number and sequence of Zn fingers might influence meiotic recombination outcome, most probably by regulating strength and specificity of DNA binding, thus in agreement with what has been reported in humans and mice [17,19,38,47]. Mice carrying the 10 allele showed, on average, a significantly lower number of COs than mice with longer Zn finger repeats. In fact, the allele *10A* was the most frequently observed in the Barcelona chromosomal polymorphism zone (16 of 18 were homozygote for *10A*). Despite this, we still could distinguish two different groups among Rb mice in terms of MLH1 mean values, irrespective of their *Prmd9* allelic composition: mice with high ($2n = 37–39$) and low diploid numbers ($2n = 28–32$). The fact that both groups did not differ in genetic background (i.e. homozygote for *10A*) suggests that additional factors (rather than genetic ones) could explain such differences. Further analysis of the population structure of the Barcelona Rb system would reveal novel *Prdm9* allelic variants, as well as possible demographic or stochastic effects underling its allelic distribution and the genetic modulation of recombination. In fact, our analysis of RPA *foci* (representative of early nodules) indicated that both wild St and Rb mice (carrying the 10 allele) presented a similar pattern in terms of overall numbers of RPA *foci* per cell detected, but different with respect to the laboratory mouse strain, suggesting that mice from the Barcelona chromosomal polymorphic zone have a unique genetic background that is affecting recombination.

Which are the evolutionary implications of our results? Here, we demonstrate that the average number of recombination events is indeed reduced in chromosomal arms involved in the Rb fusions when compared with acrocentric chromosomes, especially in proximal chromosomal regions, and that this can be related to alterations in epigenetic signatures for heterochromatinization. In mammals, recent studies have detected reduced recombination rates and gene flow within reorganized regions [8,14,16]. This has been the case of inversions, where recombination is specially reduced in heterokaryotypes [1–4]. By contrast, the effect of Rb fusions based on the state of the fusion (hetero- versus homozygous) has been less explored [5,6,8,11,12]. The fact that we observed a reduction in recombination in both heterokaryotypes and homokaryotypes suggest that fusions might evolve

differently than inversions do. Although at this stage it would be premature to discuss the evolutionary forces behind this pattern, both the absence of a fixed metacentric race and the high level of chromosomal polymorphisms that characterizes the Barcelona Rb system highlights its importance as an informative model. Further analysis of the genetic structure of this system would help us to elucidate whether suppressed recombination triggered by Rb fusions is indeed leaving a signature of genetic divergence.

4. Conclusion

Here, we show that mice with Rb fusions present a substantially reduced number of the total recombinational events as a result of a redistribution of COs in those chromosomal arms involved in Rb fusions. Moreover, the detection of novel *Prdm9* allelic variants in the Barcelona Rb polymorphic zone has permitted the examination of recombination variability observed within the population. Overall, our results suggest that changes in both number and distribution of recombination events are probably modulated by heterochromatinization disturbances produced by Rb fusions and influenced by the *Prdm9* genetic background.

5. Material and methods

A total of 31 wild male mice ($2n = 40$, 11 mice; $2n = 39$, 2 mice; $2n = 38$, 3 mice; $2n = 37$, 3 mice; $2n = 32$, 2 mice; $2n = 31$, 3 mice; $2n = 30$, 4 mice; $2n = 29$, 2 mice; $2n = 28$, 1 mouse) were live-trapped in commensal habitats from 10 different localities representative of the Barcelona Rb system (figure 1 and the electronic supplementary material, table S1). See the electronic supplementary material for details on sample processing, immunofluorescence, FISH, image processing and data analysis, and *Prdm9* genotyping.

Permission to capture was granted by the Departament de Medi Ambient of the Generaltitat de Catalunya (Spain). Animals were handled in compliance with the guidelines and ethical approval by the Comissió d'Ètica en l'Experimentación Animal y Humana (CEEAH) of the Universitat Autònoma de Barcelona and by the Departament d'Agriculta, Romaderia, Pesca, Alimentació i Medi Natural (Direcció General de Medi Natural i Biodiversitat) of the Generalitat de Catalunya.

Acknowledgements. The authors acknowledge Josep Medarde for his valuable help in the fieldwork and J. Davidian-Britton, T. J. Robinson and M. García-Caldés for insightful comments on earlier versions of the manuscript and to M. Fritzler for the sclerodactyly and telangiectasia (CREST) serum.

Funding statement. Financial support from Ministerio de Economia y Competitividad is gratefully acknowledged (CGL2010-15243 to J.V. and CGL-2010-20170 to A.R.H.). M.O.B. is supported by Instituto de Salud Carlos III (CP07/ 0258 and PI08/1185). L.C. is the beneficiary of an FPI pre-doctoral fellowship (BES-2011-047722), whereas A.A.S. is sponsored by a pre-doctoral fellowship from the Conselleria d'Educació, Cultura i Universitats (Govern Illes Balears) and Fons Social Europeu.

References

1. Rieseberg LH. 2001 Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358. (doi:10.1016/S0169-5347(01)02187-5)
2. Faria R, Navarro A. 2010 Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**, 660–669. (doi:10.1016/j.tree.2010.07.008)
3. Navarro A, Barton NH. 2003 Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science* **300**, 321–324. (doi:10.1126/science.1080600)
4. Farré M, Micheletti D, Ruiz-Herrera A. 2013 Recombination rates and genomic shuffling in human and chimpanzee: a new twist in the chromosomal speciation theory. *Mol. Biol. Evol.* **30**, 853–864. (doi:10.1093/molbev/mss272)
5. Dumas D, Britton-Davidian J. 2002 Chromosomal rearrangements and evolution of recombination: comparison of chiasma distribution patterns in standard and Robertsonian populations of the house mouse. *Genetics* **162**, 1355–1366.
6. Castiglia R, Capanna E. 2002 Chiasma repatterning across a chromosomal hybrid zone between chromosomal races of *Mus musculus domesticus*. *Genetica* **114**, 35–40. (doi:10.1023/A:1014626330022)
7. Ruiz-Herrera A, Farré M, Ponsà M, Robinson TJ. 2010 Selection against Robertsonian fusions involving housekeeping genes in the house mouse: integrating data from gene expression arrays and chromosome evolution. *Chromosome Res.* **18**, 801–808. (doi:10.1007/s10577-010-9153-8)
8. Franchini P, Colangelo P, Solano E, Capanna E, Verheyen E, Castiglia R. 2010 Reduced gene flow at pericentromeric loci in a hybrid zone involving chromosomal races of the house mouse *Mus musculus domesticus*. *Evolution* **64**, 2020–2032. (doi:10.1111/j.1558-5646.2010.00964.x)
9. Nunes AC, Catalan J, Lopez J, Ramalhinho MDG, Mathias MDL, Britton-Davidian J. 2011 Fertility assessment in hybrids between monobrachially homologous Rb races of the house mouse from the island of Madeira: implications for modes of chromosomal evolution. *Heredity* **106**, 348–356. (doi:10.1038/hdy.2010.74)
10. Förster DW, Mathias ML, Britton-Davidian J, Searle JB. 2013 Origin of the chromosomal radiation of Madeiran house mice: a microsatellite analysis of metacentric chromosomes. *Heredity* **110**, 380–388. (doi:10.1038/hdy.2012.107)
11. Merico V, Giménez MD, Vasco C, Zuccotti M, Searle JB, Hauffe HC, Garagna S. 2013 Chromosomal speciation in mice: a cytogenetic analysis of recombination. *Chromosome Res.* **21**, 523–533. (doi:10.1007/s10577-013-9377-5)
12. Bidau CJ, Giménez MD, Palmer CL, Searle JB. 2001 The effects of Robertsonian fusions on chiasma frequency and distribution in the house mouse (*Mus musculus domesticus*) from a hybrid zone in northern Scotland. *Heredity* **87**, 305–313. (doi:10.1046/j.1365-2540.2001.00877.x)
13. Merico V, Pigozzi MI, Esposito A, Merani MS, Garagna S. 2003 Meiotic recombination and spermatogenic impairment in *Mus musculus domesticus* carrying multiple simple Robertsonian translocations. *Cytogenet. Genome Res.* **103**, 321–329. (doi:10.1159/000076820)
14. Giménez MD, White TA, Hauffe HC, Panithanarak T, Searle JB. 2013 Understanding the basis of diminished gene flow between hybridizing chromosome races of the house mouse. *Evolution* **67**, 1446–1462. (doi:10.1111/evo.12054)
15. Borodin PM, Karamysheva TV, Belonogova NM, Torgasheva AA, Rubtsov NB, Searle JB. 2008 Recombination map of the common shrew, *Sorex araneus* (Eulipotyphla, Mammalia). *Genetics* **178**, 621–632. (doi:10.1534/genetics.107.079665)
16. Yannic G, Basset P, Hausser J. 2009 Chromosomal rearrangements and gene flow over time in an inter-specific hybrid zone of the *Sorex araneus* group. *Heredity* **102**, 616–625. (doi:10.1038/hdy.2009.19)
17. Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009 A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**, 373–375. (doi:10.1126/science.1163601)
18. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2010 PRDM9 is

a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840. (doi:10.1126/science.1183439)

19. Smagulova F, Gregoretti IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV. 2011 Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* **472**, 375–378. (doi:10.1038/nature09869)
20. Wegmann D *et al*. 2011 Recombination rates in admixed individuals identified by ancestry-based inference. *Nat. Genet.* **43**, 847–853. (doi:10.1038/ng.894)
21. Jeffreys AJ, Cotton VE, Neumann R, Lam KG. 2013 Recombination regulator PRDM9 influences the instability of its own coding sequence in humans. *Proc. Natl Acad. Sci. USA* **110**, 600–605. (doi:10.1073/pnas.1220813110)
22. Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012 Genetic recombination is directed away from functional genomic elements in mice. *Nature* **485**, 642–645. (doi:10.1038/nature11089)
23. Kirkpatrick M, Barton N. 2006 Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434. (doi:10.1534/genetics.105.047985)
24. Gazave E, Catalan J, Da Graça Ramalhinho M, Da Luz Mathias M, Claudia Nunes A, Dumas D, Britton-Davidian J, Auffray J-C. 2003 The non-random occurrence of Robertsonian fusion in the house mouse. *Genet. Res.* **81**, 33–42. (doi:10.1017/S001667230200602X)
25. Piálek J, Hauffe HC, Searle JB. 2005 Chromosomal variation in the house mouse. *Biol. J. Linn. Soc.* **84**, 535–563. (doi:10.1111/j.1095-8312.2005.00454.x)
26. Medarde N, López-Fuster MJ, Muñoz-Muñoz F, Ventura J. 2012 Spatio-temporal variation in the structure of a chromosomal polymorphism zone in the house mouse. *Heredity* **109**, 78–89. (doi:10.1038/hdy.2012.16)
27. Hausser J, Fedyk S, Fredga K, Searle JB, Volobouev V, Wojcik JM, Zima J. 1994 Definition and nomenclature of the chromosome races of *Sorex Araneus*. *Folia Zool.* **43**, 1–9.
28. Gündüz I, López-Fuster MJ, Ventura J, Searle JB. 2001 Clinal analysis of a chromosomal hybrid zone in the house mouse. *Genet. Res.* **77**, 41–51. (doi:10.1017/S0016672300004808)
29. Keeney S, Giroux CN, Kleckner N. 1997 Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375–384. (doi:10.1016/S0092-8674(00)81876-0)
30. Moens PB, Marcon E, Shore JS, Kochakpour N, Spyropoulos B. 2007 Initiation and resolution of interhomolog connections: crossover and non-crossover sites along mouse synaptonemal complexes. *J. Cell Sci.* **120**, 1017–1027. (doi:10.1242/jcs.03394)
31. Froenicke L, Anderson LK, Wienberg J, Ashley T. 2002 Male mouse recombination maps for each autosome identified by chromosome painting. *Am. J. Hum. Genet.* **71**, 1353–1368. (doi:10.1086/334714)
32. Wallace BM, Searle JB, Everett CA. 1992 Male meiosis and gametogenesis in wild house mice (*Mus musculus domesticus*) from a chromosomal hybrid zone; a comparison between 'simple' Robertsonian heterozygotes and homozygotes. *Cytogenet. Cell Genet.* **61**, 211–220. (doi:10.1159/000133410)
33. Matveevsky SN, Pavlova SV, Acaeva MM, Kolomiets OL. 2012 Synaptonemal complex analysis of interracial hybrids between the Moscow and Neroosa chromosomal races of the common shrew *Sorex araneus* showing regular formation of a complex meiotic configuration (ring-of-four). *Comp. Cytogenet.* **6**, 301–314. (doi:10.3897/compcytogen.v6i3.3701)
34. Manterola M *et al*. 2009 A high incidence of meiotic silencing of unsynapsed chromatin is not associated with substantial pachytene loss in heterozygous male mice carrying multiple simple Robertsonian translocations. *PLoS Genet.* **5**, e1000625. (doi:10.1371/journal.pgen.1000625)
35. Burgoyne PS, Mahadevaiah SK, Turner JM. 2009 The consequences of asynapsis for mammalian meiosis. *Nat. Rev. Genet.* **10**, 207–216. (doi:10.1038/nrg2505)
36. Hublitz P, Albert M, Peters AH. 2009 Mechanisms of transcriptional repression by histone lysine methylation. *Int. J. Dev. Biol.* **53**, 335–354. (doi:10.1387/ijdb.082717ph)
37. Parvanov ED, Petkov PM, Paigen K. 2010 Prdm9 controls activation of mammalian recombination hotspots. *Science* **327**, 835. (doi:10.1126/science.1181495)
38. Berg IL, Neumann R, Lam K-WG, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ. 2010 PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat. Genet.* **42**, 859–863. (doi:10.1038/ng.658)
39. Lynn A, Ashley T, Hassold T. 2004 Variation in human meiotic recombination. *Annu. Rev. Genomics Hum. Genet.* **5**, 317–349. (doi:10.1146/annurev.genom.4.070802.110217)
40. Bernard P, Maure JF, Partridge JF, Genier S, Javerzat JP, Allshire RC. 2001 Requirement of heterochromatin for cohesion at centromeres. *Science* **294**, 2539–2542. (doi:10.1126/science.1064027)
41. Mirouze M, Lieberman-Lazarovich M, Aversano R, Bucher E, Nicolet J, Reinders J, Paszkowski J. 2012 Loss of DNA methylation affects the recombination landscape in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **109**, 5880–5885. (doi:10.1073/pnas.1120841109)
42. Cole F, Kauppi L, Lange J, Roig I, Wang R, Keeney S, Jasin M. 2012 Homeostatic control of recombination is implemented progressively in mouse meiosis. *Nat. Cell Biol.* **14**, 424–430. (doi:10.1038/ncb2451)
43. Garcia-Cruz R, Pacheco S, Brieño MA, Steinberg ER, Mudry MD, Ruiz-Herrera A, Garcia-Caldés M. 2011 A comparative study of the recombination pattern in three species of Platyrrhini monkeys (primates). *Chromosoma* **120**, 521–530. (doi:10.1007/s00412-011-0329-6)
44. Segura J *et al*. 2013 Evolution of recombination in eutherian mammals: insights into mechanisms that affect recombination rates and crossover interference. *Proc. R. Soc. B* **280**, 20131945. (doi:10.1098/rspb.2013.1945)
45. Sans-Fuentes MA, García-Valero J, Ventura J, López-Fuster MJ. 2010 Spermatogenesis in house mouse in a Robertsonian polymorphism zone. *Reproduction* **140**, 569–581. (doi:10.1530/REP-10-0237)
46. Medarde N, Muñoz-Muñoz F, López-Fuster MJ, Ventura J. 2013 Variational modularity at the cell level: insights from the sperm head of the house mouse. *BMC Evol. Biol.* **13**, 179. (doi:10.1186/1471-2148-13-179)
47. Berg IL, Neumann R, Sarbajna S, Odenthal-Hesse L, Butler NJ, Jeffreys AJ. 2011 Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. *Proc. Natl Acad. Sci. USA* **108**, 12 378–12 383. (doi:10.1073/pnas.1109531108)

4.3.2 Supplementary information

4.3.2.1 Supplementary materials and methods

4.3.2.1.1 Animals and chromosomal characterization

A total of 31 wild male mice were live-trapped in commensal habitats from 10 different localities representatives of the Barcelona Rb system (Table S1 and Figure S1). Three males from the laboratory strain C57BL6 were also included in the analysis for comparison. Mitotic metaphase spreads were obtained from fibroblast cultures and bone marrow cells. Cell cultures were derived from fresh tissues samples and the harvesting of cells followed standard procedures. Karyotypes were determined by analyzing G-banded chromosomal preparations (Scally et al. 2012). When G-banding identification was not sufficient for chromosomal identification, we used commercial mouse chromosome-specific painting probes (Cambio, Cambridge, UK) to confirm G-banding homologies following (Ruiz-Herrera et al. 2004). The Ethics Committee of the Universitat Autònoma de Barcelona approved the animal treatment protocols adopted herein.

4.3.2.1.2 Immunoflourescence

Immunofluorescence (IF) of spermatocyte spreads was performed following (Garcia-Cruz et al. 2011). Different sets of antibodies were used: mouse monoclonal antibody against MLH1 (BD Pharmigen) for the detection of crossovers (COs), rabbit polyclonal serum against central element protein of the synaptonemal complex (SC) (SYCP3, Abcam), human calcinosis, Raynaud's phenomenon, esophageal dysfunction, sclerodactyly and telangiectasia (CREST) serum for the detection of the centromeres, mouse γH2AX (Millipore), rabbit H3K9me3 (Abcam) and mouse RPA (Abcam). Corresponding secondary antibodies were: anti-mouse conjugated with FITC, anti-mouse conjugated with Cy3, anti-rabbit Cy3 and anti-human Cy5 (all purchased from Jackson Immunoresearch).

4.3.2.1.3 Fluorescence *in situ* hybridization

With the aim of identifying the chromosomes implicated in Rb fusions, fluorescence *in situ* hybridization (FISH) was performed on previously immunostained preparations using Bacterial Artificial Chromosome (BAC) probes for specific mouse chromosomes (chromosomes 4, 9, 11, 12, 13 and 14). BAC clones where selected from the UCSC genome browser (http://genome.ucsc.edu) and purchased from CHORI (http://bacpac.chori.org) (Table S3). DNA was purified from bacterial pellets using a Midiprep extraction kit (Quiagen). FISH was performed on both metaphase chromosomes and spermatocyte spreads as previously described (Ruiz-Herrera et al. 2004; Garcia-Cruz et al. 2011), with modifications. Briefly, 1 µg of the plasmid DNA was labeled with dUTP-digoxigenin using a nick translation kit (Abbot) and ethanol precipitated with competitor DNA (COT-1 Mouse DNA, Invitrogen, 1mg/ml), salmon sperm DNA (Invitrogen, 10mg/ml) and 1/10 volume of 3mol/L sodium acetate overnight at -20ºC. The slides were dehydrated in an ethanol scale containing 70% formamide at 74ºC before, and after the denaturation. The probe was precipitated and washed twice with 70% ethanol and then diluted in 15 ul of hybridization buffer (50% deionized formamide, 10% dextran sulfate, 2XSSC and 0.5mol/L phosphate). Chromosomes were denatured at 74º at which point the probe hybridization mix was placed directly on the slides, which were cover-slipped and incubated O/N at 37ºC in a humid chamber. Washes were performed with 2X saline sodium citrate (SSC) and 50% formamide solution before applying the anti-digoxigenin conjugated with FITC.

4.3.2.1.4 Image processing and data analysis

Preparations were visualized using a Zeiss Axioskop epifluorescence microscope equipped with the appropriate filters and a charged coupled device camera (ProgRes® CS10plus, Jenoptik). Images were captured and produced by the ProgRes® software (2.7.7). Images for each specimen were taken blindly (i.e., specimens were numerically assigned irrespectively of their chromosomal composition) in order to avoid against unintentional biases in MLH1 scores.

The Micromeasure 3.3 software (http://www.colostate.edu/Depts/Biology/MicroMeasure) (Reeves 2001) was used for the analysis of recombination maps (based on the distances between adjacent MLH1 and RPA *foci*). For each chromosome analyzed, the position of each focus (MLH1 and RPA) was recorded as a relative

position (percentage of the synaptonemal complex total length) from the centromere, identified by the CREST signal in each preparation. The positions of MLH1/RPA *foci* were calculated for each chromosome using the centromere as reference point, (i.e., from the centromere to the telomere in the q-arm). Thus for comparison among chromosomes, the positions of MLH1 *foci* were expressed as the relative position of each CO to the length of the chromosome (the length of each SC was divided into 10% intervals). These data were used to construct cumulative frequency distribution plots and we compared the different distributions using a Kolmogorov-Smirnov test (KS test). Moreover, we analyzed the frequencies of MLH1 *foci* along SCs by dividing them in three regions depending on the relative distance from the centromere: (i) proximal (from the centromere to 30% of the SC) (ii) interstitial (between 30%-70% of SC) and (iii) distal (from 70% to telomeric region). For the analysis of H3K9me3 chromosomal distribution, the Micromeasure software was used to calculate the signal area (measured in micrometers) overlapping the SC for each chromosomal arm analyzed.

The statistical analysis was carried out by means of PAWS Statistics 18© and JMP 7 software package. Given that the distribution of crossovers per specimen did not follow a normal distribution (Shapiro-Wilk test, p-value≤0.001), we applied nonparametric analysis, such as Mann–Whitney U test or Kruskal–Wallis test for mean comparison and Spearman test for correlations.

4.3.2.1.5 *Prdm9* genotyping

Genomic DNA was extracted and purified from cell cultures derived from mice of the Rb polymorphism zone (n=27) following the standard phenol-chloroform protocol. We sequenced the *Prmd9* exon 12, from Zn repeat +3 towards the c-terminal domain. This exon contains the ZnF domain array that recognizes and methylates the specific DNA sequences (Mihola et al. 2009; Baudat et al. 2010). PRDM9 zinc finger arrays were amplified by PCR using ExTaq™ (Takara), with 4% DMSO, primers fl1500U20 and 2848L23 (Mihola et al. 2009) as follows: 95ºC (3 min), 30 cycles of 95ºC (30 s); 56ºC (30 s); 72ºC (90 s). PCR products were run in a 2% agarose gel in order to distinguish different haplotypes (Figure S5). When mice had alleles of the same length (length homozygotes, n=20), PCR amplified fragments were purified with the GeneJET™ PCR purification kit (Fermentas). For mice with alleles of two different lengths (length heterozygotes, n=7), bands were cut from the agarose gel and purified using the GeneJET™ Gel Extraction Kit (Fermentas). In both

cases, purified gene products were subjected to bidirectional Sanger sequencing with primers fl1822U24 and 2848L23 (Baudat et al. 2010) and reads were subsequently analyzed using the CLCbio Main Workbench program, version 6.8 (Aarhus, DK). For those animals found heterozygotes for alleles of the same length, two different softwares were used in order to distinguish allelic variants and define the haplotypes: Phase (http://stephenslab.uchicago.edu/software.html) and Champuru v1.0 (http://www.mnhn.fr/jfflot/champuru/). In any case, all new allelic variants were confirmed by re-sequencing and re-analyzed.

4.3.2.2 Supplementary Figures and Tables

Table S1: List of specimens, localities, diploid number (2n), chromosomal characteristics, mean number of MLH1 *foci*/cell (± standard deviation), number of cells analyzed in the recombination study (N) and *Prdm9* genotyping of mice from the chromosomal polymorphism area of Barcelona included in the study. St= mice with standard karyotype (2n=40). Rb= mice with Robertsonian fusions. n.a.= DNA not available for genotyping.

Specimen	Location	2n	Fusions	Mean MLH1 foci/cell ±SD	N	*Prdm9* alleles
St 1	Arbeca	40	-	21.65 ± 2.43	55	n.a.
St 2	Vacarisses	40	-	21.16 ± 2.23	89	n.a.
St 3	Castellfollit del Boix	40	-	20.90 ± 1.74	20	10A/12B
St 4	Castellfollit del Boix	40	-	20.80 ± 1.43	21	12B/12B
St 5	Castellfollit del Boix	40	-	20.68 ± 1.35	22	10A/10A
St 6	Castellfollit del Boix	40	-	20.68 ± 1.43	25	10A/12B
St 7	Castellfollit del Boix	40	-	20.57 ± 0.87	21	10A/12C
St 8	Castellfollit del Boix	40	-	20.57 ± 1.45	14	10A/12C
St 9	Castellfollit del Boix	40	-	20.27 ± 1.40	26	12B/12B
St 10	Castellfollit del Boix	40	-	20.16 ± 1.17	25	10A/10A
St 11	Santa P. de Mogoda	40	-	20.58 ± 1.02	31	10B/11B
Rb 1	Sant Sadurní d'Anoia	39	Rb (12.13)	20.26 ± 1.68	27	n.a.
Rb 2	Sant Sadurní d'Anoia	37	Rb (4.14)+2(9.11)	19.71 ± 2.22	40	10A/10A
Rb 3	Sant Sadurní d'Anoia	38	Rb (4.14)+(9.11)	19.71 ± 1.65	46	10A/10A
Rb 4	Sant Sadurní d'Anoia	37	Rb (4.14)+(9.11)+(12.13)	20.57 ± 1.90	46	n.a.
Rb 5	Sant Sadurní d'Anoia	37	Rb (4.14)+(9.11)+(12.13)	20.26 ± 1.86	47	10A /10A
Rb 6	El Papiol	38	Rb (4.14)+(12.13)	19.39 ± 1.19	37	10A /10A
Rb 7	El Papiol	38	Rb (4.14)+(12.13)	19.64 ± 1.27	11	10A/10A
Rb 8	Ametlla de Segarra	39	Rb (4.14)	18.93 ± 0.68	16	10A/10A
Rb 9	Castelldefels	29	Rb 2(4.14)+2(9.11)+2(12.13)+2(3.8)+2(5.15)+(6.10)	19.95 ± 1.67	44	10A/10A
Rb 10	Castelldefels	32	Rb 2(4.14)+(9.11)+2(12.13)+2(5.15)+(6.10)	19.92 ± 1.59	40	10A/10A
Rb 11	Castelldefels	30	Rb 2(4.14)+(9.11)+2(12.13)+2(3.8)+2(5.15)+(6.10)	19.85 ± 1.61	27	10A/10A
Rb 12	Castelldefels	30	Rb 2(4.14)+2(9.11)+2(12.13)+(3.8)+2(5.15)+(6.10)	19.59 ± 1.53	37	10A/10A
Rb 13	Castelldefels	28	Rb 2(4.14)+2(9.11)+2(12.13)+2(3.8)+2(5.15)+2(6.10)	20.23 ± 1.81	18	10A/10A
Rb 14	Castelldefels	30	Rb 2(4.14)+2(9.11)+2(12.13)+ (3.8)+2(5.15)+(6.10)	19.00 ± 0.94	19	10A/10A
Rb 15	Castelldefels	29	Rb 2(4.14)+2(9.11)+2(12.13)+2(3.8)+2(5.15)+(6.10)	18.95 ± 1.25	22	10A/10A
Rb 16	Castelldefels	30	Rb 2(4.14)+2(9.11)+2(12.13)+(3.8)+2(5.15)+(6.10)	18.73 ± 1.33	34	10A/10A
Rb 17	Castelldefels	31	Rb 2(4.14)+(9.11)+2(12.13)+(3.8)+2(5.15)+(6.10)	18.12 ± 1.70	16	10A/10A
Rb 18	Prat de Llobregat	31	Rb 2(4.14)+(9.11)+2(12.13)+2(5.15)+2(6.10)	18.78 ± 1.12	14	10A/12B
Rb 19	Prat de Llobregat	32	Rb 2(4.14)+2(9.11)+(12.13)+2(5.15)+(6.10)	19.75 ± 1.99	40	10A/12B
Rb 20	Viladecans	31	Rb 2(4.14)+2(9.11)+(12.13)+2(3.8)+(5.15)+(6.10)	18.65 ± 2.20	22	10A/10A

Table S2: Absolute number (and relative proportion) of chromosomal arms with 0, 1 and 2 MLH1 *foci* per chromosomal arm for each of the following groups: chromosomal arms involved in Rb fusions (Met), acrocentric arms from standard mice (Ac St) and acrocentric arms from Robertsonian mice (Ac Rb). Fisher's test, **p-value≤0.001.

Number of MLH1 *foci*/chr arm	Nº (%) chr arms		
	Ac St	Ac Rb	Met
0	33 (2.6%)	113 (5.11%)	129 (6.27%)
1	1042 (87.05%)	1831 (82.81%)	1799 (87.50%)
2	122 (10.19%)	267 (12.08%)	128 (6.23%)**

Table S3: List of mouse BACs used in the study.

		Genomic position [b]	
Moue Chr	BAC clone[a]	Start (bp)	End (bp)
4	RP23-173N19	108,976,146	109,191,836
9	RP23-85H15	25,021,396	25,255,853
11	RP24-362I10	31,158,081	31,334,208
12	RP23-178L21	66,948,885	67,164,057
13	RP24-248P14	66,986,750	67,156,572
14	RP23-16H17	41,847,042	42,078,689

[a] Clones obtained from the mouse library RPCI-23 (female C57BL/6 *Mus musculus*). **BACPAC Resources Center (BPRC), CHORI.**

[b] According to NCBI37/mm9.

Figure S1. Correlation between the number of MLH1 *foci* per cell and the diploid number per individual analyzed: See Table 1 for additional information on individuals analyzed (Spearman correlation test, ρ =0.75, p≤0.001).

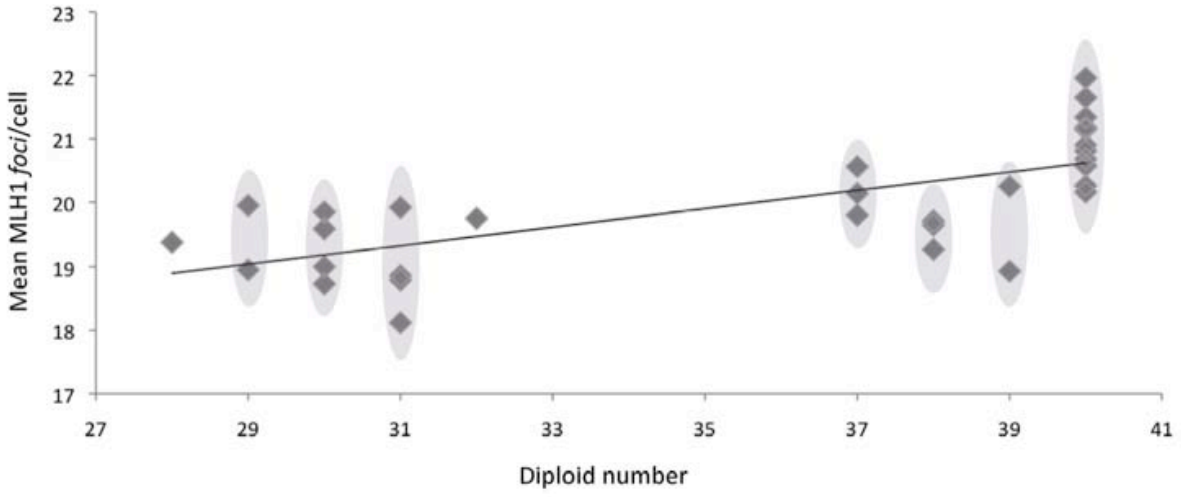


Figure S2. Analysis of the chromosomal distribution of RPA *foci*.: (A-B) Examples of immunolocalization of early meiotic recombination nodules (RPA *foci*) in mouse spermatocytes at pachynema. RPA is showed in red, centromeres in blue and the synaptonemal complexes in green. (A) Specimen with 2n=40. (B) Specimen with 2n=38, Rb (4.14)+(9.11). (C) Mean number of RPA *foci* detected per cell in the laboratory strain (Lab strain, n=37 cells, N=1 mouse), wild standard (St, n= 38 cells, N=2 mice) and Robertsonian (Rb, n=48 cells, N=3 mice) specimens. (D) Cumulative frequencies of RPA *foci* along chromosomal arms involved in Rb fusions (Met, in red, n=468 chromosomal arms) and in chromosomal arms not involved in Rb fusions (Acr, in blue, n=900 chromosomal arms). Chromosomal arms are divided into 10% intervals of SC length. Asterisk indicates statistical significance (Kruskal-Wallis test, **p-value≤0.001).

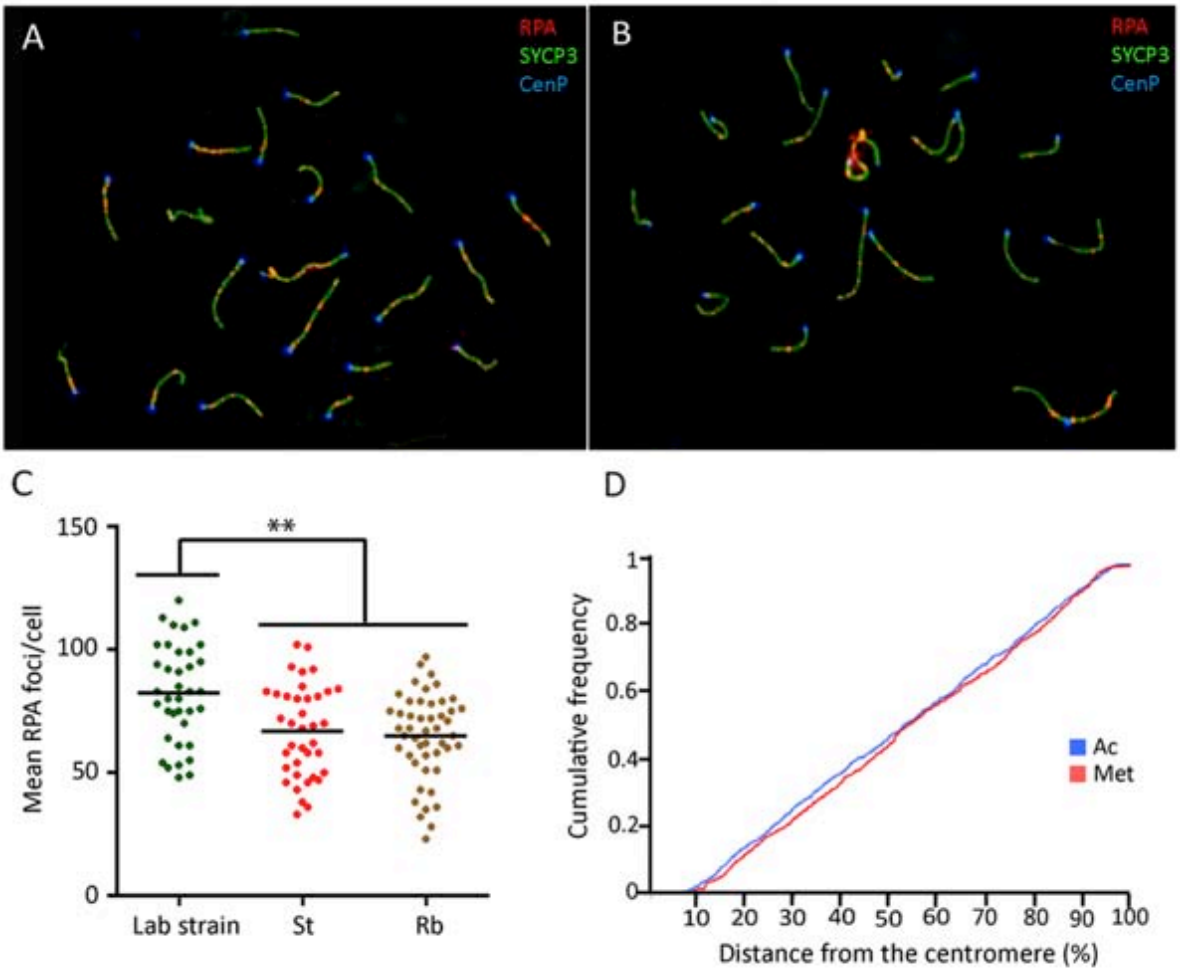


Figure S3. Chromosomal distributions of recombination events: Distribution of COs along the SC of six chromosomes involved in Rb fusions: Chr4, Chr9, Chr11, Chr12, Chr13 and Chr14. The X-axis represents the length of the synaptonemal complex (SC) from the centromere to the telomere, while the Y-axis shows the percentage of total MLH1 *foci* found per SC. Each chromosome is divided into 5% intervals of SC length. Blue lines indicate chromosomes in acrocentric form (n=38 cells analyzed for Chr4, n=25 for Chr9, n=36 for Chr11, n=31 for Chr12, n=71 for Chr13 and n=46 for Chr14). Green lines represent the distribution of MLH1 *foci* when chromosomes are involved in Rb fusions in homozygosis [(n=35 cells analyzed for Rb(4.14), n=52 Rb(9.11), n=41 Rb(12.13)] and red lines when the Rb fusions are present in heterozygosis [(n=159 cells analyzed for Rb(4.14), n=42 for Rb(9.11) and n=54 for Rb(12.13)]. Arrows in the X-axis show the position of the centromere.

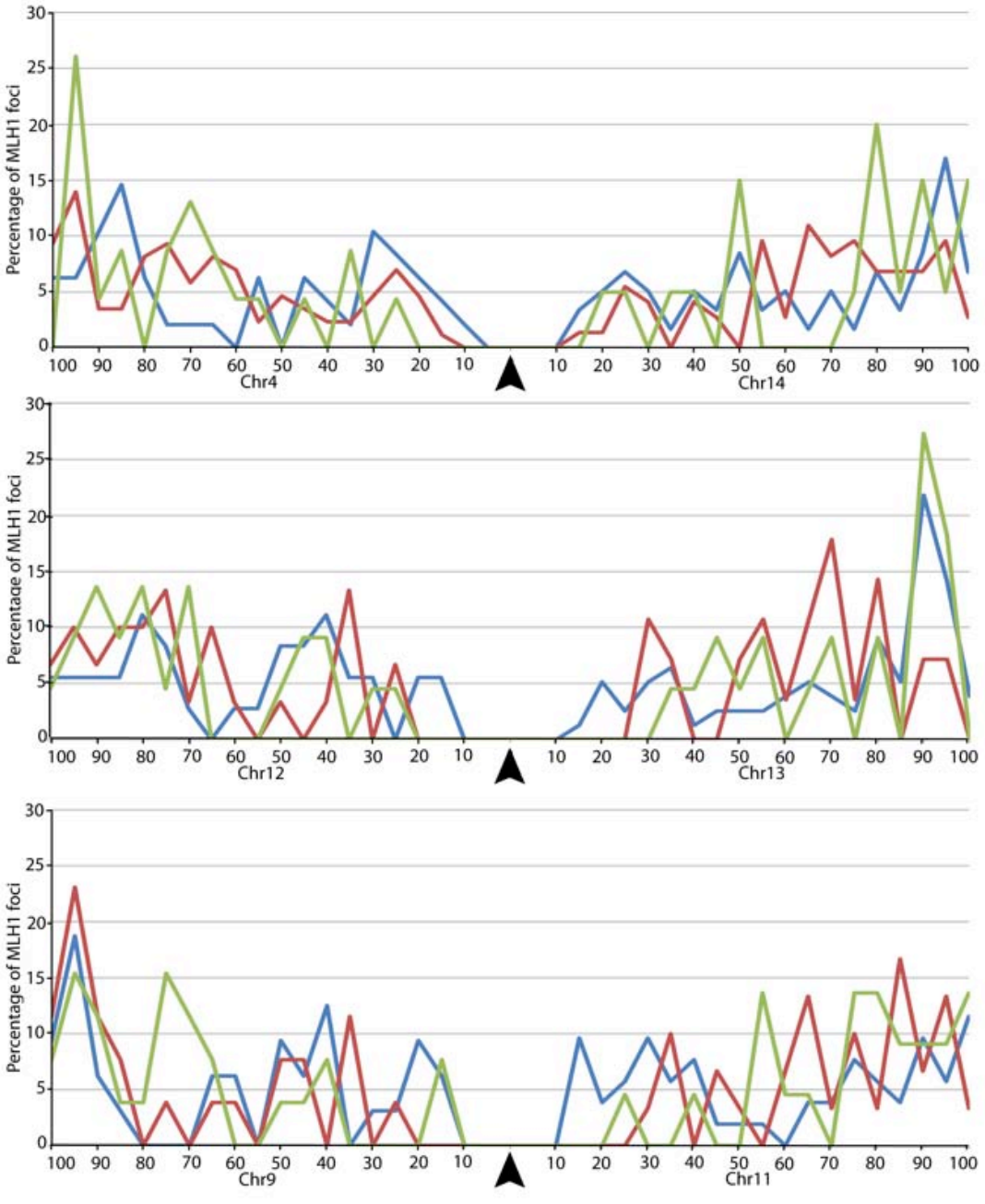


Figure S4. Distribution of γH2AX signal in Rb mouse spermatocytes: (A) Leptonema showing γH2AX signal distributed across the whole nucleus. (B) Diplonema with γH2AX signal restricted to the sex body. (C-D) Early pachynema showing how γH2AX signal is not only located in the sex chromosomes, but also across synapsed regions in both bivalents and trivalents as well as in asynapsed pericentromeric areas in trivalents. (C) Note the scattered distribution of γH2AX signal along the synaptonemal complex in synapsed regions in bivalents and trivalents. (D) γH2AX labeling is also detected in asynapsed areas in the trivalent (white arrow and inset). (E) Examples of the different configurations observed for trivalents (closed and open). SYCP3 is depicted in red and γH2AX in green. Note how γH2AX signals are not only located as large blocks in asynapsed regions but also scattered along synapsed regions. (F-G) Frequency of mid-pachynema with scattered (altered) γH2AX labeling and asynapsed XY pair in Rb mice (three specimens with diploid number ranging 2n=29-38, n=173 cells analyzed), wild standard mice (two specimens with 2n=40, n=176 cells analyzed) and the laboratory strain (one specimen with 2n=40, n=119 cells analyzed). Asterisks indicate statistical significance (Fisher's test, **p-value≤ 0.001).

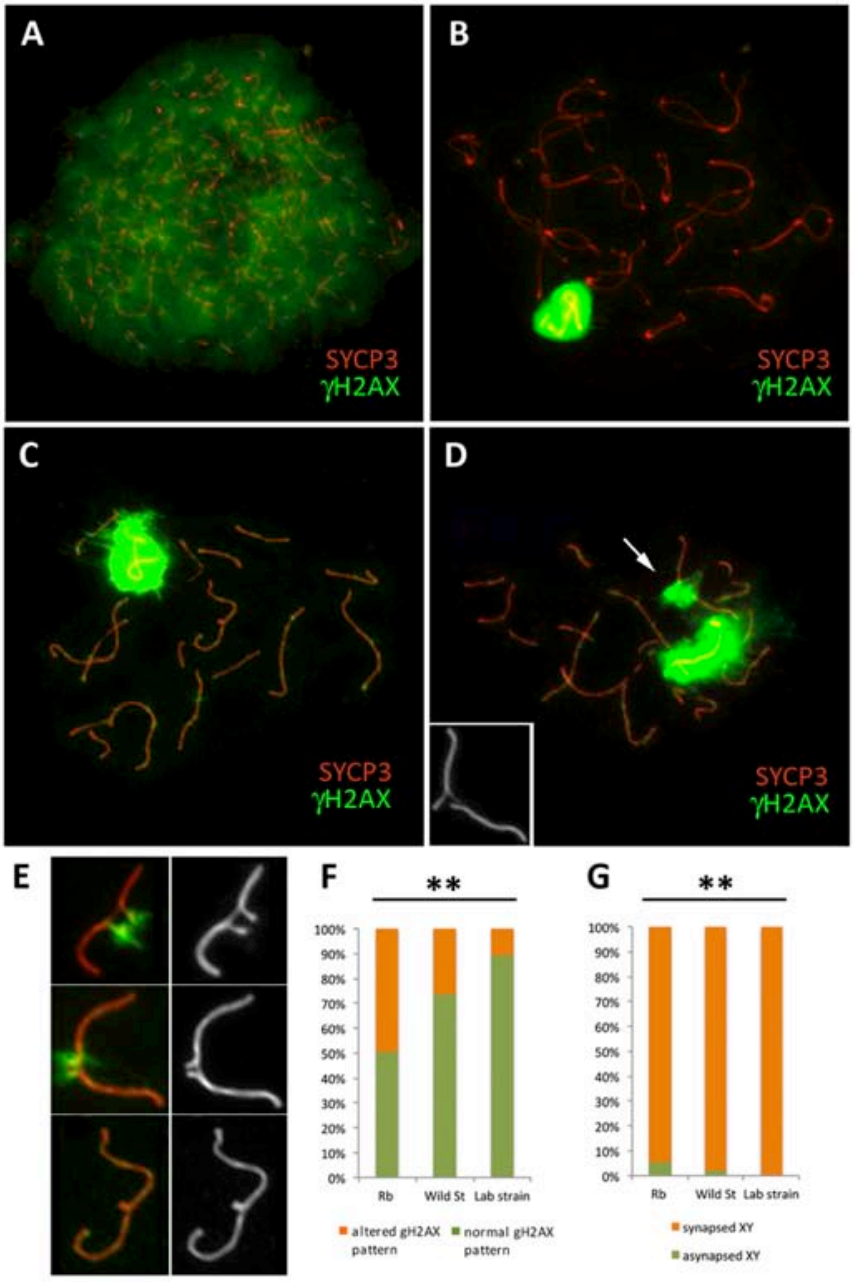
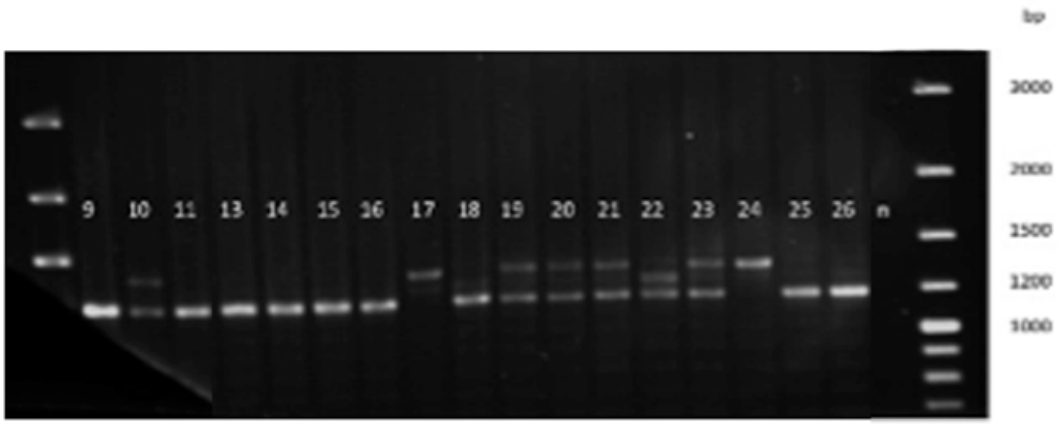


Figure S5. *Prdm9* haplotypes: Agarose gel showing a set of different haplotypes found in the Barcelona chromosomal polymorphism zone. Each well corresponds to a different specimen: St9 (9), Rb19 (10), Rb20 (11), Rb14 (13), Rb16 (14), Rb17 (15), Rb11 (16), St9 (17), St5 (18), Rt7 (19), St8 (20), St6 (21), St11 (22), St3 (23), St4 (24), Rb10 (25), Rb12 (26). See Table 1 for additional information on the individuals analyzed.



Zigure S6. *Prdm9* allelic diversity: (A) Graphic representation of the *Prdm9* reference allele described for mouse and the different alleles present in the Barcelona Rb polymorphism zone. ZnF repeats are color-coded. Allelic frequencies found in the Barcelona Rb polymorphism zone are also shown. (B) Geographical distribution of *Prdm9* alleles. Locations with standard and Rb individuals are showed in red and black, respectively (C) Correlation between *Prdm9* alleles and recombination frequency expressed in mean number of MLH1 *foci* per cell detected for each group (N/N, n=47; 10/N, n=151; 10/10, n=537) (Kruskal-Wallis test, **p-value≤0.001).

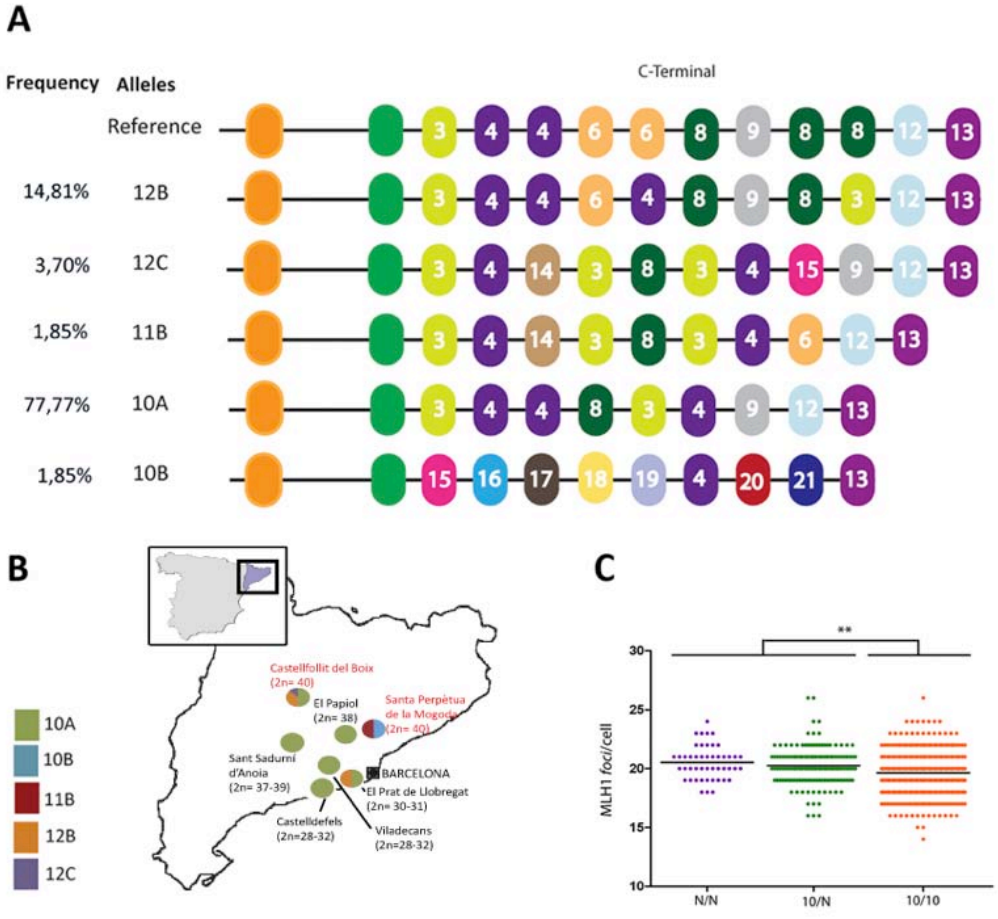


Figure S7. DNA sequence alignments of all different *Prdm9* haplotypes found in the Barcelona Rb polymorphism zone: The number of the repeat indicated in the left side of the figure indicates the type of ZnF repeats found in this study.

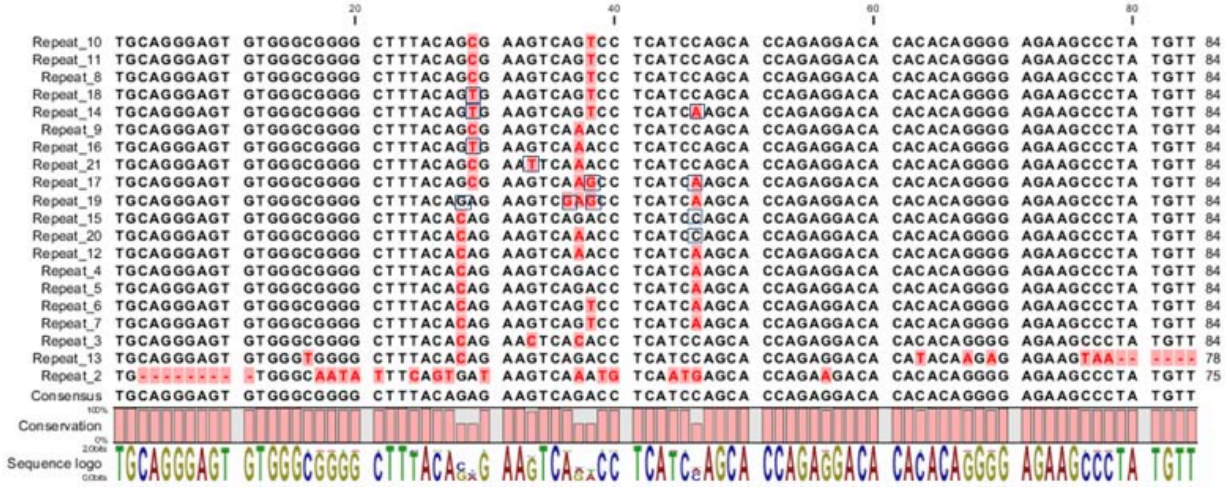


Figure S8. Haplotypes of the Zn finger domains of *Prdm9* found in the Barcelona Rb polymorphism zone compared with the alleles described in the literature (Parvanov et al. 2010): Variations in amino acids among repeats are colored in grey. Numbers on the right of each domain indicate the type of Zn repeats found in this study, based on the DNA sequence.

Zn=10

Allele 10A (this study)

CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTAKSVLIQHQRTHTGEKPYV **8**
CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTAKSNLIQHQRTHTGEKPYV **9**
CRECGRGFTQKSNLIKHQRTHTGEKPYV **12**
CRECGWGFTQKSDLIQHQRTHTREK* **13**

Zn=10

Allele 10B (this study)

CRECGRGFTQKSDLIQHQRTHTGEKPYV **15**
CRECGRGFTVKSNLIQHQRTHTGEKPYV **16**
CRECGRGFTAKSSLIKHQRTHTGEKPYV **17**
CRECGRGFTVKSVLIQHQRTHTGEKPYV **18**
CRECGRGFTEKSSLIKHQRTHTGEKPYV **19**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTQKSNLIQHQRTHTGEKPYV **20**
CRECGRGFTANSNLIQHQRTHTGEKPYV **21**
CRECGWGFTQKSDLIQHQRTHTREK* **13**

Zn=11

Allele 11B (this study)

CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTVKSVLIKHQRTHTGEKPYV **14**
CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTAKSVLIQHQRTHTGEKPYV **8**
CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTQKSVLIKHQRTHTGEKPYV **6**
CRECGRGFTQKSNLIKHQRTHTGEKPYV **12**
CRECGWGFTQKSDLIQHQRTHTREK* **13**

Allele 11A [7]

CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTQKSVLIQHQRTHTGEKPYV
CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTEKSSLIKHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTQKSSLIKHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGWGFTQKSNLIKHQRTHTGEKPYV
CRECGWGFTQKSDLIQHQRTHTREK*

Zn=12

Allele 12B (this study)

CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTQKSVLIKHQRTHTGEKPYV **6**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTAKSVLIQHQRTHTGEKPYV **8**
CRECGRGFTAKSNLIQHQRTHTGEKPYV **9**
CRECGRGFTAKSVLIQHQRTHTGEKPYV **8**
CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTQKSNLIKHQRTHTGEKPYV **12**
CRECGWGFTQKSDLIQHQRTHTREK* **13**

Allele 12A [7]

CRECGRGFTQNSHLIQHQRTHTGEKPYV
CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTQKSVLIKHQRTHTGEKPYV
CRECGRGFTQKSVLIKHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTQKSNLIKHQRTHTGEKPYV
CRECGWGFTQKSDLIQHQRTHTREK*

Allele 12C (this study)

CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTVKSVLIKHQRTHTGEKPYV **14**
CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTAKSVLIQHQRTHTGEKPYV **8**
CRECGRGFTQNSHLIQHQRTHTGEKPYV **3**
CRECGRGFTQKSDLIKHQRTHTGEKPYV **4**
CRECGRGFTQKSDLIQHQRTHTGEKPYV **15**
CRECGRGFTAKSNLIQHQRTHTGEKPYV **9**
CRECGRGFTQKSNLIKHQRTHTGEKPYV **12**
CRECGWGFTQKSDLIQHQRTHTREK* **13**

Zn= 13

Allele 13A [7]

CRECGRGFTQNSHLIQHQRTHTGEKPYV
CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTQKSVLIKHQRTHTGEKPYV
CRECGRGFTQKSVLIKHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTQKSNLIKHQRTHTGEKPYV
CRECGWGFTQKSDLIQHQRTHTREK*

Zn= 14

Allele 14A [7]

CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTQKSVLIKHQRTHTGEKPYV
CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTEKSSLIKHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTQKSVLIKHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGWGFTQKSNLIKHQRTHTGEKPYV
CRECGWGFTQKSDLIQHQRTHTREK*

Allele 14B [7]

CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTQKSVLIKHQRTHTGEKPYV
CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTQKSDLIKHQRTHTGEKPYV
CRECGRGFTAKSNLIQHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGRGFTEKSSLIKHQRTHTGEKPYV
CRECGRGFTQKSNLIKHQRTHTGEKPYV
CRECGRGFTAKSVLIQHQRTHTGEKPYV
CRECGWGFTQKSNLIKHQRTHTGEKPYV
CRECGWGFTQKSDLIQHQRTHTREK*

4.3.2.3 Supplementary references

Baudat F, Buard J, Grey C, et al (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327, 836–40.

Garcia-Cruz R, Pacheco S, Brieño MA, et al (2011) A comparative study of the recombination pattern in three species of Platyrrhini monkeys (Primates). Chromosoma 120, 521–30.

Mihola O, Trachtulec Z, Vlcek C, et al (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science 323, 373-5.

Parvanov ED, Petkov PM, Paigen K (2010) *Prdm9* controls activation of mammalian recombination hotspots. Science 327, 835.

Reeves A (2001) MicroMeasure: A new computer program for the collection and analysis of cytogenetic data. Genome 44, 439-43.

Scally A, Dutheil J, Hillier L, et al (2012) Insights into hominid evolution from the gorilla genome sequence. Nature 483:169-75

Ruiz-Herrera A, Garcia F, Frönicke L, et al (2004) Conservation of aphidicolin-induced fragile sites in Papionini (Primates) species and humans. Chromosome Res. 12, 683–90.

5.

GENERAL DISCUSSION

5.1 Genome reshuffling in Rodentia: causes and consequences of the genomic distribution of EBRs

Chromosomal rearrangements have shaped the architecture of mammalian genomes, giving as a result the current genome diversity. Therefore, analyzing the mechanisms that are involved in genome reshuffling will help us to understand the driving forces of the evolutionary process. In this way, comparative genomic studies have facilitated this task. The availability of an increasing number of sequenced genomes, together with the development of new advanced algorithms has allowed the detection, at a high resolution, of the genomic regions involved in large-scale reorganizations. Relevant advances have been achieved in the recent years that have permitted the detection of regions of homology (HSBs) between mammalian species. These include the application of different algorithms such as the *GRIMM-Synteny* (Pevzner and Tesler 2003), *Satsuma Synteny* (Grabherr et al. 2010), *Synteny Tracker* (Donthu et al. 2009) and *CASSIS* (Baudet et al. 2010). Such approaches have not only improved the resolution at which HSBs can be now characterized but also have permitted multispecies comparisons, providing with the possibility to establish genomic maps of large-scale rearrangements within several taxa (Murphy et al. 2005b; Ruiz-Herrera et al. 2006; Kemkemer et al. 2009; Lemaitre et al. 2009; Larkin et al. 2009; Farré et al. 2011; Zhang et al. 2014).

Among mammals, rodents represent the most specious taxon (with more than 2,000 species described so far) with a wide diversity of phenotypes that has been translated into a broad range of adaptations to the environment (Storz et al. 2009; Fang et al. 2014). Due to the complexity of their evolutionary history, rodents (with the exception of mouse and rat) have been often overlooked from comparative genomic studies. In fact, rodents are characterized by specific features, such as high rates of genome reshuffling (reviewed in Romanenko et al. 2012), high rates of nucleotide substitution (Wu and Li 1985), and low recombination rates (Dumont and Payseur 2011; Segura et al. 2013) when compared to other mammalian species. In the case of CRs, previous comparative studies provided relevant information on both the most parsimonious ancestral karyotype (Bourque et al. 2004; Ma et al. 2006; Graphodatsky et al. 2008; Mlynarski et al. 2010; Trifonov et al. 2010; Romanenko et al. 2012; Romanenko and Volobouev 2012) and the specific large-scale rearrangements of Rodentia (Pevzner and Tesler 2003; Zhao et al. 2004; Mlynarski et al. 2010). In the house mouse, for example, fissions and fusions are the most common specific rearrangements, while inversions are more frequently observed in rats (Zhao et al.

2004). However, the resolution at which large-scale reorganizations were detected was still poor and, therefore, the general picture of the genomic rearrangements of Rodentia at a finer scale remained to be uncovered.

In order to overcome this limitation, we have reconstructed, for the first time in the literature, a detailed map of the genomic regions involved in evolutionary reshuffling in rodents (section 4.1). This has been possible due to: (i) the use of algorithms that identify sequence homologies between genomes establishing HSBs (SS and ST, see section 3.1.1 and 3.1.2), (ii) the detection and classification of EBRs in a phylogenetic context (EBA, see section 3.1.3) and (iii) the use of an unprecedented large amount of Rodentia species representative of the major groups, such as Hystricomorpha (*H. glaber*) and Myodonta (*J. jaculus, S. galilii, M. ochorgaster, R. norvegicus* and *M. musculus*) in addition to six mammals belonging to Primates (*H. sapiens, M. mulatta* and *P. pygmaeus*), Carnivora (*F. catus*), Artyodactyla (*B. taurus*) and Perissodactyla (*E. caballus*), that served as outgroup species. This comparative approach has permitted the detection of multispecies regions of homology (HSBs and SFs, see section 3.1.2) and EBRs at a high resolution due to the pair-wise comparisons made at a sequence level. In total, we identified 3,392 HSBs, 3,142 SFs and 1,333 EBRs among all species analyzed. Our analysis showed that, overall, Rodentia species present high rates of genome reshuffling (ranging from 1.38 EBRs/Myr in *H. glaber* to 8.47 EBRs/Myr in *S. galilii* in lineage-specific EBRs) compared to other mammals (ranging from 0.50 EBRs/Myr in *P.pygmaeus* to 1.81 EBRs/Myr in *H. Sapiens*) (section 4.1; Figure 1). These results support previous studies, using both cytogenetic (Stanyon et al. 1999; Murphy et al. 2005b; Mlynarski et al. 2010; Romanenko et al. 2012) and comparative genomics approaches (Pevzner and Tesler 2003a; Bourque et al. 2004; Zhao et al. 2004; Lewin et al. 2009). But, more importantly, our approach has provided with an innovative classification of Rodentia EBRs in a phylogenetic context by the use of the EBA algorithm (section 3.1.3). This algorithm permits to classify EBRs as lineage-specific (EBRs that appeared in a specific-species lineage differentiation, such as mouse lineage-specific EBRs) or clade-specific (EBRs that appeared during a clade differentiation, such as Muroidea clade-specific EBRs). This fact has important implications for the study of genome reshuffling in rodents evolution. The phylogenetic determination of EBRs (i.e., determining when EBRs occurred) can provide valuable information on the specific features that could favor the origin and fixation of CRs. Moreover, further studies focused on the determination of the ancestral Rodentia karyotype at a fine-scale, will help us to gain insights into the directionality of the specific large-scale rearrangements (such as inversions, fusions or translocations) that occurred along Rodentia evolution and

their implication for species differentiation.

5.1.1 Functional constraints

The observation of high rates of genome reshuffling in rodent species lead us to investigate the existence of particular genomic features that might have contributed to this pattern, specifically in the EBRs that lead to the house mouse differentiation. And this group of EBRs include all clade-specific and mouse lineage-specific EBRs. First, we observed that EBRs were not homogeneously distributed across the mouse genome (section 4.1; Figure 2), as they tend to localize in specific mouse chromosomes and in specific regions within these chromosomes. This is the case of chromosome 16, which does not present any EBR. On the contrary, there were several chromosomes with a high number of EBRs. This was the case, for example, of chromosomes 17 and 18 with 11 EBRs, presenting a clustered distribution of EBRs in specific chromosomal regions. Our results add to previous observations of the same pattern in several non-rodent mammalian species (Murphy et al. 2005b; Ruiz-Herrera et al. 2006; Larkin et al. 2009; Farré et al. 2011). The presence of a non-homogeneous distribution of EBRs in the mouse genome led us to investigate the mechanisms responsible of this pattern. Are EBRs physically unstable due to sequence composition and/or genome organization, or do they represent genomic areas where the selection against breakpoints is minimal?

Previous studies comparing human, mouse and rat genomes showed that mouse EBRs are characterized by an enrichment of repetitive sequences, and more specifically SDs and TRs (Armengol et al. 2003; 2005), mirroring what it has been described for primates (Gemayel et al. 2010; Farré et al. 2011). However, evidence of the influence of additional features, such as gene content, was not available mainly due to a lack of a detailed map of EBRs within Rodentia. The delineation of such a map (section 4.1) permitted us to analyze these questions more into detail. We first detected that rodent EBRs (all clade-specific and mouse lineage-specific) presented a higher gene density than to the rest of the mouse genome (section 4.1; Figure 3B), mirroring what has been described in other mammalian species (Murphy et al. 2005b; Elsik et al. 2009; Larkin et al. 2009; Farré et al. 2011; Ullastres et al. 2014). This pattern was accompanied by a decrease in recombination rates (section 4.1.3.3). Based on previous observations (Geraldes et al. 2011; Seehausen et al. 2014; Janoušek et al. 2015), it is plausible that low recombination rates in EBRs could lead to a high genomic differentiation and the fixation of new mutations in genes related to

the species-specific phenotypes, thereby reinforcing the adaptive value of genome reshuffling. In fact, genes under strong selective constraints (i.e., genes located within EBRs or modified as the result of large-scale rearrangements) can confer selective advantage as they also show extensive divergence due to the reduction of recombination (Rieseberg et al. 1999; Yatabe et al. 2007). This has been the case, for example, of *Anopheles* where certain inversions have been associated to ecotypic speciation (Manoukis et al. 2008) such as resistance to insecticides (Mosna et al. 1958), differential mating behavior (Ayala et al. 2011) or adaptation to environmental changes (see Ayala et al. 2014 and references therein). The same pattern has been recently described in *Drosophila* where a specific inversion (3RP) was described to be involved in climate change adaptation in a recent invasion event into Australia (Rane et al. 2015). In this context, our results showing the localization of two gene families (Lcn and Hb gene families) and a gene enrichment cluster (KRAB ZnF genes) within Rodentia EBRs, can be added to such observations.

The Lcn family includes small and highly conserved extracellular proteins that bind and transport various types of ligands. In rodents, they regulate a wide variety of processes including chemical communication, reproduction, immune response, cancer development (Bratt 2000; Chamero et al. 2007; Stopková et al. 2009) and even insulin sensitivity and nutrient metabolism in obesity (Cho et al. 2011; Zhou and Rui 2013). Our analysis detected the Lcn gene family within two EBRs in mouse chromosome 2, located specifically within one Rodentia-specific EBR (containing 5 different Lcn genes) and one mouse-specific EBR (containing 1 Lcn gene). Importantly, these specific genes are mostly expressed within the epididymis (therefore related to reproductive functions) and involved in chemoreception (Suzuki et al. 2004). Strikingly, this gene family has suffered an important expansion during Rodentia evolution due to tandem duplications (Waterston et al. 2002; Suzuki et al. 2004; Grzyb et al. 2006; Stopková et al. 2009; 2014). Thus, the particular structure of duplicated genes could have played a role in the evolutionary instability of this region as it has been previously reported for the β-globin chain in rodents (Hoffmann et al. 2008). In fact, the case of the Hb subunits can also exemplify this phenomenon. The Hb family was localized within a mouse-specific EBRs which contained five genes involved in oxygen metabolism, two of them codifying for the Hb α-chain subunits. Hb genes have been generated by gene duplications during vertebrate evolution (reviewed in Storz et al. 2013)) with different haplotype polymorphisms in house mouse populations (Erhart et al. 1987) characterized by different oxygen binding affinities (D'surney and Popp, 1992). Interestingly, the α-chain is composed by three genes in rat, as a product of tandem duplications, and two genes in mouse, with the

third gene located in other chromosome due to a translocation (Storz et al. 2007), thus suggesting that this region has suffered genome reshuffling during rodents evolution.

Of special relevance is the presence of the enrichment cluster localized within EBRs containing a gene family of DNA binding proteins and transcription factors: the KRAB genes. This gene family is considered the largest of transcription factors in mammals and, although some of them present a clustered distribution across the genome, they also occur individually in specific regions (Waterston et al. 2002). The current view is that the KRAB-ZnF gene family has undergone a massive expansion during tetrapod vertebrates evolution, primarily by tandem duplications (Looman et al. 2002; Hamilton et al. 2003; Urrutia 2003; Shannon et al. 2003) to provide vertebrates with key functions that underlie their development (Urrutia 2003). This phenomenon occurred in a species-specific manner, showing differences in closely related species such as mouse and rat or human and chimpanzee (Tadepally et al. 2008). However, although their structure is well known in different mammalian species (especially in mouse and human), the KRAB-ZnF gene family is still under study as the functions are already not well described, except in few cases where their biochemical functions of its proteins are critical to their cellular roles. These include: cell differentiation and/or proliferation, apopotosis and neoplastic transformation, spermatogenesis and/or sex differentiation, among others (Urrutia 2003; Nowick et al. 2013). In our study we have detected an enrichment cluster within two mouse-specific EBRs and one Muridae-specific EBR, corresponding to three different chromosomes (11, 17 and X) and including seven ZnF genes, most of them are still not well described (such as *Zfp169, Zfp182* and *Zfp300*). Notwithstanding, one of the genes detected is *Prdm9*, a meiosis-specific gene that has been described to be involved in spermatogenesis and have been proposed as the first mammalian hybrid sterility gene (Forejt 1996; Emerson and Thomas 2009; Nowick et al. 2013) (see section 1.2.4). The presence of this gene within an evolutionary unstable region as an EBR, is in consonance with the rapid evolution at a sequence level and the strong positive selection that this gene presents (Nowick et al. 2013).

Overall, we detected that genes located within Rodentia clade-specific EBRs and the mouse lineage-specific EBRs are mostly involved in adaptation to changes of external factors pheromone detection and mating (lipocalins), oxygen metabolism (haemoglobins) and in reproductive isolation (KRAB-ZnF proteins). Therefore, we can hypothesize that the presence of these genes in unstable evolutionary regions could provide selective advantage during the evolutionary history of rodents. Additionally, the low recombination rates described in section 4.1 within EBRs

respect to the rest of the genome, also suggests for an adaptive role of the genes located within these regions as low recombination promotes higher sequence divergence that would help to establish new genic functions to adapt to the environment.

5.1.2 Chromatin structure: a new player in evolutionary genome reshuffling?

More recently, several lines of evidence have suggested the existence of additional factors, independent of the DNA sequence, that are also probably affecting genome plasticity (reviewed in Farré et al. 2015). This is exemplified by the involvement of chromatin conformation changes on the permissiveness of some regions of the genome to undergo chromosomal breakage (Lemaitre et al. 2009; Carbone et al. 2009; Véron et al. 2011; Zhang et al. 2012). Based on these findings and the recent results obtained by our group (Farré et al, 2011 and results from section 4.1), there are evidence suggesting the role of certain properties of local DNA sequences together with the epigenetic state of the chromatin in promoting changes of chromatin organization to open configurations (Farré et al. 2015). In fact, as we described in the introduction (section 1.1.5.4), mammalian genomes are not just a matter of linear DNA sequences; the DNA is packaged into a chromatin structure, the regulation of which depends on several superimposed layers of organization, making the high-order chromatin organization an important player, so far overlooked, in comparative genomic studies. At the molecular level, CRs are thought to result from errors in DSBs repair pathways when simultaneous breaks occur in close proximity within the nucleus (Lupski and Stankiewicz 2005; Meaburn et al. 2007; Korbel et al. 2007; Quinlan et al. 2010; Zhang et al. 2012). In this way, our analysis revealed that EBRs were depleted on cLADs (or regions that are linked to the nuclear envelope) (section 4.1) and thus, suggesting that EBRs tend to localize in the central regions of the nucleus also being open-chromatin regions. It is important to point out that during the preparation of this thesis Berthelot and collaborators (2015) analyzed the genomic distribution of EBRs corresponding to five mammalian and three yeast genomes by employing ancestral genome reconstructions and statistical modeling. Their results are in consonance with our observations. They observed that both chromatin structure (specifically, the open-chromatin regions) together with depletion of coding genes, could explain the functional constrains of the distribution of evolutionary fragile breakage regions within these genomes. In this way, our results provide new data on the relationship of EBRs and the chromatin

conformation, supporting the role of the nuclear structure in the formation of CRs. Further studies that incorporate the study of the epigenetic marks (such as methylations or acetylations of the different histones that integrate the nucleosome structure) and a detailed analysis of selective pressure on the role of adaptation of genes located within EBRs, will provide new insights into the processes underlying genomic rearrangements.

5.2 The Barcelona Rb system as a model for the study of CRs

Since the first models initially proposed to explain the genetic basis of evolution, much has been learned on the role of CRs in population differentiation (White et al. 1967; Rieseberg 2001; Noor et al. 2001; Faria and Navarro 2010). Two hypotheses have been put forward in order to texplain chromosomal speciation: the *hybrid dysfunction model* and the *suppressed recombination model* (see section 1.1.6). Empirical data supporting the *suppressed recombination model* in the case of inversions is abundant in the literature (Greenbaum and Reed 1984; Hale 1986; Rieseberg et al. 1995; 1999; Besansky et al. 2003; Martí and Bidau 2004; Farré et al. 2011; Ullastres et al. 2014). However, evidence for this effect in the case of other types of CRs, such as Rb fusions, is still scarce, especially in natural populations.

Rb fusions are the most widespread CRs described in mammals (Qumsiyeh 1994) including the house mouse. Since 1972, there have been numerous reports of wild-caught house mice with karyotypes containing fewer than 40 chromosomes being the first report focused on mice captured from the "Valle di Poschiavo" in southeastern Switzerland (Gropp et al. 1972). Rb metacentric populations described so far present a similar structure: a metacentric race (with a set of Rb fusions fixed in the population in homozygosis) and an hybrid zone as a product of the contact between the metacentric and standard races (Hauffe et al. 2012). The Barcelona Rb system, the focus of the present study, is an exception to this rule. It presents a striking structure where no metacentric races exists and fused chromosomes are geographically distributed in a staggered fashion leading to a progressive reduction in diploid numbers towards the centre of the distribution area. Thus, Rb fusions are not fixed in the population, being in a polymorphism state that is not the product of the mixing of previous chromosomal races (Gündüz et al. 2001; Medarde et al. 2012). In an evolutionary context, this Rb system presents a unique scenario for the study of the previous state to the fixation of Rb fusions; however, its evolutionary structure

is still not fully understood. In this context, and mainly due to the persistence of chromosomal polymorphisms, the Barcelona Rb system provides an opportunity to understand the origin and evolution of Rb fusions on the previous stage to fixation.

5.2.1 The role of telomere shortening on the origin of Rb fusions

Given the high incidence of Rb fusions in the house mouse, alteration of the telomeric structure of acrocentric chromosomes has been proposed in the literature as a possible trigger for this type of CRs (Bouffler et al. 1996; Kalitsis et al. 2006) (section 1.1.3.4). Acrocentric p-arms of the house mouse (Figure 5.1) are characterized by the presence of highly-repetitive DNA families, ranging from 50Kbp to 150Kbp (Kipling and Cooke 1990). From the terminal p-arm telomeres towards the centromere we can find: (i) a truncated highly conserved minisatellite L1 (tL1), (ii) a region of TRs with a motif of 146bp that varies in length among chromosomes, the TeLoCentric satellite (TLCSat), (iii) a Minisatellite DNA sequence (MinSat) and (iv) a Majorsatellite DNA sequence (MajSat). These two sequences (MinSat and MajSat) contribute to the centromeric and pericentromeric region in all house mouse chromosomes, except the Y chromosome (Kalitsis et al. 2006; Garagna et al. 2014, and references therein).

Figure 5.1. Structure of the p-arm region of the mouse chromosomes (except Y): Centromeric region integrates Minisatellite (MinSat) and Majorsatellite (MajSat) sequences as the binding sites of the centromeric proteins. More distally, TeLoCentric satellite (TLCSat) is followed by minisatellite L1 (tL1) and the telomeric sequences. Image adapted from Garagna and collaborators (2014).

Initial cytogenetic studies focused on the mechanistic factors underlying Rb fusions found that telomeric sequences were completely lost in the centromeric regions of fused chromosomes, although some MinSat DNA sequences were retained in the pericentromeric area (Garagna et al. 1995; Nanda et al. 1995). Later on, these results were supported in wild mice from Italy [corresponding to Poschiavo (2n=26) and Cremona (2n=22) chromosomal races] where the two acrocentric chromosomes

that fuse contribute symmetrically in the amount of MinSat DNA found in the resulting metacentric chromosome (Garagna et al. 2001). Thus, the resulting fused chromosome no only lacks the telomeric sequences but also the tL1, the TLCSat and part of the MinSat sequences (Garagna et al. 2001).

As detailed in the introduction (section 1.1.3.4), three different mechanisms have been proposed to explain the occurrence and fixation of Rb fusions in natural populations in relation to the presence / absence of telomeric sequences in the resultant metacentric chromosome (Slijepcevic 1998): (i) chromosome breakage within minor satellite sequences (ii) telomere shortening, and (iii) telomere inactivation (Figure 1.2). The highly repetitive nature of pericentromeric and telomeric sequences have been proposed as the possible mechanism of the origin of chromosomal breakage within minor satellite sequences (Holmquist and Dancis 1979; Redi and Capanna 1988; Redi et al. 1990; Page et al. 1996) as they would constitute the molecular substrate for a aberrant recombination between non-homologous chromosomes (reviewed in Garagna et al. 2014). As all acrocentric chromosomes share similar DNA sequences (Gazave et al. 2003), the involvement in Rb fusions of chromosome arms should be random in this case, having all mouse chromosomes the same probability to be fused. However, this is not the case in house mouse natural populations. Of the 171 different possible metacentric combinations, 106 have been described in wild populations (Gazave et al. 2003; Pialek et al. 2005). This indicates that not all chromosomes contribute equally to the observed chromosomal variation, begging for the need to explain this bias. Previous studies have suggested that selection against breakpoints occurring in regions that are enriched for housekeeping genes can play a role (Ruiz-Herrera et al. 2010). In this context, we investigated whether telomere shortening could explain the distribution of Rb fusions.

Initial observations in laboratory mice already detected that telomere shortening could be involved in the origin of Rb fusions (Saltman et al. 1993; Blasco et al. 1997; Slijepcevic et al. 1997). In section 4.2 we analyzed, for the first time in the literature, this phenomenon in the Barcelona Rb system (section 4.2). We detected that Rb mice (with diploid numbers ranging from 2n=29 to 2n=39) have, on average, shorter telomere lengths than standard mice (2n=40) (Figure 3C from section 4.2) and that both (Rb and standard wild mice) presented significantly shorter telomere lengths than laboratorystrain mice (Figure 3C from section 4.2). These results, although intriguing, could help us to explain why the wild mice present genome instability in the form of Rb fusions and not laboratory strain mice (as, in principle, they present the same molecular centromeric characteristics), highlighting the importance of conducting more studies focused on natural populations.

Moreover, telomeres from p-arms were, in average, shorter in length that their q-arm counterparts, a phenomenon previously described in laboratory mice (Zijlmans et al. 1997) supporting the observation that p-arms are more prone to lose telomeres than q-arms (Blasco et al. 1997). More interestingly, the proportion of short telomeres (containing less than 15Kbp) was also increased in Rb specimens, which were higher than standard individuals from the Barcelona Rb system (Figure 2B from section 4.2). But more importantly, we observed that standard mice from the Barcelona Rb system have, on average, shorter telomeric lengths but similar percentages of short telomeres than mice from laboratory strains. These results suggest that short telomeres in mice from the Barcelona Rb system could lose their capacity to protect acrocentric chromosomes from fusions (see section 1.1.3.4).

In this context, and, in the light of our results and the available literature, the occurrence of Rb fusions in the house mouse could be the result of a combination of different factors that involve: (i) erosion of telomeric repeats in the p-arms of acrocentric chromosomes and (ii) the presence of MinSat and MajSat repeat sequences at the fusion points (bringing the molecular substrate for the ill repairing of DSBs) (Garagna et al. 2002; 2014) as a result of recombination between non-homologous sequences.

Additionally, the *bouquet* configuration that occurs during prophase I, clusters all telomeres of acrocentric chromosomes together, anchoring them to the inner nuclear envelope (Scherthan et al. 1996). Therefore, this configuration would also promote a physical proximity between pericentromeric regions of acrocentric chromosomes during meiosis, and thus, the interaction between them (that would had previously lost or shorten their telomeric repeats) (Berríos et al. 2014).

Thus, telomere shortening can be considered as one of the candidates to further investigate more into detail the rate of occurrence of Rb fusions in nature. This will add to the models that explain the factors that promote the fixation of CRs dependent on drift, population structure, and the level of the selective disadvantage they confer (see Dobigny et al. 2015 and references therein). In our study we provided with an overall survey of telomere lengths considering the complete set of chromosomes as a whole. Future studies on the analysis of telomere length of specific chromosomes involved in Rb will help us to elucidate to what extent telomere length is triggering the proneness of certain chromosomes to fuse.

5.2.2 The effect of Rb fusions on fertility

Chromosomal rearrangements have been classically considered underdominant mutations associated with deleterious meiotic effects, such as problems originated during chromosome pairing and segregation, leading to subfertility or sterility (King 1993). Therefore, early investigations on the role of Rb fusions in speciation were focused on the effect of this type of CR on fertility, as the *hybrid dysfunction model* proposes (White 1969; 1978). Initial studies analyzed different parameters related with spermatogenic activity in heterokaryotypes for multiple Rb fusions, as they were expected to show subfertility phenotypes (White 1978; Gropp and Winking 1981; Redi and Capanna 1988). These studies, however, obtained contrasting results as in some cases non-disjunction was observed at a very high rates (Gropp and Winking 1981) whereas in other cases no evidence for a severe primary impairment in gametogenesis was found (Redi and Capanna 1988).

Subsequent studies in common shrews analyzing fertility features such as testes weight, Germ Cell Death (GCD) or the structure of seminiferous tubules, observed defective seminiferous tubules and lower testis weights in individuals with multiple Rb fusions in heterozygosis (Garagna et al. 1989). In the case of house mouse, negative effects on fertility have been reported in mice with high number of chromosomes implicated in Rb fusions. For example, a delay in spermatogenesis together with an increase of apoptotic cells was detected in hybrids with 9 and 8 fusions in heterozygosis (from wild-derived strains), suggesting the loss of spermatocytes during the first meiotic division (Redi et al. 1985; Merico et al. 2003; Manterola et al. 2009). However in the case of aneuploidy detection in the germ line, no significant differences were detected in mice with 8 fusions from wild-derived lanoratory strains (Manieu et al. 2014). Notwithstanding, previous studies in specimens from the Barcelona Rb system with up to three fusions in heterozygosis did not detect spermatogenesis impairment (Sans-Fuentes et al. 2010), although the number of Rb fusions and/or level of heterozygosity had an influence in several parameters as the size and shape of the sperm head (Medarde et al. 2013). Moreover, higher rates of GCD were detected in individuals with three Rb fusions in heterozygosis (Sans-Fuentes et al. 2010) or with a high number of metacentrics at the homozygous state (Medarde et al., 2015), although no subfertility was observed (Sans-Fuentes et al. 2010; Medarde et al. 2015). Nevertheless, the geographic distribution of the metacentric chromosomes along the Barcelona Rb system indicated that although spermatogenic alterations could act as partial barriers to gene flow, they are not sufficient to prevent Rb chromosomes. Consequently, and based on the available

data, subfertility effects are related with the number of fusions and the complexity of the meiotic figures (in the case of quadrivalents from monobranchial homologies) (reviewed in Garagna et al. 2014). This is the case, for example, of the hybrids of the metacentric races in Madeira, which present monobranchial homologies and showed a reduction of 50% of fertility in males (Nunes et al. 2011).

Therefore, in the case of natural Rb populations where the distribution of metacentrics is so spread, the presence of heterokaryoptypes with high number of Rb fusions or complex heterozygotes are not expected to be common. And this is in consonance with the limitations of the *hybrid dysfunction model*, as a high frequency of Rb fusions would not occur in natural populations due to their strong underdominant effects. As Dobigny and collaborators (2015) recently exposed, the higher the underdominance of CRs, the more extreme the demographic conditions required for fixation; that is, there is low probability of being recorded. On the contrary, the lower the degree of underdominance associated with a CR, the higher the probability to be present in polymorphic state. The presence of polymorphic Rb fusions in the Barcelona Rb system can be related to this latter case. In this sense, Dobigny and collaborators (2015) distinguished two categories of polymorphisms depending on the selective forces involved: (i) only moderate underdominance is play a role in the process, and (ii) underdominance is compensated by meiotic adaptations (i.e., meiotic drive) or selective advantage (i.e., favorable allelic combinations protected from recombination). In this context, the *suppressed recombination model* presents an alternative or a complement to explain this phenomenon.

5.2.2.1 Rb fusions and meiotic recombination

Are, therefore, Rb fusions affecting the first meiotic division (including the recombination process) and if so, in which degree? In order to shed light into this process, we investigated the effect of Rb fusions in the early stages of meiosis (prophase I) (section 4.3). Initial studies in wild mice populations were performed in the John O'Groats chromosomal races and its corresponding hybrid zone (Wallace et al. 1992; Bidau et al. 2001), revealing a reduction in the number of chiasmata as well as a re-distribution of the remaining chiasmata towards telomeric regions in Rb individuals. In the same line, studies on the metacentric races from Tunisia (2n=22 and 2n=40) (Dumas and Britton-Davidian 2002; Dumas et al. 2015) and Italy (Castiglia and Capanna 2002) reported the same pattern. In both cases, data obtained suggested a decrease of COs due to the presence of Rb fusions. Later on, analysis based

on the detection of MLH1 *foci* in wild derived strains from the Milano II chromosomal races, revealed lower frequency of MLH1 *foci* in both homozygote and heterozygote individuals when compared to all-acrocentric mice (Merico et al. 2003; Dumas et al, 2015). Moreover, studies in pentavalents due to multiple monobranchial homologies on hybrids from Poschiavo and Upper Valtellina populations (Italy) also presented a decrease and re-distribution of COs towards telomeres (Merico et al. 2013).

By the study of mice representative of the Barcelona Rb system, we estimated recombination rates by the analysis of the mean number of MLH1 *foci* per cell. We detected a decrease in the number of COs events in Rb mice when compared to standard animals (Figure 1D from section 4.3). Moreover, we observed that the overall decrease of COs per cell was due to a decrease of MLH1 *foci* within the metacentric chromosomes. We detected less number of metacentrics with multiple MLH1 *foci* (normally 2 MLH1 *foci* per arm) and an increase of the percentage of chromosomal arms with one MLH1 *foci*, when compared to acrocentric chromosomes. Importantly, the frequency of chromosomal arms with the absence of COs was not altered in Rb mice (Figure 1E from section 4.3). Taking together, these results not only indicate a decrease in the number of COs in metacentrics compared to acrocentric chromosomes (in both homozygotes and heterozygotes) but also that, despite this decrease in COs due to the presence of Rb fusions, no important changes have been observed in the percentage of chromosomal arms with one CO. This observation has important implications since it suggest that meiosis is not largely compromised in Rb mice from the Barcelona Rb system, as proper disjunction of chromosomes requires the presence of at least one CO per chromosomal arm (Hassold et al. 2000). Moreover, our analysis of chromosomal-specific recombination maps revealed that metacentric chromosomes presented a re-distribution of COs towards the telomeres when compared to acrocentric chromosomes (in both homozygotes and heterozygotes; Figure 2B-H and supplementary Figure S3 from section 4.3).

Dumas and Britton-Davidian (2002) initially proposed three hypotheses to explain the re-distribution of COs due to the presence of Rb fusions: (i) the loss of centromeric heterochromatin, that is known to buffer recombination suppression effects on adjacent euchromatin (Yamamoto and Miklos 1978; John and King 1985), (ii) the presence of only one kinetochore in the metacentrics that could lead to a reduction of the efficiency of microtubule attachment, and (iii) the loss of telomeric sequences which are involved to the attachment to the inner nuclear membrane leading to a modification of the spatial arrangement of chromosomes in the first meiotic prophase (Scherthan et al. 1996; Garagna et al. 2001). In the moment we initiated this study, empirical evidence on the genetic and mechanistic factors

that were shaping this process were scarce. Therefore, we further studied this phenomenon by analyzing a marker of constitutive heterochromatin, the histone H3 trimethylated at lysine 9 (section 3.4.3). Constitutive heterochromatin located within the centromere has been described to repress recombination in yeast (Ellermeier et al. 2010) and rodents, where a specific gene, the *Smc6*, has been pointed to play a role on this repression (Verver et al. 2013). By using this approach, we could observe a bigger overlapping of constitutive heterochromatin along the SCs of the metacentric chromosomes when compared to the acrocentrics (Figure 3 from section 4.3). These results suggest that centromeric interference in metacentric chromosomes is probably affecting the overall decrease in recombination observed in Rb individuals as Dumas and Britton-Davidian proposed (2002). Strikingly, we have found similar reduction and re-distribution of COs in both homozygotes and heterozygotes supporting previous observations in house mouse (Merico et al. 2003; Merico et al. 2013; Dumas et al. 2015). However, although some correspondence can be observed in the recombination effect between Rb fusions and other type of rearrangement, such as inversions, in this later case, inversions have been seen to reduce recombination only in heterozygotes (Stevison et al. 2011).

Additionally, we also analyzed the pairing dynamics of the bivalents along the first meiotic division in the specimens of the Barcelona Rb system. Previous analysis showed that trivalents can present open configurations in several cases, thus acquiring the same epigenetic configurations of the sex chromosomes in the form of Meiotic Silencing of Unsynapsed Chromosomes (MSUC) (Turner et al. 2005; Baarends et al. 2005; Manterola et al. 2009). MSUC has been proposed to constitute an important factor in triggering meiotic arrest in the case of chromosomal abnormalities (reviewed in Garagna et al. 2014). In the case of Rb fusions Manterola and collaborators (2004) analyzed this mechanism in simple multiple heterozygotes with 8 Rb fusions visualizing a high MSUC signals but no meiotic arrest. In our study we used the γH2AX as a marker for unrepaired DSBs, which is accumulated in latter stages of prophase I in asynapsed regions and the sex body. We analyzed the distribution pattern of γH2AX in mice with the presence of up to 3 trivalents per cell, which presented γH2AX signals mainly in the centromeric regions of trivalents (supplementary Figure S1F from section 4.3). However, although we detected an increase of open trivalents suggesting the persistence of non-repaired DSBs, this did not compromise the correct formation of the sex body (supplementary Figure S1G from section 4.3) allowing, as a consequence, the progression of the cell through prophase I. Therefore, as these cells finally achieve the completion of the correct synapsis, these results would be suggesting a possible delay of the pairing

dynamics of trivalents at pachytene stage. This would be in consonance with the nuclear disposition of the Rb chromosomes during meiosis. In Rb individuals, the pericentromeric region of metacentric chromosomes localize in centric regions of the nucleus in prophase I, whereas pericentromeric regions of acrocentric chromosomes are distributed to the periphery. Therefore the synapsis of the trivalents necessarily involves the movement of the metacentric to the nuclear periphery, leading to a cause of delay in the completion of synapsis (Berríos et al. 2014; reviewed in Garagna et al. 2014).

In the light of our results, we can conclude that, although Rb fusions do not have a strong effect on fertility, they affect meiotic recombination in the Barcelona Rb system, thus supporting the *suppressed recombination model*. Under this scenario, a decrease in meiotic recombination can lead to an increase in genetic divergence within the regions affected, and this effect could lead to the accumulation of genetic incompatibilities and possibly to a genetic isolation between populations (Rieseberg 2001; Noor 2001; Faria and Navarro 2010). The significant morphological (Muñoz-Muñoz et al. 2006; Sans-Fuentes et al. 2009; Muñoz-Muñoz et al. 2011; Martínez-Vargas et al. 2014) and ethological differences (Sans-Fuentes et al. 2005) reported up to now between standard and Rb mice from the Barcelona Rb system could be related with the reduction of gene flow explained by the suppressed recombination model. In fact, the non-deleterious effect of Rb fusions is in consonance with its widespread distribution in the Barcelona Rb system, leading to the discussion on the forces that may favor their positive selection. Some studies have been conducted in order to understand the possible selective advantage of Rb fusions in nature. In this context, Guerrero and Kirkpatrick (2014) proposed a theoretical model that could explain the fixation of Rb fusions by the linkage of locally adapted *loci* in reorganized chromosomes that were previously unlinked when in acrocentric form, conferring a beneficial effect, and thus, facilitating the spread of the fusions within populations. Thus, following this model, the fixation of chromosomal fusions in nature would involve local adaptation of specific genes. However, to this extend, if Rb fusions confer a selective advantage remains unclear at this stage.

5.2.2.2. *Prdm9* and the study of chromosomal evolution

At this point, the description of the genetic factors that could be playing an role on the distribution of recombination events has been taking force in the recent years. *Prdm9* has been described to be involved in the speciation process as a hybrid

sterility gene (Mihola et al. 2009; Flachs et al. 2012; Bhattacharyya et al. 2013) as its activity during meiosis involves the recruitment of the DSBs repair machinery (Baudat et al. 2010; Myers et al. 2010; Parvanov et al. 2010; Grey et al. 2011; Billings et al. 2013). Thus, changes in the ZnF sequence are translated into a re-distribution of recombination sites (Brick et al. 2012; Nowick et al. 2013). In this context, the understanding of its sequence variability and the forces that are driving its rapid evolution can bring valuable information on its role in the speciation process.

Here, and with the aim to elucidate the influence of genetic factors, such as *Prdm9*, on recombination, we analyzed the *Prdm9* allelic background in the Barcelona Rb system (section 4.3). We conducted the most comprehensive study of *Prdm9* variability in a single natural population of house mouse. We detected novel house mouse variants in the Barcelona Rb system (*10A, 10B, 11B, 12B* and *12C*) that differed in both the number of ZnF repeats (ranging from 10 to 12 repeats) and their sequence (describing 21 different repeat sequences). Thus, these results support the highly polymorphic nature of this gene (Oliver et al. 2009; Nowick et al. 2013) (Supplementary Figure S6A from section 4.3). Strikingly, the distribution of the different alleles presented a specific pattern, where the higher number of different alleles was observed in standard mice (2n=40, with 5 different alleles in two different locations). Rb mice, on the other hand, presented more sequence homogeneity, showing almost 95% frequency of the allele *10A* (2n=28-39, with two alleles in 5 different locations of the population). In fact, we observed an homogeneous presence of this allele, *10A*, in all Rb populations (supplementary Figure S6B from section 4.3). During the development of this work, two recent publications (Buard et al. 2014; Kono et al. 2014) described the *Prdm9* allelic diversity from almost 100 mice distributed across all continents (except America). This analysis permitted to observe the unique allelic characteristics of the Barcelona Rb system, as only two alleles (*11B* and *12B*) were shared with the populations described in these surveys. Importantly, the observation that the allele *10A* is present in all populations within the Rb system and it is not shared with other mouse populations analyzed, suggests a possible *in situ* origin of this allele in the Barcelona Rb system.

Moreover, our results revealed a striking correlation between recombination rates and the number of *Prdm9* ZnF, meaning that individuals with less number of ZnF repeats (10) presented less recombination rates than mice with more ZnF repeats (supplementary Figure S6C from section 4.3). These results suggest a possible role of *Prdm9* gene in the distribution of recombination sites in natural populations, supporting previous data in laboratory strain specimens (Baudat and de Massy 2007; Brick et al. 2012; Billings et al. 2013; Baker et al. 2015). It might seem

contradictory that mice with less number of ZnF repeats (*10A*) and thus, low rates of recombination could be selected in the population. However, we have to take into account that the widespread distribution of allele *10A* in Rb individuals also suggests the possibility that this specific allele would have developed different recognition sites that maybe would reflect the re-distribution of recombination sites observed in our MLH1 analysis.

Which are, then, the factors responsible for the widespread distribution of the *10A* allele in the Barcelona Rb system if it appears that mice owing the *10A* allele have, on average, less COs than mice with a different *Prdm9* allelic composition? It is possible that the *10A* allele is somehow advantageous for the population. For example, that this new allele has evolved rapidly in order to recognize with strongest affinity new DNA motifs. As we explained in the introduction (section 1.2.4), a process of biased gene conversion occurs during the repairing of COs at prophase I, giving as a consequence, changes in the sequence of the hotspots that would undergo the self-destruction of the recombination recognition sites (Boulton et al. 1997). In this way, the rapid evolution of the ZnF sequence would lead to rapid changes in the distribution of the recombination sites by recognizing with stronger affinity new DNA motifs that would compensate the loss of the primary hotspots by gene conversion. Additionally, *Prdm9* has been described to be under strong positive selection in both primate and mouse populations, resulting in many different alleles that contain substitutions in the ZnF domains at each of the three amino-acid positions that control DNA binding specificity (Thomas et al. 2009; Oliver et al. 2009; Buard et al. 2014; Schwartz et al. 2014; Kono et al. 2014). Therefore, *Prdm9 10A* allele could be somehow advantageous for mice in the Rb system, thus facilitating its selection and spread. This phenomenon could be explained, as the lack of COs around the centromere observed in Rb mice is probably the result of mechanistic disturbances during early meiosis. Therefore, it is possible that new alleles with stronger affinity to new motifs distributed towards more telomeric regions, have been developed in Rb mice in order to assure the minimum number of COs necessary to proper meiosis progression.

However, on the other hand, we cannot discard the possibility that the pattern observed in *Prdm9* allelic frequencies within the Barcelona Rb system could be the result of its specific population genetic structure, yet to be described into detail. As it has been suggested, the Barcelona Rb system is a particular Rb scenario that has been originated by primary intergradation, although secondary contact cannot be excluded (Gündüz et al. 2001; Gündüz et al. 2010). Therefore, taking into account the rapid change of ZnF sequences (Thomas et al. 2009; Nowick et al. 2013; Pratto et al.

2014b), it is plausible that older populations (2n=40) would present higher number of alleles, whereas more recent populations (Rb mice) would present a homogeneous distribution with predominance allele *10A* due to a foundational effect.

In the light of our results, therefore, further studies at a finer scale of the population structure of the Barcelona Rb system are needed in order to elucidate its role on the distribution of the *Prdm9* alleles across the population. However, future surveys on the analysis of the specific DNA motifs recognized by allele *10A* and its distribution across the genome would bring inestimable information on the characteristics of this allele that favored its widespread distribution within the Barcelona Rb system.

6.

CONCLUSIONS

1. We have reconstructed a detailed map of the genomic regions involved in evolutionary reshuffling in rodents. This has permitted the delineation of 3,392 Homologous Synteny Blocks, 3,142 Syntenic Fragments and 1,333 Evolutionary Breakpoint Regions among eleven mammalian species, including six rodent species (*Jaculus jaculus, Spalax galilii, Microtus ochorgaster, Rattus norvegicus* and *Mus musculus*). Importantly, rodents present an increased rate of genome reshuffling (estimated as the number of EBRs per Myr) when compared to the mammalian species included in this study (*Homo sapiens, Macaca mulatta, Pongo pygmaeus, Felis catus, Bos taurus* and *Equus caballus*).
2. Evolutionary Breakpoint Regions are not homogenously distributed across the mouse genome, as they tend to localize in specific mouse chromosomes and in specific chromosomal regions within these chromosomes. This genomic distribution is conditioned by specific genomic constrains (as gene enrichment and lower recombination rates) and to the nuclear chromatin structure (being located in central open-chromatin regions).
3. Rodentia clade-specific evolutionary breakpoint regions and the mouse lineage-specific evolutionary breakpoint regions are enriched for specific gene functions such as pheromone detection and mating (lipocalins), oxygen metabolism (haemoglobins) and in reproductive isolation (Krueppel-Associated Box Zinc finger proteins). In this context, EBRs can present opportunities for the development of novel functions that may promote the adaptation of species.
4. Robertsonian mice from the Barcelona Rb system have, on average, shorter telomeric lengths (in both p- and q-arms) and a higher percentage of short telomeres (only in the p-arms) than standard individuals. Moreover, standard mice have, on average, shorter telomeric lengths but similar percentages of short telomeres than mice from laboratory strains. This indicates that telomere shortening in the form of critically short telomeres is probably playing a role in the occurrence of Rb fusions in the Barcelona Rb system.
5. Robertsonian mice from the Barcelona Rb system present a significant reduction in the number of total MLH1 *foci* per cell analyzed. This reduction was due to a decrease in the number of total MLH1 *foci* detected in the metacentric chromosomes.
6. Metacentric chromosomes present a re-distribution of MLH1 *foci* towards the telomeric regions of the chromosomes, being the first crossover localized further from the centromere, when compared to acrocentric chromosomes. The construction of chromosome-specific recombination maps indicated that

this re-distribution of MLH1 *foci* in metacentric chromosomes was observed either in homozygosis (bivalents) or heterozygosis (trivalents) state.

7. The analysis of the distribution of the histone H3 trimethylated at lysine 9, an epigenetic signal for constitutive heterochromatin, indicated that the re-distribution of MLH1 *foci* in metacentric chromosomes was caused, most probably, due to an interference effect of the centromeric heterochromatin that generates a higher suppression of recombination in the adjacent euchromatin.
8. Robertsonian fusions alter the distribution of the γH2AX, a marker of non-repaired double strand breaks, in the autosomes. We detect a higher frequency of both bivalents and trivalents with asynapsed regions, indicating the persistence of regions with non-repaired DSBs through pachynema. This situation, however, did not affect the normal progression of the meiotic silencing of asynapsed chromatin that characterized sex chromosomes as we detected that the sex body showed positive signals for γH2AX in Rb mice.
9. The *Prdm9* gene presents a high allelic diversity within the different locations sampled in the Barcelona Rb system, differing in both the number and the sequence zinc finger repeats. In fact, five new *Prdm9* allelic variants ranging from 10 to 12 zinc finger repeats were detected.
10. Both the allele frequency and distribution of *Prdm9* varied among specimens and localities sampled. Standard mice presented the highest allelic diversity whereas Rb mice showed a predominant presence of the allele *10A*, which were found in all the locations sampled within the population. Additionally, the allele *10A* presents a specific zinc finger repeat sequence when compared to the rest of the mouse populations described so far, thus reinforcing the particularity of the Barcelona Rb system.
11. The number of zinc finger repeats was correlated with recombination rates. Thus, *Prdm9* could be a genetic factor that plays a role in the re-distribution of COs within the Barcelona Rb system.
12. We propose that both mechanistic (meiotic disturbances affected by the presence of Rb fusions), and genetic (*Prdm9*) factors would be playing a role in the suppression of recombination observed in the Barcelona Rb system.

7.

BIBLIOGRAPHY

Al-Jaru a, Goodwin W, Skidmore J, Khazanehdari K (2014) Distribution of MLH1 *foci* in horse male synaptonemal complex. Cytogenet Genome Res 142:87–94.

Alekseyev M a, Pevzner P a (2010) Comparative genomics reveals birth and death of fragile regions in mammalian evolution. Genome Biol 11:R117.

Allers T, Lichten M (2001) Differential Timing and Control of Noncrossover and Crossover Recombination during Meiosis. Cell 106:47–57.

Allsopp RC, Vaziri H, Patterson C, et al (1992) Telomere length predicts replicative capacity of human fibroblasts. Proc Natl Acad Sci 89:10114–8.

Anderson LK, Reeves A, Webb LM, Ashley T (1999) Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localization of MLH1 protein. Genetics 151:1569–79.

Aquino CI, Abril V V, Duarte JMB (2013) Meiotic pairing of B chromosomes, multiple sexual system, and Robertsonian fusion in the red brocket deer Mazama americana (Mammalia, Cervidae). Genet Mol Res 12:3566–74.

Armengol L, Marquès-Bonet T, Cheung J, et al (2005) Murine segmental duplications are hot spots for chromosome and gene evolution. Genomics 86:692–700.

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

Armour JAL (2006) Tandemly repeated DNA: why should anyone care? Mutat Res 598:6–14.

Arnheim N, Calabrese P, Nordborg M (2003) Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved. Am J Hum Genet 73:5–16.

Ashley T, Moses MJ, Solari AJ (1981) Fine structure and behaviour of a pericentric inversion in the sand rat, *Psammomys obesus*. J Cell Sci 50:105–19.

Attie O, Darling AE, Yancopoulos S (2011) The rise and fall of breakpoint reuse depending on genome resolution. BMC Bioinformatics 12 Suppl 9:S1.

Auffray JC, Marshall JT, Thaler L, Bonhomme F (1990) Focus on the nomenclature of European species of Mus. Mouse Genome 88:7–8.

Auton A, Fledel-Alon A, Pfeifer S, et al (2012) A fine-scale chimpanzee genetic map from population sequencing. Science 336:193–8.

Axelsson E, Webster MT, Ratnakumar A, et al (2012) Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. Genome Res 22:51–63.

Ayala D, Fontaine MC, Cohuet A, et al (2011) Chromosomal inversions, natural selection and adaptation in the malaria vector *Anopheles funestus*. Mol Biol Evol 28:745–58.

Ayala D, Ullastres A, González J (2014) Adaptation through chromosomal inversions in *Anopheles*. Front Genet 5:129.

Azzalin CM, Reichenbach P, Khoriauli L, et al (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318:798–801.

Baarends WM, Wassenaar E, van der Laan R, et al (2005) Silencing of unpaired chromatin and histone H2A ubiquitination in mammalian meiosis. Mol Cell Biol 25:1041–53.

Backström N, Forstmeier W, Schielzeth H, et al (2010) The recombination landscape of the zebra finch *Taeniopygia guttata* genome. Genome Res 20:485–95.

Bacolla A, Larson JE, Collins JR, et al (2008) Abundance

and length of simple repeats in vertebrate genomes are determined by their structural properties. Genome Res 18:1545–53.

Bailey JA, Church DM, Ventura M, et al (2004) Analysis of segmental duplications and genome assembly in the mouse. Genome Res 14:789–801.

Bailey JA, Eichler EE (2006) Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet 7:552–64.

Baker CL, Kajita S, Walker M, et al (2015) PRDM9 Drives Evolutionary Erosion of Hotspots in *Mus musculus* through Haplotype-Specific Initiation of Meiotic Recombination. PLoS Genet 11:e1004916.

Baker RJ, Bickham JW (1986) Speciation by monobrachial centric fusions. Proc Natl Acad Sci 83:8245–8.

Baker SM, Plug AW, Prolla TA, et al (1996) Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. Nat Genet 13:336–42.

Ballenger L (1999) "*Mus musculus*, animal diversity" web. http://www.biokids.umich.edu/accounts/Mus_musculus/. Web page.

Band MR, Larson JH, Rebeiz M, et al (2000) An ordered comparative map of the cattle and human genomes. Genome Res 10:1359–68.

Barski A, Cuddapah S, Cui K, et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–37.

Basheva EA, Torgasheva AA, Gomez Fernandez MJ, et al (2014) Chromosome synapsis and recombination in simple and complex chromosomal heterozygotes of tuco-tuco (*Ctenomys talarum*: Rodentia: Ctenomyidae). Chromosome Res 22:351–63.

Basset P, Yannic G, Brünner H, Hausser J (2006) Restricted gene flow at specific parts of the shrew genome in chromosomal hybrid zones. 60:1718–1730.

Bateson W (1909) Heredity and variation in modern lights. In: Darwin and modern science. Cambridge University Press, Cambridge, pp 85–101.

Baudat F, de Massy B (2007) Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Res 15:565–77.

Baudat F, Buard J, Grey C, et al (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327:836–40.

Baudet C, Lemaitre C, Dias Z, et al (2010) Cassis: detection of genomic rearrangement breakpoints. Bioinformatics 26:1897–8.

Becker TS, Lenhard B (2007) The random versus fragile breakage models of chromosome evolution: a matter of resolution. Mol Genet Genomics 278:487–91.

Beckmann JS, Estivill X, Antonarakis SE (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet 8:639–46.

Berchowitz LE, Copenhaver GP (2010) Genetic interference: don't stand so close to me. Curr Genomics 11:91–102.

Berg IL, Neumann R, Sarbajna S, et al (2011) Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. Proc Natl Acad Sci 108:12378–83.

Bernardi G (2000) Isochores and the evolutionary genomics of vertebrates. Gene 241:3–17.

Berríos S, Manieu C, López-Fenner J, et al (2014) Robertsonian chromosomes and the nuclear architecture of mouse meiotic prophase spermatocytes. Biol Res 47:16.

Berthelot C, Muffato M, Abecassis J, Roest Crollius H (2015) The 3D Organization of Chromatin Explains Evolutionary Fragile Genomic Regions. Cell Rep

10:1913–1924.

Besansky NJ, Krzywinski J, Lehmann T, et al (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. Proc Natl Acad Sci 100:10818–23.

Bhattacharyya T, Gregorova S, Mihola O, et al (2013) Mechanistic basis of infertility of mouse intersubspeci fic hybrids. Proc Natl Acad Sci 110:E468-77

Bidau CJ, Giménez MD, Palmer CL, Searle JB (2001) The effects of Robertsonian fusions on chiasma frequency and distribution in the house mouse (*Mus musculus domesticus*) from a hybrid zone in northern Scotland. Heredity 87:305–13.

Billings T, Parvanov ED, Baker CL, et al (2013) DNA binding specificities of the long zinc-finger recombination protein PRDM9. Genome Biol 14:R35.

Birney E, Stamatoyannopoulos JA, Dutta A, et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799–816.

Bishop DK, Zickler D (2004) Early Decision : Meiotic Crossover Interference prior to Stable Strand Exchange and Synapsis. Cell 117:9–15.

Blasco MA, Lee H-W, Hande MP, et al (1997) Telomere Shortening and Tumor Formation by Mouse Cells Lacking Telomerase RNA. Cell 91:25–34.

Borodin PM, Karamysheva T V, Belonogova NM, et al (2008) Recombination map of the common shrew, *Sorex araneus* (Eulipotyphla, Mammalia). Genetics 178:621–32.

Bouffler SD, Morgan WF, Pandita TK, Slijepcevic P (1996) The involvement of telomeric sequences in chromosomal aberrations. Mutat Res Genet Toxicol 366:129–135.

Boulton A, Myers RS, Redfield RJ (1997) The hotspot conversion paradox and the evolution of meiotic recombination. Proc Natl Acad Sci 94:8058–63.

Bourque G, Pevzner PA (2002) Genome-Scale Evolution: Reconstructing Gene Orders in the Ancestral Species. Genome Res 12:26–36.

Bourque G, Pevzner PA, Tesler G (2004) Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. Genome Res 14:507–16.

Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F (1993) The Evolution of House Mice. Annu Rev Ecol Syst 24:119–152.

Bratt T (2000) Lipocalins and cancer. Biochim Biophys Acta - Protein Struct Mol Enzymol 1482:318–326.

Brick K, Smagulova F, Khil P, et al (2012) Genetic recombination is directed away from functional genomic elements in mice. Nature 485:642–5.

Britton-Davidian J, Catalan J, Belkhir K (2002) Chromosomal and allozyme analysis of a hybrid zone between parapatric Robertsonian races of the house mouse: a case of monobrachial homology. Cytogenet Genome Res 96:75–84.

Britton-Davidian J, Catalan J, da Graça Ramalhinho M, et al (2005) Chromosomal phylogeny of Robertsonian races of the house mouse on the island of Madeira: testing between alternative mutational processes. Genet Res 86:171–83.

Broman KW, Murray JC, Sheffield VC, et al (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861–9.

Broman KW, Rowe LB, Churchill GA, Paigen K (2002) Crossover interference in the mouse. Genetics 160:1123–31.

Brown JD, O'Neill RJ (2010) Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. Annu Rev Genomics Hum Genet 11:291–316.

Brown KM, Burk LM, Henagan LM, Noor MAF (2004) A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura*. Evolution 58:1856–1860.

Brunschwig H, Levi L, Ben-David E, et al (2012) Fine-scale maps of recombination rates and hotspots in the mouse genome. Genetics 191:757–64.

Bruschi DP, Rivera M, Lima AP, et al (2014) Interstitial Telomeric Sequences (ITS) and major rDNA mapping reveal insights into the karyotypical evolution of Neotropical leaf frogs species (Phyllomedusa, Hylidae, Anura). Mol Cytogenet 7:22.

Buard J, Barthès P, Grey C, de Massy B (2009) Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. EMBO J 28:2616–24.

Buard J, de Massy B (2007) Playing hide and seek with mammalian meiotic crossover hotspots. Trends Genet 23:301–9.

Buard J, Rivals E, Dunoyer de Segonzac D, et al (2014) Diversity of *Prdm9* zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite. PLoS One 9:e85021.

Burma S, Chen BP, Murphy M, et al (2001) ATM phosphorylates histone H2AX in response to DNA double-strand breaks. J Biol Chem 276:42462–7.

Butlin RK (2005) Recombination and speciation. Mol Ecol 14:2621–35.

Caetano a. R (1999) A Comparative Gene Map of the Horse (*Equus caballus*). Genome Res 9:1239–1249.

Campuzano V, Montermini L, Moltò MD, et al (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271:1423–7.

Capanna E, Redi CA (1995) Whole-arm reciprocal translocation (WART) between Robertsonian chromosomes: finding of a Robertsonian heterozygous mouse with karyotype derived through WARTs. Chromosome Res 3:135–7.

Capozzi O, Carbone L, Stanyon RR, et al (2012) A comprehensive molecular cytogenetic analysis of chromosome rearrangements in gibbons. Genome Res 22:2520–8.

Capy P (1998) Evolutionary genomics. A plastic genome. Nature 396:522-3

Carbone L, Harris RA, Gnerre S, et al (2014) Gibbon genome and the fast karyotype evolution of small apes. Nature 513:195–201.

Carbone L, Vessere GM, ten Hallers BFH, et al (2006) A high-resolution map of synteny disruptions in gibbon and human genomes. PLoS Genet 2:e223.

Carbone L, Harris RA, Vessere GM, et al (2009) Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. PLoS Genet 5:e1000538.

Casacuberta E, González J (2013) The impact of transposable elements in environmental adaptation. Mol Ecol 22:1503–17.

Caspersson T, Zech L, Johansson C (1970) Analysis of human metaphase chromosome set by aid of DNA-binding fluorescent agents. Exp Cell Res 62:490–492.

Castiglia R, Capanna E (2002) Chiasma repatterning across a chromosomal hybrid zone between chromosomal races of *Mus musculus domesticus*. Genetica 114:35–40.

Castiglia R, Capanna E (2000) Contact zone between chromosomal races of *Mus musculus domesticus*. 2. Fertility and segregation in laboratory-reared and wild mice heterozygous for multiple Robertsonian rearrangements. Heredity 85:147–156.

Chai X, Nagarajan S, Kim K, et al (2013) Regulation of the boundaries of accessible chromatin. PLoS Genet 9:e1003778.

Chamero P, Marton TF, Logan DW, et al (2007) Identification of protein pheromones that promote aggressive behaviour. Nature 450:899–902.

Chang D, Duda TF (2012) Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. Mol Biol Evol 29:2019–29.

Cho KW, Zhou Y, Sheng L, Rui L (2011) Lipocalin-13 regulates glucose metabolism by both insulin-dependent and insulin-independent mechanisms. Mol Cell Biol 31:450–7.

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

Clark AG, Wang X, Matise T (2010) Contrasting methods of quantifying fine structure of human recombination. Annu Rev Genomics Hum Genet 11:45–64.

Clément Y, Arndt PF (2013) Meiotic recombination strongly influences GC-content evolution in short regions in the mouse genome. Mol Biol Evol 30:2612–8.

Clemente IC, Ponsà M, García M, Egozcue J (1990) Evolution of the Simiiformes and the phylogeny of human chromosomes. Hum Genet 84:493–506.

Cohen PE, Holloway JK (2010) Predicting gene networks in human oocyte meiosis. Biol Reprod 82:469–72.

Cole F, Kauppi L, Lange J, et al (2012) Homeostatic control of recombination is implemented progressively in mouse meiosis. Nat Cell Biol 14:424–30.

Collins I, Newlon CS (1994) Meiosis-specific formation of joint DNA molecules containing sequences from homologous chromosomes. Cell 76:65–75.

Contreras LC, Torres-Mura JC, Spotorno AE, Walker LI (1994) Chromosomes of *Octomys mimax* and *Octodontomys gliroides* and Relationships of Octodontid Rodents. J Mammal 75:768–774.

Coop G, Myers SR (2007) Live hot, die young: transmission distortion in recombination hotspots. PLoS Genet 3:e35.

Coop G, Przeworski M (2007) An evolutionary view of human recombination. Nat Rev Genet 8:23–34.

Cordaux R, Batzer M a (2009) The impact of retrotransposons on human genome evolution. Nat Rev Genet 10:691–703.

Cox DR, Burmeister M, Price ER, et al (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250:245–50.

Cox A, Ackert-Bicknell CL, Dumont BL, et al (2009) A new standard genetic map for the laboratory mouse. Genetics 182:1335–44.

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland.

Cremer M, Grasser F, Lanctôt C, et al (2008) Multicolor 3D fluorescence *in situ* hybridization for imaging interphase chromosomes. Methods Mol Biol 463:205–39.

Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2:292–301.

Crouau-Roy B, Service S, Slatkin M, Freimer N (1996) A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. Hum Mol Genet 5:1131–7.

Darwin CR (1859) On the Origin of the Species. P.F. Collier & Son,New York.

De Boer E, Stam P, Dietrich AJJ, et al (2006) Two levels of interference in mouse meiotic recombination. Proc Natl Acad Sci 103:9607–12.

De Castro E, Soriano I, Marín L, et al (2012) Nucleosomal organization of replication origins and meiotic recombination hotspots in fission yeast. EMBO J 31:124–37.

De Grouchy J, Turleau C, Roubin M, Klein M (1972) Karyotypic evolution in man and chimpanzees. A comparative study of band topographies after controlled denaturation. Ann génétique 15:79–84.

De Lange T (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev 19:2100–10.

De Massy B (2003) Distribution of meiotic recombination sites. Trends Genet 19:514–22.

De Oliveira EHC, Tagliarini MM, Rissino JD, et al (2010) Reciprocal chromosome painting between white hawk (*Leucopternis albicollis)* and chicken reveals extensive fusions and fissions during karyotype evolution of accipitridae (Aves, Falconiformes). Chromosome Res 18:349–55.

Deininger PL, Batzer MA (1999) Alu Repeats and Human Disease. Mol Genet Metab 67:183–193.

Dekker J, Marti-Renom MA, Mirny LA (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet 14:390–403.

Dekker J (2014) Two ways to fold the genome during the cell cycle: insights obtained with chromosome conformation capture. Epigenetics Chromatin 7:25.

Delprat A, Negre B, Puig M, Ruiz A (2009) The transposon Galileo generates natural chromosomal inversions in *Drosophila* by ectopic recombination. PLoS One 4:e7883.

Di-Nizo CB, Ventura K, Ferguson-Smith MA, et al (2015) Comparative chromosome painting in six species of *Oligoryzomys* (Rodentia, Sigmodontinae) and the karyotype evolution of the genus. PLoS One 10:e0117579.

Diez-Villanueva, A Malinverni R, Gel B (2015) regioneR: Association analysis of genomic regions based on permutation tests. R package version 1.0.3.

Dixon JR, Selvaraj S, Yue F, et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376–80.

Dobigny G, Britton-Davidian J, Robinson T (2015) Chromosomal polymorphism in mammals: an evolutionary perspective. Biol Rev Camb Philos Soc. On-line.

Dobzhansky T (1937) Genetics and the Origin of Species. Columbia University Press, New York.

Dobzhansky T (1951) Genetics and the Origin of Species, 3rd edn. Columbia University Press, New York.

Dobzhansky T, Sturtevant AH (1938) Inversions in the Chromosomes of *Drosophila Pseudoobscura*. Genetics 23:28–64.

Donthu R, Lewin H a, Larkin DM (2009) SyntenyTracker: a tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. BMC Res Notes 2:148.

Dos Reis M, Inoue J, Hasegawa M, et al (2012) Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc Biol Sci 279:3491–500.

Duda TF, Palumbi SR (1999) Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod Conus. Proc Natl Acad Sci 96:6820–6823.

Dumas D, Britton-Davidian J (2002) Chromosomal rearrangements and evolution of recombination: comparison of chiasma distribution patterns in standard and robertsonian populations of the house mouse. Genetics 162:1355–66.

Dumas D, Catalan J, Britton-Davidian J (2015) Reduced recombination patterns in Robertsonian hybrids between chromosomal races of the house mouse:

chiasma analyses. Heredity 114:56–64.

Dumont BL, Payseur B a (2008) Evolution of the genomic rate of recombination in mammals. Evolution 62:276–94.

Dumont BL, Payseur B a (2011) Genetic analysis of genome-scale recombination rate evolution in house mice. PLoS Genet 7:e1002116.

Dumont BL, White M a, Steffy B, et al (2011) Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps. Genome Res 21:114–25.

Dutrillaux B, Rethoré MO, Lejeune J (1975) Comparison of the karyotype of the orangutan (*Pongo pygmaeus*) to those of man, chimpazee, and gorilla. Ann génétique 18:153–61.

Dzur-Gejdosova M, Simecek P, Gregorova S, et al (2012) Dissecting the genetic architecture of F1 hybrid sterility in house mice. Evolution 66:3321–35.

Egozcue J, Aragonés J, Caballín MR, Goday C (1973a) Banding patterns of the chromosomes of man and gorilla. Ann génétique 16:207–10.

Egozcue J, Caballìn MR, Goday C (1973b) Banding patterns of the chromosomes of man and the chimpanzee. Humangenetik 18:77–80.

Eisenbarth I (2001) Long-range sequence composition mirrors linkage disequilibrium pattern in a 1.13 Mb region of human chromosome 22. Hum Mol Genet 10:2833–2839.

Elferink MG, van As P, Veenendaal T, et al (2010) Regional differences in recombination hotspots between two chicken populations. BMC Genet 11:11.

Ellermeier C, Higuchi EC, Phadnis N, et al (2010) RNAi and heterochromatin repress centromeric meiotic recombination. Proc Natl Acad Sci 107:8701–5.

Elliott B, Richardson C, Jasin M (2005) Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. Mol Cell 17:885–94. doi: 10.1016/j.molcel.2005.02.028

Elsik CG, Tellam RL, Worley KC, et al (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324:522–8.

Emerson RO, Thomas JH (2009) Adaptive evolution in zinc finger transcription factors. PLoS Genet 5:e1000325.

Fachinetti D, Folco HD, Nechemia-Arbely Y, et al (2013) A two-step mechanism for epigenetic specification of centromere identity and function. Nat Cell Biol 15:1056–66.

Fang X, Nevo E, Han L, et al (2014) Genome-wide adaptive complexes to underground stresses in blind mole rats Spalax. Nat Commun 5:3966.

Faria R, Navarro A (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends Ecol Evol 25:660–9.

Faria R, Neto S, Noor M, Navarro A (2011) Role of Natural Selection in Chromosomal Speciation. eLS

Farré M, Bosch M, López-Giráldez F, et al (2011) Assessing the role of tandem repeats in shaping the genomic architecture of great apes. PLoS One 6:e27239.

Farré M, Micheletti D, Ruiz-Herrera A (2013) Recombination rates and genomic shuffling in human and chimpanzee-a new twist in the chromosomal speciation theory. Mol Biol Evol 30:853–64.

Farré M, Robinson TJ, Ruiz-Herrera A (2015) An Integrative Breakage Model of genome architecture, reshuffling and evolution: The Integrative Breakage Model of genome evolution, a novel multidisciplinary hypothesis for the study of genome plasticity. Bioessays 37:479–88.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. Trends Genet 28:342–50.

Feder JL, Nosil P (2009) Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? Evolution 63:3061–75.

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–68.

Fischer G, James SA, Roberts IN, et al (2000) Chromosomal evolution in Saccharomyces. Nature 405:451–4.

Flachs P, Mihola O, Simeček P, et al (2012) Interallelic and intergenic incompatibilities of the *Prdm9* (*Hst1*) gene in mouse hybrid sterility. PLoS Genet 8:e1003044.

Ford CE, Hamerton JL (1956) The Chromosomes of Man. Nature 178:1020–1023. doi: 10.1038/1781020a0

Forejt J, Vincek V, Klein J, et al (1991) Genetic mapping of the t-complex region on mouse chromosome 17 including the Hybrid sterility-1 gene. Mamm Genome 1:84–91.

Forejt J (1996) Hybrid sterility in the mouse. Trends Genet 12:412–7.

Forment J V, Kaidi A, Jackson SP (2012) Chromothripsis and cancer: causes and consequences of chromosome shattering. Nat Rev Cancer 12:663–70.

Franchini P, Colangelo P, Solano E, et al (2010) Reduced gene flow at pericentromeric loci in a hybrid zone involving chromosomal races of the house mouse *Mus musculus domesticus*. Evolution 64:2020–32.

Frit P, Barboule N, Yuan Y, et al (2014) Alternative end-joining pathway(s): bricolage at DNA breaks. DNA Repair 17:81–97.

Froenicke L, Anderson LK, Wienberg J, Ashley T (2002) Male Mouse Recombination Maps for Each Autosome Identified by Chromosome Painting. Am J Hum Genet 71:1353–1368.

Frönicke L, Wienberg J, Stone G, et al (2003) Towards the delineation of the ancestral eutherian genome organization: comparative genome maps of human and the African elephant (*Loxodonta africana*) generated by chromosome painting. Proc Biol Sci 270:1331–40.

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

Froenicke L, Caldés MG, Graphodatsky A, et al (2006) Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? Genome Res 16:306–10.

Garagna S, Broccoli D, Redi CA, et al (1995) Robertsonian metacentrics of the house mouse lose telomeric sequences but retain some minor satellite DNA in the pericentromeric area. Chromosoma 103:685–92.

Garagna S, Marziliano N, Zuccotti M, et al (2001) Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. Proc Natl Acad Sci 98:171–5.

Garagna S, Page J, Fernandez-Donoso R, et al (2014) The Robertsonian phenomenon in the house mouse: mutation, meiosis and speciation. Chromosoma 123:529–44.

Garagna S, Zuccotti M, Capanna E, Redi C a (2002) High-resolution organization of mouse telomeric and pericentromeric DNA. Cytogenet Genome Res 96:125–9.

Garagna S, Zuccotti M, Searle JB, et al (1989) Spermatogenesis in heterozygotes for Robertsonian chromosomal rearrangements from natural populations of the common shrew, *Sorex araneus.* J Reprod Fertil 87:431–8.

García F, Nogués C, Ponsà M, et al (2000) Chromosomal homologies between humans and *Cebus apella* (Primates) revealed by ZOO-FISH. Mamm Genome

11:399–401.

Garcia-Cruz R, Pacheco S, Brieño MA, et al (2011) A comparative study of the recombination pattern in three species of Platyrrhini monkeys (primates). Chromosoma 120:521–30.

Gauthier P, Hima K, Dobigny G (2010) Robertsonian fusions, pericentromeric repeat organization and evolution: a case study within a highly polymorphic rodent species, *Gerbillus nigeriae*. Chromosome Res 18:473–86.

Gazave E, Catalan J, Da Graça Ramalhinho M, et al (2003) The non-random occurrence of Robertsonian fusion in the house mouse. Genet Res 81:33–42.

Geisler R, Rauch GJ, Baier H, et al (1999) A radiation hybrid map of the zebrafish genome. Nat Genet 23:86–9.

Gemayel R, Vinces MD, Legendre M, Verstrepen KJ (2010) Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. Annu Rev Genet 44:455-77.

Genome 10K Community of Scientists Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species (200) J Hered 100:659–74.

Gerton JL, DeRisi J, Shroff R, et al (2000) Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci 97:11383–90.

Giménez MD, White T a, Hauffe HC, et al (2013) Understanding the basis of diminished gene flow between hybridizing chromosome races of the house mouse. Evolution 67:1446–62.

Giraut L, Falque M, Drouaud J, et al (2011) Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. PLoS Genet 7:e1002354.

Girirajan S, Chen L, Graves T, et al (2009) Sequencing human-gibbon breakpoints of synteny reveals mosaic new insertions at rearrangement sites. Genome Res 19:178–90.

Good JM, Dean MD, Nachman MW (2008) A complex genetic basis to X-linked hybrid male sterility between two species of house mice. Genetics 179:2213–28.

Gordon JL, Byrne KP, Wolfe KH (2009) Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. PLoS Genet 5:e1000485.

Grabherr MG, Russell P, Meyer M, et al (2010) Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics 26:1145–51.

Graphodatsky AS, Yang F, Dobigny G, et al (2008) Tracking genome organization in rodents by Zoo-FISH. Chromosome Res 16:261–74.

Greenbaum IF, Reed MJ (1984) Evidence for heterosynaptic pairing of the inverted segment in pericentric inversion heterozygotes of the deer mouse (*Peromyscus maniculatus*). Cytogenet Cell Genet 38:106–11.

Gregorová S, Forejt J (2000) PWD/Ph and PWK/Ph inbred mouse strains of *Mus m. musculus* subspecies--a valuable resource of phenotypic variations and genomic polymorphisms. Folia Biol 46:31–41.

Grey C, Barthès P, Chauveau-Le Friec G, et al (2011) Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. PLoS Biol 9:e1001176.

Griffin DK, Robertson LB, Tempest HG, et al (2008) Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. BMC Genomics 9:168.

Griffiths AJ, Gelbart WM, Miller JH, Lewontin RC (1999) Modern Genetic Analysis. WH Freeman.

Groenen MAM, Archibald AL, Uenishi H, et al (2012) Analyses of pig genomes provide insight into porcine

demography and evolution. Nature 491:393–8.

Groenen MAM, Wahlberg P, Foglio M, et al (2009) A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Res 19:510–9.

Gropp a., Winking H, Zech L, Müller H (1972) Robertsonian chromosomal variation and identification of metacentric chromosomes in feral mice. Chromosoma 39:265–288.

Gropp A, Winking H (1981) Robertsonian translocations: cytology, meiosis, segregation patterns and biological consequences of heterozygosity. In: Symposium of the Zoological Society of London. pp 47, 141–181

Gropp A, Winking H, Redi C (1982) Consequences of Robertsonian heterozygosity: segregational impairment of fertility versus male limited sterility. In: Genetic control of gamete production and function. Academic Press/Grune and Stratton.

Grzyb J, Latowski D, Strzałka K (2006) Lipocalins - a family portrait. J Plant Physiol 163:895–915.

Guelen L, Pagie L, Brasset E, et al (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453:948–51.

Guénet J-L, Bonhomme F (2003) Wild mice: an ever-increasing contribution to a popular mammalian model. Trends Genet 19:24–31.

Guerrero RF, Kirkpatrick M (2014) Local adaptation and the evolution of chromosome fusions. Evolution 68:2747–56.

Gündüz I, López-Fuster MJ, Ventura J, Searle JB (2001) Clinal analysis of a chromosomal hybrid zone in the house mouse. Genet Res 77:41–51.

Gündüz İ, Pollock CL, Giménez MD, et al (2010) Staggered Chromosomal Hybrid Zones in the House Mouse: Relevance to Reticulate Evolution and Speciation. Genes 1:193–209.

Guttenbach M, Nanda I, Feichtinger W, et al (2003) Comparative chromosome painting of chicken autosomal paints 1-9 in nine different bird species. Cytogenet Genome Res 103:173–84.

Gyapay G, Schmitt K, Fizames C, et al (1996) A radiation hybrid map of the human genome. Hum Mol Genet 5:339–46.

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. J Genet 12:101–109.

Hale DW (1986) Heterosynapsis and suppression of chiasmata within heterozygous pericentric inversions of the Sitka deer mouse. Chromosoma 94:425–32.

Hamilton AT, Huntley S, Kim J, et al (2003) Lineage-specific expansion of KRAB zinc-finger transcription factor genes: implications for the evolution of vertebrate regulatory networks. Cold Spring Harb Symp Quant Biol 68:131–40.

Handel MA, Schimenti JC (2010) Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. Nat Rev Genet 11:124–36.

Harley CB, Futcher AB, Greider CW (1990) Telomeres shorten during ageing of human fibroblasts. Nature 345:458–60.

Harr B (2006) Genomic islands of differentiation between house mouse subspecies. Genome Res 16:730–7.

Harushima Y, Yano M, Shomura A, et al (1998) A High-Density Rice Genetic Linkage Map with 2275 Markers Using a Single F2 Population. Genetics 148:479–494.

Hassold T, Hansen T, Hunt P, VandeVoort C (2009) Cytological studies of recombination in rhesus males. Cytogenet Genome Res 124:132–8.

Hassold T, Sherman S, Hunt P (2000) Counting cross-overs: characterizing meiotic recombination in mammals. Hum Mol Genet 9:2409–19.

Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. Nat Rev Genet 10:551–64.

Hauffe HC, Giménez MD, Searle JB (2012) Chromosomal hybrid zones in the house mouse. In: Evolution of the house mouse. Cambridge University Press, Cambridge, pp 407–430.

Hauffe HC, Searle JB (1998) Chromosomal heterozygosity and fertility in house mice (*Mus musculus domesticus*) from Northern Italy. Genetics 150:1143–1154.

Hausser J, Fedyk S, Fredga K, et al (1994) Definition and nomenclature of chromosome races of Sorex araneus. Folia Zool 43:1–9.

Hawken RJ, Murtaugh J, Flickinger GH, et al (1999) A first-generation porcine whole-genome radiation hybrid map. Mamm Genome 10:824–30.

Hayashi K, Yoshida K, Matsui Y (2005) A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature 438:374–8.

He Z, Henricksen LA, Wold MS, Ingles CJ (1995) RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. Nature 374:566–9.

Hefferin ML, Tomkinson AE (2005) Mechanism of DNA double-strand break repair by non-homologous end joining. DNA Repair 4:639–48.

Henderson KA, Keeney S (2005) Synaptonemal complex formation: where does it start? Bioessays 27:995–8.

Hewitt GM (1988) Hybrid zones-natural laboratories for evolutionary studies. Trends Ecol Evol 3:158–67.

Heyer W-D, Ehmsen KT, Liu J (2010) Regulation of homologous recombination in eukaryotes. Annu Rev Genet 44:113–39.

Heyting C (1996) Synaptonemal complexes: structure and function. Curr Opin Cell Biol 8:389–396.

Hoffmann FG, Opazo JC, Storz JF (2008) New genes originated via multiple recombinational pathways in the beta-globin gene family of rodents. Mol Biol Evol 25:2589–600.

Holmquist GP, Dancis B (1979) Telomere replication, kinetochore organizers, and satellite DNA evolution. Proc Natl Acad Sci 76:4566–70.

Hublitz P, Albert M, Peters AHFM (2009) Mechanisms of transcriptional repression by histone lysine methylation. Int J Dev Biol 53:335–54.

Hübner R, Koulischer L (1990) Cytogenetic studies on wild house mice from Belgium. Genetica 80:93–100.

Hunter N, Kleckner N (2001) The Single-End Invasion. Cell 106:59–70.

Jabbari K, Bernardi G (1998) CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families. Gene 224:123–128.

Janoušek V, Munclinger P, Wang L, et al (2015) Functional organization of the genome may shape the species boundary in the house mouse. Mol Biol Evol 32:1208–20.

Janoušek V, Wang L, Luzynski K, et al (2012) Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. Mol Ecol 21:3032–47.

Jansen van Vuuren B, Chown SL (2006) Genetic evidence confirms the origin of the house mouse on sub-Antarctic Marion Island. Polar Biol 30:327–332.

Jeffreys AJ, Cotton VE, Neumann R, Lam KG (2013) Recombination regulator PRDM9 influences the instability of its own coding sequence in humans.

John B, King M (1985) The inter-relationship between heterochromatin distribution and chiasma

distribution. Genetica 66:183–194.

Jurka J (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci 94:1872–7.

Kaback D, Guacci V, Barber D, Mahon J (1992) Chromosome size-dependent control of meiotic recombination. Science 256:228–232.

Kalitsis P, Griffiths B, Choo KHA (2006) Mouse telocentric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. Proc Natl Acad Sci 103:8786–91.

Kauppi L, Jeffreys AJ, Keeney S (2004) Where the crossovers are: recombination distributions in mammals. Nat Rev Genet 5:413–24.

Kawakami T, Smeds L, Backström N, et al (2014) A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol Ecol 23:4035–58.

Keeney S (2008) Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. Genome Dyn Stab 2:81–123.

Keeney S, Giroux CN, Kleckner N (1997) Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family. Cell 88:375–384.

Kehrer-Sawatzki H, Cooper DN (2008a) Comparative analysis of copy number variation in primate genomes. Cytogenet Genome Res 123:288–96.

Kehrer-Sawatzki H, Cooper DN (2008b) Molecular mechanisms of chromosomal rearrangement during primate evolution. Chromosome Res 16:41–56.

Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome Res 18:30–38.

Kemkemer C, Kohn M, Cooper DN, et al (2009) Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. BMC Evol Biology 9:84.

Kent WJ, Baertsch R, Hinrichs A, et al (2003) Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc Natl Acad Sci 100:11484–9.

King M (1993) Species evolution. Cambridge University Press, Cambridge.

Kipling D, Cooke HJ (1990) Hypervariable ultra-long telomeres in mice. Nature 347:400–2.

Kloosterman WP, Tavakoli-Yaraki M, van Roosmalen MJ, et al (2012) Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. Cell Rep 1:648–55.

Kneitz B, Cohen PE, Avdievich E, et al (2000) MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. Genes & Dev 14:1085–1097.

Koehler KE, Cherry JP, Lynn A, et al (2002a) Genetic control of mammalian meiotic recombination. I. Variation in exchange frequencies among males from inbred mouse strains. Genetics 162:297–306.

Koehler KE, Millie EA, Cherry JP, et al (2002b) Sex-specific differences in meiotic chromosome segregation revealed by dicentric bridge resolution in mice. Genetics 162:1367–79.

Koepfli K-P, Paten B, O'Brien SJ (2015) The Genome 10K Project: A Way Forward. Annu Rev Anim Biosci 3:57–111.

Kolb J, Chuzhanova N a, Högel J, et al (2009) Cruciform-forming inverted repeats appear to have mediated many of the microinversions that distinguish the

human and chimpanzee genomes. Chromosome Res 17:469–83.

Kong A, Thorleifsson G, Gudbjartsson DF, et al (2010) Fine-scale recombination rate differences between sexes, populations and individuals. Nature 467:1099–103.

Kono H, Tamura M, Osada N, et al (2014) *Prdm9* polymorphism unveils mouse evolutionary tracks. DNA Res 21:315–26.

Korbel JO, Urban AE, Affourtit JP, et al (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318:420–6.

Koszul R, Fischer G (2009) A prominent role for segmental duplications in modeling eukaryotic genomes. C R Biol 332:254–66.

Kozubek S, Lukásová E, Jirsová P, et al (2002) 3D Structure of the human genome: order in randomness. Chromosoma 111:321–31.

Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor M a F (2008) Fine-scale mapping of recombination rate in *Drosophila* refines its correlation to diversity and divergence. Proc Natl Acad Sci 105:10051–6.

Kuo LJ, Yang L-X (2008) -gamma-H2AX - A Novel Biomarker for DNA Double-strand Breaks. In Vivo 22:305–309.

Kvikstad EM, Makova KD (2010) The (r)evolution of SINE versus LINE distributions in primate genomes: sex chromosomes are important. Genome Res 20:600–13

Lalley PA, Minna JD, Francke U (1978) Conservation of autosomal gene synteny groups in mouse and man. Nature 274:160–3.

Lande R (1985) The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. Heredity 54 (Pt 3):323–32.

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

Larkin DM (2010) Role of chromosomal rearrangements and conserved chromosome regions in amniote evolution. Mol Genet Microbiol Virol 25:1–7.

Larkin DM, Pape G, Donthu R, et al (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res 19:770–7.

Lawrie NM, Tease C, Hultén MA (1995) Chiasma frequency, distribution and interference maps of mouse autosomes. Chromosoma 104:308–14.

Lee J, Han K, Meyer TJ, et al (2008) Chromosomal Inversions between Human and Chimpanzee Lineages Caused by Retrotransposons. PLoS One 3:e4047.

Lee JA, Carvalho CMB, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131:1235–47.

Lee SE (2014) Single-Strand Annealing. Mol life Sci 10:1–77.

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

Lemaitre C, Zaghloul L, Sagot M-F, et al (2009) Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. BMC Genomics 10:335.

Lenormand T, Dutheil J (2005) Recombination difference between sexes: a role for haploid selection. PLoS Biol 3:e63.

Lentzios G, Stocker AJ, Martin J (1980) C-banding and chromosome evolution in some related species of Australian chironominae. Genetica 54:51–68.

Levy A, Schwartz S, Ast G (2010) Large-scale discovery of insertion hotspots and preferential integration sites of human transposed elements. Nucleic Acids Res 38:1515–30.

Lewin H a, Larkin DM, Pontius J, O'Brien SJ (2009) Every genome sequence needs a good map. Genome Res 19:1925–8.

Li L, Lejnine S, Makarov V, Langmore JP (1998) In vitro and in vivo reconstitution and stability of vertebrate chromosome ends. Nucleic Acids Res 26:2908–2908.

Li W, Freudenberg J (2009) Two-parameter characterization of chromosome-scale recombination rate. Genome Res 19:2300–7.

Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem 79:181–211.

Lieberman-Aiden E, van Berkum NL, Williams L, et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–93.

Lindgren G, Sandberg K, Persson H, et al (1998) A primary male autosomal linkage map of the horse genome. Genome Res 8:951–66.

Lipkin SM, Moens PB, Wang V, et al (2002) Meiotic arrest and aneuploidy in MLH3-deficient mice. Nat Genet 31:385–90.

Liu G, Liu J, Cui X, Cai L (2012) Sequence-dependent prediction of recombination hotspots in *Saccharomyces cerevisiae.* J Theor Biol 293:49–54.

Liu GE, Ventura M, Cellamare A, et al (2009) Analysis of recent segmental duplications in the bovine genome. BMC Genomics 10:571.

Longhese MP, Bonetti D, Guerini I, et al (2009) DNA double-strand breaks in meiosis: checking their formation, processing and repair. DNA Repair 8:1127–38.

Longo MS, Carone DM, Green ED, et al (2009) Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. BMC Genomics 10:334.

Looman C, Abrink M, Mark C, Hellman L (2002) KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. Mol Biol Evol 19:2118–30.

Lupski JR (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 14:417–22.

Lupski JR, Stankiewicz P (2005) Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet 1:e49.

Lynn A, Koehler KE, Judis L, et al (2002) Covariation of synaptonemal complex length and mammalian meiotic exchange rates. Science 296:2222–5.

Lynn A, Schrump S, Cherry J, et al (2005) Sex, not genotype, determines recombination levels in mice. Am J Hum Genet 77:670–5.

Ma J, Zhang L, Suh BB, et al (2006) Reconstructing contiguous regions of an ancestral genome. Genome Res 16:1557–65.

Manieu C, González M, López-Fenner J (2014) Aneuploidy in spermatids of Robertsonian (Rb) chromosome heterozygous mice. Chromosome Res 22:545-57.

Manoukis NC, Powell JR, Touré MB, et al (2008) A test of the chromosomal theory of ecotypic speciation in *Anopheles gambiae*. Proc Natl Acad Sci 105:2940–5.

Manterola M, Page J, Vasco C, et al (2009) A high incidence of meiotic silencing of unsynapsed chromatin is not associated with substantial pachytene loss in heterozygous male mice carrying multiple simple robertsonian translocations. PLoS Genet 5:e1000625.

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402.

Marques-Bonet T, Kidd JM, Ventura M, et al (2009) A burst of segmental duplications in the genome of the African great ape ancestor. Nature 457:877–81.

Martí DA, Bidau CJ (2004) Synapsis in Robertsonian Heterozygotes and Homozygotes of *Dichroplus Pratensis* (Melanoplinae, Acrididae) and Its Relationship with Chiasma Patterns. Hereditas 134:245–254.

Martínez-Vargas J, Muñoz-Muñoz F, Medarde N, et al (2014) Effect of chromosomal reorganizations on morphological covariation of the mouse mandible: insights from a Robertsonian system of *Mus musculus domesticus*. Front Zool 11:51.

Masly JP, Jones CD, Noor MAF, et al (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. Science 313:1448–50.

Mayr E (1942) Systematics and the Origin of Species, from the Viewpoint of a Zoologist. Harvard University Press.

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801.

McGraw LA, Davis JK, Young LJ, Thomas JW (2011) A genetic linkage map and comparative mapping of the prairie vole (*Microtus ochrogaster*) genome. BMC Genet 12:60.

McMahill MS, Sham CW, Bishop DK (2007) Synthesis-dependent strand annealing in meiosis. PLoS Biol 5:e299.

McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet 24:529–38.

Meaburn KJ, Misteli T, Soutoglou E (2007) Spatial genome organization in the formation of chromosomal translocations. Semin Cancer Biol 17:80–90.

Medarde N (2013) La zona de polimorfismo cromosómico "Barcelona" de Mus musculus domesticus Schwarz y Schwarz, 1943: dinámica espaciotemporal de su estructura y efecto de las fusiones robertsonianas sobre la espermatogénesis. Doctoral Thesis.

Medarde N, López-Fuster MJ, Muñoz-Muñoz F, Ventura J (2012) Spatio-temporal variation in the structure of a chromosomal polymorphism zone in the house mouse. Heredity 109:78–89.

Medarde N, Merico V, López-Fuster MJ, et al (2015) Impact of the number of Robertsonian chromosomes on germ cell death in wild male house mice. Chromosome Res 23:159–69.

Medarde N, Muñoz-Muñoz F, López-Fuster MJ, Ventura J (2013) Variational modularity at the cell level: insights from the sperm head of the house mouse. BMC Evol Biol 13:179.

Mellersh CS, Hitte C, Richman M, et al (2000) An integrated linkage-radiation hybrid map of the canine genome. Mamm Genome 11:120–30.

Meredith RW, Janečka JE, Gatesy J, et al (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science 334:521–4.

Merico V, Giménez MD, Vasco C, et al (2013) Chromosomal speciation in mice: a cytogenetic analysis of recombination. Chromosome Res 21:523–33.

Merico V, Pigozzi MI, Esposito A, et al (2003) Meiotic recombination and spermatogenic impairment in Mus musculus domesticus carrying multiple simple Robertsonian translocations. Cytogenet Genome Res 103:3–4.

Meuleman W, Peric-hupkes D, Kind J, et al (2013) Constitutive nuclear lamina – genome interactions are highly conserved and associated with A / T-rich sequence. 270–280.

Meyne J, Baker RJ, Hobart HH, et al (1990) Distribution

of non-telomeric sites of the (TTAGGG)n telomeric sequence in vertebrate chromosomes. Chromosoma 99:3–10.

Meznar ER, Gadau J, Koeniger N, Rueppell O (2010) Comparative linkage mapping suggests a high recombination rate in all honeybees. J Hered 101 Suppl :S118–26.

Mihola O, Trachtulec Z, Vlcek C, et al (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science 323:373–5.

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

Moens PB, Marcon E, Shore JS, et al (2007) Initiation and resolution of interhomolog connections: crossover and non-crossover sites along mouse synaptonemal complexes. J Cell Sci 120:1017–27.

Mongin E, Dewar K, Blanchette M (2009) Long-range regulation is a major driving force in maintaining genome integrity. BMC Evol Biol 9:203.

Mora L, Sánchez I, Garcia M, Ponsà M (2006) Chromosome territory positioning of conserved homologous chromosomes in different primate species. Chromosoma 115:367–75.

Morán T, Fontdevila A (2014) Genome-wide dissection of hybrid sterility in *Drosophila* confirms a polygenic threshold architecture. J Hered 105:381–96.

Mosna E, Rivosechi L, Ascher KR (1958) Studies on insecticide-resistant anophelines. I. Chromosome arrangements in a dieldrin-selected strain of *Anopheles atroparvus*. Bull World Health Organ 19:297–301.

Moyzis RK, Buckingham JM, Cram LS, et al (1988) A highly conserved repetitive DNA sequence, (TTAGGG) n, present at the telomeres of human chromosomes. Proc Natl Acad Sci 85:6622–6.

Muller HJ (1942) Isolating mechanisms, evolution and temperature. Biol Symp 6:71–125.

Munch K, Mailund T, Dutheil JY, Schierup MH (2014) A fine-scale recombination map of the human-chimpanzee ancestor reveals faster change in humans than in chimpanzees and a strong impact of GC-biased gene conversion. Genome Res 24:467–74.

Muñoz-Muñoz F, Sans-Fuentes M a., Lopez-Fuster MJ, Ventura J (2006) Variation in fluctuating asymmetry levels across a Robertsonian polymorphic zone of the house mouse. J Zool Syst Evol Res 44:236–250.

Muñoz-Muñoz F, Sans-Fuentes MA, López-Fuster MJ, Ventura J (2011) Evolutionary modularity of the mouse mandible: dissecting the effect of chromosomal reorganizations and isolation by distance in a Robertsonian system of *Mus musculus domesticus*. J Evol Biol 24:1763–76.

Murakami H, Keeney S (2008) Regulating the formation of DNA double-strand breaks in meiosis. Genes Dev 22:286–92.

Murphy WJ, Agarwala R, Schäffer AA, et al (2005a) A rhesus macaque radiation hybrid map and comparative analysis with the human genome. Genomics 86:383–95.

Murphy WJ, Larkin DM, Everts-van der Wind A, et al (2005b) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309:613–7.

Murphy WJ, Stanyon R, O'Brien SJ (2001) Evolution of mammalian genome organization inferred from comparative gene mapping. Genome Biol 2:REVIEWS0005.

Murphy WJ, Sun S, Chen Z, et al (2000) A radiation hybrid map of the cat genome: implications for comparative mapping. Genome Res 10:691–702.

Musilova P, Kubickova S, Vahala J, Rubes J (2013) Subchromosomal karyotype evolution in Equidae. Chromosome Res 21:175–87.

Myers S, Bowden R, Tumian A, et al (2010) Drive against

hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327:876–9.

Myers S, Freeman C, Auton A, et al (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. Nat Genet 40:1124–9.

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

Nadeau JH, Taylor B a (1984) Lengths of chromosomal segments conserved since divergence of man and mouse. Proc Natl Acad Sci 81:814–8.

Nanda I, Schneider-Rasp S, Winking H, Schmid M (1995) Loss of telomeric sites in the chromosomes of *Mus musculus domesticus* (Rodentia: Muridae) during Robertsonian rearrangements. Chromosome Res 3:399–409.

Näslund K, Saetre P, von Salomé J, et al (2005) Genome-wide prediction of human VNTRs. Genomics 85:24–35.

Navarro A, Barton NH (2003a) Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes. Science 300:321–4.

Navarro A, Barton NH (2003b) Accumulating Postzygotic Isolation Genes in Parapatry: a New Twist on Chromosomal Speciation. Evolution 57:447.

Navarro A, Betrán E, Barbadilla A, Ruiz A (1997) Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics 146:695–709.

Nergadze SG, Rocchi M, Azzalin CM, et al (2004) Insertion of telomeric repeats at intrachromosomal break sites during primate evolution. Genome Res 14:1704–10.

Newman TL, Tuzun E, Morrison VA, et al (2005) A genome-wide survey of structural variation between human and chimpanzee. Genome Res 15:1344–56.

Nishant KT, Rao MRS (2006) Molecular features of meiotic recombination hot spots. Bioessays 28:45–56.

Noor M a, Grams KL, Bertucci L a, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. Proc Natl Acad Sci 98:12084–8.

Noor MAF, Garfield DA, Schaeffer SW, Machado CA (2007) Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. Genetics 177:1417–28.

Nora EP, Dekker J, Heard E (2013) Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? BioEssays 35:818–828.

Nowick K, Carneiro M, Faria R (2013) A prominent role of KRAB-ZNF transcription factors in mammalian speciation? Trends Genet 29:130–9.

Nunes a C, Catalan J, Lopez J, et al (2011) Fertility assessment in hybrids between monobrachially homologous Rb races of the house mouse from the island of Madeira: implications for modes of chromosomal evolution. Heredity 106:348–56.

O'Brien SJ, Cevario SJ, Martenson JS, et al (1995) Comparative gene mapping in the domestic cat (Felis catus). J Hered 88:408–14.

O'Brien SJ, Wienberg J, Lyons LA (1997) Comparative genomics: lessons from cats. Trends Genet 13:393–9.

O'Sullivan RJ, Karlseder J (2010) Telomeres: protecting chromosomes against genome instability. Nat Rev Mol Cell Biol 11:171–81.

Oliver PL, Goodstadt L, Bayes JJ, et al (2009) Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. PLoS Genet 5:e1000753.

Onishi-Seebacher M, Korbel JO (2011) Challenges in studying genomic structural variant formation mechanisms: the short-read dilemma and beyond.

Bioessays 33:840–50.

Ohno S (1973) Ancient linkage groups and frozen accidents. Nature 244:259–62.

Oud JL, de Jong JH, de Rooij DG (1979) A sequential analysis of meiosis in the male mouse using a restricted spermatocyte population obtained by a hydroxyurea/triaziquone treatment. Chromosoma 71:237–48.

Pagacova E, Cernohorska H, Kubickova S, et al (2009) Centric fusion polymorphism in captive animals of family Bovidae. Conserv Genet 12:71–77.

Page SL, Hawley RS (2004) The genetics and molecular biology of the synaptonemal complex. Annu Rev Cell Dev Biol 20:525–58.

Page SL, Shin JC, Han JY, et al (1996) Breakpoint diversity illustrates distinct mechanisms for Robertsonian translocation formation. Hum Mol Genet 5:1279–88.

Paigen K, Szatkiewicz JP, Sawyer K, et al (2008) The recombinational anatomy of a mouse chromosome. PLoS Genet 4:e1000119.

Parada L, Misteli T (2002) Chromosome positioning in the interphase nucleus. Trends Cell Biol 12:425–32.

Pardini AT, O'Brien PCM, Fu B, et al (2007) Chromosome painting among Proboscidea, Hyracoidea and Sirenia: support for Paenungulata (Afrotheria, Mammalia) but not Tethytheria. Proc Biol Sci 274:1333–40.

Pardo-Manuel de Villena F, Sapienza C (2001) Recombination is proportional to the number of chromosome arms in mammals. Mamm Genome 12:318–22.

Parvanov ED, Petkov PM, Paigen K (2010) *Prdm9* controls activation of mammalian recombination hotspots. Science 327:835.

Pawlina K, Bugno-Poniewierska M (2012) The application of zoo-fish technique for analysis of chromosomal rearrangements in the equidae family. Ann Anim Sci 12:5–13.

Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution 58:2064–2078.

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

Perelman PL, Beklemisheva VR, Yudkin D V, et al (2012) Comparative chromosome painting in Carnivora and Pholidota. Cytogenet Genome Res 137:174–93.

Perez DE, Wu CI, Johnson NA, Wu ML (1993) Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, Odysseus (Ods). Genetics 134:261–75.

Peric-Hupkes D, Meuleman W, Pagie L, et al (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell 38:603–13.

Petes TD (2001) Meiotic recombination hot spots and cold spots. Nat Rev Genet 2:360–9.

Petkov PM, Broman KW, Szatkiewicz JP, Paigen K (2007) Crossover interference underlies sex differences in recombination rates. Trends Genet 23:539–42.

Pevzner P, Tesler G (2003a) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc Natl Acad Sci 100:7672–7.

Pevzner P, Tesler G (2003b) Genome Rearrangements in Mammalian Evolution: Lessons From Human and Mouse Genomes. Genome Research 13:37-45.

Phadnis N, Orr HA (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. Science 323:376–9.

Phillips-Cremins JE (2014) Unraveling architecture of the pluripotent genome. Curr Opin Cell Biol 28:96–104.

Piálek J, Hauffe HC, Rodríguez-Clark KM, Searle JB (2001) Raciation and speciation in house mice from the Alps: the role of chromosomes. Mol Ecol 10:613–25.

Piálek J, Hauffe HC, Searle JB (2005) The genus Mus as a model for evolutionary studies Chromosomal variation in the house mouse. 535–563.

Pittman DL, Cobb J, Schimenti KJ, et al (1998) Meiotic Prophase Arrest with Failure of Chromosome Synapsis in Mice Deficient for Dmc1, a Germline-Specific RecA Homolog. Mol Cell 1:697–705.

Plug AW, Peters AH, Xu Y, et al (1997) ATM and RPA in meiotic chromosome synapsis and recombination. Nat Genet 17:457–61.

Pollack JR, Sørlie T, Perou CM, et al (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci 99:12963–8.

Poon SSS, Martens UM, Ward RK, Lansdorp PM (1999) Telomere length measurements using digital fluorescence microscopy. Cytometry 36:267–278.

Pratto F, Brick K, Khil P, et al (2014a) Recombination initiation maps of individual human genomes. Science 346:1256442–1256442.

Presgraves DC (2010) Speciation genetics: search for the missing snowball. Curr Biol 20:R1073–4.

Proost S, Fostier J, De Witte D, et al (2012) i-ADHoRe 3.0-fast and sensitive detection of genomic homology in extremely large data sets. Nucleic Acids Res 40:e11.

Prowse KR, Greider CW (1995) Developmental and tissue-specific regulation of mouse telomerase and telomere length. Proc Natl Acad Sci 92:4818–22.

Quinlan AR, Clark RA, Sokolova S, et al (2010b) Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Res 20:623–35.

Qumsiyeh MB (1994) Evolution of number and morphology of mammalian chromosomes. J Hered 85:455–65.

Rane R V, Rako L, Kapun M, et al (2015) Genomic evidence for role of inversion 3RP of *Drosophila melanogaster* in facilitating climate change adaptation. Mol Ecol 24:2423–32.

Rao SSP, Huntley MH, Durand NC, et al (2014) A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159:1665–80.

Redi CA, Capanna E (1988) Robertsonian heterozygotes in the house mouse and the fate of their germ cells. In: The Cytogenetics of Mammalian Rearrangements. pp 315–359

Redi CA, Garagna S, Capanna E (1990) Nature's experiment with *in situ* hybridization? A hypothesis for the mechanism of Rb fusion. J Evol Biol 3:133–137.

Redi CA, Garagna S, Hilscher B, Winking H (1985) The effects of some Robertsonian chromosome combinations on the seminiferous epithelium of the mouse. J Embryol Exp Morphol 85:1–19.

Reeves A (2001) MicroMeasure: A new computer program for the collection and analysis of cytogenetic data. Genome 44:439-43.

Reig-Viader R, Vila-Cejudo M, Vitelli V, et al (2014) Telomeric repeat-containing RNA (TERRA) and telomerase are components of telomeres during mammalian gametogenesis. Biol Reprod 90:103.

Richard G-F, Kerrest A, Dujon B (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev 72:686–727.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. Trends Ecol Evol 16:351–358.

Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. Heredity 83 :363–72.

Rieseberg LH, Linder CR, Seiler GJ (1995) Chromosomal and genic barriers to introgression in Helianthus. Genetics 141:1163–71.

Robertson WRB (1916) Chromosome studies. I. Taxonomic relationships shown in the chromosomes of tettigidae and acrididae: V-shaped chromosomes and their significance in acrididae, locustidae, and gryllidae: Chromosomes and variation. J Morphol 27:179–331.

Robinson TJ, Ropiquet A (2011) Examination of hemiplasy, homoplasy and phylogenetic discordance in chromosomal evolution of the Bovidae. Syst Biol 60:439–50.

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

Rödelsperger C, Dieterich C (2010) CYNTENATOR: progressive gene order alignment of 17 vertebrate genomes. PLoS One 5:e8861.

Rogakou EP, Pilch DR, Orr a. H, et al (1998) DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. J Biol Chem 273:5858–5868.

Rogers J, Witte SM, Kammerer CM, et al (1995) Linkage mapping in Papio baboons: conservation of a syntenic group of six markers on human chromosome 1. Genomics 28:251–4.

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

Romanenko SA, Volobouev V (2012) Non-Sciuromorph rodent karyotypes in evolution. Cytogenet Genome Res 137:233–45.

Romanienko PJ, Camerini-Otero RD (2000) The Mouse Spo11 Gene Is Required for Meiotic Chromosome Synapsis. Mol Cell 6:975–987.

Ruiz-Herrera a, Nergadze SG, Santagostino M, Giulotto E (2008) Telomeric repeats far from the ends: mechanisms of origin and role in evolution. Cytogenet Genome Res 122:219–28.

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

Ruiz-Herrera A, Farré M, Ponsà M, Robinson TJ (2010) Selection against Robertsonian fusions involving housekeeping genes in the house mouse: integrating data from gene expression arrays and chromosome evolution. Chromosome Res 18:801–8.

Ruiz-Herrera A, García F, Mora L, et al (2005) Evolutionary conserved chromosomal segments in the human karyotype are bounded by unstable chromosome bands. Cytogenet Genome Res 108:161–74.

Ruiz-Herrera A, Ponsà M, García F, et al (2002) Fragile sites in human and *Macaca fascicularis* chromosomes are breakpoints in chromosome evolution. Chromosome Res 10:33–44.

Saltman D, Morgan R, Cleary ML, de Lange T (1993) Telomeric structure in cells with chromosome end associations. Chromosoma 102:121–128.

Sankoff D (2006) The signal in the genomes. PLoS Comput Biol 2:e35.

Sankoff D, Ferretti V, Nadeau JH (1997) Conserved segment identification. J Comput Biol 4:559–65.

Sankoff D, Trinh P (2005) Chromosomal breakpoint reuse in genome sequence rearrangement. J Comput Biol 12:812–21.

Sans-Fuentes MA (2004) Estudio Biologico de Mus domesticus Rutty, 1772 en una zona de polimorfismo Robertsoniano. Doctoral Thesis.

Sans-Fuentes MA, García-Valero J, Ventura J, López-Fuster MJ (2010) Spermatogenesis in house mouse in a Robertsonian polymorphism zone. Reproduction 140:569–81.

Sans-Fuentes MA, López-Fuster MJ, Ventura J, et al (2005) Effect of Robertsonian translocations on the motor activity rhythm in the house mouse. Behav Genet 35:603–13.

Sans-Fuentes MA, Muñoz-Muñoz F, Ventura J, López-Fuster MJ (2007) Rb(7.17), a rare Robertsonian fusion in wild populations of the house mouse. Genet Res 89:207–13.

Sans-Fuentes MA, Ventura J, López-Fuster MJ, Corti M (2009) Morphological variation in house mice from the Robertsonian polymorphism area of Barcelona. Biol J Linn Soc 97:555–570.

Santos-Colares MC dos, Degrandi TH, Valente VLS (2004) Cytological Detection of Male Recombination in *Drosophila willistoni*. Cytologia69:359–365.

Santucci-Darmanin S (2000) MSH4 acts in conjunction with MLH1 during mammalian meiosis. FASEB J 14:1539–1547.

Sasaki M, Lange J, Keeney S (2010) Genome destabilization by homologous recombination in the germ line. Nat Rev Mol Cell Biol 11:182–95.

Scherthan H, Cremer T, Arnason U, et al (1994) Comparative chromosome painting discloses homologous segments in distantly related mammals. Nat Genet 6:342–7.

Scherthan H, Weich S, Schwegler H, et al (1996) Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing. J Cell Biol 134:1109–25.

Schibler L, Roig A, Mahe M-F, et al (2006) High-resolution comparative mapping among man, cattle and mouse suggests a role for repeat sequences in mammalian genome evolution. BMC Genomics 7:194.

Schmidt D, Schwalie PC, Wilson MD, et al (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 148:335–48.

Schoeftner S, Blasco MA (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat Cell Biol 10:228–36.

Schwacha A, Kleckner N (1995) Identification of double holliday junctions as intermediates in meiotic recombination. Cell 83:783–791.

Schwartz JJ, Roach DJ, Thomas JH, Shendure J (2014) Primate evolution of the recombination regulator PRDM9. Nat Commun 5:4370.

Schwarz E, Schwarz HK (1943) The Wild and Commensal Stocks of the House Mouse, *Mus musculus* Linnaeus. J Mammal 24:59–72.

Searle JB (1991) A hybrid zone comprising staggered chromosomal clines in the house mouse (*Mus musculus domesticus*). Proc Biol Sci 246:47–52.

Searle JB (1993) Chromosomal hybrid zones in eutherian mammals. In: Hybrid zones and the evolutionary process. Oxford University Press, New York, pp 309–353

Segura J, Ferretti L, Ramos-onsins S, et al (2013) Evolution of recombination in eutherian mammals : insights into mechanisms that affect recombination rates and crossover interference. Proc R Soc B 280:20131945.

Sen SK, Han K, Wang J, et al (2006) Human genomic deletions mediated by recombination between Alu elements. Am J Hum Genet 79:41–53.

Sexton T, Kurukuti S, Mitchell JA, et al (2012) Sensitive

detection of chromatin coassociations using enhanced chromosome conformation capture on chip. Nat Protoc 7:1335–50.

Shannon M, Hamilton AT, Gordon L, et al (2003) Differential expansion of zinc-finger transcription factor loci in homologous human and mouse gene clusters. Genome Res 13:1097–110.

Sharp AJ, Hansen S, Selzer RR, et al (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet 38:1038–42.

Shaw CJ, Lupski JR (2004) Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet 13 Spec No:R57–64.

She X, Cheng Z, Zöllner S, et al (2008) Mouse segmental duplication and copy number variation. Nat Genet 40:909–14.

She X, Jiang Z, Clark RA, et al (2004) Shotgun sequence assembly and recent segmental duplications within the human genome. Nature 431:927–30.

Siva N (2008) 1000 Genomes project. Nat Biotechnol 26:256.

Skinner BM, Griffin DK (2012) Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity 108:37–41.

Slijepcevic P (1998) Chromosoma Focus Telomeres and mechanisms of Robertsonian fusion. Chromosoma 136–141.

Slijepcevic P, Hande MP, Bouffler SD, et al (1997) Telomere length, chromatin structure and chromosome fusigenic potential. Chromosoma 106:413–421.

Smagulova F, Gregoretti I V, Brick K, et al (2011) Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. Nature 472:375–8.

Smukowski CS, Noor M a F (2011) Recombination rate variation in closely related species. Heredity 107:496–508.

Snowden T, Acharya S, Butz C, et al (2004) hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. Mol Cell 15:437–51. 0

Solano E, Castiglia R, Capanna E (2008) Chromosomal evolution of the house mouse, *Mus musculus domesticus,* in the Aeolian Archipelago (Sicily, Italy). Biol J Linn Soc 96:194–202.

Speed RM (1982) Meiosis in the foetal mouse ovary. I. An analysis at the light microscope level using surface-spreading. Chromosoma 85:427–37.

Stanyon R, Bigoni F, Slaby T, et al (2004) Multi-directional chromosome painting maps homologies between species belonging to three genera of New World monkeys and humans. Chromosoma 113:305–15.

Stanyon R, Consigliere S, Müller S, et al (2000) Fluorescence *in situ* hybridization (FISH) maps chromosomal homologies between the dusky titi and squirrel monkey. Am J Primatol 50:95–107.

Stanyon R, Yang F, Cavagna P, et al (1999) Reciprocal chromosome painting shows that genomic rearrangement between rat and mouse proceeds ten times faster than between humans and cats. Cytogenet Cell Genet 84:150–5.

Steiner CC, Ryder OA (2013) Characterization of *Prdm9* in equids and sterility in mules. PLoS One 8:e61746.

Steiner WW, Smith GR (2005) Optimizing the nucleotide sequence of a meiotic recombination hotspot in Schizosaccharomyces pombe. Genetics 169:1973–83.

Stephens PJ, Greenman CD, Fu B, et al (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell

144:27–40.

Stevison LS, Hoehn KB, Noor M a F (2011) Effects of inversions on within- and between-species recombination and divergence. Genome Biol Evol 3:830–41.

Stopková R, Dudková B, Hájková P, Stopka P (2014) Complementary roles of mouse lipocalins in chemical communication and immunity. Biochem Soc Trans 42:893–8.

Stopková R, Hladovcová D, Kokavec J, Vyoral D (2009) Multiple roles of secretory lipocalins (Mup, Obp) in mice. 58:29–40.

Storchová R, Gregorová S, Buckiová D, et al (2004) Genetic analysis of X-linked hybrid sterility in the house mouse. Mamm Genome.

Storz JF, Baze M, Waite JL, et al (2007) Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. Genetics 177:481–500.

Storz JF, Opazo JC, Hoffmann FG (2013) Gene duplication, genome duplication, and the functional diversification of vertebrate globins. Mol Phylogenet Evol 66:469–78.

Storz JF, Runck AM, Sabatino SJ, et al (2009) Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. Proc Natl Acad Sci 106:14450–5.

Sun F, Trpkov K, Rademaker A, et al (2005) Variation in meiotic recombination frequencies among human males. Hum Genet 116:172–8.

Suzuki K, Lareyre J-J, Sánchez D, et al (2004) Molecular evolution of epididymal lipocalin genes localized on mouse chromosome 2. Gene 339:49–59.

Tadepally HD, Burger G, Aubry M (2008) Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. BMC Evol Biol 8:176.

Taffarel AL, Laudio CJ, Ardo DA (2015) Chromosome fusion polymorphisms in the grasshopper , D*ichroplus fuscus* (Orthoptera : Acrididae : Melanoplinae): Insights on meiotic effects. 112:11–19.

Takata M, Sasaki MS, Sonoda E, et al (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. EMBO J 17:5497–508.

Tanabe H, Müller S, Neusser M, et al (2002) Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proc Natl Acad Sci 99:4424–9.

Tease C, Hultén MA (2004) Inter-sex variation in synaptonemal complex lengths largely determine the different recombination rates in male and female germ cells. Cytogenet Genome Res 107:208–15.

Teeter KC, Payseur BA, Harris LW, et al (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. 67–76.

Terrenoire E, McRonald F, Halsall JA, et al (2010) Immunostaining of modified histones defines high-level features of the human metaphase epigenome. Genome Biol 11:R110.

Tesler G (2002) GRIMM: genome rearrangements web server. Bioinformatics 18:492–3.

Thomas JH, Emerson RO, Shendure J (2009) Extraordinary molecular evolution in the PRDM9 fertility gene. PLoS One 4:e8505.

Tichy H, Vucak I (1987) Chromosomal polymorphism in the house mouse (*Mus domesticus*) of Greece and Yugoslavia. Chromosoma 95:31–6.

Ting C (1998) A Rapidly Evolving Homeobox at the Site of a Hybrid Sterility Gene. Science (80-) 282:1501–

1504.

Tjio JH, Levan A (2010) The Chromosome Number of Man. Hereditas 42:1–6.

Tortereau F, Servin B, Frantz L, et al (2012) A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. BMC Genomics 13:586.

Trifonov VA, Kosyakova N, Romanenko SA, et al (2010) New insights into the karyotypic evolution in muroid rodents revealed by multicolor banding applying murine probes. Chromosome Res 18:265–75.

Tsai IJ, Burt A, Koufopanou V (2010) Conservation of recombination hotspots in yeast. Proc Natl Acad Sci 107:7847–52.

Turner JM a, Mahadevaiah SK, Fernandez-Capetillo O, et al (2005a) Silencing of unsynapsed meiotic chromosomes in the mouse. Nat Genet 37:41–7.

Turner LM, Harr B (2014) Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. Elife 3:e02504.

Turner LM, White MA, Tautz D, Payseur BA (2014) Genomic networks of hybrid sterility. PLoS Genet 10:e1004162.

Turner TL, Hahn MW, Nuzhdin S V (2005b) Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol 3:e285.

Uddin M, Sturge M, Peddle L, et al (2011) Genome-wide signatures of "rearrangement hotspots" within segmental duplications in humans. PLoS One 6:e28853.

Ullastres A, Farré M, Capilla L, Ruiz-Herrera A (2014) Unraveling the effect of genomic structural changes in the rhesus macaque - implications for the adaptive role of inversions. BMC Genomics 15:530.

Urrutia R (2003) KRAB-containing zinc-finger repressor proteins. Genome Biol 4:231.

Usdin K, Grabczyk E (2000) DNA repeat expansions and human disease. Cell Mol Life Sci 57:914–931.

Van Os H, Andrzejewski S, Bakker E, et al (2006) Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics 173:1075–87.

Vasco C, Manterola M, Page J, et al (2012) The frequency of heterologous synapsis increases with aging in Robertsonian heterozygous male mice. Chromosome Res 20:269–78.

Ventura M, Antonacci F, Cardone MF, et al (2007) Evolutionary formation of new centromeres in macaque. Science 316:243–6.

Véron AS, Lemaitre C, Gautier C, et al (2011) Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny. BMC Genomics 12:303.

Verver DE, van Pelt AMM, Repping S, Hamer G (2013) Role for rodent Smc6 in pericentromeric heterochromatin domains during spermatogonial differentiation and meiosis. Cell Death Dis 4:e749.

Veyrunes F, Dobigny G, Yang F, et al (2006) Phylogenomics of the genus Mus (Rodentia; Muridae): extensive genome repatterning is not restricted to the house mouse. Proc Biol Sci 273:2925–34.

Walker EL, Robbins TP, Bureau TE, et al (1995) Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. EMBO J 14:2350–63.

Wallace BMN, Searle JB, Everett C a. (2002) The effect of multiple simple Robertsonian heterozygosity on chromosome pairing and fertility of wild-stock house mice (*Mus musculus domesticus*). Cytogenet Genome Res 96:276–286.

Wallace BMN, Searle JB, Everett CA (1992) Male meiosis and gametogenesis in wild house mice (*Mus musculus domesticus*) from a chromosomal hybrid zone; a comparison between "simple" Robertsonian heterozygotes and homozygotes. Cytogenet Genome Res 61:211–220.

Watanabe TK, Bihoreau MT, McCarthy LC, et al (1999) A radiation hybrid map of the rat genome containing 5,255 markers. Nat Genet 22:27–36.

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

White M, Blackith R, Blackith R, Cheney J (1967) Cytogenetics of the viatica group morabine grasshoppers. I. The coastal species. Aust J Zool 15:263.

White MA, Steffy B, Wiltshire T, Payseur BA (2011) Genetic dissection of a key reproductive barrier between nascent species of house mice. Genetics 189:289–304.

White MJ. (1973) Animal Cytology and Evolution. Cambridge University Press, London.

White MJD (1969) Chromosomal Rearrangements and Speciation in Animals. Annu Rev Genet 3:75–98.

White MJD (1978) Modes of speciation. W.H.Freeman & Co.

Wicker T, Sabot F, Hua-Van A, et al (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–82

Wilfert L, Gadau J, Schmid-Hempel P (2007) Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity 98:189–97.

Wilson GM, Flibotte S, Missirlis PI, et al (2006) Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. Genome Res 16:173–81.

Wong AK, Ruhe AL, Dumont BL, et al (2010) A comprehensive linkage map of the dog genome. Genetics 184:595–605.

Woodcock CL (2006) Chromatin architecture. Curr Opin Struct Biol 16:213–20.

Wright S (1941) On the probability of fixation of reciprocal translocations. Am Nat 75:531–522.

Wu CI, Li WH (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. Proc Natl Acad Sci 82:1741–1745.

Wu ZK, Getun I V., Bois PRJ (2010) Anatomy of mouse recombination hot spots. Nucleic Acids Res 38:2346–2354.

Wurster DH, Benirschke K (1970) Indian Momtjac, Muntiacus muntiak: A Deer with a Low Diploid Chromosome Number. Science 168:1364–1366.

Yamamoto M, Miklos GL (1978) Genetic studies on heterochromatin in *Drosophila melanogaster* and their implications for the functions of satellite DNA. Chromosoma 66:71–98.

Yang F, Alkalaeva EZ, Perelman PL, et al (2003) Reciprocal chromosome painting among human, aardvark, and elephant (superorder Afrotheria) reveals the likely eutherian ancestral karyotype. Proc Natl Acad Sci 100:1062–6.

Yang H, Wang JR, Didion JP, et al (2011) Subspecific origin and haplotype diversity in the laboratory mouse. Nat Genet 43:648–55.

Yannic G, Basset P, Hausser J (2009) Chromosomal rearrangements and gene flow over time in an inter-specific hybrid zone of the S*orex araneus* group. Heredity 102:616–25.

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. Genetics 175:1883–93.

Yoshida K, Kondoh G, Matsuda Y, et al (1998) The Mouse RecA-like Gene Dmc1 Is Required for Homologous Chromosome Synapsis during Meiosis. Mol Cell 1:707–718.

Yu AM, McVey M (2010) Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res 38:5706–17.

Yunis JJ, Prakash O (1982) The origin of man: a chromosomal pictorial legacy. Science 215:1525–30.

Zakian VA (1997) Life and cancer without telomerase. Cell 91:1–3.

Zhang F, Khajavi M, Connolly AM, et al (2009) The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nat Genet 41:849–53.

Zhang G, Li C, Li Q, et al (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311–1320.

Zhang Y, McCord RP, Ho Y-J, et al (2012) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell 148:908–21.

Zhao H, Bourque G (2009) Recovering genome rearrangements in the mammalian phylogeny. Genome Res 19:934–42.

Zhao S, Shetty J, Hou L, et al (2004) Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. Genome Res 14:1851–60.

Zhou Y, Rui L (2013) Lipocalin 13 regulation of glucose and lipid metabolism in obesity. Vitam Horm 91:369–83.

Zijlmans JM, Martens UM, Poon SS, et al (1997) Telomeres in the mouse have large inter-chromosomal variations in the number of T2AG3 repeats. Proc Natl Acad Sci 94:7423–8.

Zima J, Gaichenko V a, Macholán M, et al (1990) Are Robertsonian variations a frequent phenomenon in mouse populations in Eurasia? Biol J Linn Soc 41:229–233.