DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
160

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
160

PILLE HALLAST

Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB*/*CGB*) gene clusters: diversity and divergence of young duplicated genes

TARTU UNIVERSITY
PRESS

Institute of Molecular and Cell Biology, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in molecular diagnostics) on 16.02.2009 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Supervisor: Prof. Maris Laan, PhD
Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Estonia

Opponent: Prof. John A.L. Armour, PhD
Institute of Genetics
The University of Nottingham, Queen's Medical Centre, UK

Commencement: Room No 217, 23 Riia Str., Tartu, on May 7th 2009, at 10.00.

The publication of this dissertation is granted by the University of Tartu

ISSN 1024–6479
ISBN 978–9949–19–077–5 (trükis)
ISBN 978–9949–19–078–2 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 68

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

Hallast, P.; Nagirnaja, L.; Margus, T.; Laan, M. (2005). Segmental duplications and gene conversion: Human luteinizing hormone/chorionic gonadotropin beta gene cluster. Genome Research, 15(11):1535–1546.

Hallast, P.; Rull, K.; Laan, M. (2007). The evolution and genomic landscape of *CGB1* and *CGB2* genes. Molecular and Cellular Endocrinology, 260–262: 2–11.

Hallast, P.; Saarela, J.; Palotie, A.; Laan M. (2008). High divergence in primate-specific duplicated regions: human and chimpanzee Chorionic Gonadotropin Beta genes. BMC Evolutionary Biology, 7;8: 195.

Author's contributions:

Ref. I and II – participated in experimental design, experimental and *in silico* data collection, analysis and manuscript preparation

Ref. III – participated in experimental design, conducted the experiments and analysis, wrote the first draft of the manuscript

LIST OF ABBREVIATIONS

Array-CGH	microarray based comparative genomic hybridization
BAC	bacterial artificial chromosome
bp	base pair
CG	chorionic gonadotropin
CGα	chorionic gonadotropin alpha-subunit
CGB	*chorionic gonadotropin beta* subunit gene
CNV	copy number variation
CTP	carboxyl-terminal extension
DDC	duplication-degeneration-complementation model
d_n/d_s	non-synonymous/ synonymous rate ratio (ω)
DNA	deoxyribonucleic acid
DBS	double-strand break
DSBR	double-strand break repair
EAC	escape from adaptive conflict model
eCG	equine chorionic gonadotropin
eLH	equine luteinizing hormone
FoSTeS	fork stalling and template switching model
FSH	follicle stimulating hormone
FSHB	*follicle stimulating hormone beta* subunit gene
GH	growth hormone
GTH	gonadotropic hormone
hCG	human chorionic gonadotropin
hCG-H	hyperglycosylated hCG
hCGβ	human chorionic gonadotropin hormone beta subunit
HJ	Holliday junction
IVF	*in vitro* fertilization
Kb	kilobase, 1,000 base pairs
LD	linkage disequilibrium
LH	luteinizing hormone
LHB	*luteinizing hormone beta* subunit gene
LHβ	luteinizing hormone beta subunit
MAF	minor allele frequency
Mb	megabase, 1,000,000 base pairs
MHC	major histocompatibility complex
MMIR	microhomology/microsatellite-induced replication model
MSV	multi-site variation
MSY	male-specific region
MYA	million years ago
NAHR	non-allelic homologous recombination
NF90	nuclear factor 90 protein
NHEJ	non-homologous end-joining
NTF5	*neutrophin 5* gene

NWM	New World monkeys
OWM	Old World monkeys
PCR	polymerase chain reaction
PRAME	*preferentially expressed antigen of melanoma* gene
RUVBL2	*RuvB-like 2* gene
SDSA	synthesis-dependent strand annealing
SINE	short interspersed repetitive element
snaR	small NF90-associated RNA
SNP	single nucleotide polymorphism
SRS	serial replication slippage model
ssDNA	single-stranded DNA
TSH	thyroid stimulating hormone
TSHB	*thyroid stimulating hormone beta* subunit gene
UTR	untranslated region

INTRODUCTION

The evolution of genomes has been thought to be primarily driven by single basepair mutation, chromosomal rearrangement and gene duplication (Ohno 1970). Duplications have been considered as being the key mechanism for generating new genes and biological processes that facilitated the evolution of complex organisms from primitive ones. In the vertebrate evolution, apparently two whole-genome duplications have happened (Dehal and Boore 2005), followed by a third duplication in the lineage of ray-finned fish (Jaillon et al. 2004; Postlethwait et al. 2000). In addition to the importance of whole-genome duplications in the evolution of species, smaller and more frequent duplications are also thought to be important in hominoid evolution and speciation.

Over the past few years as the genome sequences of many species have become available, it has become apparent that species- and lineage-specific duplications are far more frequent than originally believed. In comparison of human and chimpanzee, large duplications contribute considerably (2.7%) to the overall divergence compared to single base pair substitutions (1.2–1.5%) (Cheng et al. 2005; Mikkelsen et al. 2005; Watanabe et al. 2004). Segmental duplications that have >90% DNA sequence identity and range from one to several hundred kilobases, cover 5.2% of human genome (Bailey et al. 2001). In primates the functional distribution of duplicated regions has been found to be nonrandom involvolving several loci regulating immunity, reproduction, brain functions, development and adaptation. Such duplicated genomic regions are prone to non-homologous crossing over and gene conversion, which in addition to leading to concerted evolution and spread of mutations between duplicons, are often recognized as mechanisms leading to a growing number of human genomic disorders (Chen et al. 2007; Lee and Lupski 2006; Lupski 2007; Stankiewicz and Lupski 2002). The duplication architecture of a genome also influences normal phenotypic variation as 12% of human genome is estimated to consist of copy number variable regions differing between individuals (Kidd et al. 2008; Korbel et al. 2007; Redon et al. 2006; Wong et al. 2007).

However, despite the contribution of genomic rearrangements to intraspecies diversity and interspecific divergence, we still lack detailed information about the emergence, evolution and functional effect of duplicated regions.

One of the gene families that has evolved through duplication events in the primate lineage is the Gonadotropin Hormone Beta subunit family, which includes the Luteinizing hormone/ Chorionic Gonadotropin beta (*LHB/CGB*) genome cluster. The *LHB/CGB* genes encode for beta-subunit of gonadotropin hormones that have an essential role in fertility and reproduction (Moyle and Campbell 1996). The human *LHB/CGB* gene cluster (19q13.3) consists of seven highly homologous (85–99%) genes: one *LHB* and six *CGB* genes. Duplication of the ancestral *LHB* gene gave rise to a new gene *CGB*, differing from *LHB* both in time (pregnancy vs. adult lifetime) and tissue (placenta vs. pituitary) of expression as well as hormone stability (Maston and Ruvolo 2002; Policastro et al. 1986). The number of *CGB* genes differs between primate species ranging

from one in New World monkeys to up to 50 copies in gorilla (Fortna et al. 2004; Maston and Ruvolo 2002). The *LHB/CGB* genome cluster was used as a model to study the variation and evolution of primate-specific duplicated gene families.

The literature review of the present thesis gives a brief summary about the duplications of genetic material, the mechanisms leading to duplications and factors influencing their evolution. The second major overview concentrates on gonadotropin hormones and their subunit genes, functions and evolution.

The summary of the experimental research conducted in the framework of this thesis focuses on the following topics: (i) the role of gene conversion in shaping the variation and LD structure of human *LHB/CGB* genes, (ii) evolution of *LHB/CGB* clusters in human and chimpanzee and (iii) the emergence and evolution of recently duplicated genes in the cluster.

1. REVIEW OF LITERATURE

1.1 Duplications of genomic material

Duplications of genomic material can occur on two scales: duplications of the whole genome and smaller scale duplications, which occur continuously and involve individual genes and genomic segments.

It is generally accepted that whole-genome duplications drive macro-evolutionary change, but it is unclear to what extent. Genome duplications have been relatively common in plant evolutionary history and most of the angiosperms are believed to be ancient polyploids. For example Arabidopsis has experienced at least three ancient polyploidy events (Bowers et al. 2003; Vision et al. 2000). Most vertebrate lineages have undergone only one or two large-scale genome duplication events in their ~500 million year history (Gu et al. 2002a; McLysaght et al. 2002). By combining data from gene families, phylogenetic trees and genomic map position Dehal and Boore (2005) showed that there is evidence of two distinct whole-genome duplications in vertebrate evolution (the 2R hypothesis)(Dehal and Boore 2005), supporting earlier studies (reviewed in (Roth et al. 2007)) although the topic has long been heavily debated. Additional whole genome duplication has occurred in the lineage of ray-finned fish after their divergence from the land vertebrates, but before the divergence of most teleost species (Jaillon et al. 2004; Postlethwait et al. 2000). It has been speculated that the whole genome duplications have played an important role in speciation and adaptive radiation (Aburomia et al. 2003; Holland 2003; Ohno 1970). The fish-specific duplication might have contributed to the biological diversification of ray-finned fishes (Meyer and Van de Peer 2005), although many argue against this hypothesis (Donoghue and Purnell 2005). Genome duplications have also been associated with speciation in yeast and plants (De Bodt et al. 2005; Maere et al. 2005; Scannell et al. 2006; Wendel 2000).

Small-scale duplications involving distinct genomic segments are much more frequent compared to whole-genome duplications. Around 15% of genes in the human genome are believed to arise from duplication events (Lynch and Conery 2000). Nearly one third of duplicated genes are arrayed in tandem (Shoja and Zhang 2006). For *Drosophila melanogaster*, *Caenorhabditis elegans* and *Saccharomyces cervisiae* genomes gene duplicates account for 8–20% (Lynch and Conery 2000).

In human genome, segmental duplications of >1 kb in size and >90% sequence identity cover 5.2% (Bailey et al. 2002). Segmental duplications are particularly enriched at pericentromeric and subtelomeric regions: subtelomeric duplications account for 40% (Linardopoulou et al. 2005) and pericentromeric duplications for 33% of the total (She et al. 2004). Recent duplications in primate genomes have been associated with *Alu* elements as the primate-specific burst in *Alu*-repeats 35–40 million years ago could have been one critical event initiating segmental gene duplications (Bailey et al. 2003).

Lately much attention has been focused on copy number variations (CNVs), especially in the human genome. CNVs range from kilobases to megabases in size and include deletion, insertion and duplication variations between individuals (Iafrate et al. 2004; Sebat et al. 2004; Sharp et al. 2005). CNVs are enriched in duplications. Approximately 1,450 copy number variable regions, encompassing 360 Mbs or 12% of the human genome, were mapped through a study of 270 individuals from the HapMap collection by (Redon et al. 2006).

1.1.1 Molecular mechanisms leading to duplications

Several mechamisms have been proposed to potentially lead to duplications of genetic material: non-allelic homologous recombination (NAHR), non-homologous end-joining (NHEJ), fork stalling and template switching (FoSTeS), serial replication slippage (SRS), microhomology/microsatellite-induced replication (MMIR) and retrotransposition (Table 1).

Recent analyses of copy number variation breakpoints in human genome have suggested that majority of indels originate from NHEJ, NAHR and retrotransposition mechanisms. A detailed analysis of 188 human copy number variations (indels) estimated that 56% of indels originate from NHEJ events, 30% from retrotransposition events and 14% by NAHR (Korbel et al. 2007). Perry and colleagues also suggested that NHEJ may be involved in the formation of a large proportion of common human CNVs (Perry et al. 2008). Analysis of 227 indels by Kidd and colleages estimated that NAHR and NHEJ contribute approximately equally (~40%) to the rise of indels, retrotransposition accounting for ~17% of the events (Kidd et al. 2008).

Table 1. Molecular mechanisms leading to duplications

Mechanism	Molecular basis	Typical characteristics	References
NAHR	Homologous recombination between paralogous sequences	– Requires high DNA sequence identity – Involves DSBs	Reviewed in Gu et al. 2008
NHEJ	Repair mechanism for DNA DSBs	– Does not require sequence homology, can rejoin broken ends directly end-to-end – Involves DSBs	Lieber et al. 2003; Roth et al. 1985; Weterings and van Gent 2004
FoSTeS	Fork stalling or pausing and (serial) template switching during replication	– Requires microhomology for template switching – Stalling or pausing of the replication fork may be facilitated by genomic architectural elements (e.g. palindromic DNA, stem-loop structures, repeats, etc.)	Lee et al. 2007

Mechanism	Molecular basis	Typical characteristics	References
SRS	(Serial) slipped strand mispairing during replication	– Requires microhomology for template switching – Stalling or pausing of the replication fork may be facilitated by genomic architectural elements (e.g. palindromic DNA, stem-loop structures, repeats, etc.) – Replication slippage occurs on closely adjacent sites and leads to smaller rearrangements compared to FoSTeS	Chen et al. 2005a; Chen et al. 2005b; Chen et al. 2005c; Streisinger et al. 1966
MMIR	Collapse of the replication fork, repair by DNA strand invasion and reassembly of a new replication fork	– Requires microhomology or stretches of low-complexity DNA sequences (e.g. microsatellites) for strand invasion – Involves DSBs	Payen et al. 2008
Retrotrans-position	Reverse transcription of mRNA and insertion into the genome	– Loss of introns and regulatory sequences – Presence of poly A tracts and flanking short direct repeats	Reviewed in Kaessmann et al. 2009

NAHR – non-allelic homologous recombination; NHEJ – non-homologous end-joining; MMEJ – microhomology-mediated end-joining; FoSTeS – fork stalling and template switching; SRS – serial replication slippage; MMIR – microhomology/microsatellite-induced replication; DSB – double-strand break.

1.1.2 Evolutionary fates and rates of gene duplications

Duplications of individual genes, chromosomal segments or entire genomes have long been thought to be the main source of genetic novelty, including new gene functions and expression patterns. Although the discussions about occurrence and consequences of duplications are over 100 years old (reviewed in (Taylor and Raes 2004), Susumu Ohno's book "Evolution by gene duplication" from 1970 where he stated that gene duplication is the most important factor in evolution, is the most widely cited (Ohno 1970).

Three alternative outcomes have been suggested in the evolution of duplicated genes: (i) most often one of the gene copies becomes silenced by degenerative mutations (nonfunctionalization); (ii) one copy of the duplication may acquire a novel, beneficial function and become preserved by natural selection, while the other copy retains the original function (neofunctionalization); or (iii) both, the original and the duplicate gene, mutate and evolve to fulfill

complementary functions already present in the original gene (subfunctionalization) (Force et al. 1999; Hughes 1994; Ohno 1970; Walsh 1995).

The classical model proposed by Susumu Ohno identified neofunctionalization as the primary mechanism for preservation of duplicate copies (Ohno 1970). The model is based on assumption that after gene duplication, one copy is redundant and thus freed from any functional constraint. All mutations occurring in this gene copy are selectively neutral. In most cases, one copy will accumulate deactivating mutations and become a pseudogene, while the other copy retains the original function. By chance, rare beneficial mutations can arise in the gene and lead to novel function resulting in preservation of both duplicates, one with the new function and the other with the old.

Force and colleagues proposed a subfunctionalization model called duplication-degeneration-complementation (DDC) model. DDC model predicts that degenerative mutations in regulatory regions can increase rather than reduce the probability of duplicate gene preservation and the usual mechanism of duplicate gene retention is partitioning rather than the evolution of new functions (Force et al. 1999; Hughes 1994; Lynch and Force 2000). Genes may have multiple functions and evolutionarily important changes might happen primarily at the level of gene regulation rather than protein function. The mutations causing subfunctionalization are considered to be neutral, not deleterious, because the function is still performed by the other copy of the gene. Examples of subfunctionalization are the *Hoxb1* genes. The ancestral mouse *Hoxb1* expression patterns seem to have subdivided by the zebrafish *hoxb1* duplicates. The zebrafish *hoxb1b* shares the early expression pattern of mouse *Hoxb1*, in the hindbrain of gastrulating embryos, whereas *hoxb1a* shares the later expression of mouse *Hoxb1*, in a single segment of the neurulation-stage hindbrain (McClintock et al. 2001; McClintock et al. 2002).

An alternative model of subfunctionalization is the escape from adaptive conflict (EAC) that involves adaptive mutations (Des Marais and Rausher 2008; Hittinger and Carroll 2007; Hughes 2005). EAC will occur if the two functions of the ancestral gene cannot simultaneously be optimized by the natural selection. After gene duplication, the two genes can escape from the conflict, with each gene undergoing adaptive mutations that cause it to become specialized towards one of the original subfunctions. Example of the EAC model are the *S. cerevisiae* duplicated *GAL1* (sugar kinase) and *GAL3* (transcriptional inducer) genes, in *Kluyveromyces lactis* orthologous and unduplicated gene *GAL1* serves both functions.

Many studies have suggested that both subfunctionalization and neofunctionalization are important and not mutually exclusive modes of evolution of duplicated genes (He and Zhang 2005; Hughes and Liberles 2007; MacCarthy and Bergman 2007; Rastogi and Liberles 2005). Subfunctionalization of a duplicate pair may increase the probability of the fixation of a fitness-enhancing mutation (Hughes and Liberles 2007).

The rates of duplications depend on their scale. Three whole-genome duplications are currently known in vertebrate evolution (two shared by all

vertebrates and one in the teleost lineage) whereas the rate of small-scale duplications is much higher (Dehal and Boore 2005; Jaillon et al. 2004; Postlethwait et al. 2000). The average rate of origin of new gene duplicates has been estimated on the order of 0.01 per gene per million years, with rates in different species ranging from about 0.002 to 0.02, the rate of duplication of a gene being of the same order of magnitude as the rate of mutation per nucleotide site (Gu et al. 2002b; Lynch and Conery 2000; Lynch and Conery 2003). The average half-life of a gene duplicate was estimated approximately 4 million years (Lynch and Conery 2000; Lynch and Conery 2003). However, recent estimates based on human disease-causing recurring genomic rearrangements suggest that the rate of rearrangements might be several orders of magnitude greater than those of point mutations (Lupski 2007).

1.1.3 Molecular mechanisms affecting evolutionary change of duplicated genes

1.1.3.1 Non-allelic homologous recombination

Non-allelic homologous recombination between highly similar DNA sequences can increase or decrease the number of genes in a gene family (Papadakis and Patrinos 1999). NAHR events can be mediated by perfectly matching segments smaller than 50 bp (Lam and Jeffreys 2006), but more commonly 300–500 bp is necessary (Reiter et al. 1998) making segmental duplications and repetitive sequences suitable initiators of NAHR. For example, the human fetal Gγ and Aγ-globin genes resulted from a tandem duplication event of an aproximately 5.5 kb DNA fragment. The duplication resulted from an unequal homologous crossover between two related L1 long interspersed repetitive elements, located upstream and downstream of the ancestral γ-globin gene (Fitch et al. 1991; Papadakis and Patrinos 1999). Also, comparison of the human and chimpanzee genomes revealed that a large number of deletions (612 human specific and 914 chimpanzee specific deletions (Mikkelsen et al. 2005); 492 human specific deletions (Sen et al. 2006)) appeared to have resulted from recombination between two nearby *Alu* elements present in the common ancestor (Mikkelsen et al. 2005; Sen et al. 2006).

The localization of NAHR events is often nonrandom in the genome. Common features shared among NAHR hotspots include clustering within small (<1 kb) genomic regions, coincidence with apparent gene conversion events and no obvious sequence similarities with one another. It is not clear if recombination hotspots reflect cis-acting sequence motifs, positional preference for trans-acting factors, unusual non-B DNA structures or denote genomic regions more susceptible to DNA double-stranded breaks. DNA structures that have been shown to induce DSBs (such as palindromes, minisatellites and DNA transposons) have often been reported near NAHR hotspots (Inoue and Lupski 2002).

NAHR is often considered as the main mechanism causing a class of human diseases called genomic disorders. These disorders result from chromosome rearrangements, often initiated by segmental duplications that can lead to a disease through the loss or gain of a dosage-sensitive gene(s) or a disruption of a gene (reviwed in (Conrad and Antonarakis 2007; Stankiewicz and Lupski 2002)).

1.1.3.2 Gene conversion as the main form of homologous recombination

Gene conversion is a nonreciprocal transfer of sequence information between a pair of nonallelic DNA sequences (interlocus gene conversion) or allelic sequences (interallelic gene conversion). The prerequisite of gene conversion is high sequence identity between the involved regions (Liskay et al. 1987; Lukacsovich and Waldman 1999; Rubnitz and Subramani 1984). One of the interacting sequences, the donor, remains unchanged, but the other DNA sequence, the acceptor, is changed so that it gains some sequence copied from the donor.

In eukaryotes, gene conversion constitutes the main form of homologous recombination that is initiated by DNA double-strand breaks. Gene conversion mediates the transfer of genetic information from intact homologous sequence to the region that contains the DSB, and it can occur between sister chromatids, homologous chromosomes or homologous sequences on either the same chromatid or different chromosomes. During meiosis, DSBs are created by a topoisomerase-like enzyme, whereas during mitosis they can be induced by radiation, free radicals, chemicals, nucleases and may occur at stalled replication forks (Paques and Haber 1999).

Several models are used to explain gene conversion (Figure 1): (i) the double-strand break repair (DSBR) (Szostak et al. 1983); (ii) the synthesis-dependent strand-annealing (SDSA) (Ferguson and Holloman 1996; Nassif et al. 1994; Paques and Haber 1999) and (iii) the double Holliday junction dissolution model (Hastings 1988; Nassif et al. 1994).

Since both are initiated by DSBs, a number of studies have found co-localization of recombination and gene conversion activities. In the direct study of inter-allelic gene conversion by sperm typing of three human recombination hot spots (*DNA3* and *DMB2* in MHC and *SHOX* in the pseudoautosomal pairing region PAR1 in the sex chromosomes) the peak of gene conversion activity coincided with the peak of crossover rates (Jeffreys and May 2004). Similar results have been reported in mouse *Psmb9* hot spot (Guillon and de Massy 2002) and in the *rosy* locus in *D. melanogaster* (Hilliker et al. 1994). The gene conversion frequency and the local recombination rate were also found to be positively correlated in the recently published large-scale study of inter-locus gene conversion in human genome (Benovoy and Drouin 2008).

During recombination, gene conversion events are alternative solutions to crossovers. There are indications that gene conversions actually outnumber crossovers. (Jeffreys and May 2004) estimated that ~80–94% of recombination events are gene conversions. Higher gene conversion than recombination rate

has been reported for human *MS32* and *TAP2* hotspots and growth hormone gene family (Gay et al. 2007; Sedman et al. 2008). Also in *Arabiopsis thaliana*, it has been estimated that up to 90% of crossovers result in gene conversion (Haubold et al. 2002), although a more recent study found the rates of gene conversion and crossover to be approximately equal (Plagnol et al. 2006).

According to some studies, the non-allelic gene conversion has been found to show biased directionality. The proximal-to-distal gene conversion rate between two directly repeated HERV elements on human Yq was estimated as 2.4×10^{-4} to 1.2×10^{-3} per generation, the rate of distal-to-proximal about 20 times lower (Bosch et al. 2004). In human globin genes, the directionality of gene conversion has been correlated with the relative levels of expression of the participating genes (Papadakis and Patrinos 1999).

The gene conversion tracts in yeast (*S. cerevisiae*) have generally found to be slightly longer compared to mammalian cells. The yeast conversion tract lengths vary from 1–2 kb (Judd and Petes 1988; Mancera et al. 2008) to very short (<12 bp) regions (Palmer et al. 2003). Study based on 63 multigene families in yeast found the average length of converted regions to be 173 ± 220 bp (range 8 to 1,181 bp)(Drouin 2002).

In mammals, gene conversion tracts are usually short, on the order of 200 bp to 1 kb in length. In *D. melanogaster rosy* locus the mean gene conversion tract length was 352 bp (Hilliker et al. 1994). In mouse *Psmb9* hot spot gene conversion tracts were <540 bp long (Guillon and de Massy 2002), in hotspot located on chromosome 8 C-D average tract length of 371 bp (range 255 to 627 bp) was seen (Nishant et al. 2004). In human (Table 2), the estimates range from 54–132 bp for single-sperm analysis of the leukocyte antigen HLA-DPB1 locus (Zangenberg et al. 1995), to 55–290 bp for various gene conversion hotspots (Jeffreys and May 2004), 113–2,266 bp for the globin genes (Papadakis and Patrinos 1999) and to 1–1,365 bp for two Yq-located endogenous retroviral (HERV) sequences (Bosch et al. 2004).

Large-scale studies of inter-locus gene conversion have been conducted in several organisms. Gene conversion has been detected in 7.8% (69/879) of yeast *S. cerevisiae* gene pairs (Drouin 2002), in 2% (143/7,829) of nematode worm *C. elegans* gene pairs (Semple and Wolfe 1999), ~13% (340/2,641) of mouse *Mus musculus* gene pairs (Ezawa et al. 2006). In rice *Oryza satva indica*, gene conversion was detected in 9.7% (513/5,274) of studied genes (Xu et al. 2008), frequent gene conversion has also been found in *A. thaliana* genome (Mondragon-Palomino and Gaut 2005). Common features of the studies include short tract length (around few hundred base pairs), higher frequency of intra-chromosomal compared to inter-chromosomal gene conversion events, increasing frequency with increasing nucleotide identity and decreasing physical distance between gene pairs (Drouin 2002; Ezawa et al. 2006; Mondragon-Palomino and Gaut 2005; Semple and Wolfe 1999; Xu et al. 2008). A recent large-scale study of 1,434 protein coding human gene families with three or more genes found that gene conversion occurred at frequency of 0.88% (483 events of 55,050 compared gene pairs)(Benovoy and Drouin 2008).

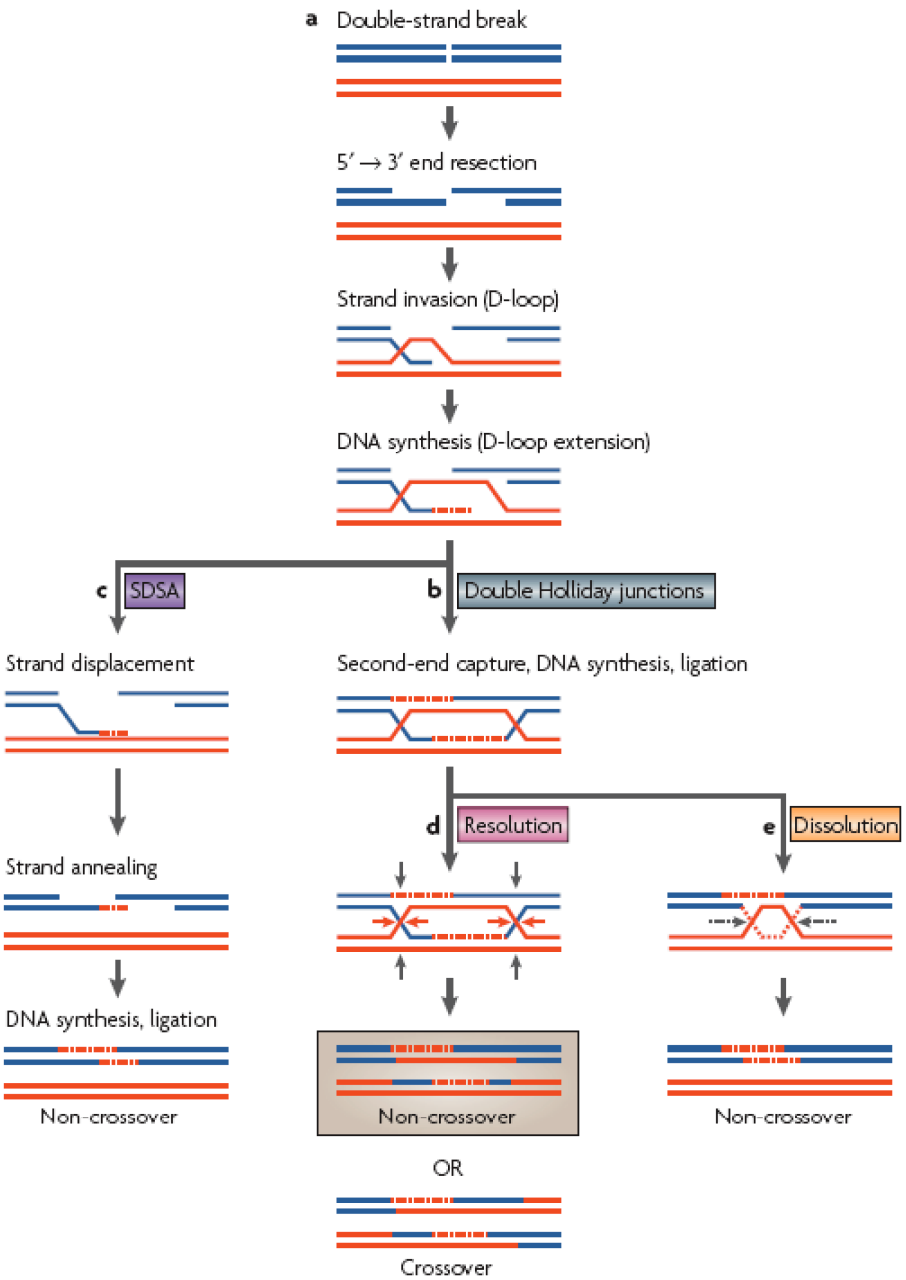

Figure 1. Gene conversion mechanisms (Chen et al. 2007). The double-strand break repair (DSBR; a–b–d), synthesis-dependent strand-annealing (SDSA; a–c) and double Holliday junction (HJ) dissolution (a–b–e) models are illustrated.

All models share a common initiating step: the 5' ends of the double-strand break are resected to form 3' ssDNA tails which actively scan the genome for homologous

sequences. One of the tails invades the homologous DNA duplex forming a displacement (D)-loop, which is then extended by DNA synthesis.

SDSA diverges from the other two pathways after D-loop extension: the invading strand and the newly synthesized DNA are displaced from the template and anneal to the other 3' end of the DSB, leading to the formation of only gene conversion events (Ferguson and Holloman 1996; Nassif et al. 1994; Paques and Haber 1999).

Otherwise, the 3' end of the DSB is captured and DNA synthesis and ligation of nicks lead to the formation of double HJs. According to the dissolution model, the double HJs are removed via convergent branch migration (indicated by dotted arrows) leading exclusively to gene conversion (Hastings 1988; Nassif et al. 1994).

In DSBR model, the resolution of the double HJs is predicted to generate an equal number of non-crossover (indicated by red arrows at both HJs) and crossover (indicated by black arrows at one HJ and red arrows at the other HJ) events (Szostak et al. 1983).

Table 2. Examples of gene conversion events in human genome.

Locus	Length of converted region	Gene conversion frequency	Detection method	Reference
MHC DMB2	55–290 bp	1.3–3.4 x 10^{-3} per sperm	Sperm typing	Jeffreys and May 2004
HLA-DPB1	54–132 bp	~1/10,000 sperm	Sperm typing	Zangenberg et al. 1995
Globin genes	113–2,266 bp	–	Bioinformatic analyis of gene sequences	Papadakis and Patrinos 1999
AZFa region	1–1,365 bp	2.4 x 10^{-4} to 1.2 x 10^{-3} per generation	Bioinformatic analyis of gene sequences	Bosch et al. 2004
MSY	–	~600 bp per newborn male	Bioinformatic analyis of gene sequences	Rozen et al. 2003
GH/CSH gene family	3–142 bp, mean 24 bp	–	Bioinformatic analyis of gene sequences	Sedman et al. 2008
1,434 protein coding gene families with ≥3 members	10 – 6,011 bp, average 371±752 bp	0.88% (483 conversion events/ 55,050 gene pairs compared)	Bioinformatic analyis of gene sequences	Benovoy and Drouin 2008

1.1.3.3 Effects of gene conversion on duplicated genomic regions

Gene conversion is a mechanism acting on genomic regions with high sequence identity, on one hand maintaining the sequence homogeneity but on the other hand producing sequence diversity.

Inter-locus gene conversion has an important role in the concerted evolution of multigene families and highly repeated DNA sequences. A hallmark of its action is that paralogous gene sequences become more closely related to each other than they are to their orthologous counterparts in closely related species (Li 1997; Li et al. 2005). In humans the first gene conversion events were reported in the globin genes (Slightom et al. 1980). Since then, the inter-locus gene conversion has been implicated in the concerted evolution of many human gene families like the Rh blood group antigen genes *RHD* and *RHCE* (Innan 2003), red (*OPN1LW*) and green (*OPN1MW*) opsin genes (Verrelli and Tishkoff 2004), olfactory receptor genes (Sharon et al. 1999), growth hormone gene family (Giordano et al. 1997), MHC (Zangenberg et al. 1995), the chemokine receptor genes *CCR2* and *CCR5* (Vazquez-Salat et al. 2007), etc.

Gene conversion can generate diversity by re-assorting variants among paralogs. For example, gene conversion is suggested to generate diversity in the major histocompatibility complex (Ohta 1991; Parham and Ohta 1996; Pease et al. 1993) and immunoglobulin gene families (Huber et al. 1993).

Gene conversion seems to favor some alleles over others leading towards higher GC content in homologous multigene families (Eyre-Walker 1993; Galtier et al. 2001; Mancera et al. 2008; Marais 2003). This process is known as biased gene conversion arising as a consequence of the GC-biased repair of A:C and G:T mismatches that are formed in heteroduplex recombination intermediates during meiosis (Brown and Jiricny 1988). Biased gene conversion toward GC in mammals has been shown between histone genes (Galtier 2003), *Hsp70* genes (Kudla et al. 2004), *Fxy* genes (Galtier and Duret 2007), Bex genes (Zhang 2008) and in *HINTW* genes in birds (Backstrom et al. 2005).

Repetitive sequences, i.e. *Alu* sequences, can initiate gene conversion. A genome-wide study attributed approximately 10–20% of the sequence variation in the *Alu* Ya5 subfamily to gene conversion (Roy et al. 2000). Recently, (Zhi 2007) estimated that 15,000–85,000 point mutations in the human genome have arisen through sequence exchanges between neighboring *Alu* elements.

There is growing evidence that gene conversion has an important role in shaping the patterns of linkage disequilibrium (LD) in the human genome (Ardlie et al. 2001; Frisse et al. 2001; Przeworski and Wall 2001; Ptak et al. 2004). Gene conversion decreases LD at small scales (e.g. over a few hundred base pairs), recombination has a greater role over longer stretches of DNA (Andolfatto and Nordborg 1998; Wiehe et al. 2000). Whereas crossing over is the major determinant of LD for distant sites, the effects of gene conversion cannot be ignored for closely linked sites.

In addition to affecting the evolution of gene families and genome variation patterns, gene conversion events have been implicated as the molecular cause of an increasing number of human diseases (reviewed in (Chen et al. 2007)). Pathogenic gene conversion often results from the transfer of genetic information from non-functional pseudogenes to their closely related functional counterparts. Some examples of such diseases are the Gaucher disease (conversion from *GBAP* to *GBA*)(reviewed in (Hruska et al. 2008)), Shwachman-

Bodian-Diamond syndrome (conversion from *SBDSP* to *SBDS*)(Nicolis et al. 2005), von Willebrand disease (conversion from *VWFP* to *VWF*)(Eikenboom et al. 1994), congenital adrenal hyperplasia (conversion from *CYP21A1P* to *CYP21A2*)(Higashi et al. 1988). In most of the cases, the conversion event results in the functional loss of the respective acceptor gene through the introduction of frameshift, aberrant splicing, nonsense mutations and so on. In nearly all known cases of disease-causing inter-locus gene conversion, the acceptor and donor genes are located on the same chromosome (Chen et al. 2007).

1.1.3.4 Balance between natural selection and relaxed mutation rate in homologous genes

The evolutionary forces behind the functional divergence of duplicated genes are largely unknown.

The classical view is that after duplication, the selection is relaxed because if one gene gains a mutation that alters the function, the remaining copy serves as a backup to retain the original function (Ohno 1970). The divergent copy is then free to aquire random substitutions, become non-functional or by chance, obtain a new tissue specificity or other function. Increasing functional and sequence divergence of gene duplicates can end the period of neutrality and lead to increased selective constraints on both genes (Force et al. 1999).

Relaxation of selection can also affect both of the copies and one possible outcome of such relaxation is subfunctionalization. Each copy may be slightly damaged by mutations to the point where both of them are necessary to perform the original function (Force et al. 1999; Lynch and Conery 2000).

Both these models predict that a duplication will be followed by accelerated evolution of one or both copies. Indeed early in their history, many gene duplications show a phase of relaxed selection or even accelerated evolution (Conant and Wagner 2003; Lynch and Conery 2000; Lynch and Conery 2003). However, these accelerations could be due to an increase of mutation rates after duplication, the relaxation of purifying selection due to the duplication of functional genes, the action of positive diversifying selection on one or both copies, or a combination of these (Conant and Wagner 2003; Kondrashov et al. 2002; Lynch and Conery 2000; Zhang et al. 2003). The apparent acceleration immediately after a duplication is insufficient to decide whether the duplication was adaptive or neutral (Kondrashov and Kondrashov 2006).

1.1.4 Duplicated gene families in primates

In addition to large-scale chromosomal rearrangements and single base pair substitutions, primate species differ from each other by segmental duplication content and in many gene families also by gene copy numbers.

It has been estimated that between human and chimpanzee, large segmental duplications (>20 kb and >94% identity) have a greater impact (2.7%) in

altering the genomic landscape compared to single-base-pair differences (1.2%) (Cheng et al. 2005). Demuth and colleagues found that the human genome contains 1,418 genes that do not have orthologs in the chimpanzee (689 gains in human, 729 losses in chimpanzee / 22,000 total genes) leading to at least 6% difference in their complement of genes (Demuth et al. 2006).

Primate specific gene duplications have involved loci regulating immunity (e.g. *MHC, beta-defensin*, *CD33rSiglec* gene clusters), reproduction (e.g. *GH*, *PRAME* genes, Y-chromosomal gene families), development and adaptation (e.g. *Opsin*, *Beta Globin* genes, *Rh blood group*, *Class 1 ADH*, *PRDM* and *FAM90A* gene families) and brain functions (*NAIP, ROCK1, USP10* and *MGC8902* genes) (Angata et al. 2004; Birtle et al. 2005; Bosch et al. 2007; Cheng et al. 2005; Dumas et al. 2007; Fitch et al. 1991; Fumasoni et al. 2007; Hunt et al. 1998; IHGSC 2004; Li et al. 2005; Oota et al. 2007; Piontkivska and Nei 2003; Popesco et al. 2006; Salvignol et al. 1993; Semple et al. 2003).

One example of an expanded gene family is the preferentially expressed antigen of melanoma (*PRAME*) gene family that in humans consists of a single gene on chromosome 22q11.22 and a cluster of genes on chromosome 1p36.21. *PRAME* and *PRAME*-like genes are actively expressed in cancers but normally manifest testis-specific expression and may thus have a role in spermatogenesis. Several segmental duplication events have occurred independently in human (at least 22 *PRAME* genes and 10 pseudogenes), chimpanzee (at least 12 *PRAME* genes and pseudogenes) and macaque (at least 8 *PRAME* genes and pseudogenes) lineages in the *PRAME* gene family (Birtle et al. 2005; Gibbs et al. 2007).

The *Growth Hormone* (*GH*) gene family expansion happened after the split of NWM and OWM/apes. Prosimians have only a single copy of *GH* genes. Independent duplications have been reported for several species leading to a different number of genes in *GH* locus: five in human, six in chimpanzee, five in macaque and eight *GH-like* genes in marmoset (Golos et al. 1993; Li et al. 2005; Wallis and Wallis 2006; Ye et al. 2005). Between human and chimpanzee only three genes are clearly orthologous: the two *GH* genes and one *placental lactogen (PL)* but none of the human and chimpanzee *placental lactogens* are clearly orthologous to those reported for the macaque (Revol De Mendoza et al. 2004; Ye et al. 2005).

MGC8902 gene shows a very rapid human lineage-specific amplification (Dumas et al. 2007; Fortna et al. 2004; Popesco et al. 2006). *MGC8902* gene encodes six primate-specific *DUF1220* protein domains highly expressed in brain regions associated with higher cognitive function and in brain showing neuron-specific expression preferentially in cell bodies and dendrites. 49 copies of *MGC8902* gene have been found in human, 10 in chimpanzee and 4 in macaque (Popesco et al. 2006). It has been estimated that 34 different human genes encode DUF1220 domains. The number of DUF1220 copies is highly expanded in humans (212 copies), reduced in African great apes, further reduced in orangutan and Old World monkeys, single-copy in nonprimate mammals, and absent in nonmammalian species (Popesco et al. 2006).

1.2 Glycoprotein hormones

Together with tyroid-stimulating hormone (tyrotrophin, TSH), the gonadotropins luteinizing hormone (lutropin, LH), chorionic gonadotropin (CG) and follicle-stimulating hormone (follitropin, FSH) belong to the family of glycoprotein hormones. LH, FSH and TSH are produced in pituitary gland, while CG originates from placenta. Members of the family are relatively large heterodimer proteins (molecular mass 30–40 kDa) formed by the non-covalent association of α-subunit, which is common to all members of the family and a hormone-specific β-subunit which determines the specific activity of each hormone (Pierce and Parsons 1981).

The human α-subunit is encoded by a singe gene localized on chromosome 6q12.21 (Moyle and Campbell 1996). The β-subunit genes are located on different chromosomes: *TSHB* on chromosome 1p13, *FSHB* on chromosome 11p13 and the *LHB/CGB* gene cluster on 19q13.32. The human *LHB/CGB* genome cluster consists of one *LHB* and 6 *CGB* genes (Fiddes and Talmadge 1984; Graham et al. 1987; Policastro et al. 1986).

The glycoprotein hormones exert their activity by binding to distinct cell surface receptors and activating adenylyl cyclase (Pierce and Parsons 1981). The receptors belong to the large family of G protein-coupled receptors, whose members all have a transmembrane domain with seven transmembrane helices connected by three extracellular and three intracellular loops and a large extracellular hormone-binding domain at the N-terminus. FSH and TSH bind to the FSH and TSH receptors, respectively, while LH and CG both bind to the same LH receptor. The LH receptor is located on chromosome 2p21(Rousseau-Merck et al. 1990a), FSH on 2p21–16 (Gromoll et al. 1994; Rousseau-Merck et al. 1993) and TSH receptor on 14q31 (Libert et al. 1990; Rousseau-Merck et al. 1990b).

1.2.1 Molecular structure of gonadotropins

Three of the glycoprotein hormones, the gonadotropins FSH, LH and CG have essential functions in the reproductive systems controlling the production of male and female gametes and the sex steroid hormones.

Gonadodropin hormones consist of α and β-subunit formed by non-covalent associations. The non-covalent interaction between the α- and β-subunits is stabilized by a segment of the β-subunit that extends like a “seat-belt” around the α-subunit and is “locked” by a disulfide bridge.

In human the mature α-subunit consists of 92 amino acid residues and contains 10 cycteines which form five intramolecular disulfide linkages and two N-linked glycosylation sites (Moyle and Campbell 1996). The β-subunits of FSH, LH and hCG contain 12 cycteine pairs forming six intrasubunit disulfide brigdes and two N-linked glycosylation sites (one for LHβ). Lengths of the

subunits vary from 111 amino acids for FSHβ to 121 for LHβ and 145 amino acids for hCGβ.

The *CGB* gene is believed to have been evolved from the *LHB* gene through a frame-shift mutation in the last exon elongating the open reading frame (ORF) into the previously 3' untranslated region. Thus, the hCGβ protein is larger than the LHβ protein, containing a carboxy-terminal extension (CTP) of 24 amino acids, with four additional O-linked glycosylation sites not present in LHβ (Fiddes and Goodman 1980; Talmadge et al. 1984). The amino acid homology between LHβ and hCGβ subunits is more than 80% (Pierce and Parsons 1981). These hCGβ-specific O-linked glycosylation sites explain the longer circulating half-life of hCG over LH. Human LH has a circulating half-life of 40–60 minutes (Diebel and Bogdanove 1978), whereas hCG has a circulating half-life of 5–6 hours (Armstrong et al. 1984; Wehmann and Nisula 1981; Yen et al. 1968). The glycosylated CTP also results in an altered route of hormone secretion: LH is secreted baso-laterally from gonadotrophic storage granules after GnRH-stimulation, whereas hCG is released constantly and apically by trophoblasts towards the maternal blood circulation system (Handwerger et al. 1987; Jablonka-Shariff et al. 2002; Lloyd and Childs 1988). Compared to other glycoproteins, hCG is the most glycosylated; 25–40% of the molecular weight comes from oligosaccharides (Cole 2007). An over-glycoylated variant of hCG called hyperglycosylated hCG (hCG-H) differs from regular hCG in size (>40 kDa vs. 36.7 kDa) and place of expression (cytotrophoblast vs. syncytiotrophoblasts) (O'Connor et al. 1998).

Crystal structures of human CG and FSH both reveal elongated molecules with similar folds for the α- and β-chains, and a cysteine-knot motif in the central core of each subunit (Fan and Hendrickson 2005; Fox et al. 2001; Lapthorn et al. 1994; Wu et al. 1994).

1.2.2 Functions of gonadotropins

1.2.2.1 Pituitary gonadotropins FSH and LH

FSH is secreted from pituitary and its principal role is to stimulate the development of the follicles that have resumed meiosis. In females, FSH targets a receptor (FSHR) expressed only on granulosa cells and induces them to proliferate. Follicular maturation requires FSH stimulation – in the absence of sufficient FSH, follicles fail to develop and ovulation does not occur (Dias et al. 2002; Moyle and Campbell 1996; Simoni et al. 1997). Together with activin, inhibin and estradiol, FSH has a role in selecting follicle(s) that will develop and produce sufficient estradiol to trigger the LH surge leading to ovulation. In males, FSH stimulates sertoli cell proliferation in testes and supports spermatogenesis (Dias et al. 2002; Themmen and Huhtaniemi 2000).

Luteinizing hormone acts on the ovary to promote the ovulation and luteinization of mature Graafian follicles and to stimulate steroidogenesis in thecal and interstitial cells. In humans LH also maintains the progesterone production from *corpus luteum*. Because LH stimulates androgen production in theca cells, thus providing substrate for granulosa cell estrogen production, LH has an important role in follicular estrogen formation and acts synergistically with FSH. In males LH promotes testosterone formation from the Leydig cells of the testis. These androgens are needed for spermatogenesis and for development of male secondary sexual characteristics (Moyle and Campbell 1996; Themmen and Huhtaniemi 2000).

1.2.2.2 Placental gonadotropin hCG

hCG is produced by early trophoblast cells and it prolongs the life of *corpus luteum*, thereby permitting the synthesis of progesterone, essential to prevent uterine rejection of the developing embryo. hCG is important in preparing the endometrium for the implantation of embryo and placentation. The hormone also has a role in modifying the local immunosupression, enabling the implantation of the embryo (Moyle and Campbell 1996). Reports of hCG's promotion of angiogenesis support the hypothesis of the embryo fostering maternal blood vessels growth via hCG for better supply of nutrients and easier release of hCG and other factors (Herr et al. 2007; Zygmunt et al. 2002). During pregnancy, hCG is produced by the fetal part of the placenta and is also crucial for male fetal sexual differentiation, as hCG stimulates fetal testosterone synthesis in the testicular Leydig cells (Clements et al. 1976; Gromoll et al. 2000; Huhtaniemi et al. 1977).

Hyperglycosylated hCG is responsible for cytotrophoblast invasion into the endometrium (O'Connor et al. 1998).

hCG is one of the first proteins synthezised after the conception and forms the basis of most pregnancy tests. After implantation, hCG is transported into the maternal bloodstream, where its concentration rapidly increases. The maximum level of hCG is reached by 9–10 weeks of pregnancy. Concentration decreases from the 10th to the 16th week of gestation, being approximately 25% of the peak concentration, thereafter the level of hCG falls to become 10% of the peak trimester value that persists throughout the remainder of the pregnancy (Hay 1988; Jameson and Hollenberg 1993).

Hyperglycosylated hCG accounts for the major proportion of hCG forms produced during trophoblast invasion at implantation early in pregnancy. hCG-H accounts for more than 50% of total hCG until the fifth week of pregnancy, thereafter declines rapidly and keeps a level of <2% in the second and third trimesters (Cole et al. 2004; O'Connor et al. 1998).

Abnormal levels of hCG have been associated to several phenotypes of complicated pregnancies. Low levels of hCG during the first trimester of pregnancy are related to miscarriage, ectopic pregnancy and failure of IVF procedure

(Gerhard and Runnebaum 1984; Letterie and Hibbert 2000; Poikkeus et al. 2002). Several chromosomal aberrations, for example trisomy of chromosome 21 and 18, are also associated with low hormone concentrations (Brizot et al. 1996; Brizot et al. 1995). The molar pregnancy and preeclampsia are related to elevated production of hCG (Gurbuz et al. 2004). High levels of hCG and especially β-subunit and their metabolites, refer to gestational trophoblastic disease. hCG expression in non-pregnant state is a sensitive and specific marker for trophoblastic tumors and many non-trophoblastic malignancies (Madersbacher et al. 1994; Marcillac et al. 1992; Reimer et al. 2000; Stenman et al. 2004).

Low hyperglycosylated hCG levels predict a risk for early pregnancy failure and preeclampsia; high mid-trimester levels predict Down syndrome pregnancies (Kovalevskaya et al. 2002; Sutton-Riley et al. 2006). hCG-H is also considered a very good marker for choriocarcinoma and testicular germ cell malignancy cases, identifying the presence of invasive disease and the need for chemotherapy. The more malignant the cells are, the more hCG has been found to be hyper-glycosylated (Cole 2007; Cole and Khanlian 2007; Valmu et al. 2006).

1.2.3 Evolution of gonadotropins

1.2.3.1 Gonadotropin genes in vertebrates

In most species of fish the gonadotropin types I and II control the gonadal development (designated GTH-I and GTH-II) and the duality of gonadotropins has been established for all teleost orders examined (Schulz et al. 2001). GTH-I has been shown to be chemically and functionally related to mammalian FSH and GTH-II to mammalian LH (Li and Ford 1998; Prat et al. 1996). However, in some fish species like chum salmon, bonito and carp, two types of α-subunits have been identified (Chang et al. 1988; Itoh et al. 1990; Koide et al. 1993).

FSH and LH are structurally and functionally conserved hormones and their genes have been cloned from all classes of vertebrate species including fishes, amphibians, reptiles and birds as well as mammals ((Watanabe et al. 2007) and references therein). At the amino acid level, the α-subunits are most highly conserved between species, the amino acid identity between the teleost fish, the European sea bass (*Dicentrarchus labrax*) and amphibians, reptiles, birds and mammals ranging 52–61% (Li and Ford 1998; Mateos et al. 2003).

In contrast to functionally conserved FSH and LH, the placenta-specific CG is evolutionarily young hormone. CG is produced in the placenta of only few mammals, notably primates and a few equines like horse (*Equus ferus caballus*), donkey (*Equus asinus*) and zebra (*Equus burchelli*). However, in horse the amino acid sequences of eCG and eLH are identical, produced by the same gene named *eLH/eCGB,* and differ only in their glycosylation patterns (Chopineau et al. 1999; Chopineau et al. 1995; Murphy and Martinuk 1991; Sherman et al. 1992). The protein harbours a carboxy-terminal peptide (CTP) like CG. In horses the ancestral *LHB* gene has remained as a single copy, but retained the pituitary-

specific expression and acquired both the C-terminal extended domain (CTP) and an additional property of placenta-specific expression. The horse CTP was most probably derived from a 10 bp deletion in exon 3, which is different from the single nucleotide deletion in the primate *CGB* gene. The same promoter region of the horse *LHB/CGB* gene serves as promoter in the pituitary and placenta. The main difference from other mammalian *LHB* promoters is a small insertion (50 bp) that gave rise to a second TATA-box, leaving the remaining promoter largely unchanged (Sherman et al. 1992).

1.2.3.2 Emergence of primate-specific *CGB* gene

In primates and equines the evolution of gonadotropins have occurred through independent evolutionary pathways (Murphy and Martinuk 1991; Sherman et al. 1992). In primates the duplication of the ancestral *LHB* gene and diversification of one of the gene copies led to a novel placentally expressed hormone CG. The novel *CGB* gene most probably evolved from the *LHB* gene through a 1 bp deletion in the last exon of the gene causing a frame shift and elongating the open reading frame into the previously 3' untranslated region (Fiddes and Goodman 1980).

The *CGB* gene first arose in the common ancestor of the anthropoid primates (New World monkeys, Old World monkeys, apes and human), after the anthropoids diverged from tarsiers (Figure 2). At least two subsequent duplication events occurred in the catarrhine primates (OWM, apes and humans), all of which possess multiple *CGB* copies (Table 3) (Maston and Ruvolo 2002). Fossil and molecular phylogenetic studies place the origin of the *CGB* gene between 50 and 34 MYA (Bailey et al. 1991). The *CGB* specific 1 bp deletion was found in the *CGB* genes of all the anthropoid species studied by (Maston and Ruvolo 2002), suggesting that the deletion occurred early after the initial *LHB* duplication, before the divergence of the catarrhines from the platyrrhines (NWM) (Maston and Ruvolo 2002). The copy numbers of *CGB* genes vary between different catarrhine primate species studied so far (Table 3).

The levels of CG in humans and monkeys reach maximal values early in pregnancy during the first trimester. CG expression is detected throughout the entire gestational period of great apes and humans (Lasley et al. 1980), only during the first trimester in Old World and New World monkeys (Munro et al. 1997), and has not been detected in prosimians (Shideler et al. 1983).

Interestingly, recent studies have shown that in New World monkeys the evolution of *LHB/CGB* genes has followed a different route compared to other primates harboring *CGB* genes (Figure 2). In the common marmoset (*Callithrix jacchus*) the evolution of *LHB/CGB* genes have been studied more thoroughly as marmoset is a commonly used animal model in medical studies. The pituitary of common marmoset only produces CG instead of luteinizing hormone, for LH mediated functions in most species (Muller et al. 2004). Although *LHB* and *CGB* genes are present at the genomic level, the *LHB* gene has become non-

functional and only *CGB* gene is expressed in both the pituitary and placenta tissues (Gromoll et al. 2003; Muller et al. 2004). Marmoset *CGB* possesses one N-glycosylation and two O-glycosylation sites and therefore shows differences both with human *CGB* and *LHB* representing a peptide like hCGβ but with a glycosylation pattern intermediate between human LHβ and hCGβ (Amato et al. 1998; Simula et al. 1995). *CGB* gene expression in marmoset pituitary is activated similar to human *LHB*. *In vivo* study of pregnant marmoset monkeys showed that during pregnancy there is no significant decrease of pituitary CG production, contrasting human LH down-regulation (Henke et al. 2007).

Table 3. Copy numbers of *CGB* genes in primates.

	Species	No. of *CGB* genes	Method of detection	Reference
1	Human (*Homo sapiens*)	6	Human genomic library screening and sequencing	Policastro et al. 1983
2	Orangutan (*Pongo pygmaeus*)	4	Southern blot	Maston and Ruvolo 2002
3	Gorilla (*Gorilla gorilla*)	~50	Array-CGH	Fortna et al. 2004; Dumas et al. 2007
4	Rhesus macaque (*Macaca mulatta*)	3	Southern blot Sequence analysis based on macaque genome project data	Maston and Ruvolo 2002; Henke and Gromoll 2008
5	Guereza monkey (*Colobus guereza*)	5	Southern blot	Maston and Ruvolo 2002
6	Dusky leaf monkey (*Presbytis obscura*)	5	Southern blot	Maston and Ruvolo 2002
7	Common marmoset (*Callithrix jacchus*)	1	Reverse-transcription PCR	Simula et al. 1995
8	Owl monkey (*Aotus trivirgaus*)	1	Southern blot	Maston and Ruvolo 2002
9	Dusky titi monkey (*Callicebus moloch*)	1	Southern blot	Maston and Ruvolo 2002
10	Ring-tailed lemur (*Lemur catta*)	0	Southern blot	Maston and Ruvolo 2002
11	Aye-aye (*Daubentonia madagascariensis*)	0	Southern blot	Maston and Ruvolo 2002
12	Philippine tarsier (*Tarsius bancanus*)	0	Southern blot	Maston and Ruvolo 2002

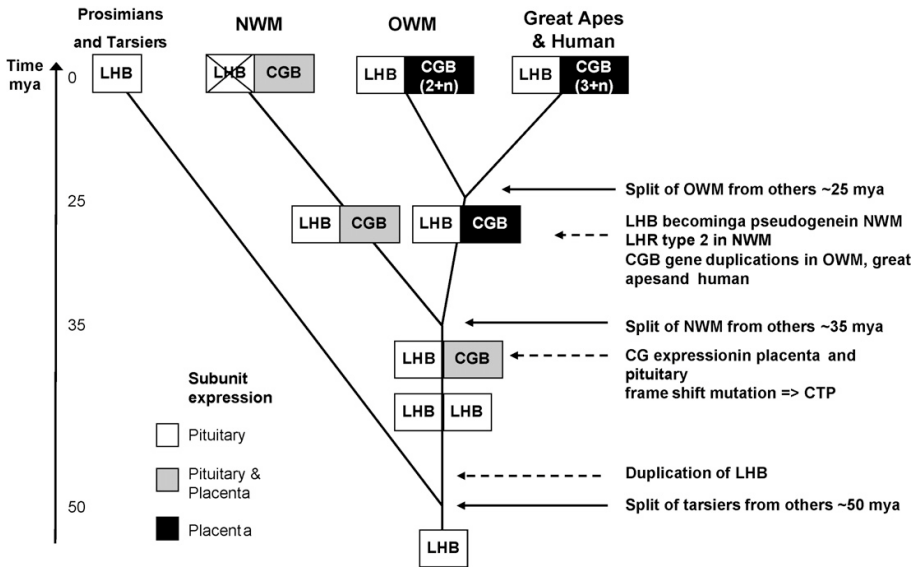

Figure 2. Schematic evolution of *LHB* and *CGB* genes in primates. NWM – New World monkeys, OWM – Old World mokeys (Henke and Gromoll 2008).

1.2.3.3 Co-evolution of CG and placental morphology

The occurrence of CG hormone is in accordance with the evolution of placental morphology (Maston and Ruvolo 2002). CG has to move from the placenta into the maternal bloodstream and then be transported to the ovary in order to act on its target, *corpus luteum*. Anthropoid primates all have hemochorial placenta, in which placental tissue is directly bathed in maternal blood, making it easy for placentally derived molecules to enter the maternal bloodstream (King 1993). Hemochorial placentation first appears in tarsiers. Strepsirrhine primates and most other mammals have an epitheliochorial placenta, in which both the uterine epithelium and the maternal vascular endothelium remain present during pregnancy (Moffett and Loke 2006). These two additional tissue layers impede the flow of large macromolecules from the placenta to the maternal blood-stream. Horses have also evolved specialized placental structures – endometrial cups – which help in the delivery of equine CG to the mare's bloodstream (Maston and Ruvolo 2002).

1.2.4 The human *LHB*/*CGB* genes

1.2.4.1 *LHB*/*CGB* gene cluster at 19q13.32

In human genome, the *LHB/CGB* gene cluster is located on chromosome 19q13.32, consisting of one *LHB* and six *CGB* genes and spanning approximately 50 kb (Figure 3)(Fiddes and Talmadge 1984; Policastro et al. 1983; Policastro et al. 1986). The beta-subunit of hCG hormone is encoded by four genes: *CGB*, *CGB5*, *CGB7* and *CGB8*. Amino acid sequence identity between the hCGβ-coding genes is 98–100% and to *LHB* 85% (Bo and Boime 1992; Hollenberg et al. 1994).

Two genes, the *CGB1* and *CGB2*, in the human *LHB/CGB* cluster do not encode for the β-subunit of hCG hormone and have long been considered to be pseudogenes. *CGB1* and *CGB2* underwent further differentiation from canonical *CGB* owing to a DNA insertion/deletion event involving the 5' UTR and upstream region of hCGβ-coding genes. These rearrangements led to a novel exon 1 (58 bp) and constitutive skipping of canonical exon 1, in which the donor splice site is disrupted. The two known exons 2 and 3 are included in the ORF but are frame-shifted, thereby leading to a hypothetical ORF of 132 codons with a new stop codon and elongated 3'UTR. The predicted protein has no similarity to hCGβ-subunit nor to any other known protein (Bo and Boime 1992; Dirnhofer et al. 1996; Hollenberg et al. 1994).

1.2.4.2 Expression of human *LHB*/*CGB* genes

Despite the high sequence similarity, *LHB* and *CGB* genes have different expression sites and patterns. *LHB* is expressed in pituitary, *CGB* genes in normal placenta and in several non-trophoblastic normal (testis, pituitary, prostate, thymus, skeletal muscle and lung) and malignant tissues (breast, ovary, bladder, lung, kidney) (Berger et al. 1994; Dirnhofer et al. 1996; Madersbacher et al. 1994; Marcillac et al. 1992; Reimer et al. 2000; Rull et al. 2008a; Stenman et al. 2004). *LHB* has a short 5'UTR (9 bp) and a consensus TATA box sequence that is located 38 bp upstream from ATG codon that is used for translation initiation. The hCGβ genes have, in contrast, a long 5'UTR, and transcription is initiated 365 bp upstream of the homologous promoter region in the *LHB* gene. The TATA box in 5'UTR of hCGβ genes is retained in non-functional state (Hollenberg et al. 1994).

In vivo studies from first trimester placenta showed that the level of expression is following: *CGB5* > *CGB* = *CGB8* > *CGB7*, *CGB1/2* (Bo and Boime 1992; Miller-Lindholm et al. 1997). Slightly different pattern of β-subunit expression in first trimester has been found by (Rull and Laan 2005): *CGB8* > *CGB5* = *CGB* >> *CGB7* and for the third trimester placentas the pattern was altered to *CGB8* ≈ *CGB5* > *CGB* >> *CGB7*. However, hetero-

geneity is found at the levels of expression of each gene in individual placentas, suggesting that the important factor during the first trimester of pregnancy is the total amount of hCGβ gene transcription and not from which gene the mRNA is derived (Miller-Lindholm et al. 1997). hCGβ has been detected already at the two-cell stage of blastocyst (Jurisicova et al. 1999).

Low expression of *CGB1* and *CGB2* mRNAs have been described in placenta, pituitary, testis and few other tissues, however no protein products have been demonstrated yet (Berger et al. 1994; Bo and Boime 1992; Dirnhofer et al. 1996; Hollenberg et al. 1994; Miller-Lindholm et al. 1997; Rull et al. 2008a; Rull and Laan 2005) and the functionality of these genes is unclear.

1.2.5 Genetic variants affecting the functions of LH and hCG hormones

So far only a few mutations have been identified in gonadotropin genes and all these have been associated with reduction or loss of function.

The only genetic alteration so far reported in the common *CGα*-subunit gene is a single $Glu^{56}Ala$ amino acid substitution in α-subunit ectopically secreted by human carcinoma. This mutated protein failed to associate with β-subunit and had a higher molecular weight compared to native α-subunit (Nishimura et al. 1986).

Three mutations have been found in human *LHB* gene ($Gly^{36}Asp$, $Glu^{54}Arg$ and $G^{536}C$ in intron 2) causing total functional inactivation of LH hormone (Table 4). The amino acid changes have been found in only two male patients; the intron 2 mutation was described in three siblings (Lofrano-Porto et al. 2007; Valdes-Socin et al. 2004; Weiss et al. 1992). $Gly^{36}Asp$ mutant *LHB* subunit showed absence of α/β heterodimerization, resulting in the absence of circulating LH (Valdes-Socin et al. 2004). In case of the $Glu^{54}Arg$ mutation, serum LH was present but devoid of biological activity because of inability to bind to LH receptor (Weiss et al. 1992). The $G^{536}C$ mutation in intron 2 disrupted the splicing of intron 2 of the *LHB* mRNA and resulted in the insertion of 236 nucleotides in the mutant transcript leading to absence of active LH hormone (Lofrano-Porto et al. 2007).

Four polymorphisms found in human *LHB* gene have been found to lead to normal or slightly decreased activity. *In vitro* experiments with $Trp^{8}Arg$ and $Ile^{15}Thr$ LH have shown higher activity than wild-type LH in bioassays, but shorter half-life in circulation (Haavisto et al. 1995; Pettersson et al. 1992; Suganuma et al. 1996). The biological function of these variants is unclear, several reports have suggested association with infertility, menstrual disturbance, spontaneous miscarriage, polycystic ovary syndrome and endometriosis (Berger et al. 2005; Elter et al. 1999; Kurioka et al. 1999; Okuno et al. 2001; Rajkhowa et al. 1995; Ramanujam et al. 1999; Takahashi et al. 1999). *In vitro* studies of the $Ala^{-3}Thr$ variant of *LHB* gene have showed a slightly different signal transduction properties compared to the wild-type LH (Jiang et al. 2002).

The Ser^{102}Gly has been found in some Asian populations like in Singapure and China and has been associated with infertility, menstrual disorders and endometriosis (Liao et al. 1998; Ramanujam et al. 1998; Ramanujam et al. 1999; Roy et al. 1996).

So far only one substitution in the *CGB* genes has been characterized leading to a Val^{79}Met change in *CGB5* gene. *In vitro* experiments showed that the assembly of α-subunit and mutated β-subunit was inefficient, although those dimers that did form had normal bioactivity (Miller-Lindholm et al. 1999).

Table 4. List of currently known mutations and polymorphisms in *LHB* and *CGB* genes.

Gene and location	Nucleotide/ amino acid change	Effect at protein level	Phenotype	Reference
LHB mutations				
Exon 2	G^{519}A/ Gly^{36}Asp	Absent bioactivity	Reduced spermatogenesis, hypoplastic Leydig cells	Valdes-Socin et al. 2004
Exon 3	G^{809}C/ Glu^{54}Arg	Absent bioactivity	Absence of spontaneus puberty, no testosterone	Weiss et al. 1992
Intron 2	G^{536}C	Absent bioactivity	Hypogonadism, azoospermia, infertility	Lofrano-Porto et al. 2007
LHB polymorphisms				
Exon 2	T^{434}G/ Trp^{8}Arg & T^{456}C/ Ile^{15}Thr	Increased *in vitro* bioactivity, decreased circulatory half-time	Slightly surpressed fertility	Berger et al. 2005; Elter et al. 1999; Haavisto et al. 1995; Kurioka et al. 1999; Pettersson et al. 1992; Suganuma et al. 1996
Exon 2	G^{404}C/ Ala^{-3}Thr	Normal	Normal	Jiang et al. 2002
Exon 3	G^{952}C/ Gly^{102}Ser	Decreased *in vitro* bioactivity	Infertility, menstrual disorders	Lamminen et al. 2002; Liao et al. 1998; Ramanujam et al. 1998; Ramanujam et al. 1999; Roy et al. 1996
CGB5 polymorphism				
Exon 3	G^{882}A/ Val^{79}Met	*In vitro* inefficient assembly of dimer		Miller-Lindholm et al. 1999

The nucleotide number was calculated according to the translation start site.

2. AIMS OF THE PRESENT STUDY

The aim of the present study was to investigate the duplicated Luteinizing hormone/ Chorionic Gonadotropin hormone beta-subunit (*LHB/CGB*) genome cluster in human and higher primates.

Primate-specific *LHB/CGB* gene cluster was used as a model to study:

1. Fine-scale sequence variation and LD structure in duplicated regions
2. Evolution of duplicated genes in closely related species
3. Emergence and evolution of novel genes created by duplication events

The specific focus of the current thesis was the following:

1. Genetic variation in the human *LHB/CGB* genes (Ref. I, II):
   - (i) The role of gene conversion in shaping the diversity and LD patterns
   - (ii) Determinants of crossover and gene conversion events
2. Evolution of *LHB/CGB* genome cluster in sister-species (Ref. II, III):
   - (i) Comparison of human and chimpanzee *LHB/CGB* genome clusters
   - (ii) Variation in substitution rates; genic and intergenic divergences
   - (iii) Impact of intra-species gene conversion in phylogeny and divergence
   - (iv) Evidence of natural selection
3. Fate of novel genes (Ref. II):
   - (i) The origin and evolutionary conservation of *CGB1* and *CGB2* genes
   - (ii) *In silico* analysis of putative *CGB1/2* promoter – implication to the gene function

3. RESULTS

3.1. The human *LHB/CGB* genome cluster: diversity, gene conversion and linkage disequilibrium (Ref. I, II)

3.1.1 The human *LHB/CGB* cluster

The genomic structure of human *LHB/CGB* cluster was constructed based on NCBI GenBank sequence, locus no. NG_000019. The human cluster is located at 19q13.32, contains seven genes (one *LHB* gene and six *CGB* genes) and spans 45,165 bp (Fiddes and Talmadge 1984; Policastro et al. 1983; Policastro et al. 1986). The cluster is flanked by *RUVBL2* and *NTF5* genes at the centromeric and telomeric sides, respectively (Figure 3). The *CGB* gene most probably arose in the primate lineage by the duplication of the ancestral *LHB* gene and the *CGB* gene copy numbers differ greatly between different primate species (Table 3).

The lengths of human *LHB/CGB* genes are following: *LHB* – 1,111 bp, *CGB5/7/8* – 1,467 bp, *CGB1/2* – 1,366 bp. All human *LHB/CGB* genes are highly identical on DNA level. The identity between the four hCGβ-subunit coding genes (*CGB, CGB5, CGB7, CGB8*) is 97%–99%, identity to the functionally divergent *LHB* ranges 92%–93% and with *CGB1* and *CGB2* genes, 85%. Identity between *CGB1* and *CGB2* genes is 97%. In addition to the genes, also the intergenic regions show very high identity ranging from 78% to 97%. The intergenic regions are rich in repetitive sequences (mainly *Alu* SINE sequences) ranging from 14.13% to 61.14%. Current structure of the *LHB/CGB* cluster reflects several past duplication events of the genes and intergenic regions. The initial duplication event probably involved also the 3' part of the flanking *neutrophin 5* (*NTF5*) gene, which is now located in three intergenic regions in the human cluster (Figure 3). The duplicated fragment of *NTF5* gene contains a χ-sequence (GCTGGTGG) that has been associated to recombination and gene conversion activity (Smith 1988). Exon 3 of *NTF5* has also given rise to five *NTF6* pseudogenes (*psNTF6A, 6G, 6B, 6G', 6A'*).

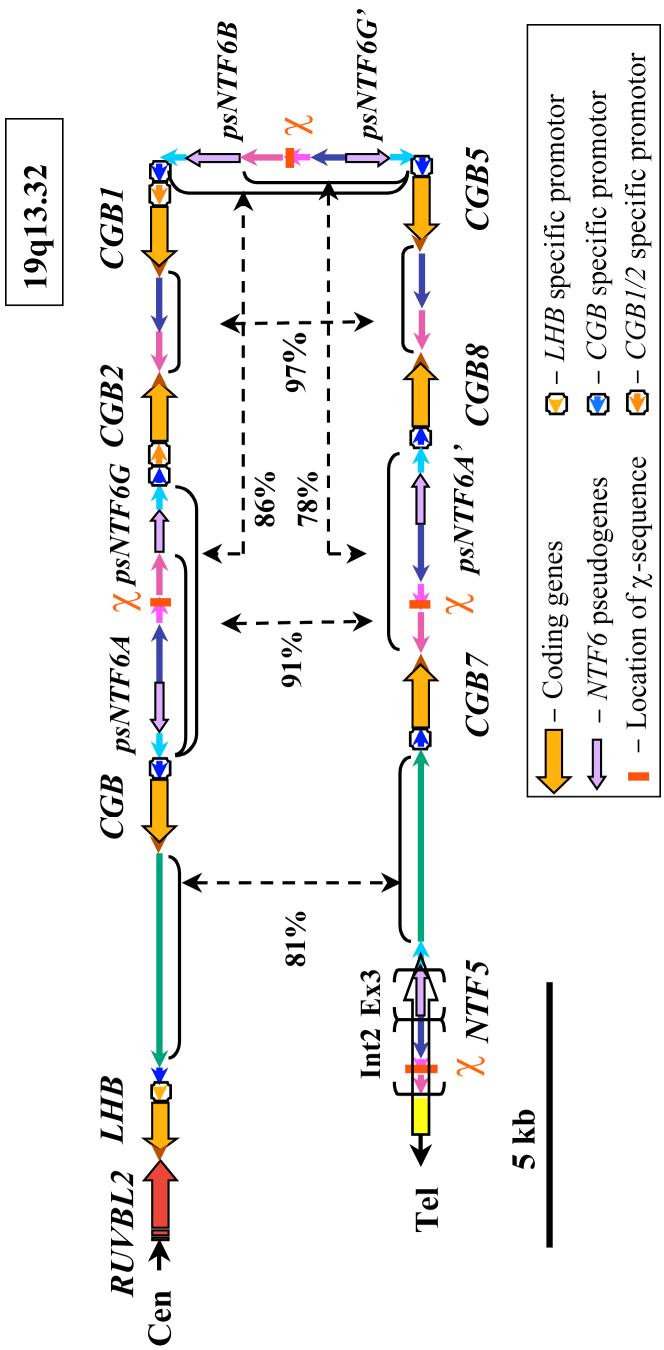

Figure 3. Structure of the human *LHB/CGB* cluster. Identical color codes refer to highly identical homologous DNA sequences within the cluster. Genes are depicted as wide arrows in the direction of transcription.

3.1.2 Human population diversity

To characterize the human population diversity of *LHB/CGB* genes, six of the genes (*LHB, CGB, CGB1, CGB2, CGB5, CGB7*) were re-sequenced in population samples from three continents: Europe (Estonians, n=47), Africa (Mandenkalu, n=23) and Asia (Chinese Han, n=25). The re-sequenced regions (including the entire coding area) ranged from 1,510 bp to 2,233 bp. In total 191 SNPs were identified from the six genes and the total re-sequenced region per individual was 10,009 bp. The number of SNPs identified per gene varied from 20 (for *CGB1*) to 50 SNPs (for *CGB7*) and the SNP density ranged from ~13/kb to ~22/kb, respectively (Table 1 in Ref. I). The sequence variation data from this study was submitted to dbSNP under accession numbers ss48399882-ss48400071 (Appendix 1). The studied genes exhibited the highest variation in the African population, Mandenkalu (Figure 4A; Table 1 and Supplemental Table S1 in Ref. I). Compared to the results of 74 genes ($\pi = 1 \times 10^{-3}$ for African Americans, $\pi = 0.8 \times 10^{-3}$ for European Americans (Crawford et al. 2004)), the nucleotide diversity (π) values of *LHB/CGB* genes in both Estonians and Mandenkalu were found up to 7 times higer (Figure 4A; Table 1 in Ref. I). Although the genes are up to 99% identical, the identified SNP patterns varied greatly between genes. The diversity of the *LHB/CGB* genes is characterized by lower variation in the central genes of the cluster (*CGB2, CGB1* and *CGB5*) and increased variation in the peripheral loci (*LHB, CGB* and *CGB7*).

Tajima's D statistic was calculated to determine if the observed patterns of diversity in the three populations are consistent with the standard neutral model (Tajima 1989). The basis of the Tajima's D value is the difference between the π (per-site heterozygosity derived from the average pairwise sequence differences) and θ (Watterson's θ, per-site heterozygosity based on the number of segregating sites) estimates: under neutral conditions $\pi = \theta$ and Tajima's D = 0. Positive Tajima's D values pointed out the excess of high-frequency SNPs for *CGB7, CGB* and *LHB* consistent with the alternative scenarios of balancing selection or population bottlenecks and subdivision. In contrast, for *CGB1*, *CGB2* and *CGB5* the Tajima's D values were mostly negative indicating higher frequency of rare variants. Enrichment of rare variants in a population is consistent with recent directional selection or population expansion (Figure 4A; Table 1 in Ref. I).

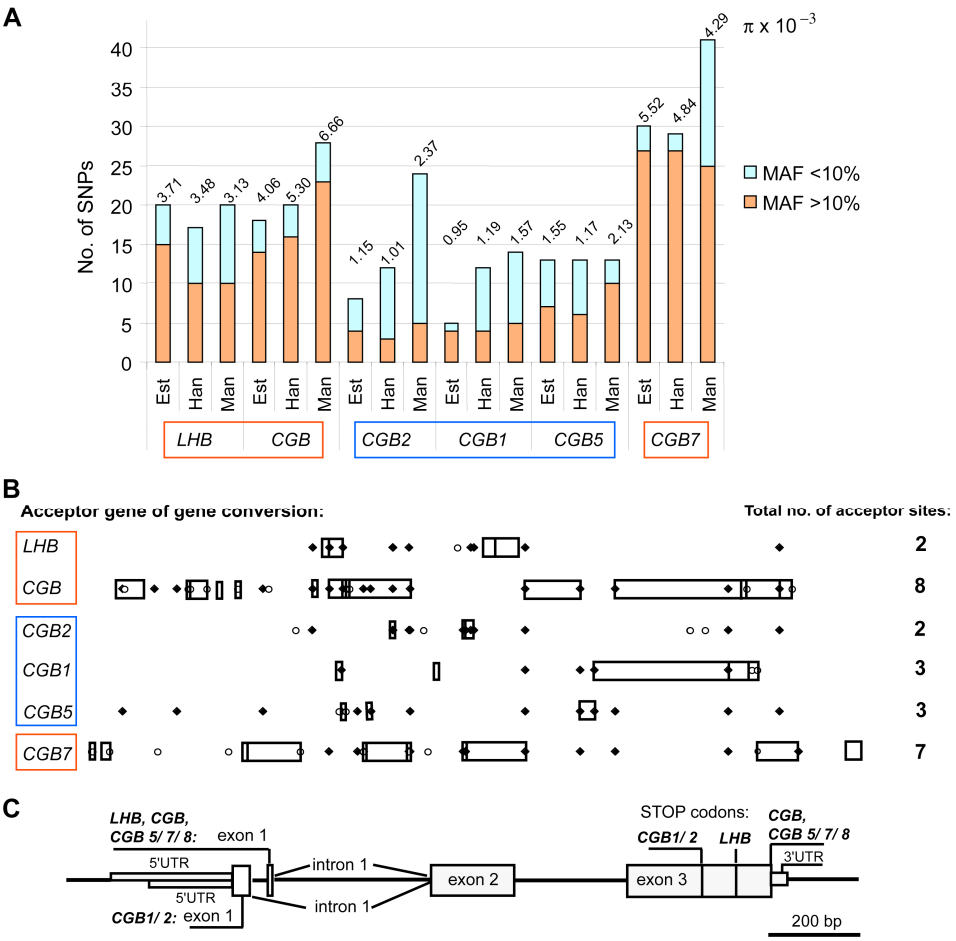

Figure 4. (A) SNPs identified in three studied population samples (Est – Estonians, n=47; Han – Chinese Han, n=25; Man – Mandenkalu, n=23). Numbers above each column show estimated nucleotide diversity per site calculated from average pairwise difference among individuals x10^{-3} (π). MAF – minor allele frequency. (B) The summary distribution of manually detected gene conversion acceptor sites and multisite variants (MSVs) within the genes in parallel with (C) the consensus exon-intron structure of *LHB/CGB* genes. Open circles denote for MSV1 (SNPs that are represented as paralogous sequence variants in other duplicons), black diamonds for MSV2 (shared SNPs between duplicons). The genes with increased variation and excess of high-frequency SNPs are marked with red squares, genes with lower variation and excess of rare SNPs with blue squares.

3.1.3 Gene conversion in human *LHB/CGB* genes

Gene conversion is known to lead to concerted evolution of duplicons and to the spread of mutations between duplicated highly identical genomic regions (Bettencourt and Feder 2002; Hurles 2001; Papadakis and Patrinos 1999). To study the potential role of gene conversion in shaping the diversity patterns in human *LHB/CGB* genes, two alternative approaches were used: (i) manual analysis of aligned haplotypes and (ii) Sawyer's gene conversion detection algorithm implemented in GENECONV program (Sawyer 1989).

For manual detection a minimum gene conversion site was defined as a region within an acceptor gene with ≥2 associated, motif-forming polymorphisms for which a potential donor gene could be defined. 25 gene conversion tracts were identified with a minimum observed tract 2–387 bp (mean 57 bp, median 23 bp) and maximum extension up to 796 bp (mean 229 bp, median 138 bp). The number of acceptor sites varied from two (for *LHB* and *CGB1*) to eight (for *CGB*)(Figure 4B-C and 5; Figure 2 and Table 2 in Ref. I). Highest numbers of gene conversion acceptor sites were identified within the two genes (*CGB7* – 7 sites and *CGB* – 8 sites) which also had the highest diversity in human populations (Figure 4A). Gene conversion might be a potential source of high diversity and positive Tajima's D values "transporting" polymorphisms from one homologous gene to the other.

Sawyer's gene conversion detection algorithm does not rely on polymorphism data, but searches for regions where pairs of sequences are unusually similar compared to overall similarity and predicts the fragments likely to have been converted between gene pairs. The GENECONV algorithm estimated 398 conversion tracts between *LHB/CGB* genes. The length of the estimated tracts ranged from 35 to 1055 bp (mean 313 bp, median 291 bp)(Supplementary Table S3 in Ref. I). *CGB2* appeared as the most active participant of gene conversion, the maximum number of estimated between-loci events reaching 49 for *CGB2-CGB* and 44 for *CGB2-CGB7* gene pairs. An association was detected between the number of conversion events estimated by GENECONV and the number of shared SNPs (MSV2) between gene pairs (Pearson's correlation coefficient 0.44, p=0.044)(Supplemental Fig. S3 in Ref. I).

3.1.4 Linkage disequilibrium in human *LHB/CGB* genes

A growing amount of studies have shown that gene conversion has a role in decreasing the extent of LD in duplicated regions (Ardlie et al. 2001; Frisse et al. 2001; Przeworski and Wall 2001; Ptak et al. 2004; Sedman et al. 2008). To study the patterns of LD in human *LHB/CGB* genes, three methods were used: (i) r^2, the correlation coefficient between alleles, (ii) the Li and Stephens (2003) "product of approximate conditionals" (PAC) likelihood method which calculates crossing-over parameter ρ and allows estimations of putative recombination hotspots (Li and Stephens 2003) and (iii) Hudson's (2001) "composite likelihood" (CL) method which allows the simultaneous estimation of crossing-over and gene conversion rate (Hudson 2001).

The r^2 showed that in all studied populations strong allelic associations at the periphery (*LHB, CGB* and *CGB7*) and breakdown towards the center (*CGB1, CGB2, CGB5*) characterize the LD structure of the cluster (Figure 5; Figure 3A-C in Ref. I). In contrast to predictions (Andolfatto and Nordborg 1998; Ardlie et al. 2001; Wiehe et al. 2000) most manually detected gene conversion acceptor sites were found to co-localize with high LD (Figure 5; Figures 3A-C in Ref. I). Thus, gene conversion may have contributed to both high diversity and high short-range LD in the peripheral (mostly acceptor genes) compared to central loci (mostly donor genes).

The average recombination rates calculated across the *LHB/CGB* genes for SNPs with MAF >10% using the Li and Stephens (2003) and Hudson's (2001) methods, fell in the range published for a large set of 74 genes (Table 3A in Ref. I)(Crawford et al. 2004). The Li and Stephens (2003) algorithm estimates the factor λ, by which the recombination rate between loci exceeds the average recombination rate in the cluster. The $\lambda > 1$ indicates increased recombination activity. The algorithm estimated a recombination "hotspot" ($\lambda > 10$) between *CGB5* and *CGB7* in all populations (λ=57.1 for Estonians, 11.6 for Han and 13.6 for Mandenkalu), a "warm spot" ($1<\lambda<10$) was identified between *CGB* and *CGB2* (λ=2.36 for Estonians, 5.47 for Hans and 4.17 for Mandenkalu). Both the hotspot and warm spot co-localized with LD breakdown on the r^2 blot and were located in regions containing the χ-sequence (Figure 3A, B and C in Ref. I).

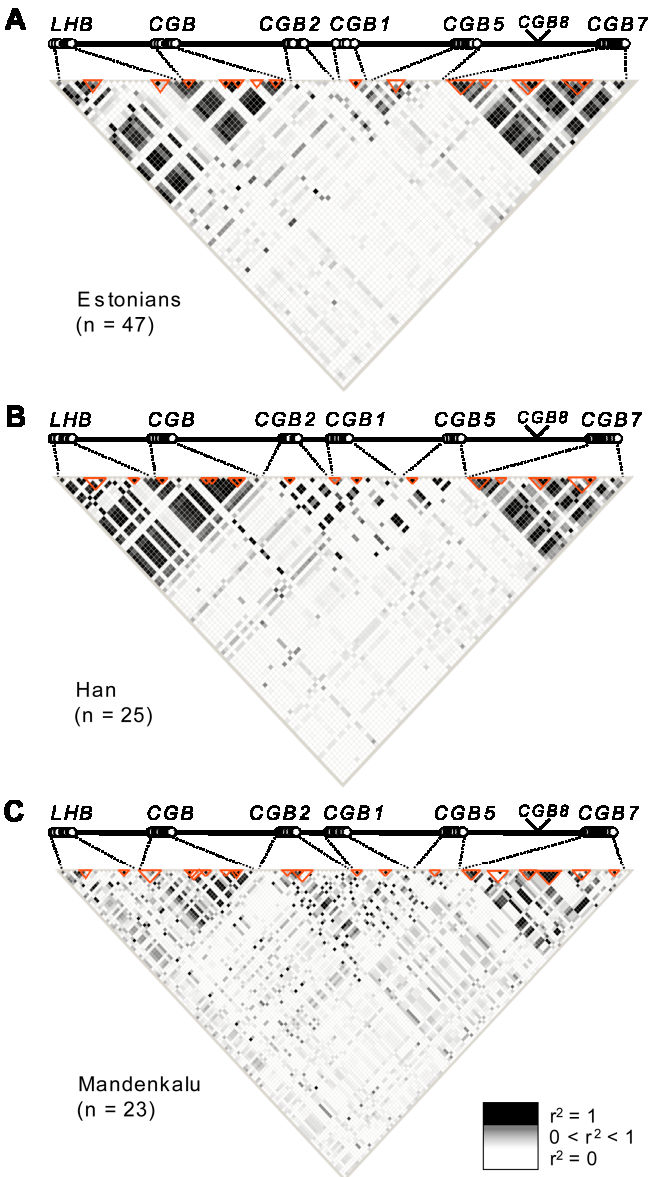

Figure 5. Linkage disequilibrium in human *LHB/CGB* genes based on r^2 and localization of manually detected gene conversion acceptor sites (A) Estonians, (B) Chinese Han and (C) Mandenkalu. Gene conversion acceptor sites detected by manual analysis of aligned haplotypes are marked by red triangles.

3.2 Human and chimpanzee *LHB/CGB* clusters (Ref. II and III)

3.2.1 Comparison of human and chimpanzee *LHB/CGB* cluster

In order to sequence the *LHB/CGB* genome cluster in the common chimpanzee, the BAC library RPCI-43 was used (BACPAC Resource Centre at the Children's Hospital Oakland Research Institute, Oakland, CA). A shotgun library was constructed based on two BAC clones (68p2 and 109b10) chosen by hybridization screening of the library and sequenced with an average redundancy of 7x using a primer-walking approach. The full sequence of the chimpanzee *LHB/CGB* genome cluster has been deposited in GenBank (accession number EU000308)(Appendix 2). The sequence characteristics of the cluster were similar between the human and chimpanzee including high GC content and high fraction of CpG islands (Table 5-A; Additional file 1 in Ref. III).

However, a substantial difference in cluster size was observed between the species: 45,165 bp in human and 39,876 bp in chimpanzee (from the end of *LHB* gene to the end of *NTF5* gene). Two highly identical, apparently orthologous segments within the cluster were identified: *RUVBL2/LHB/ intergenic region A* (Hu 7,973 bp, Ch 8,084 bp, 96% sequence identity) and the region spanning from *CGB1* to *NTF5* (Hu 28,568 bp, Ch 29,136 bp, 94.8% sequence identity)(Figure 6B). A large species-specific structural rearrangement was localized between *intergenic region A* and *CGB1* gene, resulting in discordant size of the clusters as well as species-specific number of duplicated genes, seven for human (1 *LHB* and 6 *CGB*) and six for chimpanzee (1 *LHB* and 5 *CGB* genes). In human the rearranged region (12,700 bp) harbors one hCG beta-coding gene (*CGB)* and one *CGB1/2*-like gene (*CGB2)*, in chimpanzee (6,725 bp) only a *CGB1/2*-like gene (*CGB1B)* is present in an inverted orientation compared to human.

Table 5. (A) Sequence parameters and (B) divergence values for *LHB/CGB* cluster of the current study and comparative values of previously published studies.

A – Sequence parameters			
	Current study	**Previous studies**	
	Human/ Chimpanzee	**Previous results**	**Reference**
GC%	57%/ 57%	41%	Watanabe et al. 2004
CpG islands	6.6%/ 6.1%	1–3.5%	Britten et al. 2003; Ebersberger et al. 2002
Repetitive sequences	26.9%/ 25.15%	40–50%	Lander et al. 2001; Mikkelsen et al. 2005

B – Divergence between human and chimpanzee			
	Current study	**Previous results**	**Reference**
Indels	2.7%	3–11.9%	Anzai et al. 2003; Britten 2002; Chen and Li 2001; Chen et al. 2001; Ebersberger et al. 2002; Elango et al. 2006; Hughes et al. 2005; Mikkelsen et al. 2005; Shi et al. 2003; Watanabe et al. 2004; Wetterbom et al. 2006
Nucleotide substitutions	2.3%	1.2–1.5%	
Promoters	3.22% (1–5.1%)	0.75–0.88%	
Exons	1.39% (1–1.88%)	0.51–1.09%	
Introns	2.62% (2.04–3.24%)	1.03–1.47%	
5' UTR	2.54% (0–3.83%)	1–1.41%	

3.2.2 Evidence of parallel independent duplication events in human and chimpanzee *LHB*/*CGB* genome clusters

Several lines of evidence support the scenario that independent duplication events have occurred in *LHB/CGB* clusters in human and chimpanzee (Figure 6-B,D,E). First, it was the most parsimonious solution requiring the smallest number of rearrangement events. Assuming that the ancestral state of Hu-Ch *LHB/CGB* cluster consisted of the highly identical segments, the current structure of chimpanzee cluster would be explained by only one direct duplication of a region including the *CGB1* gene and most of an *intergenic region B*. In human the current structure would have required two events: inverted duplication of the entire region from *CGB1* to *CGB5* and a direct duplication of *intergenic region C*. This hypothesis was also supported by the results of phylogenetic analysis (Figure 6-D,E; Fig. 2 in Ref. III).

These independent parallel duplications in sister species were possibly initiated by *Alu*-mediated NAHR as *Alu* SINE sequences are found at the potential rearrangement breakpoints in both species (Figure 6-C). This is consistent with several studies indicating a high frequency of non-allelic recombination events and gene conversion between *Alu* sequences leading to genomic rearrangements but also to spreading polymorphisms due to gene conversion (Roy et al. 2000; Sen et al. 2006; Zhi 2007).

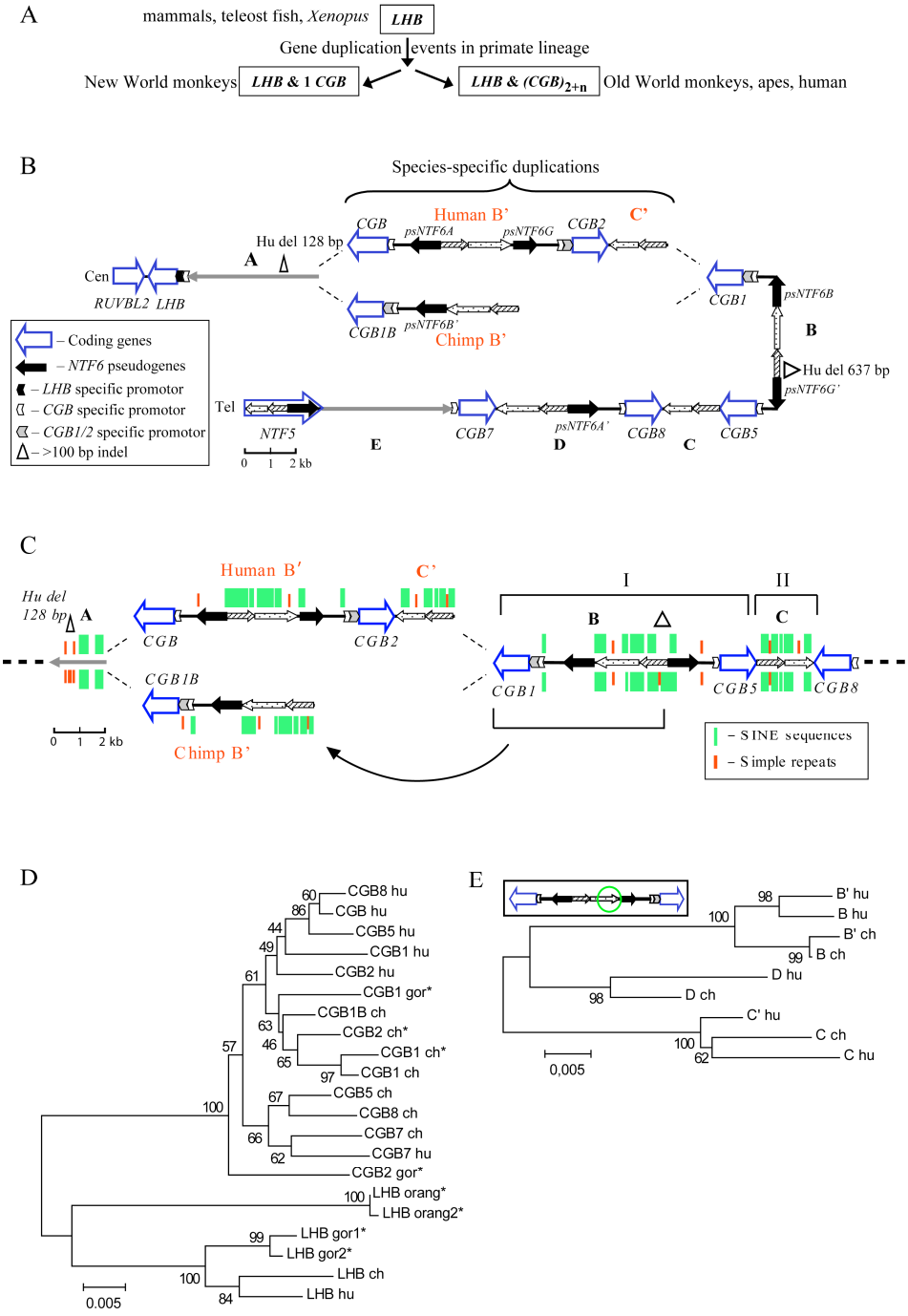

Figure 6. (A) A simplified schematic presentation of the evolution of *LHB/CGB* genes in primates. Duplication of an ancestral *LHB* gene in primates has given rise to a novel gene, *CGB*.

(B) Comparative structure of the human and chimpanzee *LHB/CGB* clusters. Identical color and pattern codes refer to the DNA segments within the cluster with highly similar sequences, direction of the DNA sequence is indicated as on the sense strand.
(C) Location of SINE and simple repeat sequences near the duplication breakpoints of human and chimpanzee *LHB/CGB* genome clusters. Brackets denote the ancestral regions that indepenently duplicated in human (upper) and chimpanzee (lower) giving rise to intergenic regions B' and C'. I and II denote the two duplications of intergenic regions B and C in human genome.
(D) Neighbor-joining trees based on genic and
(E) intergenic regions within the *LHB/CGB* cluster. The full sequences of *LHB/CGB* genes are included from human (hu), chimpanzee (ch), gorilla (gor) and orangutan (orang). (*) denotes sequences from Ref. II, others are from Ref. III. Homologous segments used for the phylogenetic tree are indicated with a green circle on a consensus structure of the intergenic regions in *LHB/CGB* cluster (boxed, from Figure 6B). Bootstrap support values are shown at the nodes (1,000 bootstrap replications).

3.2.3 Divergence and selection in orthologous regions

The total divergence in the whole orthologous region was 5% (2.7% indels, 2.3% nucleotide substitutions, total compared region covered 37,220 bp for chimpanzee and 36,541 bp for human)(Table 5-B). Previous studies have suggested that the majority of genomic divergence between human and chimpanzee comprises of indels (3.0–11.9%) compared to nucleotide substitutions (1.2–1.5%)(Anzai et al. 2003; Britten 2002; Ebersberger et al. 2002; Mikkelsen et al. 2005; Watanabe et al. 2004). Divergence of the orthologous genes *LHB, CGB1, CGB5, CGB8* and *CGB7* ranged from 1.8% in *LHB* to 2.59% in *CGB5* and *CGB8* which exceeds the estimates for single-copy genes even several times. Even after including the intraspecific population variation data of human (n = 95, from Ref I) and chimpanzee (n = 11, unpublished data), the higher divergence compared to published data for single-copy genes remained (Figure 5 and Additional file 3 in Ref. III).

In order to study the evolutionary forces acting on *LHB/CGB* orthologous genes in human and chimpanzee, the non-synonymous (d_n)/ synonymous (d_s) rate ratio (ω) was estimated using the maximum likelihood method (Goldman and Yang 1994). For four of the genes (*LHB, CGB5, CGB8* and *CGB7*) the $\omega < 1$, indicating purifying selection (Table 6). The only locus showing $\omega > 1$ was *CGB1* ($\omega = 2.658$), which would be consistent with positive selection or adaptive evolution. These results are consistent with Reference II suggesting that *CGB1/2*-like genes might be evolving towards a new function in human and chimpanzee, but not in gorilla (Gorilla gorilla) and probably are absent in the genome of orangutan (*Pongo pygmaeus*).

Table 6. Maximum likelihood estimation of ω (=d_n/d_s) values by PAML analysis and amino acid divergence in human and chimpanzee orthologous genes.

Gene	n[a]	ω[b]	AA %[c]
LHB	141	0.087	1.42
CGB1	132	2.658	3.03
CGB5	165	0.180	1.82
CGB8	165	0.099	1.21
CGB7	165	0.378	1.82

[a] number of codons in the sequence
[b] non-synonymous/ synonymous rate ratio, averaged over sites (d_n/d_s)
[c] Divergence at amino acid level (%)

3.2.4 Footprints of intraspecies gene conversion in interspecies analysis

The comparison of human and chimpanzee *LHB/CGB* genome clusters indicated a role of gene conversion in the evolution of both species.

First of all, the higher divergence between human and chimpanzee found in *LHB/CGB* genome cluster and the genes compared to unique genomic regions (Table 5-B) could possibly result from the effect of frequent intraspecies gene conversion events (Figure 3, 5 and Additional file 3 in Ref. III).

Next, the footprint of gene conversion was observed in the phylogenetic tree of human and great apes *LHB/CGB* genes as some genes clustered together within the species instead of forming separate clades for orhtologous genes (Figure 6-D).

Gene conversion might also explain the patterns of nucleotide substitutions between human and chimpanzee in the orthologous *LHB/CGB* region. Transitions (C⇔T, A⇔G) were found to contribute to 62% and 57% of the total substitutions in five orthologous *LHB/CGB* genes and in the whole orthologous region, respectively (Figure 4 in Ref. III), being ~10% lower than reported in previous studies comparing human and chimpanzee genomic regions (68.87% – 70.3%)(Anzai et al. 2003; Ebersberger et al. 2002). The ~10% lower fraction of transitions in the *LHB/CGB* genome cluster contradicts the expectations based on its genomic structure rich in CpG dinucleotides and rapidly diverging *Alu* sequences (Figure 3 and Additional File 1 in Ref. III)(Chen and Li 2001). It is generally accepted that a high proportion of transitions are C to T substitutions in CpG dinucleotides, exhibiting about 10 times higher mutation rate than the genomic average (Mikkelsen et al. 2005; Nachman and Crowell 2000). Active gene conversion between multiple duplicated and highly homologous segments might be one explanation for the higher proportion of tranversions in the *LHB/CGB* cluster as gene conversion has no preference for either transitions or transversions.

3.3 *CGB1* and *CGB2*: most recent members of the *LHB/CGB* clusters (Ref. II, III)

3.3.1 *CGB1/2* genes in human and great apes

Two genes of the human *LHB/CGB* cluster, *CGB1* and *CGB2*, have long been considered to be pseudogenes as they have been predicted to encode for a 132 aa hypothetical protein that has no similarity to hCGβ-subunit nor to any other known protein (Bo and Boime 1992). The DNA sequence identity of human *CGB1* and *CGB2* to hCGβ-subunit coding genes is 85%. The main difference between *CGB1/2* and hCGβ-genes is an inserted DNA fragment that creates a novel *CGB1/2* specific putative promoter fragment, novel exon 1 and leads to one basepair shift in the open reading frame of exons 2 and 3.

To study the evolution of *CGB1* and *CGB2* genes and look for indirect evidence of their functionality, human and three great ape species were studied (common chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*) and orangutan (*Pongo pygmaeus*)). *LHB* gene was used as a reference as it is a functional gene that exists in the genome of all vertebrates. Human-specific PCR primers were capable of amplifying the *LHB* gene and two *CGB1/2*-like genes from the genomic DNA of chimpanzee and gorilla. However, from the genomic DNA of orangutan, only *LHB,* but not the *CGB1/2*-like genes, was amplified. The sequences of the *CGB1/CGB2*-like genes (for chimpanzee and gorilla) and *LHB* genes (for chimpanzee, gorilla and orangutan) have been deposited in GenBank (accession numbers DQ238547-DQ238553)(Appendix 3).

When comparing the sequences of the human and chimpanzee's genes no changes were found leading to ORF shift or preliminary stop-codon in *CGB1* or *CGB2* (ch*CGB1B*) of the chimp. In gorilla, a 1 bp insertion in *CGB1* exon 2 and a 12 bp deletion in exon 2 of *CGB2* were identified presumably leading to the disruption of the predicted protein (Figure 2 in Ref. II).

Based on the results it was suggested that the *CGB1*/*2*-like genes have arisen among the common ancestor of African great apes as two *CGB1*/*2*-like genes were amplified from chimpanzee and gorilla but none from the orangutan. In human and chimpanzee no disruptive changes were found in the studied genes. However, in gorilla both the *CGB1/2*-like genes harbor insertion/deletion changes disrupting the predicted protein. Therefore the genes might potentially be functional in human and chimpanzee, but pseudogenes in gorilla.

3.3.2 Analysis of 5' upstream regions of human *CGB1/2* genes

The 5' upstream regions of human *CGB1* and *CGB2* genes were studied *in silico* for possible transcription factor binding sites using MatInspector 2.2 and Alibaba 2.1 programs (Cartharius et al. 2005; Grabe 2002). *CGB1* and *CGB2*

were found to be lacking 52 bp of the proximal promoter segment of hCGβ containing two Ets-2 binding sites. Other sequence motifs playing a crucial role in regulating hCGβ expression were conserved among the genes (Figure 1B and C in Ref. II).

The *in silico* anaysis predicted the novel *CGB1/2*-specific insert (736 bp for *CGB1* and 724 bp for *CGB2*) to contain several regulatory elements essential for gonadotrope expression and also several transcription factors regulating implantation and placental development (Figure 1D in Ref. II).

4. DISCUSSION

4.1 Methodological challenges in studying duplicated genomic regions

Segmentally duplicated regions of >1kb in size and >90% identity have been estimated to comprise a high proportion of human genome (5.2%) (Bailey et al. 2002; Bailey et al. 2001). Also copy-number variations (~12% of human genome and >1 kb in size) are found to be frequent (Redon et al. 2006). High DNA identities raise several aspects making the studying of such duplicated genomic regions rather challenging.

First of all, problems arise with sequence assembly of duplicated regions. Genomes of many different species have been sequenced and the availability of high-throughput sequencing systems like 454 Sequencing system and SOLiD™ System enable very fast sequencing of whole genomes. Despite the availability of large amount of data, the quality of duplicated regions in databases is often poor due to erroneous assembly of highly identical sequences. For example, the common chimpanzee's (*Pan troglodytes*) genome was already published in 2005 (Mikkelsen et al. 2005) but the database sequences of many duplicated regions, including the *LHB/CGB* cluster, are still of low quality.

The problem rises also with studies based on SNPs starting with initial SNP determination. An apparent enrichment of SNPs in segmental duplications has been found in dbSNP database (Estivill et al. 2002; Fredman et al. 2004; Sengupta et al. 2008). Part of this increase in SNP density is probably due to sequence variation between copies that are falsely interpreted as SNPs. A few years ago a new type of polymorphisms were described in duplicon regions, the multisite variations (MSVs) that are represented as paralogous sequence variants among duplicons (MSV1) or "shared" SNPs located at the same position in several duplicated regions (MSV2)(Fredman et al. 2004). In human *LHB/CGB* genes, 35%-77% of the SNPs were actually identified as being MSVs (Table 1 in Ref. I). It is important to study duplicated regions in detail to characterize the actual diversity patterns in the duplicons. In genomic regions containing copy-number variations and/or segmental duplications many of the highly heterozygous sequence variants stored in public databases are likely to be erroneously recorded MSVs (Fredman et al. 2004; Lindsay et al. 2006; Redon et al. 2006; Sengupta et al. 2008).

Also, duplicated regions are shown to harbor a large number of population-specific variants and low LD due to the activity of gene conversion making it technologically challenging to select reliable tag-SNPs and to establish genotyping methods capable of targeting unique SNPs in duplicated genes (Rull et al. 2008b; Sedman et al. 2008). To be able to design studies for duplicated regions, it is very important to know the correct structure and sequence variation of such clusters.

Several large-scale methods like array-CGH and Illumina HumanCNV370-Duo DNA Analysis BeadChip are available for studying the copy-number differences between individuals and species, but these methods are not capable of giving detailed information on rearrangements and exact structure of the regions under study.

Considering the aspects mentioned above, a resequencing method was used in the current study to characterize human population diversity of *LHB/CGB* genes and conduct a shotgun sequencing of BAC clones to obtain the chimpanzee's *LHB/CGB* genome cluster sequence.

4.2 Diversity in duplicated genes: balance between gene conversion and selection

In *LHB/CGB* genome cluster, active gene conversion may have contributed to both higher variation levels in human populations and also higher interspecies sequence divergence between human and chimpanzees compared to unique regions. However, probably a balance between gene conversion and selection exists in *LHB/CGB* cluster and also in other duplicated clusters. As these genes are very important for reproduction, spreading of non-functionalizing mutations probably cannot be tolerated. Only a few changes in the sequence of *LHB* and *CGB* coding regions that do not affect the reproductive function severely are allowed (Table 4).

Duplicated gene families are often characterized by higher diversity compared to unique genomic regions. Higher than genome average diversity estimates have been found in several duplicated regions like *SMT1A-REP* (Lindsay et al. 2006), AZFa region (Bosch et al. 2004), LCR22 (Pavlicek et al. 2005), *Growth Hormone* gene family (Esteban et al. 2007; Sedman et al. 2008). Also *MHC* region shows both high diversity within and high divergence between species (Anzai et al. 2003; Shiina et al. 2006). In many cases higher variation levels have been associated to the gene conversion activity as it can spread mutations among the members of gene families although other factors like selection are also important in influencing the diversity patterns (Papadakis and Patrinos 1999).

High diversity in *MHC* regions has been explained by different factors including both the concerted evolution of the genes by gene conversion and non-homologous recombination (Jeffreys and May 2004; Ohta 1991; Pease et al. 1993; Zangenberg et al. 1995) and selective forces acting on the biologically important loci (Hughes and Nei 1988; Hughes and Nei 1989b; Piertney and Oliver 2006). Positive selection is well known as an important factor operating on *MHC* loci to maintain *MHC* variation, higher diversity leading to better protection from various pathogens. Several studies have suggested that although gene conversion affects the *MHC* loci, its effects on *MHC* polymorphism is quite minor and the diversity is mainly generated by nucleotide substitution and

positive selection (Gu and Nei 1999; Martinsohn et al. 1999; Nei and Rooney 2005).

Several studies in mammals have indicated that non-crossovers outnumber the crossovers by several times. Also, the current study of gene conversion in *LHB/CGB* cluster found that the rate of gene conversion was several times higher than the recombination rate. The ratio of the two parameters (f) depended inversely on the conversion tract length – as the length of the tract decreased, the estimated rate increased. Including only common SNPs, the maximum likelihood estimate for conversion tract L = 30 bp ranged from 6 (Han) to 16 (Mandenka); whereas for L = 500 bp, f was 0.5 (Han) to 1.5 (Mandenka).

In literature, the ratio of gene conversion to recombination rate has been estimated 1.6 across a 28 Mbp region on chromosome 21 for a mean conversion tract length of 500 bp and f = 9.4 for the mean tract length of 50 bp (Padhukasahasram et al. 2004). Gene conversion rate was found ~1.5 times higher than recombination rate in the *MS32* hotspot on human chromosome 1 and ~9 times higher in the *TAP2* region of MHC (Gay et al. 2007). In the *growth hormone/ chorionic somatomammotropin* genes the gene conversion rate exceeded recombination rate tens to even hundreds of times (Sedman et al. 2008).

4.3 Divergence and evolution of primate-specific duplicated genes

The current study found that human and chimpanzee differ in the *CGB* gene number: humans having six *CGB* copies compared to five copies in chimpanzee. Independent duplications in human and chimpanzee explain best the contemporary structure of the cluster. It is possible that the duplications found in the human and chimpanzee *LHB/CGB* cluster are mediated by NAHR between repetitive sequences as *Alu* SINE sequences were found at the rearrangement breakpoints in both species (Figure 6-C).

It has been suggested that many of such duplications may have been facilitated by NAHR events between *Alu* sequences, expanded into millions of copies all over primate genomes (Fitch et al. 1991; Kulski et al. 1997). Detailed analyses of the boundaries of segmental duplications have revealed an enrichment of *Alu* sequences at the breakpoints and are restricted to younger subfamilies (*Alu*Y and *Alu*S) that have emerged recently during primate evolution (Bailey et al. 2003; Johnson et al. 2006a). Also the breakpoints of human copy-number variations have been found to be enriched in *Alu* repeat elements (de Smith et al. 2008).

There are several examples of independent duplication events in primate lineage. For example, the OWM and apes have three Opsin genes and are trichromats due to gene duplication at the base of OWM lineage. In NWM, the situation is more variable: most species exhibit two Opsin genes, but in the howler monkey an additional gene duplication has led to full trichromacy (Hunt

et al. 1998; Kainz et al. 1998). In *Growth Hormone* gene cluster most mammals including prosimians have a single gene (*GH1*) encoding for the pituitary growth hormone, whereas in anthropoids five to eight *GH*-related genes arisen through successive events of duplication have been described (Li et al. 2005; Ye et al. 2005). Gene duplications that gave rise to the New World monkey and Old World monkey/ hominoid *GH* gene family occurred independently after the division of NWM and OWM lineages (Li et al. 2005; Wallis and Wallis 2002). Other gene clusters with independent gene duplications in humans, apes and macaque lineages are *MHC* and testes-expressed *PRAME* genes (Birtle et al. 2005; Boyson et al. 1996; Chen et al. 1992; Gibbs et al. 2007). In addition to the structural differences between human and chimpanzee *LHB/CGB* genes reported in this study, an expansion of *CGB* genes up to 50 gene copies has been shown in gorilla (Dumas et al. 2007; Fortna et al. 2004).

Different evolutionary patterns can be seen in duplicated gene families. For example, the *MHC* class I and II gene families include a large number of loci and have been shown to evolve according to the birth-and-death process (Hughes and Nei 1989a; Nei et al. 1997). As a result, these multigene families consist of a mixture of different genes. In class II genes the rate of birth-and-death is relatively slow. It has been estimated that most of these loci originated at least 170–200 MYA (Takahashi et al. 2000). Loci orthologous to the human class II genes have been identified in addition to apes, Old World monkeys, New World monkeys and prosimians also in mice (Go et al. 2003). The class I loci have gone through the birth-and-death evolution at much faster rate: rapidly expanded and contracted in evolutionary time and the number of loci varies from six in the pig (Singer et al. 1982) to over a thousand in the African mouse (Delarbre et al. 1992). As a result there seem to be no orthologous relationships of different class I loci among different mammalian orders. The divergence occurred so recently, that even humans and New World monkeys, which diverged only about 33–35 MYA, do not share functional genes (Cadavid et al. 1997; Hughes and Nei 1989a; Watkins et al. 1990).

In addition to creating structural divergence among the species, duplications provide also the basis for diversification of gene functions. For primate-specific gene duplications, there is evidence of variability in evolutionary rates among the gene copies within and among the species, and in different selective constraints acting on different members of the gene clusters such as *MHC*, *beta-globin*, *GH/CSH*, *PRAME*, *Rh blood group*, *CD33rSiglec* and *beta-defensin* genes (Aguileta et al. 2004; Aguileta et al. 2006; Angata et al. 2004; Birtle et al. 2005; Gibbs et al. 2007; Hughes and Nei 1988; Li et al. 2005; Salvignol et al. 1993; Sedman et al. 2008; Semple et al. 2003).

4.4 Duplications as a source of new genes

After duplication gene copies can undergo three possible routes: (i) losing an original function due to accumulation of deleterious mutations (nonfunctionalization), (ii) gaining a new function under positive selection for advantageous mutations (neofunctionalization) or (iii) partitioning of original functions among the copies (subfunctionalization)(Force et al. 1999; Hughes 1994; Ohno 1970; Walsh 1995).

4.4.1 Emergence of primate-specific placental hormone CG

Duplication of the ancestral *LHB* gene in the common ancestor of anthropoid primates gave rise to novel gene *CGB* and in contemporary primate species the *CGB* copy numbers vary from one in New World primate species to six in human and even up to ~50 copies in gorilla (Table 3)(Fortna et al. 2004; Maston and Ruvolo 2002). In the *LHB/CGB* genes, the prerequisite for a rise of a novel functional gene was only a 1 bp deletion in the last exon of the duplicated *LHB* gene leading to a frame-shift and elongation of the open reading frame into the previously 3' untranslated region (Fiddes and Goodman 1980). In horses a similar C-terminal extended domain was observed by a 10 bp deletion in the third exon, which occurred independently from primate 1 bp deletion. Interestingly, the *LHB* gene is not duplicated in horses, but two hormones are obtained from the same gene by differences in post-translational modifications (Sherman et al. 1992). The novel *CGB* gene in primates encodes for the β-subunit of chorionic gonadotropin CG, a new primate-specific placental hormone. CG hormone differs from the luteinizing hormone by the time and tissue of expression, circulating half-life and biological function (Fiddes and Talmadge 1984; Moyle and Campbell 1996; Themmen and Huhtaniemi 2000).

Interestingly, in New World monkeys with only one copy of *CGB* gene the evolution has led to a totally different situation compared to Old World Monkeys, apes and human. Based on studies of common marmoset it has been shown that the novel chorionic gonadotropin hormone has completely taken over the role of luteinizing hormone. Although both *LHB* and *CGB* genes are present at the genomic level, at mRNA and protein levels the marmoset does not have *LHB* and only *CGB* gene is expressed in the pituitary and placental tissues (Gromoll et al. 2003; Muller et al. 2004; Simula et al. 1995). Thus in marmosets the CG is the only gonadotropin with luteinizing function present in the pituitary. Whether these results apply to other NWM species is to be studied further.

4.4.2 *CGB1/CGB2*-like genes: recent duplicates with unknown function

The *CGB1/CGB2*-like genes are regarded as the youngest members of the *LHB/CGB* genome cluster. In human these genes have long been considered to be pseudogenes, they differ from the hCGβ-genes by an inserted DNA fragment creating a novel *CGB1/2*-specific putative promoter fragment, novel exon 1 and leading to one basepair shifted open reading frame for exons 2 and 3.

The results of Ref. II suggest that *CGB1/2*-like gene probably arose first in the common ancestor of African great apes. However, the divergence calculations between human, chimpanzee and gorilla *CGB1/2* genes in Ref. II should be taken with caution. Based on Ref. III the human *CGB2* and chimpanzee *CGB1B* (defined as chimpanzee *CGB2* in Ref. II) are not orthologous but paralogous genes independently duplicated within both species. At least two *CGB1/2*-like genes obviously also exist in the genome of gorilla but unless the whole cluster is sequenced, the real copy number and ancestry of the *CGB1/2*-like genes in gorilla will remain unclear. Amplification of these genes failed from the genome of orangutan and the *in silico* search for the *CGB1/2*-like genes from the OWM rhesus macaque (*Macaca mulatta*, Genbank accession no. AC202849) was unsuccessful.

The mRNAs of *CGB1* and *CGB2* have been detected in human placenta, pituitary, testis, prostate, ovary, thymus, kidney and colon (Berger et al. 1994; Bo and Boime 1992; Dirnhofer et al. 1996; Hollenberg et al. 1994; Miller-Lindholm et al. 1997; Rull et al. 2008a; Rull and Laan 2005). Contribution of *CGB1* and *CGB2* to the summarized expression of the six *CGB* genes in human placenta was found to be as low as 1/1000 to 1/10,000 (Rull et al. 2008a; Rull and Laan 2005). Higher expression of *CGB1* and *CGB2* genes was detected during the first trimester of pregnancy compared to the second and third trimesters (Rull et al. 2008a; Rull and Laan 2005), consistent with *in silico* predictions of transcription factors in Ref. II suggesting that these genes might have a role in implantation and placentation. Interestingly, in testis the proportional contribution of *CGB1/2* to the total hCGβ transcript pool was as high as 1/3 (Rull et al. 2008a).

Unlike the hCG beta-coding genes that each produce a single mRNA transcript, four splice variants have been described for *CGB1* and *CGB2* (Bo and Boime 1992). The major transcript of *CGB1/CGB2* codes for a unique 5' UTR and exon 1, and one base pair shifted open reading frame (ORF) for exon 2 and exon 3. The alternative mRNAs contain additional +47, +166 or +176 bp of DNA sequence from intron 1 that alter the amino acid sequence of transcript. Approximately equal fractions of the *CGB1/2* transcript coding for a hypothetical 132 aa protein and the alternative +176 mRNA containing a premature stop-codon (60 aa) were detected in human placenta and testis, +47 form being transcribed at borderline of detection limit (Rull et al. 2008a). However, it is not clear if these transcripts might have a regulatory role at RNA-level or are translated into a protein.

Interestingly, a recent study identified two novel small genes located within the *CGB1/2* specific putative promoter region (Figure 7)(Parrott and Mathews 2007). These genes encode for non-coding RNA substrates for a nuclear factor 90 (NF90) protein and are called snaR (small NF90-associated RNA). The NF90 protein family members are double-stranded RNA binding proteins that participate in many aspects of vertebrate RNA metabolism, have been implicated in development, cell cycle and virus infection. The novel genes are located 86 bp and 85 bp upstream of 5' ends of *CGB1* and *CGB2* genes and are named *SNAR-G1* (128 bp) and *SNAR-G2* (119 bp), respectively. From a set of 19 human tissues, the snaRs were detected predominantly in testis, weak expression was found in placenta and brain. SnaR orthologs are present in chimpanzee but not in other mammals and their function is currently unknown (Parrott and Mathews 2007). Thus, the expressional activity of *CGB1* and *CGB2* in testis may result as a by-product from the transcriptional activation of *SNAR-G1* and *SNAR-G2* genes. Similarly, the transcription of *CGB1* and *CGB2* in placenta could be associated with concerted transcriptional activation of the other *CGB* genes through simultaneous chromatin modifications and/or by highly homologous DNA sequences (>95%) in the promoter regions of *CGB* genes (Rull et al. 2008a). To date, all three options are possible: *CGB1/2* genes are pseudogenes or protein coding loci or a source of regulatory RNAs. Further studies are necessary to study the functionality of *CGB1*/*2* genes.

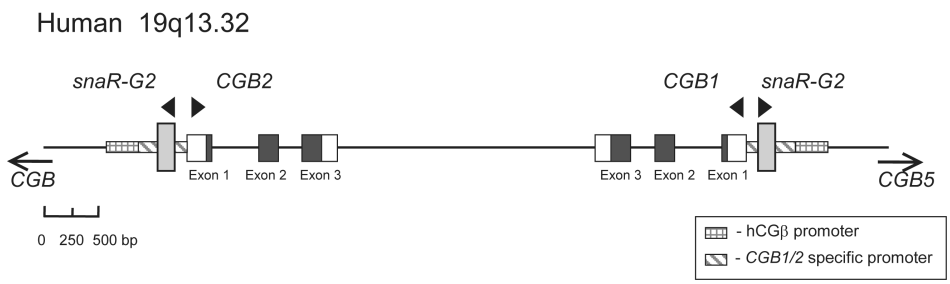

Figure 7. Location of the novel *SNAR-G1* and *SNAR-G2* genes respective to *CGB1* and *CGB2* in human *LHB/CGB* cluster. 5' and 3' UTR's are denoted as white boxes, coding segments of *CGB1/2* genes as dark gray boxes. Light gray boxes mark the novel *SNAR-G1* and *SNAR-G2* genes. The direction of transcription is indicated by black triangles.

4.4.3 Additional examples of novel functions evolving after gene duplication

Examples of novel or altered functions arising after duplication event(s) could be drawn from other gene families. For example, *globin* paralogs have explored diverse evolutionary pathways, with some functional genes retaining their original function, some having become nonfunctional and some having changed

their function and time of expression (Aguileta et al. 2004; Fitch et al. 1991). After the ancestor of the anthropoids diverged from those of prosimians, developmental expression of *γ-globins* changed from embryonic to fetal in the emerging anthropoids (Goodman et al. 1987; Tagle et al. 1988). The *γ-globin* genes continued to function as embryonic genes in rodents, lagomorphs and prosimian primates, in the ancestral anthropoid lineage however they act as fetally expressed genes (Fitch et al. 1991; Johnson et al. 2006b; Tagle et al. 1988).

Also, in the higher primates a rapid evolution of growth hormone gene has occurred. In great apes and Old World monkeys the duplicated *GH* loci, apart from a couple of *GH*-related genes, have acquired a novel function and code for chorionic somatomammotropin (*CSH* genes), also known as placental lactogen expressed in the placenta (Chen et al. 1989; Li et al. 2005; Ye et al. 2005).

A good example of how a gene duplication could be a source of novel protein functions are the two ribonuclease genes in the colobine monkey the douc langur (*Pygathrix nemaeus*). They are unique among primates in using leaves rather than fruits and insects as their primary food source, the leaves are fermented in the foregut by symbiotic bacteria. Similar to ruminants, colobines recover nutrients by breaking and digesting the bacteria with the use of various enzymes, including pancreatic *RNase1*. A recent duplication (~4 MYA) of the *RNase1* gene was identified in douc langur, that occurred after the divergence of the colobine lineage from the other Old World monkeys (Zhang et al. 2002). The novel ribonuclease is effective in degrading bacterial RNAs at lower pH values of colobine's small intestine (pH 6–7 compared to pH 7.4–8 in humans and other monkeys), whereas the *RNase1* retained its former function (Zhang et al. 2002).

In addition, evidence of parallel adaptive evolution of digestive RNases in Asian and African leaf monkeys was recently found (Zhang 2006). Colobine monkeys are separated into Asian and African clades, which diverged from each other ~13 MYA. In addition to Asian colobine the douc langur (*Pygathrix nemaeus*), Zhang studied *RNase* genes of African colobine the guereza monkey (*Colobus guereza*). The guereza monkey has three genes: *RNase1* and independently duplicated *RNase1β* and *RNase1γ* genes. It was suggested that the duplication events occurred 6.7 and 7.8 MYA The independently generated douc langur *RNase1B* and guereza *RNase1β* and *RNase1γ* genesshow very similar patterns of sequence evolution: role of positive selection leading to lower optimal pH compared to the ancestral RNase1 gene (7.4 for *RNase1*, 6.3 for *RNase1B* of douc langur, 6.7 for guereza *RNase1β* and *RNase1γ*) and losing the ability to degrade double-stranded RNA. The pancreatic *RNase1* gene was duplicated independently in Asian and African colobines and the duplicated genes subsequently experienced parallel functional changes by means of parallel amino acid replacements driven by selective pressure. Additional duplications of *RNase1* gene in Asian and African colobines have been reported by Schienman et al. (Schienman et al. 2006).

CONCLUSIONS

Following conclusions can be drawn from the current PhD thesis:

1. Duplicated and highly identical *LHB/CGB* genes are characterized by a very high diversity in three human populations (Estonians, Chinese Han and African Mandenkalu). High sequence diversity, strong LD and concentration of gene conversion acceptor sites co-localized at the periphery of the cluster (*LHB*, *CGB* and *CGB7* genes); lower diversity, breakdown of LD and gene conversion donor activity was characteristic to the central region (*CGB2*, *CGB1* and *CGB5* genes). The results indicate an important role of gene conversion in spreading polymorphisms among the duplicon copies and generating LD around them. The directionality of gene conversion events might be associated to the predicted recombination "hotspot" and "warm spot" in the vicinity of the most active acceptor genes at the periphery of the cluster.
2. Comparison of human and chimpanzee *LHB/CGB* genomic regions revealed independent duplication events in these species resulting in discordant number of *CGB* genes (6 in human, 5 in chimpanzee). A difference in *CGB* copy numbers between species indicates active dynamics of *LHB/CGB* genomic region and suggests potential susceptibility to intraspecies rearrangement events.
3. Active gene conversion may have contributed to higher interspecies sequence divergence (both genic and intergenic) and altered transition/ transversion ratio compared to single-copy loci. Probably a balance exists between gene conversion and selection in the studies genes. Despite a high sequence identity (85–99%), the individual *LHB/CGB* genes might be evolving under different selective constraints.
4. Study of the *CGB1/CGB2*-like genes in humans and great apes (chimpanzee, gorilla and orangutan) indicated that these genes may have arisen in the common ancestor of African great apes. In human and chimpanzee these genes might be functional as no ORF disturbing mutations were identified. In gorilla however, the *CGB1/CGB2*-like genes might be pseudogenes as in both of them insertion/deletion events were found disrupting the predicted protein.

Implications and further development of the study:
It is evident that duplicated genes and genomic segments have an important role in intra- and interspecies genetic variation and in a number of disorders. Detailed studies of such regions aid in understanding the mechanisms of past and ongoing gene and genome evolution. Understanding the dynamics of the duplicated gene clusters might help to predict the regions in human genome susceptible to rearrangement events and therefore prone to genomic disorders.

REFERENCES

Aburomia, R., O. Khaner, and A. Sidow. 2003. Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. *J Struct Funct Genomics* **3:** 45–52.

Aguileta, G., J.P. Bielawski, and Z. Yang. 2004. Gene conversion and functional divergence in the beta-globin gene family. *J Mol Evol* **59:** 177–189.

Aguileta, G., J.P. Bielawski, and Z. Yang. 2006. Evolutionary rate variation among vertebrate beta globin genes: implications for dating gene family duplication events. *Gene* **380:** 21–29.

Amato, F., A.P. Simula, L.J. Gameau, and R.J. Norman. 1998. Expression, characterisation and immunoassay of recombinant marmoset chorionic gonadotrophin dimer and beta-subunit. *J Endocrinol* **159:** 141–151.

Andolfatto, P. and M. Nordborg. 1998. The effect of gene conversion on intralocus associations. *Genetics* **148:** 1397–1399.

Angata, T., E.H. Margulies, E.D. Green, and A. Varki. 2004. Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. *Proc Natl Acad Sci U S A* **101:** 13251–13256.

Anzai, T., T. Shiina, N. Kimura, K. Yanagiya, S. Kohara, A. Shigenari, T. Yamagata, J.K. Kulski, T.K. Naruse, Y. Fujimori et al. 2003. Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proc Natl Acad Sci U S A* **100:** 7708–7713.

Ardlie, K., S.N. Liu-Cordero, M.A. Eberle, M. Daly, J. Barrett, E. Winchester, E.S. Lander, and L. Kruglyak. 2001. Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *Am J Hum Genet* **69:** 582–589.

Armstrong, E.G., P.H. Ehrlich, S. Birken, J.P. Schlatterer, E. Siris, W.C. Hembree, and R.E. Canfield. 1984. Use of a highly sensitive and specific immunoradiometric assay for detection of human chorionic gonadotropin in urine of normal, nonpregnant, and pregnant individuals. *J Clin Endocrinol Metab* **59:** 867–874.

Backstrom, N., H. Ceplitis, S. Berlin, and H. Ellegren. 2005. Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. *Mol Biol Evol* **22:** 1992–1999.

Bailey, J.A., Z. Gu, R.A. Clark, K. Reinert, R.V. Samonte, S. Schwartz, M.D. Adams, E.W. Myers, P.W. Li, and E.E. Eichler. 2002. Recent segmental duplications in the human genome. *Science* **297:** 1003–1007.

Bailey, J.A., G. Liu, and E.E. Eichler. 2003. An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* **73:** 823–834.

Bailey, J.A., A.M. Yavor, H.F. Massa, B.J. Trask, and E.E. Eichler. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* **11:** 1005–1017.

Bailey, W.J., D.H. Fitch, D.A. Tagle, J. Czelusniak, J.L. Slightom, and M. Goodman. 1991. Molecular evolution of the psi eta-globin gene locus: gibbon phylogeny and the hominoid slowdown. *Mol Biol Evol* **8:** 155–184.

Benovoy, D. and G. Drouin. 2008. Ectopic gene conversions in the human genome. *Genomics*.

Berger, K., A.E. Billerbeck, E.M. Costa, L.S. Carvalho, I.J. Arnhold, and B.B. Mendonca. 2005. Frequency of the allelic variant (Trp8Arg/Ile15Thr) of the luteinizing hormone gene in a Brazilian cohort of healthy subjects and in patients with hypogonadotropic hypogonadism. *Clinics* **60:** 461–464.

Berger, P., W. Kranewitter, S. Madersbacher, R. Gerth, S. Geley, and S. Dirnhofer. 1994. Eutopic production of human chorionic gonadotropin beta (hCG beta) and luteinizing hormone beta (hLH beta) in the human testis. *FEBS Lett* **343:** 229–233.

Bettencourt, B.R. and M.E. Feder. 2002. Rapid concerted evolution via gene conversion at the Drosophila hsp70 genes. *J Mol Evol* **54:** 569–586.

Birtle, Z., L. Goodstadt, and C. Ponting. 2005. Duplication and positive selection among hominin-specific PRAME genes. *BMC Genomics* **6:** 120.

Bo, M. and I. Boime. 1992. Identification of the transcriptionally active genes of the chorionic gonadotropin beta gene cluster in vivo. *J Biol Chem* **267:** 3179–3184.

Bosch, E., M.E. Hurles, A. Navarro, and M.A. Jobling. 2004. Dynamics of a human interparalog gene conversion hotspot. *Genome Res* **14:** 835–844.

Bosch, N., M. Caceres, M.F. Cardone, A. Carreras, E. Ballana, M. Rocchi, L. Armengol, and X. Estivill. 2007. Characterization and evolution of the novel gene family FAM90A in primates originated by multiple duplication and rearrangement events. *Hum Mol Genet* **16:** 2572–2582.

Bowers, J.E., B.A. Chapman, J. Rong, and A.H. Paterson. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422:** 433–438.

Boyson, J.E., C. Shufflebotham, L.F. Cadavid, J.A. Urvater, L.A. Knapp, A.L. Hughes, and D.I. Watkins. 1996. The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. *J Immunol* **156:** 4656–4665.

Britten, R.J. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc Natl Acad Sci U S A* **99:** 13633–13635.

Brizot, M.L., E. Jauniaux, A.T. McKie, F. Farzaneh, and K.H. Nicolaides. 1996. Placental mRNA expression of alpha and beta human chorionic gonadotrophin in early trisomy 18 pregnancies. *Mol Hum Reprod* **2:** 463–465.

Brizot, M.L., R.J. Snijders, J. Butler, N.A. Bersinger, and K.H. Nicolaides. 1995. Maternal serum hCG and fetal nuchal translucency thickness for the prediction of fetal trisomies in the first trimester of pregnancy. *Br J Obstet Gynaecol* **102:** 127–132.

Brown, T.C. and J. Jiricny. 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. *Cell* **54:** 705–711.

Cadavid, L.F., C. Shufflebotham, F.J. Ruiz, M. Yeager, A.L. Hughes, and D.I. Watkins. 1997. Evolutionary instability of the major histocompatibility complex class I loci in New World primates. *Proc Natl Acad Sci U S A* **94:** 14536–14541.

Cartharius, K., K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, and T. Werner. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21:** 2933–2942.

Chang, Y.S., C.J. Huang, F.L. Huang, and T.B. Lo. 1988. Primary structures of carp gonadotropin subunits deduced from cDNA nucleotide sequences. *Int J Pept Protein Res* **32:** 556–564.

Chen, E.Y., Y.C. Liao, D.H. Smith, H.A. Barrera-Saldana, R.E. Gelinas, and P.H. Seeburg. 1989. The human growth hormone locus: nucleotide sequence, biology, and evolution. *Genomics* **4:** 479–497.

Chen, F.C. and W.H. Li. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* **68:** 444–456.

Chen, J.M., N. Chuzhanova, P.D. Stenson, C. Ferec, and D.N. Cooper. 2005a. Complex gene rearrangements caused by serial replication slippage. *Hum Mutat* **26:** 125–134.

Chen, J.M., N. Chuzhanova, P.D. Stenson, C. Ferec, and D.N. Cooper. 2005b. Intrachromosomal serial replication slippage in trans gives rise to diverse genomic rearrangements involving inversions. *Hum Mutat* **26:** 362–373.

Chen, J.M., N. Chuzhanova, P.D. Stenson, C. Ferec, and D.N. Cooper. 2005c. Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. *Hum Mutat* **25:** 207–221.

Chen, J.M., D.N. Cooper, N. Chuzhanova, C. Ferec, and G.P. Patrinos. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet* **8:** 762–775.

Chen, Z.W., S.N. McAdam, A.L. Hughes, A.L. Dogon, N.L. Letvin, and D.I. Watkins. 1992. Molecular cloning of orangutan and gibbon MHC class I cDNA. The HLA-A and -B loci diverged over 30 million years ago. *J Immunol* **148:** 2547–2554.

Cheng, Z., M. Ventura, X. She, P. Khaitovich, T. Graves, K. Osoegawa, D. Church, P. DeJong, R.K. Wilson, S. Paabo et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437:** 88–93.

Chopineau, M., N. Martinat, C. Pourchet, F. Stewart, Y. Combarnous, and F. Guillou. 1999. Cloning, sequencing and functional expression of zebra (Equus burchelli) LH. *J Reprod Fertil* **115:** 159–166.

Chopineau, M., F. Stewart, and W.R. Allen. 1995. Cloning and analysis of the cDNA encoding the horse and donkey luteinizing hormone beta-subunits. *Gene* **160:** 253–256.

Clements, J.A., F.I. Reyes, J.S. Winter, and C. Faiman. 1976. Studies on human sexual development. III. Fetal pituitary and serum, and amniotic fluid concentrations of LH, CG, and FSH. *J Clin Endocrinol Metab* **42:** 9–19.

Cole, L.A. 2007. Hyperglycosylated hCG. *Placenta* **28:** 977–986.

Cole, L.A. and S.A. Khanlian. 2007. Hyperglycosylated hCG: a variant with separate biological functions to regular hCG. *Mol Cell Endocrinol* **260–262:** 228–236.

Cole, L.A., J.M. Sutton, T.N. Higgins, and G.S. Cembrowski. 2004. Between-method variation in human chorionic gonadotropin test results. *Clin Chem* **50:** 874–882.

Conant, G.C. and A. Wagner. 2003. Asymmetric sequence divergence of duplicate genes. *Genome Res* **13:** 2052–2058.

Conrad, B. and S.E. Antonarakis. 2007. Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu Rev Genomics Hum Genet* **8:** 17–35.

Crawford, D.C., T. Bhangale, N. Li, G. Hellenthal, M.J. Rieder, D.A. Nickerson, and M. Stephens. 2004. Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet* **36:** 700–706.

De Bodt, S., S. Maere, and Y. Van de Peer. 2005. Genome duplication and the origin of angiosperms. *Trends Ecol Evol* **20:** 591–597.

de Smith, A.J., R.G. Walters, L.J. Coin, I. Steinfeld, Z. Yakhini, R. Sladek, P. Froguel, and A.I. Blakemore. 2008. Small deletion variants have stable breakpoints commonly associated with alu elements. *PLoS ONE* **3:** e3104.

Dehal, P. and J.L. Boore. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* **3:** e314.

Delarbre, C., C. Jaulin, P. Kourilsky, and G. Gachelin. 1992. Evolution of the major histocompatibility complex: a hundred-fold amplification of MHC class I genes in the African pigmy mouse Nannomys setulosus. *Immunogenetics* **37:** 29–38.

Demuth, J.P., T. De Bie, J.E. Stajich, N. Cristianini, and M.W. Hahn. 2006. The evolution of mammalian gene families. *PLoS ONE* **1:** e85.

Des Marais, D.L. and M.D. Rausher. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454:** 762–765.

Dias, J.A., B.D. Cohen, B. Lindau-Shepard, C.A. Nechamen, A.J. Peterson, and A. Schmidt. 2002. Molecular, structural, and cellular biology of follitropin and follitropin receptor. *Vitam Horm* **64:** 249–322.

Diebel, N.D. and E.M. Bogdanove. 1978. Analysis of luteinizing hormone and follicle-stimulating hormone release kinetics during a dynamic secretory event, the postpartum preovulatory surge in the rat, based on quantitative changes in stored and circulating luteinizing hormone and follicle-stimulating hormone and metabolic clearance data for these hormones. *Endocrinology* **103:** 665–673.

Dirnhofer, S., M. Hermann, A. Hittmair, R. Hoermann, K. Kapelari, and P. Berger. 1996. Expression of the human chorionic gonadotropin-beta gene cluster in human pituitaries and alternate use of exon 1. *J Clin Endocrinol Metab* **81:** 4212–4217.

Donoghue, P.C. and M.A. Purnell. 2005. Genome duplication, extinction and vertebrate evolution. *Trends Ecol Evol* **20:** 312–319.

Drouin, G. 2002. Characterization of the gene conversions between the multigene family members of the yeast genome. *J Mol Evol* **55:** 14–23.

Dumas, L., Y.H. Kim, A. Karimpour-Fard, M. Cox, J. Hopkins, J.R. Pollack, and J.M. Sikela. 2007. Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res* **17:** 1266–1277.

Ebersberger, I., D. Metzler, C. Schwarz, and S. Paabo. 2002. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am J Hum Genet* **70:** 1490–1497.

Eikenboom, J.C., T. Vink, E. Briet, J.J. Sixma, and P.H. Reitsma. 1994. Multiple substitutions in the von Willebrand factor gene that mimic the pseudogene sequence. *Proc Natl Acad Sci U S A* **91:** 2221–2224.

Elter, K., C.T. Erel, N. Cine, U. Ozbek, B. Hacihanefioglu, and E. Ertungealp. 1999. Role of the mutations Trp8 => Arg and Ile15 => Thr of the human luteinizing hormone beta-subunit in women with polycystic ovary syndrome. *Fertil Steril* **71:** 425–430.

Esteban, C., L. Audi, A. Carrascosa, M. Fernandez-Cancio, A. Perez-Arroyo, A. Ulied, P. Andaluz, R. Arjona, M. Albisu, M. Clemente et al. 2007. Human growth hormone (GH1) gene polymorphism map in a normal-statured adult population. *Clin Endocrinol (Oxf)* **66:** 258–268.

Estivill, X., J. Cheung, M.A. Pujana, K. Nakabayashi, S.W. Scherer, and L.C. Tsui. 2002. Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome. *Hum Mol Genet* **11:** 1987–1995.

Eyre-Walker, A. 1993. Recombination and mammalian genome evolution. *Proc Biol Sci* **252:** 237–243.

Ezawa, K., O.O. S, and N. Saitou. 2006. Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Mol Biol Evol* **23:** 927–940.

Fan, Q.R. and W.A. Hendrickson. 2005. Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* **433:** 269–277.

Ferguson, D.O. and W.K. Holloman. 1996. Recombinational repair of gaps in DNA is asymmetric in Ustilago maydis and can be explained by a migrating D-loop model. *Proc Natl Acad Sci U S A* **93:** 5419–5424.

Fiddes, J.C. and H.M. Goodman. 1980. The cDNA for the beta-subunit of human chorionic gonadotropin suggests evolution of a gene by readthrough into the 3'-untranslated region. *Nature* **286:** 684–687.

Fiddes, J.C. and K. Talmadge. 1984. Structure, expression, and evolution of the genes for the human glycoprotein hormones. *Recent Prog Horm Res* **40:** 43–78.

Fitch, D.H., W.J. Bailey, D.A. Tagle, M. Goodman, L. Sieu, and J.L. Slightom. 1991. Duplication of the gamma-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. *Proc Natl Acad Sci U S A* **88:** 7396–7400.

Force, A., M. Lynch, F.B. Pickett, A. Amores, Y.L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151:** 1531–1545.

Fortna, A., Y. Kim, E. MacLaren, K. Marshall, G. Hahn, L. Meltesen, M. Brenton, R. Hink, S. Burgers, T. Hernandez-Boussard et al. 2004. Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol* **2:** E207.

Fox, K.M., J.A. Dias, and P. Van Roey. 2001. Three-dimensional structure of human follicle-stimulating hormone. *Mol Endocrinol* **15:** 378–389.

Fredman, D., S.J. White, S. Potter, E.E. Eichler, J.T. Den Dunnen, and A.J. Brookes. 2004. Complex SNP-related sequence variation in segmental genome duplications. *Nat Genet* **36:** 861–866.

Frisse, L., R.R. Hudson, A. Bartoszewicz, J.D. Wall, J. Donfack, and A. Di Rienzo. 2001. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am J Hum Genet* **69:** 831–843.

Fumasoni, I., N. Meani, D. Rambaldi, G. Scafetta, M. Alcalay, and F.D. Ciccarelli. 2007. Family expansion and gene rearrangements contributed to the functional specialization of PRDM genes in vertebrates. *BMC Evol Biol* **7:** 187.

Galtier, N. 2003. Gene conversion drives GC content evolution in mammalian histones. *Trends Genet* **19:** 65–68.

Galtier, N. and L. Duret. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet* **23:** 273–277.

Galtier, N., G. Piganeau, D. Mouchiroud, and L. Duret. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* **159:** 907–911.

Gay, J., S. Myers, and G. McVean. 2007. Estimating meiotic gene conversion rates from population genetic data. *Genetics* **177:** 881–894.

Gerhard, I. and B. Runnebaum. 1984. Predictive value of hormone determinations in the first half of pregnancy. *Eur J Obstet Gynecol Reprod Biol* **17:** 1–17.

Gibbs, R.A. J. Rogers M.G. Katze R. Bumgarner G.M. Weinstock E.R. Mardis K.A. Remington R.L. Strausberg J.C. Venter R.K. Wilson et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316:** 222–234.

Giordano, M., C. Marchetti, E. Chiorboli, G. Bona, and P. Momigliano Richiardi. 1997. Evidence for gene conversion in the generation of extensive polymorphism in the promoter of the growth hormone gene. *Hum Genet* **100:** 249–255.

Go, Y., Y. Satta, Y. Kawamoto, G. Rakotoarisoa, A. Randrianjafy, N. Koyama, and H. Hirai. 2003. Frequent segmental sequence exchanges and rapid gene duplication characterize the MHC class I genes in lemurs. *Immunogenetics* **55:** 450–461.

Goldman, N. and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11:** 725–736.

Golos, T.G., M. Durning, J.M. Fisher, and P.D. Fowler. 1993. Cloning of four growth hormone/chorionic somatomammotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta. *Endocrinology* **133:** 1744–1752.

Goodman, M., J. Czelusniak, B.F. Koop, D.A. Tagle, and J.L. Slightom. 1987. Globins: a case study in molecular phylogeny. *Cold Spring Harb Symp Quant Biol* **52:** 875–890.

Grabe, N. 2002. AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol* **2:** S1–15.

Graham, M.Y., T. Otani, I. Boime, M.V. Olson, G.F. Carle, and D.D. Chaplin. 1987. Cosmid mapping of the human chorionic gonadotropin beta subunit genes by field-inversion gel electrophoresis. *Nucleic Acids Res* **15:** 4437–4448.

Gromoll, J., U. Eiholzer, E. Nieschlag, and M. Simoni. 2000. Male hypogonadism caused by homozygous deletion of exon 10 of the luteinizing hormone (LH) receptor: differential action of human chorionic gonadotropin and LH. *J Clin Endocrinol Metab* **85:** 2281–2286.

Gromoll, J., T. Ried, H. Holtgreve-Grez, E. Nieschlag, and T. Gudermann. 1994. Localization of the human FSH receptor to chromosome 2 p21 using a genomic probe comprising exon 10. *J Mol Endocrinol* **12:** 265–271.

Gromoll, J., J. Wistuba, N. Terwort, M. Godmann, T. Muller, and M. Simoni. 2003. A new subclass of the luteinizing hormone/chorionic gonadotropin receptor lacking exon 10 messenger RNA in the New World monkey (Platyrrhini) lineage. *Biol Reprod* **69:** 75–80.

Gu, W., F. Zhang, and J.R. Lupski. 2008. Mechanisms for human genomic rearrangements. *Pathogenetics* **1:** 4.

Gu, X. and M. Nei. 1999. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol Biol Evol* **16:** 147–156.

Gu, X., Y. Wang, and J. Gu. 2002a. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat Genet* **31:** 205–209.

Gu, Z., A. Cavalcanti, F.C. Chen, P. Bouman, and W.H. Li. 2002b. Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. *Mol Biol Evol* **19:** 256–262.

Guillon, H. and B. de Massy. 2002. An initiation site for meiotic crossing-over and gene conversion in the mouse. *Nat Genet* **32:** 296–299.

Gurbuz, A., A. Karateke, M. Mengulluoglu, A. Gedikbasi, M. Ozturkmen, C. Kabaca, and Z. Sahinoglu. 2004. Can serum HCG values be used in the differential diagnosis of pregnancy complicated by hypertension? *Hypertens Pregnancy* **23:** 1–12.

Haavisto, A.M., K. Pettersson, M. Bergendahl, A. Virkamaki, and I. Huhtaniemi. 1995. Occurrence and biological properties of a common genetic variant of luteinizing hormone. *J Clin Endocrinol Metab* **80:** 1257–1263.

Handwerger, S., S.P. Wilson, L. Tyrey, and P.M. Conn. 1987. Biochemical evidence that human placental lactogen and human chorionic gonadotropin are not stored in cytoplasmic secretion granules. *Biol Reprod* **37:** 28–32.

Hastings, P.J. 1988. Recombination in the eukaryotic nucleus. *Bioessays* **9:** 61–64.

Haubold, B., J. Kroymann, A. Ratzka, T. Mitchell-Olds, and T. Wiehe. 2002. Recombination and gene conversion in a 170-kb genomic region of Arabidopsis thaliana. *Genetics* **161:** 1269–1278.

Hay, D.L. 1988. Placental histology and the production of human choriogonadotrophin and its subunits in pregnancy. *Br J Obstet Gynaecol* **95:** 1268–1275.

He, X. and J. Zhang. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169:** 1157–1164.

Henke, A. and J. Gromoll. 2008. New insights into the evolution of chorionic gonadotrophin. *Mol Cell Endocrinol* **291:** 11–19.

Henke, A., C.M. Luetjens, M. Simoni, and J. Gromoll. 2007. Chorionic gonadotropin beta-subunit gene expression in the marmoset pituitary is controlled by steroidogenic factor 1, early growth response protein 1, and pituitary homeobox factor 1. *Endocrinology* **148:** 6062–6072.

Herr, F., N. Baal, K. Reisinger, A. Lorenz, T. McKinnon, K.T. Preissner, and M. Zygmunt. 2007. HCG in the regulation of placental angiogenesis. Results of an in vitro study. *Placenta* **28 Suppl A:** S85–93.

Higashi, Y., A. Tanae, H. Inoue, T. Hiromasa, and Y. Fujii-Kuriyama. 1988. Aberrant splicing and missense mutations cause steroid 21-hydroxylase [P-450(C21)] deficiency in humans: possible gene conversion products. *Proc Natl Acad Sci U S A* **85:** 7486–7490.

Hilliker, A.J., G. Harauz, A.G. Reaume, M. Gray, S.H. Clark, and A. Chovnick. 1994. Meiotic gene conversion tract length distribution within the rosy locus of Drosophila melanogaster. *Genetics* **137:** 1019–1026.

Hittinger, C.T. and S.B. Carroll. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449:** 677–681.

Holland, P.W. 2003. More genes in vertebrates? *J Struct Funct Genomics* **3:** 75–84.

Hollenberg, A.N., R.G. Pestell, C. Albanese, M.E. Boers, and J.L. Jameson. 1994. Multiple promoter elements in the human chorionic gonadotropin beta subunit genes distinguish their expression from the luteinizing hormone beta gene. *Mol Cell Endocrinol* **106:** 111–119.

Hruska, K.S., M.E. LaMarca, C.R. Scott, and E. Sidransky. 2008. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Hum Mutat* **29:** 567–583.

Huber, C., K.F. Schable, E. Huber, R. Klein, A. Meindl, R. Thiebe, R. Lamm, and H.G. Zachau. 1993. The V kappa genes of the L regions and the repertoire of V kappa gene sequences in the human germ line. *Eur J Immunol* **23:** 2868–2875.

Hudson, R.R. 2001. Two-locus sampling distributions and their application. *Genetics* **159:** 1805–1817.

Hughes, A.L. 1994. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci* **256:** 119–124.

Hughes, A.L. 2005. Gene duplication and the origin of novel proteins. *Proc Natl Acad Sci U S A* **102:** 8791–8792.

Hughes, A.L. and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335:** 167–170.

Hughes, A.L. and M. Nei. 1989a. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol Biol Evol* **6:** 559–579.

Hughes, A.L. and M. Nei. 1989b. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci U S A* **86:** 958–962.

Hughes, T. and D.A. Liberles. 2007. The pattern of evolution of smaller-scale gene duplicates in mammalian genomes is more consistent with neo- than subfunctionalisation. *J Mol Evol* **65:** 574–588.

Huhtaniemi, I.T., C.C. Korenbrot, and R.B. Jaffe. 1977. HCG binding and stimulation of testosterone biosynthesis in the human fetal testis. *J Clin Endocrinol Metab* **44:** 963–967.

Hunt, D.M., K.S. Dulai, J.A. Cowing, C. Julliot, J.D. Mollon, J.K. Bowmaker, W.H. Li, and D. Hewett-Emmett. 1998. Molecular evolution of trichromacy in primates. *Vision Res* **38:** 3299–3306.

Hurles, M.E. 2001. Gene conversion homogenizes the CMT1A paralogous repeats. *BMC Genomics* **2:** 11.

Iafrate, A.J., L. Feuk, M.N. Rivera, M.L. Listewnik, P.K. Donahoe, Y. Qi, S.W. Scherer, and C. Lee. 2004. Detection of large-scale variation in the human genome. *Nat Genet* **36:** 949–951.

IHGSC. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431:** 931–945.

Innan, H. 2003. A two-locus gene conversion model with selection and its application to the human RHCE and RHD genes. *Proc Natl Acad Sci U S A* **100:** 8793–8798.

Inoue, K. and J.R. Lupski. 2002. Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet* **3:** 199–242.

Itoh, H., K. Suzuki, and H. Kawauchi. 1990. The complete amino acid sequences of alpha subunits of chum salmon gonadotropins. *Gen Comp Endocrinol* **78:** 56–65.

Jablonka-Shariff, A., V. Garcia-Campayo, and I. Boime. 2002. Evolution of lutropin to chorionic gonadotropin generates a specific routing signal for apical release in vivo. *J Biol Chem* **277:** 879–882.

Jaillon, O., J.M. Aury, F. Brunet, J.L. Petit, N. Stange-Thomann, E. Mauceli, L. Bouneau, C. Fischer, C. Ozouf-Costaz, A. Bernot et al. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431:** 946–957.

Jameson, J.L. and A.N. Hollenberg. 1993. Regulation of chorionic gonadotropin gene expression. *Endocr Rev* **14:** 203–221.

Jeffreys, A.J. and C.A. May. 2004. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nat Genet* **36:** 151–156.

Jiang, M., T. Lamminen, P. Pakarinen, J. Hellman, P. Manna, R.J. Herrera, and I. Huhtaniemi. 2002. A novel Ala(-3)Thr mutation in the signal peptide of human luteinizing hormone beta-subunit: potentiation of the inositol phosphate signalling pathway and attenuation of the adenylate cyclase pathway by recombinant variant hormone. *Mol Hum Reprod* **8:** 201–212.

Johnson, M.E., Z. Cheng, V.A. Morrison, S. Scherer, M. Ventura, R.A. Gibbs, E.D. Green, and E.E. Eichler. 2006a. Recurrent duplication-driven transposition of DNA during hominoid evolution. *Proc Natl Acad Sci U S A* **103:** 17626–17631.

Johnson, R.M., T. Prychitko, D. Gumucio, D.E. Wildman, M. Uddin, and M. Goodman. 2006b. Phylogenetic comparisons suggest that distance from the locus control region guides developmental expression of primate beta-type globin genes. *Proc Natl Acad Sci U S A* **103:** 3186–3191.

Judd, S.R. and T.D. Petes. 1988. Physical lengths of meiotic and mitotic gene conversion tracts in Saccharomyces cerevisiae. *Genetics* **118:** 401–410.

Jurisicova, A., M. Antenos, K. Kapasi, J. Meriano, and R.F. Casper. 1999. Variability in the expression of trophectodermal markers beta-human chorionic gonadotrophin, human leukocyte antigen-G and pregnancy specific beta-1 glycoprotein by the human blastocyst. *Hum Reprod* **14:** 1852–1858.

Kaessmann, H., N. Vinckenbosch, and M. Long. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet* **10:** 19–31.

Kainz, P.M., J. Neitz, and M. Neitz. 1998. Recent evolution of uniform trichromacy in a New World monkey. *Vision Res* **38:** 3315–3320.

Kidd, J.M., G.M. Cooper, W.F. Donahue, H.S. Hayden, N. Sampas, T. Graves, N. Hansen, B. Teague, C. Alkan, F. Antonacci et al. 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature* **453:** 56–64.

King, B.F. 1993. Development and structure of the placenta and fetal membranes of nonhuman primates. *J Exp Zool* **266:** 528–540.

Koide, Y., H. Itoh, and H. Kawauchi. 1993. Isolation and characterization of two distinct gonadotropins, GTHI and GTHII, from bonito (Katsuwonus plelamis) pituitary glands. *Int J Pept Protein Res* **41:** 52–65.

Kondrashov, F.A. and A.S. Kondrashov. 2006. Role of selection in fixation of gene duplications. *J Theor Biol* **239:** 141–151.

Kondrashov, F.A., I.B. Rogozin, Y.I. Wolf, and E.V. Koonin. 2002. Selection in the evolution of gene duplications. *Genome Biol* **3:** RESEARCH0008.

Korbel, J.O., A.E. Urban, J.P. Affourtit, B. Godwin, F. Grubert, J.F. Simons, P.M. Kim, D. Palejev, N.J. Carriero, L. Du et al. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318:** 420–426.

Kovalevskaya, G., S. Birken, T. Kakuma, N. Ozaki, M. Sauer, S. Lindheim, M. Cohen, A. Kelly, J. Schlatterer, and J.F. O'Connor. 2002. Differential expression of human chorionic gonadotropin (hCG) glycosylation isoforms in failing and continuing pregnancies: preliminary characterization of the hyperglycosylated hCG epitope. *J Endocrinol* **172:** 497–506.

Kudla, G., A. Helwak, and L. Lipinski. 2004. Gene conversion and GC-content evolution in mammalian Hsp70. *Mol Biol Evol* **21:** 1438–1444.

Kulski, J.K., S. Gaudieri, M. Bellgard, L. Balmer, K. Giles, H. Inoko, and R.L. Dawkins. 1997. The evolution of MHC diversity by segmental duplication and transposition of retroelements. *J Mol Evol* **45:** 599–609.

Kurioka, H., K. Takahashi, M. Irikoma, M. Okada, T. Ozaki, T. Ueda, and K. Miyazaki. 1999. Diagnostic difficulty in polycystic ovary syndrome due to an LH-beta-subunit variant. *Eur J Endocrinol* **140:** 235–238.

Lam, K.W. and A.J. Jeffreys. 2006. Processes of copy-number change in human DNA: the dynamics of {alpha}-globin gene deletion. *Proc Natl Acad Sci U S A* **103:** 8921–8927.

Lamminen, T., M. Jiang, P.R. Manna, P. Pakarinen, H. Simonsen, R.J. Herrera, and I. Huhtaniemi. 2002. Functional study of a recombinant form of human LHbeta-subunit variant carrying the Gly(102)Ser mutation found in Asian populations. *Mol Hum Reprod* **8:** 887–892.

Lapthorn, A.J., D.C. Harris, A. Littlejohn, J.W. Lustbader, R.E. Canfield, K.J. Machin, F.J. Morgan, and N.W. Isaacs. 1994. Crystal structure of human chorionic gonadotropin. *Nature* **369:** 455–461.

Lasley, B.L., J.K. Hodges, and N.M. Czekala. 1980. Monitoring the female reproductive cycle of great apes and other primate species by determination of oestrogen and LH in small volumes of urine. *J Reprod Fertil Suppl* **Suppl 28:** 121–129.

Lee, J.A., C.M. Carvalho, and J.R. Lupski. 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131:** 1235–1247.

Lee, J.A. and J.R. Lupski. 2006. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* **52:** 103–121.

Letterie, G.S. and M. Hibbert. 2000. Serial serum human chorionic gonadotropin (hCG) levels in ectopic pregnancy and first trimester miscarriage. *Arch Gynecol Obstet* **263:** 168–169.

Li, M.D. and J.J. Ford. 1998. A comprehensive evolutionary analysis based on nucleotide and amino acid sequences of the alpha- and beta-subunits of glycoprotein hormone gene family. *J Endocrinol* **156:** 529–542.

Li, N. and M. Stephens. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165:** 2213–2233.

Li, W.H. 1997. *Molecular Evolution*. Sinauer Associates, Inc., Sunderland.

Li, Y., C. Ye, P. Shi, X.J. Zou, R. Xiao, Y.Y. Gong, and Y.P. Zhang. 2005. Independent origin of the growth hormone gene family in New World monkeys and Old World monkeys/hominoids. *J Mol Endocrinol* **35:** 399–409.

Liao, W.X., A.C. Roy, C. Chan, S. Arulkumaran, and S.S. Ratnam. 1998. A new molecular variant of luteinizing hormone associated with female infertility. *Fertil Steril* **69:** 102–106.

Libert, F., E. Passage, A. Lefort, G. Vassart, and M.G. Mattei. 1990. Localization of human thyrotropin receptor gene to chromosome region 14q3 by in situ hybridization. *Cytogenet Cell Genet* **54:** 82–83.

Lieber, M.R., Y. Ma, U. Pannicke, and K. Schwarz. 2003. Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* **4:** 712–720.

Linardopoulou, E.V., E.M. Williams, Y. Fan, C. Friedman, J.M. Young, and B.J. Trask. 2005. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437:** 94–100.

Lindsay, S.J., M. Khajavi, J.R. Lupski, and M.E. Hurles. 2006. A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am J Hum Genet* **79:** 890–902.

Liskay, R.M., A. Letsou, and J.L. Stachelek. 1987. Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics* **115:** 161–167.

Lloyd, J.M. and G.V. Childs. 1988. Differential storage and release of luteinizing hormone and follicle-releasing hormone from individual gonadotropes separated by centrifugal elutriation. *Endocrinology* **122:** 1282–1290.

Lofrano-Porto, A., G.B. Barra, L.A. Giacomini, P.P. Nascimento, A.C. Latronico, L.A. Casulari, and A. da Rocha Neves Fde. 2007. Luteinizing hormone beta mutation and hypogonadism in men and women. *N Engl J Med* **357:** 897–904.

Lukacsovich, T. and A.S. Waldman. 1999. Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. *Genetics* **151:** 1559–1568.

Lupski, J.R. 2007. Genomic rearrangements and sporadic disease. *Nat Genet* **39:** S43–47.

Lynch, M. and J.S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290:** 1151–1155.

Lynch, M. and J.S. Conery. 2003. The evolutionary demography of duplicate genes. *J Struct Funct Genomics* **3:** 35–44.

Lynch, M. and A. Force. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154:** 459–473.

MacCarthy, T. and A. Bergman. 2007. The limits of subfunctionalization. *BMC Evol Biol* **7:** 213.

Madersbacher, S., C. Kratzik, R. Gerth, S. Dirnhofer, and P. Berger. 1994. Human chorionic gonadotropin (hCG) and its free subunits in hydrocele fluids and neoplastic tissue of testicular cancer patients: insights into the in vivo hCG-secretion pattern. *Cancer Res* **54:** 5096–5100.

Maere, S., S. De Bodt, J. Raes, T. Casneuf, M. Van Montagu, M. Kuiper, and Y. Van de Peer. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A* **102:** 5454–5459.

Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L.M. Steinmetz. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454:** 479–485.

Marais, G. 2003. Biased gene conversion: implications for genome and sex evolution. *Trends Genet* **19:** 330–338.

Marcillac, I., F. Troalen, J.M. Bidart, P. Ghillani, V. Ribrag, B. Escudier, B. Malassagne, J.P. Droz, C. Lhomme, P. Rougier et al. 1992. Free human chorionic gonadotropin beta subunit in gonadal and nongonadal neoplasms. *Cancer Res* **52:** 3901–3907.

Martinsohn, J.T., A.B. Sousa, L.A. Guethlein, and J.C. Howard. 1999. The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* **50:** 168–200.

Maston, G.A. and M. Ruvolo. 2002. Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol Biol Evol* **19:** 320–335.

Mateos, J., E. Mananos, G. Martinez-Rodriguez, M. Carrillo, B. Querat, and S. Zanuy. 2003. Molecular characterization of sea bass gonadotropin subunits (alpha, FSHbeta, and LHbeta) and their expression during the reproductive cycle. *Gen Comp Endocrinol* **133:** 216–232.

McClintock, J.M., R. Carlson, D.M. Mann, and V.E. Prince. 2001. Consequences of Hox gene duplication in the vertebrates: an investigation of the zebrafish Hox paralogue group 1 genes. *Development* **128:** 2471–2484.

McClintock, J.M., M.A. Kheirbek, and V.E. Prince. 2002. Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. *Development* **129:** 2339–2354.

McLysaght, A., K. Hokamp, and K.H. Wolfe. 2002. Extensive genomic duplication during early chordate evolution. *Nat Genet* **31:** 200–204.

Meyer, A. and Y. Van de Peer. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* **27:** 937–945.

Mikkelsen, T.S., L.W. Hillier, E.E. Eichler, M.C. Zody, D.B. Jaffe, S.P. Yang, W. Enard, I. Hellmann, K. Lindblad-Toh, and T.K. Altheide. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437:** 69–87.

Miller-Lindholm, A.K., E. Bedows, C.F. Bartels, J. Ramey, V. Maclin, and R.W. Ruddon. 1999. A naturally occurring genetic variant in the human chorionic gonadotropin-beta gene 5 is assembly inefficient. *Endocrinology* **140:** 3496–3506.

Miller-Lindholm, A.K., C.J. LaBenz, J. Ramey, E. Bedows, and R.W. Ruddon. 1997. Human chorionic gonadotropin-beta gene expression in first trimester placenta. *Endocrinology* **138:** 5459–5465.

Moffett, A. and C. Loke. 2006. Immunology of placentation in eutherian mammals. *Nat Rev Immunol* **6:** 584–594.

Mondragon-Palomino, M. and B.S. Gaut. 2005. Gene conversion and the evolution of three leucine-rich repeat gene families in Arabidopsis thaliana. *Mol Biol Evol* **22:** 2444–2456.

Moyle, W. and R. Campbell. 1996. Gonadotropins. In *Reproductive Endocrinology, Surgery, and Technology* (eds. E. Adashi J. Rock, and Z. Rosenwaks). Lippincott-Raven Publishers, Philadelphia.

Muller, T., M. Simoni, E. Pekel, C.M. Luetjens, R. Chandolia, F. Amato, R.J. Norman, and J. Gromoll. 2004. Chorionic gonadotrophin beta subunit mRNA but not luteinising hormone beta subunit mRNA is expressed in the pituitary of the common marmoset (Callithrix jacchus). *J Mol Endocrinol* **32:** 115–128.

Munro, C.J., L.S. Laughlin, J.C. Illera, J. Dieter, A.G. Hendrickx, and B.L. Lasley. 1997. ELISA for the measurement of serum and urinary chorionic gonadotropin concentrations in the laboratory macaque. *Am J Primatol* **41:** 307–322.

Murphy, B.D. and S.D. Martinuk. 1991. Equine chorionic gonadotropin. *Endocr Rev* **12:** 27–44.

Nachman, M.W. and S.L. Crowell. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156:** 297–304.

Nassif, N., J. Penney, S. Pal, W.R. Engels, and G.B. Gloor. 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol Cell Biol* **14:** 1613–1625.

Nei, M., X. Gu, and T. Sitnikova. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci U S A* **94:** 7799–7806.

Nei, M. and A.P. Rooney. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* **39:** 121–152.

Nicolis, E., A. Bonizzato, B.M. Assael, and M. Cipolli. 2005. Identification of novel mutations in patients with Shwachman-Diamond syndrome. *Hum Mutat* **25:** 410.

Nishant, K.T., H. Ravishankar, and M.R. Rao. 2004. Characterization of a mouse recombination hot spot locus encoding a novel non-protein-coding RNA. *Mol Cell Biol* **24:** 5620–5634.

Nishimura, R., J. Shin, I. Ji, C.R. Middaugh, W. Kruggel, R.V. Lewis, and T.H. Ji. 1986. A single amino acid substitution in an ectopic alpha subunit of a human carcinoma choriogonadotropin. *J Biol Chem* **261:** 10475–10477.

O'Connor, J.F., N. Ellish, T. Kakuma, J. Schlatterer, and G. Kovalevskaya. 1998. Differential urinary gonadotrophin profiles in early pregnancy and early pregnancy loss. *Prenat Diagn* **18:** 1232–1240.

Ohno, S. 1970. *Evolution by gene duplication*. Springer, New York.

Ohta, T. 1991. Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. *Proc Natl Acad Sci U S A* **88:** 6716–6720.

Okuno, A., S. Komori, K. Sakata, Y. Nakata, Y. Tsuji, K. Koyama, and T. Kameda. 2001. Genetic analysis of a variant luteinizing hormone in an infertile woman. *Arch Gynecol Obstet* **265:** 148–150.

Oota, H., C.W. Dunn, W.C. Speed, A.J. Pakstis, M.A. Palmatier, J.R. Kidd, and K.K. Kidd. 2007. Conservative evolution in duplicated genes of the primate Class I ADH cluster. *Gene* **392:** 64–76.

Padhukasahasram, B., P. Marjoram, and M. Nordborg. 2004. Estimating the rate of gene conversion on human chromosome 21. *Am J Hum Genet* **75:** 386–397.

Palmer, S., E. Schildkraut, R. Lazarin, J. Nguyen, and J.A. Nickoloff. 2003. Gene conversion tracts in Saccharomyces cerevisiae can be extremely short and highly directional. *Nucleic Acids Res* **31:** 1164–1173.

Papadakis, M.N. and G.P. Patrinos. 1999. Contribution of gene conversion in the evolution of the human beta-like globin gene family. *Hum Genet* **104:** 117–125.

Paques, F. and J.E. Haber. 1999. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **63:** 349–404.

Parham, P. and T. Ohta. 1996. Population biology of antigen presentation by MHC class I molecules. *Science* **272:** 67–74.

Parrott, A.M. and M.B. Mathews. 2007. Novel rapidly evolving hominid RNAs bind nuclear factor 90 and display tissue-restricted distribution. *Nucleic Acids Res* **35:** 6249–6258.

Pavlicek, A., R. House, A.J. Gentles, J. Jurka, and B.E. Morrow. 2005. Traffic of genetic information between segmental duplications flanking the typical 22q11.2 deletion in velo-cardio-facial syndrome/DiGeorge syndrome. *Genome Res* **15:** 1487–1495.

Payen, C., R. Koszul, B. Dujon, and G. Fischer. 2008. Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. *PLoS Genet* **4:** e1000175.

Pease, L.R., R.M. Horton, J.K. Pullen, and T.J. Yun. 1993. Unusual mutation clusters provide insight into class I gene conversion mechanisms. *Mol Cell Biol* **13:** 4374–4381.

Perry, G.H., A. Ben-Dor, A. Tsalenko, N. Sampas, L. Rodriguez-Revenga, C.W. Tran, A. Scheffer, I. Steinfeld, P. Tsang, N.A. Yamada et al. 2008. The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* **82:** 685–695.

Pettersson, K., Y.Q. Ding, and I. Huhtaniemi. 1992. An immunologically anomalous luteinizing hormone variant in a healthy woman. *J Clin Endocrinol Metab* **74:** 164–171.

Pierce, J.G. and T.F. Parsons. 1981. Glycoprotein hormones: structure and function. *Annu Rev Biochem* **50:** 465–495.

Piertney, S.B. and M.K. Oliver. 2006. The evolutionary ecology of the major histocompatibility complex. *Heredity* **96:** 7–21.

Piontkivska, H. and M. Nei. 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. *Mol Biol Evol* **20:** 601–609.

Plagnol, V., B. Padhukasahasram, J.D. Wall, P. Marjoram, and M. Nordborg. 2006. Relative influences of crossing over and gene conversion on the pattern of linkage disequilibrium in Arabidopsis thaliana. *Genetics* **172:** 2441–2448.

Poikkeus, P., V. Hiilesmaa, and A. Tiitinen. 2002. Serum HCG 12 days after embryo transfer in predicting pregnancy outcome. *Hum Reprod* **17:** 1901–1905.

Policastro, P., C.E. Ovitt, M. Hoshina, H. Fukuoka, M.R. Boothby, and I. Boime. 1983. The beta subunit of human chorionic gonadotropin is encoded by multiple genes. *J Biol Chem* **258:** 11492–11499.

Policastro, P.F., S. Daniels-McQueen, G. Carle, and I. Boime. 1986. A map of the hCG beta-LH beta gene cluster. *J Biol Chem* **261:** 5907–5916.

Popesco, M.C., E.J. Maclaren, J. Hopkins, L. Dumas, M. Cox, L. Meltesen, L. McGavran, G.J. Wyckoff, and J.M. Sikela. 2006. Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* **313:** 1304–1307.

Postlethwait, J.H., I.G. Woods, P. Ngo-Hazelett, Y.L. Yan, P.D. Kelly, F. Chu, H. Huang, A. Hill-Force, and W.S. Talbot. 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* **10:** 1890–1902.

Prat, F., J.P. Sumpter, and C.R. Tyler. 1996. Validation of radioimmunoassays for two salmon gonadotropins (GTH I and GTH II) and their plasma concentrations throughout the reproductivecycle in male and female rainbow trout (Oncorhynchus mykiss). *Biol Reprod* **54:** 1375–1382.

Przeworski, M. and J.D. Wall. 2001. Why is there so little intragenic linkage disequilibrium in humans? *Genet Res* **77:** 143–151.

Ptak, S.E., K. Voelpel, and M. Przeworski. 2004. Insights into recombination from patterns of linkage disequilibrium in humans. *Genetics* **167:** 387–397.

Rajkhowa, M., J.A. Talbot, P.W. Jones, K. Pettersson, A.M. Haavisto, I. Huhtaniemi, and R.N. Clayton. 1995. Prevalence of an immunological LH beta-subunit variant in a UK population of healthy women and women with polycystic ovary syndrome. *Clin Endocrinol (Oxf)* **43:** 297–303.

Ramanujam, L., W.X. Liao, A.C. Roy, S.C. Ng, and S.S. Ratnam. 1998. Molecular variants of luteinizing hormone in three populations of Southeast Asia. *Hum Hered* **48:** 232–234.

Ramanujam, L.N., W.X. Liao, A.C. Roy, A. Loganath, H.H. Goh, and S.C. Ng. 1999. Association of molecular variants of luteinizing hormone with menstrual disorders. *Clin Endocrinol (Oxf)* **51:** 243–246.

Rastogi, S. and D.A. Liberles. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol Biol* **5:** 28.

Redon, R., S. Ishikawa, K.R. Fitch, L. Feuk, G.H. Perry, T.D. Andrews, H. Fiegler, M.H. Shapero, A.R. Carson, W. Chen et al. 2006. Global variation in copy number in the human genome. *Nature* **444:** 444–454.

Reimer, T., D. Koczan, H. Muller, K. Friese, A. Krause, H.J. Thiesen, and B. Gerber. 2000. Human chorionic gonadotrophin-beta transcripts correlate with progesterone receptor values in breast carcinomas. *J Mol Endocrinol* **24:** 33–41.

Reiter, L.T., P.J. Hastings, E. Nelis, P. De Jonghe, C. Van Broeckhoven, and J.R. Lupski. 1998. Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. *Am J Hum Genet* **62:** 1023–1033.

Revol De Mendoza, A., D. Esquivel Escobedo, I. Martinez Davila, and H. Saldana. 2004. Expansion and divergence of the GH locus between spider monkey and chimpanzee. *Gene* **336:** 185–193.

Roth, C., S. Rastogi, L. Arvestad, K. Dittmar, S. Light, D. Ekman, and D.A. Liberles. 2007. Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. *J Exp Zoolog B Mol Dev Evol* **308:** 58–73.

Roth, D.B., T.N. Porter, and J.H. Wilson. 1985. Mechanisms of nonhomologous recombination in mammalian cells. *Mol Cell Biol* **5:** 2599–2607.

Rousseau-Merck, M.F., M. Atger, H. Loosfelt, E. Milgrom, and R. Berger. 1993. The chromosomal localization of the human follicle-stimulating hormone receptor gene (FSHR) on 2p21-p16 is similar to that of the luteinizing hormone receptor gene. *Genomics* **15:** 222–224.

Rousseau-Merck, M.F., M. Misrahi, M. Atger, H. Loosfelt, E. Milgrom, and R. Berger. 1990a. Localization of the human luteinizing hormone/choriogonadotropin receptor gene (LHCGR) to chromosome 2p21. *Cytogenet Cell Genet* **54:** 77–79.

Rousseau-Merck, M.F., M. Misrahi, H. Loosfelt, M. Atger, E. Milgrom, and R. Berger. 1990b. Assignment of the human thyroid stimulating hormone receptor (TSHR) gene to chromosome 14q31. *Genomics* **8:** 233–236.

Roy, A.C., W.X. Liao, Y. Chen, S. Arulkumaran, and S.S. Ratnam. 1996. Identification of seven novel mutations in LH beta-subunit gene by SSCP. *Mol Cell Biochem* **165:** 151–153.

Roy, A.M., M.L. Carroll, S.V. Nguyen, A.H. Salem, M. Oldridge, A.O. Wilkie, M.A. Batzer, and P.L. Deininger. 2000. Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res* **10:** 1485–1495.

Rozen, S., H. Skaletsky, J.D. Marszalek, P.J. Minx, H.S. Cordum, R.H. Waterston, R.K. Wilson, and D.C. Page. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423:** 873–876.

Rubnitz, J. and S. Subramani. 1984. The minimum amount of homology required for homologous recombination in mammalian cells. *Mol Cell Biol* **4:** 2253–2258.

Rull, K., P. Hallast, L. Uuskula, J. Jackson, M. Punab, A. Salumets, R.K. Campbell, and M. Laan. 2008a. Fine-scale quantification of HCG beta gene transcription in human trophoblastic and non-malignant non-trophoblastic tissues. *Mol Hum Reprod* **14:** 23–31.

Rull, K. and M. Laan. 2005. Expression of beta-subunit of HCG genes during normal and failed pregnancy. *Hum Reprod* **20:** 3360–3368.

Rull, K., L. Nagirnaja, V.M. Ulander, P. Kelgo, T. Margus, M. Kaare, K. Aittomaki, and M. Laan. 2008b. Chorionic Gonadotropin Beta gene variants are associated with recurrent miscarriage in two European populations. *J Clin Endocrinol Metab*.

Salvignol, I., A. Blancher, P. Calvas, W.W. Socha, Y. Colin, J.P. Cartron, and J. Ruffie. 1993. Relationship between chimpanzee Rh-like genes and the R-C-E-F blood group system. *J Med Primatol* **22:** 19–28.

Sawyer, S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol* **6:** 526–538.

Scannell, D.R., K.P. Byrne, J.L. Gordon, S. Wong, and K.H. Wolfe. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440:** 341–345.

Schienman, J.E., R.A. Holt, M.R. Auerbach, and C.B. Stewart. 2006. Duplication and divergence of 2 distinct pancreatic ribonuclease genes in leaf-eating African and Asian colobine monkeys. *Mol Biol Evol* **23:** 1465–1479.

Schulz, R.W., H.F. Vischer, J.E. Cavaco, E.M. Santos, C.R. Tyler, H.J. Goos, and J. Bogerd. 2001. Gonadotropins, their receptors, and the regulation of testicular functions in fish. *Comp Biochem Physiol B Biochem Mol Biol* **129:** 407–417.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi et al. 2004. Large-scale copy number polymorphism in the human genome. *Science* **305:** 525–528.

Sedman, L., B. Padhukasahasram, P. Kelgo, and M. Laan. 2008. Complex signatures of locus-specific selective pressures and gene conversion on Human Growth Hormone/Chorionic Somatomammotropin genes. *Hum Mutat*.

Semple, C. and K.H. Wolfe. 1999. Gene duplication and gene conversion in the Caenorhabditis elegans genome. *J Mol Evol* **48:** 555–564.

Semple, C.A., M. Rolfe, and J.R. Dorin. 2003. Duplication and selection in the evolution of primate beta-defensin genes. *Genome Biol* **4:** R31.

Sen, S.K., K. Han, J. Wang, J. Lee, H. Wang, P.A. Callinan, M. Dyer, R. Cordaux, P. Liang, and M.A. Batzer. 2006. Human genomic deletions mediated by recombination between Alu elements. *Am J Hum Genet* **79:** 41–53.

Sengupta, M., A. Ray, M. Chaki, M. Maulik, and K. Ray. 2008. SNPs in genes with copy number variation: a question of specificity. *J Genet* **87:** 95–97.

Sharon, D., G. Glusman, Y. Pilpel, M. Khen, F. Gruetzner, T. Haaf, and D. Lancet. 1999. Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. *Genomics* **61:** 24–36.

Sharp, A.J., D.P. Locke, S.D. McGrath, Z. Cheng, J.A. Bailey, R.U. Vallente, L.M. Pertz, R.A. Clark, S. Schwartz, R. Segraves et al. 2005. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* **77:** 78–88.

She, X., J.E. Horvath, Z. Jiang, G. Liu, T.S. Furey, L. Christ, R. Clark, T. Graves, C.L. Gulden, C. Alkan et al. 2004. The structure and evolution of centromeric transition regions within the human genome. *Nature* **430:** 857–864.

Sherman, G.B., M.W. Wolfe, T.A. Farmerie, C.M. Clay, D.S. Threadgill, D.C. Sharp, and J.H. Nilson. 1992. A single gene encodes the beta-subunits of equine luteinizing hormone and chorionic gonadotropin. *Mol Endocrinol* **6:** 951–959.

Shideler, S.E., N.M. Czekala, K. Benirschke, and B.L. Lasley. 1983. Urinary estrogens during pregnancy of the ruffed lemur (Lemur variegatus). *Biol Reprod* **28:** 963–969.

Shiina, T., M. Ota, S. Shimizu, Y. Katsuyama, N. Hashimoto, M. Takasu, T. Anzai, J.K. Kulski, E. Kikkawa, T. Naruse et al. 2006. Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in humans via hitchhiking diversity. *Genetics* **173:** 1555–1570.

Shoja, V. and L. Zhang. 2006. A roadmap of tandemly arrayed genes in the genomes of human, mouse, and rat. *Mol Biol Evol* **23:** 2134–2141.

Simoni, M., J. Gromoll, and E. Nieschlag. 1997. The follicle-stimulating hormone receptor: biochemistry, molecular biology, physiology, and pathophysiology. *Endocr Rev* **18:** 739–773.

Simula, A.P., F. Amato, R. Faast, A. Lopata, J. Berka, and R.J. Norman. 1995. Luteinizing hormone/chorionic gonadotropin bioactivity in the common marmoset (Callithrix jacchus) is due to a chorionic gonadotropin molecule with a structure intermediate between human chorionic gonadotropin and human luteinizing hormone. *Biol Reprod* **53:** 380–389.

Singer, D.S., R.D. Camerini-Otero, M.L. Satz, B. Osborne, D. Sachs, and S. Rudikoff. 1982. Characterization of a porcine genomic clone encoding a major histocompatibility antigen: expression in mouse L cells. *Proc Natl Acad Sci U S A* **79:** 1403–1407.

Slightom, J.L., A.E. Blechl, and O. Smithies. 1980. Human fetal G gamma- and A gamma-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21:** 627–638.

Smith, G.R. 1988. Homologous recombination in procaryotes. *Microbiol Rev* **52:** 1–28.

Stankiewicz, P. and J.R. Lupski. 2002. Genome architecture, rearrangements and genomic disorders. *Trends Genet* **18:** 74–82.

Stenman, U.H., H. Alfthan, and K. Hotakainen. 2004. Human chorionic gonadotropin in cancer. *Clin Biochem* **37:** 549–561.

Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye. 1966. Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harb Symp Quant Biol* **31:** 77–84.

Suganuma, N., K. Furui, F. Kikkawa, Y. Tomoda, and M. Furuhashi. 1996. Effects of the mutations (Trp8 --> Arg and Ile15 --> Thr) in human luteinizing hormone (LH) beta-subunit on LH bioactivity in vitro and in vivo. *Endocrinology* **137:** 831–838.

Sutton-Riley, J.M., S.A. Khanlian, F.W. Byrn, and L.A. Cole. 2006. A single serum test for measuring early pregnancy outcome with high predictive value. *Clin Biochem* **39:** 682–687.

Szostak, J.W., T.L. Orr-Weaver, R.J. Rothstein, and F.W. Stahl. 1983. The double-strand-break repair model for recombination. *Cell* **33:** 25–35.

Tagle, D.A., B.F. Koop, M. Goodman, J.L. Slightom, D.L. Hess, and R.T. Jones. 1988. Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol* **203:** 439–455.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585–595.

Takahashi, K., T. Ozaki, M. Okada, H. Kurioka, H. Kanasaki, and K. Miyazaki. 1999. Increased prevalence of luteinizing hormone beta-subunit variant in patients with premature ovarian failure. *Fertil Steril* **71:** 96–101.

Takahashi, K., A.P. Rooney, and M. Nei. 2000. Origins and divergence times of mammalian class II MHC gene clusters. *J Hered* **91:** 198–204.

Talmadge, K., N.C. Vamvakopoulos, and J.C. Fiddes. 1984. Evolution of the genes for the beta subunits of human chorionic gonadotropin and luteinizing hormone. *Nature* **307:** 37–40.

Taylor, J.S. and J. Raes. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* **38:** 615–643.

Themmen, A.P.N. and I.T. Huhtaniemi. 2000. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. *Endocr Rev* **21:** 551–583.

Valdes-Socin, H., R. Salvi, A.F. Daly, R.C. Gaillard, P. Quatresooz, P.M. Tebeu, F.P. Pralong, and A. Beckers. 2004. Hypogonadism in a patient with a mutation in the luteinizing hormone beta-subunit gene. *N Engl J Med* **351:** 2619–2625.

Valmu, L., H. Alfthan, K. Hotakainen, S. Birken, and U.H. Stenman. 2006. Site-specific glycan analysis of human chorionic gonadotropin beta-subunit from malignancies and pregnancy by liquid chromatography--electrospray mass spectrometry. *Glycobiology* **16:** 1207–1218.

Vazquez-Salat, N., N. Yuhki, T. Beck, S.J. O'Brien, and W.J. Murphy. 2007. Gene conversion between mammalian CCR2 and CCR5 chemokine receptor genes: a potential mechanism for receptor dimerization. *Genomics* **90:** 213–224.

Verrelli, B.C. and S.A. Tishkoff. 2004. Signatures of selection and gene conversion associated with human color vision variation. *Am J Hum Genet* **75:** 363–375.

Vision, T.J., D.G. Brown, and S.D. Tanksley. 2000. The origins of genomic duplications in Arabidopsis. *Science* **290:** 2114–2117.

Wallis, O.C. and M. Wallis. 2002. Characterisation of the GH gene cluster in a new-world monkey, the marmoset (Callithrix jacchus). *J Mol Endocrinol* **29:** 89–97.

Wallis, O.C. and M. Wallis. 2006. Evolution of growth hormone in primates: the GH gene clusters of the New World monkeys marmoset (Callithrix jacchus) and white-fronted capuchin (Cebus albifrons). *J Mol Evol* **63:** 591–601.

Walsh, J.B. 1995. How often do duplicated genes evolve new functions? *Genetics* **139:** 421–428.

Watanabe, H., A. Fujiyama, M. Hattori, T.D. Taylor, A. Toyoda, Y. Kuroki, H. Noguchi, A. BenKahla, H. Lehrach, R. Sudbrak et al. 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* **429:** 382–388.

Watanabe, N., J. Hatano, K. Asahina, T. Iwasaki, and S. Hayakawa. 2007. Molecular cloning and histological localization of LH-like substances in a bottlenose dolphin (Tursiops truncatus) placenta. *Comp Biochem Physiol A Mol Integr Physiol* **146:** 105–118.

Watkins, D.I., Z.W. Chen, A.L. Hughes, M.G. Evans, T.F. Tedder, and N.L. Letvin. 1990. Evolution of the MHC class I genes of a New World primate from ancestral homologues of human non-classical genes. *Nature* **346:** 60–63.

Wehmann, R.E. and B.C. Nisula. 1981. Metabolic and renal clearance rates of purified human chorionic gonadotropin. *J Clin Invest* **68:** 184–194.

Weiss, J., L. Axelrod, R.W. Whitcomb, P.E. Harris, W.F. Crowley, and J.L. Jameson. 1992. Hypogonadism caused by a single amino acid substitution in the beta subunit of luteinizing hormone. *N Engl J Med* **326:** 179–183.

Wendel, J.F. 2000. Genome evolution in polyploids. *Plant Mol Biol* **42:** 225–249.

Weterings, E. and D.C. van Gent. 2004. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair (Amst)* **3:** 1425–1435.

Wiehe, T., J. Mountain, P. Parham, and M. Slatkin. 2000. Distinguishing recombination and intragenic gene conversion by linkage disequilibrium patterns. *Genet Res* **75:** 61–73.

Wong, K.K., R.J. deLeeuw, N.S. Dosanjh, L.R. Kimm, Z. Cheng, D.E. Horsman, C. MacAulay, R.T. Ng, C.J. Brown, E.E. Eichler et al. 2007. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* **80:** 91–104.

Wu, H., J.W. Lustbader, Y. Liu, R.E. Canfield, and W.A. Hendrickson. 1994. Structure of human chorionic gonadotropin at 2.6 A resolution from MAD analysis of the selenomethionyl protein. *Structure* **2:** 545–558.

Xu, S., T. Clark, H. Zheng, S. Vang, R. Li, G.K. Wong, J. Wang, and X. Zheng. 2008. Gene conversion in the rice genome. *BMC Genomics* **9:** 93.

Ye, C., Y. Li, P. Shi, and Y.P. Zhang. 2005. Molecular evolution of growth hormone gene family in old world monkeys and hominoids. *Gene* **350:** 183–192.

Yen, S.S., O. Llerena, B. Little, and O.H. Pearson. 1968. Disappearance rates of endogenous luteinizing hormone and chorionic gonadotropin in man. *J Clin Endocrinol Metab* **28:** 1763–1767.

Zangenberg, G., M.M. Huang, N. Arnheim, and H. Erlich. 1995. New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nat Genet* **10:** 407–414.

Zhang, J. 2006. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat Genet* **38:** 819–823.

Zhang, J., Y.P. Zhang, and H.F. Rosenberg. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet* **30:** 411–415.

Zhang, L. 2008. Adaptive evolution and frequent gene conversion in the brain expressed X-linked gene family in mammals. *Biochem Genet* **46:** 293–311.

Zhang, P., Z. Gu, and W.H. Li. 2003. Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol* **4:** R56.

Zhi, D. 2007. Sequence correlation between neighboring Alu instances suggests post-retrotransposition sequence exchange due to Alu gene conversion. *Gene* **390:** 117–121.

Zygmunt, M., F. Herr, S. Keller-Schoenwetter, K. Kunzi-Rapp, K. Munstedt, C.V. Rao, U. Lang, and K.T. Preissner. 2002. Characterization of human chorionic gonadotropin as a novel angiogenic factor. *J Clin Endocrinol Metab* **87:** 5290–5296.

SUMMARY IN ESTONIAN

Inimese ja šimpansi *Luteiniseeriva hormooni/ Koorion Gonadotropiini beeta*-subühiku *(LHB/CGB)* geeniklaster: liigisisene varieeruvus ning noorte duplitseeritud geenide lahknemine sõsar-liikides

Geneetilise materjali duplitseerumist peetakse üheks peamiseks evolutsiooni mehhanismiks, olles aluseks uute geenide ja bioloogiliste protsesside tekkimises. Arvatakse, et primitiivsete eluvormide evolutsioon keerulisemate suunas on olnud võimalik tänu duplikatsioonisündmuste toimumisele eellasgenoomides. Selgroogsete evolutsioonis on teadaolevalt toimunud kaks kogu genoomi haaravat geneetilise materjali kahekordistumist. Kolmas sarnane sündmus on toimunud ainult kiiruimsete kalade klassis. Kogu genoomi duplikatsioonid annavad alusmaterjali nii bioloogilise mitmekesisuse kui ka uute liikide tekkimiseks. Harvadest laiaulatuslikest kogu genoomi hõlmavatest duplikatsioonidest on sagedasemad lühemaid genoomi fragmente "paljundavad" evolutsioonisündmused.

Käesolevaks ajaks on paljude liikide genoomid sekveneeritud ning on selgunud, et DNA fragmentide duplikatsioonid on toimunud ja toimuvad sagedamini, kui varem arvatud. Ka inimese ja šimpansi vahelistest geneetilise materjali erinevustest hõlmavad suured (>20 000 aluspaari ja >94% DNA järjestuse sarnasususega) duplitseerunud genoomilõigud rohkem (2.7%) kui ühe nukleotiidilised erinevused (1.2–1.5%). Primaatide evolutsiooni käigus "paljundatud" geenid osalevad mitmetes reproduktsiooni, immuunsust, arengut, adaptatsiooni ja aju funktsioone reguleerivates protsessides.

Inimese genoomis on nn. segmentaalseid duplikatsioone umbes 5.2% (ulatusega tuhat kuni mitusada tuhat aluspaari ja >90% DNA järjestuse sarnasusega). Erinevalt unikaalsetest genoomi piirkondadest on sellised väga kõrge DNA järjestuse homoloogiaga piirkonnad altid mitte-alleelsele homoloogilisele rekombinatsioonile ja geenikonversioonile. Mitte-alleelne homoloogiline rekombinatsioon võib põhjustada genoomsete piirkondade ümberkorraldusi (deletsioone, insertsioone, inversioone). Geenikonversioon soodustab mutatsioonide levikut geenikoopiate vahel ja/või homogeniseerib suguluses olevate geenide järjestusi. Lisaks olulisele rollile multigeensete perekondade evolutsioonis, on mõlema mehhanismi osalust näidatud ka mitmete geneetiliste haiguste tekkel.

Lisaks nn. duplikaatpiirkondade struktuuri ja DNA järjestuste erinevustele liikide võrdlusel, võib neis genoomsetes piirkondades olla varieeruvusi ka ühe liigi indiviidide vahel. Näiteks moodustab inimese genoomist koguni ~12% selline geneetiline materjal, mille olemasolu või puudumine populatsioonis varieerub.

Kuigi on selge, et genoomsete segmentide, s.h. geenide duplikatsioonidel on oluline roll liigisiseses ja liikide vahelises varieeruvuses, pole nende tekkimine, evolutsioon ja funktsionaalsed tagajärjed kuigi hästi mõistetud.

Luteiniseeriva hormooni/ Koorion Gonadotropiini beeta-subühiku (*LHB/ CGB*) geeniperekond on tekkinud primaatide liinis mitmete järjestikuste duplikatsioonisündmuste tagajärjel. *LHB/CGB* geenid kodeerivad gonadotroopsete hormoonide beeta subühikut. Gonadotroopsetel hormoonidel on väga oluline roll reproduktiivsüsteemis, olles asendamatud sugurakkude normaalse arengu ning eduka raseduse tagamisel.

Inimese *LHB/CGB* geeniklaster asub kromosoomis 19q13.32 ning koosneb seismest väga kõrge DNA sarnasusega (85–99%) geenist – ühest *LHB* ning kuuest *CGB* geenist. *CGB* geen on tekkinud primaatide liinis ilmselt *LHB* geeni duplikatsiooni ning divergeerumise tagajärjel. *CGB* geen erineb oma eellasest nii ekspressiooni aja (rasedus vs. täiskasvanu iga), koe (platsenta vs. ajuripats) kui hormooni eluea (LH ~1 h vs. hCG ~6 h) poolest. *CGB* geenide koopiate arv varieerub erinevates primaatide liikides: ühest Uue Maailma ahvidel kuni umbes 50'ni gorillal. Antud uuringus kasutati *LHB/CGB* geeniperekonda kui mudelit primaatide-spetsiifiliste duplikaatgeenide varieeruvuse ja evolutsiooni uurimiseks.

Doktoritöö eksperimentaalne osa keskendus järgmistele küsimustele: (i) geenikonversiooni roll inimese *LHB/CGB* geenide DNA varieeruvuse ja alleelse aheldatuse mustrite kujundamises, (ii) inimese ja šimpansi *LHB/CGB* genoomiklastrite võrdlus ning (iii) *LHB/CGB* klastri evolutsiooniliselt kõige nooremate geenide *CGB1* ja *CGB2* iseloomustamine.

Uurimistöö peamised tulemused võib kokku võtta järgnevalt:

1. Duplitseeritud *LHB/CGB* geenide resekveneerimisel kolmes inimpopulatsioonis (Eesti; Hiina Han; Aafrika Mandenkalu) selgus, et DNA järjestuse varieeruvus uuritud geenides on kuni 7x kõrgem, kui unikaalsetes geenides. Geeniklastri servades paiknevaid geene (*LHB*, *CGB* ja *CGB7*) iseloomustavad väga kõrge DNA järjestuse varieeruvus, tugev geenisisene alleelne aheldatus ning aktseptori roll geenikonversiooni protsessis. Klastri keskel asuvate geenide (*CGB2*, *CGB1* ja *CGB5*) varieeruvus on madalam ehk umbes inimese genoomi keskmisel tasemel, alleelse aheldatuse ulatus lühem kui klastri äärtes paiknevates geenides ning geenikonversiooni sündmustes osalevad nad peamiselt doonorina. Saadud tulemused viitavad geenikonversiooni olulisele rollile polümorfismide levitajana geenikoopiate vahel. Geenikonversiooni toimumise suuna võib määrata rekombinatsioonilise aktiivsuse "kuumade punktide" asukoht kõige aktiivsemate aktseptorgeenide läheduses geeniklastri servades.
2. Inimese ja šimpansi *LHB/CGB* genoomiklastrite võrdlemisel selgus, et mõlemas liigis on toimunud sõltumatud duplikatsioonid, mille tagajärjel *CGB* geenide arv liikide vahel erineb: inimesel on kuus *CGB* geenikoopiat, šimpansil viis. Mõlemal liigil on duplitseerunud sama genoomne piirkond, ainult et erinevas ulatuses. *CGB* geenide koopiaarvude erinevused nii inimese ja šimpansi kui ka muude seni uuritud primaatide

liikide vahel osutavad aktiivse dünaamikaga genoomsele piirkonnale, mis on soodus pinnas ka liikide sisestele genoomsetele ümberkorraldustele.

3. Aktiivne liigisisene geenikonversioon on ilmselt: (a) viinud inimese ja šimpansi *LHB/CGB* geeniklastri (nii geenide kui intergeensete alade) kõrge lahknevuseni võrreldes unikaalsete genoomsete piirkondadega ja (b) muutnud transitsiooniliste/ transversiooniliste mutatsioonide suhet uuritud regioonis. *LHB/CGB* geenide evolutsiooni võti on ilmselt tasakaal geenikonversiooni ja loodusliku valiku vahel. Kuigi *LHB/CGB* geenid on väga sarnased (DNA järjestuse sarnasus 85–99%), võib üksikute geenide evolutsioon toimuda erinevate loodusliku valiku survete all.
4. Inimese ja inimahvide *CGB1/CGB2*-sarnaste geenide resekveneerimise tulemused viitavad, et need geenid on tekkinud evolutsiooni mastaabis suhteliselt hiljuti, ilmselt Aafrika inimahvide ühisel eellasel. Inimesel ja šimpansil võib tegemist olla funktsionaalsete geenidega, kuna ühtegi avatud lugemisraami muutvat mutatsiooni geenide kodeerivates järjestustes ei leitud. Gorillal on aga *CGB1/CGB2*-sarnased geenid arvatavasti pseudogeenid, sest mõlemas geenis leiti insertsioone/ deletsioone, mis rikuvad ennustatud valgu järjestuse.

Geenide ja genoomi segmentide duplikatsioonidel on oluline osa liikide siseses ja liikide vahelises geneetilises varieeruvuses ning mitmete haiguste kujunemises. Selliste piirkondade detailne uurimine aitab mõista toimunud ning käimasolevaid geenide ja genoomide evolutsiooni mehhanisme. Detailsemad teadmised duplitseeritud geeniklastrite dünaamikast võivad olla abiks, et ennustada selliseid piirkondi inimese genoomis, mis on altid ümberkorraldustele ja seetõttu võivad osaleda genoomsete haiguste tekkimises.

ACKNOWLEDGEMENTS

I am most grateful to my supervisor Prof. Maris Laan for the support, guidance and help through the whole M.Sc and Ph.D studies. She was always supportive, her ideas and comments have been invaluable in my work.

I would like to thank all the co-authors of the papers, which this dissertation is based on.

I am very grateful to Prof. Aarno Palotie, for the opportunity to do research in his group in Finnish Genome Center in University of Helsinki and for all the help, support and discussions. I also thank Janna Saarela, Susanna Rosas, Maija Wessman, Pekka Ellonen, Verneri Anttila, Mari Kaunisto, Annika Sarahonka, Anu Yliperttula and all the other people from Biomedicum Helsinki for all the help and for creating friendly atmosphere.

I am especially thankful to all the people in our laboratory:
Liina Nagirnaja and Kärt Tomberg for reading my manuscripts, thesis and for all the comments and discussions; Siim Sõber and Tarmo Annilo for the discussions and help with experimental problems; Katrin Kepp for all the candies, Coca-Cola, sauna-nights and support; Kristiina Rull for her energy, help in the medical terms and just for being a very good model in many aspects; Piret Kelgo, Liis Uusküla, Margus Putku, Marina Grigorova, Elin Org, Jaana Männik, Laura Sedman and other members of our lab for the great atmosphere; Viljo Soo and Tõnu Margus for the very valuable technical support.

My special thanks go to Mari, Ingrid, Signe, Tõnis, Siim, Kadri, Kristiina and Monika for sharing the path.

I owe many thanks to Krista Liiv, Tiiu Rootslane, Annely Kukk, Aivar Torp, Jüri Parik and many other people in the Institute of Molecular and Cell Biology for technical and administrative help.

I also wish to thank all my colleagues from Department of Biotechnology and Department of Bioinformatics for all the help.

Finally, my greatest gratitude goes to Peeter, to my friends and family for all the love and support.

This study was financially supported by personal stipends from Kristjan Jaak Stipend Program, World Federation of Scientists', Graduate School in Biomedicine and Biotechnology, Centre for International Mobility CIMO, University of Helsinki Bilateral student exchange program, Estonian Students Fund in the USA, Estonian-Revelia Academic Fund Harald Raudsepp Stipend and Artur Lind Gene Technology Stipend Fund.

APPENDIX

1. The human population sequence variation data of *LHB/CGB* genes from Ref. I is deposited in dbSNP database under accession numbers:
   - *LHB* – ss48399882 – ss48399902
   - *CGB* – ss48399909 – ss48399943
   - *CGB1* – ss48399944 – ss48399963
   - *CGB2* – ss48399964 – ss48399997
   - *CGB5* – ss48399998 – ss48400022
   - *CGB7* – ss48400023 – ss48400071
   - *CGB8* – ss48399818 – ss48399832
   - *CGB5* – *CGB8* intergenic region – ss48399833 – ss48399849
   - *CGB8* – *CGB7* intergenic region – ss48399850 – ss48399881

2. The full chimpanzee *LHB/CGB* genome cluster sequence from Ref. III is deposited in NCBI GenBank, accession number EU000308.

3. The sequences of the *CGB1/CGB2*-like and *LHB* genes for chimpanzee, gorilla and orangutan from Ref. II are deposited in GenBank under accession numbers:
   - Chimpanzee *CGB1/CGB2*-like genes – DQ238547, DQ238549
   - Gorilla *CGB1/CGB2*-like genes – DQ238548, DQ238550
   - Chimpanzee *LHB* gene – DQ238551
   - Gorilla *LHB* gene – DQ238552
   - Orangutan *LHB* gene – DQ238553

PUBLICATIONS

CURRICULUM VITAE

Pille Hallast

Date of birth:	12 July 1980, Tartu, Estonia
Address:	Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu Riia str. 23, 51010, Tartu, Estonia
Phone:	+372 737 5009
E-mail:	pille.hallast@ut.ee

Education:

1986–1995	Luunja Secondary School
1995–1998	Hugo Treffner Gymnasium
1998–2002	B.Sc University of Tartu, Institute of Molecular and Cell Biology
2002–2004	M.Sc University of Tartu, Institute of Molecular and Cell Biology
2004–2009	Ph.D student in Department of Biotechnology, University of Tartu, Institute of Molecular and Cell Biology

Professional employment:

2000–2002	Asper Biotech Ltd., laboratory assistant
2002–2005	Asper Biotech Ltd., project manager
2004	University of Tartu, laboratory assistant
2007–2008	Estonian Biocentre, specialist
2008–	University of Tartu, extraordinary researcher

Scientific work:

During my B.Sc studies I was involved in a project studying the potential candidate genes for mood disorders and depression. In my further studies I have focused on studying primate-specific duplicated genes, specifically the *Luteinizing hormone/ Chorionic Gonadotropin beta* subunit gene family. I have concentrated on the intra-specific and inter-specific DNA variation patterns and evolution of the duplicated genes in human and great apes.

Publications:

1. **Hallast P**, Saarela J, Palotie, Laan M (2008) „High divergence in primate-specific duplicated regions: human and chimpanzee Chorionic Gonadotropin Beta genes" BMC Evol Biol. 7;8:195.
2. Rull K, **Hallast P**, Uusküla L, Jackson J, Punab M, Salumets A, Campbell RK, Laan M. (2008) "Fine-scale quantification of HCG beta gene transcription in human trophoblastic and non-malignant non-trophoblastic tissues". Mol Hum Reprod. 14(1):23–31.
3. **Hallast P**, Rull K and Laan M. (2007) "The evolution and genomic landscape of *CGB1* and *CGB2* genes". Mol Cell Endocrinol. 2;260–262:2–11
4. **Hallast P**, Nagirnaja L, Margus T, Laan M. (2005) "Segmental duplications and gene conversion: human luteinizing hormone/ chorionic gonadotropin beta gene cluster". Gen Res. 15(11):1535–46
5. Kõks S, Nikopensius T, Koida K, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, **Hallast P**, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. (2005) "Analysis of SNP profiles in patients with major depressive disorder." Int J Neuropsychopharmacol. 1: 1–8.
6. Maron E, Nikopensius T, Koks S, Altmae S, Heinaste E, Vabrit K, Tammekivi V, **Hallast P**, Koido K, Kurg A, Metspalu A, Vasar E, Vasar V, Shlik J. (2005) "Association study of 90 candidate gene polymorphisms in panic disorder". Psychiatr Genet. 15(1):17–24.
7. Koido K, Kõks S, Nikopensius T, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, **Hallast P**, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. (2005) "Polymorphisms in wolframin (WFS1) gene are possibly related to increased risk for mood disorders". Int J Neuropsychopharmacol. 8(2):235–44.

ELULOOKIRJELDUS

Pille Hallast

Sünniaeg ja koht:	12. juuli 1980, Tartu, Eesti
Aadress:	Biotehnoloogia õppetool, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool Riia 23, 51010, Tartu, Eesti
Telefon:	+372 737 5009
E-mail:	pille.hallast@ut.ee

Hariduskäik:

1986–1995:	Luunja Keskkool
1995–1998:	Hugo Treffneri Gümnaasium
1998–2002:	Tartu Ülikooli Bioloogia-Geograafia teaduskond, geenitehnoloogia eriala, B.Sc
2002–2004:	Tartu Ülikooli Bioloogia-Geograafia teaduskond, geenitehnoloogia eriala, M.Sc
2004–2009:	Tartu Ülikooli Loodus- ja tehnoloogia teaduskond, geenitehnoloogia eriala, doktorant

Erialane teenistuskäik:

2000–2002:	Asper Biotech Ltd., laborant
2002–2005:	Asper Biotech Ltd., projektijuht
2004:	Tartu Ülikool, laborant
2007–2008:	Eesti Biokeskus, spetsialist
2008-	Tartu Ülikool, erakorraline teadur

Teadustegevus:

Bakalaureuseõpingute raames osalesin projektis, mis tegeles meeleoluhäirete ja depressiooni potentsiaalsete kandidaatgeenide otsimisega. Edaspidises teadustöös olen keskendunud primaatide-spetsiifiliste duplikaatgeenide uurimisele, täpsemalt *Luteiniseeriva hormooni/ Koorion Gonadotropiini beeta*-subühiku geeniperekonnale. Mind huvitavad inimese ja inimahvide duplikaatgeenide liikide sisesed ja liikide vahelised DNA järjestuse varieeruvused ja evolutsioon.

Publikatsioonid:

1. **Hallast P**, Saarela J, Palotie, Laan M (2008) „High divergence in primate-specific duplicated regions: human and chimpanzee Chorionic Gonadotropin Beta genes" BMC Evol Biol. 7;8:195.
2. Rull K, **Hallast P**, Uusküla L, Jackson J, Punab M, Salumets A, Campbell RK, Laan M. (2008) "Fine-scale quantification of HCG beta gene transcription in human trophoblastic and non-malignant non-trophoblastic tissues". Mol Hum Reprod. 14(1):23–31.
3. **Hallast P**, Rull K and Laan M. (2007) "The evolution and genomic landscape of *CGB1* and *CGB2* genes". Mol Cell Endocrinol. 2;260–262:2–11
4. **Hallast P**, Nagirnaja L, Margus T, Laan M. (2005) "Segmental duplications and gene conversion: human luteinizing hormone/ chorionic gonadotropin beta gene cluster". Gen Res. 15(11):1535–46
5. Kõks S, Nikopensius T, Koida K, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, **Hallast P**, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. (2005) "Analysis of SNP profiles in patients with major depressive disorder." Int J Neuropsychopharmacol. 1: 1–8.
6. Maron E, Nikopensius T, Koks S, Altmae S, Heinaste E, Vabrit K, Tammekivi V, **Hallast P**, Koido K, Kurg A, Metspalu A, Vasar E, Vasar V, Shlik J. (2005) "Association study of 90 candidate gene polymorphisms in panic disorder". Psychiatr Genet. 15(1):17–24.
7. Koido K, Kõks S, Nikopensius T, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, **Hallast P**, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. (2005) "Polymorphisms in wolframin (WFS1) gene are possibly related to increased risk for mood disorders". Int J Neuropsychopharmacol. 8(2):235–44.

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro*. Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous crassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan.** Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.

41. **Sulev Ingerpuu.** Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.
42. **Veljo Kisand.** Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa.** Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa.** Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik.** Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo.** Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo.** Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots.** Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero.** Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees.** Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks.** Cholecystokinin (CCK) — induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and erotonin. Tartu, 1999, 123 p.
52. **Ebe Sild.** Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva.** Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna.** Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro.** Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane.** Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm.** Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg.** Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild.** The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu.** Studies of the TOL plasmid transcription factor XylS. Tartu 2000. 88 p.

61. **Dina Lepik.** Modulation of viral DNA replication by tumor suppressor protein p53. Tartu 2000. 106 p.
62. **Kai Vellak.** Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu 2000. 122 p.
63. **Jonne Kotta.** Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu 2000. 160 p.
64. **Georg Martin.** Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000. 139 p.
65. **Silvia Sepp.** Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira.** On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000. 96 p.
67. **Priit Zingel.** The role of planktonic ciliates in lake ecosystems. Tartu 2001. 111 p.
68. **Tiit Teder.** Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu 2001. 122 p.
69. **Hannes Kollist.** Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu 2001. 80 p.
70. **Reet Marits.** Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu 2001. 112 p.
71. **Vallo Tilgar.** Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002. 126 p.
72. **Rita Hõrak.** Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002. 108 p.
73. **Liina Eek-Piirsoo.** The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002. 74 p.
74. **Krõõt Aasamaa.** Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002. 110 p.
75. **Nele Ingerpuu.** Bryophyte diversity and vascular plants. Tartu, 2002. 112 p.
76. **Neeme Tõnisson.** Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002. 124 p.
77. **Margus Pensa.** Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003. 110 p.
78. **Asko Lõhmus.** Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003. 168 p.
79. **Viljar Jaks.** p53 — a switch in cellular circuit. Tartu, 2003. 160 p.
80. **Jaana Männik.** Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003. 140 p.
81. **Marek Sammul.** Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003. 159 p

82. **Ivar Ilves.** Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003. 89 p.
83. **Andres Männik.** Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003. 109 p.
84. **Ivika Ostonen.** Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003. 158 p.
85. **Gudrun Veldre.** Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003. 199 p.
86. **Ülo Väli.** The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004. 159 p.
87. **Aare Abroi.** The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004. 135 p.
88. **Tiina Kahre.** Cystic fibrosis in Estonia. Tartu, 2004. 116 p.
89. **Helen Orav-Kotta.** Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004. 117 p.
90. **Maarja Öpik.** Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004. 175 p.
91. **Kadri Tali.** Species structure of *Neotinea ustulata*. Tartu, 2004. 109 p.
92. **Kristiina Tambets.** Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004. 163 p.
93. **Arvi Jõers.** Regulation of p53-dependent transcription. Tartu, 2004. 103 p.
94. **Lilian Kadaja.** Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004. 103 p.
95. **Jaak Truu.** Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004. 128 p.
96. **Maire Peters.** Natural horizontal transfer of the *pheBA* operon. Tartu, 2004. 105 p.
97. **Ülo Maiväli.** Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004. 130 p.
98. **Merit Otsus.** Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004. 103 p.
99. **Mikk Heidemaa.** Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004. 167 p.
100. **Ilmar Tõnno.** The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004. 111 p.
101. **Lauri Saks.** Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004. 144 p.

102. **Siiri Rootsi.** Human Y-chromosomal variation in European populations. Tartu, 2004. 142 p.
103. **Eve Vedler.** Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover.** Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 126 p.
105. **Helen Udras.** Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005. 100 p.
106. **Ave Suija.** Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005. 162 p.
107. **Piret Lõhmus.** Forest lichens and their substrata in Estonia. Tartu, 2005. 162 p.
108. **Inga Lips.** Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005. 156 p.
109. **Kaasik, Krista.** Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005. 121 p.
110. **Juhan Javoiš.** The effects of experience on host acceptance in ovipositing moths. Tartu, 2005. 112 p.
111. **Tiina Sedman.** Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005. 103 p.
112. **Ruth Aguraiuja.** Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005. 112 p.
113. **Riho Teras.** Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 106 p.
114. **Mait Metspalu.** Through the course of prehistory in india: tracing the mtDNA trail. Tartu, 2005. 138 p.
115. **Elin Lõhmussaar.** The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006. 124 p.
116. **Priit Kupper.** Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006. 126 p.
117. **Heili Ilves.** Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006. 120 p.
118. **Silja Kuusk.** Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006. 126 p.
119. **Kersti Püssa.** Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006. 90 p.
120. **Lea Tummeleht.** Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006. 94 p.
121. **Toomas Esperk.** Larval instar as a key element of insect growth schedules. Tartu, 2006. 186 p.

122. **Harri Valdmann.** Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers.** Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli.** Gata3 and Gata2 in inner ear development. Tartu, 2007. 123 p.
125. **Kai Rünk.** Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007. 143 p.
126. **Aveliina Helm.** Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007. 89 p.
127. **Leho Tedersoo.** Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007. 233 p.
128. **Marko Mägi.** The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007. 135 p.
129. **Valeria Lulla.** Replication strategies and applications of Semliki Forest virus. Tartu, 2007. 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007. 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007. 171 p.
132. **Tatjana Krama.** Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007.
133. **Signe Saumaa.** The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007. 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007. 96 p.
135. **Priit Kilgas.** Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007. 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007. 95 p.
137. **Kärt Padari.** Protein transduction mechanisms of transportans. Tartu, 2008. 128 p.
138. **Siiri-Lii Sandre.** Selective forces on larval colouration in a moth. Tartu, 2008. 125 p.
139. **Ülle Jõgar.** Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008. 99 p.
140. **Lauri Laanisto.** Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008. 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008. 105 p.

142. **Birgot Paavel.** Bio-optical properties of turbid lakes. Tartu, 2008. 175 p.
143. **Kaire Torn.** Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg.** Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd.** Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag.** Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.
147. **Ulvi Karu.** Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm.** Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks.** Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu.** Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš.** Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko.** Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast.** Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats.** Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova.** The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik.** Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2008, 112 p.
158. **Kaja Kiiver.** Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja.** Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.