The Effects of Mutation and Selection on the Rate and Pattern of Molecular Evolution in Birds

BY

SOFIA BERLIN
Dissertation presented at Uppsala University to be publicly examined in Friessalen, EBC, Uppsala, Thursday, September 30, 2004 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

By comparing sequence diversity and divergence on sex chromosomes one can study how the rate of evolution in affected by mutation and/or selection. The rate of mutation in male biased, meaning that relatively more mutations are created in the male germ line than in the female. Since the male mutation bias (um) most likely is a consequence of the difference in cell divisions between male and female germ lines, life history characters that affect this difference should covary with um. Indeed, we found a positive correlation between estimates of um and increased generation times and increased intensity of sperm competition. We have also found that estimates of um varied significantly between gametologous introns located on the sex chromosomes. This could be a consequence of the variation in substitution rates between loci.

Population genetics theory predicts that both positive and negative selection reduce genetic diversity around a selected locus at a distance determined by the rate of recombination. Consequently, a non-recombining chromosome, like the female specific W chromosome in birds, selection is expected to have a large impact on sequence diversity. Indeed, in a large sequence screening we found only one segregating site among 7643 base pairs sequenced in 47 chicken females. Furthermore, we also found that deleterious substitutions are fixed in a higher rate for W- than Z-linked sequences, which is in agreement with the lack of recombination and strong genetic drift due to the low effective population size.

Rarely non-synonymous mutations are beneficial for an individual, but when it happens, the mutation is positively selected and rapidly reaches fixation in a population. We have found that positive selection has been acting on the female reproductive protein, zona pellucida c in birds. This rapid evolution is likely a mechanism to prevent hybridisation.

Keywords: substitution rate, diversity, selection, male mutation bias, birds

Sofia Berlin, Department of Evolution, Genomics and Systematics, Evolutionary Biology, Norbyv. 18C, Uppsala University, SE-752 36 Uppsala, Sweden

© Sofia Berlin 2004

ISSN 1104-232X
ISBN 91-554-6021-6
urn:nbn:se:uu:diva-4516 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-4516)
To My Family
List of Papers


\(^1\)The first two authors contributed equally to this study.


IV  Berlin, S. and Ellegren, H. Chicken W: a homogenous chromosome in a highly variable genome. Submitted manuscript.

V  Berlin, S. and Smith, N.G.C. Adaptive evolution of ZPC, a female reproductive protein in diverse vertebrate species. Submitted manuscript.

Paper I is printed with permission form the publisher.
Contents

Introduction...................................................................................................11
Nucleotide substitutions ...........................................................................11
Substitution rates ..................................................................................12
Selection ...............................................................................................13
Rate variation in neutral DNA.............................................................14
Nucleotide diversity .................................................................................16
Selection ...............................................................................................18
Sex chromosome evolution ......................................................................19
The Y and W........................................................................................21
Z- and W-linked genes – Gametologs .................................................24
Nucleotide diversity and sex chromosomes ........................................26
The male mutation bias ........................................................................27
The strength of the male mutation bias................................................29
Obstacles with the evolutionary approach...........................................30
The study organisms – Birds ..................................................................32
Research aims ..........................................................................................33

Present Investigation.....................................................................................34
Paper I. Life history and the male mutation bias ......................................34
Methods ...............................................................................................35
Results .................................................................................................35
Discussion............................................................................................36
Paper II. Local mutation rate heterogeneity and the male mutation bias .37
Methods ...............................................................................................38
Results .................................................................................................39
Discussion............................................................................................40
Paper III. Accumulation of deleterious mutations on the female-specific
W chromosome in birds .........................................................................40
Methods ...............................................................................................41
Results .................................................................................................41
Discussion............................................................................................42
Paper IV. The chicken W chromosome – a homogenous chromosome in a
highly variable genome .........................................................................43
Methods and Results............................................................................43
Discussion............................................................................................44
Paper V. Adaptive evolution of ZPC, a female reproductive protein in diverse vertebrate species

Methods

Results

Discussion

Acknowledgements

References
Abbreviations

\( \mu \)  Mutation rate
\( K \)  Divergence or substitutions per site
\( T \)  Time of divergence
\( K_a \)  Non-synonymous substitutions per non-synonymous site
\( K_s \)  Synonymous substitutions per synonymous site
\( \omega \)  \( K_a/K_s \)
\( n_f \)  Cell divisions in female germ line
\( n_m \)  Cell divisions in male germ line
\( \pi \)  Nucleotide diversity
\( N_e \)  Effective population size
\( \alpha_m \)  Male-to-female rate ratio of mutations
\( C \)  Male-to-female ratio of germ cell divisions
Introduction

Nucleotide substitutions

When studying evolution at the molecular level the basic quantity is the number of nucleotide differences or DNA substitutions between two sequences. This knowledge can for instance be used to estimate the rate of evolution, time of divergence and to reconstruct the phylogentic relationship between species (Li 1997). After two lineages separate from each other, each of them begins to accumulate substitutions and they diverge from each other. A common way of describing the divergence between two sequences is the number of different nucleotides (substitutions) per site (K). In the first phase of divergence of two sequences, most new substitutions will increase the difference between them. With time, multiple substitutions (multiple hits) may happen at the same site. Therefore, in order to study the dynamics of nucleotide substitutions one must take into account the different probabilities of substitutions of one nucleotide by another given a certain divergence time.

A number of such mathematical models have been proposed of which Jukes and Cantor’s one-parameter model is the simplest (Jukes and Cantor 1969). In this model all nucleotides have the same frequency and probability to change to any of the other. However, a complicating factor is that the direction of mutations is non-random, therefore, certain sequences require more complicated models. A parameter-rich model is the one of Tamura-Nei (Tamura and Nei 1993); with its six parameters it takes into account variable
nucleotide frequencies, two transition frequencies and one transversion frequency.

Protein coding and non-coding sequences are generally treated separately, since they most often evolve with different rates. There are several models for estimating substitutions in protein coding sequences. In most of these methods, synonymous (silent) and non-synonymous substitutions (amino acid altering) are analysed separately. To correctly estimate synonymous and non-synonymous divergences one must take into account the number of sites that potentially produce mutations of either of the two classes. The accurate procedure to measure synonymous divergence ($K_s$) is therefore to divide the number of synonymous substitutions with the number of synonymous sites and to measure non-synonymous divergence ($K_a$) divide the number of non-synonymous substitutions with the number of non-synonymous sites.

Substitution rates

The basic quantity in molecular evolution is the rate of nucleotide substitution or the rate with which sequences evolve. This quantity can be used to study substitution rate variation between organisms as well as within organismal genomes, e.g. between genes. The substitution rate is often defined as the number of substitutions per site per year and can be estimated by dividing $K$ with $2T$.

A common way of studying substitution rate variation between genomic regions is to compare divergences using sequences from the same species. By this one can ignore the time of divergence and lineage-specific mutation rates. The latter might be important if different lineages evolve with different rates. For instance, it has been proposed that organisms with a short generation time should have a faster rate of evolution, since they go through more rounds of cell division per unit time than organisms with a long generation time. This is called the generation time effect, which might explain why rodents evolve faster than primates (Waterston et al. 2002; Wu and Li 1985).
The rate of substitution is dependent on the rate of mutation and the probability of fixation (Hartl and Clark 1997). If the neutral mutation rate is \( \mu \) per generation, the number of mutations arising in a diploid population with size \( N \) is \( 2N\mu \) per generation. The probability of fixation is equal to the initial frequency of a new neutral mutation, which is \( 1/2N \). Remarkably simple, for neutral mutations the rate of substitution (\( K \)) equals the rate of mutation, since \( K = 2N\mu \times 1/2N \). Consequently, substitutions in selectively neutral sequences can be used to infer mutation rates and patterns of mutations.

Selection

Non-synonymous substitution rates are usually much lower than synonymous rates, although it can vary a lot among different genes. Some proteins, like histones and ribosomal proteins are extremely conserved (Li 1997), while proteins involved in reproduction (Swanson and Vacquier 2002) and immune defence (Hughes and Yeager 1998) evolve rapidly. Rate variation both between synonymous and non-synonymous sites and among proteins is usually attributed to differences in the intensity of purifying selection. Mutations that result in a substitution of an amino acid often tend to have a deleterious effect of the function of a protein. As a result most non-synonymous changes will be eliminated from the population by purifying selection, and the rate of fixation for these mutations is reduced compared to neutral mutations. However, sometimes a non-synonymous mutation improves the function of a protein. Such mutations are positively selected and the rate of fixation elevated. One can study selection at the molecular level by using the ratio between \( K_a \) and \( K_s \), referred to as the \( \omega \) parameter (Yang 1998; Yang and Nielsen 1998). \( K_s \) in this case, is a measure of the neutral mutation rate (Hurst 2002). For an amino acid change that is deleterious, \( \omega < 1 \). An \( \omega \) parameter significantly higher than one is evident of positive Darwinian selection (Yang and Bielawski 2000).
Non-coding DNA and synonymous sites in protein coding DNA have been thought to be neutrally evolving, but evidence suggests that this is not always the case. It has for instance been demonstrated in introns that nucleotides close to exon-introns junctions are functionally constraint, most likely due to conservation of splice and donor sites (Chamary and Hurst 2004; Halligan et al. 2004). Sequence conservation has also been observed inside introns (Shabalina and Spiridonov 2004). The function of such conserved nucleotides is not entirely known but it has been speculated that it somehow involves RNA (Shabalina and Spiridonov 2004). One type of constraint on synonymous sites is the codon usage bias, or the unequal use of codons encoding the same amino acid (Eyre-Walker 1991). It is believed that the codon usage is correlated with the relative abundance of tRNA molecules that interact and translates the codons. Indeed, in *D. melanogaster, E. coli* and yeast (*Saccharomyces cerevisiae*) the codon usage is biased towards codons that encode the most abundant tRNA’s for each amino acid (Akashi 1997). However, the codon usage bias varies among genes and seems to correlate with protein abundance.

One important thing to bear in mind is that even weak selection on non-coding DNA and synonymous sites will violate the assumption of neutrality if the population size is large, as is the case for *D. melanogaster* and *E. coli* (Hartl and Clark 1997).

Rate variation in neutral DNA

Several studies have reported a variation in substitution rates in presumably neutrally evolving DNA among different genomic regions (Waterston et al. 2002; Wolfe et al. 1989). Such variation has been observed at various scales, from between single nucleotides up to among chromosomes. A well-known example of substitution rate variation among nucleotides is the faster substitution rate of transitions compared to transversions (Li 1997). Such rate variation is also affected by the sequence context, which is the influence of
neighbouring nucleotides on the substitution rate (Ellegren et al. 2003). The best-known and strongest sequence context effect is the hypermutability of a methylated cytosine in a CpG dinucleotide, which increases the mutations from C to T. Ebersberger et al. (2002), reported that transitions at CpG sites accounted for 28% of all substitutions, whereas CpG sites only constituted 3.5% of the analysed sequence in an extensive study of sequence divergence between humans and chimpanzees.

Another scale of substitution rate variation is that within chromosomes. There seems to be regions of local similarity at scales > 1 Mb when comparing rates at synonymous sites in neighbouring genes (Lercher et al. 2001; Malcom et al. 2003).

The largest scale of variation is that between chromosomes. It has been shown that the mammalian chromosomes 19 and 21 are the chromosomes with the highest substitution rates, while the X chromosome is the chromosome with the lowest rate (Ebersberger et al. 2002; Lercher et al. 2001; Waterston et al. 2002).

Clearly, there are large-scale variations in substitution rates, but it is not entirely clear what genomic features are causing this variation. However, a well-known phenomenon is the positive correlation between substitution rates and GC content (Hurst and Williams 2000; Smith and Hurst 1999). Mammalian and bird genomes are characterised by the existence of long regions (>300 kb) of relatively homogenous GC content, termed isochores (Eyre-Walker and Hurst 2001). Studies on several mammalian lineages have shown that isochores are not stable over time and are decaying, hence the base composition is not stable over time (Duret et al. 2002; Smith et al. 2002). In particular, excess of GC to AT substitutions have been observed in regions of high GC content. When isochores are decaying, a positive correlation between substitution rates and GC content is expected, if GC to AT mutations are more common than AT to GC and both rates are constant across the genome (Piganeau et al. 2002).
In addition to base composition, it has been shown that recombination rates correlate with substitution rates, which suggests that recombination is somehow mutagenic (Hellmann et al. 2003; Lercher and Hurst 2002). However, the relative importance of base composition and recombination is not clear as they are strongly correlated and it is possible that recombination influences GC content (Meunier and Duret 2004). An observation in line with this is the differences in substitution rates between macro- and micro-chromosomes in birds. The micro-chromosomes are GC-rich and chicken-human divergences at synonymous sites are also higher for micro-chromosomes than intermediate and macro-chromosomes (International Chicken Genome Sequencing Consortium). This could have arisen because of the higher recombination rate for micro-chromosomes, given recombination is mutagenic.

Another possibility that could explain the variation in synonymous substitution rates is a clustering of genes with similar expression patterns (Williams and Hurst 2000; Williams and Hurst 2002). In a recent study it was found that broadly expressed genes had low intronic substitution rates (Webster et al. 2004). As an explanation it was suggested that such genes are more likely to be expressed in the germ line and therefore have reduced mutation rates due to the transcription coupled repair (TCR) mechanism (Svejstrup 2002). The TCR mechanism is an additional repair mechanism that is working in actively transcribed regions.

Nucleotide diversity

In neutrally evolving sequences new mutations are either fixed or lost in a population through a process known as random genetic drift. The mean time to fixation of new mutations is dependent on the population size (N) and is approximated to 4N generations (Hartl and Clark 1997). In a large population there is a great input of new mutations, which slowly drift to loss or
fixation and therefore levels of diversity are usually high. In small populations fewer mutations enter the population and drift is more pronounced, consequently levels of diversity is normally reduced (Aquadro 1992). However, to correct for the fact that not all individuals reproduce, the effective population size ($N_e$) is used. Most often $N_e$ is much smaller than $N$ and might for instance be influenced by sex ratio, fluctuating population size and mating system (Hartl and Clark 1997). The expected diversity ($\theta$) at equilibrium depends on the mutation rate, $\mu$ and $N_e$. The parameter $\theta = 4 N_e \mu$ is known as the population mutation rate.

The amount of polymorphism in a sample is commonly measured as the number of segregating sites, corrected for sequence length, $\theta_W$ (Watterson 1975). Another measure that takes into account the frequencies of the variants at each segregating site is the average nucleotide diversity, or $\pi$ (Nei 1987). This is a measure of the average number of differences per site between two randomly chosen sequences in the sample. Both $\theta_W$ and $\pi$ are approximations of the population mutation rate, $\theta$. If there are no effects of population subdivision, population growth or selection, the two measures should be the same (Tajima 1989). In the case of population growth and selection, there will be relatively many rare genetic variants (Li 1997). $\theta_W$ weights rare genetic variants relatively highly, while $\pi$ is an average and thus less sensitive to rare variants.

Naturally, organisms differ in their amount of intraspecific sequence variation, as do populations within species. In humans, $\pi$ has been estimated to 0.074 %, which corresponds to one polymorphic site in 1000 base pairs (Sachidanandam et al. 2001). This is lower than in many other species and could be a result of the small effective population size in humans (Li and Sadler 1991). In contrast, in Drosophila melanogaster, $\pi$ averaging over several populations has been estimated to 1.58 % (Andolfatto 2001). $\pi$ in chicken has been estimated to 0.65 %, which corresponds to one polymorphic site every 39 base pair (Sundstrom et al. 2004).
Selection

Positive and negative selection affect the rate of fixation for new mutations, and consequently the level of polymorphism. Both positive and negative selection tend to reduce genetic variation (Li 1997). In the case of positive selection, selected alleles will rapidly go to fixation and the genetic variation will be reduced around the selected site at a distance that is determined by the rate of recombination (Nachman 2001). This event is known as a selective sweep or genetic hitchhiking and describes a reduction in genetic variation after a selective event. A short period of time after a selective sweep, the allele frequency spectrum will be skewed towards rare alleles (Braverman et al. 1995). This is because after the fixation of a selected allele, new variation will arise through the input of new mutations and new mutations are initially at low frequencies in a population (Li 1997). A negatively selected mutation will also tend to reduce genetic variation at linked sites through a process called background selection (Figure 1) (Charlesworth et al. 1993; Nachman 2001).
Sex chromosome evolution

A wide range of species including mammals, birds, reptiles, amphibians, insects and plants has differentiated sex chromosomes (Charlesworth 2002; Lucchesi 1999; Marshall Graves and Shetty 2001; Schmid and Steinlein 2001). The majority of sex chromosome systems are either organized as XX/XY with male heterogamety (Figure 2a) or ZZ/ZW with female heterogamety (Figure 2b). Most likely, the sex chromosomes were once a homologous pair of autosomes but during the course of evolution one of the proto-sex chromosomes became differentiated from the other (Charlesworth and Charlesworth 2000). This could happen if suppression of recombination between the proto-sex chromosomes for some reason is selectively advantageous. Consider a hypothetical autosomal dominant sex-determining locus where M+ is a dominant allele causing male development (i.e. male genotype = M+M- and female = M-MM+). Genes that are advantageous for males but not
for females, so called sexually antagonistic genes, then accumulate close to $M^+$. Unless there is recombination between $M^+$ and these genes, they will only be inherited from father to son. This selectively advantageous situation, with no recombination, keeps these genes from being present in females. One can hypothesise that the male specific region gradually increases in size when more and more sexually antagonistic genes accumulate, eventually leading to recombination being restricted to the pseudoautosomal region(s) where recombination is necessary for correct segregation of the chromosomes at meiosis. The chromosomes carrying $M^+$ will continue to recombine in females, but the male specific chromosome ($F^+$) will never exchange any material with another chromosome (Rice 1996).

Empirical evidence supporting the above-described scenario for the evolution of sex chromosomes is the finding of ‘evolutionary strata’ along the X chromosomes in humans (Lahn and Page 1999) and mice (Sandstedt and Tucker 2004). By measuring divergence at synonymous sites ($K_s$) between related genes on the X- and Y-chromosomes, Lahn and Page (1999) discovered that $K_s$ fell into four distinct groups, or ‘strata’. Interestingly, the groups corresponded to the order of the genes along the X chromosome. They hypothesised that sequential inversions of segments along the Y chromosome could have formed the strata by the termination of recombination between groups of continuous X chromosome genes and their Y linked homologs. Similar to what has been observed for X-linked genes in mammals, there is evidence for a stepwise termination of recombination between the Z- and W-chromosomes (Handley et al. 2004; Lawson Handley et al. 2004). In this study $K_s$ were estimated between five genes located on both the Z- and W-chromosomes and results suggest the presence of at least two evolutionary strata.
The Y and W

The Y chromosome in mammals and the W chromosome in birds are sex specific and never undergo interchromosomal recombination in their sex specific regions. A typical Y chromosome contains two distinct domains, the pseudoautosomal region (PAR), where recombination occurs with corresponding PAR of the X and the male specific region (MSY) (Skaletsky et al. 2003). Correspondingly, the W chromosome contains a female specific region outside the PAR. The most prominent feature of the mammalian Y and the avian W is that they are genetically degenerated in their sex specific regions, which is evident from their small sizes and low gene contents (Skaletsky et al. 2003; Smith and Burt 1998). For comparison, the X and Z are large gene rich chromosomes (Marshall Graves and Shetty 2001; Smith and Burt 1998). Furthermore, Y chromosome degeneration is also evident from the dosage compensation of X-linked gene so that the activity of most X-linked genes is approximately the same in males and females.
Dosage compensation also seems to occur in birds, although the mechanism is not known (Ellegren 2002).

There are a number of evolutionary forces that might trigger the degeneration of Y and W chromosomes. First of all, once low recombination between the young sex chromosomes has evolved, genes on the non-recombining part of Y or W are in a very peculiar genetic situation, since they are ‘doomed’ to permanent heterozygosity (Charlesworth and Charlesworth 2000). This makes natural selection very inefficient, since deleterious mutations are hidden by the ‘healthy’ gene-copy on the non-degenerated chromosome. As a consequence, the likelihood of accumulation of deleterious mutations on Y and W increases (Charlesworth and Charlesworth 2000). Elevated K_a/K_s-ratios of Y-linked genes compared to X-linked in mammals, is in agreement with this hypothesis (Wyckoff et al. 2002).

Another important feature of Y- and W-linked genes are their small effective population sizes (N_e) compared to X, Z and autosomes. There are \( \frac{1}{3} \) as many copies of Y-linked (and W) genes as X-linked (and Z) and \( \frac{1}{4} \) as many copies as autosomeal. This makes Y- and W-linked deleterious mutations especially likely to become fixed by genetic drift (Charlesworth and Charlesworth 2000). Another factor that is thought to affect N_e of sex chromosomes is a skew in mating success among males (Sundstrom et al. 2004). This will reduce N_e for the male specific chromosome further (Y and Z).

In summary, lack of recombination and low population sizes make sex specific chromosomes vulnerable to the accumulation of deleterious mutations. This may happen through various population genetic processes. One such process is called ‘Muller’s ratchet’ (Felsenstein 1974; Muller 1964). This involves the process of stochastic loss of the least mutation-loaded chromosome by genetic drift. Since there is no recombination and back mutations are quite unlikely, the chromosome carrying the fewest deleterious
mutations cannot be restored. The speed at which the ratchet moves is dependent upon \( N_e \) and the rate of mutations. If everything else is equal, the process should go faster as the population size gets smaller. However, the ratchet can move rapidly even in very large populations if there are many mutations with small effects on fitness (Charlesworth and Charlesworth 2000).

Another process to affect the rate of fixation of deleterious mutations on a sex specific chromosome is called background selection (Charlesworth et al. 1993). This process has already been mentioned since it might affect genetic variability around a selected site. This process is very similar to Muller’s ratchet, but background selection can operate with selection coefficients and population sizes where the ratchet cannot, and it can reduce the fitness of Y- and W-chromosomes in large populations over long periods of time. A new mutation arising on a chromosome carrying a deleterious mutation will soon be eliminated from the population and has therefore non-zero chance of fixation. Thus, the effective population size of a non-recombining chromosome is therefore reduced to include only those free of deleterious mutations and consequently the rate of fixation of mildly deleterious mutations is accelerated, and fixation of advantageous mutations retarded.

A third process named the Hill-Robertson effect is operating as the lack of recombination makes the formation of the fittest combination of alleles impossible (McVean and Charlesworth 2000). This takes place if new selectively advantageous mutations arise independently at two sites and without recombination it is not possible for the two alleles to be located on the same chromosome. It is believed that the Hill-Robertson effect needs considerable time to erode the fitness of an evolving Y- or W-chromosome because new mutations might have to reach fixation at numerous sites along the chromosome. One interesting consequence of the Hill-Robertson effect is that the rate of adaptive evolution should be higher for genes with higher than average rates of recombination (Marais and Charlesworth 2003).
The fourth and last process that might contribute to the degeneration of Y- and W-chromosomes has already been mentioned and is the hitchhiking of advantageous mutations. In the absence of recombination all linked variation around a selected site will go to fixation and thereby reduce the genetic variability in the population. This can obviously also contribute to the degeneration of a non-recombining chromosome if the selective sweep also drag a deleterious mutation to fixation. The efficacy of this process has been questioned since the favourable mutations must be as strongly selected as the deleterious mutations to have a good chance of fixation.

Z- and W-linked genes – Gametologs

Z and W linked copies of a gene share a common ancestry from the time the sex chromosomes were an ordinary pair of autosomes but after the cessation of recombination they began to evolve independently. Such genes located on both sex chromosomes are termed gametologs (Garcia-Moreno and Mindell 2000).

At the point of writing there are eight known genes shared between the avian Z- and W-chromosomes (Table 1). Of these genes, the molecular evolution of ATP5A1 (Carmichael et al. 2000) and CHD1 (Ellegren and Fridolfsson 1997; Fridolfsson and Ellegren 2000) have been studied in some detail. Unlike the other known genes on the avian sex chromosomes, HINTW (Hori et al. 2000; O'Neill et al. 2000; Pace and Brenner 2003) is present in multiple copies. The phenomenon of multiplicity has also been reported for Y-linked genes (Skaletsky et al. 2003). Since these genes are predominantly expressed in the testis it has been proposed that some male specific function is driving the evolution of such genes, for instance sperm competition among males. Obviously sperm competition cannot explain why the female specific HINTW is present in multiple copies. Parts of HINTW is evolving under positive selection (Ceplitis and Ellegren 2004). Positive selection has also observed for genes involved in male reproduction and it has been hypothe-
sised that such rapid evolution is a consequence of sexual selection on male reproductive traits (Swanson and Vacquier 2002; Wyckoff et al. 2000). It is however not clear how such predictions would apply to genes on the W chromosome, unless sexual selection also operates in females. Moreover, it has been proposed that gene conversion between repeats of multi-copy genes on non-recombining chromosomes provide a mechanism to conserve gene function through evolutionary times (Skaletsky et al. 2003). This theory is supported by the fact that multiple copy genes are more frequently found on the Y and W chromosomes, which are more prone to accumulate deleterious mutations due to the lack of inter-chromosomal recombination. Finally, one more gene is known on the W chromosome, FET1, with a possible role in ovarian development (Reed and Sinclair 2002). A Z chromosome gametolog has not yet been discovered.

Table 1. Gametologous genes shared between the Z and W chromosomes in birds.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>No. of copies</th>
<th>Stratum</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD1</td>
<td>Z, W = 1</td>
<td>1</td>
<td>(Ellegren 1996; Griffiths et al. 1996; Griffiths and Korn 1997)</td>
</tr>
<tr>
<td>SPIN</td>
<td>Z, W = 1</td>
<td>1</td>
<td>(Itoh et al. 2001)</td>
</tr>
<tr>
<td>UBAP2</td>
<td>Z, W = 1</td>
<td>2</td>
<td><a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a></td>
</tr>
<tr>
<td>ATP5A1</td>
<td>Z, W = 1</td>
<td>2</td>
<td>(Carmichael et al. 2000)</td>
</tr>
<tr>
<td>HINT</td>
<td>Z = 1, W &gt;40</td>
<td>1</td>
<td>(Hori et al. 2000; O'Neill et al. 2000; Pace and Brenner 2003)</td>
</tr>
<tr>
<td>MADH2</td>
<td>Z, W = 1</td>
<td>2</td>
<td><a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a></td>
</tr>
<tr>
<td>IDN2</td>
<td>Z, W = 1</td>
<td>2</td>
<td><a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a></td>
</tr>
<tr>
<td>RASA1</td>
<td>Z, W = 1</td>
<td>1</td>
<td><a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a></td>
</tr>
</tbody>
</table>
Nucleotide diversity and sex chromosomes

As indicated above, \( N_e \), mutation and selection affect the level of polymorphism. It is therefore expected that neutral sequence variation for autosomal (A) and sex linked sequences should differ as the chromosomal classes also differ with respect to these underlying population parameters. In a random mating population with an even sex ratio, autosomes are expected to have the highest diversity because of the larger effective population size compared to the sex chromosomes. Specifically, Y and W sequences are expected have \( \frac{1}{4} \) the diversity of autosomes, and X and Z \( \frac{3}{4} \). Similarly, Y sequences should have \( \frac{1}{5} \) the diversity of X and W sequences to have \( \frac{1}{5} \) the diversity of Z.

Mutation rates typically differ between autosomes and sex chromosomes as well as between the sex chromosomes. As is explained in more detail below, the chromosome that spend most time in the male germ line is expected to accumulate more mutations and consequently display the highest diversity, that is, \( Y > A > X \) and \( Z > A > W \).

Both positive and negative selection will have different effects on autosomes and sex chromosomes. Since the consequences of selection on the level of diversity is strongly correlated with the rate of recombination both genetic hitchhiking and background selection will drastically reduce diversity on the non-recombining Y and W chromosomes. Similarly, hitchhiking and background selection are expected to more strongly affect Z- and X-linked genes compared to autosomal, since Z and X chromosomes do not recombine in the heterogametic sexes.

Nucleotide diversity in the non-recombining region of the human Y chromosome (NRY) has been estimated to 0.01-0.015 % (Sachidanandam et al. 2001; Shen et al. 2000), which is about 20% of the autosomal average (Sachidanandam et al. 2001). The human X chromosome has a higher diversity, with \( \pi \) estimated to 0.045 % (Sachidanandam et al. 2001), which is 60% of the autosomal diversity. \( \pi \) for Z-linked sequences in chicken, has been
estimated to 0.2% (Sundstrom et al. 2004), which is about 30% of the autosomal value. Although correcting for differences in $N_e$ and mutation rates, Z diversity remained lower than A and Sundstrom, Webster, and Ellegren (2004) concluded that the reduction of diversity on Z compared to A must be a consequence of natural selection. This is likely due to larger effects of selective sweeps on the Z chromosomes due to lower recombination rates compared to autosomes. This is supported by heterogeneity in diversity levels among Z-linked loci.

The male mutation bias

Mutations are primarily thought to arise due to errors in DNA replication at cell division (Haldane 1935; Miyata et al. 1987). According to this hypothesis, more mutations should arise in the male germ line compared to the female germ line, assuming there are more cell divisions in male germ lines compared to female germ lines (Li et al. 2002). The ultimate way to test this would be to compare sequences of parents and their offspring. However, a reliable estimate of the number of DNA-changes arising between generations would require extensive sequencing, considering that the mutation rate is in the range of $10^{-9}$ per nucleotide (Li 1997). Therefore another method is needed to estimate DNA-changes and as already mentioned one can use substitutions between species.

In 1987, Miyata et al. proposed a method known as the evolutionary approach for estimating the male to female mutation rate ratio, $\alpha_m$. The basic rationale of the evolutionary approach is to compare the substitution rates of the two sex chromosomes or of a sex chromosome and autosomes (A). This method can be applied to any organism with heterogametic sex chromosomes irrespective of which sex is the heterogametic (Figure 3). In birds, a W chromosome spends all its time in female germ line, while a Z-chromosome spends $\frac{1}{2}$ of its time in female germ line and $\frac{1}{2}$ in male germ
line. Assuming errors in DNA replication is a major source of mutations and that there are more cell divisions in spermatogenesis than oogenesis, it is expected that a chromosome that spends more time in the male germ line (Z chromosome) should accumulate relatively more mutations than a female specific chromosome (W chromosome). Therefore the Z chromosome should evolve faster than the W chromosome.

From estimates of substitution rates for Z- and W-linked sequences ($K_z$ and $K_w$) one can calculate $\alpha_m$ by taking into account the time each chromosome spends in each sex (Miyata et al. 1987). The expected mutation frequency per generation for the Z chromosome relative to the W chromosome is $(2/3 \alpha_m + 1/3):1$, and with a little rearrangement this yields an estimator of $\alpha_m = \frac{1}{2} (3K_z/K_w - 1)$.

---

**Figure 3.** Time chromosomes spend in male and female germ lines in species with male (a) and female heterogamety (b).
The strength of the male mutation bias

In order to test if estimates of $\alpha_m$ is in line with what to expect from the differences in numbers of cell divisions between the sexes one needs to know a parameter known as $C$ (Hurst and Ellegren 1998). $C = n_m/n_f$, where $n_m$ is the number of cell divisions in the male germ line and $n_f$ is the number of cell divisions in the female germ line. If DNA replication is responsible for the majority of mutations, $C$ and $\alpha_m$ should be positively correlated. Rough anatomical estimates of $C$ in mammals are consistent with such a correlation. In humans, female germ cells undergo 24 divisions in total, whereas male germ cells at the age of 20 have undergone 150 cell divisions and as many as 600 at the age of 40 (Hurst and Ellegren 1998). Thus, with a mean age of reproduction of 20 years, $C$ is six and this is close to estimates of $\alpha_m$ (Makova and Li 2002; Shimmin et al. 1993). Similar calculations of $C$ in rodents predict $\alpha_m$ to be approximately two, which has also been observed (Chang et al. 1994; Gibbs et al. 2004). Interestingly, a whole genome analysis of substitutions of the X-chromosome and autosomes in rodents also suggest $\alpha_m$ to be approximately two (Makova et al. 2004). $C$ has also been estimated in birds, based on information on Japanese quail and chicken (Kahn and Quinn 1999). Female germ cells are thought to undergo approximately 20 cell divisions per generation, while male germ cells in an average male at an average age of time of reproduction have undergone approximately 89 germ cell divisions. According to these figures, $\alpha_m$ is expected to be about 4 in galliform birds. Estimates of $\alpha_m$ in birds have so far revealed somehow different results, depending on the species used to estimate Z and W divergences. The values range between 2 and 6 (Carmichael et al. 2000; Ellegren and Fridolfsson 1997; Garcia-Moreno and Mindell 2000; Kahn and Quinn 1999). However, $\alpha_m$ has been estimated to 2.5 using divergences between chicken and turkey (Axelsson et al. 2004) indicating that $\alpha_m$ is less than expected from the difference in cell divisions between the sexes for galliforms.
Obstacles with the evolutionary approach

**Reduction of mutation rates on the hemizygote chromosome**

A second process thought to affect mutation rate variation between the sex chromosomes is the adaptive reduction of mutation rates on the X and Z chromosomes due to hemizygous exposure of recessive deleterious mutations in mammalian males and avian females (McVean and Hurst 1997). It means that higher substitution rates in Y- compared to X-linked sequences could not be explained by a male biased mutation rate alone. This might be a problem when estimating $\alpha_m$ using X- and Y-linked sequences in mammals, which then would be overestimated. Empirical support for a reduced mutation rate on X comes from studies of human and mouse (Lercher et al. 2001; McVean and Hurst 1997), however the current evidence for such an adaptive reduction in the X-linked mutation rate of mammals is weak. For example, in a human-chimpanzee comparison of genomic sequences, the reduction in X-linked substitution rates can be explained by a male mutation bias alone (Ebersberger et al. 2002). A similar conclusion was also reached from extensive human-mouse and mouse-rat comparisons using synonymous substitution rates (Malcom et al. 2003).

This issue has also been studied using divergences between Z, W and autosomal sequences in birds (Axelsson et al. 2004). If the mutation rate on the Z chromosome was reduced, $\alpha_m$ would be underestimated. This was tested by estimating $\alpha_m$ using divergences between Z/A, Z/W and A/W. If the differences in substitution rates between the chromosomal classes are only due to a male biased mutation rate, no differences between the estimates are expected. However, if Z chromosome mutation rate is reduced, $\alpha_m$ estimated from A/W should be higher than estimates using Z/A and Z/W. As the $\alpha_m$ estimates of the three comparisons did not differ significantly there was no evidence for a specific reduction in the Z chromosome mutation rate.
Ancestral polymorphism

When Makova and Li (2002) used substitutions in non-coding DNA in humans and greater apes for estimating $\alpha_m$ they observed an interesting pattern. They found that $\alpha_m$ estimated from internal branches were higher than $\alpha_m$ estimated from terminal branches, or in other words $\alpha_m$ estimated from closely related species were lower than estimates obtained from more distantly related species. A probable explanation is the difference in the amount of ancient polymorphism for X- and Y-linked sequences. The average divergence between two species is equal to $\pi + 2\mu t$, where $\pi$ is the average divergence between two sequences in the ancestral population (ancient nucleotide diversity) at the time of speciation (Figure 4). The polymorphism for Y linked sequences is usually very low, compared for the situation for X linked sequences (Hellborg and Ellegren 2004). Consequently, pre-existing polymorphism for X chromosome sequences can be high enough to substantially underestimate the Y/X ratio and consequently $\alpha_m$. Pre-existing polymorphism will only bias the estimates when using closely related species when divergences are low. Therefore too closely related species are not suitable for estimating $\alpha_m$. The corresponding situation for Z/W sex chromosomes is that the Z/W ratio and $\alpha_m$ will be overestimated.
Figure 4. The difference in preexisting polymorphism for Z (X) linked and W (Y) linked sequences.

The study organisms – Birds

Birds are and have been extensively studied both in terms of ecology and evolution and in recent years there has been a growing interest in avian genetics. Chicken serves as the primary genetic model organism in avian genetics and at the time of writing the chicken genome sequence is about to be published (Chicken genome sequencing consortium). Chicken belongs to the order Galliformes, which is a basal group in the bird phylogeny (Garcia-Moreno et al. 2003). In this thesis, sequence variation has been studied in several species from this order.

Furthermore, sequence variation has also been studied in a variety of species from several other bird orders, using genetic markers developed in chicken. Hence, the similarities between the chicken genome and the genome of other species have made it possible to extract and use the detailed information available for the chicken to the other species used in this thesis.

This thesis is just at the starting point of putting the chicken genome data to use. Clearly, the publication of the chicken genome opens up incredible possibilities for the study of molecular evolution and genetics in birds.
Research aims

The general aim of this thesis is to study effects of mutation and selection on divergence and diversity on sex linked and autosomal genes in birds. Specifically we studied:

- If the strength of the male mutation bias in birds varies with life history characters and if estimates of the male mutation bias using the evolutionary method is sensitive to what sequences are used to estimate divergences.

- The degeneration of genes on the W chromosome compared to the Z chromosome.

- The degree of nucleotide diversity on the female specific W chromosome in chicken compared to the nucleotide diversity on the Z chromosome.

- If positive Darwinian selection drives the evolution of the female reproductive protein, ZPC, in birds and other vertebrate species.
Present Investigation

Paper I. Life history and the male mutation bias

The fact that more mutations are created in the male germ line compared to the female germ line is now well established but proper quantification and understanding of the extent of this bias are in focus of current research (Li et al. 2002). The main aim of the study was to test if the strength of the male mutation bias ($\alpha_m$) was sensitive to factors thought to affect the ratio of male to female cell divisions ($C$).

If the male mutation bias is a result of the higher number of cell divisions in the male ($n_m$) than in the female ($n_f$) germ line, those life history characters that are expected to affect the ratio of male germline cell divisions to female germline cell divisions ($C$), should also affect the male to female mutation bias ($\alpha_m$). Two such life history traits are sperm competition and generation time. Sperm competition may affect $C$, because an increase in sperm competition induces increased sperm production. Increased sperm production is based on an increase in $n_m$, which in turn increases $C$. It was thus expected that the intensity of sexual selection, as manifested in sperm production, was positively correlated with $\alpha_m$. The second life history character, the generation time, was thought to affect $\alpha_m$ since the longer the generation time, the greater the expected excess in $n_m$. 

34
Methods

31 species were included and prior to the analyses split into three groups: (1) 22 species of Passeriformes, (2) six species of Charadriiformes + Procellariiformes and (3) three species of Galliformes + Anseriformes. Parsimony and Bayesian analyses were used to determine the correct phylogenies for each group using a concatenated alignment of Z and W data.

For any set of sequences, a point estimate of $D_m$ was obtained using the total Z and W tree lengths. Confidence intervals were estimated of $D_m$ by nonparametric bootstrapping. 1000 bootstrapped tree lengths were generated for both Z and W sequences. This produced 1000 independent estimates of $K_z/K_w$ and hence $D_m$.

In order to investigate whether certain lineages or clades had evolved at different rates since their divergence from a common ancestor, relative rate tests were performed. Such analyses made it possible to determine if the differences in $D_m$ between different clades in the passerine phylogeny were due to differences in $K_z$ or $K_w$.

Frequencies of extra-pair paternity (EPP) were used as a measure of sexual selection for each species and age at first breeding were used as a proxy of generation time.

Results

The data demonstrated considerable variation in $D_m$ between different clades of birds. Within the order Passeriformes, $D_m$ in clades 1, 2 and 3 was estimated to be 3.48, 2.43 and 1.70. $D_m$ in Passeriformes 1 was significantly greater than $D_m$ in both Passeriformes 2 ($p = 0.04$) and Passeriformes 3 ($p = 0.005$ by bootstrapping), although Passeriformes 2 did not have significantly higher $D_m$ than Passeriformes 3 ($p = 0.12$). In the orders Charadriiformes + Procellariiformes, $D_m$ was estimated to be 4.74 and in Galliformes + Anseri-
forms $\alpha_m$ was 1.77. $\alpha_m$ in Charadriiformes + Procellariiformes was significantly greater than $\alpha_m$ in Galliformes + Anseriformes ($p < 0.001$ by bootstrapping).

EPP frequency was much more pronounced and variable in passerines in comparison with the other bird species included in the study. Passeriformes 1, 2 and 3 had average EPP frequencies of 28%, 23% and 14%, although the differences were not significant (Mann Whitney U tests, $p > 0.05$). However, the ranking of mean EPP values for clades matched the ranking of $\alpha_m$ for these clades. For instance, Passeriformes 3 with the lowest mean EPP frequency also had the lowest value of $\alpha_m$ (1.70). This indicated that EPP may positively covary with $\alpha_m$ in Passeriformes.

The Charadriiformes + Procellariiformes species had a mean generation time of 5.1 years and the Galliformes + Anseriformes species had a mean generation time of 1.8 years, which was a significance difference (Mann Whitney U test, $p = 0.031$). The latter group also had a significantly lower mean $\alpha_m$ (1.77 vs. 4.74) so there was positive covariation between generation time and $\alpha_m$.

The relative rate test revealed as predicted that the Z-rate was significantly higher in Passeriformes 1 than in Passeriformes 3 (rate ratio 1.47, $p = 0.008$ by bootstrapping) whereas the difference between Passeriformes 1 and Passeriformes 2 was non-significant (rate ratio 1.09, $p = 0.241$). However, the W chromosome rate was low in both comparisons (rate ratios 0.49 and 0.50, $p < 0.001$).

Discussion

The study showed that clades differed considerably in the relative extent of the male mutation bias (1.7-4.7). There was indication that both life history characters; generation time and extra-pair paternity, affected the male mutation bias in the direction predicted from germ cell biology. Males should
produce more gametes, and therefore have more cell divisions, as generation
time and sperm competition increase.

This opens a new and interesting perspective to the theory of evolution of
sexual selection. Sexual selection arises from competition among individuals
of the chosen sex for access to individuals of the choosy sex, and females
(choosy sex) may either obtain direct or indirect fitness benefits from their
mate choice. Indirect fitness benefits may consist of genetically based attrac-
tiveness of sons or genetic viability of offspring. The maintenance of indirect
fitness benefits in the presence of intense directional selection by choosy
females poses a theoretical problem, because characters subject to intense
directional selection tend to go to fixation leaving little genetic variation. An
increased male mutation bias in birds with frequent extra-pair paternity
could suggest that mutational input increases in species subject to intense
sexual selection as a direct consequence of sperm competition.

One reservation was the results from the relative rate tests. An important
assumption concerning the predicted effects of life history characters on $\alpha_m$
was that variation in C was solely due to the number of male germline cell
divisions. This means that $K_z$ but not $K_w$ should vary in accordance with $\alpha_m$
variation. However, this prediction did not hold in one comparison, Passeri-
formes clades 1 and 3, where the significantly higher $\alpha_m$ in clade 1 was due
to a lower $K_w$ rather than higher $K_z$. The consequence of this observation
remains unclear.

Paper II. Local mutation rate heterogeneity and the male
mutation bias

The main aim of paper II was to explore variation in estimates of $\alpha_m$ among
gametologous introns to see if heterogeneity in substitution rates among
introns affected the estimates. Several studies have reported regional varia-
tion in substitution rates in putatively neutrally evolving DNA among different genomic regions (Ellegren et al. 2003). Such variation has also been observed for sequences on the sex chromosomes in birds (Axelsson et al. 2004).

Recent studies have demonstrated functional constraints in introns, which would obscure the important assumption of neutrality for the use of substitution rates to estimate mutation rate (Chamary and Hurst 2004; Halligan et al. 2004). In particular, nucleotides close to exon-intron boundaries seem to be functionally constraint, but also nucleotides inside introns can have some regulatory function and are therefore more conserved than neutrally evolving nucleotides (Shabalina and Spiridonov 2004). In order to investigate sequence conservation, divergences in intron ends were compared with divergences inside introns. In addition, conserved blocks inside the introns were searched for using Z- and W-alignments. The hypothesis was that conserved blocks in the Z/W-alignments reflect functional constraints. Z and W sequences are so divergent, therefore it is expected under a neutral model of evolution that random substitutions have erased most sequence similarity.

In addition, the use of five species made it possible to test the molecular clock hypothesis by studying lineage specific substitution rates (Gillespie 1991; Yi et al. 2002).

Methods

In total 18 Z- and W-linked gametologous introns were sequenced in five species from the order Galliformes; chicken (Gallus gallus), common quail (Coturnix coturnix), red-legged partridge (Alectoris rufa), black grouse (Tetrao tetrix) and turkey (Meleagris gallopavo).

Substitution rates using parsimony were estimated in the intron ends in non-overlapping sequence blocks at various distances from the intron-exon boundaries and compared it with the rates in the interior of the introns with Fisher’s exact tests. Conserved blocks were also searched for inside the in-
trons using alignments of Z- and W-linked intron sequences. The aggrega-
tions of nucleotides in the real data were compared with simulated data.

\( \alpha_m \) was estimated following the same procedures as in paper I using \( K_z \)
and \( K_w \) estimated using total branch lengths. Heterogeneity in substitution
rates was tested among introns for each chromosomal class using G-tests.
ANOVA’s were used to test for variation in estimates of \( \alpha_m \) between the
introns.

Lineage specific effects on substitution rates and estimates of \( \alpha_m \) were
tested by relative rate tests using concatenated alignments with all introns
combined.

Results

Substitution rates were significantly lower in the intron edges, therefore the
first and last 20 base pairs of each introns were removed prior to the analy-
ses. The conserved blocks that were observed in the Z- and W-alignments
were not more common than would be possible by random aggregation of
bases. However, to reduce the possibility that some nucleotides were evolv-
ing under negative selection, conserved blocks that were 90 % identical or
more were removed. These block lengths varied between 3 to 57 nucleotides
and in total 576 nucleotides were removed.

There was statistically significant variation in \( K_z \) among introns (G test:
\( G_{15} = 36.7, p < 0.0013 \)) as well as in \( K_w \) among introns (G test: \( G = 40.2, p =
0.0004 \)). Moreover, estimates of \( \alpha_m \) based on individual introns differed sig-
ificantly (ANOVA: \( F_{13} =3.44, p = 0.014 \)). Estimates of \( \alpha_m \) ranged between
1.53 and 3.39, with a mean of 2.31.

The \( C. \ coturnix \) lineage had a significantly higher substitution rate com-
pared to the other four lineages, which could be an effect of the short gener-
ation time. There was no clear trend for lineage effects on \( \alpha_m \).
Discussion

The study demonstrated significant mutation rate variation between introns along the Z and W chromosomes. This local mutation rate variation affected estimates of $\alpha_m$, which also varied significantly among the introns. Therefore, to properly quantify the extent of the male mutation bias one should collect sequence data from various regions along the sex chromosomes. Mutation rate variation between introns could explain some of the variation in $\alpha_m$ in studies using different bird species (Carmichael et al. 2000; Ellegren and Fridolfsson 1997; Kahn and Quinn 1999).

The fast evolution of the C. coturnix lineage could be a consequence of the much shorter generation time than the other four species.

Paper III. Accumulation of deleterious mutations on the female-specific W chromosome in birds

The main aim of paper III was to compare substitution patterns in coding DNA for Z-linked and W-linked sequences. The substitutions were divided in non-synonymous (amino acid changing) and synonymous (silent) changes and an outgroup was used to infer the changes. The main purpose was to study differences in $K_a/K_s$-ratios ($K_a = $ non-synonymous substitutions per non-synonymous site, $K_s = $ synonymous substitution per synonymous site) between Z- and W-branches in a phylogeny. A comparison of $K_a/K_s$-ratios between Y and X linked genes using human-rodent alignments demonstrated elevated $K_a/K_s$ levels for Y-linked genes (Wyckoff et al. 2002). It was expected that the W chromosome would accumulate more deleterious mutations than the Z chromosome and autosomes because it is non-recombining and has a low effective population size ($\frac{1}{4}$ of autosomes and $\frac{1}{2}$ of Z-chromosomes) (Charlesworth and Charlesworth 2000).
Methods

Protein-coding sequences from three gametologous gene pairs: SPINZ/W, UBAP2Z/W and MADH2Z/W in chicken (Gallus gallus) and turkey (Meleagris gallopavo) and from a fourth gene pair: CHD1Z/W in chicken and zebra finch (Taeniopygia guttata) were included in this study. In addition, Mus musculus orthologous were used as outgroup sequences. UBAP2Z, MADH2Z/W and SPINZ/W were sequenced in turkey, while the other sequences except for chicken MADH2W (http://genome.ucsc.edu/) were collected in Genbank. Substitution rate patterns were analysed with a maximum likelihood approach using the codeml program in the PAML package. The data was analysed with different models that allowed for heterogeneity in K_a/K_s among branches in a phylogeny. The simplest model (one-ratio model) assumes the same K_a/K_s-ratio for all branches. The most parameter rich model, the free-ratio model, contains as many K_a/K_s-ratios as branches, with the two-ratio and the three-ratio models in between.

Results

In total, 1581 codons from the four genes were included in the maximum likelihood analyses. Results from the concatenated alignment implied that the K_a/K_s-ratio for W-linked genes was approximately six times higher than the corresponding value for Z-linked genes, although the ratio varied a lot among genes. A chi-square test revealed that the difference in the ratios of non-synonymous/synonymous substitutions between Z- and W-linked genes was highly significant (chi square = 38.32, p<0.0001). The values were for Z-linked genes, 16:292 and for W-linked, 50:162, which show an excess of non-synonymous substitutions for W-linked genes compared to Z-linked genes. Likelihood ratio tests indicated that the free-ratio, the three-ratio and the two-ratio models represent significantly better fit to the data than the one-ratio model (p<0.0001). However, the free-ratio and the three-ratio
models were not significantly better than the two-ratio model, indicating that the two-ratio model best described the data.

Discussion

This study shows that W-linked genes evolve in a different manner than Z-linked genes, suggesting that W-linked genes have and are accumulating more non-synonymous substitutions than Z-linked genes. Since the W-chromosome is not recombining and has a low effective population size compared to other chromosomes, this pattern is expected due to a higher chance of fixation of deleterious mutations by a reduction in effectiveness of purifying selection and stronger genetic drift. These results are in agreement with the rapid evolution found for genes located on the non-recombining Y-chromosome (Tucker et al. 2003; Wyckoff et al. 2002). Various population genetic processes might be involved in further reducing $N_e$ of W-chromosomes. Two such processes are background selection and genetic hitchhiking. Background selection is occurring since a new mutation arising on a chromosome carrying a deleterious mutation will soon be eliminated from the population and has therefore non-zero chance of fixation. Thus, the effective population size of a non-recombining chromosome is therefore reduced to include only those free of deleterious mutations and consequently the rate of fixation of mildly deleterious mutations is accelerated. Genetic hitchhiking or selective sweep happens in the absence of recombination, since all linked variation around a selected site will go to fixation. This can contribute to the degeneration of a non-recombining chromosome if the selective sweep also drag a deleterious mutation to fixation.
Paper IV. The chicken W chromosome – a homogenous chromosome in a highly variable genome

The Y chromosome of organisms with male heterogamety is expected to show reduced levels of nucleotide diversity as the effective population size is one-fourth that of autosomes. However, studies in a wide range of species show that Y chromosome diversity is lower than expected even when differences in effective population size is taken into account (Sachidanandam et al. 2001; Shen et al. 2000). This may be explained by skewed reproductive success among males, leading to low male effective population size, or by a strong role of selection in shaping levels of nucleotide diversity in non-recombining chromosomes (Charlesworth and Charlesworth 2000). To test these hypotheses in a system with female heterogamety we estimated nucleotide diversity in the female-specific W chromosome of domestic chicken.

Methods and Results

Thirteen introns from three different W-linked genes (CHD1W, UBAP1W, SPINW) were sequenced in 47 female chickens from 10 breeds. In total, 7643 base pairs of non-coding DNA were analysed for each W-chromosome. Remarkably, only one single segregating site (C to A) was found in this screening, with the rare allele at a frequency of 0.43. The polymorphic site was at position 389 in CHD1W intron 11. When taking sex-specific mutation rates and differences in effective population size into account, the observed degree of W chromosome polymorphism was 28-fold lower than expected for SNP frequency and 13-fold lower than expected for estimates of nucleotide diversity (autosomes, 6.5x10^{-3}; W, 7.0x10^{-5}).
Discussion

A recent study of nucleotide diversity in chicken revealed extensive genetic variability for autosomal, non-coding DNA with one segregating site every 39 base pairs and $\pi$ estimated to $6.5\pm0.3\times10^{-3}$ (Sundstrom et al. 2004). The frequency of SNP’s on autosomes is 50 times higher than that on W and nucleotide diversity is 23 times higher for autosomes than W chromosomes, after correcting for the difference in $N_e$ and sex specific mutation rates. The chromosome most similar to the W is the Y chromosome in mammals. Nucleotide diversity ($\pi$) in the non-recombining region of the human Y chromosome has been estimated to $1.0-1.5\times10^{-4}$, which is about 20% of the autosomal average, whereas a 25% reduction would be predicted from a 4:1 relationship in $N_e$.

Clearly, polymorphism levels on the chicken W chromosome appear much lower than expected from comparisons with chicken autosomes. What evolutionary processes could cause such low levels? It has been hypothesized that the low levels of nucleotide diversity on the mammalian Y chromosome is at least partly due to lower $N_e$ of males than of females. Since this is not applicable on the W chromosome, our data seem best explained by a strong role of selection in shaping levels of W chromosome variability in the domestic chicken. In the lack of recombination, positive selection for an advantageous mutation leading to its fixation will lead to the fixation of all linked sites, which is the entire chromosome outside the pseudoautosomal region. In addition, background selection arising in connection with negative selection for deleterious mutations will reduce genetic diversity over the whole chromosome.
Paper V. Adaptive evolution of ZPC, a female reproductive protein in diverse vertebrate species

The aim of paper V was to study positive Darwinian selection in a female reproductive gene called Zona Pellucida C (ZPC). ZPC encodes a protein (ZPC) that is part of the membrane surrounding the vertebrate egg. This membrane is an insoluble extracellular matrix, which is called zona pellucida (ZP) in mammals, perivitelline membrane (pvm) in birds and chorion in fish. In chicken, the pvm is composed of an inner and an outer layer separated by a thin membrane (Bausek et al. 2000). The inner layer of the chicken pvm consists of two proteins, of which one is a homologue to the mammalian sperm receptor protein, ZPC. For successful fertilization to occur, sperm must first bind to the pvm, a process that is species-specific, and then penetrate it (O’Rand 1988). The aim was to study the molecular evolution of ZPC, in particular testing for signals of positive selection in six species of birds from the order Galliformes. In addition, 23 chicken individuals were sequenced to test for selection using polymorphism data. A number of tests were performed using the polymorphism data on its own or in combination with divergence data. In addition, the protein was analysed in a number of fish and mammalian species, to study the generality of molecular evolution of ZPC in vertebrates.

Methods

A phylogenetic based maximum likelihood approach was used to test for positive selection by comparing the ratio of nonsynonymous (K_a) and synonymous substitution rates (K_s). The distributions of positively selected codons and the K_a/K_s ratios along the avian ZPC sequence were determined and compared to the patterns in mammalian ZPC. Recent studies have shown that analyses of positive selection using maximum likelihood ratio tests might be too liberal when data sets contain too little sequence variation.
Therefore simulation studies were performed to test the finding of positive selection in avian ZPC. Secondly, the sequence variation within species was compared to the sequence variation between species to test for positive selection (McDonald-Kreitman and HKA tests). The intraspecific sequence data was also used on its own to test for recent selective sweeps (Tajima’s D and H tests). In addition to the bird data, ZPC sequences were obtained from four fish and ten mammalian species from public sequence databases that were analysed using the maximum likelihood approach described above.

Results

The complete coding sequence of the ZPC gene in six bird species were amplified (chicken, common quail, pheasant, turkey, black grouse and sage grouse) as well as in 23 chicken individuals. Under a model of no variation in $K_a/K_s$ among codons or branches, average $K_a/K_s = 0.13$. This low value indicates that the predominant mode of selection on the ZPC is purifying selection and it is consistent with the finding that domains within the ZPC gene are very well conserved across vertebrates. Although much of the gene is highly conserved, there are high $K_a/K_s$ ratios in three of the nine exons (exons one, eight and nine), indeed the $K_a/K_s$ ratio of exon nine suggests positive selection, since it was greater than one. A comparison of a model that do not allow for variation of $K_a/K_s$ ratios among codons (M0) with one that does (M3) using a maximum likelihood ratio tests was highly significant, suggesting that the $K_a/K_s$ indeed varies among amino acid sites in the protein. Furthermore, parameter estimates under M3 suggest the presence of 12 sites under positive selection. Seven out of the twelve positively selected sites were situated in exon nine, four in exon one and one in exon eight. These results were confirmed with the simulations.

A HKA test was performed to test for heterogeneity between intraspecific variation in chicken and interspecific divergence with the turkey outgroup using all combined autosomal loci from Sundstrom, Webster and Ellegren.
(2004) and ZPC. The HKA test did not show significant deviation from neutrality (p = 0.60). The nucleotide diversity in ZPC introns is \( \pi = 8.9 \times 10^{-3} \), higher by a factor of 1.37 than the average for several other intronic loci sequenced in the same individuals (\( \pi = 6.5 \times 10^{-3} \)) (Sundstrom et al. 2004). The relative increase in polymorphism is nearly matched by the relative increase in divergence: the ZPC chicken-turkey intronic divergence of 0.129 is higher by a factor of 1.40 than the average of 0.092 from the study of Sundstrom, Webster and Ellegren (2004). A McDonald-Kreitman test was performed separately for all species pairs between chicken and the five other species. In all five cases, the neutrality index indicated positive selection, but the only significant MK test was with the chicken-pheasant pair (Fisher’s exact test 1-T, p = 0.01).

The analysis of the evolution of ZPC in fish and mammals gave similar results as in birds but much stronger indications of the actions of positive selection in the protein, with 45 and 16 sites respectively under positive selection.

Discussion

It was demonstrated that some sites in ZPC are evolving under positive Darwinian selection in the bird order galliformes, as well as in a number of fish and mammalian species. The principle finding of this study is thus that ZPC is subject to positive selection in diverse vertebrate groups. For both mammals and birds, positively selected sites are found at both ends of the gene, while the central portion is particularly well conserved. If similar patterns are found across three classes of vertebrates, this suggests that the form of selection is general and not specific to a particular environment or even particular details of reproductive biology. There is some evidence from the patterns of evolution in ZPC that sperm-egg interactions are involved: the sperm-binding region of ZPC has been identified in mammals to be near the carboxy-end of the ZPC polypeptide, and in mammals, birds and fish a high
proportion, although not all, of the positively selected sites are found in or near this region.

In order to maintain the sperm-egg interaction at optimum and mediate successful fertilisation, an advantageous mutation in the ZPC polypeptide must be countered by an adaptive change in the sperm. As a consequence, co-evolution between sperm and egg proteins may be associated with the evolution of reproductive isolation and speciation, in which selection favours barriers to hybridisation. Thus, ZPC evolution may be driven by a fundamental link between speciation and adaptive evolution.
Acknowledgements

I would like to thank all former and present people at the Department of Evolutionary Biology for creating a friendly and creative atmosphere that I have enjoyed so much. In particular, I would like to thank:

Hans for giving me the opportunity to do my PhD with you. I have always enjoyed working with you and it is very encouraging to know that our research is right at the forefront of this field. I am very grateful because you let me be independent, which for me is very creative. Also, you can handle my bursts of both anger and joy.

Nick for help with analyses and manuscripts.

Annika for being such a good friend. I am really appreciating our conversations, both in good and bad times. You have with a never-ending enthusiasm asked me to have coffee, although I am not the most social of people. Good luck with your new job and cat!

My office mates Eva and Anna-Karin. I think we have an optimal blend of quiet moments (read when we work) and moments when we chat about this and that. Eva, maybe some day we can have a horse together, although I am not too sure about a racehorse! Thanks to Anna-Karin for ‘sometimes’ bringing me coffee.

Hannah. You are the most helpful person I know and we’ve managed not to take our work too seriously. We were often cynical about both research and people. That was fun!

Carles. I don’t know how many times we have started to train again. For some reason we always start but we never continue. Without you, my time here would have been so much more boring.

Niclas B for your enthusiasm about research. Also thanks to Erik, Mikael and Matt, the other members of the ‘Mol Evol’ group and the former members, Helene and Lori for a stimulating scientific environment but most importantly a friendly atmosphere. Although research can be very competitive we have managed to work as a group and that is definitely a winning concept!
Frank, Cia, Susanne and Karin for being friendly and for your help on small but important things. Sometimes I have asked for help with for instance fixing coffee for the course and you have always been willing to help.

Gunilla K, Gunilla A, Anneli, Anki and Katrin who have ran my sequences throughout the years. Especially thanks to Gunilla K for your willingness to help in case of emergency when I needed my sequences quickly (as was the case this spring) and also for keeping the lab nice and organised.

Also thanks to Sidde and Caroline for long friendship and lots of fun. If I am not mistaken I think Sidde first of all realised that Niclas and I would fit together.

Fredrik and Maria also for long friendship and lots of fun.

Ullen for the longest of friendships, in fact from the day I was born. Although we became very different we will always keep in contact

Erik ‘The Straw’. For lots of fun at the golf course. Although perhaps not so much fun during my outbursts…

Lena K for being the best mother-in-law in the world and for your endless support, kindness, generosity and for being a good friend.

Ad for your generosity, fun parties and strong drinks.

Lars A for coping with me although I can be a bit obstinate.

My parents, Lena and Kalle. For the freedom you have given me, for always being there for me and for so much more.

My sister Johanna. For taking care of me in all situations and for always being there for me. Also thanks to Klas for keeping track of my sister.

Niclas, my great love. For ‘teaching’ me all about love and for believing in me and most importantly for believing in us.

Loppan for being the best cat in the world! (Yes, I am serious).
References


Lercher MJ, Williams EJ, Hurst LD (2001) Local similarity in evolutionary rates extends over whole chromosomes in human-rodent and mouse-
rat comparisons: implications for understanding the mechanistic ba-
Mass
129:513-23
Dev 12:650-656.
Makova KD, Li WH (2002) Strong male-driven evolution of DNA se-
Makova KD, Yang S, Chiaromonte F (2004) Insertions and deletions are
male biased too: a whole-genome analysis in rodents. Genome Res
14:567-73
Malcom CM, Wyckoff GJ, Lahn BT (2003) Genic mutation rates in mam-
mals: local Similarity, chromosomal heterogeneity, and X-versus-
up adaptive evolution. Curr Biol 13:R68-70
Marshall Graves JA, Shetty S (2001) Sex from W to Z: Evolution of verte-
brate sex chromosomes and sex determining genes. J Exp Zool
290:449-462.
ence between weakly selected mutations on patterns of molecular
McVean GT, Hurst LD (1997) Evidence for a selectively favourable reduc-
tion in the mutation rate of the X chromosome. Nature 386:388-392
Meunier J, Duret L (2004) Recombination drives the evolution of GC-
content in the human genome. Mol Biol Evol 21:984-90
Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T (1987) Male-
driven molecular evolution: a model and nucleotide sequence analy-
Muller HJ (1964) The Relation of Recombination to Mutational Advance.
Mutat Res 106:2-9
Nachman MW (2001) Single nucleotide polymorphisms and recombination
New York
O’Neill M, Binder M, Smith C, Andrews J, Reed K, Smith M, Millar C,
Lambert D, Sinclair A (2000) ASW: a gene with conserved avian W-
linkage and female specific expression in chick embryonic gonad.
Dev Genes Evol 210:243-249.
Smith NG, Hurst LD (1999) The causes of synonymous rate variation in the rodent genome. Can substitution rates be used to estimate the sex bias in mutation rate? Genetics 152:661-73
Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256-76
A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine*. (Prior to October, 1985, the series was published under the title “Abstracts of Uppsala Dissertations from the Faculty of Medicine”.)