Aus dem Lehrstuhl für Molekularbiologie
im Biomedizinischen Centrum (BMC)
Institut der Ludwig-Maximilians-Universität München
Vorstand: Prof. Dr. rer. nat. Peter B. Becker

Identification and characterization of a protein complex involved in chromatin architecture and the evolution of *Drosophila* species

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften (*Dr. rer. nat.*)
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von

Andrea Lukacs

aus Rom

2023

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Betreuer:	Prof. Dr. rer. nat. Axel Imhof
Zweitgutachter:	PD Dr. rer. nat. Kai Hell
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	25 März 2024
Tag der Abgabe:	05 April 2023

Eidesstattliche Versicherung (Affidavit)

I hereby declare, that the submitted thesis entitled

Identification and characterization of a protein complex involved in chromatin architecture and the evolution of Drosophila species

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Munich, 05 April 2023

Ort, Datum

Andrea Lukacs

Unterschrift Doktorandin/Doktorand

To my younger self and his dreams of exploring the secrets of Nature, that one way or another brought me here

To my mother, my father, Bea, Lorenzo and my friends

To my tante Almut for the constant support

To my undefeatable feline companions Alice and Picki

“In the midst of winter, I found there was within me an invincible summer"

Albert Camus

“Nothing in Biology makes sense, if not in the light of Evolution”

Theodosius Dobzhansky

Table of Contents

Abbreviations

3C	Chromosome conformation capture
3D	Three dimensional
AP-MS	Affinity purification coupled with mass spectrometry (aka IP-MS)
aa	Amino acid
Act-Gal4	Gal4 driver expressed under the control of a constitutive actin promoter
BEAF32	Boundary Element Associated Factor 32
BESS	BEAF, Su-var(3-7), Stonewall like (domain)
bp	Basepair(s)
CD	Chromo domain
CSD	Chromo-shadow domain
Cenp-A	Centromeric protein A (aka CenH3, centromeric histone H3 variant, gene)
CENP-A	Centromeric protein A
CENP-C	Centromeric protein C
CENP-N	Centromeric protein N
Cen-H3	Centromere specific histone H3 (aka CENP-A)
ChIP-Seq	Chromatin Immunoprecipitation coupled with DNA Sequencing
CP-190	Centrosomal Protein 190
CTCF	CCCTC-binding factor
D.mau	*Drosophila mauritiana*
D.mel	*Drosophila melanogaster*
D.sim	*Drosophila simulans*
DamID	DNA adenine methyltransferase identification
DAPI	4',6-diamidino-2-phenylindole
Df(1)Hmr-	*Hmr* null mutant with a large N-terminal deletion including the promoter
DMM	Dobzhansky-Muller model
DCC	Dosage compensation complex
DNA	Deoxyribonucleic acid
E371K	Point mutation with substitution from glutamic acid to lysine at codon 371
FC	Follicle cells
FISH	Fluorescent *in situ* hybridization
FSC	Follicle stem cells
GCC	Germline cyst cell
gfzf	Glutathione-S-transferase-containing FLYWCH zinc finger protein (gene)
GFZF	Glutathione-S-transferase-containing FLYWCH zinc finger protein (prot)
GSC	Germline stem cells
G527A	Point mutation with substitution from glycine to alanine at codon 527
H2A	Histone 2A

H2B	Histone 2B
H3	Histone 3
H4	Histone 4
H3K9me2/3	Histone H3 lysine 9 di- or tri- methylation
H3K4me	Histone 3 lysine 4 methylation
H3K27me3	Histone 3 lysine 27 trimethylation
H4K20me	Histone 4 lysine 20 methylation
HeT-A	Healing Transposon
HI	Hybrid incompatibility
Hid	Head involution defective (activator of apoptosis)
HiC	High-throughput chromosome conformation capture
HJURP	Holiday junction recognition protein
Hmr/HMR	Hybrid male rescue (gene/protein)
Hmr^{dc}	*Hmr* mutant with a C-terminal deletion, presented here for the first time
Hmr^{1}	*Hmr* hypomorphic mutant with hybrids rescue phenotype
Hmr^{2}	*Hmr* mutant with two point mutations and with hybrids rescue phenotype
Hmr^{3}	*Hmr* hypomorphic mutant with hybrids rescue phenotype
Hmr-	generally referring to *Hmr* loss of function mutants or to *Df(1)Hmr-*
HP1a	Heterochromatin Protein 1a
HP3	Heterochromatic protein 3 (aka LHR)
Hs	Heat-shock
HTT	HeT-A, TART and TAHRE
IF	Immunofluorescence (microscopy)
IP	Immunoprecipitation
IP-MS	Immunoprecipitation coupled with mass spectrometry (aka AP-MS)
IRTG	Graduate program Integrated Research Training Group
Kb	Kilobase(s)
kDa	Kilodalton
Kc cells	*Drosophila* Kc cell line
LAD	Lamina associated domain
Lhr/LHR	Lethal hybrid rescue (gene/protein)
Lhr^{1}	*Lhr* hypomorphic mutant with hybrids rescue phenotype
Lhr^{2}	*Lhr* mutant with a precise deletion in a 16aa conserved region
LTR	Long Terminal Repeat (retrotransposons)
M1BP	Motif 1 Binding Protein
MADF	Myb/SANT-like in ADF1 (domain)
Mb	Megabase(s)
mod(mdg4)	Modifier of mdg4
MS	Mass spectrometry
NAD	Nucleolus associated domain
NGS	Next Generation Sequencing

Nlp/NLP	Nucleoplasmin-like protein (gene/protein)
Nph/NPH	Nucleophosmin (gene/protein)
OdsH	Odysseus-site homeobox
Prod	Proliferation disruptor
PEV	Position effect variegation
PcG	Polycomb-group proteins
PCH	Pericentromeric heterochromatin
piRNA	Piwi-interacting RNA
PRDM9	PR domain-containing protein 9
PTM	Post-translational modification (here mostly referred to histone tails)
QBM	Graduate School of Quantitative Biosciences Munich
Ras	Rat sarcoma (small GTPase involved in cell growth and proto-oncogene)
RNA Pol II	RNA Polymerase II
RNA	Ribonucleic acid
RNAi	RNA interference
SET	Su(var)3-9, Enhancer of Zeste, Trithorax (domain)
SL2 cells	Schneider's *Drosophila* Line 2 (cell culture)
SRM	Super-Resolution Microscopy
Su(var)3-9	Suppressor of variegation 3-9
Su(var)205	Suppressor of variegation 205 (aka HP1a)
Su(Hw)	Suppressor of Hairy wings
TAD	Topologically associated domain
TAHRE	Telomere Associated and HeT-A Related
TART	Telomere Associated Retrotransposon
TEs	Transposable elements
TSS	Transcriptional start site
UAS-Gal4	Genetic system borrowed from yeast to ectopically induce gene expression
UAS-Ras	Ras oncogene expressed under a Gal4-induced UAS enhancer
Zhr	Zygotic hybrid rescue

Publications

This cumulative thesis includes two published articles in peer-reviewed journals.
We described a protein complex involved in the formation of two *Drosophila* species, we dissected the role of different parts of the complex in the pure species and interspecific hybrids by using mutants to interfere with the complex formation, and, eventually, used an artificially induced system to reproduce hybrids-like conditions to perform high-throughput experiments that would have been otherwise not possible.
In addition, we describe and name for the first time two uncharacterized proteins and find the first evidence for a molecular interaction between two hybrid incompatibility proteins, for which previously only a genetic interaction was described.
My main work of thesis is included in the first publication, while the second publication stems from a successful collaboration with Dr. Jacob Cooper and Prof. Dr. Nitin Phadnis from the University of Utah.

In addition, as an integration to the work presented here, I also recommend reading the published work not included in this thesis, to which I contributed as a co-author. That will help frame my thesis within its context in chromatin biology and speciation and give a more complete overall picture.

Published work included in this thesis:

Andrea Lukacs, Andreas W. Thomae, Peter Krueger, Tamas Schauer, Anuroop V. Venkatasubramani, Natalia Y. Kochanova, Wasim Aftab, Rupam Choudhury, Ignasi Forne, & Axel Imhof (2021). *The integrity of the HMR complex is necessary for centromeric binding and reproductive isolation in Drosophila*. PLoS Genetics, 17(8), 1–27. https://doi.org/10.1371/journal.pgen.1009744

Jacob C. Cooper, **Andrea Lukacs.**, Shelley Reich, Tamas Schauer, Axel Imhof, & Nitin Phadnis (2019). *Altered localization of hybrid incompatibility proteins in Drosophila.* Molecular Biology and Evolution, 1–10. https://doi.org/10.1093/molbev/msz105

Published work to which I contributed, not included but relevant to this thesis:

Natalia Y. Kochanova, Tamas Schauer, Mathias G. Presley, **Andrea Lukacs**, Andreas Schmidt, Andrew Flatley, Aloys Schepers, Andreas W. Thomae, & Axel Imhof (2020). *A multi-layered structure of the interphase chromocenter revealed by proximity-based biotinylation*. Nucleic Acids Research, 16, 1–18. https://doi.org/10.1093/nar/gkaa145

Thomas A. Gerland, Bo Sun, Pawel Smialowski, **Andrea Lukacs**, Andreas W. Thomae, & Axel Imhof (2017). *The Drosophila speciation factor HMR localizes to genomic insulator sites*. PLoS ONE, 12(2), 1–18. https://doi.org/10.1371/journal.pone.0171798

Summary

Centromeres and pericentromeric heterochromatin are involved in a number of essential functions including cell division, silencing of repetitive DNA elements and spatial organization of the genome. The clustering of these regions around chromocenters during interphase is required for their proper function and involves a complex and finely tuned network of protein-protein interactions.

Despite their critical importance, centromeric and pericentromeric chromatin are poorly conserved even among closely related species and are, in fact, often involved in the formation of species. This apparent paradox is often the result of intrinsically fast-evolving repetitive DNA elements embedded into heterochromatin. The rapid evolution of such repetitive elements poses a threat to cellular fitness thereby exerting a selective pressure on the silencing machinery responsible for their management. Such process typically leads to the rapid coevolution of heterochromatin proteins acting like suppressors of fast-evolving DNA elements. This coevolution can, in turn, lead two populations once belonging to the same species to diverge by developing species-specific heterochromatin, that can eventually function as a postzygotic barrier between the two newly generated species.

The fruit fly sibling species *Drosophila melanogaster* (*D.mel*) and *Drosophila simulans* (*D.sim*) have been a privileged model for studying the formation of species since over a century and provide an excellent example of how hybrid incompatibilities can arise within chromatin. Hybrids from *D. melanogaster* mothers and *D. simulans* fathers fail to develop because of the detrimental genetic interaction of the three hybrid incompatibility (HI) genes *Hmr*, *Lhr* and *gfzf*.

Genetic studies of the three HI genes have revealed pleiotropic phenotypes, with *Hmr* and *Lhr* mutations disrupting oogenesis, female fertility and repetitive elements silencing and *gfzf* mutants having a broad spectrum of phenotypes including defects in cell cycle regulation.

In addition to their genetic interaction, HMR and LHR proteins interact within a complex network of protein-protein interactions that is important for the architecture and function of centromeric and pericentromeric chromatin. For GFZF, on the other hand, few molecular data are available and there is no evidence of molecular interactions with the other two HI genes.

Although the last decade has brought enormous progress in the field, the molecular details underlying both the divergent evolution of these hybrid incompatibility factors in the respective pure species and their genetic interactions in interspecific hybrids are still poorly understood.

A major twist in the field has been the finding that the expression of HMR and LHR proteins has diverged during the evolution in *D.mel* and *D.sim* species and that the two proteins are both overrepresented in hybrids. However, while these findings revealed the importance of a proper quantitative balance of HMR and LHR, the HMR/LHR protein-protein interaction network could only be studied in overexpressing conditions so far.

Characterizing the HMR/LHR protein complex in native conditions could therefore pave the way, on the one hand for understanding how these two proteins interact to mediate normal

pericentromeric and centromeric functions in pure species, and, on the other hand, how their interaction network is altered in hybrids where they are overexpressed.

To address these questions, in the first publication, we used affinity purification coupled with Mass Spectrometry (AP-MS) and revealed for the first time the existence of a stable six-subunit HMR/LHR protein complex in native conditions. In addition to HMR and LHR, the complex contains the two nucleolar proteins nucleoplasmin (NLP) and nucleophosmin (NPH), as well as the two non-characterized proteins, CG33213 and CG4788. For these last two proteins our publication provided the first molecular characterization and we named them Buddy Of HMR 1 (BOH1) and Buddy Of HMR 2 (BOH2), respectively. In addition, as a resource for the field, we published a detailed description of the intricate network of interactions (interactome) involving each complex component.

After identifying the complex we went further and generated two different mutants targeting two different *Hmr* domains, to dissect how HMR interacts with other complex components and how disrupting such interactions affects HMR localization and function. Our results suggest that the integrity of the HMR/LHR complex is necessary for both HMR physiological function in pure species and its toxic function in hybrids.

Next, we started from the HMR/LHR native complex and induced HMR/LHR overexpression to mimic a hybrid background in a cell culture system and asked how the HMR protein-protein interaction network is altered upon their overexpression. A range of new chromatin interactors, from architectural proteins like insulators to zinc-finger DNA binding proteins, appear to specifically interact when HMR is in excess, suggesting that these may be binding with low affinity and therefore only observed when HMR amount is not limiting, such as in hybrids.

Finally, we set out to study HMR subnuclear localization with respect to CENP-A and HP1a, a centromeric and a pericentromeric marker, respectively. Our findings allow us to build a model that reconciles previous controversies and suggests that HMR is neither centromeric nor pericentromeric but it is rather sitting in the middle, bridging these two types of chromatin by forming a complex that interacts with both.

In the second publication, we focused on the third and less known HI factor and asked how GFZF localizes in *D.mel*, *D.sim* and hybrids and how it molecularly interacts with HMR to cause hybrid incompatibility. We used *in situ* hybridization in polytene chromosomes to describe for the first time GFZF localization in *D.mel*, *D.sim* and hybrids. In addition, here we show the first evidence of a molecular interaction between GFZF and HMR by looking at their respective localization in both polytene chromosomes and cell lines where HMR was present in physiological amounts or overexpressed with LHR (hybrid-mimicking condition). Strikingly, while the two HI proteins HMR and GFZF occupy distinct and non-overlapping territories in pure species, their localization merges in the hybrid background.

Zusammenfassung

Zentromere und perizentromerisches Heterochromatin sind an einer Reihe wesentlicher Funktionen beteiligt, darunter die Zellteilung, die Stilllegung repetitiver DNS-Elemente und die räumliche Organisation des Genoms. Die Bündelung dieser Regionen um die Chromozentren während der Interphase ist für ihre ordnungsgemäße Funktion erforderlich und benötigt ein komplexes und fein abgestimmtes Netz von Protein-Protein-Interaktionen.

Trotz ihrer entscheidenden Bedeutung sind das zentromerische und das perizentromerische Chromatin selbst bei eng verwandten Arten oft nur wenig konserviert und an der Entstehung der Arten beteiligt. Dieses scheinbare Paradoxon ist oft das Ergebnis von sich schnell entwickelnden repetitiven DNS-Elementen, die in das Heterochromatin eingebettet sind. Die schnelle Evolution solcher repetitiven Elemente stellt eine Bedrohung für die zelluläre Fitness dar und übt dadurch einen Selektionsdruck auf die Stilllegung-Maschinerie aus, der zur schnellen Koevolution von Heterochromatin-Proteinen führt, die wie Suppressoren wirken. Diese Koevolution kann dazu führen, dass zwei Populationen, die einst zur selben Art gehörten, artspezifisches Heterochromatin entwickeln, das wiederum als postzygotische Barriere zwischen den beiden neu entstandenen Arten fungieren kann.

Die Fruchtfleigen Schwesterarten, *Drosophila melanogaster* (*D.mel*) und *Drosophila simulans* (*D.sim*) sind seit über einem Jahrhundert ein bevorzugtes Modell für die Untersuchung der Artbildung und liefern ein hervorragendes Beispiel dafür, wie Hybridinkompatibilitäten innerhalb des Heterochromatins entstehen können. Hybride aus *D. melanogaster*-Müttern und *D. simulans*-Vätern sind aufgrund der schädlichen genetischen Interaktion der drei Hybrid-Inkompatibilitätsgene (HI) *Hmr*, *Lhr* und *gfzf* nicht lebensfähig.

Genetische Untersuchungen der drei HI-Gene haben pleiotrope Phänotypen ergeben, wobei *Hmr*- und *Lhr*-Mutationen die Oogenese, die weibliche Fertilität und die Unterdrückung der Expression von repetitiven Elementen stören und *gfzf*-Mutanten ein breites Spektrum an Phänotypen aufweisen, einschließlich einer fehlerhaften Regulierung des Zellzyklus.

Zusätzlich zu ihrer genetischen Interaktion wirken die HMR- und LHR-Proteine innerhalb eines komplexen Netzwerks von Protein-Protein-Interaktionen, die für die Architektur und Funktion des zentromerischen und perizentromerischen Chromatins wichtig sind. Für GFZF hingegen liegen nur wenige molekulare Daten vor, und es gibt keine Hinweise auf molekulare Wechselwirkungen mit den beiden anderen HI-Genen.

Obwohl das letzte Jahrzehnt enorme Fortschritte auf diesem Gebiet gebracht hat, sind die molekularen Details, die sowohl der divergenten Entwicklung dieser Hybridunverträglichkeitsfaktoren in den jeweiligen reinen Arten als auch ihren genetischen Interaktionen in interspezifischen Hybriden zugrunde liegen, noch immer kaum verstanden.

Eine wichtige Wendung auf diesem Gebiet war die Feststellung, dass die Expression von HMR- und LHR-Proteinen während der Evolution in den reinen Arten *D.mel* und *D.sim* divergiert hat und dass beide Proteine in Hybriden überrepräsentiert sind. Während diese Ergebnisse die Bedeutung eines angemessenen quantitativen Gleichgewichts von HMR und LHR aufzeigten,

konnte das HMR/LHR-Protein-Protein-Interaktionsnetzwerk bisher nur unter überexprimierenden Bedingungen untersucht werden.
Die Charakterisierung des HMR/LHR-Proteinkomplexes unter nativen Bedingungen könnte daher den Weg ebnen, um einerseits zu verstehen, wie diese beiden Proteine interagieren, um die normale perizentromerische und zentromerische Funktion in reinen Spezies zu vermitteln, und andererseits, wie ihr Interaktionsnetzwerk in Hybriden, in denen sie überexprimiert werden, verändert wird.
Um diese Fragen zu klären, haben wir in der ersten Veröffentlichung eine Affinitätsreinigung in Verbindung mit Massenspektrometrie (AP-MS) durchgeführt und zum ersten Mal die Existenz eines stabilen HMR/LHR-Proteinkomplexes mit sechs Untereinheiten unter nativen Bedingungen nachgewiesen. Neben HMR und LHR enthält der Komplex die beiden nukleolaren Proteine Nucleoplasmin (NLP) und Nucleophosmin (NPH) sowie die beiden nicht charakterisierten Proteine CG33213 und CG4788. Für die beiden letztgenannten Proteine haben wir hier die erste molekulare Charakterisierung vorgenommen und sie Buddy Of HMR 1 (BOH1) bzw. Buddy Of HMR 2 (BOH2) genannt. Darüber hinaus haben wir eine detaillierte Beschreibung des komplizierten Interaktionsnetzwerks der einzelnen Komponenten des Komplexes vorgelegt.
Nachdem wir den Komplex identifiziert hatten, gingen wir noch einen Schritt weiter und verwendeten zwei verschiedene Mutanten, die zwei unterschiedliche *Hmr*-Domänen verändern, um zu untersuchen, wie sie mit anderen Komponenten des Komplexes interagieren und wie eine Unterbrechung dieser Interaktionen die Lokalisierung und Funktion des HMR Proteins beeinflusst. Unsere Ergebnisse legen nahe, dass die Integrität des HMR/LHR-Komplexes sowohl für die physiologische Funktion des HMR in reinen Arten als auch für seine toxische Funktion in Hybriden notwendig ist.
Als Nächstes gingen wir vom nativen HMR/LHR-Komplex aus und induzierten eine HMR/LHR-Überexpression, um einen hybriden Hintergrund in einem Zellkultursystem zu imitieren, und untersuchten, wie sich das HMR-Protein-Protein-Interaktionsnetzwerk bei seiner Überexpression verändert. Eine Reihe neuer Chromatin-Interaktoren, von architektonischen Proteinen wie Isolatoren bis hin zu Zink-Finger-DNS-Bindungsproteinen, scheinen spezifisch zu interagieren, wenn HMR im Überschuss vorhanden ist, was darauf hindeutet, dass diese möglicherweise mit geringer Affinität binden und daher nur beobachtet werden, wenn die HMR-Menge nicht limitierend ist, wie z. B. in Hybriden.
Schließlich untersuchten wir die subnukleare Lokalisierung von HMR in Bezug auf CENP-A und HP1a, einem zentromerischen bzw. perizentromerischen Marker. Unsere Ergebnisse erlauben es uns, ein Modell zu erstellen, das frühere Kontroversen in gewisser Weise ausräumt und nahelegt, dass HMR weder zentromerisch noch perizentromerisch ist, sondern eher in der Mitte sitzt und diese beiden Arten von Chromatin überbrückt, indem es einen Komplex bildet, der mit beiden interagiert.
In der zweiten Publikation konzentrierten wir uns auf den dritten und weniger bekannten HI-Faktor und fragten, wie GFZF in *D.mel*, *D.sim* und Hybriden lokalisiert ist und wie er mit HMR molekular interagiert, um Hybridinkompatibilität zu verursachen. Mit Hilfe der *In-situ*-Hybridisierung von Polyten-Chromosomen konnten wir erstmals die Lokalisierung von GFZF in *D.mel*, *D.sim* und

Hybriden beschreiben. Hier konnten wir den ersten Nachweis für eine molekulare Interaktion zwischen GFZF und HMR liefern, indem wir ihre jeweilige Lokalisierung sowohl in Polyten-Chromosomen als auch in Zelllinien untersuchten, in denen HMR in physiologischen Mengen vorhanden war oder mit LHR überexprimiert wurde (Hybrid-ähnliche Bedingungen). Auffallend ist, dass die beiden HI-Proteine HMR und GFZF in reinen Spezies unterschiedliche und sich nicht überschneidende Territorien besetzen, während ihre Lokalisierung im Hybrid-Hintergrund verschmilzt.

1. Introduction

1.1. Chromatin organizes genome structure and regulates its function

1.1.1. Nucleosomes are the basic structural unit of chromatin

Eukaryotic genomes are packaged together with different protein and RNA molecules into a complex structure called chromatin. This macromolecular architecture responds to the need of fitting DNA molecules as large as the almost 2 meters long human genome into a nucleus spanning only a few micrometers in diameter. On the other hand, chromatin must be dynamic enough to continuously reorganize and allow access to the genetic information, in order to regulate its expression, replication and repair.

Histones are the prevalent chromatin proteins and they form, together with DNA, repetitive structures called nucleosomes that resemble beads-on-a-string under the electron microscope, and that are the basic structural units of chromatin (Fig. 1). The histone proteins H3, H4, H2A and H2B are organized in an octamer, constituted by a tetramer with two molecules each of H3 and H4, held between two dimers formed by H2A and H2B (Luger et al., 1997). A nucleosome consists of 147 base pairs (bp) of DNA wrapped in 1.65 left-handed superhelical turns around a histone octamer (Luger et al., 1997). The N-terminal parts of histones, protruding as tails from the octamer structure, can be modified (histone post-translational modifications or PTMs), and the combination of these modifications has been proposed to form the histone code, a defined regulatory system that allows a fine modulation of the chromatin structure and the expression of the underlying genes (Jenuwein & Allis, 2001). These modifications are handled by a specific set of proteins: "writers", "readers" and "erasers", that are responsible for the deposition, the recognition, and the removal of histone marks, respectively, and whose expression can in turn be regulated (reviewed in Cosgrove, 2012). Indispensable for giving chromatin its dynamic properties are other proteins, chromatin remodelers, that modulate nucleosomes spacing and phasing through eviction, deposition and sliding (reviewed in Baldi, Korber & Becker 2020). Further regulation is provided by different histones variants with different properties that can substitute conventional histones within nucleosomes (reviewed in Baldi, Korber & Becker 2020). Besides nucleosomes and their modulators, chromatin also includes a large number of other proteins involved among others in transcription, replication and DNA damage repair. A steadily increasing number of small and large non-coding RNAs has also been found to participate in both chromatin structure (Schubert et al., 2012) and active complexes (Akhtar et al., 2000; Franke & Baker, 1999).

1.1.2. Chromatin has a complex and dynamic structure

On top of nucleosomes, chromatin is organized into higher order structures that can span large portions of the genome and allow the compartmentalization and coregulation of defined genomic regions (Fig. 1).

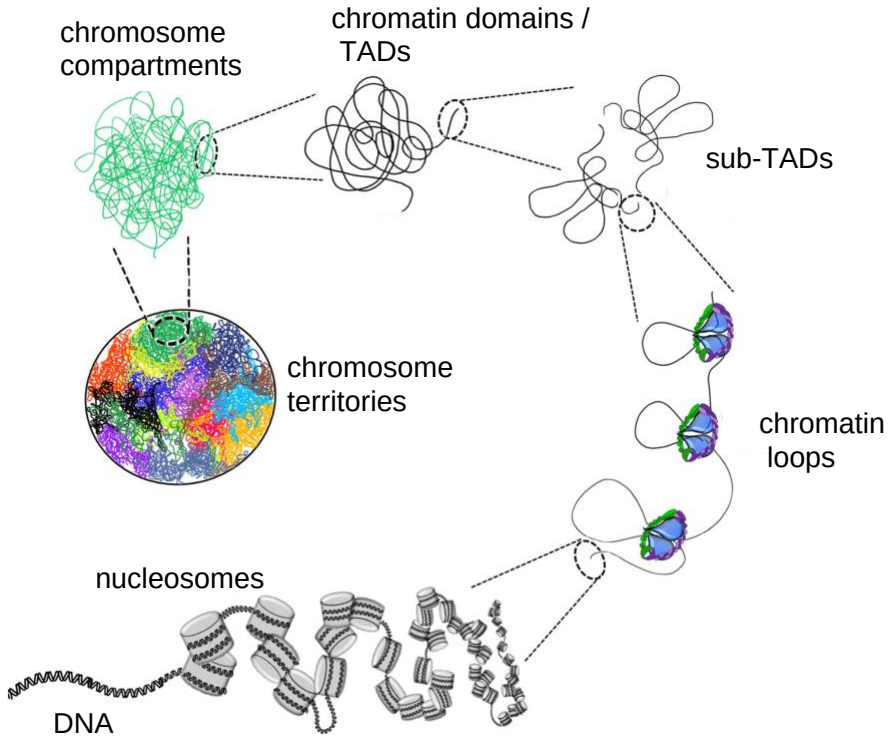


Fig. 1. The modular organization of chromatin. DNA is wrapped into histones that are, in turn, organized into higher order structures. Architectural proteins like insulators, condensin and cohesin can form chromatin loops. At a higher scale, chromatin domains of similar composition are usually organized into topologically associated (TADs). Specific parts of chromosomes like centromeres and telomeres, are characterized by specific chromatin types and form chromosome compartments. At the nuclear scale, each chromosome occupies a specific territory (Figure adapted from Magaña-Acosta & Valadez-Graham, 2020).

Almost one century ago, chromatin was first described to form two distinct types: on the one hand heterochromatin, intensely stained, highly condensed and genetically inert throughout the cell cycle and, on the other hand, euchromatin, gene-rich and with a looser structure that becomes invisible due to decondensation and lower staining contrast during mitosis (Heitz 1928).
These two types of chromatin have long been thought to underlie two well defined and separated structural organizations, with heterochromatin organized in a regular higher order structure, the 30 nm fiber, and euchromatin being mostly unstructured (reviewed in Sanulli & Narlikar, 2020; Baldi, Korber & Becker, 2020). This model has been challenged by several observations in higher

resolution thanks to more recent technologies in both genomics (Chromosome Conformation Capture, HiC) and imaging (Super-Resolution Microscopy, SRM). In fact, chromatin appears to be far less regularly structured, with even constitutive heterochromatin being much more dynamic than previously thought (reviewed in Di Stefano & Cavalli 2022; Sanulli & Narlikar, 2020; Baldi, Korber & Becker, 2020).
The emerging picture is a complex structural organization of the genome that is tightly connected to its function.

1.1.3. The linear genome is partitioned into chromatin domains

After the early subdivision into euchromatin and heterochromatin (Heitz 1928), nearly a century of studies using different methods and technologies have allowed looking at chromatin from different perspectives, describing both its linear and its spatial organization (reviewed in Schwartz & Cavalli, 2018; Bickmore & Van Steensel, 2013).
Before the genomics era kicked in, cytological studies in polytene chromosomes from salivary glands of the fruit fly *Drosophila melanogaster*, have allowed a systematic pre-genomic mapping of proteins by fluorescent *in situ* hybridization (FISH) (reviewed in Schwartz & Cavalli, 2018). These studies revealed the existence of a well-defined banding pattern along chromosomes as well as a specific binding for chromatin proteins, thereby demonstrating that interphase chromosomes are partitioned into stable chromatin domains on the linear scale. Despite their invaluable contribution to the understanding of chromatin organization, polytene chromosomes come from a polyploid tissue, making these observations difficult to generalize.
The subsequent advent of genomics as well as the continuous progress in microscopy towards super-resolution methods, allowed to expand these studies to more conventional cell types and learn general principles of chromatin organization in both the linear and the three-dimensional scale (3D) (Fig. 1) (reviewed in Schwartz & Cavalli, 2018; Bickmore & Van Steensel, 2013). Methods such as Chromatin Immunoprecipitation (ChIP-Seq; Ren et al., 2000; Barski et al., 2007; Robertson et al., 2007) and DNA adenine methyltransferase identification (DamID; van Steensel & Henikoff, 2000), relying on Next Generation Sequencing (NGS) techniques, were systematically used to map the genomic localization of chromatin proteins in high resolution.
With these genomic tools available, one could ask if chromatin could be partitioned into functionally defined domains by unique combinations of chromatin proteins and histones PTMs. A pioneering attempt to answer this question measured the genomic distribution of 53 non-histone chromatin proteins with DamID in *D. melanogaster* cultured cells (Filion, van Bemmel, Braunschweig et al., 2010). The resulting binding profiles were used to partition the *Drosophila* genome into domains characterized by five different *chromatin flavors*, corresponding to specific combinations of chromatin proteins, to which they assigned specific colors. Yellow and red to transcriptionally active regions (euchromatin), blue and green to facultative and constitutive heterochromatin, respectively, and black to a chromatin type mostly devoid of specific marks,

whose function remains rather obscure to date. A later work using ChIP-Seq profiles of histone marks, largely confirmed this model, although with a deeper genome partitioning into nine chromatin states (Kharchenko et al., 2011).
In conclusion, at the *linear scale*, distinct combinatorial patterns of chromatin factors partition the genome into distinct *chromatin domains* (also called epigenomic domains) (Fig. 1).
At the *chromosome scale*, however, these linear chromatin domains are unevenly distributed (Fig. 1). Large islands of green constitutive heterochromatin marked by the heterochromatin hallmark heterochromatin protein 1 (HP1a) and the histone PTM H3K9me2/3 typically accumulate in proximity to centromeres (the chromosome's kinetochore assembly sites) and to the chromosome's ends (telomeres). Throughout the rest of the genome yellow/red active euchromatin alternates with black/void chromatin domains and blue chromatin domains characterized by the binding of Polycomb-group proteins (PcG) associated with H3K27me3 (Filion, van Bemmel, Braunschweig et al., 2010; reviewed in Schwartz & Cavalli, 2018).

1.1.4. Higher order chromatin architecture is modular

The segmentation of the linear (or one-dimensional) genome reflects in many ways its three-dimensional architecture in the nuclear space.
Visual approaches such as immunofluorescence and DNA FISH have repeatedly shown that heterochromatin, with very few known exceptions (Solovei et al., 2009), associates with nuclear lamina at the nuclear periphery (Akhtar & Gasser, 2007; Falk et al., 2019) and with nucleoli (Padeken & Heun, 2014), a membraneless nuclear organelle where ribosome biogenesis occurs. In contrast, euchromatin and actively transcribed genes are typically localized in the nuclear interior (Falk et al., 2019). The purification and sequencing of DNA sequences associated with specific nuclear features like the nucleolus (Guelen et al., 2008) and the nuclear lamina (Nemeth et al., 2010; van Koningsbruggen et al., 2010) largely supported these observations with the newly-defined nucleolus-associated domains (NADs) and lamina-associated domains (LADs) mostly containing green HP1a heterochromatin.
The understanding of the basic principles underlying genome folding was dramatically boosted by the development of Hi-C, a technology based on the ‘‘3C’’ (chromosome conformation capture) method that allows the genome-wide mapping of chromatin contacts (Lieberman-Aidenet al. 2009).
This method associated with increasingly deeper sequencing revealed that metazoan genomes fold into distinct sub-megabase-sized modules, called physical domains or *topologically associated domains* (TADs) (Fig. 1), characterized by strong genomic interactions within them and sharply defined borders where the interactions are strongly depleted (reviewed in Schwartz & Cavalli, 2017). Although not universal, TADs have a surprising degree of conservation in evolution and among different cell types within the same species (Dixon et al. 2012).

The three-dimensional organization of the genome into TADs shows a remarkable agreement with its linear organization into chromatin domains, suggesting that TAD formation is mostly driven by homotypic or “like-with-like” interactions (Fig. 1) (Sexton & Cavalli, 2015; Schwartz & Cavalli, 2017). Chromatin domains of the same types, such as actively transcribed (red/yellow), HP1a heterochromatin (green) and Polycomb (blue), tend to establish contacts with each-other, within TAD borders. In addition to these intra-TAD contacts, similar chromatin domains also preferentially establish longer range inter-TADs contacts with each other.
An important role in the establishment, maintenance or reinforcement of TADs seem to be played by insulator elements and proteins, architectural factors typically enriched at TAD boundaries equipped with a notable ability to form long-distance connections (Schwartz & Cavalli, 2017). In addition, TAD borders are enriched in nucleosome-free regions and actively-transcribed housekeeping genes, where transcription may contribute to enforce the borders (Sexton & Cavalli, 2015; van Steensel & Furlong, 2019).
A further hierarchical order of chromatin organization are chromosome territories (Fig. 1). At the base of these structures seems to be long-range interactions among TADs that form domains in the range of several megabase pairs (Mb). Chromosome territories are detected by fluorescent *in situ* DNA hybridization with whole chromosome probes: each chromosome occupies a distinct portion of the nuclear space (Schwartz & Cavalli 2018; Bickmore & Van Steensel 2013).

1.1.5. Chromatin architecture orchestrates genome function

Early evidence that chromosome architecture can influence gene function came from genetic studies in flies, when Muller described the position effect variegation (PEV) for the *white* gene (Muller, 1930). This phenomenon was later also demonstrated for many other genes and showed that heterochromatin can exert a silencing effect for genes that are relocated in its proximity (reviewed in Schwartz & Cavalli, 2018).
In addition to these short-range effects involving flanking regions, many examples have shown long-range effects of higher order chromatin structure on distant loci, even on different chromosomes. This is the case for transvection in *Drosophila*, a phenomenon in which long-range chromatin contacts resulting in gene regulation depend on somatic homologous chromosome pairing (Lewis 1954).
Another notable example of long-range interactions are insulators (Gerasimova et al., 2000). In this case, the effects are more complex than directly inducing or repressing the specific contact loci, and they rather modulate gene expression by optimizing spatial organization (Gomez-Diaz & Corces, 2014).
Since these early genetic works, much more evidence has accumulated supporting the idea that chromatin architecture orchestrates genome function. Homotypic interactions tend to bring similar chromatin domains together in the nuclear space thereby promoting a more efficient co-regulation and co-expression of the underlying genes. TADs typically enclose chromatin domains of the

same type (Sexton & Cavalli, 2015; Schwartz & Cavalli, 2018).

At a larger scale, chromatin domains cluster together in the nuclear space, thereby promoting their own co-regulation. This is the case for several *nuclear bodies*, including nucleoli, which are involved in ribosome biogenesis, chromocenters, which are pericentric heterochromatin foci, Polycomb-group bodies and transcriptional factories associated with RNA Pol II (Erdel & Rippe, 2018; Sanulli & Narlikar, 2020). An increasing body of evidence is supporting the idea that such nuclear bodies result from self-organizing phase separation mechanisms with local chromatin features acting as nucleation sites (Erdel & Rippe, 2018; Sanulli & Narlikar, 2020). Notably, the vast majority of these bodies are observable during interphase and disassemble during cell division. Self-organizing phase separation mechanisms with local chromatin features acting as nucleation sites could facilitate their reliable and faithful reestablishment (Erdel & Rippe, 2018).

Chromatin domains are thought to function as anchor points that help establish nuclear architecture by restricting the position of the chromosomes within the three-dimensional space of the nucleus. However, while the nuclear localization of a locus correlates with its transcription, this does not seem to translate into a direct role in gene regulation. Instead, nuclear localization appears to act downstream of transcription factors and chromatin context for gene regulation and to be more important to reduce the ‘‘noise’’ of upstream regulation events by ‘‘locking in’’ a particular transcriptional status (Padeken & Heun, 2014; Schwartz & Cavalli, 2018; van Steensel & Furlong, 2019).

Additionally, while genome organization influences gene expression, the opposite is also true, as there is clear evidence that transcription affects genome organization (reviewed in van Steensel & Furlong, 2019).

A particularly relevant role in organizing the 3D genome is played by the HP1-marked green constitutive heterochromatin, which will be the focus of the next section.

1.2. Heterochromatin is critical for genome structure and function

Heterochromatin is fundamental for various aspects of genome function: (i) it mediates gene silencing, (ii) protects genome integrity by repressing recombination and mobilization of selfish repetitive elements, (iii) helps organizing specialized structures such as centromeres and telomeres, and (iv) forms dense clusters that function as anchor points for the genome, thereby regulating nuclear morphology and rigidity (Allshire & Madhani, 2017; Sanulli & Narlikar, 2020).

1.2.1. Heterochromatin formation and spreading involves a tightly balanced network of protein-protein interactions

Constitutive heterochromatin (referred to as heterochromatin hereafter) is characterized by H3K9me2/3 stretches bound by HP1a (green chromatin) and is particularly concentrated at telomeric and pericentromeric regions characterized by repetitive DNA elements, including both satellite DNA and transposable elements (Allshire & Madhani, 2017; Nicetto & Zaret, 2019).
Depending on the cell type and the stage of development, different heterochromatic domains from the same - but also from different - chromosomes coalesce in the 3D nuclear space and localize around the nucleolus or at the nuclear periphery (Fawcett, 1966, Padeken & Heun, 2014; Schwartz & Cavalli, 2018; Bickmore & van Steensel, 2013). The localization of heterochromatin to such matrices could then act as a constraint to organize the rest of the genome. As a consequence, heterochromatin is concentrated, which in turn enforces silencing. This compartmentalization does not seem to regulate transcription *per se*, but, instead, to stabilize the transcriptional ‘‘ground-state’’ of a gene, acting downstream of transcription factors and chromatin context, to reduce upstream regulatory events (Padeken & Heun, 2014; Schwartz & Cavalli, 2018; Bickmore & van Steensel, 2013).
Some additional HP1a heterochromatin islands are found in euchromatic regions and have been shown to also cluster with pericentromeric heterochromatin in flies (Lee et al., 2020).
Heterochromatin establishment involves a histone PTM reader-writer coupling mechanism (reviewed in Allshire and Madhani, 2017). In *Drosophila* and humans, the H3K9 methyltransferases SU(VAR)3-9 is equipped with both the reader and the writer domains, an N-terminal chromodomain and a C-terminal Su(var)3-9, Enhancer of Zeste, Trithorax (SET) domain, respectively. Methylation of H3K9 by the SET domain enforces the further recruitment of more SU(VAR)3-9 through the binding of H3K9me2/3 by its chromodomain. HP1a protein, in turn is recruited to these methylated histone tails through its chromodomain and can dimerize or recruit other effector proteins through its chromoshadow domain (a protein-protein interaction domain). This positive loop allows heterochromatin to spread from its nucleation site in a DNA sequence independent manner.
Heterochromatin spreading has been proposed to respond to a mass-action law requiring the formation of large macromolecular complexes involving a large number of proteins (Tartof et al. 1989). A surplus of unassembled heterochromatin components is required for spreading. Therefore, heterochromatin formation and maintenance is the result of the cooperative efforts of many chromatin components that are tightly connected in a network of protein-protein interactions and whose balance must be carefully tuned within the nucleus. The relative concentration of multiple heterochromatin components determines the extent to which heterochromatin would silence the genes immediately adjacent. Overexpression can result in deleterious spreading and silencing. On the other hand, when free-roaming heterochromatin factors are limiting, their recruitment to sites of spreading can weaken the repression at the sites where they are recruited from (Schwartz & Cavalli, 2018; Allshire & Madhani, 2017).

A functional consequence of heterochromatin spreading has been described in position effect variegation in *Drosophila:* genes placed in proximity can be accidentally silenced because of heterochromatin spreading **(**Schwartz & Cavalli, 2018**)**.
In addition to this reader-writer mechanism involving HP1a and H3K9me2/3, heterochromatin spreading may also involve the HP1a biophysical propensity to phase-separate and therefore form membraneless compartments, which would allow efficient local concentration of heterochromatin proteins and the control of molecules traffic (Erdel & Rippe, 2018; Sanulli & Narlikar, 2020).
Despite its importance in genome architecture and function, heterochromatin spreading must be carefully limited, and this is achieved through boundary elements **(**Erdel & Rippe, 2018**)**: heterochromatin boundaries are typically enriched for insulator proteins but also for actively transcribed genes.
An important example of this tradeoff are centromeres, where a tight balance is necessary between centromeric chromatin marked by the H3 histone variant CENP-A (centromeric protein A) and HP1a-marked pericentromeric heterochromatin. On the one hand, pericentromeric chromatin constitutes a buffer zone preventing centromeres from invading other parts of the genome and resulting in *de novo* kinetochore formation (Olszak et al., 2011; Steiner & Henikoff, 2015; Fukagawa and Earnshaw 2013). On the other hand, heterochromatin must be prevented from spreading into CENP-A chromatin (Allshire & Madhani, 2017).
Despite heterochromatin being mostly silenced, a basal level of transcription resulting in the production of small non-coding RNAs is necessary for its establishment and maintenance. This system is particularly well understood in *Drosophila*, where piRNAs (PIWI-interacting RNAs) are transcribed from transposons relics and are used to recognize complementary sequences of nascent TE transcripts to initiate heterochromatin formation at these loci (Liu et al., 2020; Allshire & Madhani, 2017; Gu et al., 2013).

1.2.2. Heterochromatin and the nucleolus

The nucleolus is the archetype of the nuclear body and its membraneless compartmentalization is at least partially involving liquid-liquid phase separation (Brangwynne et al., 2011). Well visible inside the nucleus, nucleoli have been described across different eukaryotic phyla and are typically organized around repetitive sequences of ribosomal DNA (rDNA) (van Steensel & Furlong, 2020; Padeken & Heun, 2014). Here rRNA transcription and processing as well as ribosome biogenesis occur. The repetitive nature of rDNA allows the concomitant transcription of multiple rDNA genes but also poses a threat to the stability of the locus. Therefore, a tightly regulated structure is required to prevent internal recombination. Only rDNA copies that are looped inside the nucleolus are transcribed while silenced copies occupy its periphery where they accumulate repressive marks and they are coated by a shell of silencing factors and constitutive heterochromatin.

Besides their function as ribosome factories, nucleoli are critical for the structural organization of the genome as their periphery serves as an anchoring point for centromeres and pericentromeric heterochromatin (van Steensel & Furlong, 2020; Padeken & Heun, 2014). Similarly to what resulted from microscopy experiments, sequencing of nucleoli associated domains (NADs) has shown that nucleoli associate mostly with satellite DNA from centromeric and pericentromeric regions or from gene-poor and silent chromatin (Nemeth et al., 2010; Van Koningsbruggen et al., 2010).
Interestingly, it appears that NADs overlap substantially with LADs (lamina associated domains) suggesting that they are at least partially constituted of the same repressive chromatin that is stochastically positioned either at the nuclear lamina or at the nucleoli (van Steensel & Furlong 2019).

1.2.3. Heterochromatin and centromeres

The centromere was traditionally defined as the observable constriction that forms on mitotic chromosomes (Flemming 1882) and is now defined as the genomic region that serves as a platform for kinetochore assembly and chromosomal segregation (Fukagawa & Earnshaw 2014; Talbert & Henikoff, 2020). Despite its essential role in mitosis, centromere function extends well beyond: while the kinetochore only transiently assembles during cell division, the centromere persists as a stable domain throughout the cell cycle (Padeken et al., 2013).
In most organisms, centromeres are not genetically specified by a “magic DNA sequence” but rather epigenetically determined by the presence of nucleosomes containing the centromeric histone H3 variant CENP-A (Fukagawa & Earnshaw 2014; Talbert & Henikoff, 2020). Across evolution CENP-A-containing domains can span in size (reviewed in Fukagawa & Earnshaw 2014; Talbert & Henikoff, 2020) from one single nucleosome in the 125 bp “point centromere” of budding yeast to the entire length of the chromosome in the “holocentromeres” that are found in few animal and plant species. Besides these exceptions, most eukaryotes carry “regional centromeres” constituted of multiple CENP-A-containing nucleosomes confined in a specific region of the chromosome that goes from few kb in *S. pombe* to few Mb in humans.
The crystal structure of CENP-A-containing nucleosomes shows that rather than wrapping the canonical 147 bp of DNA, they only contain 121 bp of DNA (Tachiwana et al., 2011). This peculiar nucleosome structure is unable to bind the linker histone H1 and form the canonical higher order nucleosome fiber. Instead, Centromeric Protein N (CENP-N) has been proposed (Zhou et al., 2022) to act as a “centromere-specific linker histone” that on one side specifically reads the histone variant on CENP-A nucleosomes, while interacting on the other side with the DNA on a neighboring nucleosome, to potentially form unique chromatin structures at the centromere. This specialized chromatin structure built upon the CENP-A nucleosome is necessary for the assembly of the kinetochore.
In regional centromeres, CENP-A-containing nucleosomes are interspersed among canonical

nucleosomes (in a 1:6-1:25 ratio) and occupy a genomic region constituted by large arrays of repetitive elements (satellites or retrotransposons) (Fukagawa & Earnshaw 2014; Talbert & Henikoff 2020). Because DNA sequence doesn't seem to be essential for the formation of CENP-A domains, the presence of flanking repetitive sequence could provide a buffer zone that stabilizes centromeres and prevent deleterious effect in case of a drift in kinetochore positioning (Fukagawa & Earnshaw 2013). On both sides, centromeres are surrounded by pericentromeric heterochromatin that is characterized by satellite monomers covered in constitutive heterochromatin marks such as H3K9me2/3 and HP1a (Fukagawa & Earnshaw 2013). Although pericentromeric heterochromatin is not essential for centromere formation and maintenance it has been proposed to provide an additional layer of stability while also preventing centromeres spreading into euchromatic sites (Olszak et al., 2011; Steiner & Henikoff, 2015; Fukagawa & Earnshaw 2013).

1.2.4. Heterochromatin and telomeres

Chromosome ends are not just blunt edges but are rather constituted by specialized structures called telomeres (reviewed in Chakravarti et al., 2021; Casacuberta, 2017). The need for specialized structures protecting from chromosome end fusion and consequent genetic aberrations, was first postulated and then proved already in the 30s by Barbara McClintock and HJ Muller (Creighton & McClintock 1931; Muller, 1938). Telomeres are composed of repetitive nucleotide sequences that form a ''cap structure'' that maintains chromosome integrity (Chakravarti et al., 2021).

In addition, because of the "end replication problem", i.e. the cellular DNA polymerases being incapable of moving in a 3' to 5' direction, chromosome ends require specialized replication mechanisms to prevent chromosome loss at each replication cycle (Olovnikov, 1973; Watson, 1972). In most eukaryotes, a specialized polymerase, called telomerase, takes care of telomeres elongation by carrying its own RNA template (Chakravarti et al., 2021). An interesting exception to this rule are *Drosophila* telomeres, which adopt a telomerase-independent elongation mechanism (Pardue & DeBaryshe, 2008; Casacuberta, 2017, Abad et al., 2004). *Drosophila* telomeres are composed of several kilobases (kb) of head-to-tail arrays of three specialized non-Long Terminal Repeat (non-LTR) retrotransposons, Healing Transposon (HeT-A), Telomere Associated Retrotransposon (TART) and Telomere Associated and HeT-A Related (TAHRE) (collectively HTT). In this organism, telomere elongation is achieved with targeted retrotransposition of these three very special retroelements that are always transposing in the same orientation, exclusively at the end of the chromosomes (Pardue & DeBaryshe, 2008; Casacuberta, 2017; Abad et al., 2004).

Despite being actively transcribed for their own maintenance and retrotransposition, *Drosophila* HTT telomeric repeats are embedded into HP1a heterochromatin, which is necessary to maintain telomeres at a proper length and prevent their fusion (Casacuberta, 2017).

1.2.5. Pericentromeric heterochromatin and centromeres coalesce at chromocenters

Pericentromeric heterochromatin coalesce at the periphery of the nucleolus during interphase forming visible nuclear bodies called chromocenter because of their characteristic intense staining with a DNA-binding dye (DAPI, 4',6-diamidino-2-phenylindole) in fluorescence microscopy (Fig. 2) (Probst & Almouzni, 2008; Erdel & Rippe, 2018).

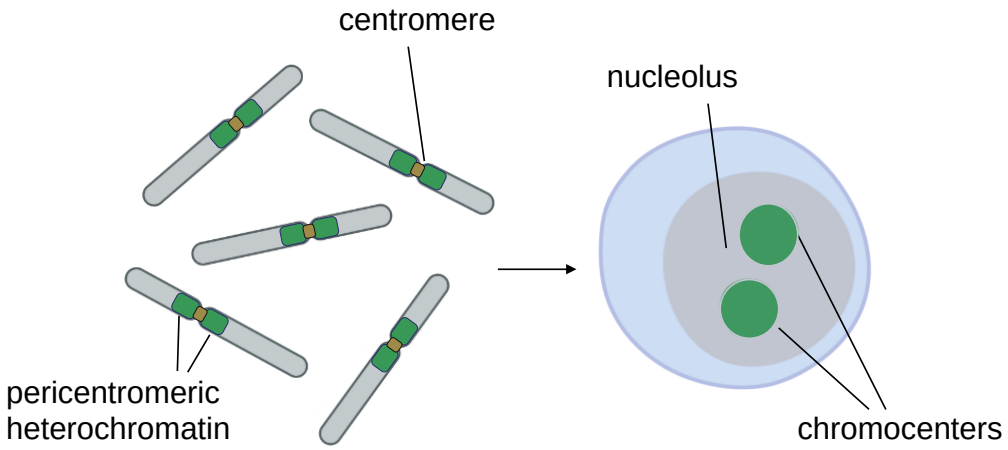


Fig. 2. Chromocenters formation. During interphase, pericentromeric heterochromatin from different chromosomes clusters at the periphery of nucleoli along with centromeric chromatin, forming chromocenters (model based on (Padeken & Heun, 2014; Jagannathan et al., 2019)).

The mechanisms of chromocenter formation are dependent on both the integrity of the nucleolus (Peng & Karpen, 2007) and of centromeric clusters (Padeken et al., 2013). In *Drosophila*, during interphase, centromeres cluster together at the periphery of the nucleolus and seem to function as an anchoring point to drag pericentromeric heterochromatin and form chromocenters (Padeken & Heun, 2014; Bickmore & van Steensel, 2013).

The clustering of centromeres at nucleoli, in turn, is dependent on several chromatin proteins (Padeken et al., 2013; Padeken & Heun, 2014; Anselm et al., 2018).

Nucleophosmin-like Protein (NLP) oligomerizes with Nucleophosmin (NPH) and together they bind to Hybrid Male Rescue (HMR) (Anselm et al., 2018). Binding to HMR is necessary for NLP recruitment to centromeres (Anselm et al., 2018). NLP in turn is necessary for centromeric clustering and interacts with both the centromeric histone CENP-A and the nucleolar protein Modulo (Padeken et al., 2013).

Interestingly on superresolution microscopy, NLP and HMR do not overlap with CENP-A nor with HP1a (Anselm et al., 2018; Kochanova et al., 2021). Instead, they fill the space between centromeres and pericentromeric heterochromatin, possibly functioning as an anchor or a bridge for tethering these two structures together. The role of HMR and NLP interactions in bridging these two types of chromatin together has been investigated in this thesis and is the subject of the publication reported here (Lukacs et al., 2021, Publication 1).
In addition to its architecture, centromeric clustering has also been shown to be necessary - although not sufficient - for proper pericentromeric heterochromatin silencing (Padeken et al., 2013).
In *Drosophila*, the clustering of centromeric and pericentromeric heterochromatin at chromocenters cooperates to create interchromosomal cross-links that bundle different chromosomes together (Fig. 2). While the clustering of different centromeres together seems to be mostly an epigenetic feature, including factors like CENP-A, NLP, HMR and Modulo (Padeken et al., 2013; Padeken & Heun, 2014; Anselm et al., 2018), for pericentromeric heterochromatin the process seems to also involve the binding of chromosome-specific satellite DNA sequences (Jagannathan et al., 2019).
Whatever the mechanism that allows the formation of such macromolecular structures, chromocenter integrity is important for normal cell function and its disruption leads to micronuclei formation and cell death (Jagannathan et al., 2019) and seems to be also involved in the causing lethality in interspecific hybrids (Jagannathan & Yamashita, 2021).

1.3. Chromatin is involved in the formation of species

1.3.1. Evolution of species and postzygotic isolation

In “On the Origin of Species”, Darwin put the continuous process of formation and extinction of species at the center of the evolution of biological diversity on our planet (Darwin 1859).
About 80 years later, in his foundational work “Genetics and the Origin of Species”, Dobzhansky provided a modern synthesis of Darwinism and Mendelism (Dobzhansky 1937), recognizing in genetic variation the raw material for natural selection to shape biodiversity.
At the core of his new vision is the idea that a continuum of genetic variability within populations provides the base for adaptations to different environments that can in turn result in the formation of separate species, thereby interrupting this continuity (Dobzhansky 1937; Lewontin 1997). In his view, biological variation and species discontinuity can be described by thinking of species as groups of organisms with unique constellations of genes separated from each other (Fig. 3). Each unique constellation of genes is only fit and adaptive in a specific environment. In a hypothetical multidimensional space with all possible gene combinations, there would be adaptive peaks where a particular combination of genes is suitable to a particular habitat and there would be maladaptive valleys where the gene combination is unfit (Fig. 3). In this view the variation is

represented by the infinite possible combinations of genes and the discontinuity from the fact that each species occupies only a specific adaptive peak and is instead surrounded by maladaptive valleys that separates it from other species. As a consequence, hybridization among different species often results in the reshuffling of two adaptive constellations of genes generating intermediate genotypes most likely falling into a maladaptive valley.

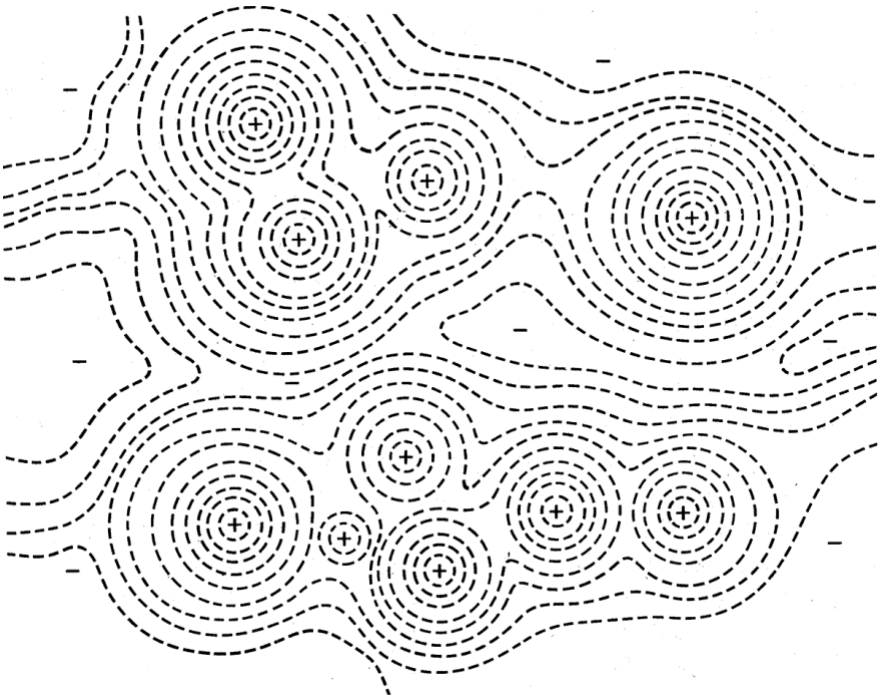


Fig. 3. Adaptive peaks and valleys in the field of gene combinations according to Dobzhansky (illustration from Dobzhansky's "Genetics and the Origin of Species", 1937). The contour lines encompass genotypes with comparable adaptive values. Populations occupy adaptive peaks (+) while the maladaptive valleys (-) represent genotypic combinations with low adaptive value and therefore unfit.

What are the mechanisms and the driving forces for the formation of species has remained one of the most central questions in Evolutionary Biology. In what is the most widely accepted definition of species, Mayr defined them as "groups of interbreeding natural populations that are reproductively isolated from other such groups" (Mayr, 1942).

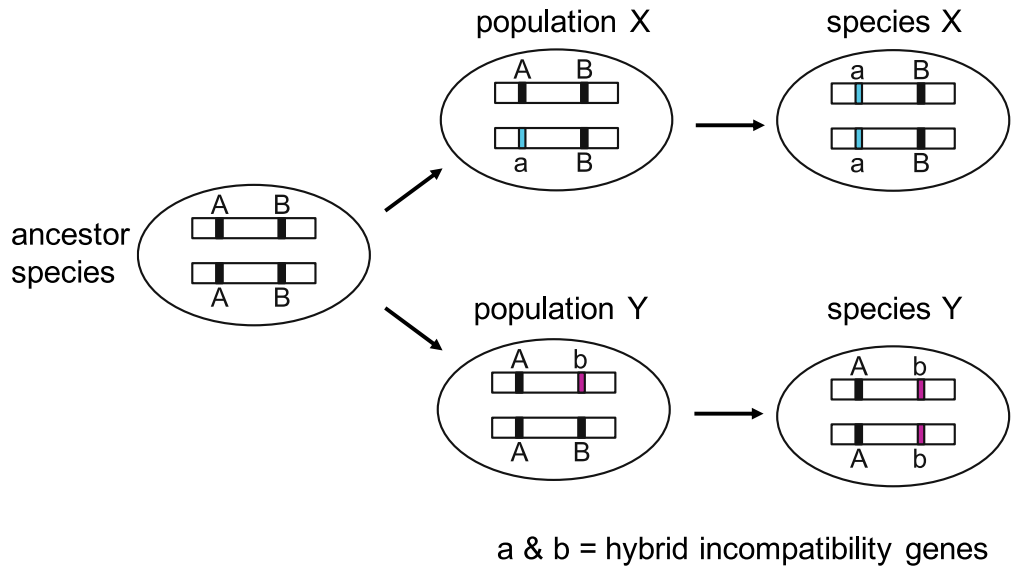
The formation of species, or speciation, is therefore a mechanism that involves the formation of reproductive barriers that create a discontinuity between two different populations that once belonged to the same species. This discontinuity in the genotypic and phenotypic variation is assured by reproductive isolation, the interruption of the genetic flow between two different species. Without such reproductive barriers, specific adaptations would be lost in favor of intermediate phenotypes (Seehausen et al., 2014; Orr, Masly & Phadnis, 2007; Coyne & Orr 1989).

Mechanisms leading to reproductive isolation can be divided in pre-mating and post-mating (Seehausen et al., 2014; Orr, Masly & Phadnis, 2007; Coyne & Orr 1989). The first group includes extrinsic forces such as ecological or behavioral barriers that prevent the two different species from mating and therefore results in the complete absence of interspecific hybrids. In post-mating isolation, instead, the two species can mate but the resulting hybrids are sterile, lethal or maladapted to their environment. Post-mating isolation can be further divided into pre-zygotic or postzygotic isolation, with barriers that prevent fertilization or that act after fertilization, respectively.

1.3.2. Hybrid incompatibilities are common postzygotic barriers in speciation

Postzygotic isolation is intrinsic to the two species involved and is often characterized by negative epistatic interactions between genetic loci that are incompatible in interspecific hybrids (Fig. 4) (Sawamura 2011, Seehausen et al., 2014; Orr, Masly & Phadnis, 2007; Coyne & Orr 1989). The most popular theoretical base for these observations is provided by the Dobzhansky-Muller model (DMM) (Fig. 4) (Dobzhansky, 1936; Muller and Pontecorvo, 1942). In this model an ancestral population with genotype AABB is split into two populations. While in one population the locus A diverges into a, in the other population the locus B diverges into b. Over time these new alleles get fixed in the respective populations (Fig. 4A). The new alleles a and b are incompatible in the same organism and are therefore named hybrid incompatibility (HI) genes (Fig. 4B). This causes no harm in pure species, since the diverged alleles are never present together, but leads to negative epistatic interactions in hybrids between these two species.

A



B

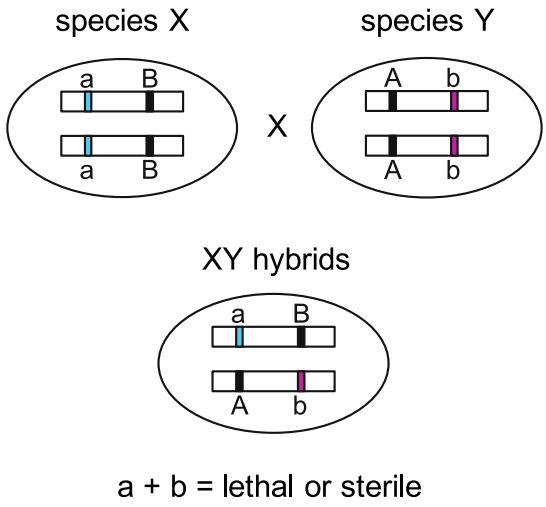


Fig. 4. The Dobzhansky-Muller model for postzygotic isolation. In two populations once belonging to the same ancestor species, two genes A and B can diverge because of positive selection or simply by genetic drift. The resulting alleles a and b can be fixed in the respective populations (A). When the two diverged alleles a and b genetically interact and are incompatible in the same genetic background (i.e. in interspecific hybrids), they are called hybrid incompatibility genes, and they constitute a reproductive barrier that makes the two new species postzygotically isolated (B).

The divergence of the ancestral genes into HI genes can be simply caused by genetic drift but it is more often accompanied by a signature of positive selection indicating that specific adaptations

are involved (Sawamura 2011; Seehausen et al., 2014; Orr, Masly & Phadnis, 2007; Coyne & Orr 1989).
Due to its simplicity, the DMM has been applied universally across different taxa to explain postzygotic isolation, but speciation is often involving more complex and multigenic mechanisms. Multiple observations support the idea that genomic conflicts can lead to HI formation (Sawamura 2011, Seehausen et al., 2014; Orr, Masly & Phadnis, 2007). Genomic conflicts can arise from meiotic drive and sexual selection as well as an arms race between selfish DNA elements and their suppressors.

1.3.3. Heterochromatin is fast evolving and drives the evolution of hybrid incompatibilities and the formation of species

Many of the genomic conflicts that lead to the formation of HIs, have been described to happen at heterochromatin.
The genomic sequences embedded into heterochromatin, including both satellite DNA and transposable elements, due to their repetitive nature are intrinsically prone to evolve rapidly by replication slippage, unequal crossovers or transposition (Liu et al., 2020; Ausio et al., 2019). The resulting expansion or contraction of copy numbers as well as sequence divergence can lead to an impressive divergence in these regions both within and among closely related species (Naville et al., 2019; Clark et al., 2007). Consistently, a meta-analysis of different types of cancers showed that up to 40% of the mutations found in cancers occurred in H3K9me3 marked chromatin (Schuster-Bockler et al., 2012).
Within such an evolutionary dynamic environment, selfish DNA elements can expand their genetic baggage at the expense of the overall fitness of the organism. This poses a threat to cellular fitness, thereby constituting a selective pressure that paves the way for the rapid evolution of heterochromatin proteins acting as suppressor. This arms race between selfish genetic elements and cellular defense machinery can lead two populations once belonging to the same species to develop species-specific heterochromatin. This adaptations in turn can function as a postzygotic barrier since interspecific hybrids might lack the appropriate apparatus to silence the genome from the other species (Sawamura, 2011; Orr, Masly & Phadnis, 2007; Seehausen et al., 2014; Roach, Ross & Malik, 2013; Kursel & Malik, 2018).
Surrounded by pericentromeric heterochromatin, centromeres constitute a hub for genomic conflicts. A phenomenon called centromeric drive has been observed in *Drosophila* (Henikoff, Ahmad & Malik,, 2011; Roach, Ross & Malik, 2013; Kursel & Malik, 2018), where centromeres are competing to distort chromosome segregation in order to bias their own inclusion into the oocyte and assure their own transmission to the next generation, thereby avoiding the evolutionary dead-end of the polar body. In response, the rapid evolution of centromere-associated proteins rescues the meiotic imparity caused by selfish centromeric DNA.

As a result of centromeric drive, many centromeric proteins are under positive selection and many have been described to cause hybrid incompatibility, some of which are listed below.

One such example is the centromeric histone *Cenp-A*. Whereas most core histones are under purifying selection and widely conserved across evolution, *Cenp-A* has undergone adaptive evolution and incompatibilities between its variants developed even between closely related species (Maheshwari et al., 2015; Malik and Henikoff, 2001).

Drosophila has been serving as a popular model for the study of hybrid incompatibility.

In this context, the hybridization of *D. melanogaster* and *D. simulans* (*D. mel* and *D. sim*, respectively) has long been studied and brought tremendous progress to the field (Sturtevant, 1919; Barbash, 2010). Three chromatin proteins encoded by the genes *Hmr*, *Lhr* and *gfzf* cause hybrid incompatibility in the cross between *D.mel* mothers and *D.sim* fathers. The products of these genes are the main object of this thesis and will be discussed extensively in the next paragraphs.

Notably, the opposite cross between these two species, involves a species-specific pericentromeric region (zygotic hybrid rescue or *Zhr*), the 359bp satellite repeats, that is present in the *D.mel* X but absent in *D.sim*. The female hybrids from this cross die during embryogenesis because of lagging *D.mel* X-chromosomes, resulting in mitotic defects (Ferree and Barbash, 2009; Sawamura and Yamamoto, 1997).

Another example consistent with this scenario is the case of the *OdsH* (*Odysseus-site homeobox*) that encodes a chromatin protein that binds satellite DNA in a species-specific fashion in *D.sim* and *D. mauritiana* (Bayes and Malik, 2009; Ting et al., 1998). In hybrids, OdsH erroneously binds and decondenses the Y chromosome heterochromatin of *D. simulans*, causing male sterility (Bayes and Malik, 2009).

Further supporting the role of (peri)centromeric chromatin in reproductive isolation, hybrids of both *D.mel*/*D.sim* and *D.sim/D.mau*, whose peri-centromeric satellite repeats and the respective binding proteins have diverged in the pure species, undergo chromocenter disruption and micronuclei formation which is thought to be the cause for the observed developmental arrest (Jagannathan & Yamashita, 2021).

Besides *Drosophila*, heterochromatin has also been shown to cause HIs in vertebrates as it is the case for the speciation of sticklefish from the Japan Sea and the Pacific Ocean, that involved the rapid evolution of a heterochromatic protein with consequent divergence in its binding ability to pericentromeric H4K20me and loss of *gypsy* retrotransposons silencing (Yoshida et al., 2018).

In the two species of frogs *Xenopus laevis* and *X. tropicalis*, hybrids fail to develop because of centromere incompatibilities that seem not to be dependent on centromeric DNA sequence but rather on divergent evolution of CENP-A and its chaperone HJURP (Kitaoka et al., 2022).

The first mammalian HI gene was identified in mouse and is the H3K4 trimethyltransferase PR domain-containing protein 9 (PRDM9), that causes sterility in hybrid males by inducing meiotic arrest (Mihola et al., 2009).

1.4. Chromatin proteins HMR, LHR and GFZF lead to speciation in *D. melanogaster and D. simulans*

1.4.1. One century of *D. melanogaster* and *D. simulans* hybrids

The system that probably contributed the most to the current understanding of the genetics of postzygotic isolation is the hybridization between the two sibling species *Drosophila melanogaster* (*D.mel*) and *Drosophila simulans* (*D.sim*) (Fig. 5). Over one century ago, Sturtevant first crossed these two species and documented the outcome. With his great disappointment the resulting F1 hybrids were either unviable or infertile, depending on the direction of the cross (Sturtevant 1919). His disappointment stems from the fact that their post-reproductive isolation makes the mechanistic study of hybrid incompatibility extremely complex (reviewed in Barbash 2010).

The adult hybrid offspring of *D.mel* females and *D.sim* males is constituted exclusively by females (Fig. 5), while the reciprocal cross results in the reciprocal outcome, with only male offspring surviving to adulthood. The death of the lethal sex occurs during development, at larval stages or as pupae, respectively (reviewed in Barbash 2010).

Regardless of these limitations, over the years, complex and ingenious genetic studies have led to a much deeper understanding of the underlying genetic interactions and to the identification of several hybrid incompatibility genes (reviewed in Barbash, 2010). While the outcome of the two crosses seems to be governed by at least partially independent mechanisms involving different master genes, it is remarkable that in both cases chromatin plays a central role.

Despite the undoubted interest in understanding differences and commonalities between the two different crosses, the work of this thesis has focused on the cross involving *D.mel* females and *D.sim* males (Fig. 5).

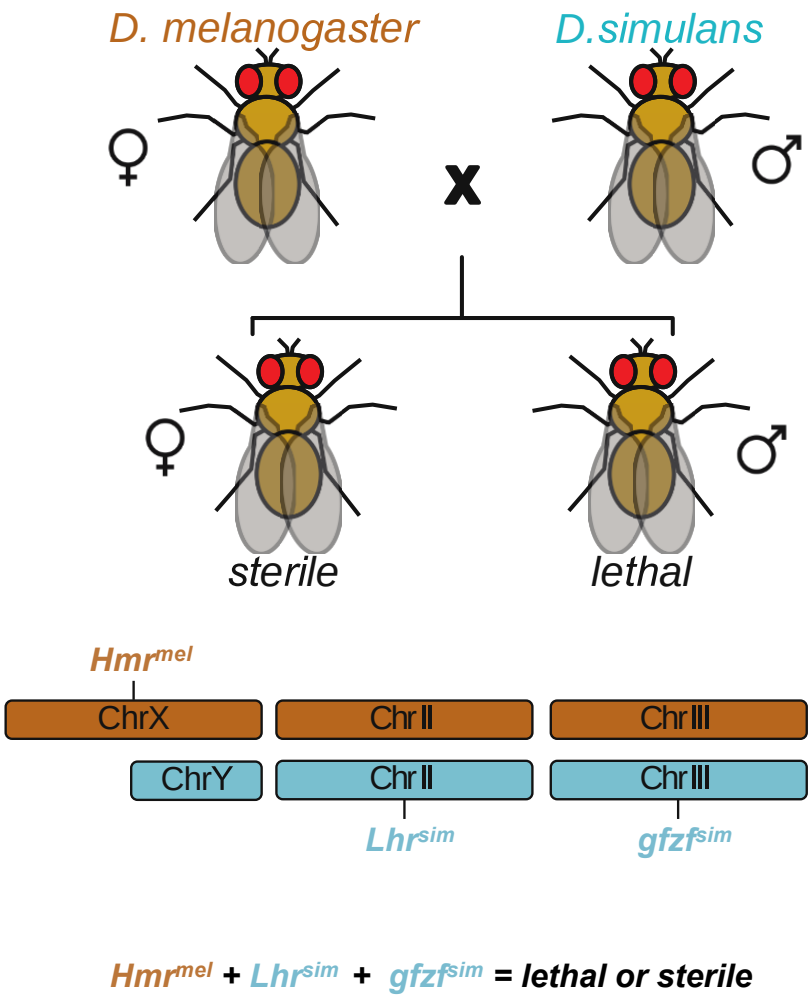


Fig. 5. Postzygotic isolation in *Drosophila*. The three hybrid incompatibility genes *Hmr*^{mel}, *Lhr*^{sim} and *gfzf*^{sim} genetically interact to cause hybrid male lethality and female sterility in crosses between *D. melanogaster* mothers and *D. simulans* fathers.

1.4.2. Three hybrid incompatibility genes genetically interact

In crosses between *D.mel* females and *D.sim* males, three genes have been identified to date whose loss of function or reduced expression was able to rescue hybrid male viability. In other words, the mutation of each of these genes suppresses lethality in F1 hybrid males. These are *hybrid male rescue* from *D.mel* (*Hmr*) (Hutter and Ashburner, 1987; Barbash et al., 2003), *lethal hybrid rescue* from *D.sim* (*Lhr*) (Watanabe, 1979; Brideau et al., 2006), and *GST-containing FLYWCH zinc-finger protein* from *D.sim* (*gfzf*) (Phadnis et al., 2015) (Fig. 5). These three genes engage in detrimental genetic interactions that cause hybrid incompatibility and all of them must be simultaneously present in hybrids to cause lethality. *Hmr-mel* is lethal to hybrids while *Hmr-*

sim is not and, conversely, both the *D.sim* alleles of *Lhr* and *gfzf* cause lethality while their *D.mel* orthologs do not (Fig 5).

Despite being one of the best characterized models of post-reproductive isolation, the molecular details underlying *D.mel/D.sim* hybridization have started to be uncovered only recently.

Strikingly but not quite surprisingly considering what discussed above, all three genes encode for chromatin binding proteins and all have undergone fast evolution (Barbash et al., 2004; Maheshwari et al., 2008, Phadnis et al., 2015), probably reflecting an adaptation to genomic alterations.

While the three genes genetically interact and their mutual presence is required for hybrid lethality, a biochemical interaction is well characterized only between HMR and LHR proteins (discussed in detail in the next sections; Brideau et al., 2006; Thomae et al., 2013; Satyaki et al., 2014). The molecular nature of the connection between either of these two proteins and GFZF remains unclear and constitutes an interesting open question.

While they are essential for the lethality of F1 interspecific hybrids, in *D. melanogaster*, *Hmr* and *Lhr* genes are not essential for viability (Watanabe, 1979; Hutter and Ashburner, 1987). The mutation of *gfzf*, instead, results in lethality in the early stages of larval development (Provost et al., 2006). Details of the intraspecific phenotypes resulting from *Hmr* and *Lhr* mutations will be discussed in the following paragraphs, while this paragraph will focus on gfzf.

The product of *gfzf* is a protein containing 4 FLYWCH zinc finger domains and two GST domains (Fig. 6) (Dai et al., 2004). Multiple genetic screenings have confirmed the essential role of *gfzf*, that has been found to be involved in cell-cycle regulation (Ambrus et al., 2009), DNA damage induced cell-cycle checkpoints (Kondo and Perrimon, 2011), Ras/MAPK signaling (Ashton-Beaucage et al., 2014), and Polycomb complex regulation (Gonzalez et al., 2014). All these results point to a rather pleiotropic function of GFZF that could be potentially explained by its broad genomic binding profile (Baumann et al., 2018). Using chromatin immunoprecipitation, Baumann and colleagues showed that GFZF binds along with the Motif 1 binding protein (M1BP) to the transcriptional start sites (TSS) of several genes relevant to the screenings discussed above, possibly as a transcriptional coactivator.

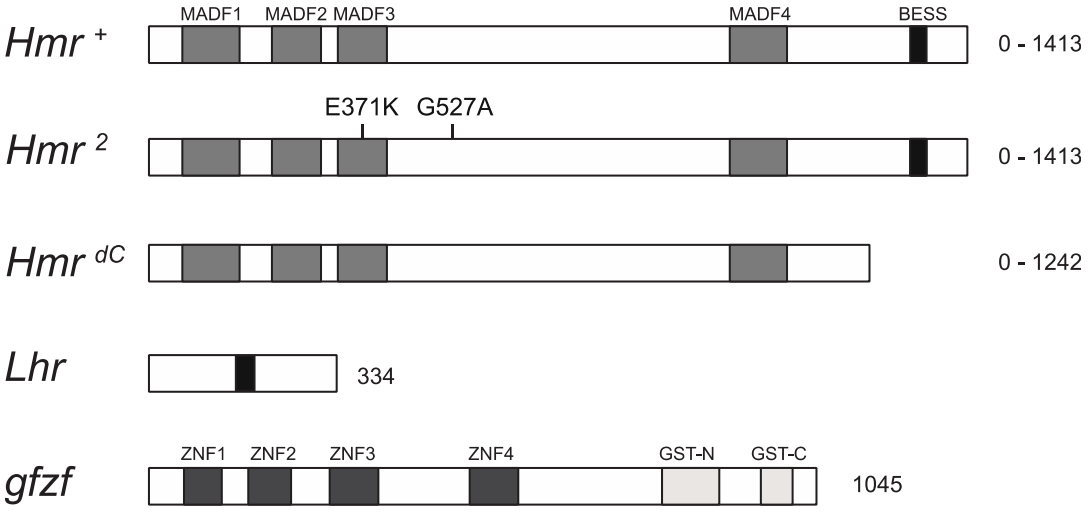


Fig. 6. Gene models of the three HI genes. *Hmr* is characterized by four MADF domains and one C-terminal BESS domain shared with *Lhr*. *gfzf* contains four 4 FLYWCH zinc finger domains and two GST domains. The two *Hmr* mutants used in this thesis encompass either two point mutations (Hmr^{2}) or a deletion of the C-terminus (Hmr^{dc}).

Similarly to other hybrid incompatibility genes, *gfzf* has evolved under positive selection. Consistently with its divergent evolution between the two species, signs of positive selection are found in the *D.sim* but not in the *D.mel* lineage and in particular in the FLYWCH zinc finger domains (Phadnis et al., 2015).

1.4.3. *Hmr* and *Lhr* encode for two mutually interacting chromatin proteins

Hmr encodes for a large protein carrying four domains of the Myb/SANT-like domain in ADF1 (MADF) family and one BEAF, Su(var)3-7 and Stonewall like (BESS) domain in its C-terminus (Fig. 6) (Brideau et al., 2006; Maheshwari et al., 2008).
Lhr was previously identified as a member of constitutive heterochromatin and HP1a interactor and named *HP3* (Greil et al., 2008). The product of the *Lhr* gene encodes for a small protein that also contains a BESS domain (Fig. 6) (Brideau et al., 2006).
HMR and LHR physically interact in a protein complex involving the heterochromatin hallmark HP1a and their interaction has been speculated to occur directly through their shared BESS domain (Brideau et al., 2006; Thomae et al., 2013; Satyaki et al., 2014).
Like *gfzf* and other hybrid incompatibility genes, *Hmr* and *Lhr* have all evolved rapidly under recurrent positive selection.
However, unlike *gfzf* that is conserved up until mammals, *Hmr* and *Lhr* are *Drosophila*-specific. The same is true for the family of MADF-BESS containing proteins, consisting of over 16 members in *D.mel* (Shukla et al., 2014). This family has evolved and expanded rapidly during *Drosophila*

evolution, presumably through gene duplications and contains several genes involved in hybrid incompatibility (Shukla et al., 2014).
Hmr divergence in *Drosophila* is particularly remarkable in its MADF domains. MADF domains were first identified in ADF1 as highly positively charged domains mediating DNA binding (Cutler et al., 1998; England et al., 1992). Notably, three of the four *Hmr* MADF domains have diverged significantly in their ionic properties among *Drosophila* species. This is particularly true for *Hmr* MADF3 that in several species, including *D.mel*, has acquired a negative charge, suggesting that *Hmr* MADFs may combine both DNA and chromatin binding properties (Maheshwari et al., 2008). *Lhr* has also extensively diverged and, interestingly, the *D.sim* ortholog like in its sister species *D. mauritiana* and *D. sechellia*, contains a 16–amino acid insertion, interrupting a potential leucine zipper domain, differently from *D.mel.* and outgroup species (Brideau et al., 2006).

Similarly to the other HI gene *gfzf*, *Hmr* and *Lhr* genes have been isolated from fly stocks that could produce viable male hybrids in *D.mel / D.sim* crosses (Fig. 5) (Phadnis et al., 2015; Barbash et al., 2003; Brideau et al., 2006). This has led to the identification of several naturally occurring or induced *Hmr* and *Lhr* mutant alleles. For *Hmr*, these include the hypomorphic mutations Hmr^{1} and Hmr^{3}, that are characterized by strongly reduced *Hmr* expression (Hutter and Ashburner, 1987; Barbash et al., 2003), and *Df(1)Hmr-* that results from a deletion encompassing part of the promoter region and the first two exons of the *Hmr* gene (Barbash and Lorigan, 2007). Another *Hmr* allele, Hmr^{2}, is of particular interest, in that it carries two point mutations, the first of which, E371K, occurs in the unusually negatively charged third MADF domain (Fig. 6) (Hutter et al., 1990; Barbash et al., 2003; Aruna et al., 2009). This glutamic acid is conserved in all 14 known *Hmr* orthologs and may mediate important protein-protein interactions necessary for HMR localization at chromocenters as suggested by its nuclear mis-localization in these mutants (Maheshwari et al., 2008; Thomae et al., 2013). In addition to its effects on chromatin binding and hybrid lethality, Hmr^{2} also affects female fertility (Aruna et al., 2009), suggesting that HMR proper chromatin localization is important in both pure-species and hybrids. Having a point mutation in a domain putatively important for chromatin interactions and localization, makes Hmr^{2} a particularly promising mutant for a fine dissection of the role of HMR molecular interactions in both intra- and inter-specific phenotypes.
For *Lhr* instead the *D. simulans* Lhr^{1} allele is a hypomorph while the Lhr^{2} consists of a precise deletion of the above mentioned *D.sim*-specific 16-aa insertion, suggesting that this sequence is responsible for hybrid lethality (Brideau et al., 2006).

1.4.4. *Hmr* and *Lhr* mutants reveal key functions in oogenesis and female fertility, silencing of repetitive elements at heterochromatic sites and telomere homeostasis

Hmr and *Lhr* mutants have shown, beside their effect in interspecific hybrids presumably caused by a gain-of-function, that these genes are important for a number of intraspecific processes, despite not being essential for viability (Watanabe, 1979; Hutter and Ashburner, 1987).
Flies mutants lacking *Hmr* or *Lhr* display a severe loss of female fertility and a mildly reduced longevity (Aruna et al., 2009; Satyaki et al., 2014), a dramatic derepression of several families of transposable elements and satellite DNA (Thomae et al., 2013; Satyaki et al., 2014) and defects in sister chromatid detachment during anaphase (Blum et al., 2017). Similar phenotypes have been observed reducing HMR and LHR levels in cultured *Drosophila* SL2 cells by RNAi treatment (Thomae et al., 2013).
Many of the repetitive elements that are upregulated upon *Hmr* and *Lhr* depletion origin from heterochromatic regions at or around centromeres and telomeres where the two proteins have been suggested to play important structural roles (Thomae et al., 2013; Satyaki et al., 2014; Anselm et al., 2019; Kochanova et al., 2020). Among the most mis-expressed elements upon depletion of *Hmr* or *Lhr* are the telomeric repeats *Het-A*, *TART* and *TAHRE*. Consistently with the essential role of these repetitive elements in telomere cap maintenance in *Drosophila*, their overexpression is accompanied by an increase in telomere length in *Hmr/Lhr* mutants (Satyaki et al., 2014; Andreyeva et al., 2005; Raffa et al., 2011). Another element that is significantly upregulated in *Hmr* mutants is the retrotransposon *gypsy*, whose 5' contains an insulator sequence that has been shown to be bound by HMR along with a complex of insulator proteins (Gerland et al., 2017; Pai et al., 2004; Gerasimova et al., 1995; Georgiev and Gerasimova, 1989; Parkhurst et al., 1988).
D. melanogaster females mutants for *Hmr* and *Lhr*, display an overall loss of fertility (Aruna et al., 2009; Satyaki et al., 2014). A detailed study of the effects of *Hmr* mutation in ovaries (Aruna et al., 2009) has shown that this results in the production of less and occasionally mis-shapen eggs and in a lower survival of the progeny with reduced egg hatchability and adult eclosion. In addition, *Hmr-* ovaries are smaller than the wildtype and with a reduced number of ovarioles. *Hmr* mutant ovarioles can be grouped in a wildtype- and a stub-like class, the first being indistinguishable from the wildtype and the second with dramatic defects in both shape and structure (Aruna et al., 2009). The important function of *Hmr* in ovaries and the fact that a class of ovarioles are morphologically intact make them an ideal system for molecular studies of *Hmr* mutants in flies.

1.4.5. HMR and LHR nuclear localization has been highly debated

HMR and LHR proteins are interacting in a complex and colocalize in the nucleus (Thomae et al., 2013; Satyaki et al., 2014). Notably, different studies have shown slight but substantial differences

in their nuclear localization in different tissues or cell types, leading to a heated debate on whether they are *bona-fide* centromeric proteins or rather sit at heterochromatic loci along with HP1a.
Earlier studies using ectopically expressed tagged proteins have shown LHR localizing at HP1a domains in cultured Kc cells (Greil et al., 2007) and in salivary glands polytene chromosomes (Brideau et al., 2006). A genomic mapping of LHR by DamID has described LHR as part of classic "green" pericentromeric heterochromatin (Filion, van Bemmel, Braunschweig et al., 2010). Similarly, a later study in embryos has shown both LHR and HMR co-localizing with HP1a (Satyaki et al., 2014).
The subsequent development of monoclonal antibodies directed against the endogenous proteins allowed a more physiological mapping of HMR and LHR. In cultured SL2 cells nuclei and mitotic cells of the larvae wing imaginal discs, HMR and LHR signal accumulates in sharp foci overlapping with centromeric markers (Thomae et al., 2013; Kochanova et al., 2019). The use of super-resolution instead of confocal microscopy, however, revealed the details of an interdigitated structure, where HMR is not strictly overlapping with CENP-A but rather bordering it (Anselm et al., 2018; Kochanova et al., 2020).
A fine mapping of HMR by chromatin immunoprecipitation brought complementary results, showing that HMR is often found at the border of HP1a domains, supporting a model where HMR is not strictly centromeric nor strictly pericentromeric but rather sits in between these two domains (Gerland et al., 2017). Furthermore, this study also revealed a bimodal binding profile of HMR with two classes of binding sites. One class of HMR binding sites is located at the border of HP1a domains that are often found at promoters or TSS. Notably, while this class is enriched in the typical constitutive heterochromatin-like pericentromeric regions and the heterochromatic chromosome 4, it is also distributed along chromosome arms. A second class of HMR binding sites is located at *gypsy*-like insulator elements that are typically bound by the insulator complex comprising the proteins SU(HW), CP-190, MOD(MDG4) and BEAF32 and are important elements for genome organization. Interestingly both classes of binding sites are candidates for constituting TAD boundaries and a recent correlation analysis has indeed revealed that HMR is enriched at TAD boundaries (Kochanova et al., 2019).
The apparent discrepancies in nuclear localization discussed above may be explained by different experimental conditions (i.e. different expression level of the endogenous proteins and the ectopically expressed ones, different antibodies, etc) or may reflect cell-type- or cell-cycle-specific differences in chromatin organization within the nucleus. In the light of the novel findings, taking a closer look at the HMR stainings in embryos performed by Satyaki and colleagues (Satyaki et al., 2014), HMR colocalization with HP1a seems to be at best only partial, with a pattern that rather resembles the interdigitating structure described later by super-resolution microscopy (Anselm et al., 2018; Kochanova et al., 2020).
In further support to a model where HMR is sitting at chromocenters in between centromeres and pericentromeric heterochromatin, Blum and colleagues have used larval brain cells to perform a detailed analysis of HMR and LHR localization with respect to CENP-A, arriving to the conclusion that HMR and LHR do not localize at centromeres but rather at chromocenter (Blum et al., 2017).

The chromocenter model would also explain the differences observed between the genomic profiles and the nuclear localization, obtained by ChIP-Seq and by IF, respectively. The ChIP-Seq profiles highlighted an array of binding sites that are not visible in immunofluorescent experiments and this may be due to coalescence of these domains at the chromocenter in the 3D space of the nucleus. This hypothesis is supported by the fact that euchromatic HP1a loci cluster together with pericentromeric heterochromatic HP1a sites at chromocenters in interphase nuclei (Lee et al., 2020).
In addition to what described above, HMR and LHR localization has been studied in polytene chromosomes from *D.mel* salivary glands, where they colocalize together with HP1a at chromocenters and at telomeres (Brideau et al., 2006; Thomae et al., 2013).
All together most observations are explained with a model that sees HMR and LHR residing at the chromocenter between centromeric and pericentromeric heterochromatin.
However, two main issues emerge:

1. All studies were performed in different tissues, cell types and experimental conditions, making an accurate comparison difficult. This is calling for a unifying model. Ovaries could offer the right framework to reconcile all these sparse evidence in a unique experimental system.
2. All experiments discussed were performed in *D.mel* with the only exceptions of studies that looked at the localization of ectopically expressed *D.sim* orthologs in a *D.mel* background (Thomae et al., 2013 and Satyaki et al., 2014). A localization analysis in *D.sim* and hybrids has been hindered by the paucity of genetic tools and by the inherent difficulties in performing molecular studies in hybrids.
3. While the localization of HMR and LHR relative to each other has been investigated, nothing is known about their localization relative to the third HI gene, *gfzf*.

1.4.6. HMR biochemically interacts with LHR and is critical for keeping chromocenters together

In *Drosophila melanogaster* HMR and LHR proteins physically interact in immunoprecipitation and yeast-two-hybrid assays (Thomae et al., 2013; Satyaki et al., 2014) and this interaction has been speculated to be mediated by their shared BESS motif, a protein-protein interaction domain (Brideau et al., 2006; Maheshwari et al., 2008). The two proteins are also interdependent, with the depletion of either of them resulting in the destabilization of the other (Thomae et al., 2013; Satyaki et al., 2014).
In addition to their mutual interaction, HMR and LHR have been shown to interact with several other chromatin proteins, including classic heterochromatin proteins like HP1a, the histone chaperones NLP and NPH and some insulator proteins (Thomae et al., 2013; Anselm et al., 2018). NLP and NPH are oligomerizing to form a pentamer that has been shown to be critical for clustering centromeres at the periphery of nucleoli (Padeken et al., 2014; Anselm et al., 2018).

HMR, in turn, is necessary to recruit NLP/NPH to centromeres and has thereby been proposed to bring together pericentromeric heterochromatin and centromeres at the chromocenters (Anselm et al., 2018). Additionally, HMR has been shown to interact with the centromeric protein CENP-C that seems to be responsible for HMR recruitment to centromeres and the formation of HMR foci (Thomae et al., 2013; Kochanova et al., 2020). Upon CENP-C knockdown centromeres decluster and HMR diffuses to heterochromatin (Kochanova et al., 2020).
Interestingly, HMR binding, particularly upon its overexpression, seems to have an inverse correlation with the binding of a subunit of the condensin complex (Kochanova et al., 2020), which is critical for chromatin condensation in higher order 3D structures (Rosin et al., 2018).
All this evidence together suggests that HMR is involved in a complex network of interactions and plays an important role in bringing together centromeric and pericentromeric chromatin to help the formation of the intricate structure of the interphase chromocenter.
Further supporting the delicate balance that is necessary to keep chromocenter bundled and functional, a recent work (Jagannathan and Yamashita, 2021) has shown that HMR and LHR, the very same factors that are important for keeping chromocenter together in pure species, interfere with its formation and are leading to its disruption in hybrids. This apparently paradoxical combination of phenotypes is attributable to a gain-of-function of the hybrid incompatibility factors in the hybrid background and is further highlighting the vital importance of HMR and LHR in the context of pericentromeric and centromeric chromatin. Remarkably, the authors propose that chromocenter disruption and the consequent formation of micronuclei resulting in cell death, could be the main reason for hybrid lethality (Jagannathan and Yamashita 2021).

1.4.7. HMR and LHR diverged in their intraspecific dosage and are overdosed in hybrids

While most molecular characterization of *Hmr* and *Lhr* has been done in *D. melanogaster*, little is known about their molecular function in *D. simulans* and hybrids.
Experiments using *D. simulans Hmr* and *Lhr* transgenic alleles in *D. melanogaster* cells have shown that their ability to interact as well as their localization to centromeric foci and chromocenters are conserved (Thomae et al., 2013; Satyaki et al., 2014). Similarly, the fertility defects observed in *Hmr* mutants are partially rescued by *Hmr* orthologs from the sibling species *D. simulans* and *D. mauritiana* (Aruna et al., 2009). On the other hand, this rescue is only partial as old females are not rescued by orthologous alleles and *D.sim Hmr* alleles also seem to have diverged in the kind of TEs they are able to repress (Aruna et al., 2009; Satyaki et al., 2014).
When Thomae et al. set out to identify species-specific differences between *Hmr* and *Lhr* orthologs the only significant and yet striking difference they could find was not qualitative but quantitative (Thomae et al., 2013). They showed that HMR and LHR proteins are expressed at different and reciprocal levels in the two species: much more HMR than LHR in *D.mel* and virtually

no HMR along with highly expressed LHR in *D.sim*. In hybrids, instead, both orthologs are strongly expressed and their overdosage has been proposed to be responsible for their gain of function that leads to hybrid lethality (Thomae et al., 2013). A localization analysis using polytene chromosomes has shown that HMR and LHR largely mislocalize from specific heterochromatic foci to numerous presumably euchromatic locations along chromosome arms (Thomae et al., 2013). The observed acquisition of new binding sites along the genome, suggests that HMR and LHR, when present in non-limiting amounts, can either localize to new genomic regions or acquire new binding partners for whom they have lower affinity, or both.

In general, the fact that *Hmr* and *Lhr* mutations result mostly in pleiotropic effects, spanning from the fertility and mitotic defects to the derepression of repetitive elements, indicate that such phenotypes could result from a number of underlying molecular mechanisms. This is consistent with a model that sees HMR and LHR as architectural proteins that are critical in organizing the genome in general and the heterochromatic regions around centromeres and telomeres in particular. As other architectural heterochromatic proteins (Tartof et al. 1989; Allshire & Madhani, 2017; Schwartz & Cavalli, 2018), their dosage is key in order to maintain a healthy balance between heterochromatin and euchromatin and to organize the genome.

Given the key role of HMR/LHR expression in pure species and hybrids, it is intriguing to ask how is the HMR/LHR core complex composed in pure species and how is it affected by the overexpression of both proteins in hybrids. In this perspective a study of HMR/LHR interactions and genomic localization in native conditions and upon overexpression would be of particular interest.

1.4.8. *Drosophila* oogenesis is an ideal model to reconcile previous controversial observations on HMR localization

As mentioned above, different studies using different *Drosophila* cell types, tissues or developmental stages have yielded a variety of different localizations for HMR. In particular, the discussion over the centromeric or pericentromeric localization of HMR has been a matter of debate over the last decade (Thomae et al., 2013; Satyaki et al., 2014; Wei et al., 2014; Blum et al. 2017; Gerland et al. 2017; Kochanova et al. 2020; Anselm et al., 2018), even with claims that the centromeric localization was artifactual (Blum et al., 2017). However, all these studies have been performed in different conditions that are hard to compare, making it hard to drive any meaningful conclusion in one direction or the other.

Drosophila oogenesis offers an ideal system to study chromatin and nuclear dynamics as different cell types are present over different cell cycle phases and differentiation states.

Ovaries are also the main adult tissues where the *Hmr* expression is strong and significant (Brown et al., 2014; Leader et al., 2018) suggesting a critical function for *Hmr*. This is also witnessed by the fact that ovaries are also the tissues that are most dramatically affected by *Hmr* mutation, which causes a massive upregulation of repetitive elements and a number of functional and

morphological defects resulting in a major drop in female fertility (Aruna et al., 2009; Satyaki et al., 2014).

Oogenesis is the process that leads from the pluripotent germline stem cells to the differentiated gametes. The major chromatin rearrangements that happen during this process, make ovaries an ideal model system to study chromatin and nuclear dynamics.

Oogenesis in *Drosophila* (reviewed in Kirilly & Xie, 2007; Bastock & Johnston, 2008; Hudson & Cooley, 2014) adult females takes place in one pair of ovaries, each of which is constituted by ~16 parallel tubes called ovarioles containing the developing egg chambers (Fig. 7).

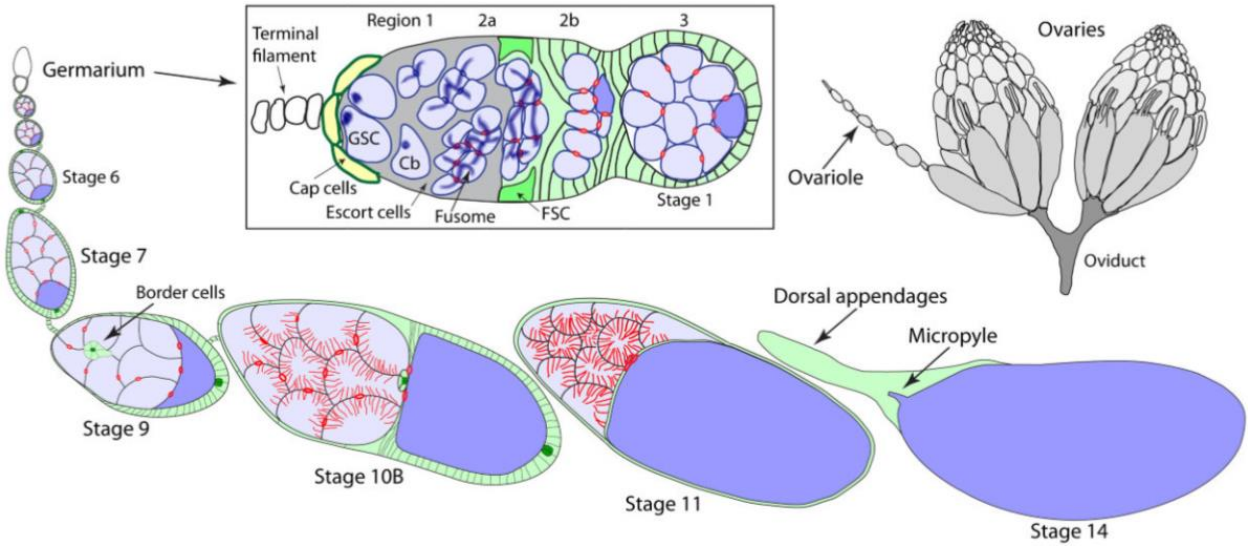


Fig. 7. Ovaries structure and oogenesis in *Drosophila*. An adult female fly has one pair of ovaries, each of which is in turn subdivided in about 16 ovarioles that contains the egg chamber. Egg development proceeds in an anterior-posterior direction, with the most proximal part containing the germarium with germline stem cells, from which progressively more mature egg chambers are succeeding (Figure adapted from Hudson & Cooley, 2014).

Development starts in the anterior tip, called germarium, and runs in an anterior to posterior direction (Fig. 7). The germarium carries germline stem cells (GSC) that divide asymmetrically to produce another stem cell and a germline cyst cell (GCC). The GCC then undergoes four mitotic cycles with incomplete cytokinesis resulting in a cyst of 16 interconnected cells, one of which will be along the way specified as the oocyte and initiate meiosis. The other 15 become nurse cells that will feed the oocyte with maternally contributed RNAs and proteins. Nurse cells switch their mitotic program to endoreplication and result in a highly polyploid genome. In parallel, in region 2, somatic follicle stem cells (FSC) self-renew and produce follicle cells (FC) that will encapsulate the 16 cells cyst in region 3. The encapsulated egg chamber will then stem out of the germarium and go through 14 additional developmental stages that will bring to the production of a mature oocyte. Similarly to nurse cells, follicle cells also undergo a major cell-cycle reprogramming: at

the end of stage six they stop having mitotic divisions and have three endoreplication cycles instead, also resulting in polyploid genomes (stages 7-10a).

Progression of these different cell types through their maturation requires extensive genomic reorganizations and different cell cycle stages are observable in parallel within the same tissues. This offers the unique opportunity to detect HMR localization in different conditions within the same experiment. Part of the work presented in Publication 1, focuses on the analysis of HMR localization in *Drosophila* oogenesis, in different cell types characterized by different cell-cycle states, bringing important new insights that allow to unify previous observations in a unique coherent model.

2. Aims of the thesis

As extensively discussed in the introduction, pericentromeric heterochromatin is a hub for the evolution of hybrid incompatibilities that act as postzygotic barriers in the formation of species.
The formation of the two fly sibling species *Drosophila melanogaster* and *Drosophila simulans* offers an ideal system to study the role of heterochromatin in reproductive isolation, as in this context three chromatin proteins have been shown to play a master role in the formation of postzygotic barriers.
The three hybrid incompatibility genes *Hmr*, *Lhr* and *gfzf* diverged during *D.mel* and *D.sim* evolution and genetically interact to cause lethality in male and infertility in female interspecific hybrids (Fig. 5). The molecular mechanisms underlying these incompatibilities are still largely unknown and many open questions still remain on both their divergent functions in pure species and their novel detrimental functions in hybrids.

1. HMR/LHR quantitative balance has been shown to be critical for their function but, so far, all analyses of their protein-protein interaction network were done using an ectopic expression system prone to overexpression. How do HMR and LHR interact in native conditions?
2. HMR has been proposed to mediate chromocenter bundling through a complex network of protein-protein interactions. How are HMR and LHR interacting with each other and with their other interaction partners? How are different interactions mediating different functions?
3. The nuclear localization of HMR and LHR has remained controversial, partially because experiments were performed in different tissues, cell types and experimental conditions, making an accurate comparison difficult. How are HMR and LHR localizing throughout the cell cycle and in different cell types?
4. The inherent instability of hybrids makes them extremely difficult to study and has hindered a detailed molecular analysis of this system. How are HMR and LHR molecular functions, including interaction networks and localization, changed in hybrids background?
5. The most recently discovered player, GFZF, is still poorly characterized at the molecular level in *D.mel*, *D.sim* and hybrids alike. How does it localize in these systems?
6. Despite their genetic interaction, no molecular interaction has been shown involving HMR or LHR with GFZF. How do they interact to cause lethality in hybrids?

In the two publications presented in this thesis we addressed these questions.

In Publication 1 we addressed questions 1-4.
1. We took advantage of antibodies targeting native epitopes in HMR and LHR, and used cultured *Drosophila* SL2 cells to perform AP-MS experiments in order to measure the interactome of both

HMR and LHR. We used this method to identify for the first time an HMR/LHR native protein complex. Additionally, we performed AP-MS experiments for all the complex components, therefore both validating our findings and identifying a larger protein-protein interaction network also involving subunit-specific interactors.
2. We used genetic tools to dissect HMR. We created an *Hmr* C-terminal mutant encompassing its presumably LHR-binding BESS domain, asking whether this would disrupt the HMR/LHR interaction and potentially other interactions. In addition, we used the naturally occurring mutation Hmr^{2} that carries a point mutation in the very conserved third MADF domain of *Hmr*, that is unusually negatively charged and has been speculated to be important for mediating protein-protein interactions. Using both these mutants we could ask: How each mutation affects HMR interaction network and complex composition? How is this affecting HMR localization in both cells and fly ovaries? How is HMR function affected by each of these mutations in pure species? How is hybrid lethality affected?
3. We chose to use ovaries as a model system to study HMR localization and try to reconcile previous contrasting observations in a unique model. In fact, *Drosophila* oogenesis in ovaries offers an ideal system to study HMR nuclear dynamics in different cell types over different cell cycle phases and differentiation states. Moreover, ovaries are the main adult tissues where the *Hmr* expression is strong and the most dramatically affected by *Hmr* mutation. We asked how HMR localizes along with CENP-A and HP1a in tissues that go from actively mitotically cycling to arrested in interphase to polyploid.
4. We used a cell line where we could induce HMR/LHR overexpression to mimic the observed condition in hybrids. The main advantage of this hybrid-like system being that it is easy to handle as well as up-scalable to a level that allows performing high-throughput experiments like ChIP-Seq and AP-MS, to obtain a detailed profile of HMR genomic localization and interaction network, respectively.

In Publication 2, we addressed questions 5 and 6.
5. We used *in situ* hybridization in polytene chromosomes to analyze for the first time GFZF localization in *D.mel*, *D.sim* and hybrids. In the same system we also looked at the localization respective to HMR.
6. In addition we used the hybrid-like HMR/LHR overexpressing cell lines to analyze GFZF localization with respect to HMR in native (not overexpressing) and hybrid (overexpressing) conditions.

3. Publication 1

Contributions to "The integrity of the HMR complex is necessary for centromeric binding and reproductive isolation in *Drosophila*"

The study was conceived by Axel Imhof, Andreas Thomae and myself. I performed all experiments with the following collaborations or contributions: the analysis of repetitive elements in ovaries was performed in collaboration with Anuroop Venkatasubramani, the analysis of localization in SL2 cells was performed in collaboration with Natalia Kochanova and Andreas Thomae, the analysis of localization in ovaries was performed in collaboration with Andreas Thomae and the staining protocol was adjusted based on suggestions by Kenneth Boerner, the analysis of genomic localization by ChIP-Seq was performed in collaboration with Peter Krueger, Tamas Schauer and Angelika Zabel, the analysis of the interaction proteomes was performed in collaboration with Peter Krueger, Ignasi Forne, Andreas Schmidt and Marc Borath. Writing of the original manuscript was done by Axel Imhof and myself. All co-authors participated to some extent in the editing and reviewing of the manuscript.

Andrea Lukacs, Andreas W. Thomae, Peter Krueger, Tamas Schauer, Anuroop V. Venkatasubramani, Natalia Y. Kochanova, Wasim Aftab, Rupam Choudhury, Ignasi Forne, & Axel Imhof (2021). *The integrity of the HMR complex is necessary for centromeric binding and reproductive isolation in Drosophila*. PLoS Genetics, 17(8), 1–27. https://doi.org/10.1371/journal.pgen.1009744

4. Publication 2

Contributions to "Altered localization of hybrid incompatibility proteins in *Drosophila*"

The part of the study performed in flies was conceived by Nitin Phadnis and Jacob Cooper, the part performed in cells was conceived by Axel Imhof and myself with Nitin Phadnis and Jacob Cooper. In particular, the idea of using HMR-overexpressing cell lines to mimic HMR localization in hybrids and therefore studying the altered genomic localization of HMR *in vitro*, in a system suitable to be up-scaled for high throughput experiments (ChIP-Seq in Publication 2 and AP-MS in Publication 1), was conceived by Axel Imhof and myself. The experimental design was conceived by Axel Imhof and myself and the bioinformatic analysis and interpretation was performed by Tamas Schauer and myself. Fig. 8, Fig. S8, Fig. S9, Fig. S10 and Fig. S11 were conceived and prepared by Tamas Schauer and myself. I contributed partially to the writing and extensively to the editing and reviewing of the original manuscript.

Jacob C. Cooper, **Andrea Lukacs.**, Shelley Reich, Tamas Schauer, Axel Imhof, & Nitin Phadnis (2019). *Altered localization of hybrid incompatibility proteins in Drosophila.* Molecular Biology and Evolution, 1–10. https://doi.org/10.1093/molbev/msz105

5. Discussion

Understanding the molecular mechanisms underlying the evolution of species requires, on the one hand, the study of the divergent evolution of species-specific adaptations in pure species and, on the other hand, of how these adaptations interact in hybrids to cause reproductive isolation. In the publications presented in this thesis we addressed both questions, providing novel insights that may, in turn, open the field to new questions and constitute the base for new promising research paths.

HMR and LHR reside in a native six subunit core complex in *D. melanogaster* (Pub. 1)

To address the question of how the two hybrid incompatibility proteins HMR and LHR interact in a protein complex in physiological conditions, we performed AP-MS experiments using antibodies targeting native epitopes in *Drosophila* SL2 cells. The resulting six subunit complex includes, besides HMR and LHR, the nucleolar proteins NLP and NPH, and the two proteins of unknown functions that we named Buddy Of HMR 1 (BOH1, previously CG33213) and Buddy Of HMR 2 (BOH2, previously CG4788). Besides the complex identification, with the first publication, we provide as a resource the interactome for all the complex components. Although we didn't include HP1a in the HMR/LHR core complex because it is consistently underrepresented with respect to the other complex subunits, the heterochromatin hallmark protein is indeed interacting with all complex components individually and is still to be considered an additional *bona fide* interactor.

HMR contains two functionally important protein-protein interaction modules (Pub. 1)

HMR has been proposed to be an important actor in centromeric and pericentromeric chromatin architecture by mediating their clustering at chromocenters, at the nucleolar periphery.

After having identified the HMR/LHR core complex components, we could speculate that HMR interaction with NLP and NPH may mediate its binding to centromeric clusters while the interaction with LHR and the accessory component HP1a could mediate the interaction with pericentromeric heterochromatin. We could test this hypothesis by using two mutants, one deleting the BESS-domain-containing *Hmr* C-terminus (Hmr^{dc}) and the other containing a point mutation in a highly conserved MADF domain in the *Hmr* N-terminus (Hmr^{2}) (Fig. 6). This genetic approach allowed us to selectively disrupt HMR interactions by depleting either its interaction with LHR and HP1a (Hmr^{dc}) or with the rest of the complex (Hmr^{2}).

While both mutations result in the disruption of the typical HMR centromeric-proximal clusters in SL2 cells interphase, our studies in adult female ovaries support the idea that the deletion of *Hmr* C-terminus, and therefore the loss of its interaction with LHR and HP1a, selectively disrupts HMR ability to bind heterochromatin but not centromeres.

These results support a model where HMR acts as a bridge between centromeric and pericentromeric heterochromatin, through its interaction modules with NLP/NPH and HP1a/LHR, respectively (Fig. 9A). In this context it will be interesting to study in greater detail the role of the

two newly characterized proteins BOH1 and BOH2 in stabilizing the complex and interacting with the two chromatin types.
In addition to affecting HMR protein-protein interactions and nuclear localization, the two mutants also disrupted both HMR physiological functions in pure species - including silencing of repetitive elements and female fertility - and non-physiological toxic functions that lead to hybrid lethality.
All together our results suggest that the integrity of the HMR core complex, with both its interaction modules, is necessary for its ability to bind centromeres and pericentromeric heterochromatin and mediate their clustering at chromocenters. Interfering with the HMR complex formation and therefore with its structural function could, in turn, result in downstream effects in both pure species and hybrids, with consequent loss of retrotransposon silencing and fertility, and suppression of hybrid lethality, respectively.

A unifying view of HMR localization at the boundaries of pericentromeric and centromeric chromatin (Pub. 1)

To redeem the long-lasting discussion on HMR localization - partially due to the lack of an organic experimental setup allowing an accurate comparison of the different observations - we used *Drosophila melanogaster* oogenesis as a model, and co-stained HMR with CENP-A and HP1a as centromeric and pericentromeric markers, respectively, in tissues that go from actively mitotically cycling to arrested in interphase to polyploid.
Our findings support a model that reconciles previous controversies and suggests that HMR is neither centromeric nor pericentromeric but rather localizes at the boundaries of these two types of chromatin, where it may act as a bridge by forming a complex that interacts with both.

Excess of HMR and LHR interact with novel chromatin factors including GFZF (Pub. 1)

Having identified the HMR/LHR core complex, also gave us the opportunity to ask how this is affected when HMR/LHR balance is altered by their overexpression, as it is observed in the interspecific hybrid background. Using an inducible system, we overexpressed HMR/LHR to mimic their status in hybrids. This system offers the great advantage that it is easy to manipulate and to up-scale enough to perform high-throughput experiments.
When we looked at the HMR interactome upon overexpression, we identified a number of new chromatin interactors, from architectural proteins such as insulators to zinc-finger DNA binding proteins. Notably, among the novel HMR-interactors we observed the hybrid incompatibility protein GFZF, therefore providing a further molecular base for the genetic interaction between these two HI factors. The described HMR supplemental interactors may be binding with low affinity, therefore being masked when HMR availability is limited and restricted to the core complex, and pop up when HMR is present in excess, such as in hybrids.

GFZF localization differs in *D.mel* and *D.sim* (Pub. 2)

To shed some light on the molecular function of the third and most recently discovered HI gene, *gfzf*, we used *in situ* hybridization in polytene chromosomes to perform for the first time a comparative study of GFZF protein localization in *D.mel*, *D.sim* and hybrids. The most obvious

observed difference is at telomeres, where GFZF displays a divergent binding behavior in the two species. This species-specific binding pattern does not seem to be driven by differences in the GFZF protein sequence but rather by the dramatic divergence in copy number of the telomeric retrotransposon arrays in the two species.

HMR co-localizes to GFZF loci in hybrids and upon overexpression (Pub. 2)

Despite their genetic interaction in hybrids, the molecular connection between GFZF and the other two HI proteins remains mysterious.

Our co-staining of GFZF and HMR on polytene chromosomes show that these two proteins bind in close proximity at telomeres but without overlapping. Similarly, our comparisons of genome-wide binding profiles obtained by ChIP-Seq of GFZF and HMR in *D.mel* cell lines show a sometimes proximal but rarely overlapping pattern.

The observation that HMR protein mapping on polytene chromosomes results in a different localization from other tissues, mostly telomeric and only weakly at chromocenter, is not new to this study and is probably the result of the peculiar nature of this tissue, where the chromocenter and pericentric heterochromatin are under-replicated with respect to chromosome arms and telomeres. While it is important to keep this in mind, it does not prevent polytene chromosomes from being a valuable model system, in particular to perform comparative studies.

In contrast to what we observed in pure species, in interspecific hybrids, we found the first evidence of HMR and GFZF co-localization. On polytene chromosomes, HMR invades GFZF loci both at chromosome ends and at several newly acquired hybrid-specific euchromatic binding loci at chromosome arms. These findings are consistent with our genome-wide analysis of HMR upon overexpression of HMR and LHR, a system that we used to mimic the hybrid condition, where we show a stark increase in HMR and GFZF co-localization with respect to the native HMR profile.

The fact that the newly acquired co-localization of HMR and GFZF can be reproduced in cell lines where only the *D.mel* orthologs of HMR and GFZF are present, suggests that this is not driven by a divergence in the sequence of the two proteins in *D.mel* and *D.sim*, but rather by a natural affinity of the surplus of HMR for GFZF. This idea is further supported by our HMR interactome in HMR/LHR overexpressing cells described above, where we showed that HMR-GFZF interaction is specifically gained upon overexpression. Alternatively, or additionally, the extension of HMR into the GFZF regions in hybrids or in overexpressing cells could reflect the loss of boundaries between the two regions.

Is the qualitative and quantitative composition of the HMR complex affecting chromosome topology?

Our results reflect the importance of both the composition and the balance of the HMR/LHR complex in tethering centromeric and pericentromeric chromatin at chromocenters in normal physiological conditions.

In our model we suggest that HMR presents two interaction modules that allow it to act as a bridge at chromocenters (Fig. 9A). The disruption of either of these modules results in an impairment of HMR function and of proper chromocenter formation. In hybrids and upon overexpression, HMR

spreads along chromosome arms acquiring new binding sites. We speculate that the new euchromatic HMR-binding sites may be tethered at chromocenters along with pericentromeric heterochromatin resulting in a dysfunctional three-dimensional configuration that could in turn cause hybrid lethality (Fig. 9B).
It would be interesting, in the future, to test these hypotheses by performing topological studies with chromosome conformation capture techniques (HiC), therefore allowing to test how intra and inter-chromosomal contacts are affected by HMR complex disruption or by HMR/LHR overexpression.

Challenges for tackling the molecular secrets of hybrid incompatibilities

Studying hybrids presents some inherent limitations and has traditionally required creative experimental solutions. This is due to the fact that they are effectively an evolutionary dead-end. Reproductive isolation makes them infertile or lethal and, therefore, impossible to maintain as a stable stock. As a result, it is only possible to perform experiments with very limited starting material, making large scale high-throughput experiments almost impossible.
Genomic localization in hybrids has been performed in polytene chromosomes also in other publications before ours. Despite polytene chromosomes being an extremely valuable system and allowing to perform localization studies even with very limited starting material, they present some obvious limitations. On the one hand, the mapping resolution is very low compared to genome-wide NGS techniques and, on the other hand, they constitute a very special tissue from which it is difficult to drive general conclusions. In addition, chromocenters, one of the key features in *Drosophila* hybrid incompatibilities, are under-replicated in polytene chromosomes, resulting in a bias towards other genomic regions.
In this work, we circumvent the problem by using an artificial system that mimics the hybrid condition, by co-overexpressing HMR and LHR in *D.mel* cultured cells. This system allowed us to perform high throughput experiments to study both the interactome and the genome-wide localization of HMR.
Although using this cell inducible system has allowed us to obtain important insights that are largely comparable to our observations in polytene chromosomes, we have to keep in mind that we are not studying a real hybrid background and only the *D.mel* genome, transcriptome and proteome are present, with the possible consequence of missing important interactions among species-specific adaptations.

Newly developed tools for large-scale and high-throughput molecular studies in hybrids

During my thesis work, in addition to the published work presented here, I have developed two new strategies to study hybrids. As these results are not part of this thesis I will only briefly mention them and will not enter in the details of the characterization and the preliminary results that I obtained.
The first system was developed in a joint collaboration with Jacob Cooper and Nitin Phadnis from the University of Utah in the US and with Yuri Schwartz and Tania Kahn from the Umea University in Sweden. In the time I visited the Schwartz lab at the Umea University and in the following

months in Munich, we generated primary hybrid cell lines from embryos of *D. melanogaster* mothers and *D. simulans* fathers immortalized through the overexpression of the Ras proto-oncogene (Fig. 8A).
In a second collaboration with Alessandro Scacchetti from Peter Becker's lab, we used a genetic trick to facilitate the notoriously laborious virgin collection process for crosses. Thanks to a fly stock carrying a Y-linked inducible pro-apoptotic gene, we could selectively obtain an exclusively *D.mel* female virgin population to cross with *D.sim* fathers (Fig. 8B). This technique allowed us to significantly reduce the time required to collect flies for the genetic crosses as well as to reduce potential collection errors.

These two new methods that significantly reduce the time and the effort necessary to collect hybrids, will allow in the future to perform large-scale high-throughput experiments and to unravel important molecular mechanisms underlying hybrid incompatibilities.

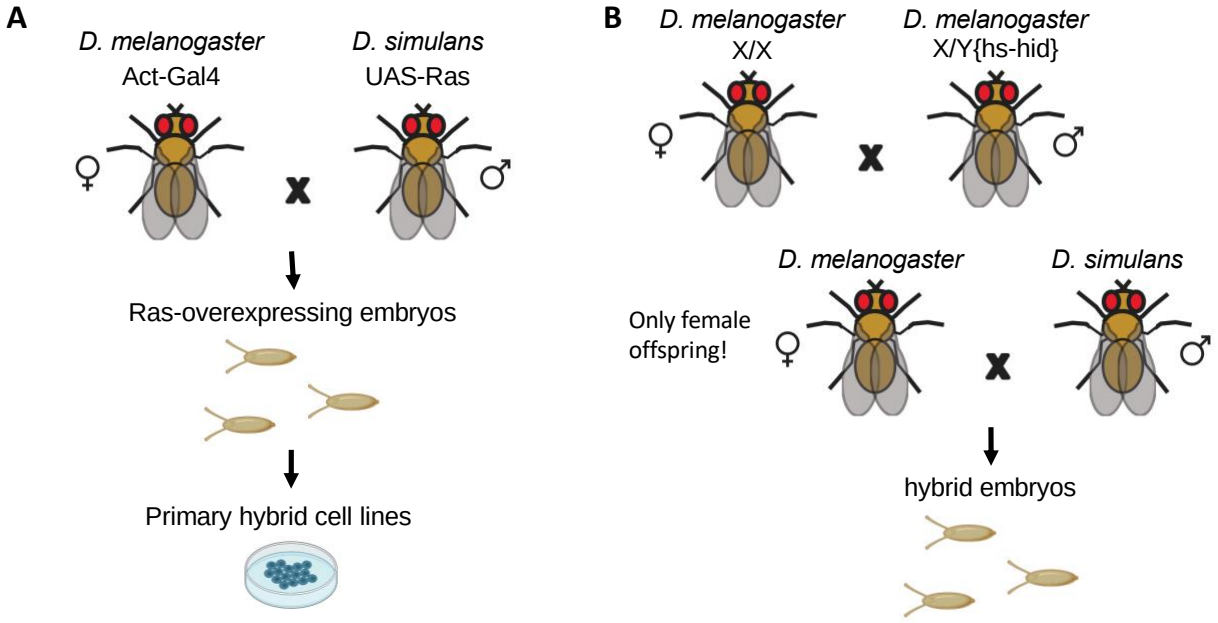


Fig. 8. Two methods for large-scale and high-throughput studies in hybrids. (A) Generation of stable embryonic hybrid cell lines from offspring of *D.mel* mothers and *D.sim* fathers. Hybrid cell lines were generated by crossing *D.mel* mothers with *D.sim* fathers and collecting embryos overexpressing Ras through a UAS-Gal4 system. Embryos were then cultivated *in vitro* for several months and throughout multiple generations, until stable cell lines were obtained. (B) *D.mel* wildtype females are crossed to males carrying a heat-shock inducible pro-apoptotic gene. In the offspring, males mortality is induced by heat-shock during development resulting in only virgin female offspring. *D.mel* female offspring is crossed with *D.sim* males and wildtype hybrids are collected. This process can be performed in large scale due to the ease in collection.

Working model

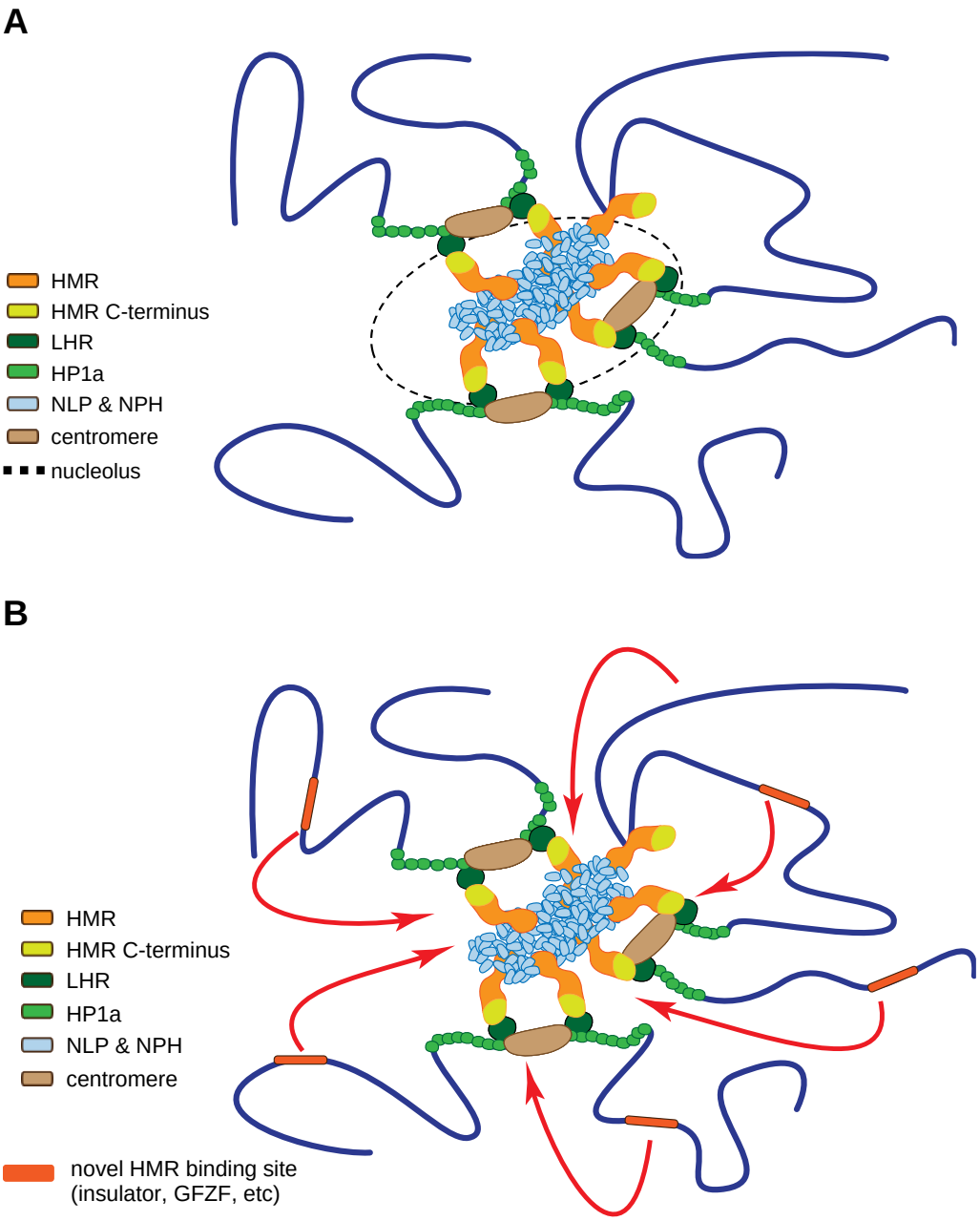


Fig. 9. Model of the role of the HMR complex at chromocenters in physiological conditions (A) and upon HMR/LHR overexpression or in hybrids (B). HMR/LHR complex forms a bridge between pericentromeric heterochromatin and centromeric chromatin that facilitates their coalescence at chromocenters, at the nucleolar periphery (A). Upon overexpression of HMR/LHR and in interspecific hybrids, HMR mislocalizes to novel binding sites along chromosome arms (B). We propose that the novel HMR binding sites could be tethered by the HMR/LHR complex at chromocenters giving rise to an overcondensed architecture (B) that is incompatible with normal cellular functions. One could further speculate that mutants interfering with HMR/LHR complex formation could, in turn, release this dysfunctional condensation rescuing hybrids viability.

Bibliography

Abad, J.P.; De Pablos, B.; Osoegawa, K.; De Jong, P.J.; Martin-Gallardo, A.; Villasante, A. TAHRE, a novel telomeric retrotransposon from Drosophila melanogaster, reveals the origin of Drosophila telomeres. Mol. Biol. Evol. 2004, 21, 1620–1624.

Akhtar, A., Gasser, S. The nuclear envelope and transcriptional control. *Nat Rev Genet* **8**, 507–517 (2007). https://doi.org/10.1038/nrg2122

Akhtar, A., Zink, D. & Becker, P. Chromodomains are protein–RNA interaction modules. *Nature* 407, 405–409 (2000). https://doi.org/10.1038/35030169

Allshire, R. C., & Madhani, H. D. (2017). Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2017.119

Ambrus A. M., V. I. Rasheva, B. N. Nicolay, and M. V. Frolov, 2009 Mosaic Genetic Screen for Suppressors of the de2f1 Mutant Phenotype in Drosophila. Genetics 183: 79–92. https://doi.org/10.1534/genetics.109.104661

Andreyeva, E.N., Belyaeva, E.S., Semeshin, V.F., Pokholkova, G.V., and Zhimulev, I.F. (2005). Three distinct chromatin domains in telomere ends of polytene chromosomes in Drosophila melanogaster Tel mutants. J Cell Sci 118, 5465-5477.

Anselm, E., Thomae, A. W., Jeyaprakash, A. A., & Heun, P. (2018). Oligomerization of Drosophila Nucleoplasmin-Like Protein is required for its centromere localization. *Nucleic Acids Research*, *46*(21), 11274–11286. https://doi.org/10.1093/nar/gky988 PMID: 30357352

Ashton-Beaucage D., C. M. Udell, P. Gendron, M. Sahmi, M. Lefrançois, et al., 2014 A Functional Screen Reveals an Extensive Layer of Transcriptional and Splicing Control Underlying RAS/MAPK Signaling in Drosophila. PLoS Biol 12: e1001809. https://doi.org/10.1371/journal.pbio.1001809

Aruna, S., Flores, H. A., & Barbash, D. A. (2009). Reduced fertility of Drosophila melanogaster Hybrid male rescue (Hmr) mutant females is partially complemented by Hmr orthologs from sibling species. *Genetics*, *181*(4), 1437–1450. https://doi.org/10.1534/genetics.108.100057 PMID: 19153254

Ausio, J., Drinnenberg, I. A., Berger, F. De, Elsa, S. J., Andersen, P. R., Ausio, J., Bickmore, A., Blackwell, A. R., Erwin, D. H., Gahan, J. M., Gaut, B. S., Harvey, H., Henikoff, S., Kao, J. Y., Kurdistani, S. K., Lemos, B., Levine, M. T., Warnecke, T., & Wolfe, K. H. (2019). *EvoChromo :*

towards a synthesis of chromatin biology and evolution. *October*. https://doi.org/10.1242/dev.178962

Baldi, S., Korber, P., & Becker, P. B. (2020). Beads on a string—nucleosome array arrangements and folding of the chromatin fiber. *Nature Structural and Molecular Biology*, *27*(2), 109–118. https://doi.org/10.1038/s41594-019-0368-x

Barbash DA, Siino DF, Tarone AM, Roote J. 2003. A rapidly evolving MYB-related protein causes species isolation in Drosophila. Proc Natl Acad Sci U S A. 100(9):5302–5307

Barbash, D. A., Awadalla, P., & Tarone, A. M. (2004). Functional divergence caused by ancient positive selection of a Drosophila hybrid incompatibility locus. *PLoS Biology*, *2*(6), 839–848. https://doi.org/10.1371/journal.pbio.0020142

Barbash, D. A., and J. G. Lorigan, 2007 Lethality in Drosophila mel- anogaster/Drosophila simulans species hybrids is not associated with substantial transcriptional misregulation. J. Exp. Zool. B Mol. Dev. Evol. 308B: 74–84

Barbash, D. A. (2010). Ninety years of Drosophila melanogaster hybrids. *Genetics*, *186*(1), 1–8. https://doi.org/10.1534/genetics.110.121459

Bastock R, Johnston DS. Drosophila oogenesis. Curr Biol. 2008; 18: R1082–R1087. https://doi.org/10.1016/j.cub.2008.09.011

Baumann, D., Dai, M.-S., Lu, H., & Gilmour, D. (2018). GFZF, a Glutathione S-Transferase Protein Implicated in Cell Cycle Regulation and Hybrid Inviability, Is a Transcriptional Coactivator Douglas. *Molecular and Cellular Biology*, 1–16. https://dx.doi.org/10.1128/mcb.00476-17

Bayes, J.J., and Malik, H.S. (2009). Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. Science 326, 1538-1541.

Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823–837. doi: 10.1016/j.cell.2007.05.009

Bickmore, W. A., & Van Steensel, B. (2013). Genome architecture: Domain organization of interphase chromosomes. *Cell*, *152*(6), 1270–1284. https://doi.org/10.1016/j.cell.2013.02.001

Blum, J. A., Bonaccorsi, S., Marzullo, M., Palumbo, V., Yamashita, Y. M., Barbash, D. A., & Gatti, M. (2017). The hybrid incompatibility genes Lhr and Hmr are required for sister chromatid

detachment during anaphase but not for centromere function. *Genetics*. https://doi.org/10.1534/genetics.117.300390 PMID: 29046402

Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid- like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. Proc. Natl Acad. Sci. USA 108, 4334–4339 (2011).

Brideau NJ, Flores HA,Wang J,Maheshwari S,Wang X, BarbashDA. 2006. Two Dobzhansky–Muller genes interact to cause hybrid lethality in Drosophila. Science 314(5803):1292–1295

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., Wan, K.H., Yu, C., Zhang, D., Carlson, J.W., Cherbas, L., Eads, B.D., Miller, D., Mockaitis, K., Roberts, J., Davis, C.A., Frise, E., Hammonds, A.S., Olson, S., Shenker, S., Sturgill, D., Samsonova, A.A., Weiszmann, R., Robinson, G., Hernandez, J., Andrews, J., Bickel, P.J., Carninci, P., Cherbas, P., Gingeras, T.R., Hoskins, R.A., Kaufman, T.C., Lai, E.C., Oliver, B., Perrimon, N., Graveley, B.R., Celniker, S.E. (2014). Diversity and dynamics of the Drosophila transcriptome. Nature 512(7515): 393--399. http://dx.doi.org/10.1038/nature12962

Casacuberta, E. (2017). Drosophila: Retrotransposons making up telomeres. *Viruses*, *9*(7). https://doi.org/10.3390/v9070192

Chakravarti, D., LaBella, K. A., & DePinho, R. A. (2021). Telomeres: history, health, and hallmarks of aging. *Cell*, *184*(2), 306–322. https://doi.org/10.1016/j.cell.2020.12.028

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N., et al. (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature 450, 203-218

Cooper JC, **Lukacs A**, Reich S, Schauer T, Imhof A, Phadnis N. Altered Localization of Hybrid Incompatibility Proteins in Drosophila. Larracuente AM, editor. Molecular Biology and Evolution. 2019; 36: 1783–1792. https://doi.org/10.1093/molbev/msz105

Cosgrove, M.S. (2012). Writers and readers: deconvoluting the harmonic complexity of the histone code. Nat Struct Mol Biol 19, 739-740

Coyne, J. A., & Orr, H. A. (1989). *Patterns of Speciation in Drosophila*.

Creighton, H.B., and McClintock, B. (1931). A Correlation of Cytological and Genetical Crossing-Over in Zea Mays. Proc. Natl. Acad. Sci. USA17, 492–497.

Cutler G, Perry KM, Tjian R. 1998. Adf-1 is a nonmodular transcription factor that contains a TAF-binding Myb-like motif. Mol Cell Biol. 18:2252–2261.

Dai M.-S., X.-X. Sun, J. Qin, S. M. Smolik, and H. Lu, 2004 Identification and characterization of a novel Drosophila melanogaster glutathione S- transferase-containing FLYWCH zinc finger protein. Gene 342: 49–56. https://doi.org/10.1016/j.gene.2004.07.043

Darwin, Charles (1859). *On the Origin of Species By Means of Natural Selection.*

Di Stefano, M., & Cavalli, G. (2022). Integrative studies of 3D genome organization and chromatin structure. *Current Opinion in Structural Biology*, *77*, 102493. https://doi.org/10.1016/j.sbi.2022.102493

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li et al., 2012 Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485: 376–380.

Dobzhansky, T. (1936). Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids. Genetics 21, 113-135.

Dobzhansky, Th., 1937, 1941, 1951 Genetics and the Origin of Species. Ed. 1, 2, 3. Columbia University Press, New York.

England BP, Admon A, Tjian R. 1992. Cloning of Drosophila transcription factor Adf-1 reveals homology to Myb oncoproteins. Proc Natl Acad Sci USA. 89:683–687

Erdel, F., & Rippe, K. (2018). Formation of Chromatin Subcompartments by Phase Separation. *Biophysical Journal*, *114*(10), 2262–2270. https://doi.org/10.1016/j.bpj.2018.03.011

Falk, M., Feodorova, Y., Naumova, N. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019). https://doi.org/10.1038/s41586-019-1275-3

Fawcett, D.W. 1966. On the occurrence of a fibrous lamina on the inner aspect of the nuclear envelope in certain cells of vertebrates. Am.J.Anat. 119: 129–145.

Ferree, P.M., and Barbash, D.A. (2009). Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. PLoS Biol 7, e1000234.

Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., Brugman, W., de Castro, I. J., Kerkhoven, R. M., Bussemaker, H. J., & van Steensel, B. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell*, *143*(2), 212–224. https://doi.org/10.1016/j.cell.2010.09.009

Flemming, W. Zellsubstanz, Kern und Zelltheilung (in German) (F.C.W. Vogel, Leipzig, 1882).

Franke, A., and Baker, B.S. (1999). The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. Mol Cell 4, 117-122.

Fukagawa, T. & Earnshaw, W. C. The centromere: chromatin foundation for the kinetochore machinery. Dev. Cell 30, 496–508 (2014)

Georgiev, P.G., and Gerasimova, T.I. (1989). Novel genes influencing the expression of the yellow locus and mdg4 (gypsy) in Drosophila melanogaster. Mol Gen Genet 220, 121-126.

Gerasimova, T. I., K. Byrd, and V. G. Corces, 2000 A chromatin insulator determines the nuclear localization of DNA. Mol. Cell 6: 1025–1035.

Gerasimova, T.I., Gdula, D.A., Gerasimov, D.V., Simonova, O., and Corces, V.G. (1995). A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. Cell 82, 587-597.

Gerland, T. A., Sun, B., Smialowski, P., **Lukacs, A.**, Thomae, A. W., & Imhof, A. (2017). The Drosophila speciation factor HMR localizes to genomic insulator sites. *PLoS ONE*, *12*(2), 1–18. https://doi.org/10.1371/journal.pone.0171798 PMID: 28207793

Gomez-Diaz, E., and Corces, V. G. 2014 Architectural proteins: regulators of 3D genome organization in cell fate. Trends Cell Biol. 24: 703–711.

Gonzalez I., J. Mateos-Langerak, A. Thomas, T. Cheutin, and G. Cavalli, 2014 Identification of Regulators of the Three-Dimensional Polycomb Organization by a Microscopy-Based Genome-wide RNAi Screen. Mol. Cell 54: 485–499. https://doi.org/10.1016/j.molcel.2014.03.004

Greil, F., De Wit, E., Bussemaker, H. J., & Van Steensel, B. (2007). HP1 controls genomic targeting of four novel heterochromatin proteins in Drosophila. *EMBO Journal*, *26*(3), 741–751. https://doi.org/10.1038/sj.emboj.7601527

Gu, T. & Elgin, S. C. Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in Drosophila. PLoS Genet. 9, e1003780 (2013)

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eu- ssen, B.H., de Klein, A., Wessels, L., de Laat, W., and van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453, 948–951. Hakim,

Heitz, E., 1928 Das heterochromatin der Moose. I. Jb. wiss. Bot. 69: 762–818.

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

Hudson AM, Cooley L. Methods for studying oogenesis. Methods. Elsevier Inc.; 2014;68(1):207–17.

Hutter P, Ashburner M. 1987. Genetic rescue of inviable hybrids between Drosophila melanogaster and its sibling species. Nature 327(6120):331–333

Hutter, P., J. Roote and M. Ashburner, 1990 A genetic basis for the inviability of hybrids between sibling species ofDrosophila. Genetics 124: 909–920.

Kharchenko, P.V., A.A. Alekseyenko, Y.B. Schwartz, et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471: 480–485. https://doi.org/10.1038/nature09725

Kelleher ES, Edelman NB, Barbash DA. Drosophila Interspecific Hybrids Phenocopy piRNA-Pathway Mutants. Noor MAF, editor. PLoS Biology. 2012; 10: e1001428. https://doi.org/10.1371/journal.pbio.1001428 PMID: 23189033

Kirilly D, Xie T. The Drosophila ovary: an active stem cell community. Cell Res. 2007; 17: 15–25. https://doi.org/10.1038/sj.cr.7310123 PMID: 17199109

Kitaoka, M., Smith, O. K., Straight, A. F., & Heald, R. (2022). Molecular conflicts disrupting centromere assembly contribute to Xenopus hybrid inviability. *BioRxiv*. https://doi.org/10.1101/2022.03.06.483208

Kochanova, N. Y., Schauer, T., Mathias, G. P., **Lukacs, A.**, Schmidt, A., Flatley, A., Schepers, A., Thomae, A. W., & Imhof, A. (2020). A multi-layered structure of the interphase chromocenter revealed by proximity-based biotinylation. *Nucleic Acids Research*, *16*, 1–18. https://doi.org/10.1093/nar/gkaa145 PMID: 32182352

Kondo S., and N. Perrimon, 2011 A Genome-Wide RNAi Screen Identifies Core Components of the G2-M DNA Damage Checkpoint. Sci. Signal. 4: rs1–rs1. https://doi.org/10.1126/scisignal.2001350

Kursel, L. E., & Malik, H. S. (2018). The cellular mechanisms and consequences of centromere drive. *Current Opinion in Cell Biology*, *52*, 58–65. https://doi.org/10.1016/j.ceb.2018.01.011

Jagannathan, M., Cummings, R., & Yamashita, Y. M. (2019). The modular mechanism of chromocenter formation in Drosophila. *ELife*, *8*. https://doi.org/10.7554/eLife.43938

Jagannathan, M., & Yamashita, Y. M. (2021). Defective Satellite DNA Clustering into Chromocenters Underlies Hybrid Incompatibility in Drosophila . *Molecular Biology and Evolution*, *38*(11), 4977–4986. https://doi.org/10.1093/molbev/msab221

Jenuwein T, Allis CD. Translating the histone code. Science. 2001 Aug 10;293(5532):1074-80. https://doi:10.1126/science.1063127

Leader, D.P., Krause, S.A., Pandit, A., Davies, S.A., Dow, J.A.T. (2018). FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. https://doi.org/10.1093/nar/gkx976

Lee, Y. C. G., Ogiyama, Y., Martins, N. M. C., Beliveau, B. J., Acevedo, D., Wu, C. -tin., Cavalli, G., & Karpen, G. H. (2020). Pericentromeric heterochromatin is hierarchically organized and spatially contacts H3K9me2 islands in euchromatin. *PLOS Genetics*, *16*(3), e1008673. https://doi.org/10.1371/journal.pgen.1008673

Lewis, E. B., 1954 The theory and application of a new method of detecting chromosomal rearrangements in Drosophila mela- nogaster. Am. Nat. 88: 225–239.

Lewontin, R. C. (1997). Dobzhansky's genetics and the origin of species: Is it still relevant? *Genetics*, *147*(2), 351–355.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293.

Liu, J., Ali, M., & Zhou, Q. (2020). *Establishment and evolution of heterochromatin*. 1–19. https://doi.org/10.1111/nyas.14303

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251-260.

Maheshwari, S., Tan, E. H., West, A., Franklin, F. C. H., Comai, L. and Chan, S. W. L. (2015). Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genet. 11, e1004970. https://doi:10.1371/journal.pgen.1004970

Maheshwari, S., Wang, J., & Barbash, D. A. (2008). Recurrent positive selection of the Drosophila hybrid incompatibility gene Hmr. *Molecular Biology and Evolution*, *25*(11), 2421–2430. https://doi.org/10.1093/molbev/msn190

Magaña-Acosta M and Valadez-Graham V (2020) Chromatin Remodelers in the 3D Nuclear Compartment. Front. Genet. 11:600615. doi: 10.3389/fgene.2020.600615

Malik, H. S. and Henikoff, S. (2001). Adaptive evolution of Cid, a centromere- specific histone in Drosophila. Genetics 157, 1293-1298.

Mayr, E. (1942). Systematics and the Origin of Species (New York: Columbia University Press).

Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J.C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science 323, 373- 375.

Muller, H. J., 1930 Types of visible variations induced by X-rays in Drosophila. J. Genet. 22: 299–334.

Muller, H. (1938). The remaking of chromosomes. Collecting Net 8, 182–198.

Muller, H.J., and Pontecorvo, G. (1942). Recessive genes causing interspecific sterility and other disharmonies between Drosophila melanogaster and simulans. Genetics 27, 157.

Naville, M., S. Henriet, I. Warren, et al. 2019. Massive changes of genome size driven by expansions of non- autonomous transposable elements. Curr. Biol. 29: 1161– 1168.e1166.

Nemeth, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Peterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Langst, G. (2010). Initial genomics of the human nucleolus. PLoS Genet. 6, e1000889.

Nicetto, D., & Zaret, K. S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. *Current Opinion in Genetics & Development*, *55*, 1–10. https://doi.org/10.1016/j.gde.2019.04.013

Olovnikov, A.M. (1973). A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological signif- icance of the phenomenon. J. Theor. Biol. 41, 181–190.

Olszak, A. M., Van Essen, D., Pereira, A. J., Diehl, S., Manke, T., Maiato, H., Saccani, S., & Heun, P. (2011). Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nature Cell Biology*, *13*(7), 799–808. https://doi.org/10.1038/ncb2272

Orr, H. A., Masly, J. P., & Phadnis, N. (2007). Speciation in Drosophila: From phenotypes to molecules. *Journal of Heredity*, *98*(2), 103–110. https://doi.org/10.1093/jhered/esl060

Padeken, J., & Heun, P. (2014). Nucleolus and nuclear periphery: Velcro for heterochromatin. *Current Opinion in Cell Biology*, *28*(1), 54–60. https://doi.org/10.1016/j.ceb.2014.03.001

Padeken, J., Mendiburo, M. J., Chlamydas, S., Schwarz, H. J., Kremmer, E., & Heun, P. (2013). The Nucleoplasmin Homolog NLP Mediates Centromere Clustering and Anchoring to the Nucleolus. *Molecular Cell*, *50*(2), 236–249. https://doi.org/10.1016/j.molcel.2013.03.002

Pai, C.Y., Lei, E.P., Ghosh, D., and Corces, V.G. (2004). The centrosomal protein CP190 is a component of the gypsy chromatin insulator. Mol Cell 16, 737-748

Pardue, M.L.; DeBaryshe, P.G. Drosophila telomeres: A variation on the telomerase theme. Fly 2008, 2, 101–110.

Parkhurst, S.M., Harrison, D.A., Remington, M.P., Spana, C., Kelley, R.L., Coyne, R.S., and Corces, V.G. (1988). The Drosophila su(Hw) gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. Genes Dev 2, 1205-1215.

Peng JC, Karpen GH: H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. Nat Cell Biol 2007, 9:25-35

Phadnis, N., Baker, E. P., Cooper, J. C., Frizzell, K. A., Hsieh, E., de la Cruz, A. F. A., Shendure, J., Kitzman, J. O., & Malik, H. S. (2015). An essential cell cycle regulation gene causes hybrid inviability in Drosophila. *Science*, *350*(6267), 1552–1555. https://doi.org/10.1126/science.aac7504

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

Provost E, Hersperger G, Timmons L, Ho WQ, Hersperger E, Alcazar R, Shearn A. 2006. Loss-of-function mutations in a glutathione S-trans- ferase suppress the prune-Killer of prune lethal interaction. Genetics 172(1):207–219

Raffa, G.D., Ciapponi, L., Cenci, G., and Gatti, M. (2011). Terminin: a protein complex that mediates epigenetic maintenance of Drosophila telomeres. Nucleus 2, 383-391.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, et al. Genome-wide location and function of DNA binding proteins. Science. 2000;290(5500):2306–2309. doi:10.1126/science.290.5500.2306.

Roach, K. C., Ross, B. D., & Malik, H. S. (2013). Rapid evolution of centromeres and centromeric/kinetochore proteins. *Oxford Press*, *15*(1), 583–605. https://doi.org/10.1093/acprof

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods. 2007;4(8):651–657. doi: 10.1038/nmeth1068.

Rosin,L.F., Nguyen,S.C. and Joyce,E.F. (2018) Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in Drosophila nuclei. PLoS Genet., 14, e1007393.

Sanulli, S., & Narlikar, G. J. (2020). Liquid-like interactions in heterochromatin : Implications for mechanism and regulation. *Current Opinion in Cell Biology*, *64*, 90–96. https://doi.org/10.1016/j.ceb.2020.03.004

Satyaki PRV, Cuykendall TN, Wei KH-C, Brideau NJ, Kwak H, Aruna S, Ferree PM, Ji S, Barbash DA. 2014. The Hmr and Lhr hybrid incom- patibility genes suppress a broad range of heterochromatic repeats. PLoS Genet. 10(3):e1004240 https://doi.org/10.1371/journal.pgen.1004240 PMID: 24651406

Sawamura, K. (2011). Chromatin Evolution and Molecular Drive in Speciation. *International Journal of Evolutionary Biology*, *2012*, 1–9. https://doi.org/10.1155/2012/301894

Sawamura, K., and Yamamoto, M.T. (1997). Characterization of a reproductive isolation gene, zygotic hybrid rescue, of Drosophila melanogaster by using minichromosomes. Heredity 79, 97-103.

Schubert, T., Pusch, M.C., Diermeier, S., Benes, V., Kremmer, E., Imhof, A., and Langst, G. (2012). Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. Mol Cell 48, 434-444.

Schuster-Bockler, B. & B. Lehner. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature 488: 504–507.

Schwartz, Y. B., & Cavalli, G. (2017). *Three-Dimensional Genome Organization and*. *205*(January), 5–24. https://doi.org/10.1534/genetics.115.185132

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G. P., Bank, C., Bränström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Foote, A. D., Franchini, P., Jiggins, C. D., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Shukla, V., Habib, F., Kulkarni, A., and Ratnaparkhi, G.S. (2014). Gene duplication, lineage-specific expansion, and subfunctionalization in the MADF-BESS family patterns the Drosophila wing hinge. Genetics 196, 481-496.

Solovei I, Kreysing M, Lanctôt C, Kösem S, Peichl L, Cremer T, Guck J, Joffe B. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Cell. 2009 Apr 17;137(2):356-68. https://doi.org/10.1016/j.cell.2009.01.052

Steiner, F. A., & Henikoff, S. (2015). Diversity in the organization of centromeric chromatin. *Current Opinion in Genetics and Development*, *31*, 28–35. https://doi.org/10.1016/j.gde.2015.03.010

Sturtevant A. H., 1919 A New Species Closely Resembling Drosophila Melanogaster. Psyche J. Entomol. 26: 153–155. https://doi.org/10.1155/1919/97402

Tachiwana, H. et al. Crystal structure of the human centromeric nucleosome containing CENP-A. Nature 476, 232–235 (2011).

Talbert, P. B., & Henikoff, S. (2020). What makes a centromere ? *Experimental Cell Research*, *389*(2), 111895. https://doi.org/10.1016/j.yexcr.2020.111895

Thomae AW, Schade GOM, Padeken J, Borath M, Vetter I, Kremmer E, Heun P, Imhof A. 2013. A Pair of Centromeric Proteins Mediates Reproductive Isolation in Drosophila Species. Developmental Cell. 2013; 27: 412– 424. https://doi.org/10.1016/j.devcel.2013.10.001 PMID: 24239514

Ting, C.T., Tsaur, S.C., Wu, M.L., and Wu, C.I. (1998). A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282, 1501-1504.

van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G.J., Ariyurek, Y., den Dunnen, J.T., and Lamond, A.I. (2010). High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol. Biol. Cell 21, 3735– 3748.

van Steensel, B. & Furlong, E. E. M. (2019). The role of transcription in shaping the spatial organization of the genome. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-019-0114-6

van Steensel, B., and S. Henikoff, 2000 Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat. Biotechnol. 18: 424–428.

Watanabe TK. 1979. A gene that rescues the lethal hybrids between. Drosophila melanogaster and D. simulans. Japan Journal of Genetics 54:325–331

Watson, J.D. (1972). Origin of concatemeric T7 DNA. Nat. New Biol. 239, 197–201.

Wei, K. H. C., Clark, A. G., & Barbash, D. A. (2014). Limited gene misregulation is exacerbated by allele-specific upregulation in lethal hybrids between Drosophila melanogaster and Drosophila simulans. *Molecular Biology and Evolution*, *31*(7), 1767–1778. https://doi.org/10.1093/molbev/msu127

Yoshida, K., Ishikawa, A., Toyoda, A., Shigenobu, S., Fujiyama, A., & Kitano, J. (2018). Functional divergence of a heterochromatin-binding protein during stickleback speciation. *Molecular Ecology*, *28*. https://doi.org/10.1111/mec.14841

Zhou, K., Gebala, M., Woods, D., Sundararajan, K., Edwards, G., Krzizike, D., Wereszczynski, J., Straight, A. F., & Luger, K. (2022). CENP-N promotes the compaction of centromeric chromatin. *Nature Structural and Molecular Biology*, *29*(4), 403–413. https://doi.org/10.1038/s41594-022-00758-y

Acknowledgements

Since the acknowledgements in the publications presented here are already covering the scientific part, this section will be somehow atypical in that I want to dedicate a major part of it to all the non-scientific parts that have been running in parallel to my PhD.
My PhD has been a long journey, sometimes exciting and rewarding, sometimes depressing and tough, but it has definitely been worth it. Like I mentioned in the opening pages of this thesis, my younger self has always been dreaming of spending his life researching and diving deep into the secrets of Nature and Life. Therefore, I can only feel incredibly lucky and grateful for having had the opportunity to work and grow here in this department but also campus and scientific community in general. I am thankful for having been working for many years at the edges of scientific knowledge within a lively and critical environment and state of the art technology available. None of the difficulties I have encountered during this path would make me regret having chosen this path. Each of these challenges has been a learning opportunity and has contributed to make me not only a better scientist, but also a more mature, analytical and aware person.
There is an uncountable number of people I would like to thank that walked on my side through this incredible journey. I am very likely to forget some of them just because so many people have crossed my path and shared something with me. A bit of their knowledge, experience, know-how, of their time, or just a chat, a smile, a coffee or a beer. Or a meal in a restaurant. I owe so much to this department, to the people that have been working here, to all the opportunities it offered me. It has been my second house for many years.
I first of all want to thank my supervisor, Prof. Dr. Axel Imhof, for giving me the opportunity to work in this exciting scientific environment, for having always given me the trust and the freedom to be the leader of my own projects, to be an independent thinker and bring forward my own ideas, to build collaborations, to try and make mistakes, and for having given me the resources and the support to carry on my research. I am very thankful for having had the opportunity to work on a project I fell in love with from the very first time I heard it. It was during a conference in Spetses, Greece, and back then I was working as a research assistant in Rome. I still vividly remember when I attended his charming presentation and first heard of chromatin and speciation in *Drosophila*. I understood without hesitation that I wanted to work on that project, at the interface between Chromatin and Evolutionary Biology, trying to unveil the molecular mechanisms underlying the big evolution question of how species form. And I remember enquiring with Prof. Dr. Axel Imhof for a PhD position after a long and exciting scientific discussion in the early morning bus to the airport. I am grateful it worked out. Over the years we had fruitful and exciting discussions as well as disagreements and conflicts, successful and unsuccessful projects, but I have always been aware of the mutual appreciation and respect.
I am deeply grateful to Prof. Dr. Peter Becker, that has been throughout my PhD not the head of the department and a brilliant scientist but first of all a leader and a role model. I have learnt from him what the very essence of leadership means: leading by example, being kind and open to

anyone, and yet firm and determined in pursuing major objectives. This is all witnessed by the department where a successful and critical scientific approach coexists with a friendly and collaborative atmosphere.
I want to thank Carolin, Edith Mentele, Barbara, Daisy, Tatyana for several different reasons, but in particular I want to thank Edith Mueller, who has been a constant reassuring presence. It took me a couple of weeks to understand that behind a somehow grumpy surface there was the sweetest and most helpful person I could hope for.

I want to thank the whole Imhof lab throughout the years.
For being a fundamental scientific support and critical voice, in particular during the first phase of my PhD, I want to thank the Speciation Team in particular: my friends Thomas Gerland, Natalia Kochanova, Bo Sun and then also Ulku Uzun and Peter Krueger.
Thanks to Andi Thomae who, despite some initial frictions, has given me a “crash course” on fly genetics, has been present throughout my PhD and with time we also learnt to understand each other. With him I shared many inspiring and productive scientific discussions that have led to ideas and collaborations and with whom I am happy to share three publications.
I want to thank all the students that I had the pleasure to supervise and from whom I learnt a lot about teaching and leadership: Egle, Merve and Peter, in particular.
The many “Imhofini” labmates that then turned into friends, from Thomas, Natalia and Victor, to Bo, Wasim, Anu, Magdalini, Vera, Rupam and Marco.
Thanks to the Mass Spec team that I had a long collaboration with. Starting from Andreas and the very ambitious complexome project, to Ignasi, that eventually helped me measure most of my samples, but also Marc and Pierre. And finally Marco. And thanks to Shibo for the nice scientific and non-scientific discussions.
I want to thank my TAC members: Prof. Dr. Peter Becker, Prof. Dr. Nicolas Gompel, Prof. Dr. Julien Gagneur, Dr. Tobias Straub.

I want to thank the many awesome and brilliant scientists for the discussions that helped me grow and sharpen my scientific mind (and not only), from the students to the professors.
I want to thank the numerous people I have collaborated with and, in this context, particularly those collaborations that didn’t end up in a publication. Among others Andi Thomae, Tamas Schauer, Thomas Gerland, Natalia Kochanova, Anu Venkatasubramani, Victor Solis, Alessandro Scacchetti, Catherine Regnard, Kenneth Boerner, Alessandro Angerilli, Alessandro Baldi, and many many more.
Tamas, in particular, who co-authored all of my publications but was also an occasional non-scientific buddy for running and for partying.

I also want to thank my external collaborators, Prof. Dr. Yuri Schwartz and Dr. Tania Kahn from the Umea University in Sweden, that made me feel at home when they hosted me for a collaboration period in their lab that resulted in the generation of primary hybrid cell lines.

I want to thank Prof. Dr. Nitin Phadnis and Dr. Jacob Cooper from the University of Utah with whom I collaborated for the second publication presented in this thesis. Finally, I want to thank Prof. Dr. Patrick Heun and Dr. Eduard Anselm from the University of Edinburgh, UK for the collaborations in terms of discussions, ideas and resources.

I want to thank all the BMC people that through the years have been there as colleagues, travel companions and many of them also as friends, gathered first in the BMC coffee team, or the Octopi or the BMC running club. Alessandro Angerillli, my previous study companion and benchmate in Rome and then cofounder of the BMC coffee team. And then Daniil, Sophia, Tom, Sylvain, Catherine, Filippo, Lisa, Nadia, Irina, Ale Scacchetti, Magdalo, Anu, Shahaf, Andrea Terrasi, Martina Peritore, Iris, Petra, and many others.
I want to thank Catherine et Filippo for having been mentors, pillars and friends over the years.

Thanks to Sophia, Elizabeth and Brone that read this manuscript and gave me helpful comments and precious help.

I am incredibly thankful for having been part of the two amazing graduate schools I had the luck to be part of: QBM and IRTG.
Their constantly helpful, friendly and supportive managers and administrators Elizabeth and Markus but also Michael, Filiz, Mara and Julia before.
QBM and IRTG with all the outstanding training and programs. That really built up not only a beautiful network but also my career and my hard and soft skills. All these networking events and the retreats and the opportunity to travel to Cambridge and Heidelberg and to perform a period of collaboration in Sweden.
QBM has been like a family, particularly during the first year when I was new in Munich and experiencing its school-like intensive programme has been sometimes hard to reconcile with lab work but also extremely rewarding in terms of education and scientific growth in the direction of a quantitative approach to science and biology. And also for the many classmates that I met there and that became great friends and that I am still in contact with: Rahmi, Madlin, Linda, Vicente, Nanni, Elena, Gokcen, Daniela, Victor, Ellie, Alena, Laia, Brone…
And the IRTG network with Marco La Fortezza, Andrea Storti, Alessio, Thomas the Dutch, Vanessa, Claudia, Martina P and many more.

Last but not least, all the life that has been going on out of my scientific community and that sometimes, I now clearly realize, I have sacrificed too much.
Grazie alla mia meravigliosa famiglia, a mamma, a papa', a Bea e Lorenzo, ed anche a Libero, che ci sono sempre stati e mi hanno sempre supportato indipendentemente dalle alterne fortune. Grazie a mia nonna che ha sempre creduto in me. Grazie a Giorgio Camilloni e ad Elisa Cesarini per avermi iniziato alla ricerca ed alla biologia molecolare. Grazie Carmelo Frisina, a Robertino Zanoni, a Roberto Argano ed ai molti altri professori che hanno alimentato in me la fiamma della curiosità scientifica.

Danke Tante Almut für deine Unterstützung.
Merci Ariane pour avoir partagé un gros morceau de mon doctorat. Merci d'avoir été là d'une façon ou de l'autre, dans les moments heureux et difficiles, avec la distance, les differents humeurs et les dépressions, merci d'avoir supporté mes soirées et mes week-ends au labo.
The many places that I called home and the beautiful flatmates that were with me, some of them will be among my best friends for the rest of my life like Marco, Casto, Yann, Joachim, Carlos, Stephan, Mikel, Elena and Iman.
Thanks to the Palombelle, Andrea and Elena M, Elena B, Silvia, Ale, Alessio e Vero.

And thanks to the friends that have been there in the last couple of years, the most difficult ones, and that offered me a safe space or just a smile. Thanks to Almut and the mountains, to Sophia, to Judith that was there during the last writing frenzy, to Brone and the zoo, to Daniil and Schroedi, to Rahmi Deniz and Didem, to Martina and Antonio, to Victor and Lisa, to Mina. Thanks to Andreea, that now has to stop smoking.
Thanks to Kathrin for the beautiful time together, for the inspiring walks, for the positive potato, for the support I may have overlooked or given for granted, thank you for seeing my potential and for believing in me.
Each and every of these people shared a bit with me and helped me build this beautiful experience. I would like to thank all of them for the small to large contributions that they had during one of the most incredible adventures of my life so far.