



UPPSALA  
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Medicine 1038*

# Meiotic Recombination in Human and Dog

*Targets, Consequences and Implications for Genome  
Evolution*

JONAS BERGLUND



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2014

ISSN 1651-6206  
ISBN 978-91-554-9057-7  
urn:nbn:se:uu:diva-233195

Dissertation presented at Uppsala University to be publicly examined in B41, BMC, Husargatan 3, Uppsala, Thursday, 20 November 2014 at 13:15 for the . The examination will be conducted in English. Faculty examiner: assistant professor Nadia Singh (Department of Biological Sciences, North Carolina State University).

### **Abstract**

Berglund, J. 2014. Meiotic Recombination in Human and Dog. Targets, Consequences and Implications for Genome Evolution. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1038. 43 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9057-7.

Understanding the mechanism of recombination has important implications for genome evolution and genomic variability. The work presented in this thesis studies the properties of recombination by investigating the effects it has on genome evolution in humans and dogs.

Using alignments of human genes with chimpanzee and macaque orthologues we studied substitution patterns along the human lineage and scanned for evidence of positive selection. The properties mirror the situation in human non-coding sequences with the fixation bias 'GC-biased gene conversion' (gBGC) as a driving force in the most rapidly evolving regions. By assigning candidate genes to distinct classes of evolutionary forces we quantified the extent of those genes affected by gBGC to 20%. This suggests that human-specific characters can be prompted by the fixation bias of gBGC, which can be mistaken for selection.

The gene PRDM9 controls recombination in most mammals, but is lacking in dogs. Using whole-genome alignments of dog with related species we examined the effects of PRDM9 inactivation. Additionally, we analyzed genomic variation in the genomes of several dog breeds. We identified that non-allelic homologous recombination (NAHR) via sequence identity, often GC-rich, creates structural variants of genomic regions. We show that these regions, which are also found in dog recombination hotspots, are a subset of unmethylated CpG-islands (CGIs). We inferred that CGIs have experienced a drastic increase in biased substitution rates, concurrent with a shift of recombination to target these regions. This enables recurrent episodes of gBGC to shape their distribution.

The work presented in this thesis demonstrates the importance of meiotic recombination on patterns of molecular evolution and genomic variability in humans and dogs. Bioinformatic analyses identified mechanisms that regulate genome composition. gBGC is presented as an alternative to positive selection and is revealed as a major factor affecting allele configuration and the emergence of accelerated evolution on the human lineage. Characterization of recombination-induced sequence patterns highlights the potential of non-methylation and establishes unmethylated CGIs as targets of meiotic recombination in dogs. These observations describe recombination as an interesting process in genome evolution and provide further insights into the mechanisms of genomic variability.

*Keywords:* recombination, biased gene conversion, CpG island, copy number variation, substitutions, methylation

*Jonas Berglund, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.*

© Jonas Berglund 2014

ISSN 1651-6206

ISBN 978-91-554-9057-7

urn:nbn:se:uu:diva-233195 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-233195>)

*What we are is evolution's gift to us,  
what we become is our gift to evolution.*

Modified from Eleanor Powell

Cover: A Holliday junction before branch migration and resolution.

© 2006 Daniel D. Brown at Laughing Mantis Studio

Adapted, modified and extended.

# List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I **Berglund, J.**, Pollard, K.S., Webster, M.T. (2009) Hotspots of biased nucleotide substitutions in human genes. *PLoS Biology*, 7:e1000026. doi: 10.1371/journal.pbio.1000026.
- II Ratnakumar, A., Mousset, S., Glémin, S., **Berglund, J.**, Galtier, N., Duret, L., Webster, M.T. (2010) Detecting positive selection within genomes: the problem of biased gene conversion. *Philosophical Transactions of the Royal Society B Biological Sciences*, 365:2571-2580. doi:10.1098/rstb.2010.0007.
- III **Berglund, J.**, Nevalainen, E.M., Molin, A-M., Perloski, M., The LUPA Consortium, André, C., Zody, M.C., Sharpe, T., Hitte, C., Lindblad-Toh, K., Lohi, H., Webster, M.T. (2012) Novel origins of copy number variation in the dog genome. *Genome Biology*, 13:R73. doi: 10.1186/gb-2012-13-8-r73.
- IV **Berglund, J.**, Quílez, J., Arndt, P.F., Webster, M.T. (2014) Germline methylation patterns determine the distribution of recombination events in the dog genome. *Submitted.*

Reprints were made with permission from the respective publishers.

Additional co-authored publications completed during PhD research (not included in the thesis)

- I Lamichhaney, S.\* , Martínez Barrio, Á.\* , Rafati, N.\* , Sundström, G.\* , Rubin, C.-J., Gilbert, E.R., **Berglund, J.**, Wetterbom, A., Laikre, L., Webster, M.T., Grabherr, M., Ryman, N., Andersson, L. (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *PNAS*, 109:19345-19350. doi: 10.1073/pnas.1216128109.
- II Molin, A.-M., **Berglund, J.**, Webster, M.T., Lindblad-Toh, K. (2014) Genome-wide copy number variant discovery in dogs using the CanineHD genotyping array. *BMC Genomics*, 15:210. doi: 10.1186/1471-2164-15-210.
- III Ramírez, O., Olalde, I., **Berglund, J.**, Lorente-Galdos, B., Hernández-Rodríguez, J., Quílez, J., Webster, M.T., Wayne, R.K., Lalueza-Fox, C., Vilá, C., Marques-Bonet, T. (2014) Analysis of structural diversity in wolf-like canids reveals post-domestication variants. *BMC Genomics*, 15:465. doi: 10.1186/1471-2164-15-465.
- IV Lamichhaney, S.\* , **Berglund, J.\*** , Sällman Almén, M., Maqbool, K., Grabherr, M., Martínez Barrio, Á., Promerova, M., Rubin, C.-J., Wang, C., Zamani, N., Grant, B.R., Grant, P.R., Webster, M.T., Andersson, L. (2014) Evolution of Darwin's finches and their beaks revealed by whole genome sequencing. *Submitted*.

\* These authors contributed equally.

# Contents

Introduction .....	11
Evolutionary advantage of recombination .....	12
Characterization of recombination .....	13
Mechanism of meiotic recombination .....	13
Recombination hotspots .....	16
Control of recombination .....	17
Effects of recombination .....	18
Meiotic drive .....	18
GC-biased gene conversion .....	19
Recombination can generate CNVs .....	21
Dog as a model for recombination .....	23
The dog in genetics .....	23
Control of recombination .....	24
Targets of recombination .....	24
CNVs in dogs .....	25
Aims .....	26
Publications .....	27
PAPER I .....	27
Motivation .....	27
Results .....	27
Conclusion .....	28
PAPER II .....	28
Motivation .....	28
Results .....	29
Conclusion .....	30
PAPER III .....	30
Motivation .....	30
Results .....	31
Conclusion .....	31
PAPER IV .....	32
Motivation .....	32
Results .....	32
Conclusion .....	33

Concluding remarks and future perspectives .....34

Acknowledgements .....36

References .....38

# Abbreviations

aCGH	Array comparative genomic hybridization
ADCYAP1	Adenylate cyclase activating polypeptide 1
Bp	Basepair
CGI	CpG-island
CNV	Copy number variant
CO	Crossover
dN/dS	Non-synonymous-to-synonymous substitution rate
DNA	Deoxyribonucleic acid
DSB	Double-stranded break
DSBR	Double-stranded break repair
gBGC	GC-biased gene conversion
GC*	Equilibrium GC-content
HAR	Human accelerated region
Kb	Kilobases
L1	LINE1 – long interspersed nuclear element 1
LD	Linkage disequilibrium
NAHR	Non-allelic homologous recombination
NCO	Non-crossover
PAR	Pseudo-autosomal region
PRDM9	Positive-regulatory domain containing 9
PSG	Positively selected gene
S	Strong nucleotide (G and C)
SW	Strong-to-weak, GC-to-AT
SD	Segmental duplication
SDSA	Synthesis-dependent strand annealing
SNP	Single nucleotide polymorphism
ssDNA	Single-stranded DNA
W	Weak nucleotide (A and T)
WS	Weak-to-strong, AT-to-GC



# Introduction

Recombination is a vital biological process that forms the connections between homologous chromosomes during meiosis in sexual diploid eukaryotes in a process called synapsis. This has important implications for genomic integrity, evolution and disease (Handel & Schimenti 2010). The existence of recombination was first demonstrated through experiments with genetic crosses of maize, which demonstrated that genetically linked traits could be naturally uncoupled (McClintock 1950). Without recombination, genes on the same chromosome are passed through generations as a single linked block. Recombination unlinks the alleles of genes as homologous chromosomes separate during the first division of meiosis. This has important implications for efficient selection and genomic variability.

There are several important roles of recombination, but two essential functions can be distinguished: first, synapsis is required for correct segregation of chromosomes; second, recombination increase the genetic diversity within a population by reshuffling alleles into new combinations upon which natural selection can act (Coop & Myers 2007; Paigen & Petkov 2010).

Recombination rates vary at different scales - across genomes, between sexes and between taxa, and are in particular known to preferentially occur in specialized regions known as 'hotspots' (Kauppi et al. 2004). The intensity of recombination is largely heterogeneous and the peak rates of hotspots can vary many orders of magnitude (Jeffreys et al. 2001; McVean et al. 2004), and have large effects on genome evolution.

Beside crucial roles in meiosis, cell division and shuffling of alleles, evidence suggests that recombination can also affect genome evolution by a process called GC-biased gene conversion (gBGC). This process can be seen as an indirect consequence of recombination and may play a significant role in genome evolution (Galtier & Duret 2007; Webster & Hurst 2012). gBGC is a non-adaptive neutral recombination-associated process that favors the fixation of G and C nucleotides (Galtier et al. 2001; Galtier & Duret 2007), with dynamics similar to positive selection (Nagylaki 1983). This indicates that recombination may also result in segregation distortion, which can lead to the proliferation of deleterious mutations.

Another GC-biased feature of vertebrate genomes is the existence of short regions with increased frequency of CpG dinucleotides termed 'CpG-islands' (CGIs) (Bird 1980). These regions to a large extent lack DNA methylation, which suppresses a natural mutagenic tendency of methylated CpG

dinucleotides and enables a higher CpG-content. Recombination have been observed to target CGIs in several mammalian species (Brick et al. 2012; Auton et al. 2012, 2013).

Recombination requires a degree of sequence identity and is commonly mediated by homology between alleles of the same locus resulting in homologous recombination. Recombination that relies solely on sequence identity can occur between non-homologous alleles and is called ectopic exchange or non-allelic homologous recombination (NAHR). NAHR can result in large-scale structural variants of deletions and duplications known as ‘copy number variants’ (CNVs). CNVs have a major impact on genome evolution, and have been implicated in several human diseases (Hurles et al. 2008).

The work in this thesis focuses on the effects of recombination driving genome evolution in humans and dogs via gBGC and generation of CNVs and investigates the interaction of recombination with CGIs.

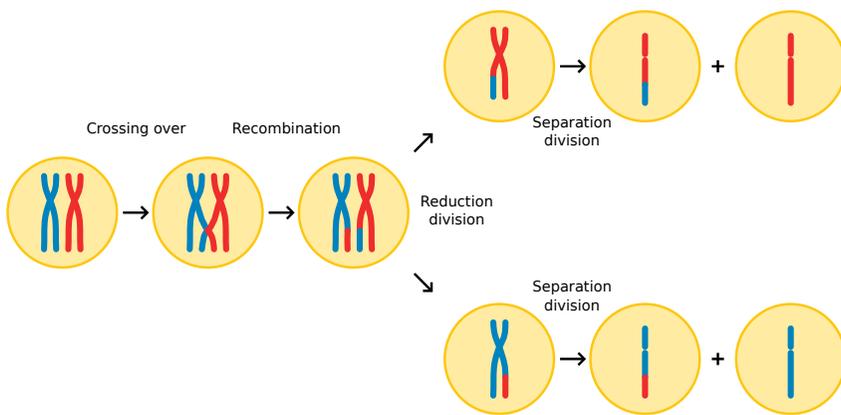
## Evolutionary advantage of recombination

The key to the prevalence of recombination can lie within another controversial question in evolutionary biology, which is the existence of sex. In asexual populations all individuals can produce offspring, while in sexual populations only one sex is capable to produce offspring. This reduces the population representation for each generation and is referred to as the ‘two-fold cost of sex’ (Smith 1978). Two generally accepted theories explain the controversial maintenance of recombination and sex: 1) recombination breaks down linkage and reduces selection interference between mutations, which enables selection locally with subsequent fixation of single advantageous mutations and 2) recombination increases diversity by generating new haplotypes by shuffling genotypes, which reduces infection susceptibility and repels disease (Hartfield & Keightley 2012). Thus there seem to be an intricate interplay between selection and recombination, where fitness variation interacts with selection for recombination. It is interesting to note however that in most higher eukaryotes recombination rates are close to one event per chromosome arm, which is the minimum required for correct disjunction of chromosomes during meiosis (Webster & Hurst 2012) (Fig. 1). This suggests that there is not strong selective pressure on the number of recombination events due to their evolutionary advantage in higher eukaryotes.

# Characterization of recombination

## Mechanism of meiotic recombination

The process of recombination involves multiple controlled events including alignment of DNA strands, precise breakage of each strand, genetic exchange between the strands, and ligation of the recombined molecules. This occurs at high frequency throughout the genome and requires a high degree of accuracy to avoid aberrant segregation. One major role of recombination is to aid correct segregation of chromosomes during cell division, and thus recombination occurs in all cells. However, recombination that defines genomic patterns of inheritance occurs in the germline during meiosis (Fig. 1).



*Figure 1.* Recombination ensuring correct segregation of chromosomes during meiosis, which is one of the major roles of recombination.

Meiosis begins by replication of chromosomes followed by the synapsis process to form pairs of chromatids, known as bivalents, connected by their centromeres. Recombination then begins by alignment of homologous chromosomes, known as tetrads, followed by a DNA double stranded break (DSB) introduced into one of the four chromatids, leaving two free ends. The ends are converted to ssDNA by resection of the 5' strands to leave 3' overhangs. One of the ends crosses over and invades the homologous region of a non-sister chromatid in a structure called displacement loop (D-loop) reminiscent of its visual appearance. The intersecting homologous strands form a structure known as 'Holliday junction' from the pioneering model of the recombination process proposed by Robin Holliday in 1964 (Holliday 1964). The invading DSB end is subsequently repaired by DNA synthesis during D-loop extension where the opposite chromatid acts as a template. (Fig. 2).

Recombination can then continue in either of two distinct pathways: I) synthesis-dependent strand annealing (SDSA) or II) double-stranded break repair (DSBR), which are finally resolved as either non-crossovers (NCOs) or crossovers (COs). NCO events result in patch-recombinants of chromosomes with exchange of a short homologous sequence without the exchange of flanking genetic material, while CO events result in splice-recombinants of chromosomes with exchange of flanking chromosomal sequences. This causes COs to have a larger impact on variation than NCOs by breaking up chromosomes in haplotype blocks that are useful features to aid genetic mapping. (Haber et al. 2004; Paigen & Petkov 2010)

- I The SDSA pathway generates around 90% of all recombination events and yields predominantly NCOs. In the SDSA model strand invasion occurs only on one side of the DSB as described above, and the D-loop is then extended by DNA synthesis. The newly synthesized DNA strand is displaced and anneals back to the other DSB end, where DNA synthesis is followed by ligation of nicks. SDSA leaves the homologous sister chromatid untouched in the recombination process.
- II The DSBR pathway generates around 10% of all recombination events and is predominantly resolved as COs. In the DSBR model the second end of the DSB is captured by the second non-sister chromatid of the D-loop, and both DSB ends are subsequently repaired by DNA synthesis during D-loop extension. The reciprocal invasion and ligation of nicks results in a crossover structure with two intersecting D-loops that forms double Holliday junctions. This double D can be either dissolved or resolved.
  - Dissolution generates NCO events identical to the SDSA pathway, by displacement of the sealed D-loops that anneals back to the sister chromatid.
  - Resolution is the main outcome of the DSBR pathway, and generates predominantly COs. One of the Holliday junctions is cleaved and the sealed strands anneal back to the sister chromatid, while the other Holliday junction undergoes branch migration along the DNA before it is resolved as a CO or NCO event.

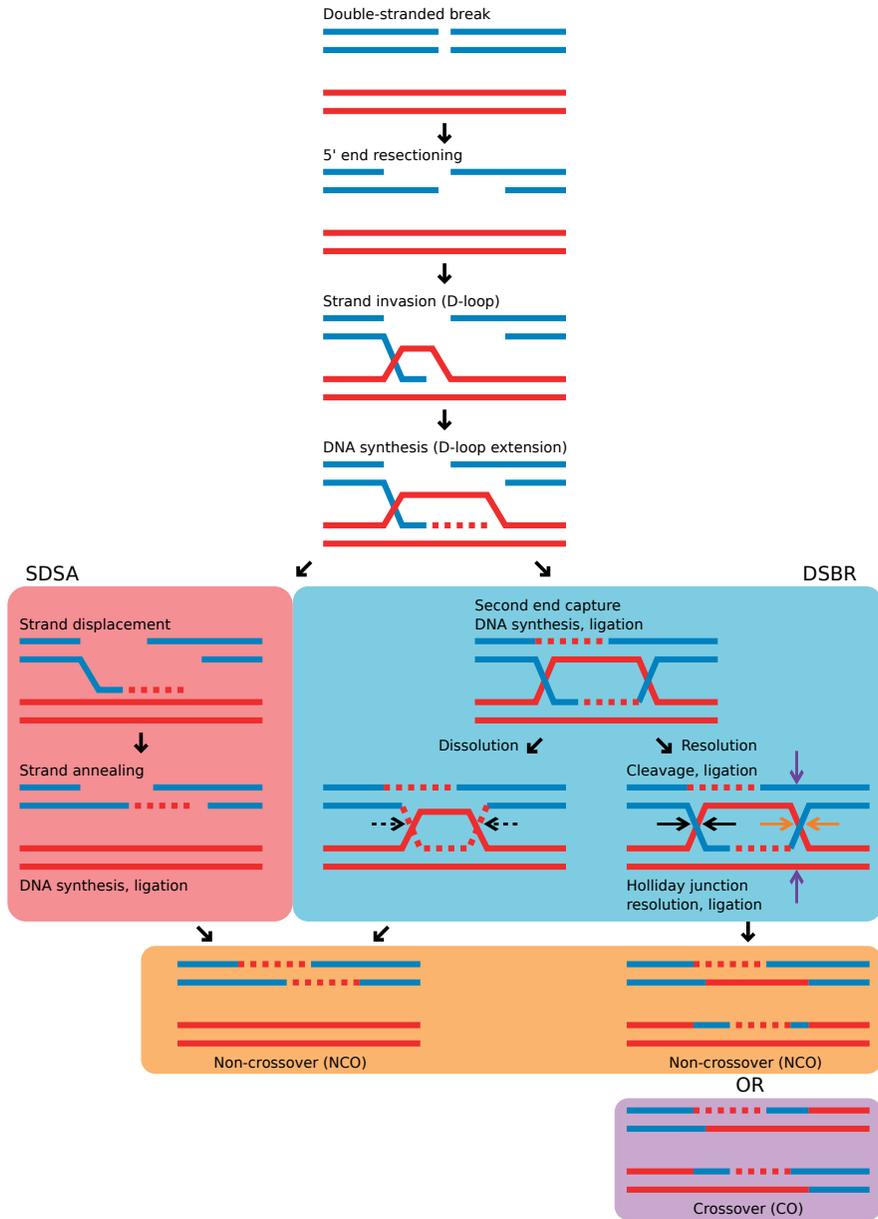
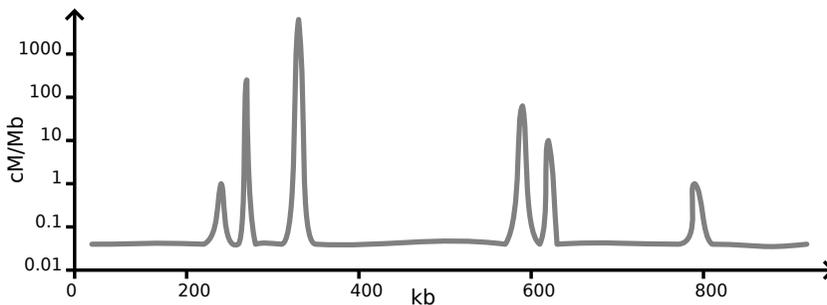


Figure 2. Mechanisms in the two pathways of meiotic recombination and their products.

## Recombination hotspots

Studies of human genetic variation revealed the existence of haplotype blocks, consistent with highly punctuate COs (Gabriel et al. 2002). This suggests that recombination events are not spaced uniformly across the genome, but occur at preferred sites for DSB introduction. COs have been further analyzed and tend to cluster in distinct regions 1-2 kilobases (kb) in length known as ‘hotspots’ of recombination (Jeffreys et al. 2001; McVean et al. 2004), with surrounding regions essentially devoid of recombination activity (Fig. 3).



*Figure 3.* Schematic representation of typical hotspot activity distribution.

Hotspots can be analyzed in mammals using pedigrees, sperm typing or by the analysis of linkage disequilibrium (LD) (Paigen & Petkov 2010). Pedigree studies can map actual recombination events on a broad scale, but are limited to the small number of recombination events that occur during each generation. Sperm typing is restricted to assay rates in single known hotspot locations in a germline, and provides exclusively male-specific rates. Coalescent modeling of LD utilizing population genetics data is a method to map hotspots at fine-scale and to quantify recombination rates. This approach does not allow for a direct count of recombination events, but for the reconstruction of the evolutionary process generating LD along the genome.

A genome-wide map of the locations of human recombination hotspots has been created in the HapMap project (Myers et al. 2005; Frazer et al. 2007). More than 30,000 hotspots have been identified in the human genome, spaced on average 50-100 kb apart. The peak recombination rate can span many orders of magnitude, and potentially have major effects on genome evolution. Although regional recombination rates do not vary to a large extent, hotspots are known to be transient (Winckler et al. 2005), as hotspot locations in the human genome do not match those in chimpanzee (Myers et al. 2005). This limits the local impact of hotspots, while maintaining their regional effect.

## Control of recombination

The spatial distribution of hotspots argues for specific targets for induction of the recombination process. In search of underlying sequence features related to hotspots, recent studies of human hotspots have revealed an enrichment of a degenerate 13-bp sequence motif CCNCCNTNNCCNC (Myers et al. 2008), which is estimated to be associated with more than 40% of hotspots genome-wide. The discovery of this recombination-attracting motif originates from two separate routes of research mapping genome-wide hotspot activity. Allelic variation at SNPs within this motif (Jeffreys & Neumann 2002) and allelic variation within the human PRDM9 gene (Baudat et al. 2010) were both correlated with genome-wide hotspot activity. The association between PRDM9 and the motif proposes an important role of the gene in determining genome-wide locations of hotspots.

PRDM9 is a human zinc-finger protein recently identified and subsequently computationally predicted to bind to the hotspot motif (Fig. 4) and cause a histone modification that acts as a marker for the initiation of recombination (Myers et al. 2010). Comparison of the PRDM9 gene in several mammals indicates rapid evolution due to positive selection, which suggests rapid turnover of the motif recognition sites (Oliver et al. 2009). Comparison of human and chimpanzee orthologs of PRDM9 supports this, since they are predicted to bind completely different sequence motifs (Myers et al. 2010). Furthermore, the chimpanzee interspecific allele variation at PRDM9 is high, and no specific sequence motif is enriched in chimpanzee hotspots (Auton et al. 2012).

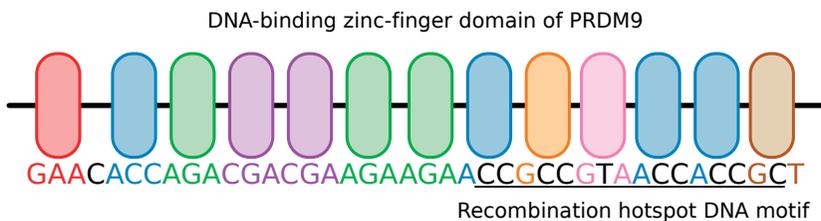


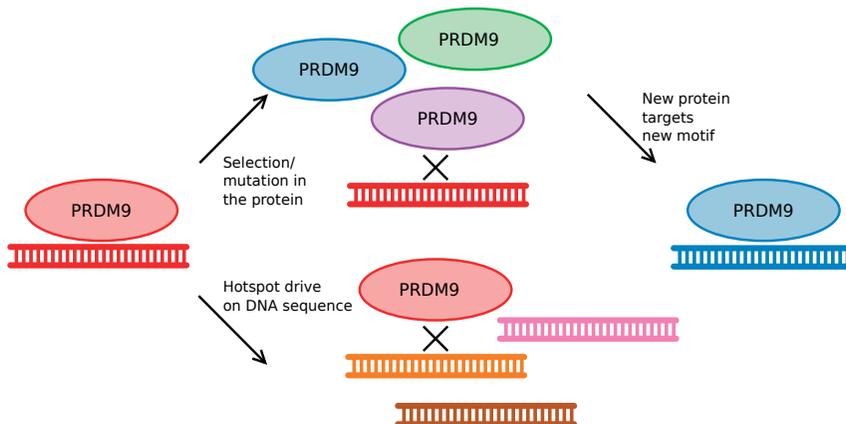
Figure 4. PRDM9 binding to the hotspot motif to begin initiation of recombination.

# Effects of recombination

## Meiotic drive

The rapid turnover of the hotspot motif recognition site is a result of a force that over-transmits certain alleles in gametes referred to as 'meiotic drive'. Meiotic drive has been experimentally observed in hotspots, in which transmission of the non-recombinogenic allele is favored over the recombinogenic allele, referred to here as 'hotspot drive' (Jeffreys & Neumann 2002). This process exists because the DSB during recombination occurs at the allele that attracts recombination, and is repaired by synthesizing DNA with the less recombinogenic allele as template. Thus, hotspot drive predicts that hotspots will gradually disappear from the genome, as the active form of the motif is converted into the inactive form. The observation that hotspots are, in fact, common is known as the 'hotspot conversion paradox' (Boulton et al. 1997).

The rapid evolution of the PRDM9 gene provides a simple yet elegant solution to the hotspot conversion paradox: hotspots do not disappear from the genome; instead their locations regularly shift during evolution due to alterations in the hotspot motif recognized by PRDM9 (Myers et al. 2010) (Fig. 5). In addition, it is suggested that PRDM9 is an important speciation gene in mammals, capable of producing genomic incompatibilities between incipient species (Oliver et al. 2009; Mihola et al. 2009). These findings suggest that the recombination landscape is dynamic and that hotspot drive is an important force in genome evolution.



*Figure 5.* Meiotic drive under-transmitting the hotspot motif and rapid evolution of the PRDM9 DNA-binding domain. This results in recognition of new motifs and explains the dynamics of recombination hotspots.

## GC-biased gene conversion

A related form of meiotic drive that occurs in recombination hotspots is known as ‘GC-biased gene conversion’ (gBGC), which refers to the biased transmission of G and C (S, strong) over A and T (W, weak) alleles (Mancera et al. 2008; Duret & Galtier 2009). The gBGC model predicts an increased fixation of G:C alleles in regions of high recombination. It occurs when homologous chromosomes heterozygous for A:T and G:C aligns during meiosis. Any aberrant base-pairing in the heteroduplex tract formed around the DSB during recombination is preferentially repaired to G:C over A:T, which increases the number of G and C containing alleles (Fig. 6). This unidirectional transfer of genetic material leads to an increased probability of fixation of G:C (Webster et al. 2003; Webster & Smith 2004). There are several lines of evidence that support this prediction:

1. GC-content correlates with recombination rate in a wide range of eukaryotes (Fullerton et al. 2001; Kong et al. 2002; Birdsell 2002; Duret & Galtier 2009), most likely because recombination drives the evolution of GC-content by biasing the substitution pattern towards WS mutations (Meunier & Duret 2004; Webster et al. 2005; Dreszer et al. 2007; Duret & Arndt 2008).
2. Both yeast and primate cell lines show evidence of biased base mismatch repair towards incorporation of G and C nucleotides (Brown & Jiricny 1989; Mancera et al. 2008).
3. Analyses of human SNP frequency distributions together with human diversity and human-chimpanzee divergence, reveals evidence for a fixation bias of S nucleotides, which is stronger in regions of elevated recombination (Duret et al. 2002; Webster et al. 2003; Webster & Smith 2004; Duret & Arndt 2008; Katzman et al. 2011).
4. Regions of elevated recombination are GC-rich (Spencer 2006) and contain clusters of GC-biased substitutions (Dreszer et al. 2007; Galtier & Duret 2007).
5. Sequences with high rates of gene conversion have elevated GC-content (Galtier 2003; Montoya-Burgos et al. 2003).
6. Elevated WS substitution rates show a much stronger association with male than female recombination rate (Webster et al. 2005; Dreszer et al. 2007; Duret & Arndt 2008); a correlation unlikely to result from selection, which would not predict a sex-specific pattern.
7. Direct evidence from analyses of NCO gene conversion events in human pedigrees show a strong allelic bias of 70% in favor of transmitting G:C alleles (Williams et al. 2014).

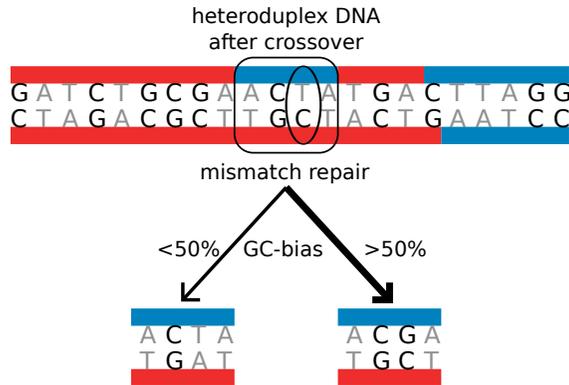


Figure 6. GC-biased gene conversion (gBGC). The biased transmission of G:C over A:T alleles during base excision repair in heteroduplex DNA.

### **gBGC can affect functional evolution**

Investigating the forces that drive genome evolution is important to understand species adaptations, and is commonly performed with genomic scans to identify regions under positive selection. A strategy for identifying the genes and genomic regions under positive selection along a lineage is based on the assumption that positive natural selection has accelerated the evolution of certain species-specific traits that makes the species unique. However, processes like gBGC have properties that may affect coding sequences and drive protein evolution in the absence of natural selection, which could interfere with common interpretations in scans of positive selection.

An important approach to understand human-specific biology is to identify regions in our genome that show evidence of accelerated evolution but are highly conserved in other vertebrates. This is based on the assumptions that conserved regions under purifying selection across vertebrates are functional, and such functional regions with accelerated rates of sequence evolution along one branch evolve under positive selection. A phylogenetic investigation of sequences with high rates of evolution across a large range of eukaryotic species demonstrated that a GC-substitution bias is a general characteristic of highly diverged sequences (Capra & Pollard 2011). Both positive selection and gBGC can give rise to bursts of accelerated sequence evolution. While positive selection leads to accumulation of substitutions for increased fitness, gBGC promotes the fixation of WS substitutions regardless of fitness effects. Strong gBGC, which is known to induce specific mutational patterns, could therefore drive sequence evolution by overpowering the effects of negative selection and lead to the fixation of non-beneficial changes, and thereby represent a genomic 'Achilles' heel' (Galtier & Duret 2007).

### **gBGC affects human accelerated regions**

Several studies have performed genomic scans to identify regions of accelerated human evolution. From an alignment of 17 vertebrates, 49 conserved elements with strong evidence for human-specific acceleration were identified, termed 'human accelerated regions' (HARs) (Pollard et al. 2006). HARs are enriched near neurodevelopmental genes, and are mostly noncoding, but many are predicted to have an RNA secondary structure. The most accelerated regions (HAR1-2) are implicated in controlling human morphology and cognitive abilities (Pollard et al. 2006). Although positive selection is likely to play a role in the evolution of HARs, additional non-selective processes are also implicated.

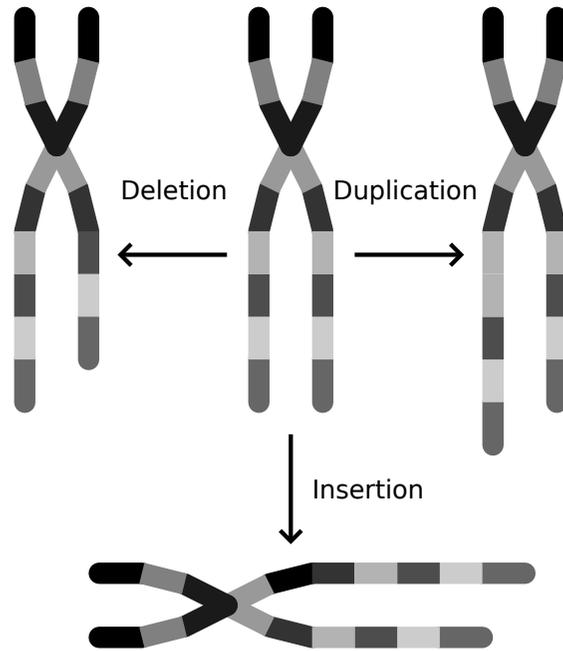
HARs show a significant enrichment for WS biased substitutions on the human lineage, which is not predicted under positive selection and differs from the genomic average of substitutions, which is predominated by SW substitutions. The top 49 HARs have twice as many WS substitutions than the reversed rate. The top 5 HARs show even more bias with 35 WS vs. 3 SW substitutions, and the top three HARs have only WS substitutions. The extremely GC-biased substitution pattern that often also extends into the flanking region, which is not inferred to be functional, supports the gBGC model of evolution for HARs. Another observation consistent with the gBGC hypothesis is that HARs tend to occur near human recombination hotspots and in subtelomeric and other regions characterized by high recombination rates.

The relative effects of selection and gBGC on the evolution of HARs have been determined to be 76% and 19% respectively, where gBGC is estimated to be strongest in the most accelerated HARs (Kostka et al. 2012). The processes of selection and gBGC may interact, but in some situations gBGC can override selection and cause the fixation of deleterious mutations (Galtier & Duret 2007; Galtier et al. 2009). These studies of HARs provide strong support for an important role of gBGC in accelerated evolution, in contrast to the common interpretation that positive selection drives the evolution in all of these regions.

### **Recombination can generate CNVs**

As mentioned above, searches for enrichment of sequence motifs in human hotspots have identified an enrichment of a common 13-bp degenerate motif CCNCCNTNNCCNC with direct evidence for a role in hotspot activity (Myers et al. 2008). This hotspot motif is recognized by PRDM9 that initiates the recombination machinery (Myers et al. 2010). Based on the implication of the 13-mer motif in allelic crossover activity during meiosis, it was also investigated for roles in other forms of recombination-associated genome rearrangements. The motif is indeed enriched in breakpoints of struc-

tural sequence variants like segmental duplications (SDs) (Myers et al. 2008; Mills et al. 2011), which implicates DSBs formed in this way during deletion, duplication and insertion events referred to as ‘copy number variants’ (CNVs) (Fig. 7).



*Figure 7.* Types of copy number variation polymorphisms.

Generation of large-scale structural genome rearrangements such as CNVs accounts for a large proportion of genetic polymorphism and have important roles in genome evolution and disease. SDs and CNVs have been mapped in several species, including human (Iafrate et al. 2004; Sharp et al. 2005; Tuzun et al. 2005; Conrad et al. 2006; Goidts et al. 2006; Redon et al. 2006; McCarroll et al. 2006; Perry et al. 2006, 2008), chimpanzee (Perry et al. 2006), mouse (Graubert et al. 2007; She et al. 2008) and rat (Guryev et al. 2008). A common approach has been the array comparative genomic hybridization (aCGH) technique, with probes spaced uniformly across target regions, which can be complete genomes. With this technique test and reference samples compete to hybridize to the array, and copy numbers in test samples are evaluated relative to the reference sample. The advent of next generation sequencing has further improved theoretical resolution.

A variety of molecular mechanisms are implicated in CNV formation (Hastings et al. 2009), with non-allelic homologous recombination (NAHR) as a major source of structural variation in regions of extended homology. In contrast to common recombination between homologous alleles, NAHR occurs between genomic regions with enough sequence identity, regardless of allelic proximity. CNVs associated with NAHR tend to be clustered in the genome, and CNVs are enriched in the vicinity of segmental duplications (Mills et al. 2011). This suggests that regions of local sequence homology are hotspots of CNV formation by the process of NAHR (Sharp et al. 2005; Graubert et al. 2007; She et al. 2008).

Significant association between CNVs and elevated recombination rates in birds supports a major role of recombination-associated processes in genome evolution (Völker et al. 2010). In humans, the frequency distribution of CNVs shows signals of purifying selection, suggesting that a significant proportion of CNVs have harmful effects (Conrad et al. 2006). While CNVs are associated with a number of genetic disorders (Hurles et al. 2008), there is also evidence for beneficial roles of CNVs, such as adaptive variation in copy number of the amylase gene in response to diet (Perry et al. 2007), and variation in susceptibility to malaria (Flint et al. 1986) and HIV/AIDS (Gonzalez et al. 2005).

## Dog as a model for recombination

### The dog in genetics

In contrast to natural populations dogs have a unique breeding history, which in many ways enhances the possibilities to explore genetic variation. Several features make the dog an ideal genetic model for studying the genetic basis of phenotypic variation, behavioral traits and disease. Dogs were domesticated thousands of years ago to support human needs, and have since accompanied humans in a close relationship. Due to intense selection for mainly appearance and behavior it is one of the most phenotypically diverse mammals. They do not only share living space with humans, but also food resources and to large extent disease prevalence with similar manifestation. The demographic history of dogs with periodic population bottlenecks during domestication and breed creation has created a unique genomic architecture with large haplotype blocks. This genetic architecture provides an unparalleled opportunity to map genetic traits. (Lindblad-Toh et al. 2005)

## Control of recombination

In many mammals initiation of recombination starts with the action of PRDM9, which recognizes a specific hotspot motif and trimethylates histone H3K4 (Myers et al. 2010). However, sequencing of PRDM9 orthologs in several dog breeds and carnivores demonstrates an inactivation of PRDM9 early in canid evolution (Oliver et al. 2009; Muñoz-Fuentes et al. 2011; Axelsson et al. 2012), suggesting recombination is controlled differently in dogs. However, fine-scale recombination maps have shown that dogs do have hotspots (Axelsson et al. 2012; Auton et al. 2013). In lack of the candidate gene responsible for recognition of the hotspot motif, dog hotspots are found on a similar scale in regions characterized by GC-richness (Axelsson et al. 2012), particularly CpG-richness (Auton et al. 2013), rather than by PRDM9 binding motifs.

## Targets of recombination

A general property of vertebrate genomes is that they are particularly deficient in CpG dinucleotides due to a specific mutagenic property; methylated cytosines in a CpG-context have a tendency of spontaneous deamination of C to T (Bird 1980). Since DNA methylation is common, this effect decreases the CpG-frequency genome-wide. However, vertebrate genomes display a fine-scale compositional non-uniformity in CpG-content, and regions of hundreds of bp of locally increased CpG-content have been observed, so called 'CpG-islands' (CGIs) (Bird 1980). CGIs can maintain a higher CpG-content due to reduced SW mutation rate, which is achieved via lack of DNA methylation, which suppresses the natural deamination tendency.

Despite absence of PRDM9, dog recombination rates are highly elevated around trimethylation marks. This association is explained by the presence of CGIs with elevated recombination rates intersecting the marks (Auton et al. 2013). A correlation between recombination and CGIs is supported by analyses in other species. Chimpanzees also lack a common PRDM9 binding motif and show elevated recombination rates around CGIs (Auton et al. 2012). Recombination in PRDM9 knock-out mice is initiated at functional genomic elements like enhancers and promoters, which are rich in CGIs (Brick et al. 2012). Investigations of CGI density in several species show how dog clearly deviates from other mammals and have higher levels of CGI density comparable to birds and fishes (Han et al. 2008), which also lack an active copy of PRDM9 (Oliver et al. 2009).

## CNVs in dogs

Recombination has been associated with genomic instability and CNV formation in human (Myers et al. 2008; Mills et al. 2011). Interestingly, genomic rearrangements during canid genome evolution have involved GC-rich regions like those found in dog recombination hotspots (Webber & Ponting 2005; Axelsson et al. 2012). This suggests that recombination can generate structural variation in the dog genome. The advantages of dog as a model for unraveling genetic disorders have propelled the mapping of phenotypic traits in the dog genome. Traits are mostly apportioned into distinct breeds. Strong artificial selection behind the unique collections of characteristics makes investigating them important for uncovering the genetic basis of phenotypic variation in dogs. Several traits have been mapped in distinct breeds (Karlsson et al. 2007; Karlsson & Lindblad-Toh 2008; Parker et al. 2010; Sutter et al. 2007) and structural variation is implicated in a number of these traits (Olsson et al. 2011; Parker et al. 2009; Salmon Hillbertz et al. 2007), which mirrors the situation in human (Hastings et al. 2009; Hurles et al. 2008). Thus, in dog much phenotypic variation is likely attributable to CNVs and recombination is important for generating CNVs, which has major effects on phenotypic variation.

# Aims

gBGC have an established importance in genome evolution, where it has major effects on non-coding regions (Pollard et al. 2006). Interestingly, gBGC acts similarly but independently to selection, which raises the possibility that gBGC could also affect functional sequences. Therefore, as a complement to general whole-genome scans, paper I focuses on the effect of gBGC on rapidly evolving coding regions in the human genome. We specifically investigate if the often-overlooked process of gBGC can affect the divergence of genes.

Functional regions are regularly interrogated for genes evolving under positive selection. If gBGC can also affect coding regions, this opens up for an interesting and unexplored interaction between gBGC and positive selection. Paper II expands on the previous work by further analyzing and quantifying the effect gBGC has on the evolution of genes and inferences of positive selection. This is done by implementing methods pioneered in paper I.

Another interesting question regarding the control and consequences of recombination in dogs relates to the lack of PRDM9 that is normally responsible for determining locations of DSBs. We ask which regions of the genome recombination targets without PRDM9. DSBs are required for structural changes, and paper III extends the research of recombination effects on genome evolution by undertaking the most comprehensive analysis of structural variation in dogs to date. We develop a new method for quantifying CNVs from aCGH data and deploy it to identify CNVs and analyze their distribution, characteristics and mechanisms of formation.

It has previously been shown that hotspots in dog match CGIs (Auton et al. 2013), which suggests that these could be targets of recombination when PRDM9 is not available. In paper IV we investigate if they are actually unmethylated or whether they are just the result of gBGC. We ask whether recombination without PRDM9 moves to unmethylated regions, or if gBGC control substitution patterns and increase the CpG-content and is responsible for generating the signal in the genome detected as CGIs.

# Publications

## PAPER I

*Hotspots of biased nucleotide substitutions in human genes.*

### Motivation

Several non-coding elements have been identified that show high levels of conservation in mammals but accelerated evolution along the human lineage (Pollard et al. 2006; Prabhakar et al. 2006; Kim & Pritchard 2007; Bird et al. 2007); ‘human accelerated regions’ (HARs). These elements are candidates for human-specific adaptation and thus potential targets of positive selection. Interestingly, gBGC is a recombination-associated molecular drive that favors the fixation of GC mutations and has population dynamics similar to positive selection. Efforts to determine the underlying evolutionary forces behind the acceleration hold gBGC as the most likely cause, and suggest that recombination hotspots can contribute to accelerated evolution via lineage-specific bursts of biased substitutions.

Some protein-coding sequences, like duplicated gene families and the pseudo-autosomal region (PAR), also show patterns of evolution consistent with gBGC (Montoya-Burgos et al. 2003; Galtier & Duret 2007). The Fxy gene in house mouse (*M. musculus*) partly resides within the highly recombining PAR region of the X-chromosome where it has accumulated 28 amino acid replacement substitutions, which all are WS. This provides evidence that gBGC can offset the effects of purifying selection and drive protein evolution by increasing ratios of non-synonymous to synonymous rates of base substitution (dN/dS) or by fixation of weakly deleterious mutations (Galtier & Duret 2007; Duret & Galtier 2009; Galtier et al. 2009) and ultimately lead to false inferences in tests of positive selection (Yang 1998; McDonald & Kreitman 1991; Goldman & Yang 1994).

### Results

We first identified exons with evidence of locally increased rate of nucleotide substitution on the human lineage to look for accelerated evolution. Sequences were obtained from a dataset of nearly 85,000 exons from 1:1:1 human-chimpanzee-macaque orthologous genes (Gibbs et al. 2007). The

most accelerated exons displayed a biased pattern of base substitutions, which extended into flanking regions. Both GC-content and recombination rates were also increased in these regions. Thus, clusters of nucleotide substitutions in single exons of genes tend to be biased. A comparison of mutations and substitutions revealed discordant patterns of nucleotide divergence and polymorphism in accelerated exons, suggesting a fixation bias consistent with gBGC.

We also noted an excess of non-synonymous substitutions in accelerated exons. The most divergent exon in genes with elevated dN/dS ratios (typically interpreted in terms of positive selection) also exhibit a substitution bias and are enriched near hotspots. The same association between non-synonymous changes and substitution bias was seen in genes identified in McDonald-Kreitman-tests. We implemented and deployed a theoretical framework to predict the effect of a bias towards fixation of WS mutations, modeled as a selective coefficient, on coding sequences under purifying selection. The model confirmed that gBGC could generate an increased dN/dS ratio that under strong effects with high coefficients could reach significant results above one, which serves as a rough indication of positive selection, and thus mimic positive selection.

## Conclusion

Whole-genome scans of accelerated evolution highlighted the effect of recombination on patterns of genome evolution in coding sequences, which mirrors that observed in mostly non-coding HARs. WS biased substitution patterns in accelerated regions are not only confined to non-coding sequences but can also affect coding sequences. The bias extends into flanking regions and display high levels of recombination as well as GC-content. Clusters of substitutions with these characteristics are evidence of a strong effect of gBGC operating on the most rapidly evolving coding sequences in our genome. We show that this process can also lead to increased fixation of replacement amino acid changes driving protein evolution, and may have a biased influence on tests of positive selection.

## PAPER II

*Detecting positive selection within genomes: the problem of biased gene conversion.*

### Motivation

A major goal in evolutionary genetics is to identify loci under positive selection, which are predicted to contribute to species-specific adaptation. A

common approach is to identify genes with accelerated rates of evolution along a particular branch of a phylogeny. However, as shown in paper I, several molecular mechanisms (Hurst 2009), like recombination-associated gBGC, affect fixation probabilities of WS mutations by biasing the transmission probability of G and C alleles in both coding and non-coding regions in many taxa (Duret & Galtier 2009; Galtier & Duret 2007; Galtier et al. 2009; Mancera et al. 2008). And indeed, a substantial fraction of instances of human-specific acceleration of functional non-coding regions were previously attributed to positive selection (Pollard et al. 2006; Prabhakar et al. 2006; Bird et al. 2007; Kim & Pritchard 2007), but have subsequently been associated with gBGC (Galtier & Duret 2007; Galtier et al. 2009). Tests of positive selection commonly use the ratio of non-synonymous (dN) to synonymous (dS) substitution rates (Yang 1998), which are expected to be more robust to gBGC than simple acceleration tests of average substitution rates. However, as shown in paper I, gBGC can also lead to an increase of the dN/dS ratio and mimic the effect of adaptive evolution (Galtier et al. 2009). Positive selection and gBGC can be modeled in the same way, but distinguished by their properties. gBGC is associated with regions of high recombination rate and generates WS biased substitution patterns that extends into flanking regions, while these properties are not expected under positive selection that operates exclusively on functional sites. By applying these criteria it should be possible to quantify the proportion of positively selected genes (PSGs) that could be falsely assigned and are indeed products of gBGC.

## Results

A scan of primate orthologous genes (Kosiol et al. 2008) identified genes with branch-specific elevated substitution rates along the human and chimpanzee lineages. In general these genes exhibit elevated recombination rates, increased equilibrium GC-content (GC\*) and are enriched near hotspots and subtelomeric regions, which is consistent with gBGC generating accelerated evolution in a subset of these genes.

A complementary scan identified genes with elevated dN/dS ratios, generally more robust to gBGC, along branches of the primate phylogeny. These genes partly subset those from the acceleration test and also show elevated recombination rates, elevated GC\* and are enriched near hotspots and terminal chromosome bands consistent with evolution via gBGC. To determine the proportion of these branch-specific PSGs that could potentially be due to gBGC, a maximum likelihood model was implemented. The model assigns genes to either of two classes with distinct GC\*; gBGC-free and gBGC-affected with low and high GC\* respectively. The gBGC-affected class was assigned a 20% excess of coding branch-PSG regions, indicating that around 20% of human branch-PSGs may be due to gBGC.

An extreme case among the human-PSGs is ADCYAP1 with dN/dS ratio above 2. All substitutions in ADCYAP1 are WS, a bias that extends into flanking regions. On top of that the gene is located within a recombination hotspot, providing excellent conditions for strong gBGC. Theoretical modeling of ADCYAP1 evolution was performed to test the prediction that gBGC can increase dN/dS ratios. Testing a wide range of parameters in the gBGC model predicts that several scenarios that invoke gBGC alone, absent of positive selection, are plausible explanations behind the evolution of ADCYAP1.

## Conclusion

Patterns of nucleotide evolution in coding sequences across the primate phylogeny show clear signatures of gBGC, both among accelerated sequences and genes with elevated dN/dS ratios. Theoretical modeling predicts 20% of PSGs to be generated under the influence of gBGC. The example of ADCYAP1 clearly illustrates the potential effect of gBGC on the evolution of coding sequences, and consequently questions the results of previous scans for positive selection in a variety of species.

## PAPER III

*Novel origins of copy number variation in the dog genome.*

### Motivation

CNVs involve deletions, duplications and insertions of DNA segments up to several megabases in length, and account for a significant proportion of genomic variation (Hurles et al. 2008). In humans, CNVs have been associated with both beneficial and detrimental phenotypic traits (Hurles et al. 2008). Human CNVs are enriched in segmental duplications (SDs) and breakpoints are enriched for the recombination targeted PRDM9 binding motif (Mills et al. 2011), which implicates DSBs formed in this way during CNV creation via NAHR. However, meiotic DSB formation seems to be differently controlled in dog (Oliver et al. 2009; Muñoz-Fuentes et al. 2011). Dogs lack an active copy of PRDM9, and hotspots are instead enriched for peaks of locally elevated GC-content that may be targets of DSBs (Axelsson et al. 2012). This is supported by the involvement of GC-rich regions in genome rearrangements during canid evolution (Webber & Ponting 2005). This ultimately indicates that GC-peaks are important targets of recombination and often involved in rearrangements. Three previous studies have mapped CNVs in dogs (Chen et al. 2009; Nicholas et al. 2009, 2011), and confirmed an enrichment in regions of segmental duplication. Undertaking the most compre-

hensive CNV discovery effort in dogs to date, together with investigating a large range of breeds, we examine the genome-wide distribution and characterization of CNVs and their breakpoints at unprecedented resolution.

## Results

We confirm that the applied method largely agrees with previous efforts to identify CNVs. We show that CNVs are in general biased away from genes, but surprisingly genes involved in olfactory transduction are enriched in duplications, possibly reflecting positive selection for dog-specific traits.

We confirmed an enrichment of CNVs in SDs, where they occur at higher allelic frequencies, are more complex, tend to be longer and are more likely to overlap genes (Nicholas et al. 2011). Scans for sequence features identified an overrepresentation of L1 repeats, GC-peaks and stretches of perfect homology in CNV breakpoints. The enrichments of L1 repeats is most pronounced for younger elements, and mirror the pattern seen between human CNVs (Kim & Pritchard 2007; Korbel et al. 2007) and SDs (Redon et al. 2006), where the likelihood of a SD being associated with a CNV was highly correlated to its sequence identity to the duplicated copy. The enrichment of GC-peaks is more than two-fold and rapidly decays with increasing distance from the breakpoint. The length of stretches of perfect homology is four times longer than expected, and in extreme cases exceeds 1 kb. The association between CNV breakpoints and SDs, L1 repeats, GC-peaks and extended regions of perfect homology in dogs suggest that CNV breakpoints may often occur at recombination hotspots, and these features promote structural variation by NAHR.

Analysis of the distribution of CNVs among breeds reveals a frequency distribution qualitatively similar to those expected for neutral polymorphisms. The majority of CNVs were found in multiple breeds, with on average only one private CNV per breed. The relative lack of breed-specific CNVs suggests that instances of those involved in breed-specific phenotypes must be rare. The extent to which patterns of CNV variation can be used to infer population structure between breeds was explored by constructing a phylogeny based on CNV allele sharing. Similar to large-scale SNP analyses (vonHoldt et al. 2010; Vaysse et al. 2011), this indicated that the CNVs have a strong phylogenetic signal, grouping dogs into breeds and to some extent breed type.

## Conclusion

The results from this study highlight a general enrichment of CNVs in regions of SDs. They also suggest that many CNVs are generated by NAHR events directed towards GC-peaks, which are also enriched in dog recombination hotspots. In support of a strong role of NAHR in dog CNV formation,

we also identify associations between CNV breakpoints and L1 elements and long stretches of sequence homology. This comprehensive catalogue of CNVs will be useful for future studies to uncover the genetic basis of complex traits in dogs.

## PAPER IV

*Germline methylation patterns determine the distribution of recombination events in the dog genome.*

### Motivation

Recombination events in eukaryotes are localized to narrow regions termed recombination hotspots (Petes 2001; Paigen & Petkov 2010), with short evolutionary lifespan (Winckler et al. 2005). In vertebrates these recombination hotspots are often enriched for binding motifs of the PRDM9 gene (Baudat et al. 2010; Parvanov et al. 2010; Cheung et al. 2010; Myers et al. 2010). Thus, PRDM9 is inferred to be responsible for initiating a majority of the recombination events. However, dogs and other canid taxa are some of the only mammals that lack an active copy of PRDM9 (Oliver et al. 2009; Muñoz-Fuentes et al. 2011; Axelsson et al. 2012). The dog genome is therefore an ideal natural model to study how recombination events are distributed in the absence of PRDM9 and the effects on genome evolution. Hotspots in dogs show an enrichment for GC-rich elements instead of specific sequence motifs (Axelsson et al. 2012; Auton et al. 2013). The challenge still remains to characterize these regions and determine if they are GC-rich because unmethylated sequences high in GC attract recombination, or if they become GC-rich because of the recombination-associated process gBGC, or if both these processes are involved simultaneously. Another related feature that distinguishes the dog genome from other mammals is an increased density of CGIs (Han et al. 2008). By analyzing CGIs and other GC-rich regions we can discover how recombination events affect patterns of nucleotide substitutions and shape the distribution of CGIs in the dog genome.

### Results

We first show that the GC-rich elements identified in dog hotspots and CNVs are a subset of the predicted CGIs. Employing the dense hotspot map from re-sequencing data in dogs (Auton et al. 2013) we confirm a strong association between CGIs and these hotspots. The association is strongest for specifically promoter-associated CGIs. Simultaneously, the recombination rates are increased in both hotspot- and promoter-associated CGIs. This suggests a strong connection between recombination and increased GC-richness.

The cause of association between CGIs and recombination was investigated by analyzing the mutation rates across the dog genome in general and in the strongly associated regions in particular. CpG mutation rates were reduced in hotspots, CGIs and promoters. This reduction was stronger in regions where these features intersect with each other. We also show that the reduced CGI mutation rate creates a WS over SW bias in hotspots. These results clearly show that CGIs are unmethylated and that this property of CGIs promotes recombination.

We next investigated substitution patterns along the dog lineage compared to the panda lineage with a substitution model that explicitly incorporates the CpG mutability (Arndt et al. 2003). Panda is the closest related species with an active PRDM9. Both species show the expected patterns consistent with lack of methylation and high recombination in hotspots, CGIs and promoters. Like mutation rates, they are additionally reduced in intersecting regions. However, the species differ in other substitution rates in CGIs. Dogs show highly increased substitution rates in general, and WS rates in particular, while panda does not. This is a signature of gBGC acting in the dog but not the panda genome. Clearly, different evolutionary forces are acting in CGIs in their respective genomes. These substitution patterns predict a dramatic increase in GC\* in dogs, while panda displays a more moderate level of GC\*. Again, promoter-associated CGIs show the most marked difference.

Analyses of CGI distribution in the dog genome reveal a relative increase in intergenic regions compared to panda, and an increasing association with hotspots. This likely reflects the relocation of hotspots to unmethylated regions and their reinforcement via gBGC, which strengthens our inference of them as CGIs. Recombination rates in CGIs are also increased relative to panda, especially in hotspots and promoters, linking recombination to CGIs.

## Conclusion

The main trends of a) higher substitution rate in dog, especially in CGIs b) suppressed CpG rates in both species and c) increased WS rate and GC\* in dog CGIs is consistent with a joint effect of non-methylation and gBGC. In absence of PRDM9, recombination sites in dogs have moved to unmethylated CGIs where they remain and increase GC\* via gBGC. The highly accelerated evolution in dog CGIs, which have important functions as TFBS, could have significance for functional evolution in dog. The link between recombination and CGIs indicates gBGC as a likely cause of the dog CGI expansion and suggests promoter-associated CGIs as most efficient at promoting recombination as a result of lower levels of methylation.

## Concluding remarks and future perspectives

Meiotic recombination clearly has remarkable effects on sequence evolution. The work conducted in this thesis provides insight into the various effects on genomic variability associated with recombination. In order to achieve this we performed several detailed analyses to reveal and explain the role of underlying factors on molecular evolution to enhance the understanding of the interplay between recombination and genomic features. CGIs emerge as important targets of recombination and predict an accumulation in regions of strong gBGC with increased rate of structural variation and substitutions.

Paper I and II characterize the mutational properties of regions identified as being under the influence of strong positive selection, and demonstrates that gBGC can be mistaken as signs of positive selection and lead to false inferences in selection scans. Thus, several protein-coding changes in fast evolving genes are prompted by the biased fixation of AT-to-GC mutations, which is inherent of the genomic architecture and completely distinct to adaptation. The methods implemented in these papers enable estimates of the contribution of gBGC and could be utilized to characterize the genetic profile of genes. Considering the substantial portion of regions undergoing rapid evolution that is affected by gBGC, it is imperative to incorporate estimates of the contribution of this force in selection tests.

Recently, direct measurement of the over-transmission of G:C alleles from pedigrees have been successfully demonstrated in humans (Williams et al. 2014). However, the variation in this transmission bias is still unclear. Does it resemble recombination and is locally increased in certain regions, and how is that controlled? Is the distortion different between species, and could it be reversed? It is of eminent interest to increase the number of observations of recombination and analyze other species to investigate this pattern under a phylogenetic framework. With continuously decreasing costs to sequence whole genomes it is important to further quantify the skew in transmission proportions by analyzing larger pedigrees.

Paper III and IV demonstrate the importance of CGIs on genomic variability and genome evolution in the dog genome. Both CNVs and recombination hotspots in dog are enriched for peaks of elevated GC-content identified as CGIs. Paper III characterizes structural variation in dogs, and consolidates a map of CNVs to aid future gene mapping studies. However, new techniques emerge and with next-generation sequencing it is now theoretically possible to achieve maximum resolution and map CNVs down to single

basepairs, which have successfully been done in several species, including pooled dog and wolf samples (Axelsson et al. 2013). The potential adverse effects of CNVs decline with smaller segments, and diminish the strength of purifying selection. Thus, the greatly improved mapping resolution predicts more breed-specific CNVs in contrast to the finding in paper III. Therefore, in search of breed-defining genetics to aid trait mapping, it is of uttermost importance to analyze the large amount of variation captured by segments in between kb and bp in size.

In paper IV we efficiently bridge the gap between the evolutionary process of gBGC and the sequence feature of CGIs by demonstrating the interplay via a common connection to recombination hotspots in the dog genome. Few mutations provide a viable natural knockout model to analyze the effects of inactivation of a gene as the example of PRDM9. The theory of CGIs as targets of recombination without PRDM9 is entirely based on analyses in dogs. However, it is plausible to consider additional alternative mechanisms of recombination initiation. For example, genes with similar activity could have taken over the role of PRDM9. Thus, related species should be interrogated for the distribution of CGIs and compare those with maps of recombination to further cement the importance of CGIs in genome evolution.

The work in this thesis highlights the importance of additional characterization of genes under positive selection. It expands current research and uncovers convincing associations between genomic features and proposes key roles of recombination in genome evolution.

# Acknowledgements

The work presented in this thesis was conducted at the department of medical biochemistry and microbiology at Uppsala University in Sweden. I would like to express my gratitude to several people who I have had the opportunity to work with during these years and who had important parts in the completion of this thesis; they include **supervisors, colleagues, co-authors, collaborators, administration, friends, relatives and family**, and in particular I would like to acknowledge the following people:

My main supervisor **Matt** for your scientific guidance and valuable advice, you are a source of knowledge and motivation. You patiently pushed (and pulled) me towards this day. Clarity and focus are key concepts in your supervision. Special thanks also to my co-supervisors **Leif** and **Kerstin** for your broad experience, genuine dedication and scientific excellence, which always inspire an exhausted mind and encourage persistence.

I would also like to thank past and present **research group members and office mates** for science, humor and socialization. Especially **Abhi** for our pronounced perception differences and related discussions, **Andreas W.** for support with software, **Matteo** for chats about politics and football but also beer and burgers and watching football, and **Olaf** for thesis feedback and your truly inspiring research commitment. You have all made my working hours endurable, and provided necessary breaks for body and mind.

Other members of the lab I would like to thank include **Eva** and **Ulla** for always helping with any imaginable question regarding lab issues, and for your happy and helpful attitudes and **Cecilia** for excellent management. And thanks to **D11:3** for the Friday fikas!

Any help related to management could always be obtained from the administration of IMBIM. In particular **Barbro**, who was a living dictionary when it came to paperwork, and **Rehné** shouldered that responsibility with worthiness. Keep supporting PhD students in that invaluable manner, just like **IPhA**, which I would like to thank for organizing various social occasions. Thanks to **Andreas K.** and **Andreas D.** for company during dry teaching hours in the wet-lab and to the **IMBIM-people** participating in the circle-trainings at the BMC gym with **coach**.

Work would not work without fantastic friendship. With **Alex** I completed my first Swedish Classic, my second Swedish Classic and is pursuing my third Swedish Classic. Thanks for your heroic efforts in hours of organized training, well planned and executed preparations and excellent company during hours of pain. Remember, next year a Super Classic it is! Warm appreciations also to **Freyja** and your twisted mind for helping me recover from my adventures with **Alex** both physically and mentally, by exercise, feeding and mindfulness. Swimming with you also helped me relax during intense thesis writing. Your much-appreciated hospitality combined with cooking appetite satisfied my eating appetite at several occasions. Thanks also for letting me cry at your place during hard times in my life, I deeply treasure your friendship. Thanks also to **Iris** for inviting me to the first game night of many, I am always in a good mood around you and your happy, energetic aura. You opened your arms and showed what real friendship is by taking me to your home in Greece during a fantastic vacation, never stop caring for others. Thanks to **Axel** for work enthusiasm and beer selection and to **Agnese** for cheering me up with your spontaneous mood and letting me poke in your garden, and to both of you for really great visits to your homes, memorable orienteering sessions in the forest and countless swims in the lake during nice days. You have each and everyone enriched my existence in many ways.

The same goes for **Marcin**, **Fabiana**, **Marta** and **Doreen**, it has been a pleasure to learn to know you. Your pets are adorable creatures and I thank you for the interactions I have had with them and for dinners at your places. **Marcin** I have never been to your place, and for me your rabbits count as pets. Thanks to **Fabiana** for lunch walks and talks, **Marta** for occasional activity trips to Fjällnora and **Doreen** for my involvement in your move; I love your carpet! Thanks to all my friends for Sunday brunches.

Last but not least I would also like to express my sincere gratitude towards my beloved family. **Dad**, you have always been a strong and adamant character in my life, your integrity has inspired me to go my own way and pursue my dreams. **Mom**, you have always been an extremely supportive, kind and warm-hearted person, which have taught me valuable lessons about empathy and to see the world through the eyes of others. As my parents you have always provided a physical and mental sanctuary. Thanks to my twin sister **Erika** for accompanying me through life from our first breaths, you have always been by my side through thick and thin. Thanks to my brother **Andreas** for growing up as my loyal sidekick always ready to rumble, together we could move mountains. As my siblings you have always been a constant source of annoyance, appreciation and endless love.

# References

- Arndt PF, Burge CB, Hwa T. 2003. DNA Sequence Evolution with Neighbor-Dependent Mutation. *J. Comput. Biol.* 10:313–322. doi: 10.1089/10665270360688039.
- Auton A et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. *Science.* 336:193–198. doi: 10.1126/science.1216872.
- Auton A et al. 2013. Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs. *PLoS Genet.* 9:e1003984. doi: 10.1371/journal.pgen.1003984.
- Axelsson E et al. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature.* 495:360–364. doi: 10.1038/nature11837.
- Axelsson E, Webster MT, Ratnakumar A, Ponting CP, Lindblad-Toh K. 2012. Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. *Genome Res.* 22:51–63. doi: 10.1101/gr.124123.111.
- Baudat F et al. 2010. PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. *Science.* 327:836–840. doi: 10.1126/science.1183439.
- Bird AP. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* 8:1499–1504.
- Bird CP et al. 2007. Fast-evolving noncoding sequences in the human genome. *Genome Biol.* 8:R118. doi: 10.1186/gb-2007-8-6-r118.
- Birdsell JA. 2002. Integrating Genomics, Bioinformatics, and Classical Genetics to Study the Effects of Recombination on Genome Evolution. *Mol. Biol. Evol.* 19:1181–1197.
- Boulton A, Myers RS, Redfield RJ. 1997. The hotspot conversion paradox and the evolution of meiotic recombination. *Proc. Natl. Acad. Sci.* 94:8058–8063.
- Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature.* 485:642–645. doi: 10.1038/nature11089.
- Brown TC, Jiricny J. 1989. Repair of base-base mismatches in simian and human cells. *Genome Natl. Res. Counc. Can. Génome Cons. Natl. Rech. Can.* 31:578–583.
- Capra JA, Pollard KS. 2011. Substitution Patterns Are GC-Biased in Divergent Sequences across the Metazoans. *Genome Biol. Evol.* 3:516–527. doi: 10.1093/gbe/evr051.
- Chen W-K, Swartz JD, Rush LJ, Alvarez CE. 2009. Mapping DNA structural variation in dogs. *Genome Res.* 19:500–509. doi: 10.1101/gr.083741.108.
- Cheung VG, Sherman SL, Feingold E. 2010. Genetic Control of Hotspots. *Science.* 327:791–792. doi: 10.1126/science.1187155.
- Conrad DF, Andrews TD, Carter NP, Hurler ME, Pritchard JK. 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* 38:75–81. doi: 10.1038/ng1697.

- Coop G, Myers SR. 2007. Live Hot, Die Young: Transmission Distortion in Recombination Hotspots. *PLoS Genet.* 3:e35. doi: 10.1371/journal.pgen.0030035.
- Dreszer TR, Wall GD, Haussler D, Pollard KS. 2007. Biased clustered substitutions in the human genome: The footprints of male-driven biased gene conversion. *Genome Res.* 17:1420–1430. doi: 10.1101/gr.6395807.
- Duret L, Arndt PF. 2008. The Impact of Recombination on Nucleotide Substitutions in the Human Genome. *PLoS Genet.* 4:e1000071. doi: 10.1371/journal.pgen.1000071.
- Duret L, Galtier N. 2009. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annu. Rev. Genomics Hum. Genet.* 10:285–311. doi: 10.1146/annurev-genom-082908-150001.
- Duret L, Semon M, Piganeau G, Mouchiroud D, Galtier N. 2002. Vanishing GC-Rich Isochores in Mammalian Genomes. *Genetics.* 162:1837–1847.
- Flint J et al. 1986. High frequencies of  $\alpha$ -thalassaemia are the result of natural selection by malaria. *Nature.* 321:744–750. doi: 10.1038/321744a0.
- Frazer KA et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 449:851–861. doi: 10.1038/nature06258.
- Fullerton SM, Carvalho AB, Clark AG. 2001. Local Rates of Recombination Are Positively Correlated with GC Content in the Human Genome. *Mol. Biol. Evol.* 18:1139–1142.
- Gabriel SB et al. 2002. The Structure of Haplotype Blocks in the Human Genome. *Science.* 296:2225–2229. doi: 10.1126/science.1069424.
- Galtier N. 2003. Gene conversion drives GC content evolution in mammalian histones. *Trends Genet.* 19:65–68. doi: 10.1016/S0168-9525(02)00002-1.
- Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* 23:273–277. doi: 10.1016/j.tig.2007.03.011.
- Galtier N, Duret L, Glémin S, Ranwez V. 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet.* 25:1–5. doi: 10.1016/j.tig.2008.10.011.
- Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-Content Evolution in Mammalian Genomes: The Biased Gene Conversion Hypothesis. *Genetics.* 159:907–911.
- Gibbs RA et al. 2007. Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. *Science.* 316:222–234. doi: 10.1126/science.1139247.
- Goidts V et al. 2006. Identification of large-scale human-specific copy number differences by inter-species array comparative genomic hybridization. *Hum. Genet.* 119:185–198. doi: 10.1007/s00439-005-0130-9.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11:725–736.
- Gonzalez E et al. 2005. The Influence of CCL3L1 Gene-Containing Segmental Duplications on HIV-1/AIDS Susceptibility. *Science.* 307:1434–1440. doi: 10.1126/science.1101160.
- Graubert TA et al. 2007. A High-Resolution Map of Segmental DNA Copy Number Variation in the Mouse Genome. *PLoS Genet.* 3:e3. doi: 10.1371/journal.pgen.0030003.
- Guryev V et al. 2008. Distribution and functional impact of DNA copy number variation in the rat. *Nat. Genet.* 40:538–545. doi: 10.1038/ng.141.
- Haber JE, Ira G, Malkova A, Sugawara N. 2004. Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. *Philos. Trans. R. Soc. B Biol. Sci.* 359:79–86.

- Handel MA, Schimenti JC. 2010. Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. *Nat. Rev. Genet.* 11:124–136. doi: 10.1038/nrg2723.
- Han L, Su B, Li W-H, Zhao Z. 2008. CpG island density and its correlations with genomic features in mammalian genomes. *Genome Biol.* 9:R79. doi: 10.1186/gb-2008-9-5-r79.
- Hartfield M, Keightley PD. 2012. Current hypotheses for the evolution of sex and recombination. *Integr. Zool.* 7:192–209. doi: 10.1111/j.1749-4877.2012.00284.x.
- Hastings PJ, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy number. *Nat. Rev. Genet.* 10:551–564. doi: 10.1038/nrg2593.
- Holliday R. 1964. A mechanism for gene conversion in fungi. *Genet. Res.* 5:282–304. doi: 10.1017/S0016672300001233.
- Hurles ME, Dermitzakis ET, Tyler-Smith C. 2008. The functional impact of structural variation in humans. *Trends Genet.* 24:238–245. doi: 10.1016/j.tig.2008.03.001.
- Hurst LD. 2009. Genetics and the understanding of selection. *Nat. Rev. Genet.* 10:83–93. doi: 10.1038/nrg2506.
- Iafate AJ et al. 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* 36:949–951. doi: 10.1038/ng1416.
- Jeffreys AJ, Kauppi L, Neumann R. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* 29:217–222. doi: 10.1038/ng1001-217.
- Jeffreys AJ, Neumann R. 2002. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat. Genet.* 31:267–271. doi: 10.1038/ng910.
- Karlsson EK et al. 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* 39:1321–1328. doi: 10.1038/ng.2007.10.
- Karlsson EK, Lindblad-Toh K. 2008. Leader of the pack: gene mapping in dogs and other model organisms. *Nat. Rev. Genet.* 9:713–725. doi: 10.1038/nrg2382.
- Katzman S, Capra JA, Haussler D, Pollard KS. 2011. Ongoing GC-Biased Evolution Is Widespread in the Human Genome and Enriched Near Recombination Hot Spots. *Genome Biol. Evol.* 3:614–626. doi: 10.1093/gbe/evr058.
- Kauppi L, Jeffreys AJ, Keeney S. 2004. Where the crossovers are: recombination distributions in mammals. *Nat. Rev. Genet.* 5:413–424. doi: 10.1038/nrg1346.
- Kim SY, Pritchard JK. 2007. Adaptive Evolution of Conserved Noncoding Elements in Mammals. *PLoS Genet.* 3:e147. doi: 10.1371/journal.pgen.0030147.
- Kong A et al. 2002. A high-resolution recombination map of the human genome. *Nat. Genet.* 31:241–247. doi: 10.1038/ng917.
- Korbel JO et al. 2007. Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science.* 318:420–426. doi: 10.1126/science.1149504.
- Kosiol C et al. 2008. Patterns of Positive Selection in Six Mammalian Genomes. *PLoS Genet.* 4:e1000144. doi: 10.1371/journal.pgen.1000144.
- Kostka D, Hubisz MJ, Siepel A, Pollard KS. 2012. The Role of GC-Biased Gene Conversion in Shaping the Fastest Evolving Regions of the Human Genome. *Mol. Biol. Evol.* 29:1047–1057. doi: 10.1093/molbev/msr279.
- Lindblad-Toh K et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature.* 438:803–819. doi: 10.1038/nature04338.

- Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*. 454:479–485. doi: 10.1038/nature07135.
- McCarroll SA et al. 2006. Common deletion polymorphisms in the human genome. *Nat. Genet.* 38:86–92. doi: 10.1038/ng1696.
- McClintock B. 1950. The Origin and Behavior of Mutable Loci in Maize. *Proc. Natl. Acad. Sci. U. S. A.* 36:344–355.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*. 351:652–654. doi: 10.1038/351652a0.
- McVean GAT et al. 2004. The Fine-Scale Structure of Recombination Rate Variation in the Human Genome. *Science*. 304:581–584. doi: 10.1126/science.1092500.
- Meunier J, Duret L. 2004. Recombination Drives the Evolution of GC-Content in the Human Genome. *Mol. Biol. Evol.* 21:984–990. doi: 10.1093/molbev/msh070.
- Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A Mouse Speciation Gene Encodes a Meiotic Histone H3 Methyltransferase. *Science*. 323:373–375. doi: 10.1126/science.1163601.
- Mills RE et al. 2011. Mapping copy number variation by population-scale genome sequencing. *Nature*. 470:59–65. doi: 10.1038/nature09708.
- Montoya-Burgos JI, Boursot P, Galtier N. 2003. Recombination explains isochores in mammalian genomes. *Trends Genet.* 19:128–130. doi: 10.1016/S0168-9525(03)00021-0.
- Muñoz-Fuentes V, Di Rienzo A, Vilà C. 2011. Prdm9, a Major Determinant of Meiotic Recombination Hotspots, Is Not Functional in Dogs and Their Wild Relatives, Wolves and Coyotes. *PLoS ONE*. 6:e25498. doi: 10.1371/journal.pone.0025498.
- Myers S et al. 2010. Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. *Science*. 327:876–879. doi: 10.1126/science.1182363.
- Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A Fine-Scale Map of Recombination Rates and Hotspots Across the Human Genome. *Science*. 310:321–324. doi: 10.1126/science.1117196.
- Myers S, Freeman C, Auton A, Donnelly P, McVean G. 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.* 40:1124–1129. doi: 10.1038/ng.213.
- Nagyilaki T. 1983. Evolution of a large population under gene conversion. *Proc. Natl. Acad. Sci.* 80:5941–5945.
- Nicholas TJ et al. 2009. The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res.* 19:491–499. doi: 10.1101/gr.084715.108.
- Nicholas TJ, Baker C, Eichler EE, Akey JM. 2011. A high-resolution integrated map of copy number polymorphisms within and between breeds of the modern domesticated dog. *BMC Genomics*. 12:414. doi: 10.1186/1471-2164-12-414.
- Oliver PL et al. 2009. Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa. *PLoS Genet.* 5:e1000753. doi: 10.1371/journal.pgen.1000753.
- Olsson M et al. 2011. A Novel Unstable Duplication Upstream of HAS2 Predisposes to a Breed-Defining Skin Phenotype and a Periodic Fever Syndrome in Chinese Shar-Pei Dogs. *PLoS Genet.* 7:e1001332. doi: 10.1371/journal.pgen.1001332.

- Paigen K, Petkov P. 2010. Mammalian recombination hot spots: properties, control and evolution. *Nat. Rev. Genet.* 11:221–233. doi: 10.1038/nrg2712.
- Parker HG et al. 2009. An Expressed Fgf4 Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs. *Science.* 325:995–998. doi: 10.1126/science.1173275.
- Parker HG, Shearin AL, Ostrander EA. 2010. Man's Best Friend Becomes Biology's Best in Show: Genome Analyses in the Domestic Dog. *Annu. Rev. Genet.* 44:309–336. doi: 10.1146/annurev-genet-102808-115200.
- Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 Controls Activation of Mammalian Recombination Hotspots. *Science.* 327:835–835. doi: 10.1126/science.1181495.
- Perry GH et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* 39:1256–1260. doi: 10.1038/ng2123.
- Perry GH et al. 2006. Hotspots for copy number variation in chimpanzees and humans. *Proc. Natl. Acad. Sci.* 103:8006–8011. doi: 10.1073/pnas.0602318103.
- Perry GH et al. 2008. The Fine-Scale and Complex Architecture of Human Copy-Number Variation. *Am. J. Hum. Genet.* 82:685–695. doi: 10.1016/j.ajhg.2007.12.010.
- Petes TD. 2001. Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.* 2:360–369. doi: 10.1038/35072078.
- Pollard KS et al. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature.* 443:167–172. doi: 10.1038/nature05113.
- Pollard KS et al. 2006. Forces Shaping the Fastest Evolving Regions in the Human Genome. *PLoS Genet.* 2:e168. doi: 10.1371/journal.pgen.0020168.
- Prabhakar S, Noonan JP, Pääbo S, Rubin EM. 2006. Accelerated Evolution of Conserved Noncoding Sequences in Humans. *Science.* 314:786–786. doi: 10.1126/science.1130738.
- Redon R et al. 2006. Global variation in copy number in the human genome. *Nature.* 444:444–454. doi: 10.1038/nature05329.
- Salmon Hillbertz NHC et al. 2007. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat. Genet.* 39:1318–1320. doi: 10.1038/ng.2007.4.
- Sharp AJ et al. 2005. Segmental Duplications and Copy-Number Variation in the Human Genome. *Am. J. Hum. Genet.* 77:78–88. doi: 10.1086/431652.
- She X, Cheng Z, Zöllner S, Church DM, Eichler EE. 2008. Mouse segmental duplication and copy number variation. *Nat. Genet.* 40:909–914. doi: 10.1038/ng.172.
- Smith M. 1978. *The Evolution of Sex* | Evolutionary biology | Cambridge University Press. <http://www.cambridge.org/us/academic/subjects/life-sciences/evolutionary-biology/evolution-sex> (Accessed October 2, 2014).
- Spencer CCA. 2006. Human polymorphism around recombination hotspots. *Biochem. Soc. Trans.* 34:535. doi: 10.1042/BST0340535.
- Sutter NB et al. 2007. A Single IGF1 Allele Is a Major Determinant of Small Size in Dogs. *Science.* 316:112–115. doi: 10.1126/science.1137045.
- Tuzun E et al. 2005. Fine-scale structural variation of the human genome. *Nat. Genet.* 37:727–732. doi: 10.1038/ng1562.
- Vaysse A et al. 2011. Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. *PLoS Genet.* 7:e1002316. doi: 10.1371/journal.pgen.1002316.
- Völker M et al. 2010. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Res.* 20:503–511. doi: 10.1101/gr.103663.109.

- vonHoldt BM et al. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*. 464:898–902. doi: 10.1038/nature08837.
- Webber C, Ponting CP. 2005. Hotspots of mutation and breakage in dog and human chromosomes. *Genome Res*. 15:1787–1797. doi: 10.1101/gr.3896805.
- Webster MT, Hurst LD. 2012. Direct and indirect consequences of meiotic recombination: implications for genome evolution. *Trends Genet*. 28:101–109. doi: 10.1016/j.tig.2011.11.002.
- Webster MT, Smith NGC. 2004. Fixation biases affecting human SNPs. *Trends Genet*. 20:122–126. doi: 10.1016/j.tig.2004.01.005.
- Webster MT, Smith NGC, Ellegren H. 2003. Compositional Evolution of Noncoding DNA in the Human and Chimpanzee Genomes. *Mol. Biol. Evol*. 20:278–286. doi: 10.1093/molbev/msg037.
- Webster MT, Smith NGC, Hultin-Rosenberg L, Arndt PF, Ellegren H. 2005. Male-Driven Biased Gene Conversion Governs the Evolution of Base Composition in Human Alu Repeats. *Mol. Biol. Evol*. 22:1468–1474. doi: 10.1093/molbev/msi136.
- Williams AL et al. 2014. Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *bioRxiv*. 009175. doi: 10.1101/009175.
- Winckler W et al. 2005. Comparison of Fine-Scale Recombination Rates in Humans and Chimpanzees. *Science*. 308:107–111. doi: 10.1126/science.1105322.
- Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol*. 15:568–573.

# Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Medicine 1038*

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: [publications.uu.se](http://publications.uu.se)  
urn:nbn:se:uu:diva-233195



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2014