JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE

98

Mervi Ahlroth

The Chicken Avidin Gene Family

Organization, Evolution and Frequent Recombination

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2001

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 98

Mervi Ahlroth

The Chicken Avidin Gene Family

Organization, Evolution and Frequent Recombination

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa (S212)
marraskuun 17. päivänä 2001 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Natural Sciences of the University of Jyväskylä,
in Auditorium S212, on November 17, 2001 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2001

The Chicken Avidin Gene Family

Organization, Evolution and Frequent Recombination

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 98

Mervi Ahlroth

The Chicken Avidin Gene Family

Organization, Evolution and Frequent Recombination

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2001

Editors
Jukka Särkkä
Department of Biological and Environmental Science, University of Jyväskylä
Pekka Olsbo, Marja-Leena Tynkkynen
Publishing Unit, University Library of Jyväskylä

URN:ISBN:978-951-39-8337-6
ISBN 978-951-39-8337-6 (PDF)
ISSN 1456-9701

ISBN 951-39-1083-0
ISSN 1456-9701

Jyväskylä University Printing House,
Jyväskylä and ER-Paino, Lievestuore 2001

To my beautiful family

ABSTRACT

Ahlroth, Mervi K.
The chicken avidin gene family. Organization, evolution and frequent recombination
Jyväskylä: University of Jyväskylä, 2001, 73p.
(Jyväskylä Studies in Biological and Environmental Science
ISSN 1456-9701; 98)
ISBN 951-39-1083-0
Yhteenveto: Kanan avidiinigeeniperhe. Organisaatio, evoluutio ja tiheä rekombinaatio
Diss.

Avidin is a biotin-binding protein found in chicken eggs. The gene encoding avidin (*AVD*) belongs to a family of multiple homologous genes, five of which (avidin-related genes *1-5* or *AVR*s *1-5*) were previously cloned and sequenced. In this study, the chromosomal organization of the *AVD* and *AVR* genes was characterized. The gene family was localized on Zq21. Different sets of *AVR* genes were found upon screening genomic libraries, and two novel *AVR* genes (*AVR6* and *AVR7*) were cloned and sequenced. The order and relative orientations of the *AVR* and *AVD* genes were mapped. A detailed evolutionary analysis of the gene sequences revealed that gene conversion plays an important role in the evolution of the gene family. Interestingly, the conversion showed a directional bias: *AVD* was able to convert the *AVR*s, but not vice versa.

Recombinant AVR proteins were produced and preliminary analysis of their functional and structural properties were performed. Differences were found in some biochemical and structural characteristics of the AVRs as compared to avidin. The AVR proteins were shown to bind biotin, with AVR1 and AVR2 exhibiting partially reversible binding.

The number of *AVD* and *AVR* genes was shown to differ between individuals, and also between cells within individuals. The molecular mechanism underlying the copy-number fluctuation seemed to be unequal crossing-over and/or unequal sister chromatid exchange. The frequent occurrence of both gene conversion and unequal sequence exchange supports the model according to which conversion and crossing-over are coupled processes. The high frequency of recombination observed suggests that copy-number lability may be more common among gene families than previously thought. Finally, the avidin gene family provides an excellent model for studying the molecular mechanisms of recombination events.

Key words: Fiber-FISH; fluorescence *in situ* hybridization; *Gallus gallus*; gene conversion; molecular evolution; multigene family; unequal crossing-over.

M. Ahlroth, University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35, FIN-40351 Jyväskylä, Finland

Author's address Mervi Ahlroth
Department of Biological and Environmental Science
University of Jyväskylä
P.O. Box 35
FIN-40351 Jyväskylä
Finland
e-mail: mervi.ahlroth@cc.jyu.fi

Supervisor Professor Markku S. Kulomaa
Department of Biological and Environmental Science
University of Jyväskylä
P.O. Box 35
FIN-40351 Jyväskylä
Finland

Reviewers Professor Howard Jacobs
Institute of Medical Technology
University of Tampere
FIN-33014 University of Tampere
Finland

Professor Olli Lassila
Department of Medical Microbiology
University of Turku
Kiinamyllynkatu 13
FIN-20520 Turku
Finland

Opponent Professor Alan Schulman
Institute of Biotechnology
University of Helsinki
P.O. Box 56, Viikinkaari 6
FIN-00014 University of Helsinki
Finland

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications, which will be referred to in the text by their Roman numerals. In addition, some unpublished results are described.

I Ahlroth, M. K., Kola, E. H., Ewald, D., Masabanda, J., Sazanov, A., Fries, R., Kulomaa, M. S. 2000: Characterization and chromosomal localization of the chicken avidin gene family. Anim. Genet. 31(6): 367-375.

II Ahlroth, M. K., Grapputo, A., Laitinen, O. H., Kulomaa, M. S. 2001: Sequence features and evolutionary mechanisms in the chicken avidin gene family. Biochem. Biophys. Res. Comm. 285: 734-741.

III Ahlroth, M. K., Ahlroth, P., Kulomaa, M. S. 2001: Copy-number fluctuation by unequal crossing-over in the chicken avidin gene family. Biochem. Biophys. Res. Comm. 288: 400-406.

IV Laitinen, O. H., Hytönen, V., Ahlroth, M. K., Pentikäinen, O. T., Gallagher, C., Nordlund, H., Ovod, V., Marttila, A., Porkka, E., Heino, S., Johnson, M. S., Airenne, K., and Kulomaa, M. S. 2001: Chicken avidin-related proteins (AVRs) show altered biotin-binding and physico-chemical properties as compared to avidin. Submitted.

RESPONSIBILITIES OF MERVI AHLROTH IN THE ARTICLES OF THIS THESIS

Article I: I screened the Clontech library, and the gridded library was screened together with Dr. Dagmar Ewald. The cosmid mapping and gene cloning experiments were planned and carried out together with Eija Kola, M.Sc., who did her master's thesis on these studies. The metaphase-FISH was carried out by Drs. Julio Masabanda and Alexei Sazanov. I wrote the article.

Article II: The study was planned together with Dr. Alessandro Grapputo. Alessandro did the phylogenetic, GENECONV and nucleotide substitution analyses, while I did the SimPlot analysis and most of the writing. Olli Laitinen, M.Sc., provided information about the putative AVR proteins.

Article III: I planned and carried out the whole study, except for the statistical analyses that were done together with Dr. Petri Ahlroth. I also wrote the article.

Article IV: I produced the *AVR6* and *AVR7* cDNAs from the corresponding genes and produced the protein expression constructs for them. I also participated in writing of the article.

All these work were carried out under the supervision of Prof. Markku Kulomaa.

ABBREVIATIONS

ApoVLDLII	apo very-low-density-lipoprotein II
AVD	avidin
AVD	avidin gene
AVR	avidin-related protein
AVR	avidin-related gene
bp	base pair
CD	cluster of differentiation
cDNA	complementary DNA
CR1	chicken repeat 1
DSB	double-strand break
DSBR	double-strand break repair
dsDNA	double-stranded DNA
EST	expressed sequence tag
FISH	fluorescence *in situ* hybridization
Ig	immunoglobulin
ILT	Ig-like transcript
kb	kilobase pair
KIR	killer cell inhibitory receptor
LRC	leukocyte receptor complex
MHC	major histocompatibility complex
mRNA	messenger RNA
NHEJ	non-homologous end-joining
PCR	polymerase chain reaction
LTR	long terminal repeat
rDNA	ribosomal DNA
rRNA	ribosomal RNA
SCE	sister-chromatid exchange
STS	sequence-tagged site
TCR	T-cell antigen receptor
tRNA	transfer RNA

1 INTRODUCTION

The genomes of eukaryotic organisms include numerous multigene families. The human genome, for example, contains more than 3 000 protein-encoding gene families, the number of genes per family ranging from two to several hundred (Li et al. 2001). The largest gene families are those encoding structural RNAs, such as rRNAs (~400 genes) and tRNAs (~1300 genes) (Li 1997). A gene family is defined as a group of genes that are of the same evolutionary origin, i.e. are derived from duplications of a common ancestor gene (Lewin 2000). Thus, members of a gene family are similar in nucleotide sequence and, very often, in function. An example of a highly conserved gene family is the histone family. There are five histone genes in most species: H1, H2A, H2B, H3 and H4. They all share the same function of wrapping chromosomal DNA around them to form packed chromatin, and their nucleotide sequences are probably the most conserved among eukaryotic taxa (Li 1997).

A gene superfamily is defined as a group of genes that exhibit functional similarity but have diverged in nucleotide sequence over long periods of time so that their relatedness is no longer as evident as within gene families (Li 1997, Strachan & Read 1999). A classical textbook example is the immunoglobulin superfamily. The superfamily comprises at least the immunoglobulins (Ig), T-cell antigen receptors (TCR), major histocompatibility complex (MHC) proteins, natural killer cell inhibitory receptors (KIR), and various CD proteins, like CD4 and CD8. All these molecules serve immunological functions, and share a common "Ig-fold" in their three-dimensional protein structure (Strachan & Read 1999).

Multigene families offer means to study the evolution of living organisms. They facilitate the timing of species diversification by allowing comparisons of between-species and within-species diversification (Lewin 2000). Gene families have also been invaluable in revealing the mechanisms of evolution of new functions, such as mutations and exon shuffling. Furthermore, the identification of related motifs in gene and protein structures facilitates deciphering the function of novel genes.

In addition to resolving the past, gene families are invaluable in studying current changes in genomes. Evolution is an ongoing process, and the genome is by no means a constant entity. We are currently living exciting times in terms of human biology: the sequence of the human genome has been revealed (see the human genome supplements of *Nature* vol. 409 [no. 6822], 2001 and *Science* vol. 291 [no. 5507], 2001). The next task is to learn to understand the meaning of the genome information. One aspect is to understand the degree of lability of the genome. Evidence is accumulating from studies on transposable elements and repetitive regions that the genome is in constant flux. Because of the slow reproductive cycle of man, as well as the practical and ethical issues involved in handling material of human origin, we need to use comparative genomics using model organisms to understand the rate and extent of changes in eukaryotic genomes.

Our studies have provided clues about genome instability in one model organsim, the chicken. We have studied the chicken avidin gene family, which consists of the avidin gene (*AVD*) and a number of avidin-related genes (*AVRs*). Our studies revealed the chromosomal organization of the gene family (I), and showed that the gene family is highly prone to recombination events such as gene conversion (II) and unequal crossing-over (III). The structure and organization of the gene family seems to render the gene family susceptible to recombination. Unequal crossing-over was shown to result in extremely frequent somatic variation in gene number within individuals. The gene number differed also between individuals, indicating that germ-line recombination occurs as well. There are very few previous examples of copy-number variation within, or even between, individuals for relatively small, protein-encoding gene families, such as the avidin gene family.

Our results also lend support for some models postulated to govern the general processes of genetic recombination and gene diversification. First, the high frequency of both gene conversion and unequal crossing-over within the avidin gene family (II, III) strongly favors the view that the two processes are initiated by the same mechanism, and are subsequently resolved into either outcome. An appealing model is the DNA double-strand break repair (DSBR) model, according to which migrating Holliday junctions are formed during the repair process and are subsequently resolved to produce either recombinant (cross-over) or non-recombinant (gene conversion) products (Fig. 4). Second, the high frequency of recombination in the avidin gene family in B-cells may be a side product of immunoglobulin gene hyperconversion. This favors the transcription-linked model of somatic hypermutation, according to which hypermutation is not strictly restricted to Ig genes, but that all transcribed genes in B-cells are susceptible to hypermutation.

Furthermore, we studied the characteristics of the avidin-related proteins (AVRs; IV). We found drastic changes in the amino acid sequences of the proteins that still allowed formation of functional oligomeric assemblies similar to avidin. On the other hand, a single amino acid substitution in the biotin-binding pocket had a major effect on binding affinity. The avidin protein family therefore provides an excellent model system for studying the structure-function relationships of proteins, a relationship that is currently poorly understood. The AVR proteins may also be valuable as biochemical tools, similarly to avidin.

2 REVIEW OF THE LITERATURE

2.1 Multigene families and their evolution

Multigene families arise by consecutive duplications initiated by an ancestral gene. The fate of the gene copies depends on the prevailing selection pressure and genetic drift. According to Fryxell (1996), gene duplication events *per se* occur rather frequently, but only a minor fraction of the duplicates are retained and may develop new functions of selective advantage. Others have suggested that a proportion as high as 50% of the duplicates may lead to functional divergence (Nadeau & Sankoff 1997, reviewed by Wagner 1998). In any case, at least half of the duplicates become lost or functionless due to accumulation of mutations and genetic drift.

Whatever the proportion of surviving gene copies, differentiation following gene multiplication is thought to be the most important mechanism for the evolution of new functions (Li 1997). As an example, the hemoglobins and myoglobin have diverged to specialize in oxygen transport in blood and oxygen storage in muscles, respectively (Fig. 1). Comparably, the immunoglobulin family members have retained a common function with slightly differing gene products, contributing to the antibody variety needed for the survival of the individual. In addition to the C, V, D and J subfamilies constituting the antibodies, the immunoglobulin superfamily includes the MHC and T-cell receptor subfamilies which, although further differentiated, still exhibit related functions of antigen recognition (Nei 1987, Li 1997, Roitt 1997, Lewin 2000). Conservation of the original function of the genes may also be selected for when production of larger amounts of the gene product is advantageous, as is the case in many species with ribosomal RNA genes and histone genes (Li 1997, Lewin 2000).

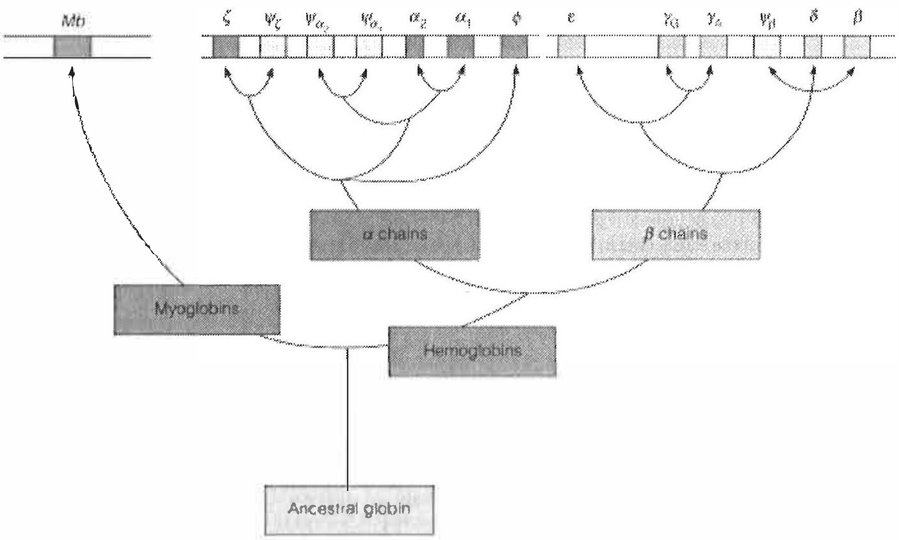

FIGURE 1 Evolution of the human globin gene family. Myoglobin and the two different forms of hemoglobin (α and β) are derived from a common ancestor. The α-genes are located on chromosome 16, β-genes on chromosome 11, and myoglobin on chromosome 22. The genes marked with ψ are pseudogenes, i.e. have lost their functionality due to accumulation of mutations. (from Brooker 1999).

2.1.1 The arrangement and copy number of genes in multigene families

The gene copies often remain close to their ancestor at the same chromosomal position, forming a linked gene family in which genes may be arranged tandemly or in random orientations (clustered organization; Graham 1995, Strachan & Read 1999). Alternatively, the copies may disperse throughout the genome via transpositional events, crossing-over or translocations between chromosomes. The chicken and human histone gene families exhibit an intermediate type of organization: most of the genes are located in two linked clusters on a single chromosome (chromosome 1 in the chicken and chromosome 6 in the human), and additional genes are dispersed on other chromosomes (Nakayama et al. 1993, Albig & Doenecke 1997, Strachan & Read 1999). In general, it appears that clustered arrangement is favored for small (with <50 members) and tandem arrangement for large (>50) gene families (Graham 1995).

Rearrangement mechanisms such as unequal crossing-over (see below) are much less likely to occur between dispersed genes than between linked genes (Dover 1982, Graham 1995). Furthermore, it has been suggested that the organization of genes in a cluster can inter-convert between tandem and random arrangements (Graham 1995). Homogenization by unequal crossing-over and gene conversion (see below) is thought to act more powerfully on tandem genes than on clustered genes. Thus, clustered genes may be more susceptible to diversifying forces. Regardless of organization and copy-number, chromosomal position also affects the evolutionary rate: the distal (telomeric) regions of chromosomes are more susceptible to recombination and translocation (Perry & Ashworth 1999).

In addition to the structural and evolutionary aspects, the organization of multiple gene copies may also have important functional implications. For example, the tandem arrangement of genes in the human β-globin gene family reflects the temporally regulated expression of different genes in the array during development (Hanscombe et al. 1991). However, the expressional organization is not conserved cross-species: in the chicken, for example, the genes are not organized according to the timing of expression (Dolan et al. 1981, Reitman et al. 1993).

The gene copy number can vary widely in different gene families within a species as well as between different species (Table 1). In the chicken, for example, the ovalbumin gene has duplicated only twice, forming a gene family consisting of three members (Royal et al. 1979), whereas the rRNA genes are present in 80-700 tandemly arranged copies (Stevens 1996, Su & Delany 1998). The rRNA genes often exhibit wide copy-number variation (Krystal et al. 1981, Long & Dawid 1980, Warburton et al. 1976), and variation in rDNA repeat size has also been observed in the chicken and human (Wellauer 1979, Delany & Krupkin 1999, Delany 2000, Kuo et al. 1996). Large gene families are generally unstable, being amenable to expansion, contraction and dispersion (Jeffreys & Harris 1982). High-copy number gene families can tolerate rearrangements and fluctuations better than low-copy number families: the probability of total gene loss by unequal crossing-over or deletion is greater in low copy-number families.

Differences in gene copy-numbers between different individuals have been observed also for some smaller, protein-encoding gene families. Recent evidence suggests that the chicken MHC genes exhibit lability in gene number (Afanassieff et al. 2001). Also the number of the human haptoglobin-related genes (*Hpr*) has been found to vary between individuals, as well as between homologous chromosomes within individuals (Maeda et al. 1986). Interestingly, the *Hpr* gene copy-number variation correlated with unusual hemoglobin gene arrangements, even though the *Hpr* and hemoglobin genes are located on different chromosomes (Maeda et al. 1986). The hemoglobin genes themselves can exhibit copy-number expansion and contraction (Harano et al. 1985). The human leukocyte receptor complex (LRC) on chromosome 19q13.4 exhibits variation in copy numbers of the killer cell inhibitory receptor (KIR) and Ig-like transcript (ILT) genes (Wilson et al. 2000). However, all these cases exemplify germ-line recombination leading to copy-number variation between individuals. In contrast, little evidence exists for somatic copy-number fluctuation within individuals. Asexual blood stages of the malaria parasite *Plasmodium falciparum* have been found to undergo somatic unequal crossing-over in their virulence factor genes (Freitas-Junior et al. 2000). The KIR genes

TABLE 1 Copy numbers and arrangements of some human and chicken multigene families. ARKdb refers to the farm animal database hosted by the Roslin Institute (http://www.thearkdb.org/).

gene family	organism	copy number	distribution in the genome	references
histones	human	61	two main clusters on chr. 6, dispersed copies on chr. 1, 4, 7, 11, 17 and 22	Albig & Doenecke 1997, Strachan & Read 1999
histones	chicken	43	40 genes in clusters *a* and *b* on chr. 1, 3 genes elsewhere, H5 on chr. 1	Grandy & Dogson 1987, Nakayama et al. 1993, Stevens 1996, ARKdb
28S, 18S and 5.8S rRNA	human	350	nucleolar organizers on chr. 13, 14, 15, 21 and 22	Evans et al. 1974, Li 1997
rRNA	chicken	80-≥700	a single nucleolar organizer on microchromosome 17	Stevens 1996, Su & Delany 1998
tRNA (40 cytoplasmic subfamilies)	human	1300	dispersed	Li 1997
globins	human	14	a-chains (7 copies) chr. 16, b-chains (6 copies) chr. 11, myoglobin (1 copy) chr. 22	Li 1997
globins	chicken	8	3 α-hemoglobins on microchromo-some 14, 4 β-hemoglobins on macrochromosome 1, myoglobin?	Stevens 1996 and ARKdb
olfactory receptor genes	human	~1000	25 clusters (6-155 genes / cluster) dispersed over all chromosomes except for chr. 20 and Y	Dryer 2000, Glusman et al. 2001, Zozulya et al. 2001

might represent another example, suggested by the non-uniform fiber-FISH mapping results reported by Suto et al. (1998) as well as by the diversity of KIR molecules found within individuals (Kwon et al. 2000, Uhrberg et al. 2001). In the germ line, the KIR genes have been found to undergo recombination to produce polymorphism (Shilling et al. 1998)

2.2 Nucleotide substitutions

The substitution of a nucleotide by another is denoted as transition, when the substitution involves changing a purine into another purine, e.g. A↔G, or a pyrimidine into another pyrimidine, e.g. C↔T (Lewin 2000). Transversion, on the other hand, involves changing a purine into a pyrimidine or vice versa. A nucleotide substitution occurring in a protein-coding region may cause an amino acid change. In this case, the substitution is called nonsynonymous

(Strachan & Read 1999, Table 2). If the change leads to substitution of the original amino acid with another amino acid, the mutation is said to be a missense mutation. If the nucleotide substitution changes the codon into a translation stop codon, the mutation is defined as a nonsense mutation. In missense mutations, if the substituted amino acid is similar in biochemical properties to the original amino acid, it is called a homologous, or conservative, amino acid substitution. Comparably, a radical change, for example from glycine to proline, is a nonhomologous (nonconservative) amino acid substitution. A substitution occurring in the third codon position is often synonymous (or silent), i.e. it does not lead to an amino acid change (Table 2). Synonymous substitutions are frequently encountered, since they do not confer selective disadvantage and are therefore easily tolerated (Li 1997).

Mutations accumulate in duplicated genes in the absence of selection. The mutations accumulate randomly across the genes and their flanking regions, eventually destroying the function of the genes and rendering them pseudogenes. In the presence of selection towards restoring the gene activity, on the other hand, mutations preferentially accumulate in non-coding regions such as introns and flanking non-regulatory regions, as well as silent third code positions in exons (Jeffreys & Harris 1982). Thus, sequence comparisons between genes of common ancestry usually show greater homology over exons and regulatory elements than non-coding regions. However, there are some examples of related genes that show higher degree of conservation over introns than over exons. An example is discussed below (section 2.6).

There are also differences in the substitution rates in different gene families. The histone genes constitute probably the most conserved gene family across eukaryotic species. Between rodents and human, for example, there are no nonsynonymous substitutions in histone 3 and histone 4 genes (Li 1997). The γ-interferon gene, in contrast, shows a nonsynonymous substitution rate of

TABLE 2 The nomenclature of nucleotide substitutions.

mutation	**classification at nucleotide level**	**classification at amino acid level**	**type of amino acid change**
Lys AAA ↓ Lys AAG	synonymous (silent) transition	-	no change
Lys AAA ↓ Glu GAA	nonsynonymous missense transition	nonhomologous	positively charged → negatively charged
Lys AAA ↓ Arg AGA	nonsynonymous missense transition	homologous	positively charged, similar side chain structures
Lys AAA ↓ stop TAA	nonsynonymous nonsense transversion	-	no corresponding amino acid

3×10^{-9} / site / year, whereas the average rate is about 0.74×10^{-9} / site / year (Li 1997). The differences obviously reflect the functional constraints set by the specific function of each protein.

2.3 Inversions

Inversion may operate at the level of large chromosomal segments or short DNA regions of only few base pairs long (Li 1997, Brooker 1999). Inversions of large chromosome segments, found in about 2% of humans, result from the formation of "inversion loops" in meiotic cells (Griffiths et al. 1993, Brooker 1999). Comparably, the mechanism underlying small, few-base pair inversions is thought to be a local hairpin-loop secondary structure formation on the DNA, possibly associated with DNA replication (Leach 1994). Inversions are often mediated by direct or inverted repeats in DNA. The looping back of DNA and pairing of the repeat sequences followed by DNA breakage and rejoining results in inversion of the segment between the participating repeats (Purandre & Patel 1997, Fig. 2). Therefore, genes flanked by repeat sequences may be susceptible to recurrent changes in orientation. Inversion of one or more genes within a tandem array has been found to suppress unequal recombination (Graham 1995 and references therein). This may be the first step in the transformation of a tandem array into a stable, randomly oriented gene cluster.

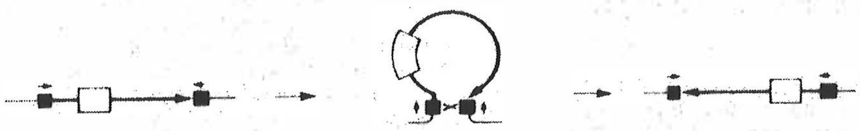

FIGURE 2 A mechanism for producing inversions in DNA. (From Jeffreys & Harris 1982.)

2.4 Recombination

In meiotic cell division, the duplicated homologous chromosomes (one paternal and one maternal chromosome) pair to segregate into different daughter cells. Crossing-over, the exchange of genetic material, occurs during this pairing (Fig. 3a). In the human male meiosis, for example, an average of 49 crossovers occur per cell (Strachan & Read 1999 and references therein). The homologous chromosomes normally do not pair in mitosis. On rare occasions, however, pairing does occur and may result in mitotic crossing-over (Brooker 1999).

Subsequent cell divisions produce batches of tissue containing different allele combinations.

Current models suggest that crossing-over results from double-strand DNA break repair (DSBR), and that the double-stranded breaks are formed during DNA replication (Lopez et al. 1992, Haber 1998 and 1999, Borde et al. 2000, Marians 2000) In DSBR, migrating Holliday junctions are formed and, depending on the way in which the junctions are resolved, either recombination or gene conversion (see below) results (reviewed in Martinsohn et al. 1999, Haber 1999 and Brooker 1999; Fig. 4). The formation of Holliday junctions requires extensive sequence homology between the participating regions, and these kind of events are therefore termed homologous recombination.

Recombination can also occur via non-homologous recombination pathways. One of these pathways is non-homologous end-joining (NHEJ), that is thought to represent an alternative way to repair double-strand breaks (Critchlow & Jackson 1998 and references therein). In this process, the broken DNA is rescued by ligating the DNA ends back together or to another DSB site. This process does not require any sequence homology, and thus often leads to chromosomal translocations. Site-specific recombination, on the other hand, is a form of non-homologous recombination that requires a short stretch of sequence homology (a dozen base pairs or so, usually constituting defined recombination signal sequences, RSS) to bring the recombining gene segments together (Grawunder & Harfst 2001). Integration of viral genomes into host cell genomes and the immunoglobulin gene rearrangements are examples of site-specific recombination (Brooker 1999).

2.4.1 Unequal crossing-over

In genomic regions where repetitive sequences occur, the chromatids may pair asymmetrically. Crossing-over in an asymmetrically paired region results in changes in gene copy numbers, reducing the copy number in one chromatid and increasing it in the other (Lewin 2000; Fig. 3c). In cases of large tandem repeats dispersed to different chromosomes, such as the rRNA genes, unequal crossing-over may happen also between nonhomologous chromosomes (Krystal 1981, Li 1997).

Unequal crossing-over may result in fairly large fluctuations in gene numbers, as observed for rRNA and tRNA genes in many species (Nei 1987). In fact, tandem genes seem to undergo unequal crossing-over more or less continuously. Unequal crossing-over homogenizes entire large repeats rather than just single genes within the repeat, and is therefore the most powerful mechanism counteracting mutations among large tandem gene families (Graham 1995). On some occasions, on the other hand, the occurrence of crossing-over breakpoints within gene sequences can produce effects of clinical relevance. For example, the human Lepore and Kenya variants of hemoglobin are produced by unequal crossing-over and cause thalassemic phenotypes (Griffiths et al. 1993).

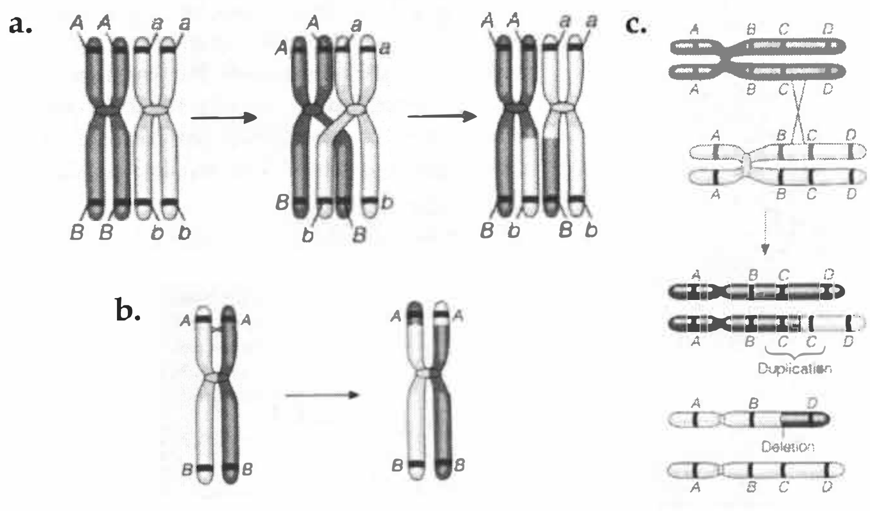

FIGURE 3 The different forms of crossing-over. **a.** Recombination between homologous chromosomes. **b.** Sister-chromatid exchange within a replicated chromosome. **c.** Reciprocal duplication and deletion resulting from unequal crossing-over. (Modified from Brooker 1999.)

2.4.2 Unequal sister-chromatid exchange

Unequal sister-chromatid exchange (SCE) is a phenomenon similar to unequal crossing-over, but in this case the crossing-over event occurs between sister chromatids of a single duplicated chromosome (Fig. 3b). Unequal SCE may occur in mitotic as well as meiotic cell division (Brooker 1999, Martinsohn et al. 1999). The overall frequency of unequal sister-chromatid exchange (~1.5%) is much lower than that of crossing-over between homologous chromosomes, but about the same frequency as that of gene conversion (Li 1997). However, in experiments involving gene duplications in yeast, intrachromosomal recombination has been observed to occur at a rate 10-100 times higher than recombination between homologous chromosomes (Martinsohn et al. 1999 and references therein). Obviously, the rates of different recombination mechanisms vary between different sequence repeats and genomic contexts, as well as between different cell types and organisms (Martinsohn et al. 1999 and references therein).

2.4.3 Gene conversion

Gene conversion is another form of homologous recombination. Conversion can occur between dispersed gene copies (ectopic gene conversion; Li 1997 and references therein, Liao 2000), also in central chromosomal regions, and is often thought to act as a correction mechanism suppressing the occurrence of deleterious mutations in gene alleles. In this process, a stretch of DNA in the "acceptor" gene is exchanged for homologous sequence in the "donor" gene. Mechanistically, gene conversion is thought to result from the alternative resolution of two Holliday junctions in DSBR, although other models have also been suggested (Szostak et al. 1983, Elliott et al. 1998, Brooker 1999; Fig. 4). This view is supported by the fact that as many as half of the conversion events have been observed to be associated with recombination (Martinsohn et al. 1999 and references therein). Conversion (as well as recombination) preferentially occurs in transcribed genes (Graham 1995), suggesting the requirement for open (euchromatic) chromatin conformation.

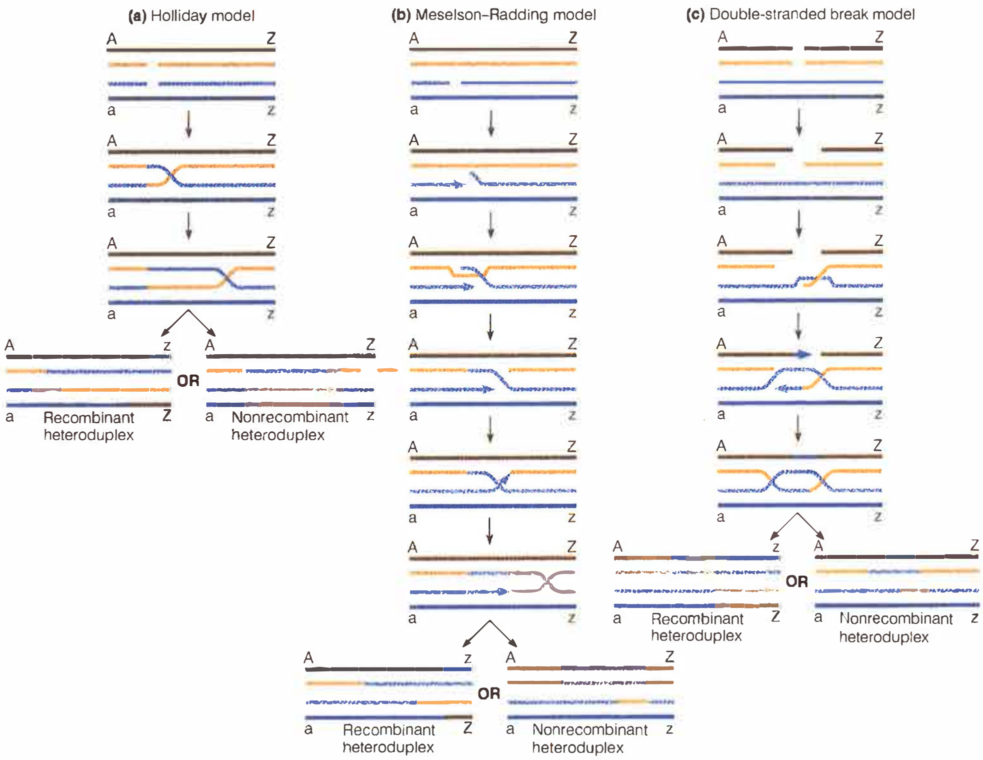

FIGURE 4 Models for the mechanism of recombination and gene conversion. (From Brooker 1999.)

Gene conversion leads to gradual homogenization of the participating sequences without changes in gene copy number (Li 1997). It is thought to be the main mechanism of homogenization among clustered gene families where unequal crossing-over is suppressed (Graham 1995). The converted stretches are generally short, with the smallest ones being just a few base-pairs (Wheeler et al. 1990). Thus, conversion homogenizes only portions of genes, leaving the remainder free to diverge. In some genes, certain regions seem to be particularly resistant to homogenization. In the human red and green color pigment genes, for example, the intron regions are more homogeneous than the exon regions although the involvement of gene conversion in the evolution of the genes is evident (Li 1997 and references therein). In this case, there has probably been a selective pressure towards restoring differences in the exon sequences, since homogenization of the coding sequences would reduce the spectrum of color vision. Furthermore, in the major histocompatibility (MHC) genes, gene conversion actually acts to preserve the diversity between the different clusters of MHC genes by shuffling sequence variants between the genes (Graham 1995, Martinsohn et al. 1999 and references therein). The chicken immunoglobulin diversity is also produced in a similar way, by copying sequences from pseudogenes into the Ig variable regions by gene conversion (see below).

Gene conversion plays an important role in the evolution of the human hemoglobin genes (reviewed by Papadakis & Patrinos 1999). Interestingly, the conversion events among the hemoglobins clearly exhibit polarity, i.e. the gene physically located on the 5′-side converts the gene on the 3′-side. The conversions also often follow the "master-slave" rule, namely that a gene expressed at high level ("master") converts the one expressed at lower level ("slave"; Papadakis & Patrinos 1999). The borders of the conversion tracts are often marked by special DNA sequence structures, such as *Chi* recombination sequences or purine/pyrimidine or polypurine tracts. Secondary DNA structures formed by these sequences, as well as blocks of sequence divergence, efficiently inhibit branch migration and thus gene conversion (Wheeler et al. 1990, Papadakis & Patrinos 1999).

2.5 Somatic hypermutation

The immunoglobulin (Ig) variable regions, together with the MHC molecules (section 2.8), probably exhibit the largest variability among eukaryotic proteins. The V(D)J recombination of Ig genes, as well as class switch recombination in Ig heavy chain genes, are well understood and serve as classical textbook examples of site-specific recombination (Brooker 1999, Strachan & Read 1999, Grawunder & Harfst 2001). Somatic hypermutation of the Ig variable regions, on the contrary, is not yet well understood. Somatic hypermutation is the main mechanism creating diversity in the secondary repertoire of antibodies, after the first round of affinity selection following antigen challenge (reviewed by Neuberger & Milstein 1995, Storb 1996, and Jacobs & Bross 2001). The hypermutation occurs during the short period of maturation of B-cells in the germinal centers of peripheral lymphoid organs (Storb 1996, Jacobs & Bross 2001). In man and mouse, somatic hypermutation involves mainly single nucleotide substitutions and, more seldom, single nucleotide insertions and deletions (Storb 1996). Somatic hypermutation is characterized by transition bias (nonsynonymous substitutions predominate), strand bias (the transcribed strand is preferentially mutated), and the occurrence of mutational hotspots, especially serine AGY codons (Neuberger & Milstein 1995, Storb 1996, Ching et al. 2000). Several possible mechanisms for the hypermutation have been suggested, and the strongest evidence favors dependence on transcription (Storb 1996, Tian & Alt 2000, Bachl et al. 2001, Jacobs & Bross 2001) and/or DNA break repair (Goodman & Tippin 2000, Kong & Maizels 2001).

Unlike in humans and most other animals, the immunoglobulin variable region diversification in chicken is achieved in a partially antigen-dependent manner, creating a large pool of preimmune antibodies instead of highly affinity-selected antibodies (Reynaud et al. 1987 and references therein, Stevens 1996). The diversification occurs in the bursa of Fabricius, which is the primary site of B-cell maturation in the chicken. The mechanism of diversification is also different: intrachromosomal gene conversion ("hyperconversion") instead of hypermutation (Reynaud et al. 1987, Carlson et al. 1990). The proximity of the chicken V genes to each other (2.4 kb as opposed to ≥10 kb in humans) probably potentiates gene conversion in *cis* (Carlson et al. 1990, Storb 1996). In the rabbit, both gene conversion and somatic hypermutation occur (Schiaffella et al. 1999).

It is currently not clear if the hypermutation is restricted to the immunoglobulin genes only (Kelsoe 1999). Storb (1996) suggested that all heavily transcribed genes in the maturating B-cells may be susceptible to hypermutation. This assumption follows the model according to which a mutator factor is recruited to transcribed genes by the transcriptional initiation complex (Storb 1996). Support for this hypothesis has recently emerged: the gene for the transcription factor BCL-6 has been found to exhibit elevated mutation frequency in hypermutating B cells (Shen et al. 1998, Pasqualucci et al. 1998). However, no other examples have as yet been found (Shen et al. 2000, Jacobs & Bross 2001). The more recent models suggest that the mutations are caused by repair of (transcription-induced) DNA double-strand breaks by (an) error-prone DNA polymerase(s) (Ninio 2000, Dominguez et al. 2000, Tissier et al. 2000, Jacobs & Bross 2001).

Somatic hypermutation may also occur in other cell types. Zheng et al. (1994) reported hypermutation of T-cell antigen receptor (TCR) α genes in germinal center T-cells. Another study (McHeyzer-Williams & Davis 1995) did not support their findings, but the controversy may be due to the loss of mutated T-cells by apoptotic death in the latter study (Storb 1996). Moreover, extremely high frequencies of somatic diversification were observed for a novel antigen receptor-like molecule in the spleen of nurse shark (Greenberg et al. 1995).

2.6 The chicken MHC genes

The major histocompatibility complex (MHC) proteins are both structurally and functionally related to immunoglobulins, and are therefore classified as members of the Ig superfamily (Roitt 1997). In the chicken, the MHC genes are located in two physically linked, but genetically unlinked, clusters (Briles et al. 1993, Miller et al. 1996, Afanassieff et al. 2001). The clusters are denoted *B* (classical) and *Rfp-Y* (for restriction fragment pattern-Y), and are located on the microchromosome 16 (Miller et al. 1996). The genetic linkage disequilibrium of the two loci results from the highly recombinogenic nucleolar organizer region (NOR) located between them (Miller et al. 1996). The NOR consists of ~145 rRNA genes, and is the only NOR in the chicken (Miller et al. 1996).

In a recent study, Afanassieff et al. (2001) examined in detail the *Rfp-Y* locus. Interestingly, the number of class I genes seemed to vary among *Rfp-Y* haplotypes. At least one of the class I genes of the locus, *YFV*, was shown to be functional. It exhibited considerable sequence variability, including frequent nonconservative amino acid substitutions in the antigen binding region, suggesting a related but specialized function as compared to the classical *B* locus genes. The gene was found to be expressed in a wide range of tissues, excluding the brain, heart and spleen. The structural and functional properties of the chicken MHC genes therefore distinctively resemble those of the avidin gene family (see below).

2.7 Avidin

2.7.1 The function and expression of avidin

Avidin is a protein that has long been used as a tool in biochemical sciences because of its strong binding to biotin (Wilchek & Bayer 1988, 1990 and 1999). The variety of avidin-biotin applications is vast and nowadays includes medical *in vivo* and diagnostic applications (Walker et al. 1996, Schetters 1999, Wilbur et al. 1999, Guttinger et al. 2000) Avidin is composed of four subunits, each being able to bind one molecule of biotin (Green 1975, Stevens 1991). The subunits are thought to be identical polypeptides encoded by a single gene. In the chicken, avidin is produced in the oviduct of egg-laying hen (Hertz & Sebrell 1942) under the influence of progesterone hormone (Hertz et al. 1949, O'Malley 1967, Tuohimaa et al. 1989 and references therein). The protein is secreted into developing egg, where it constitutes about 0.05% of egg white proteins (Green 1975, Stevens 1991). Avidin is thought to act as a defense protein protecting the developing chick embryo against bacterial infection (Tuohimaa et al. 1989 and references therein, Stevens 1991). It is thought to kill invading bacteria by depriving them of biotin. Furthermore, avidin has been shown to attach to bacterial cells (Korpela et al. 1984). The attachment may physically interfere with bacterial infection (Korpela et al. 1984).

Avidin is also produced in other tissues of both male and female chicken after bacterial or viral infection, tissue trauma or treatment with toxic chemicals (Elo et al. 1979a,b; Elo & Korpela 1984, Korpela et al. 1982, Kunnas et al. 1993), as well as in chicken cell cultures (Korpela et al. 1983, Korpela 1984, Ahlroth 2001 and unpublished). In a recent publication, Zerega and coworkers (2001) reported a novel role for avidin. They found that avidin is expressed in the skeletal muscle and growth plate hypertrophic cartilage of the developing chicken embryo. This finding together with their other results suggest that avidin assists the terminal differentiation of chondrocytes and myoblasts. Furthermore, Huberman and coworkers (2001) showed that avidin exhibits pseudocatalytic activity, being able to enhance the hydrolysis of bound biotinyl *p*-nitrophenyl.

The kinetics of avidin expression under hormonal induction has been studied in both tissue culture and hormone-treated chicks. O'Malley (1967) observed avidin synthesis after 6 hours of progesterone induction in tissue culture using minced pieces of chick oviduct. In estrogen-primed animals *in vivo*, a single injection of progesterone resulted in the occurrence of avidin after 10 h (Korenman & O'Malley 1968). Korenman and O'Malley also showed that avidin induction is sensitive to the protein synthesis inhibitor cycloheximide, thus involving *de novo* avidin synthesis. In 1969, O'Malley and McGuire showed the first evidence that the induction actually occurs at the transcriptional level: they observed accumulation of novel nuclear RNAs after progesterone induction in the chick oviduct. Chan et al. (1973) confirmed that the appearance of avidin-specific mRNA precedes avidin protein accumulation. They observed avidin mRNA first after 6 h of progesterone administration, with maximal concentrations after 18-24 h. Kunnas et al. (1992) observed avidin mRNA at 8 h and a maximum at 16-24 h from progesterone stimulation. Taken together, avidin transcription is induced in a delayed mode. Thus, the avidin gene probably belongs to the delayed primary response or secondary response class of steroid-inducible genes (Dean & Sanders 1996, Kunnas et al. 1992). As putatively functional progesterone response elements have thus far not been found flanking the *AVD* gene (unpublished), the latter model seems more likely. If correct, the induction would involve the activation by the progesterone receptor of a primary response gene, probably encoding another transcription factor, and the newly synthesized factor would in turn activate transcription of the *AVD* gene. The primary response gene may act in a cell- or tissue-specific fashion.

The induction of avidin production under inflammation is independent of progesterone, and can be abolished by treatment with anti-inflammatory drugs (Nordback et al. 1982, Niemelä 1985). The induction apparently occurs via NF-κB (Eija Kola, personal communication). Inflammation-induced avidin has thus far been detected in all tested chicken tissues except the brain, and is thought to act as a local acute defense protein in injured and infected tissues (Elo et al. 1980, Elo & Korpela 1984, Klasing 1998). The progesterone and inflammation-induced avidins appear similar (Kulomaa et al. 1981), and are thought to be encoded by a single gene operating under differential regulation schemes (Kunnas et al. 1993). However, in some cases the antigenic structures of the two forms have been reported to differ slightly (Korpela et al. 1982).

2.7.2 The avidin gene family

The cDNA encoding avidin in the chicken oviduct, cloned by Gope et al. (1987), was used as a probe in initial attempts to clone the complete avidin gene from genomic libraries. Instead of avidin, five genes homologous to it were found and named avidin-related genes *1-5* (*AVR*s *1-5*; Keinänen et al. 1988 and Keinänen et al. 1994). Thus, the existence of a family of avidin genes became evident. The *AVR*s were found to be highly conserved, being 94-100% identical to each other (Keinänen et al. 1994). The genes were 1113 bp long, consisting of four exons and three introns with well-defined junctions and associated splicing signals. They all contained putative promoter sequences at their 5′-ends and polyadenylation signals at their 3′-ends suggesting that the genes may be functional. Indeed, RT-PCR experiments have detected low amounts of mRNAs encoded by *AVR2* in the oviduct and *AVR3* in the intestine under inflammation (Kunnas et al. 1993) as well as *AVR1* in the chicken macrophage cell line HD11 in culture (unpublished). It is not currently known if the *AVR*s are expressed at the protein level. The production of the *AVR* transcripts is apparently not due to "leaky" transcription, since the mRNAs occurred only after inflammation-inducing conditions (Kunnas et al. 1993 and unpublished). Even in the tested tissues, the major fraction of transcribed avidin sequences was encoded by the *AVD* gene, supporting the suggestion that the progesterone and inflammation-induced avidins are products of the same gene.

Subsequently, the *AVD* gene was cloned using PCR on chromosomal DNA (Wallén et al. 1995). Comparing the molecular structures of the *AVD* and *AVR* genes revealed that the exon-intron structure of *AVD* was identical to that of the *AVR*s. The overall identity between *AVD* and the *AVR*s *1-5* was 91-95%, the differences consisting of nonrandomly distributed point mutations and a 6-bp deletion in the *AVR*s compared to *AVD* in exon 2. The 5′-flanking regions that have thus far been sequenced appear rather similar between the different *AVR*s, whereas the upstream sequence of the *AVD* gene differs radically (Keinänen 1994, Wallén et al. 1996 and unpublished observations), suggesting a differential regulation pattern for *AVD* and the *AVR*s. Differential regulation mechanisms possibly do operate under inflammation, suggested by the considerable differences in the levels of induction of *AVD* and the *AVR*s.

The putative AVR1-AVR5 proteins showed 74-81% identity to AVD and 85-100% identity to each other. The amino acids important to biotin binding were conserved, suggesting that the proteins are able to bind biotin (Keinänen et al. 1994).

2.7.3 Evolution of the avidin gene family

Comparisons between the sequences of the different avidin gene family members suggest that *AVD* and the *AVR*s have diverged relatively early. Duplication of the ancestral *AVD* gene has produced the first *AVR* copy, marked by the 6-bp deletion. The *AVR* copy has then been more susceptible to mutation and duplication, resulting in a growing family of tandemly arranged *AVR* genes. (Wallén et al. 1995). The most recently diverged genes are obviously *AVR4* and *AVR5*, since their coding regions are 100% identical. Subsequent gene conversion events may have counteracted mutations so that the sequences of the *AVR*s have gradually been homogenized. The involvement of gene conversion mechanisms in the evolution of the *AVD* gene family is supported by the notion that intron sequences are more conserved (97% on average) between *AVD* and the *AVR*s than exon sequences (90% on average, Wallén et al. 1995). Even though it is not currently known if the *AVR*s are expressed at protein level, the nonrandom distribution of point mutations suggests that there may be, or has been at some point during evolution, selective pressure acting also on the exon sequences of the *AVR*s (Wallén 1996).

Between-species evolutionary analyses cannot be performed, since the avidin gene (as well as possible avidin-related genes) has not been studied in any other eukaryotic species than the domestic chicken. However, avidin-like biotin-binding proteins are known to exist in various other oviparous species (Hertz & Sebrell 1942, Botte & Granata 1977, Korpela et al. 1981). Limited information about the biotin-binding proteins BBP-I and BBP-II found in chicken egg yolk suggests that these proteins, although sharing the function of biotin binding, do not exhibit extensive homology to avidin (Meslar et al. 1978). In contrast, the sea urchin fibropellins have an avidin-like domain which, despite sequence conservation, does not appear to bind biotin (Hunt & Barker 1989, Laitinen et al. 1999). The bacterial streptavidins have been extensively characterized and their genes have been cloned (Argaraña et al. 1986, Bayer et al. 1995). Streptavidins are highly similar to avidin in function and quaternary structure, despite low nucleotide sequence similarity (Livnah et al. 1993). The characteristics of the various avidin-like proteins will be discussed in more detail in the forthcoming doctoral theses of Olli Laitinen and Ari Marttila.

2.7.4 CR1 elements among the avidin gene family

The 5′-flanking regions of *AVR*s *4* and *5* have been shown to contain chicken repeat 1 (CR1) elements (Wallén et al. 1996). The CR1 elements are repetitive sequences belonging to the non-LTR class of retrotransposons, present in 7 000-30 000 copies dispersed throughout the chicken genome (Stumph et al. 1984, Silva & Burch 1989, Stevens 1996). The elements are often associated with DNaseI hypersensitive sites flanking functional genes, implicating a role for the CR1 elements in regulation of transcription (Stumph et al. 1984, Sanzo 1984 and references therein). In the case of *AVR4* and *AVR5*, the elements are located at 1.4-2.1 kb upstream of the genes and have a deletion at a site corresponding to a silencer element present in the CR1 elements of the chicken lysozyme (Baniahmad 1987) and apoVLDLII (Ryan et al. 1994) genes. In contrast, a site corresponding to a putative enhancer element in the apoVLDLII CR1 element is present in CR1*AVR4* and 5. Whether the CR1*AVR4* and CR1*AVR5* elements have any regulatory role is not currently known, since no transcripts corresponding to *AVR*s *4* or *5* have been detected. The presence of CR1 elements upstream of the other *AVR* genes or *AVD* is currently unknown.

3 AIMS OF THE STUDY

The five original *AVR* genes (*AVRs 1-5*) were cloned from two partially overlapping genomic clones (Keinänen et al. 1988, Keinänen et al. 1994) suggesting that the *AVR* genes are clustered. However, their organization, as well as their location relative to the avidin gene, remained elusive. Information on the organization of the genes was expected to provide clues on their evolution and function, especially since the functional importance of the *AVR* genes was unknown. It was also not clear if all the avidin-related genes had been cloned or if there were still more to be found. The aims of this study were to:

1. clone all *AVR* genes and reveal the chromosomal localization and organization of the gene family.

2. elucidate the molecular mechanisms acting on the evolution and maintainance of the gene family.

3. perform a preliminary analysis of the AVR proteins.

4 SUMMARY OF MATERIALS AND METHODS

The materials and methods are described in detail in the original publications (I-III).

4.1 Gene cloning and organization (I)

Two separate chicken genomic cosmid libraries (a commercial library from Clontech and a gridded library by Buitkamp et al. 1998) were screened for members of the avidin gene family. For screening the whole avidin gene family, avidin cDNA probe was used as a probe. Replicas of the gridded library were also probed with an avidin gene-specific oligonucleotide probe MA2 (I, Fig. 1).

To identify the genes present in each cosmid clone, the cosmid DNAs were mapped by carefully designed restriction enzyme mapping and Southern hybridization experiments. The novel *AVR* genes, *AVR6* and *AVR7*, were subcloned from the gridded library (cosmid clones O07-04 and C21-154; I, Fig. 2a). Novel alleles for the *AVD* gene were subcloned from both libraries. The subclone 1H4 (from cosmid 1-1-1) was sequenced completely and was later denoted as *AVDa2* (for *AVD* allele 2, II). Subclones 13H6 (from 13-1-1-1) and *AVD*H3 (from cosmid L09-154) were sequenced partially. An allele for *AVR2* (*AVR2a2*, II) was subcloned from the Clontech library (cosmid clone 1-1-1) and sequenced completely.

To determine the orientations of the genes, a "terminal PCR" approach was designed. Primers internal to the *AVR* genes were used alone or in combination with vector primers to determine the orientations and distances of the genes (I, and Fig. 5). The results were verified by restriction enzyme mapping of the cosmid clones and, in some cases, by partial sequencing.

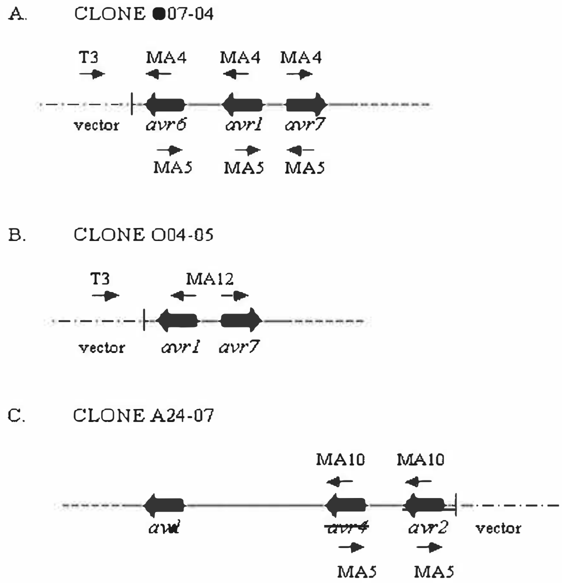

FIGURE 5 Examples of PCR strategies used to determine the order and relative orientations of the genes. A. Determination of the relative orientations of *AVR1*, *AVR6* and *AVR7*. B. Determination of the order of *AVR1* and *AVR7* in the cluster. C. Determination of the relative orientations of *AVR2* and *AVR4* (or *AVR5* on 1-1-1).

4.2 Localization by metaphase FISH (I)

The cytological localization of the avidin gene family was performed by Drs. Julio Masabanda and Alexei Sazanov (Technische Universität München, Freising-Weihenstephan, Germany). Briefly, different cosmid clones containing the *AVD* and *AVR* genes were labeled and hybridized to metaphase spreads prepared from female chicken fibroblast cultures. The hybridization signal was assigned to a chromosome band by calculating the fractional length from p-terminus (FL_{pter}).

4.3 Evolutionary sequence analysis (II)

The sequences of the *AVD* and *AVR* genes, now including the novel *AVR*s (*AVR*s *6* and *7*) as well as the *AVD* and *AVR2* alleles (*AVDa2* and *AVR2a2*), were examined in detail to reveal the evolution of the gene family. The program SimPlot (version 2.5) was used to visualize the "patchwork pattern", suggestive of gene conversion, of the sequences. Conversion tracts were further traced using the program GENECONV.

To investigate the phylogenetic histories of the genes, the program PLATO (partial likelihoods assessed through optimization) was used. A phylogenetic tree was constructed using the maximum likelihood (ML) method together with a substitution model obtained using the program Modeltest. The method revealed regions of putative recombination, gene conversion, selection or differential mutation rate. The phylogenetic analyses were conducted using the program PAUP* (Phylogenetic Analysis Using Parsimony [*and Other Methods]).

4.4 Gene copy-number assessment by fiber-FISH (III)

The fiber-FISH procedure was modified from Heiskanen et al. (1994) (see also Heiskanen et al. 1995, Heiskanen et al. 1996, Hellsten et al. 1995, Horelli-Kuitunen et al. 1999). Briefly, mononuclear white blood cells were obtained from blood samples of six chickens from two different breeds (LSL and a countryside breed denoted M). The LSL individuals were all females, whereas one of the M individuals was male (M5). The cells were mixed with low-melting point (LMP) agarose and solidified to form cell blocks. Blocks were also prepared from a chicken macrophage-type cell line, HD11. The blocks were treated with proteinase K to lyse the cells, and fiber slides were prepared by melting small pieces of the blocks in microwave oven and drawing the DNA on the slide with another glass slide. The cosmid clone K18-233 (I) was used as a control.

The cosmid clone K18-233 was also used as a probe for the whole avidin gene family locus (Fig. 6). For detecting the genes only (without intervening or flanking regions), *AVD* or both *AVD* and the *AVR* genes were used as probe. The fiber slides were hybridized with both probes simultaneously. The probes were detected using three layers of antibody or streptavidin-fluorophore conjugates. The signals were documented using a fluorescence microscope equipped with a digital camera. The number of red gene-specific signals overlapping green whole-locus signals was counted from 100 fibers from each individual as well as from the control and the cell line HD11. Statistical analysis was performed using the SPSS for Windows software.

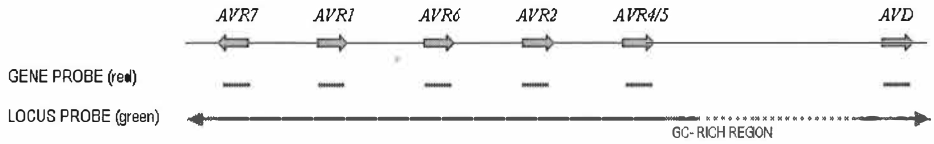

FIGURE 6 A schematic representation of the probes used in fiber-FISH.

4.5 Characterization of the avidin-related proteins (IV)

The putative AVR sequences, translated *in silico*, were compared to each other and to avidin by multiple sequence alignment. The theoretical isoelectric point of the AVRs and avidin were determined by using the GCG-package program Peptidesort. Molecular modeling was performed using the BODIL modeling environment.

The cDNAs for the *AVR*s were produced either by *in vitro* transcription and splicing (*AVR*s *1-4/5*), or by producing the cDNAs *in vivo* in transfected cells (*AVR*s *6* and *7*). The cDNAs were cloned into pFASTBAC1, and virus vectors for producing the AVR proteins were constructed and amplified according to the Bac-To-Bac system instructions (Gibco BRL). Recombinant AVR proteins were produced in Sf9 insect cells using biotin-free medium as previously reported (Airenne et al. 1997). Proteins were purified from the cells using affinity chromatography on a 2-iminobiotin column (AVRs 1, 3, 4/5, 6, 7) or biotin (AVRs 1, and 2) agarose.

The biotin-binding characteristics of the AVRs were studied using the IASyS optical biosensor as previously described in Laitinen et al. (1999) and Marttila et al. (1998).

The heat stability of the AVRs was studied using the following method: The individual AVRs were mixed with denaturing SDS-PAGE sample buffer in the presence or absence of biotin. The samples were then incubated at different temperatures, subjected to SDS-PAGE, and the relative proportions of tetrameric and monomeric forms were detected. The same assay was also performed using nonreducing conditions (with β-mercaptoethanol omitted from the sample buffer) to examine the presence of intersubunit cysteine bridges. Sensitivity to proteinase K was studied both in the absence and in the presence of biotin as described in Laitinen et al. (1999), and the isoelectric point of each AVR was determined by isoelectric focusing as in Marttila et al. (1998). The glycosylation patterns of the AVRs were studied by treating the proteins with Endo Hf or PNGase F glycosidase (New England Biolabs).

One preparation of commercial polyclonal anti-avidin and two different preparations of monoclonal anti-avidin were tested for their ability to recognize AVR proteins. The tests were performed using ELISA. Briefly, the wells of a microtiter plate were coated with avidin or AVR proteins, followed by incubation with the antibody preparations. Detection was performed using two layers of antibodies and AP-mediated colorimetry.

5 REVIEW OF THE RESULTS

5.1 Cloning and deducing the organization of the genes (I)

5.1.1 Screening of the cosmid libraries and identification of the genes

The positive cosmid clones identified by screening the libraries were mapped by restriction enzyme and hybridization analysis. Results were combined from several experiments to unambiguously identify the genes present in each cosmid clone.

The Clontech library was shown to contain only three genes: *AVR2*, *AVR5* and *AVD*. More than 10^6 clones were screened, which theoretically covered 32 equivalents of the chicken genome. Therefore, it is very unlikely that there were additional genes that were missed in the screening process. The cosmid clones from the Clontech library gave the first evidence that *AVD* is located in the same cluster with the *AVR*s. Furthermore, the terminal position of *AVD*, as well as the positions of *AVR*s 2 and *5*, were revealed.

In the gridded library (Buitkamp et al. 1998), the *AVD* gene and five different *AVR* genes were found. However, *AVR*s *3* and *5* were missing, whereas two novel genes, *AVR*s *6* and *7*, were found. The positive cosmid clones (nine altogether) were perfectly overlapping and showed no evidence of rearrangements. Remarkably, one clone (K18-233) contained the whole avidin gene family (I, Fig. 2a). The arrangement of *AVD* and *AVR2* were in agreement with that seen in the Clontech library. However, *AVR4* was found in the gridded library in the position corresponding to *AVR5* in the Clontech library.

The novel genes *AVR6* and *AVR7* were found to be 92% identical to *AVD* and 95-99% identical to each other and to the previously cloned *AVR* genes (I, Table 2). On the basis of their positions in the gene cluster, one of them could possibly represent an allele of the *AVR*3 gene, considering that *AVR*s *1-3* were originally cloned from a single genomic lambda clone (Keinänen et al. 1988). However, sequence differences between *AVR3* and *AVR*s *6* and *7* suggested that they are different genes.

An allele of the *AVR2* gene was subcloned from cosmid 1-1-1 and sequenced. The sequence differed by 0.6% from the previously cloned allele (Keinänen et al. 1994). Differences were also observed between the *AVD* alleles subcloned from cosmids 1-1-1 and 13-1-1-1 from the same library, indicating that these cosmid clones originated from different Z chromosomes. Both sequences also differed from the previously published *AVD* gene sequence (Wallén et al. 1995). Another *AVD* allele was cloned from cosmid L09-154 of the gridded library. The main differences between the various *AVD* alleles were found at the four-nucleotide inversion point present in the first intron of the *AVD* and *AVR* genes (I, Fig. 4). The different forms of this inversion site are shown in figure 7.

```
a.              166           179
      AVD       tcttc actg cagTG
      AVDa2     tcttc gtca cagTG
      13H6      tcttc attg cagTG
      AVDH3     tcttc attg cagTG

b.    AVR1      tcttc gtca cagTG
      AVR2      tcttc actg cagTG
      AVR2a2    tcttc attg cagTG
      AVR3      tcttc attg cagTG
      AVR4      tcttc attg cagTG
      AVR5      tcttc attg cagTG
      AVR6      tcttc gtca cagTG
      AVR7      tcttc gtca cagTG
```

FIGURE 7 The inversion site in intron 1 of the *AVD* and *AVR* genes. **a.** Comparison of different alleles of *AVD*. *AVD*=original *AVD* sequence by Wallén et al. (1995), *AVD*a2=*AVD* subclone 1H4 from cosmid 1-1-1, 13H6=*AVD* subclone from cosmid 13-1-1-1, *AVD*H3=*AVD* subclone from cosmid L09-154. **b.** Comparison of the different *AVR* genes, including the *AVR2* allele 2 (*AVR*2a2). Intron sequence is written in lowercase, and the first two nucleotides of exon 2 in uppercase (see also I, Fig. 4).

5.1.2 The organization of the avidin gene cluster.

By combining the results from the Southern blot and the "terminal PCR" experiments (as well as sequencing in some cases), a map of the organization of the *AVD* and *AVR* genes could be drawn (Fig. 8; also I, Fig. 2b). All the other genes were arranged tandemly except for *AVR7* which was located in an inverted orientation at one end of the cluster. The distances between the *AVR* genes were 2.8 kb, with the exception of *AVR*s *1* and *7* for which the distance was 2.5 kb. The *AVD* gene was located 9 kb downstream of the *AVR* cluster (Fig. 8; also I, Fig. 2b), the intervening region containing GC-rich sequences (unpublished results).

The cosmid clones covered about 100 kb of chromosomal DNA, with the avidin gene cluster contained within a 27-kb region (I, Fig. 2). Since the cosmid clones of the gridded library extended 35-40 kb beyond the avidin gene cluster in both directions (I, Fig. 2a), it is very likely that the whole avidin gene family was included in this cluster.

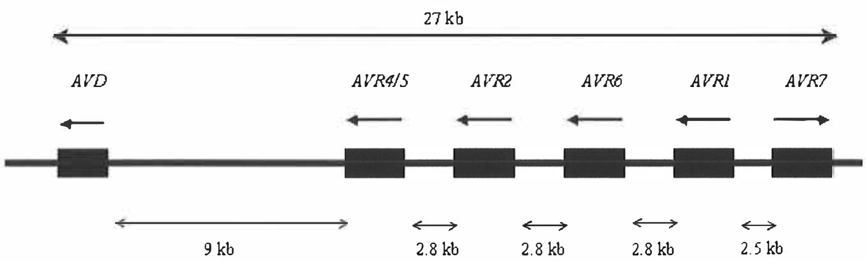

FIGURE 8 The organization of the avidin gene family.

5.1.3 Chromosomal localization.

A single locus was observed for the avidin gene family. The hybridization signals were assigned on the long arm, telomeric region of chromosome Z (I, Fig. 3). The signals were located at an average FL_{pter} of 0.83, placing the avidin gene cluster at band Zq21 (Table 3).

TABLE 3 Positioning of the avidin gene locus onto chromosomal band Zq21.

Probe clone	Genes probed for	Chromosomal position	$FL_{pter} \pm SD$
13-1-1-1	*AVD*	Zq21	0.84 ± 0.037
1-1-1	*AVD, AVR2,5*	Zq21	0.85 ± 0.026
C21-154	*AVR1,7*	Zq21	0.82 ± 0.025
F18-104	*AVR1,2,4,6,7*	Zq21	0.80 ± 0.028
A24-07	*AVD, AVR2,4*	Zq21	0.82 ± 0.023
K18-233	*AVD, AVR1,2,4,6,7*	Zq21	0.85 ± 0.029
Average FL_{pter}			0.83

5.2 Evolutionary Sequence Analysis (II)

5.2.1. Nucleotide variation

A total of 125 nucleotide substitutions were found between the *AVD* and *AVR* gene sequences, most of which were in exons. This figure does not include the 6-bp deletion in exon 2 of the *AVR*s as compared to *AVD*, nor the 1-bp insertion in intron 2 of *AVR*s *1* and *2* (Keinänen et al. 1994, Wallén et al. 1995). The corrected distance between *AVD* and the *AVR*s ranged from 8 to 20%, and from 1 to 10% among the *AVR* genes. Different regions of the genes are clearly subject to differential mutation rates. Exon 2 exhibited the highest level of divergence between *AVD* and the *AVR*s. The most conserved exon, on the other hand, was exon 1. Exon 1 encodes the signal peptide that targets AVD for secretion. As already noted in previous studies (Wallén et al. 1995), the introns are, on average, more conserved between *AVD* and the *AVR*s than exons.

Remarkable differences were found in the transition/transversion ratio between the exons and introns: transitions were up to 17 times more frequent in introns (II, Table 1A). Moreover, most of the nucleotide substitutions between all the genes occurred in the first or second codon position (Wallén et al. 1995 and current study), and about 50% of these substitutions lead to an amino acid change. This pattern of substitution clearly indicated that different gene regions are subject to different models of evolution. The Spatial Phylogeny Variation (SPV) analysis again pointed out the curious features of exon 2.

5.2.2. Gene conversion

The alignment created using SimPlot revealed the "patchwork pattern" of the *AVD* and *AVR* genes (II, Fig. 2), which strongly suggests that gene conversion plays a major role in modifying the gene family. For example, the figure showed the distinct similarity of *AVR*s *4* and *5* to *AVD* in exon 3, as well as the extreme conservation of intron 2 in all the genes. The putative gene conversion tracts were determined in more detail with GENECONV. Ten putative conversion tracts were observed, four of which occurred between *AVD* and *AVR* genes (Fig. 9 and II, Table 2). One tract, between *AVR3* and *AVR7*, covered most of the length of the genes, whereas the other tracts were generally short (22-271 bp).

5.2.3. Phylogenetic analyses

The phylogenetic analysis grouped the *AVR* genes together, with *AVD* being the most divergent gene (II, Fig. 3). Among the *AVR* group, the most closely related genes were *AVR*s *1* and *6*, while *AVR4/5* diverged most from the other *AVR* genes (II, Fig. 3). Because of the anomalous nature of exon 2, separate phylogenetic analysis was conducted for this region. Indeed, when exon 2 was excluded (II, Fig. 4A), *AVR3* resulted as the outgroup of a cluster formed by *AVR*s *1*, *6* and *7*. Analysis of exon 2 alone, on the other hand, placed *AVR7* as closely related to *AVR6*, while *AVR*s *3* and *4/5* resulted as being very closely related to each other (II, Fig. 4B). All these data suggest that exon 2 is subject to diversification whereas intron 2 is homogenized by gene conversion.

FIGURE 9 The approximate positions of putative conversion tracts. The figure represents an *AVR* gene (with exons numbered). The solid lines denote conversion events between the different *AVR* genes, whereas dashed lines indicate conversion of an *AVR* gene by *AVD*.

TABLE 4 Compilation of the main results of the evolutionary sequence analysis.

- Phylogenetic relationships: In general, the AVRs are more closely related to each other than to AVD, with AVR4 and AVR5 having the closest resemblance to AVD. Exon 2 seems to evolve in a manner different from the rest of the genes.

- Sequence conservation: On average, introns are better conserved than exons between AVD and the AVRs. Exon 1, encoding the secretion signal peptide, is the most conserved, and exon 2 is the most divergent region. Several gene conversion tracts localize in exon 2 → diversification by gene conversion?

- Transition/transversion ratio: Transitions are drastically more common in exons than in introns. About 50% of the substitutions are nonsynonymous.

- Conversion bias: The AVRs are converted by AVD, but not vice versa → involvement of the "master-slave" conversion rule?

5.3 Gene copy-number assessment by fiber-FISH (III)

The fiber-FISH results supported the deduced organization of the avidin gene family: *AVD* was located at one end the cluster, followed by a gap and the *AVR*s clustered close to each other. However, the total gene count, as well as the *AVD* count, differed significantly between individuals (Figs. 10 and 11, and

III, Table 1). Interestingly, the *AVD* count correlated with the *AVR* count (III, Table 1). The "standard", or most frequent, gene number was 4 or 5 *AVR*s and one *AVD*, depending on the individual (Fig. 10 and III, Fig. 1).

The gene counts differed widely also within individuals (III, Fig. 2). The counts from each individual were distributed according to the Gaussian curve, suggesting reciprocal sequence exchange as the mechanism behind the variation.

Duplications of the whole locus were also observed. The duplicated clusters were seen in all possible orientations: head-to-head, tail-to-tail, and tandem. In some cases, two *AVR*-clusters were seen connected by a single *AVD*.

Gene signals were frequently missing on the control slides (cosmid K18-233). This template was known to contain 5 *AVR*s and one *AVD* (I). Correspondingly, whereas all the genomic samples showed Gaussian distribution for the gene counts, the control counts showed a distribution highly biased to the left (III, Fig. 1).

5.4 Characterization of the AVR proteins (IV)

5.4.1 Sequence analysis and molecular modeling

The biotin binding residues were well conserved in all AVRs (Keinänen et al. 1994 and IV), showing only three amino acid changes (IV, Fig. 1). With the exception of AVR7, the AVRs lacked the glycosylation site (Asn17) of avidin (IV, Fig. 1). Instead, they exhibited different patterns of alternate glycosylation sites (NXS/T, 2-3 sites/AVR), and were therefore expected to be more heavily glycosylated than avidin. All AVRs, except for AVR2, had a third cysteine in addition to the two found in avidin (IV, Fig. 1). In AVRs 1, 3, 6 and 7, the third cysteine was located at the beginning of β5, at the same side of the barrel as the "normal" cysteines (those found in avidin). In AVR4/5, the extra cysteine was located near to the C-terminus of the protein, also close to the other cysteines. There was a two-amino acid deletion in all AVRs as compared to avidin, located in a loop at the end of the barrel opposite to the biotin-binding pocket. The subunit interface regions showed a variable number of amino acid substitutions compared to avidin. Interestingly, the largest interface, between subunits 1 and 4, showed numerous amino acid substitutions. In avidin, a total of 22 residues confer intersubunit contacts within this interface. The AVR proteins showed nine substitutions within this region, seven of which were found in all AVRs and two in all but AVR4/5 (IV, Table 1).

Molecular modeling suggested that the AVRs differ very little from avidin in tertiary and quaternary structure, despite the amino acid differences. Even the numerous substitutions in the interface regions should allow proper

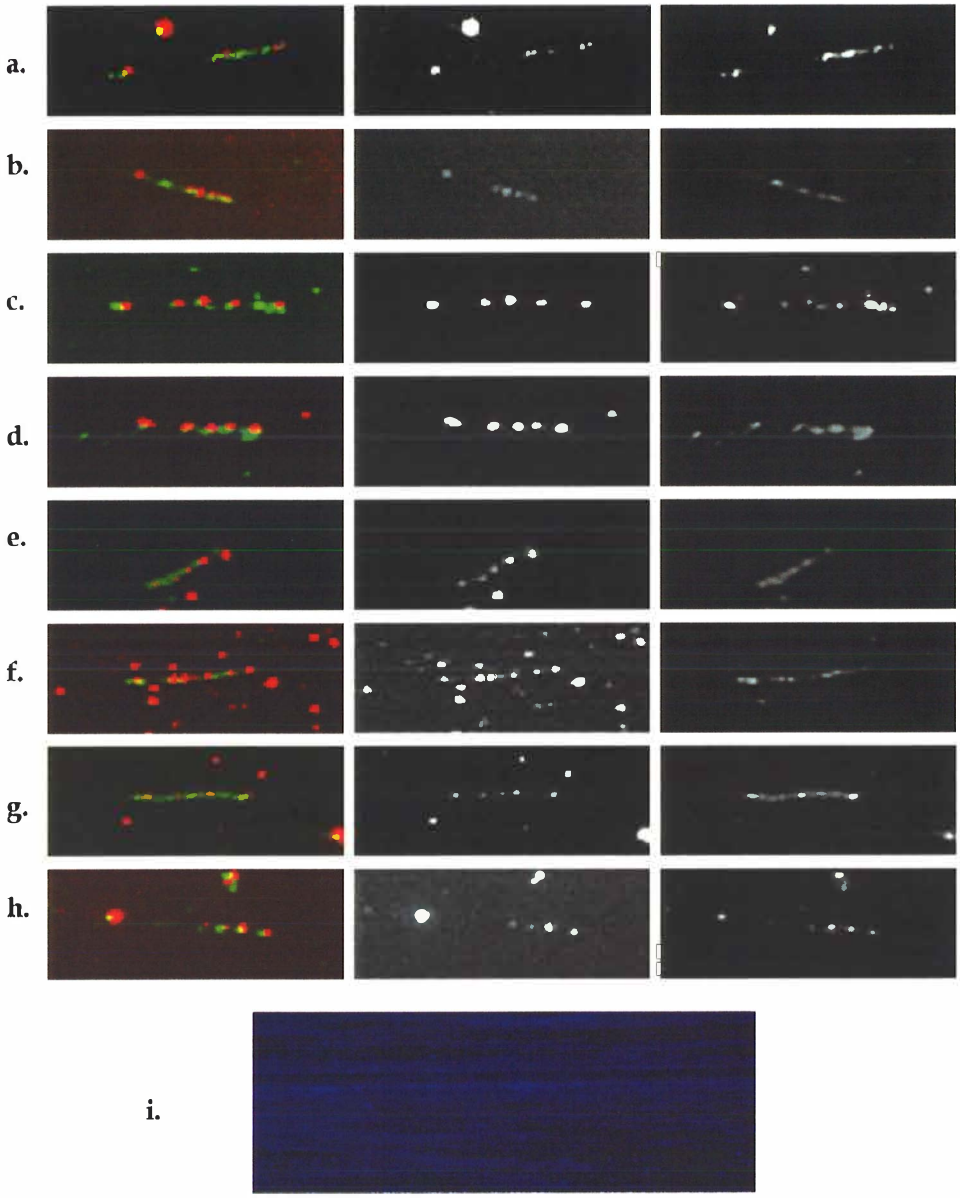

FIGURE 10 **a.-h.** Representative fibers from different individuals. **a.** Control DNA (cosmid K18-233) showing 5 *AVRs* + 1 *AVD*. **b.** Cell line HD11; 4 *AVRs* + 1 *AVD*. **c.** M4; 4 *AVRs* + 1 *AVD*. **d.** M5; 4 *AVRs* + 1 *AVD*. **e.** M6; 4 *AVRs* + 1 *AVD*. **f.** LSL1; 5 *AVRs* + 1 *AVD*. **g.** LSL2; 4 *AVRs* + 1 *AVD*. **h.** LSL3; 4 *AVRs* + 1 *AVD*. On each row, separate grayscale images for the locus signal (rightmost image) and the gene signals (center) are shown. The leftmost image represents the two channels pseudocolored and superimposed. **i.** DAPI-stained extended DNA on a microscope slide.

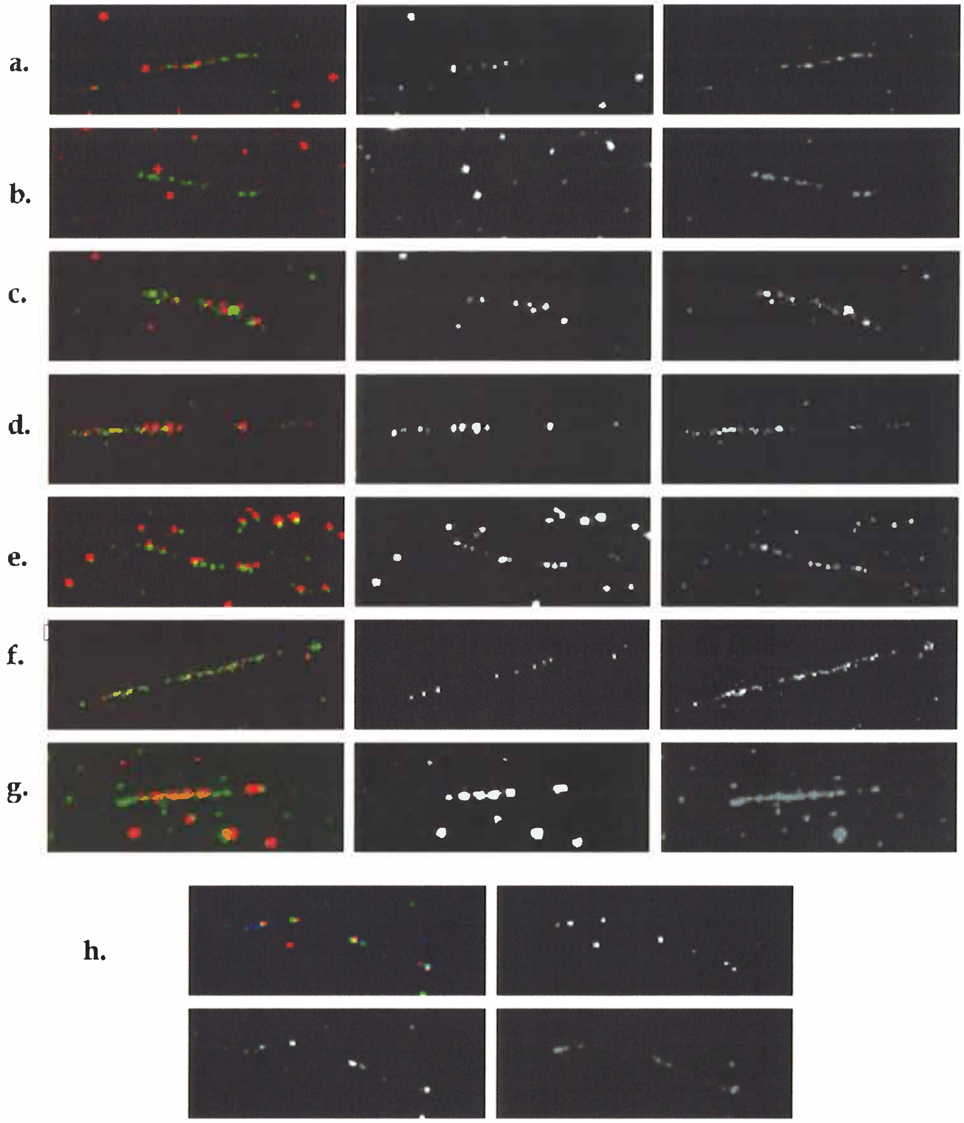

FIGURE 11 Examples of fibers showing various rearrangements. **a.** LSL1; 6 *AVRs* + 0 *AVD*. **b.** LSL2; 1 *AVR* + 0 *AVD*. **c.** LSL3; 5 *AVRs* + 2 *AVDs*. **d.** M6; 9 *AVRs* + 1 *AVD*, with possibly one more *AVR* on the right side of the *AVD*. **e.** LSL1; two clusters (4 *AVRs* + 1 *AVD* and 3 *AVRs* + 1 *AVD*) oriented head-to-head. **f.** M6; 5 *AVRs* + 3 *AVDs*. **g.** Cell line HD11; 5 *AVRs* + 1 *AVD*. **h.** M4; two clusters (2 *AVRs* + 1 *AVD* and 3 *AVRs* + 1 *AVD*) oriented head-to-head. For this fiber, also the image representing the DAPI-stained DNA is shown (lower right corner).

construction of the tetramers. However, the third cysteines found in the AVRs (all but AVR2) might have more significant effects on the higher-order structures. The extra cysteines may form disulfide bridges between subunits in AVR4/5, producing crosslinked dimers. In the other AVRs, the extra cysteines may form disulfide bridges between different tetramers, possibly causing aggregation of the proteins. The biotin-binding properties of the AVRs were also expected to be similar to avidin, with the exception of AVR2 in which the substitution of Lys111 by Ile was expected to change the shape of the binding pocket and therefore to affect biotin binding.

5.4.2 Analysis of the recombinant AVR proteins

All AVR proteins were successfully expressed in insect cells and were purified to homogeneity by 2-iminobiotin or biotin agarose affinity chromatography. Interestingly, AVR2 could not be purified at all and AVR1 only poorly on 2-iminobiotin agarose. Instead, they were purified on biotin agarose and could be eluted under acidic conditions. All AVR preparations showed multiple bands in SDS-PAGE. The different bands were found to be differentially glycosylated forms, since treatment with endo H_f glycosidase eliminated the higher-molecular weight bands (IV, Fig. 6). The proteins showed remarkable heat stability, with a portion of tetramers remaining intact in the presence of biotin even upon boiling. The AVRs also showed remarkable resistance against proteolysis, as judged by proteinase K treatment results. Again, the stability was higher in the presence of biotin. Taken together, the AVRs showed stability similar to or even greater than avidin (not shown).

Molecular modeling predicted that AVR4/5 can form intersubunit disulfide bridges. However, non-reducing SDS-PAGE showed that AVRs 1, 3, 4/5, 6 and 7 all have a tendency to form dimers (IV, Table 3). The proteins were observed to dissociate from tetrameric to dimeric and further into monomeric states along increasing temperature (IV, Fig. 5). However, AVR2 disintegrated from tetramers directly into monomers, similarly to native avidin (IV, Table 3).

In functional tests, AVRs 3, 4/5, 6, and 7 showed irreversible biotin binding, similarly to avidin. In contrast, AVR1 exhibited 25% and AVR2 90-95% reversibility following addition of free biotin (IV, Table 2), consistent with the purification results. Because of the extremely high biotin-binding affinity, a dissociation constant could not be determined except for AVR1 (K_d=2.4 x 10^{-8}) and AVR2 (K_d=8.3 x 10^{-7}). Binding to 2-iminobiotin in the IASyS cuvette could only be observed for AVR4/5, with a binding affinity similar to avidin (K_d=2 x 10^{-8}) (IV, Table 2). The lack of binding for the other AVRs was surprising, since most of them were originally purified on 2-iminobiotin agarose. However, the relatively short linker between 2-iminobiotin and the activated group of the IASyS cuvette may sterically inhibit the binding.

In immunological analyses, polyclonal rabbit anti-avidin recognized the AVRs weaker than avidin. Neither of the two monoclonal avidin antibodies tested recognized any AVRs (IV, Fig. 7).

TABLE 5 Compilation of the characteristics of the AVR proteins.

	AVD	AVR1	AVR2	AVR3	AVR 4/5	AVR6	AVR7	Interpretation
Isoelectric point	10.4	7.3	4.7	10.2	10.0	7.3	7.3	Some AVRs are neutral or acidic. Implications for altered cell binding or other functions in tissues?
No. of glycosylation sites	1	3	2	2	3	3	3	The AVRs are more heavily glycosylated than AVD. Implications for altered cell binding or other functions in tissues?
No. of cysteine residues	2	3	2	3	3	3	3	AVRs 1 and 3-7 may exhibit inter-monomeric disulfide bridges.
Occurrence of dimeric forms after boiling	no	yes (40%)	no	yes (50%)	yes (50%)	yes (50%)	yes (20%)	Inter-monomeric disulfide bridges in AVRs 1 and 3-7 render dimeric forms highly stable. In general, the AVRs are even more stable than AVD.
Occurrence of tetrameric forms after boiling	no	yes (20%)	no	no	no	no	no	
Reversibility of biotin binding	none	18%	94%	3%	2%	5%	3%	The replacement of Lys-111 by Ile in AVR2 renders biotin binding essentially reversible.
2-iminobiotin binding	+++	+	-	++	++	++	++	AVRs 3-7 (and AVR1) can bind 2-iminobiotin, as judged by affinity purification (no binding in IaSys measurements).
Recognition by polyclonal anti-avidin	++++	+	(+)	+	++	++	++	The AVRs are immunologically distinct from AVD
Recognition by monoclonal anti-avidins	++++	(+)	-	-	(+)	-	-	

6 DISCUSSION

These studies were conducted to reveal the characteristics of the chicken avidin gene family in detail. Chicken genomic cosmid libraries were screened in order to clone all members of the gene family and to be able to deduce the arrangement of the genes. The gene sequences were closely examined to reveal the evolutionary aspects concerning the gene family. Fluorescence *in situ* hybridization studies were performed on metaphase chromosomes to reveal the location and distribution of the gene family members in the chicken genome. The hybridization studies were also applied to extended chromatin fibers to verify the total number and organization of the genes and to assess their possible copy-number fluctuation. Finally, the characteristics of the avidin-related proteins were studied, both by sequence analysis and molecular modeling, as well as by expressing them as recombinant proteins.

6.1 Characteristics and evolution of the avidin gene family (I, II)

6.1.1 Organization (I)

According to our results, the avidin gene family comprises the *AVD* gene, which is single-copy in almost all instances, and a variable number of *AVR* genes arranged as a repeated array within a region of 27 kb of chromosomal DNA (I, III). The gene cluster is located telomerically on the chicken sex chromosome Z, on band Zq21 (I, Fig. 3). The avidin gene is located at one end of the array, followed by a space of 9 kb and the *AVR* cluster with intergenic distances of 2.5-2.8 kb. In the clusters characterized in this study, all other genes were arranged tandemly in the same orientation except *AVR7*, which was inverted (I, Fig. 2b).

The localization result explains why two different alleles of the *AVD* gene were found in the Clontech library, whereas only one was isolated from the gridded library. In chickens the female is the heterogametic sex (ZW), whereas males have a pair of usually nonidentical Z chromosomes (ZZ) (Stevens 1996). The Clontech library was made from the DNA of a male chicken, thus possessing two sets of the *AVD/AVR* genes. This library can therefore provide information on the degree of polymorphism between alleles of each gene within an individual. Indeed, partial sequencing revealed differences in the two *AVD* alleles from the Clontech library (Fig. 7). The gridded library, on the other hand, was made from DNA of a female chicken (Buitkamp et al. 1998), thus possessing only a single allele of each gene. The gridded library therefore ensured cloning of nonallelic *AVR* copies.

Inversions. The reversed orientation of *AVR7* is an exception among the otherwise tandem arrangement of the genes in the avidin family (I, Fig. 2b). Graham (1995) suggests that the organization of a gene family can interconvert between tandem arrangement and randomly oriented cluster. However, the *AVD/AVR* genes do not seem to be particularly prone to inversion, since identical orientations have been observed in three different libraries for *AVR*s *2* and *4/5* with respect to each other (I, Fig. 2a and Wallén, unpublished). Also, the orientation of the *AVD* gene was identical in both the Clontech and the gridded library (I, Fig. 2a). Since this suggestion is based on studies on only a few haplotypes, further studies of the orientation of the genes in different individuals, using PCR methods, would settle the issue.

In contrast, some inversion mechanism operates frequently within the coding regions of the *AVD* and *AVR* genes: there is a four-nucleotide inversion point in the first intron, the sequence of which varies between different genes as well as between different alleles of each gene. This inversion point has been sequenced from three alleles of the *AVD* gene: in the original avidin gene (Wallén et al. 1995) the sequence was ACTG, whereas in the alleles characterized in this study both the inverted form GTCA as well as the mutated form ATTG were found (Fig. 7). Comparably, in the original *AVR2* sequence by Keinänen et al. (1988) the sequence of this inversion point was ACTG, instead of ATTG found in the allele characterized here (Fig. 7). Since only three different forms of the inversion point have been observed, it may be possible that gene conversion affects this region, preventing it from mutating further.

Locus organization vs. expression pattern. Functional implications of the tandem arrangement of multiple gene copies were discussed in section 2.1.1. Interestingly, *AVD* is the only gene in its family that is expressed at considerable amounts. As the expression patterns of the *AVR* genes await in-depth studying, it remains to be seen if the arrangement of the genes is correlated with their function.

6.1.2 Nucleotide sequence variation and gene conversion (I, II)

The nucleotide sequence differences between alleles of the same gene were found to be about 2% for *AVD* and 0.6% for *AVR2*. At the amino acid level, the two allelic variants deduced for avidin show three differences (Fig. 12). The differences are located at the N-terminal part of the mature peptide, at β1 (Thr→Asp), loop 2 (Arg→Lys), and β3 (Ile→Thr) (Fig. 12 and IV, Fig. 1). Since AVD variants with slightly differing antigenic structures have been reported (Korpela et al. 1982), it may be that some degree of amino acid changes, evidenced by the reduced binding by the avidin antibody, can be tolerated without disturbing the biotin binding activity. Furthermore, Huang & DeLange (1971) reported heterogeneity (Ile or Thr) at position 34 in their AVD amino acid sequence, supporting the idea.

```
1H4     MVHATSPLLLLLLLSLALVAPGLSARKCSLTGKWDNDLGSNMTIGAVNSKGEFTGTYTTA
AVD     MVHATSPLLLLLLLSLALVAPGLSARKCSLTGKWTNDLGSNMTIGAVNSRGEFTGTYITA
        ********************************** **************.******* **

1H4     VTATSNEIKESPLHGTQNTINKRTQPTFGFTVNWKFSESTTVFTGQCFIDRNGKEVLKTM
AVD     VTATSNEIKESPLHGTQNTINKRTQPTFGFTVNWKFSESTTVFTGQCFIDRNGKEVLKTM
        ************************************************************

1H4     WLLRSSVNDIGDDWKATRVGINIFTRLRTQKE
AVD     WLLRSSVNDIGDDWKATRVGINIFTRLRTQKE
        ********************************
```

FIGURE 12 Comparison of the amino acid sequences of two *AVD* allelic variants. 1H4: *AVD* subclone from cosmid 1-1-1 (current study); *AVD*: original *AVD* sequence from Wallén et al. (1995). Gaps indicate nonhomologous substitutions, and the dot designates a homologous substitution.

The fact that the allelic differences were smaller in *AVR2* than *AVD* suggests that gene conversion acts frequently on the *AVR* genes, slowing their nucleotide substitution rate and thus preserving the homogeneity of the *AVR* sequences. In contrast, *AVD* seems to be well protected from becoming homogenized with the *AVR*s. As intrachromosomal recombination has been observed to decrease with increasing distance between the participating repeats (Martinsohn et al. 1999), the separation of *AVD* from the *AVR*s by 9 kb may represent an efficient barrier against gene conversion and possibly crossing-over (see below). It must be noted, however, that these assumptions are based on very small sequence data of *AVD* alleles, and are thus only speculative.

Closer inspection of the *AVR* sequences further suggests that gene conversion and/or recombination play a major role in modifying the gene family. For example, the sequence of *AVR3* is identical to that of *AVR4/5* in the first exon and intron, and switches then towards the other *AVR*s (Fig. 13 and I, Fig. 4). The switch strongly suggests that the 5'-end of the gene has been recombined with or converted by *AVR4/5* or, alternatively, the 3'-end of the gene has been converted by the other *AVR*s. Interestingly, *AVR4/5* is >96%

identical to *AVD* through exon 3 (II, Fig. 2). Considering the fact that *AVR4/5* shows the 6-bp deletion characteristic for all *AVRs*, it may be that these genes have partially been converted by *AVD*. This hypothesis necessitates the assumption that the conversion process exhibits polarity, so that *AVD* can convert *AVRs* but not *vice versa*. Considering the expression patterns of the *AVD* and *AVR* genes, the "master-slave" rule, i.e. that the gene expressed in higher level converts the gene expressed in lower level (Papadakis & Patrinos 1999), is an appealing model for explaining the directionality for gene conversion in the avidin gene family.

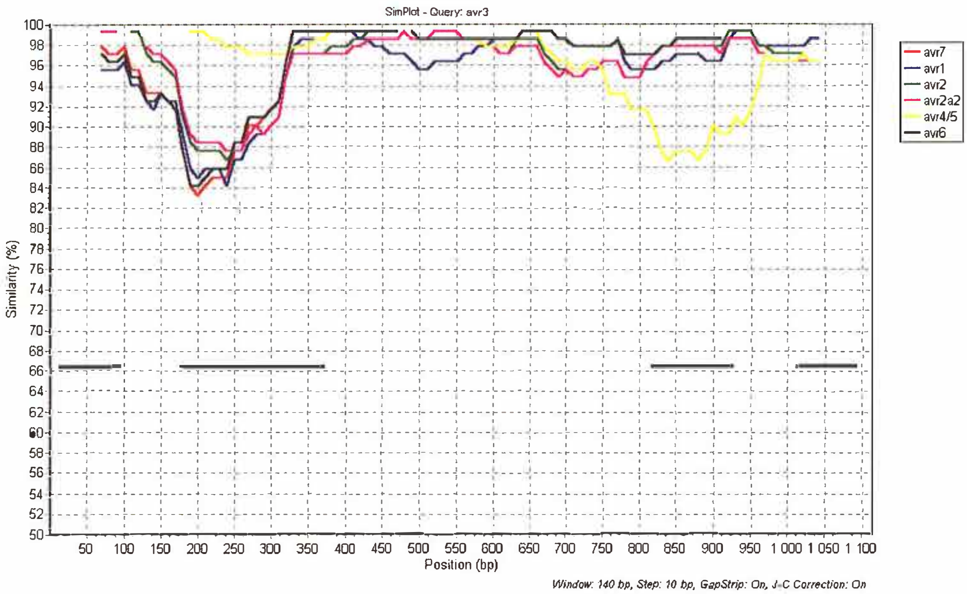

FIGURE 13 Comparison of the *AVR* genes using the SimPlot program. *AVRs 1, 2* (including the allele *AVR2a2*), *4/5, 6* and *7* were plotted against *AVR3* to visualize the patchwork-like constitution of the genes. *AVR4/5* is highly similar to *AVR3* in its 5′-region, whereas in exon 3 *AVR4/5* diverges from all the other *AVRs*. The approximate positions of exons are represented by bars below the curves. (See also II, Fig. 2.)

The SimPlot figures (Fig. 13 and II, Fig. 2) visualize the fact that intron 2 is the most conserved region in the *AVD* and *AVR* genes. Thus, intron 2 seems to represent a hotspot for double recombination or, more likely, gene conversion. The 5′-border of the conversion tract is probably at the end of exon 2, slightly downstream from the highly diverse sequence regions and the 6-bp deletion (I, Fig. 4). The 3′-border is more difficult to address, but seems to lie at around position 750 according to the SimPlot figures (Fig. 13 and II, Fig. 2). Thus, the putative conversion track begins at the end of exon 2 and ends before the beginning of exon 3.

6.2 Copy-number fluctuation (III)

Tandemly repeated sequences are thought to be susceptible to copy-number fluctuations and rearrangements by mechanisms such as unequal crossing-over or unequal sister chromatid exchange (Li 1997). Telomeric position of the repeated sequences further facilitates rearrangements (Perry & Ashworth 1999). The avidin gene family meets both criteria, suggesting that the family may be subject to accelerated evolutionary rate by recombinational mechanisms. Furthermore, not only the coding sequences but also the intergenic sequences among the avidin gene family appear to be conserved (unpublished), making unequal crossing-over events highly plausible. On the basis of our findings, we had reason to believe that some rearrangement mechanism(s) may be constantly acting on the *AVD* gene cluster. This view was supported by the facts that none of the 11 cosmid clones analyzed in this study (I) contained the previously cloned *AVR*3 gene (or any gene attributable as an allele of *AVR3*; I, Fig. 2a), but contained two previously uncharacterized *AVR* genes instead. Furthermore, *AVR*s *4* and *5* could not be found together in any of the cosmid clones in this study, *AVR4* being present in three cosmids from the gridded library and *AVR5* being found in cosmid 1-1-1 from the Clontech library (I, Fig. 2a). On the contrary, Keinänen et al. (1994) originally cloned both genes from a single genomic lambda clone of 20 kb using a library made from female chicken DNA, ruling out the possibility that the original *AVR*4 and 5 genes were each other's alleles. Thus, it seemed possible that the number and combination of *AVR* genes found in different individuals can vary. This copy-number fluctuation hypothesis was extended to suggest variation also within individuals. As unequal sister chromatid exchange can, in principal, occur in any mitotic cell division, an individual could be a mosaic of cells harboring different copy numbers. For gene families such as the *AVD* family, unequal crossing-over could thus occur considerably more frequently than average.

6.2.1 Evidence for fluctuation by unequal crossing-over

The hypothesis was tested using the fiber-FISH method. Indeed, the number of *AVD* and *AVR* genes was found to vary within individuals as well as between different individuals (Figs. 10 and 11 and III, Fig. 2). The gene counts were distributed according to normal (Gaussian) distribution (III, Fig. 1), suggesting that the mechanism underlying the variation is unequal crossing-over. By definition, an increase in gene number on one locus is accompanied by a decrease on the other locus in unequal sequence exchange, as observed with the *AVD* and *AVR* genes. Jeffreys et al. (1998) showed a similar distribution for the minisatellite MS32 repeat number in human sperm cells. Since sperm cells are haploid, the mechanism underlying the minisatellite repeat fluctuation was meiotic recombination, mainly gene conversion and, to a lesser extent, unequal crossing-over (equal crossing-over was also observed). Since the chicken chromosome Z harboring the *AVD* gene family is present in only one copy in the female, the recombination mechanism in the female must have been intrachromosomal gene conversion and/or unequal sister-chromatid exchange. In the male, unequal crossing-over between the two copies of chromosome Z as well as unequal SCE could occur. Surprisingly, the male did not show greater variation in gene number than the females, as might be expected because of the second level of possible recombination in the male.

The number of *AVR* genes varied more frequently than the number of putative *AVD* copies (III). This together with the observed bias in gene conversion (*AVD* converts the *AVR*s but not vice versa; II) suggests that *AVD* is somehow protected from recombination. The 9-kb region separating *AVD* from the *AVR*s is apparently GC-rich (unpublished) and is frequently lost upon cloning. It has also been observed to recombine with plasmid vector sequences (Ahlroth 2001). Therefore, this region may act as a hotspot for recombination, mediating rearrangements within the *AVR* cluster while isolating *AVD* from most of the recombination events.

6.2.2 Technical considerations

Resolution. The main technical problem associated with the fiber-FISH technique is considered to be the loss of signals due to inefficient hybridization of the probes (Florijn et al. 1996, Suto et al. 1998). Due to the helical conformation of DNA, for example, not all genes are equally accessible to the probes. As the DNA fiber is attached to the glass slide at every turn, the signal array resembles a "pearls in a chain" arrangement rather than a continuous stretch (Figs. 10 and 11 and III, Fig. 2). In the current application, the technique was operating near the limits of its resolution. A resolution limit of 1 kb has been proposed for fiber-FISH, determined by the optical resolution of the microscope (0.2-0.35 μm) assuming a condensation degree of 1 kb/0.34 μm for extended DNA (Florijn et al. 1996 and references therein). The detection limit is lower, however, since sequence-tagged sites (STS) or expressed sequence tags (EST) have been mapped by fiber-FISH using probes as short as 200 bp (Florijn et al. 1996, Horelli-Kuitunen et al. 1999). Thus, signals of 200 bp separated by 1 kb should be possible to detect under optimal conditions. However, the detection efficiency decreases with decreasing probe size, being 70-90% for probes >400 bp and 30% for probes of 200-250 bp according to Florijn et al. (1996). Furthermore, Horelli-Kuitunen et al. (1999) found that detecting two or three ESTs simultaneously was even less efficient, about 15%.

Evaluating the fiber-FISH results. In the current study, each gene (1.1 kb) was represented by a single signal dot, and the intergenic distance of 2.5-2.8 kb was sufficient to separate the gene signals only on highly extended fibers (Figs. 10 and 11 and III, Fig. 2). A considerable amount of variation thus represented technical errors, necessitating the analysis of large numbers of loci (100 fibers/sample). However, fibers showing gene signals exceeding the "default" gene count (4 or 5 *AVR*s + 1 *AVD*, depending on the individual; Fig. 10 and III, Fig. 1) can be considered more reliable than those showing missing signals. This is because the control target, cosmid K18-233, very seldom showed extra gene signals (5% of the cases) whereas signals fewer than 6 (total gene count) were frequently observed (56% of the cases, III, Fig. 1a). Also Florijn et al. (1996) found that nonspecific hybridization producing extra signals was rare in fiber-FISH. Surprisingly, the authentic chicken or cell line samples showed less technical variation than the control in the current study. This may be due to the excess background staining of the fibers on the control slides.

The degree of true gene number variation as opposed to technical variation in the authentic samples can be inferred from the diagrams in Figure 1 (III). For each case of increased gene number, there must be a corresponding count of decreased gene number. Comparably, excess column heights on the left side of the peak can be removed and the heights adjusted to correspond to those on the right side of the peak. For some individuals, however, the values above the peak exceed the values below the peak (HD11, M5 and LSL3; III, Fig. 1). In these cases, genes may have been gained by long-tract gene conversion. Gene conversion over tracts as long as 10 kb has been observed (Martinsohn et al. 1999 and references therein). Thus, clusters of *AVD* and/or *AVR* genes may be gained by "copying" from one locus to another, leaving the gene number in the donor locus unchanged.

Locus duplications and other rearrangements. In addition to copy loss and gain, several cases of duplication of the whole locus were observed (Fig. 11 and III, Fig. 2). Duplicated gene clusters were found in all kinds of orientations: head-to-head (2), tail-to-tail (2) and tandem (1), suggesting that there may be several kinds of mechanisms involved. Tandem duplications, for example, might result from long-tract conversion rather than crossing-over. The duplicates nearly always contained different numbers of *AVR* genes (Figs. 11e, 11h and III, Fig. 2f), suggesting that deletion or gain of gene copies is frequently associated with locus duplication. Furthermore, three fibers showed a duplicated *AVR*-cluster connected by a single *AVD*. Another three fibers suggested a single *AVR* cluster surrounded by two *AVD*s, one on both sides of the cluster. However, in these cases it was not possible to rule out the possibility that one of the putative *AVD*s was actually an *AVR* gene. Altogether, 11 locus duplications were thus seen among the 800 fibers examined in this study, suggesting a duplication frequency of 1.4%. However, there is an evident source of error with respect to the duplications: they may be artificially formed by intertwining of two DNA fibers. Regardless of this possibility, some duplications, especially those showing a single *AVD* connecting two *AVR* clusters, are probably authentic.

6.2.3 Cellular basis of recombination frequency

The genomic DNAs for fiber-FISH were prepared from white blood cell preparations, and the cell line HD11 represents macrophage-type cells (Beug et al. 1979). As described, evidence of frequent recombination was observed in these preparations. In contrast, genomic libraries (the two libraries screened in this study, I, Fig. 2a, as well as the library used by Keinänen et al. 1988 and Keinänen et al. 1994) showed no evidence of somatic recombination. This discrepancy can be explained by cell-type differences. The libraries were prepared from solid tissues (liver or muscle), for which excessive recombination frequencies have not been observed. For example, Högstrand and Böhme (1994) found elevated levels of gene conversion between the MHC II-class genes $E\beta^d$ and $A\beta^k$ genes in mouse sperm cells, whereas the same genes in liver cells showed no conversion events. In contrast, Reynaud et al. (1987) showed that exceptionally high gene conversion frequencies occur in the forming B-cells in the bursa of Fabricius, producing the immunoglobulin light chain diversity in the chicken. The observed copy-number fluctuation in the *AVD* gene family in white blood cells suggests that the gene family is highly prone to recombinatorial events in cellular environments where recombination-associated factors are available.

6.3 Characterization of the AVR proteins (IV)

The existence of the avidin-related genes has been puzzling, since their biological function, if any, has remained a mystery. The existence of AVR proteins in the chicken could not be studied because of the lack of antibodies that would differentiate the putative AVRs from avidin. However, since the AVR genes show all the characteristics of functional genes, it seemed very likely that the proteins are expressed under at least some circumstances. Furthermore, amino acid sequence analyses and structure predictions suggested that the putative AVR proteins are functional and highly similar to avidin (Keinänen et al. 1994, Keinänen 1994, Wallén 1996 and III).

In this study, we decided to produce recombinant AVR proteins to shed more light on their structure, function and putative roles. Recombinant AVRs might also be useful for biotechnical applications. Furthermore, we wanted to reveal in detail the structural basis for the differences between avidin and the AVRs by sequence analysis and computer modeling.

6.3.1 Properties of the AVRs

The AVRs are remarkably similar to avidin in amino acid sequence (Keinänen et al. 1994, Table 5 and IV, Fig. 1). Both sequence analysis and molecular modeling as well as biochemical characterization of the recombinant AVRs showed that they differ from avidin in several respects, such as isoelectric point and glycosylation pattern. The isoelectric points of the AVRs are generally lower than that of avidin, and the AVRs are more heavily glycosylated. That the tryptophan (W70, 97 and 110), tyrosine (Y33) and phenylalanine (F72 and 79) residues important for biotin binding (Livnah et al. 1993) are conserved in all AVR sequences (Fig. 14 and IV, Fig. 1) suggested that the AVR proteins are able to bind biotin. Indeed, the recombinant AVR proteins produced in this study exhibited biotin binding. However, the binding was reversible in AVR2, a property highly interesting in the view of protein structure-function

TABLE 5 Amino acid sequence identity (%) between *AVD* and the putative *AVR* proteins.

	AVD	AVR1	AVR2	AVR3	AVR4	AVR5	AVR6
AVR1	77	100					
AVR2	74	95	100				
AVR3	77	92	91	100			
AVR4	81	85	85	91	100		
AVR5	81	85	85	91	100	100	
AVR6	76	96	94	95	86	86	100
AVR7	77	95	93	94	85	85	99

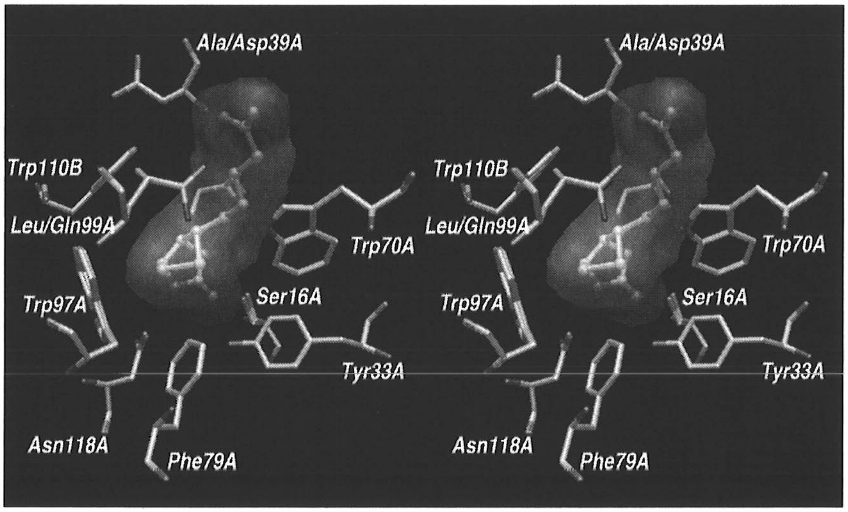

FIGURE 14 Computer stereo model of the biotin-binding pocket (with biotin bound) of AVD and AVRs. (Courtesy of Olli Pentikäinen, unpublished.)

relationships. This property probably results from the conversion of Lys111 to Ile, as this substitution is likely to alter the shape of the hydrophobic binding pocket (IV, Fig. 3). AVR1 also showed partially reversible binding, but the structural basis for this is currently unknown.

Another structurally interesting feature in all AVRs (except for AVR2) is the existence of a third cysteine in addition to the two disulfide bridge-forming cysteins found also in avidin. The extra cysteines could form novel disulfide bridges to link different subunits or even different tetramers to form larger units or aggregates of AVRs. Aggregates of AVRs may offer scaffolds for embryonic development, similarly to the sea urchin fibropellins that are thought to form dimers or higher order oligomeric structures to promote protein-protein interactions during embryogenesis (Bisgrove et al. 1991, Bisgrove & Raff 1993). The recent finding that avidin is involved in cell differentiation (Zerega et al. 2001) supports such possible roles for the AVRs. Our non-reducing SDS-PAGE results as well as molecular modeling showed that inter-monomer cysteine bridges constituting dimeric patterns can indeed occur.

Investigation of the subunit interface regions revealed numerous differences in the AVRs as compared to avidin (IV, Table 1). The extremely high stability of the AVR tetramers is therefore surprising. The amino acid substitutions may be complementary; while an amino acid change in one subunit decreases the interface affinity, the coincident mutation in the other subunit may restore it. Alternatively, a water molecule can act to fill in the gap produced by the conversion of bulky charged or polar residues into smaller ones, and the hydrogen bond therefore forms via a water molecule (Janin 1999).

In conclusion, the recombinant AVRs are functional proteins that show properties similar to, but in some respects distinct from, avidin. The AVRs can be expected to have anti-inflammatory functions similar to avidin in the chicken, due to their considerably well-conserved biotin-binding capacities. It is not yet known if the AVRs bind to bacteria, similarly to avidin (Korpela et al. 1984), but the differences in glycosylation patterns and other biochemical properties might play a role in bacterial attachment. Therefore, it is possible that the different AVRs broaden the range of host defense. The high stability of the AVRs and their resistance against proteolytic enzymes may be advantageous for fulfilling the anti-inflammatory functions. The AVRs may also provide advantages over avidin and streptavidin in biochemical applications. Due to its reversible biotin binding, for example, AVR2 could be used for affinity purification under mild elution conditions.

7 CONCLUSIONS

The current study resulted in detailed characterization of the chicken avidin gene family. The genomic organization of the gene family was revealed and the genes were shown to undergo frequent rearrangements by gene conversion and unequal crossing-over. Furthermore, a preliminary characterization of recombinant avidin-related proteins was performed.

The gene copy-number in the avidin gene family was found to differ between individuals, implying that germ-line recombination occurs leading to segregation of different combinations of the genes to the progeny. Furthermore, fiber-FISH experiments were used to evidence somatic gene copy-number fluctuation. These experiments were, however, limited to white blood cells, which exhibit extraordinarily high levels of somatic recombination. Nevertheless, the results show that the *AVD* gene family is indeed highly prone to recombination. The high degree of homology between the genes and the telomere-proximal location of the locus probably render the gene family amenable to recombination. The significance of the frequent recombination remains unclear. It may be that the *AVD* and *AVR* genes in white blood cells represent merely "passengers", recombining because they are prone to do so whenever the recombination machinery is functional. For example, the hypermutation that diversifies the immunoglobulin variable genes in maturating B cells has been suggested to induce mutations in also other genes that are transcribed in the same cells. Evidence for this suggestion is available for only one gene thus far. The avidin gene family may represent another example, and may thus strengthen our understanding of the somatic hypermutation mechanism.

On the other hand, *AVD* and some *AVR* genes are expressed in the macrophage-type HD11 cells, as well as in several tissues of the chicken in response to inflammation. As *AVD* is thought to function as an anti-inflammatory agent, it may be that the somatic recombination plays a role in the chicken immune system, forming a variety of anti-inflammatory molecules of selective advantage. Gene conversion might act a role in such diversification.

This suggestion is supported by the finding that the AVR proteins are functional in biotin binding, but show biochemical and structural properties slightly different from avidin.

Previous studies have shown repeat number fluctuation for large, mainly RNA-encoding gene families containing hundreds of copies, or for noncoding sequences such as minisatellites. The current study has shown frequent fluctuation for a relatively small protein-encoding gene family. The results showed both copy gain and loss, suggesting that unequal crossing-over is the main recombination mechanism within this gene family. The occurrence of frequent gene conversion as well supports the view that gene conversion and crossing-over are coupled processes. These data are therefore valuable for understanding genetic recombination events.

The avidin gene family provides an excellent model system to study the molecular basis of recombination in detail. A cell line widely used in recombination studies is the chicken DT40 cell line, from which several mutant and knockout lines deficient in various recombination factors have been produced (Sonoda et al. 2001). By screening the different DT40 mutants for recombination in the avidin gene family, the factors responsible for the phenomenon can be deciphered.

This study also gives insight into general issues such as the stability of genomes. As the human and several other genome projects are reaching or have already reached their finals, knowledge about the frequency of genomic rearrangements, with special emphasis on protein-coding sequences, is needed to evaluate the integrity of the genome maps, as well as proteome maps in the future.

Acknowledgements

This study was carried out at the Department of Biological and Environmental Science, University of Jyväskylä. The study was financially supported by the University of Jyväskylä, The Finnish Cultural Fund and the Academy of Finland.

I wish to thank my supervisor, Prof. Markku Kulomaa, for his support and encouragement through all these years. I appreciate very much the kindness you have always showed me. I also thank Kuku for giving me the opportunity to work very independently, giving me totally free hands with respect to lab work as well as my other duties (i.e. teaching and lab and research management).

I also wish to thank the official referees of this thesis, Prof. Olli Lassila, M.D., and Prof. Howy Jacobs, Ph.D., for their constructive comments on the thesis manuscript.

My dearest (and weirdest) colleaques have probably been the most important factors keeping me going on through numerous difficult times. The frendship of Marja Tiirola, Piia Karisola, Eija Kola and Hong Wang has been invaluable -thanks, girls! Special thanks goes to Eija also for her contribution in the first part of this study. The company of the boys, Olli Laitinen, Ari Marttila and Henri Nordlund, has been a lot of fun. Vesa Hytönen deserves a special acknowledgement for his major contribution in the work described in manuscript IV. During the earlier years, Kari Airenne and Varpu Marjomäki were key persons of our group, and they have been very helpful also after relocating to other groups.

I am deeply indebted to the person who keeps the whole lab going, and without whom we would be totally lost: Irene Helkala. Ine has taught me almost everything about lab work, and most nobaly, she has patiently done it over and over again. Thank you, Ine!

Furthermore, I wish to thank the co-authors of publication I, especially Dagmar Ewald and Julio Masabanda for their contributions in screening the gridded library and performing the metaphase FISH, respectively. Special thanks also go to Alessandro Grapputo and Maxine Iversen; Ale for performing the numerous analyses for manuscript II and Maxine for language revisions. Katri Kärkkäinen also gave numerous helpful comments when preparing manuscript II. Nina Horelli-Kuitunen, Ph.D. (now at MedixDiacor Laboratoriopalvelut Oy, Espoo), taught me the fiber-FISH technique. Without her kind help, I obviously wouldn't have been able to produce manuscript III - and probably wouldn't have finished my thesis, either.

My warmest thanks also goes to all the other, both former as well as current, "molecular recognition" people. Allu and Pirjo have helped me a lot whenever Ine has not been around, and Marjatta and Anna-Liisa have made my life a lot easier in many ways. In addition, there are numerous people I've been involved with in highly interesting, more or less scientific, discussions and/or sports activities (such as badminton, aerobics, horse riding and downhill skiing -the skiing trip to Lappland that we did with Maija, Sanna and Eve was especially memorable!).

I also owe my sincerest thanks to my mother, Aino. She has been an ever sacrificing grandmom, helping me out whenever I needed her -not just babysitting, but also supporting me in numerous ways. I love you, mom! Pekka, my father, always taught me to believe in myself. I sure needed that belief many times, so thank you, dad.

Finally, I thank my dearest family: Pete, Ranja and Ripsa. Pete is largely responsible for my career as a scientist. He originally taught me scientific thinking --and has recently had to cope with it. Thank you for standing all my nervous brakdowns! Ranja and Ripsa are are my pride and joy. The unconditional love that I get from them is the most beautiful thing I could ever imagine. Pete, Ranja and Ripsa --this work is dedicated to you.

YHTEENVETO (Résumé in Finnish)

Kanan avidiinigeeniperhe. Organisaatio, evoluutio ja tiheä rekombinaatio.

Avidiini on biotiinia sitova proteiini, jota luonnossa esiintyy kananmunassa sekä useissa kanan kudoksissa tulehdustilanteissa. Voimakkaan biotiininsitomiskykynsä vuoksi avidiinia on jo kauan käytetty biokemiallisena työkaluna; avidiinin avulla voidaan esimerkiksi detektoida biotiinilla leimattuja komponentteja eri sovelluksissa. Avidiinia koodaavan geenin (*AVD*) lisäksi kanalla on useita avidiinin kaltaisia geenejä (avidin-related genes, *AVR*), jotka ovat hyvin samankaltaisia sekä keskenään että avidiiniin nähden. Viisi avidiinin kaltaista geeniä (*AVR1*-*AVR5*) oli löydetty aiemmin, mutta geenejä epäiltiin olevan vielä lisää. Avidiinin kaltaisten geenien oletettiin sijaitsevan lähellä toisiaan, mutta itse avidiinigeenin sijainnista ei ollut tietoa.

Tässä tutkimuksessa selvitettiin avidiinigeeniperheen koostumus, sijainti ja organisaatio. Kaikkien geeniperheen jäsenten havaittiin sijaitsevan yhdessä paikassa (lokuksessa). Lokus sijaitsi kanan sukupuolikromosomi Z:ssa lähellä kromosomin telomeeriä, kohdassa Zq21. Kahdesta genomisesta kirjastosta, joita tutkimuksessa seulottiin, löydettiin eri yhdistelmä *AVR*-geenejä. Kaksi uutta *AVR*-geeniä (*AVR6* ja *AVR7*) saatiin kloonattua ja sekvensoitua. Geenien keskinäinen järjestys ja orientaatio selvitettiin restriktioentsyymikartoituksilla ja PCR-kokeilla. Avidiinigeenin havaittiin sijaitsevan geenijonon toisessa päässä n. 9000 emäsparin etäisyydellä lähimmästä *AVR*-geenistä. *AVR*-geenit puolestaan sijaitsivat lähellä toisiaan (2500-2800 emäsparin välein), järjestäytyneinä yhtä geeniä lukuunottamatta samansuuntaisesti.

Avidiini- ja *AVR*-geenien nukleotidisekvenssejä tarkasteltiin geenien evoluutiotaustan selvittämiseksi. Sekvensseistä löydettiin useita mielenkiintoisia piirteitä. Esimerkiksi nukleotidimuutoksissa transitioita löytyi huomattavan paljon transversioihin nähden. Edelleen sekvensseistä löytyi vahvoja viitteitä siitä, että geenien välillä tapahtuu geenikonversiota. Tietokoneohjelmien avulla löydettiinkin useita mahdollisia konversiojaksoja geenien sekvensseistä. Erityisen mielenkiintoinen oli havainto, että konversio näyttäisi olevan suunnattua, eli sitä tapahtuu lähinnä avidiinigeenistä *AVR*-geeneihin päin. Geenikonversio näyttäisi pyrkivän säilyttämään geenien toisen intronin muuttumattomana, kun taas toinen eksoni on erityisen altis mutaatioille. Vastaavia piirteitä on aiemmin havaittu varsin harvoista geeneistä.

Eri geeniyhdistelmien löytyminen eri genomisista kirjastoista antoi aihetta olettaa että *AVD*- ja *AVR*-geenien lukumäärä saattaisi vaihdella laajemmaltikin eri yksilöiden välillä. Hypoteesia testattiin fiber-FISH-menetelmää käyttäen. Tulokset osoittivat että geeniluku todellakin vaihtelee yksilöiden välillä, ja että tämän lisäksi vaihtelua esiintyy jopa saman yksilön eri solujen välillä. Vaihtelu voidaan selittää epätasaisella rekombinaatiolla (unequal crossing-over), jota

tässä tapauksessa voi siis tapahtua sekä meioosin yhteydessä sukusolujen muodostuessa, että mitoosin yhteydessä somaattisissa soluissa. Se, että sekä geenikonversiota että rekombinaatiota tapahtuu samoissa geeneissä ilmeisen usein, tukee mallia jonka mukaan nämä kaksi prosessia ovat yhden geneettisen tapahtuman vaihtoehtoisia lopputuloksia. Esimerkiksi DNA:n kaksoisjuosteen katkosta korjaava mekanismi voi todennäköisesti tuottaa lopputulokseksi joko konversion tai rekombinaation. Edelleen, on ehdotettu että immunoglobuliinigeenien hypermutaatiomekanismi saattaisi vaikuttaa myös muihin geeneihin samassa solussa. Tälle mallille on löydetty tukea ainoastaan yhdessä tapauksessa tähän mennessä. Avidiinigeeniperhe saattaisi olla toinen esimerkkitapaus, jossa immunoglobuliinigeenien hyperkonversio saa "sivutuotteena" aikaan muidenkin geenien lisääntynyttä konversiota tai rekombinaatiota. Näin ollen tämän tutkimuksen tulokset tuovat lisää tietoa geneettisen rekombinaation mekanismeista ja yleisyydestä, ja antavat aihetta olettaa että aitotumallisten eliöiden genomit ovat muuntelevaisempia kuin tähän saakka on ajateltu.

Lopuksi tässä tutkimuksessa tuotettiin rekombinanttisia avidiininkaltaisia proteiineja (AVR) ja selvitettiin niiden ominaisuuksia. AVR-proteiinien todettiin olevan toiminnallisesti ja rakenteellisesti hyvin samanlaisia kuin avidiini, joskin eroja löytyi proteiinien glykosylaatiossa, isoelektrisen pisteen arvoissa sekä mahdollisesti rikkisiltarakenteissa. Ehkä mielenkiintoisin havainto oli se, että jotkut AVR-proteiinit sitoivat biotiinia reversiibelisti, toisin kuin itse avidiini. Tällä hetkellä ei tiedetä, esiintyykö AVR-proteiineja kanan kudoksissa luonnollisesti. Proteiinien ominaisuudet antavat aihetta olettaa, että ne saattaisivat toimia osana kanan puolustusjärjestelmää. Tulevaisuudessa AVR-proteiinien ominaisuuksia voidaan ehkä käyttää hyväksi kehitettäessä parannuksia olemassaolevaan avidiini-biotiiniteknologiaan.

REFERENCES

Afanassieff, M., Goto, R. M., Ha, J., Sherman, M. A., Zhong, L., Auffray, C., Coudert, F., Zoorob, R. & Miller, M. M. 2001. At least one class I gene in restriction fragment pattern-Y (*Rfp-Y*), the second MHC gene cluster in the chicken, is transcribed, polymorphic, and shows divergent specialization in antigen binding region. J. Immunol. 166: 3324-3333.

Ahlroth, M. K. 2001. Characterization of the chicken avidin gene family. Licentiate's thesis. Department of Biological and Environmental Science, University of Jyväskylä, Finland.

Airenne, K. J., Oker-Blom, C., Marjomäki, V. S., Bayer, E. A., Wilchek, M. & Kulomaa, M. S. 1997. Production of biologically active recombinant avidin in baculovirus-infected insect cells. Protein Expr. Purif. 9: 100-108.

Albig, W. & Doenecke, D. 1997. The human histone gene cluster at the D6S105 locus. Hum. Genet. 101: 284-294.

Argaraña, C. E., Kuntz, I. D., Birken, S., Axel, R. & Cantor, C. R. 1986. Molecular cloning and nucleotide sequence of the streptavidin gene. Nucleic Acids Res. 14: 1871-1882.

Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M. & Olsson, C. 2001. Increased transcription levels induce higher mutation rates in a hypermutating cell line. J. Immunol. 166: 5051-5057.

Baniahmad, A., Muller, M., Steiner, C. & Renkawitz, R. 1987. Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. EMBO J. 6(8): 2297-2303.

Bayer, E. A., Kulik, T., Adar, R. & Wilchek, M. 1995. Close similarity among streptavidin-like, biotin-binding proteins from Streptomyces. Biochim. Biophys. Acta 1263: 60-66.

Beug, H., von Kirchbach, A., Döderlein, G., Conscience, J.-F. & Graf, T. 1979. Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation. Cell 18: 375-390.

Bisgrove, B. W., Andrews, M. E. & Raff, R. A. 1991. Fibropellins, products of an EGF repeat-containing gene, form a unique extracellular matrix structure that surrounds the sea urchin embryo. Dev. Biol. 146: 89-99.

Bisgrove, B. W. & Raff, R. A. 1993. The SpEGF III gene encodes a member of the fibropellins: EGF repeat-containing proteins that form the apical lamina of the sea urchin embryo. Dev. Biol. 157: 526-538.

Borde, V., Goldman, A. S. H. & Lichten, M. 2000. Direct coupling between meiotic DNA replication and recombination initiation. Science 290: 806-809.

Botte, V. & Granata, G. 1977. Induction of avidin synthesis by RNA obtained from lizard oviducts. J. Endocrinol. 73: 535-536.

Briles, W. E., Goto, R. M., Auffray, C. & Miller, M. M. 1993. A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. Immunogenetics 37: 408-414.

Brooker, R. J. 1999. Genetics: analysis and principles. Benjamin/Cummings, Menlo Park, CA, USA.

Buitkamp, J., Ewald, D., Schalkwyk, L., Weiher, M., Masabanda, J., Sazanov, A., Lehrach, H. & Fries, R. 1998. Construction and characterization of a gridded chicken cosmid library with four-fold genomic coverage. Anim. Genet. 29: 295-301.

Carlson, L. M., McCormack, W. T., Postema, C. E., Humphries, E. H. & Thompson, C. B. 1990. Templated insertions in the rearranged chicken IgL V gene segment arise by intrachromosomal gene conversion. Genes Dev. 4: 536-547.

Chan, L., Means, A. R. & O'Malley, B. W. 1973. Rates of induction of specific translatable messenger RNAs for ovalbumin and avidin by steroid hormones. Proc. Natl. Acad. Sci. USA 70(6): 1870-1874.

Ching, A. K-K., Li, P-S., Chan, W-Y., Ma, C-H., Lee, S. S-T., Lim, P-L. & Chui, Y-L. 2000. Strand bias in Ig somatic hypermutation is determined by signal sequence within the variable region. Int. Immunol. 12(9): 1245-1253.

Critchlow, S. E. & Jackson, S. P. 2001. 1998. DNA end-joining: from yeast to man. Trends Biochem. Sci. 23: 394-398.

Dean, D. M. & Sanders, M. M. 1996. Ten years after: reclassification of steroid-responsive genes. Mol. Endo. 10(12): 1489-1495.

Delany, M. E. 2000. Patterns of ribosomal gene variation in elite commercial chicken pure line populations. Anim. Genet. 31: 110-116.

Delany, M. E. & Krupkin, A. B. 1999. Molecular characterization of ribosomal gene variation within and among NORs segregating in specialized populations of chicken. Genome 42: 60-71.

Dolan, M., Sugarman, B. J., Dodgson, J. B. & Engel, J. D. 1981. Chromosomal arrangement of the chicken beta-type globin genes. Cell 24: 669-677

Dominguez, O., Ruiz, J. F., Lain de Lera, T., Garcia-Diaz, M., Gonzalez, M. A., Kirchhoff, T., Martinez, A. C., Bernad, A. & Blanco, L. 2000. DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. EMBO J. 19: 1731-1742.

Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. Nature 299: 111-117.

Dryer, L. 2000. Evolution of odorant receptors. Bioessays 22: 803-810.

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. & Jasin, M. 1998. Gene conversion tracts from double-strand break repair in mammalian cells. Mol. Cell. Biol. 18: 93-101.

Elo, H. A. 1980. Occurrence of avidin-like biotin-binding capacity in various vertebrate tissues and its induction by tissue injury. Comp. Biochem. Physiol. 67B: 221-224.

Elo, H. A. & Korpela, J. 1984. The occurrence and production of avidin: a new conception of the high-affinity biotin-binding protein. Comp. Biochem. Physiol. 78B(1): 15-20.

Elo, H. A., Kulomaa, M. S. & Tuohimaa, P. J. 1979a. Avidin induction by tissue injury and inflammation in male and female chickens. Comp. Biochem. Physiol. 62B: 237-240.

Elo, H. A., Kulomaa, M. S. & Tuohimaa, P. J. 1979b. Progesterone-independent avidin induction in chick tissues caused by tissue injury and inflammation. Acta Endocrinol. 90: 743-752.

Elo, H. A., Räisänen, S. & Tuohimaa, P. J. 1980. Induction of antimicrobial biotin-binding egg white protein (avidin) in chick tissues in septic *Escherichia coli* infection. Experientia 36: 312-313.

Evans, H. J., Buckland, R. A. & Pardue, M. L. 1974. Location of the genes coding for 18S and 28S ribosomal RNA in the human genome. Chromosoma (Berl.) 48: 405-426.

Florijn, R. J., van de Rijke, F. M., Vrolijk, H., Blonden, L. A., Hofker, M. H., den Dunnen, J. T., Tanke, H. J., van Ommen, G. J. & Raap, A. K. 1996. Exon mapping by fiber-FISH or LR-PCR. Genomics 38: 277-282.

Freitas-Junior, L. H., Bottius, E., Pirrit, L. A., Deitsch, K. W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T. E. & Scherf, A. 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. Nature 407: 1018-1022.

Fryxell, K. J. 1996. The coevolution of gene family trees. Trends Genet. 12: 364-369.

Glusman, G., Yanai, I., Rubin, I. & Lancet, D. 2001. The complete human olfactory subgenome. Genome Res. 11: 685-702.

Goodman, M. F. & Tippin, B. 2000. Sloppier copier DNA polymerases involved in genome repair. Cur. Opin. Genet. Dev. 10: 162-168.

Gope, M. L., Keinänen, R. A., Kristo, P. A., Conneely, O. M., Beattie, W. G., Zaruki-Schulz, T., O'Malley, B. W. & Kulomaa, M. S. 1987. Molecular cloning of the chicken avidin cDNA. Nucleic Acids Res. 15(8): 3595-3606.

Graham, G. J. 1995. Tandem genes and clustered genes. J. Theor. Biol. 175: 71-87.

Grandy, D. K. & Dogson, J. B. 1987. Structure and organization of the chicken H2B histone gene family. Nucleic Acids Res. 15(3): 1063-1080.

Grawunder, U. & Harfst, E. 2001. How to make ends meet in V(D)J recombination. Curr. Opin. Immunol. 13: 186-194.

Green, N. M. 1975. Avidin. Adv. Prot. Chem. 29: 85-133.

Greenberg, A. S., Avila, D., Hughes, M., Hughes, A., McKinney, E. C. & Flajnik, M. F. 1995. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374: 168-173.

Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C. & Gelbart, W. M. 1993. An introduction to genetic analysis. W.H. Freeman and company, New York.

Guttinger, M., Guidi, F., Chinol, M., Reali, E., Veglia, F., Viale, G., Paganelli, G., Corti, A. & Siccardi, A. G. 2000. Adoptive immunotherapy by avidin-driven cytotoxic T lymphocyte-tumor bridging. Cancer Res. 60: 4211-4215.

Haber, J. E. 1998. Meiosis: avoiding inappropriate relationships. Curr. Biol. 8: R832-R835.

Haber, J. E. 1999. DNA recombination: the replication connection. Trends Biochem. Sci. 24: 271-232.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. & Grosveld, F. 1991. Importance of globin gene order for correct developmental expression. Genes Dev. 5: 1387-1394.

Harano, K., Harano, T., Kutlar, F. & Huisman, T. H. 1985. Gamma-globin gene triplication and quadruplication in Japanese newborns. Evidence for a decreased in vivo expression of the 3'-A gamma-globin gene. FEBS Lett. 190: 45-49.

Heiskanen, M., Karhu, R., Hellsten, E., Peltonen, L., Kallioniemi, O. P. & Palotie, A. 1994. High resolution mapping using fluorescence in situ hybridization to extended DNA fibers prepared from agarose-embedded cells. Biotechniques 17: 928-929, 932-933.

Heiskanen, M., Hellsten, E., Kallioniemi, O. P., Mäkelä, T. P., Alitalo, K., Peltonen, L. & Palotie, A. 1995. Visual mapping by fiber-FISH. Genomics 30(1): 31-6.

Heiskanen, M., Kallioniemi, O. & Palotie, A. 1996. Fiber-FISH: experiences and a refined protocol. Genet. Anal. 12(5-6):179-84

Hellsten, E., Vesa, J., Heiskanen, M., Mäkelä, T. P., Järvelä, I., Cowell, J. K., Mead, S., Alitalo, K., Palotie, A. & Peltonen, L. 1995. Identification of YAC clones for human chromosome 1p32 and physical mapping of the infantile neuronal ceroid lipofuscinosis (INCL) locus. Genomics 25(2): 404-12.

Hertz, R. & Sebrell, W. H. 1942. Occurrence of avidin in the oviduct and secretions of the genital tract of several species. Science 96: 257.

Hertz, R., Dhyse, F. G. & Tullner, W. M. W. 1949. Elevation of biotin activity in the serum of estrogen treated chicks; relationship to hormone-induced avidin formation in the oviduct. Endocrinology 45(5): 451-454.

Högstrand, K. & Böhme, J. 1994. A determination of the frequency of gene conversion in unmanipulated mouse sperm. Proc. Natl. Acad. Sci. USA 91: 9921-9925.

Horelli-Kuitunen, N., Aaltonen, J., Yaspo, M. L., Eeva, M., Wessman, M., Peltonen, L. & Palotie, A. 1999. Mapping ESTs by fiber-FISH. Genome Res. 9: 62-71.

Huang, T.-S. & DeLange, R. J. 1971. Egg white avidin. II. Isolation, composition and amino acid sequences of the tryptic peptides. J. Biol. Chem. 246(3): 686-697.

Huberman, T., Eisenberg-Domovich, Y., Gitlin, G., Kulik, T., Bayer, E. A., Wilchek, M. & Livnah, O. 2001. Chicken avidin exhibits pseudo-catalytic properties. Biochemical, structural, and electrostatic consequences. J. Biol. Chem. 276: 32031-32039.

Jacobs, H. & Bross, L. 2001. Towards an understanding of somatic hypermutation. Curr. Opin. Immunol. 13: 208-218.

Janin, J. 1999. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. Structure Fold Des. 7: R277-279.

Jeffreys, A. J. & Harris, S. 1982. Processes of gene duplication. Nature 296: 9-10.

Jeffreys, A. J., Neil, D. L. & Neumann, R. 1998. Repeat instability at human minisatellites arising from meiotic recombination. EMBO J. 17: 4147-4157.

Keinänen, R. A., Laukkanen, M.-L. & Kulomaa, M. S. 1988. Molecular cloning of three structurally related genes for chicken avidin. J. Steroid Biochem. 30(1-6): 17-21.

Keinänen, R. A., Wallén, M. J., Kristo, P. A., Laukkanen, M. O., Toimela, T. A., Helenius, M. A. & Kulomaa, M. S. 1994. Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-5. Eur. J. Biochem. 220: 615-621.

Keinänen, R. A. 1994. Molecular cloning and structure of the chicken avidin cDNA and five avidin-related genes. Ph.D. thesis. Acta Universitatis Tamperensis ser A vol. 411. University of Tampere, Finland.

Kelsoe, G. 1999. V(D)J hypermutation and receptor revision: coloring outside the lines. Curr. Opin. Immunol. 11: 70-75.

Klasing, K. C. 1998. Avian macrophages: regulators of local and systemic immune responses. Poultry Sci. 77: 983-989.

Kong, Q. & Maizels, N. 2001. DNA breaks in hypermutating immunoglobulin genes: evidence for a break-and-repair pathway of somatic hypermutation. Genetics 158: 369-378.

Korenman, S. G. & O'Malley, B. W. 1968. Progesterone action: regulation of avidin biosynthesis by hen oviduct in vivo and in vitro. Endocrinology 83: 11-17.

Korpela, J. 1984. Chicken macrophages synthesize and secrete avidin in culture. Eur. J. Cell Biol. 33: 105-111.

Korpela, J. K., Kulomaa, M. S., Elo, H. A. & Tuohimaa, P. J. 1981. Biotin-binding proteins in eggs of oviparous vertebrates. Experientia 37: 1065-1066.

Korpela, J., Kulomaa, M., Tuohimaa, P. & Vaheri, A. 1982. Induction of avidin in chickens infected with the acute leukemia virus OK 10. Int. J. Cancer 30: 461-464.

Korpela, J., Kulomaa, M., Tuohimaa, P. & Vaheri, A. 1983. Avidin is induced in chicken embryo fibroblasts by viral transformation and cell damage. EMBO J. 2(10): 1715-1719.

Korpela, J., Salonen, E.-M., Kuusela, P., Sarvas, M. & Vaheri, A. 1984. Binding of avidin to bacetria and to the outer membrane porin of Escherichia coli. FEMS Microbiol. Lett. 22: 3-10.

Krystal, M., D'Eustachio, P., Ruddle, F. H. & Arnheim, N. 1981. Human nucleolus organizers on nonhomologous chromosomes can share the same ribosomal gene variants. Proc. Natl. Acad. Sci. USA 78: 5744-5748.

Kulomaa, M. S., Elo, H. A., Niemelä, A. O. & Tuohimaa, P. J. 1981. Similarity of biotin-binding activity and immunoreactivity in chicken oviduct and non-oviduct avidin. Biochim. Biophys. Acta 670: 207-213.

Kunnas, T. A., Joensuu, T. K., Viitala, K. K., Sopanen, P., Tuohimaa, P. & Kulomaa, M. S. 1992. Induction of avidin messenger ribonucleic acid in the chick oviduct by progesterone and other steroids. Endocrinology 130(6): 3421-3426.

Kunnas, T. A., Wallen, M. J. & Kulomaa, M. S. 1993. Induction of chicken avidin and related mRNAs after bacterial infection. Biochim. Biophys. Acta 1216: 441-445.

Kuo, B. A., Gonzalez, I. L., Gillespie, D. A. & Sylvester, J. E. 1996. Human ribosomal RNA variants from a single individual and their expression in different tissues. Nucleic Acids Res. 24: 4817-4824.

Kwon, D., Chwae, Y. J., Choi, I. H., Park, J. H., Kim, S. J. & Kim, J. 2000. Diversity of the p70 killer cell inhibitory receptor (KIR3DL) family members in a single individual. Mol. Cells 10: 54-60.

Laitinen, O. H., Airenne, K. J., Marttila, A. T., Kulik, T., Porkka, E., Bayer, E. A., Wilchek, M. & Kulomaa, M. S. 1999. Mutation of a critical tryptophan to lysine in avidin or streptavidin may explain why sea urchin fibropellin adopts an avidin-like domain. FEBS Lett. 461: 52-8.

Leach, D. R. F. 1994. Long DNA palindomes, cruciform structures, genetic instability and secondary structure repair. BioEssays 16(12): 893-900.

Lewin, B. 2000. Genes VII. Oxford University Press Inc., New York.

Li, W.-H. 1997. Molecular evolution. Sinauer Associates, Inc., Sunderland, Massachusetts.

Li, W. H., Gu, Z., Wang, H. & Nekrutenko, A. 2001. Evolutionary analyses of the human genome. Nature 409: 847-849.

Liao, D. 2000. Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. J. Mol. Evol. 51: 305-317.

Livnah, O., Bayer, E. A., Wilchek, M. & Sussman, J. L. 1993. Three-dimensional structures of avidin and the avidin-biotin complex. Proc. Natl. Acad. Sci. USA 90: 5076-5080.

Long, E. O. & Dawid, I. B. 1980. Repeated genes in eukaryotes. Annu. Rev. Biochem. 49: 727-764.

Lopez, B. S., Corteggiani, E., Bertrand-Mercat, P. & Coppey, J. 1992. Directional recombination is initiated at a double strand break in human nuclear extracts. Nucleic Acids Res. 20(3): 501-506.

Maeda, N., McEvoy, S. M., Harris, H. F., Huisman, T. H. & Smithies, O. 1986. Polymorphisms in the human haptoglobin gene cluster: chromosomes with multiple haptoglobin-related (Hpr) genes. Proc. Natl. Acad. Sci. USA 83: 7395-7399.

Maizels, N. 1987. Diversity achieved by diverse mechanisms: gene conversion in developing B cells of the chicken. Cell 48: 359-360.

Marians, K. J. 2000. Replication and recombination intersect. Curr. Opin. Genet. Dev. 10: 151-156.

Martinsohn, J. T., Sousa, A. B., Guethlein, L. A. & Howard, J. C. 1999. The gene conversion hypothesis of MHC evolution: a review [published erratum appears in Immunogenetics 2000 Jun; 51(7): 613]. Immunogenetics 50: 168-200.

Marttila, A. T., Airenne, K. J., Laitinen, O. H., Kulik, T., Bayer, E. A., Wilchek, M. & Kulomaa, M. S. 1998. Engineering of chicken avidin: a progressive series of reduced charge mutants. FEBS Lett. 441: 313-317.

McHeyzer-Williams, M. G. & Davis, M. M. 1995. Antigen-specific development of primary and memory T cells in vivo. Science 268: 106-111.

Meslar, H. W., Camper, S. A. & White, H. B. 1978. Biotin-binding protein from egg yolk. A protein distinct from egg white avidin. J. Biol. Chem. 253(19): 6979-6982.

Miller, M. M., Goto, R. M., Taylor, R. L., Jr., Zoorob, R., Auffray, C., Briles, R. W., Briles, W. E. & Bloom, S. E. 1996. Assignment of *Rfp-Y* to the chicken major histocompatibility complex/NOR microchromosome and evidence for high-frequency recombination associated with the nucleolar organizer region. Proc. Natl. Acad. Sci. USA 93: 3958-3962.

Nadeau, J. H. & Sankoff, D. 1997. Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics 147: 1259-1266.

Nakayama, T., Takechi, S. & Takami, Y. 1993. The chicken histone gene family. Comp. Biochem. Physiol. 104B(4): 635-639.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Neuberger, M. S. & Milstein, C. 1995. Somatic hypermutation. Curr. Opin. Immunol. 7: 248-254.

Niemelä, A. O. 1985. Regulation of avidin accumulation by prostaglandins in chick oviduct culture. J. Steroid Biochem. 24: 709-713.

Ninio, J. 2000. Illusory defects and mismatches: why must DNA repair always be (slightly) error prone? Bioessays 22: 396-401.

Nordback, I., Joensuu, T. & Tuohimaa, P. 1982. Effects of glucocorticoids and disodium cromoglycate on avidin production in chick tissues. J. Endocrinol. 92: 283-291.

O'Malley, B. W. 1967. In vitro hormonal induction of a specific protein (avidin) in the chick oviduct. Biochemistry 6: 2546-2551.

O'Malley, B. W. & McGuire, W. L. 1969. Progesterone-induced synthesis of a new species of nuclear RNA. Endocrinology 84(1): 63-68.

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

Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R. S., Klein, U., Kuppers, R., Rajewsky, K. & Dalla-Favera, R. 1998. BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. Proc. Natl. Acad. Sci. USA 95: 11816-11821.

Perry, J. & Ashworth, A. 1999. Evolutionary rate of a gene affected by chromosomal position. Curr. Biol. 9: 987-989.

Purandre, S. M. & Patel, P. I. 1997. Recombination hot spots and human disease. Genome Res. 7(8): 773-786.

Ramsden, D. A., van Gent, D. C. & Gellert, M. 1997. Specificity in V(D)J recombination: new lessons from biochemistry and genetics. Curr. Opin. Immunol. 9: 114-120.

Reitman, M., Grasso, J. A., Blumenthal, R. & Lewit, P. 1993. Primary sequence, evolution, and repetitive elements of the Gallus gallus (chicken) B-globin cluster. Genomics 18: 616-626.

Reynaud, C. A., Anquez, V., Grimal, H. & Weill, J. C. 1987. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell 48: 379-388.

Roitt, I. M. 1997. Essential Immunology. Blackwell Science, Oxford, UK.

Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J. L., LeMeur, M., Bregegegre, F., Gannon, F., LePennec, J. P., Chambon, P. & Kourilsky, P. 1979. The ovalbumin gene region: common features in the organisation of three genes expressed in chicken oviduct under hormonal control. Nature 279: 125-132.

Ryan, A. K., Schrader, T. J., Wright, R. B., Buchanan, L. & Deeley, R. G. 1994. Characterization of protein interactions with positive and negative elements regulating the apoVLDLII gene. DNA Cell Biol. 13(10): 987-999.

Sanzo, M., Stevens, B., Tsai, M.-J. & O'Malley, B. W. 1984. Isolation of a protein fraction that binds preferentially to chicken middle repetitive DNA. Biochemistry 23: 6491-6498.

Schetters, H. 1999. Avidin and streptavidin in clinical diagnostics. Biomol. Eng. 16: 73-78.

Schiaffella, E., Sehgal, D., Anderson, A. O. & Mage, R. G. 1999. Gene conversion and hypermutation during diversification of VH sequences in developing splenic germinal centers of immunized rabbits. J. Immunol. 162: 3984-3995.

Shen, H. M., Peters, A., Baron, B., Zhu, X. & Storb, U. 1998. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science 280: 1750-1752.

Shen, H. M., Michael, N., Kim, N. & Storb, U. 2000. The TATA binding protein, c-Myc and survivin genes are not somatically hypermutated, while Ig and BCL6 genes are hypermutated in human memory B cells. Int. Immunol. 12: 1085-1093.

Shilling, H. G., Lienert-Weidenbach, K., Valiante, N. M., Uhrberg, M. & Parham, P. 1998. Evidence for recombination as a mechanism for KIR diversification. Immunogenetics 48: 413-416.

Silva, R. & Burch, J. B. E. 1989. Evidence that chicken CR1 elements represent a novel family of retroposons. Mol. Cell. Biol. 9(8): 3563-3566.

Sonoda, E., Takata, M., Yamashita, Y. M., Morrison, C. & Takeda, S. 2001. Homologous recombination in vertebrate cells. Proc. Natl. Acad. Sci. USA 98(15): 8388-8394.

Stevens, L. 1991. Egg white proteins. Comp. Biochem. Physiol. [B] 100: 1-9.

Stevens, L.1996. Avian biochemistry and molecular biology. Cambridge University Press, Cambridge.
Storb, U. 1996. The molecular basis of somatic hypermutation of immunoglobulin genes. Curr. Opin. Immunol. 8: 206-214.
Strachan, T. & Read, A. P. 1999. Human Molecular Genetics. Bios Scientific Publishers Ltd, Oxford, UK.
Stumph, W. E., Hodgson, C. P., Tsai, M.-J. & O'Malley, B. W. 1984. Genomic structure and possible retroviral origin of the chicken CR1 repetitive DNA sequence family. Proc. Natl. Acad. Sci. USA 81: 6667-6671.
Su, M. H. & Delany, M. E. 1998. Ribosomal RNA gene copy number and nucleolar-size polymorphisms within and among chicken lines selected for enhanced growth. Poult Sci 77: 1748-1754.
Suto, Y., Ishikawa, Y., Kasahara, M., Kasai, F., Yabe, T., Akaza, T. & Juji, T. 1998. Gene arrangement of the killer cell inhibitory receptor family on human chromosome 19q13.4 detected by fiber-FISH. Immunogenetics 48: 235-241.
Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. 1983. The double-strand-break repair model for recombination. Cell 33: 25-35.
Tian, M. & Alt, F. W., 2000. Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro. J. Biol. Chem. 275: 24163-24172.
Tissier, A., McDonald, J. P., Frank, E. G. & Woodgate, R. 2000. Poliota, a remarkably error-prone human DNA polymerase. Genes Dev. 14: 1642-1650.
Tuohimaa, P., Joensuu, T., Isola, J., Keinänen, R., Kunnas, T., Niemelä, A., Pekki, A., Wallén, M., Ylikomi, T. & Kulomaa, M. 1989. Development of progestin-specific response in the chicken oviduct. Int. J. Dev. Biol. 33: 125-134.
Uhrberg, M., Valiante, N. M., Young, N. T., Lanier, L. L., Phillips, J. H. & Parham, P. 2001. The repertoire of killer cell Ig-like receptor and CD94:NKG2A receptors in T cells: clones sharing identical alpha beta TCR rearrangement express highly diverse killer cell Ig-like receptor patterns. J. Immunol. 166: 3923-3932.
Wagner, A. 1998. The fate of duplicated genes: loss or new function? BioEssays 20: 785-788.
Walker, L., Kulomaa, M. S., Bebok, Z., Parker, W. B., Allan, P., Logan, J., Huang, Z., Reynolds, R. C., King, S. & Sorscher, E. J. 1996. Development of drug targeting based on recombinant expression of the chicken avidin gene. J. Drug Targeting 4: 41-49.
Wallén, M. 1996. The chicken avidin gene family. Ph.D. thesis. Acta Universitatis Tamperensis ser A vol. 515, University of Tampere, Finland.
Wallén, M. J., Laukkanen, M. O. & Kulomaa, M. S. 1995. Cloning and sequencing of the chicken egg-white avidin-encoding gene and its relationship with the avidin-related genes Avr1-Avr5. Gene 161: 205-209.
Wallén, M. J., Keinänen, R. A. & Kulomaa, M. S. 1996. Two chicken repeat one (CR1) elements lacking a silencer-like region upstream of the chicken

avidin-related genes Avr4 and Avr5. Biochim. Biophys. Acta 1308: 193-196.

Warburton, D., Atwood, K. C. & Henderson, A. S. 1976. Variation in the number of genes for rRNA among human acrocentric chromosomes: correlation with frequency of satellite association. Cytogenet. Cell. Genet. 17: 221-230.

Wellauer, P. K. & Dawid, I. B. 1979. Isolation and sequence organization of human ribosomal DNA. J. Mol. Biol. 128: 289-303.

Wheeler, C. J., Maloney, D., Fogel, S. & Goodenow, R. S. 1990. Microconversion between murine H-2 genes integrated into yeast. Nature 347: 192-194.

Wilbur, D. S., Pathare, P. M., Hamlin, D. K., Stayton, P. S., To, R., Klumb, L. A., Buhler, K. R. & Vessella, R. L. 1999. Development of new biotin/streptavidin reagents for pretargeting. Biomol. Eng. 16: 113-118.

Wilchek, M. & Bayer, E. A. 1988. The avidin-biotin complex in bioanalytical applications. Anal. Biochem. 171: 1-32.

Wilchek, M. & Bayer, E. A. 1990. Introduction to avidin-biotin technology. Methods Enzymol. 184: 5-13.

Wilchek, M. & Bayer, E. A. 1999. Foreword and introduction to the book (strept)avidin-biotin system. Biomol. Eng. 16: 1-4.

Wilson, M. J., Torkar, M., Haude, A., Milne, S., Jones, T., Sheer, D., Beck, S. & Trowsdale, J. 2000. Plasticity in the organization and sequences of human KIR/ILT gene families. Proc. Natl. Acad. Sci. USA 97: 4778-4783.

Zerega, B., Camardella, L., Cermelli, S., Sala, R., Cancedda, R. & Descalzi Cancedda, F. 2001. Avidin expression during chick chondrocyte and myoblast development in vitro and in vivo: regulation of cell proliferation. J. Cell Sci. 114: 1473-1482.

Zheng, B., Xue, W. & Kelsoe, G. 1994. Locus-specific somatic hypermutation in germinal centre T cells. Nature 372: 556-559.

Zozulya, S., Echeverri, F. & Nguyen, T. 2001. The human olfactory receptor repertoire. Genome Biol 2, (6): research0018.1-0018.12. (http://genomebiology.com/2001/2/6/research/0018.1).

ORIGINAL PAPERS

I

Characterization and chromosomal localization of the chicken avidin gene family

by

Ahlroth, M. K., Kola, E. H., Ewald, D., Masabanda, J., Sazanov, A., Fries, R. & Kulomaa, M. S.

Anim. Genet. 31(6): 367-375, 2000

https://doi.org/10.1046/j.1365-2052.2000.00681.x

II

Sequence features and evolutionary mechanisms in the chicken avidin gene family

by

Ahlroth, M. K., Grapputo, A., Laitinen, O. H. & Kulomaa, M. S.

Biochem. Biophys. Res. Comm. 285: 734-741, 2001

https://doi.org/10.1006/bbrc.2001.5163

Biochemical and Biophysical Research Communications **285**, 734–741 (2001)
doi:10.1006/bbrc.2001.5163, available online at http://www.idealibrary.com on IDEAL®

Sequence Features and Evolutionary Mechanisms in the Chicken Avidin Gene Family

Mervi K. Ahlroth,[1] Alessandro Grapputo, Olli H. Laitinen, and Markku S. Kulomaa
Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

Received June 13, 2001

The chicken avidin gene family comprises the avidin gene (*avd*) and several homologous avidin-related genes (*avrs*). The sequences of the *avr* genes are nearly identical to each other but exhibit nonrandomly distributed, frequently nonsynonymous nucleotide substitutions compared to *avd*. In this study, we determined the genetic distances and the phylogeny of the *avd* and *avr* genes and found differences between different exons and introns. Our results suggest the involvement of biased gene conversion in the evolution of the genes. Furthermore, one of the genes was identified as a putative fusion gene. The occurrence of both gene conversion and recombination supports the models suggesting a common initiation mechanism for conversion and crossing-over. The existence of avidin-related proteins (AVRs) is currently unknown, but the putative AVRs are expected to bind biotin similarly to avidin. However, the observed sequence differences may affect the stability and glycosylation patterns of the putative AVR proteins. © 2001 Academic Press

Key Words: **avidin; biotin binding; gene family; evolution; gene conversion; conversion bias; recombination; crossing-over; inversion.**

The chicken avidin has long been used as a biochemical tool because of its strong binding to biotin (1, 2). Avidin is produced in the oviduct of egg-laying hens (3) under the influence of the progesterone hormone (4–6 and references therein). The protein is secreted into the developing egg, where it constitutes about 0.05% of egg white proteins (7, 8). Avidin is thought to act as a defense protein protecting the developing chick embryo against bacterial infection (6 and references therein, 8). It is believed to exert its function by depriving the bacteria of biotin. Avidin may also be involved in acute phase defense in adult chickens, since it is induced by inflammation in several tissues of both male and female chickens (9, 10).

Avidin has been found in the egg white of several avian species (3) as well as in the egg jelly of frogs (3) and in the oviducts of lizards and frogs (3, 11). Furthermore, fibropellins from sea urchins contain a C-terminal domain that shares considerable homology with avidin (12). In contrast, bacterial streptavidins produced by *Streptomyces avidinii* and *S. venezuelae*, although functionally identical and highly similar in their 3D-fold to avidin, are quite different in primary sequence (13–15). Green (16) speculated that since avidin/streptavidin-like proteins have not been found in other microorganisms, streptavidin may be a product of horizontal transfer. Alternatively, avidin and streptavidin may have diverged from an ancient lineage. Avidin has not been found in any mammalian species (3, 17).

The avidin gene family is located telomerically on the chicken sex chromosome Z, within a region of 27 kb on band Zq21 (18). The avidin gene is located at one end of the array, followed by a space of 9 kb and the *avr* cluster with intergenic distances of 2.5–2.8 kb (Fig. 1). Seven different *avr* genes have been characterized (*avr1–avr5*, 19, 20; *avr6–avr7*, 18). The genes are arranged tandemly in the same orientation, except *avr7* that is in the reverse orientation (Fig. 1). The *avrs* are highly conserved, being 94–100% identical to each other (18, 20). The genes consist of four exons and three introns with well-defined junctions and associated splicing signals. They all contain putative promoter sequences at their 5′-ends and polyadenylation signals at their 3′ ends suggesting that the genes may be functional. Indeed, RT-PCR experiments have detected low amounts of mRNAs encoded by *avr2* in the oviduct and *avr3* in the intestine under inflammation (21) as well as *avr1* in the chicken macrophage cell line HD11 in culture (unpublished). It is not currently known if the *avrs* are expressed at the protein level.

The *avd* gene was originally cloned by Wallén *et al.* (22). Comparison of the molecular structures of the *avd* and *avr* genes revealed that the exon-intron structure

[1] To whom correspondence and reprint requests should be addressed at Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35, FIN-40351 Jyväskylä, Finland. Fax: +358-14-260 2221. E-mail: mervi.ahlroth@cc.jyu.fi.

0006-291X/01 $35.00

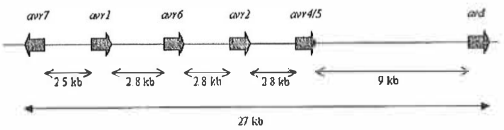

FIG. 1. Organization of the avidin gene family. Modified from Ahlroth *et al.* (18).

of *avd* is identical to that of the *avrs*. The overall identity between *avd* and the *avrs 1–7* was 91–95%, the differences consisting of nonrandomly distributed point mutations and a 6-bp deletion in the *avrs* compared to *avd* in exon 2. Furthermore, a 4-nt inversion hotspot was found in intron 1 of the genes. This inversion site was sequenced from three alleles of the *avd* gene. In the original avidin gene the sequence was ACTG (22), whereas in the alleles characterized by Ahlroth *et al.* (18) both the inverted form GTCA as well as the mutated form ATTG were found. Comparably, in the original *avr2* gene characterized by Keinänen *et al.* (19, 20) the sequence of this site was ACTG, while in the allele characterized by Ahlroth *et al.* (18) it was ATTG. The other *avr* genes show either GTCA (*avrs 1, 6,* and *7*) or ATTG (*avrs 3* and *4*).

The 5′-flanking regions of the different *avrs* thus far sequenced appear similar, whereas the upstream sequence of the *avd* gene differs radically (20, 23 and unpublished observations), suggesting a differential regulation pattern for *avd* and the *avrs*. Differential regulation mechanisms possibly do operate under inflammation, suggested by the considerable differences in the levels of induction of *avd* and the *avrs*. The fact that the major part of inflammation-induced avidin transcript is encoded by the *avd* gene (instead of *avrs*) also supports the suggestion that the progesterone and inflammation-induced avidins are products of the same gene. The 5′-flanking regions of *avrs 4* and *5* have been shown to contain chicken repeat 1 (CR1) elements (23). The CR1 elements are repetitive sequences belonging to the non-LTR class of retrotransposons and are present in 7000–30,000 copies dispersed throughout the chicken genome (24, 25). In the case of *avr4* and *avr5*, the elements are located at 1.4–2.1 kb upstream of the genes (23). The presence of CR1 elements upstream of the other *avr* genes and *avd* is currently unknown. However, more CR1 elements are expected to be found among the gene family, considering the even distribution of the *avr* genes as well as restriction enzyme mapping results suggestive of a repetitive nature for the intergenic regions as well as the coding regions (unpublished).

The involvement of gene conversion in the evolution of the *avd* gene family was originally inferred from the observation that intron sequences are more conserved between *avd* and the *avrs* than exon sequences (22). The homogeneity of the gene family is intriguing in the sense that there is no apparent requirement for restoring the *avr* sequences. Therefore, homogenization may be merely the consequence of high recombinogenicity of the locus. On the other hand, while the expression of the *avrs* at protein level is currently unknown, the nonrandom distribution of point mutations suggests that there may be, or has been at some point in evolution, selective pressure acting also on the exon sequences of the *avrs* (26). The aim of this paper is to resolve the evolutionary features of the gene family. The *avd* and *avr* gene sequences are examined closely to reveal any evidence for recombination and other rearrangement mechanisms that might have acted in the evolution, and might still be acting in the maintenance, of the gene family. Furthermore, the putative consequences of the sequence differences at the protein level are discussed.

MATERIALS AND METHODS

The nucleotide sequences used for gene comparisons can be found in the EMBL database under the following accession numbers (followed by the appropriate reference): *avd*, L27818 (22); *avr1-5*, Z21611, Z21554, Z21612, Z22883, and Z22882 (20); *avr6-7*, AJ237658, and AJ237659 (18). Additional alleles for *avd* (*avda2*, AJ311647) and *avr2* (*avr2a2*, AJ311648) were included in the analysis (unpublished). The sequences of *avrs 4* and *5* are identical except for a single substitution in the 5′-UTR (20). All the gene sequences contain four exons and three introns organized as follows (numbering according to *avd*): exon 1 (beginning from the translation start codon; total length 81 bp, nucleotides 1–81), intron 1 (84 bp, 82–165), exon 2 (211 bp, 166–376), intron 2 (426 bp, 377–801), exon 3 (121 bp, 802–922), intron 3 (87 bp, 923–1009), and exon 4 (43 bp, 1010–1052). In addition, part of the 3′-untranslated region (3′-UTR; 55 bp, 1053–1107) was included in the analysis.

We used the program SimPlot (version 2.5) to create a similarity plot to visualize the genetic diversity along the sequences. In this case, the sequences were analyzed starting from 13 bp upstream of the translation start codon. A sliding window of 140 bp with a step of 10 bp was used in the analysis.

To investigate if different regions along the *avd* and *avr* genes have different phylogenetic histories, we applied the program PLATO (partial likelihoods assessed through optimization; 27). PLATO utilizes a sliding window of varying size to find regions of alignment that do not fit with the null phylogenetic hypothesis. The null phylogenetic hypothesis in our case was the phylogenetic tree obtained with the maximum likelihood (ML) method together with a substitution model that best fitted the sequences. The substitution model was the General Time Reversible model with rate heterogeneity (GTR + G + I) obtained using the program Modeltest (28). Regions with a significant departure (indicated by the *Z* value) from the null hypothesis could be a result of recombination, gene conversion, selection or differential mutation rate. The phylogenetic analyses were conducted with the neighbor-joining method and the GTR + G + I model of substitution using Paup* (29).

Conversion tracts among the genes were traced using the program GENECONV (30–32). GENECONV is a statistical method to find possible gene conversion events in an alignment. The program assigns *P* values based on permutation (Sim P-values) with multiple comparison correction.

TABLE 1

(A) Above the Diagonal Are the Pairwise Genetic Distances Calculated with the GTR + G + I Model of Substitution among the *avd* and *avr* Genes, While below the Diagonal Is the Transitions/Transversions Ratio among Introns and Exons, Respectively; (B) Overall Mean p-Distances among *avd/avr* Genes for Exons and Introns Separately

(A)	*avd*	*avda2*	*avr1*	*avr2a1*	*avr2a2*	*avr3*	*avr5*	*avr6*	*avr7*
avd	—	0.0204	0.2052	0.1836	0.1668	0.1302	0.0829	0.1891	0.1833
avda2	11–7	—	0.1622	0.1692	0.1503	0.1284	0.0867	0.1364	0.1211
avr1	4.5–1.3	4.3–1.2	—	0.0400	0.0374	0.0545	0.1061	0.0194	0.0268
avr2a1	12–1.3	5.5–1.3	3.2–3	—	0.0056	0.0484	0.0779	0.0303	0.0359
avr2a2	nc–1.3	9–1.3	4–3	4–nc	—	0.0478	0.0767	0.0300	0.0356
avr3	17–1.1	nc–1.1	4.3–2.3	4.5–1.9	11–1.9	—	0.0394	0.0373	0.0254
avr5	12–0.9	4.5–0.9	2.8–1.9	3–2.1	6–2.1	4.5–1.4	—	0.0834	0.0799
avr6	9–1.2	11–1.1	4–2	4–1.8	7–1.8	11–2.8	4–1.8	—	0.0098
avr7	10.5–1.2	10–1.1	5.5–2	4.3–1.8	7.5–1.8	4–2	4.3–1.5	nc–2	—

(B)	Ex1	Int1	Ex2	Int2	Ex3	Int3	Ex4	3′-UTR
all	0.014	0.043	0.111	0.020	0.084	0.017	0.105	0.009
avrs	0.016	0.046	0.051	0.018	0.056	0.015	0.062	0.011

RESULTS

Nucleotide Variation

The sequence alignment of 1108 bp showed a total of 125 variable sites (11%), of which 89 (19.5%) were in exons and 36 (5.5%) were in introns and 3′-UTR. In addition, a 6-bp deletion occurred in exon 2 of the *avrs* compared to *avd*, and *avrs 1* and *2* also showed a 1-bp insertion in intron 2 (20, 22). Corrected sequence divergence values, calculated with the GTR + G + I model, are shown in Table 1A. The corrected distance between *avd* and the *avrs* ranged from 8 to 20%, and from 1 to 10% among the *avr* genes. Nucleotide substitutions were clustered rather than randomly distributed, suggesting that certain regions of the genes are subjected to different mutation rates. Comparing the divergence separately for each exon and intron revealed striking differences in their substitution rates. The overall mean uncorrected distance for each exon and intron is shown in Table 1B. Exon 2 showed the greatest divergence when comparing *avd* to *avrs* (23% between *avd* and *avr1*). Between different *avr* genes, exon 2 showed divergences ranging from 0.5 to 8.8%. The most conserved exon was exon 1, showing an overall average divergence of 1.6%. Exon 1 encodes the signal peptide, which is responsible for the extracellular targeting of AVD. The average diversity of introns ranged from 1.8% (between *avd* and *avr2a2*) to 3.4% (between *avd* and *avr7*). Between the different *avr* genes, the intron diversities ranged from 0.9% (between the two *avr2* alleles) to 2.9% (between *avr1* and *avr4*). (Pairwise comparison tables for each exon and intron are not shown).

The transition/transversion ratio was also considerably different between the exons and introns in pairwise comparisons, transitions being much more frequent in introns (Table 1A). Between *avd* and *avr3*, for example, the TS/TV ratio among the exons was close to one while transitions were 17 times more common between the introns. Moreover, most of the nucleotide substitutions between *avd* and the *avrs* as well as among the different *avr* genes occurred in the first or second codon position (22 and current study) and about 50% of the substitutions lead to an amino acid change. This pattern of substitution clearly indicates that different gene regions are subject to different models of evolution. To better identify regions that are evolving in an anomalous way, we conducted a Spatial Phylogeny Variation (SPV) analysis using PLATO. Although we used a model with rate heterogeneity and gamma shape, we identified an anomalous region extending over most of the length of exon 2 (nucleotide positions 171–347; Z value = 11.3, $P < 0.05$).

Gene Conversion

Inspection of the *avd* and *avr* sequences using SimPlot revealed a "patchwork pattern" of the genes (Fig. 2), strongly suggesting that gene conversion plays a major role in modifying the gene family. From Fig. 2, it is evident that exon 2 is the most divergent and intron 2 the most conserved region. The figure also shows that *avrs 4* and *5* are similar to the other *avrs* in all parts except exon 3, where they follow *avd* rather than the other *avrs*. Table 2 shows the possible conversion events in more detail, estimated with GENECONV. We observed ten putative gene conversion tracts. Four conversion events occurred between *avd* and *avr* genes. Two of them involved the 5′-end of the gene extending over the region including exon 1 and intron 1 (the first 180 bp). A putative gene conversion event between *avda2* and *avr4/5* involved mostly exon 3. One putative gene conversion tract, between *avr3* and *avr7*, concerned most of the length of the genes, extending from the middle of exon 2 (base 257) to the end of the alignment (total tract length 852 bp). The other gene conversion events involved shorter fragments, ranging from 22 to 271 bp. All conversion tracts between different *avr* genes involved exon 2, except for one tract (between *avrs 3* and *6*) that involved the terminal 271 bp of the genes.

Phylogenetic Analyses

The phylogenetic trees were constructed using the neighbor-joining method and the GTR + G + I model of substitution. The gamma shape parameter was 0.6645 and the proportion of sites unable to accept substitutions was 0.6412. The phylogenetic analysis grouped the *avr* genes together, with *avd* being the most divergent gene (Fig. 3). Among the *avr* group, the most closely related genes were *avrs 1* and *6*, while *avr4/5* diverged most from the other *avr* genes (Fig. 3). We conducted separate phylogenetic analyses for exon 2 and for the rest of the alignment because exon 2 showed anomalous features as indicated by the SVP analysis in PLATO. Indeed, the phylogenetic trees differed from each other with respect to the relationships of the *avr* genes (Fig. 4). When exon 2 was excluded form the phylogenetic analysis (Fig. 4A), *avr3* resulted the outgroup of a cluster formed by *avrs 1, 6,* and *7*. When we analyzed exon 2 separately, *avr7* resulted as being closely related to *avr6*, while *avrs 3* and *4/5* resulted very closely related to each other (Fig. 4B).

DISCUSSION

In this study, we analyzed the sequences of chicken avidin and seven avidin-related genes. The organization of the gene family has previously been determined and the inverted orientation of *avr7* discussed by Ahlroth *et al.* (18). Here, we discuss the 4-nt inversion hotspot in intron 1 of the genes (18) in detail. Since only three different forms of the inversion site (ACTG, GTCA, and ATTG) have been observed, it is possible that gene conversion affects this region, preventing it from mutating further. Inversion may operate at the level of large chromosomal segments or shorter DNA regions. The smallest inversions observed are just a few nucleotides long. Inversions of large chromosome segments, found in about 2% of humans, result from the formation of "inversion loops" in meiotic cells (33). Comparably, the mechanism underlying small, few base pair inversions is likely to be a local hairpin-loop secondary structure formation on the DNA, possibly associated with DNA replication (34). On the other hand, microconversions similar to those presented by Wheeler *et al.* (35) for the murine H-2 genes may generate frequent few base-pair changes, sometimes accompanied by point mutations. In the *avd* and *avr* genes, the sequence ACTG is repeated several times along intron 1 and at the beginning of exon 2, either directly or in complementary or reverse-complementary mode (not shown). Furthermore, a complementary form for ATTG (TAAC) is located adjacent to the 5′-splice signal of intron 1. These repeats may serve as templates for conversion leading to the observed variants of the 4-nt site.

The current study supports the view that the *avrs* have diverged from *avd* (18, 26; Fig. 3). We observed that from the total number of nucleotide substitutions (including the 6-bp deletion in exon 2) found in the *avrs* compared to *avd*, 37% are present in all *avr* genes. Second, *avd* is separated by 9 kb from the *avrs*, as opposed to the close proximity (2.5–2.8 kb) between the *avrs*. Geraghty *et al.* (36) observed that the members of HLA class I gene family tend to lie proximally to their putative parent sequences. The mechanism of HLA class I gene duplications remains obscure, but their findings suggested the involvement of *Alu* sequences flanking the genes. In any case, duplications of the HLA genes appeared to have occurred individually rather than in large block duplications. The evolution of *avr* genes may have followed similar pathways. An initial duplication of *avd* produced the first *avr* copy (putatively *avr4/5*), leading to further duplication events possibly effected by flanking CR1 elements. The phylogenetic relationships of the *avrs* are reflected by their positions on the chromosome, similarly to the HLA genes. For example, *avrs 1, 6,* and *7* are located next to each other on the chromosome in the order *avr7* (outermost)–*avr1*–*avr6* (Fig. 1). Correspondingly, the phylogenetic trees position *avrs 1* and *6* as sister genes and *avr7* as closely related to them (Fig. 3). Furthermore, the phylogenetic relationships are in agreement with the gene conversion events analyzed by GENECONV. GENECONV detects fragments that exhibit significantly higher similarity between two genes than the overall similarity along their whole lengths. Thus, the lack of conversion tracts between *avrs 1, 6* and *7* indicates that they are closely related rather than homogenized by conversion. In contrast, a highly significant conversion tract is found between *avr2a2* and *avr4/5* (Table 2), which are adjacent genes. The chromosomal position of *avr3* is unknown, since the gene was absent in the genomic clones used by Ahlroth *et al.* (18) to map the organization of the gene family. However, GENECONV detects a significant tract of identity between *avr3* and *avr7* for most of the length of the genes (257–1108; Table 2), and it may be that in this case recombination has occurred instead of conversion. That the identity between the two genes along the region 257–1108 is actually 100% supports this suggestion. Comparably, the 5′ end (1–255) of *avr3* is identical to *avr4/5* (not shown). Thus, *avr3* probably represents a fusion gene, produced by recombination between *avr4/5* and *avr7*. The recombination break-

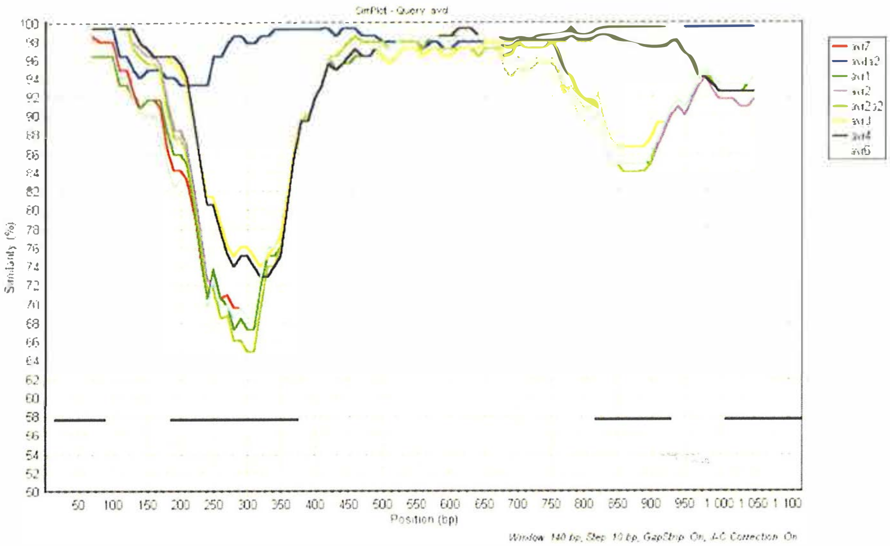

FIG. 2. Comparison of the *avr* gene sequences against *avd* using the SimPlot program. Alleles for *avd* (*avda2*) and *avr2* (*avr2a2*) are also included in the comparison. The approximate positions of exons are represented by bars below the curves. Note that *avr4* represents also the identical *avr5*, which was omitted. Intron 2 is obviously the most conserved region between *avd* and the *avrs*. The *avrs* diverge from *avd* toward the end of intron 2, with the exception of *avr4/5*, which follows the *avd* sequence through exon 3 to finally join the other *avrs* in intron 3.

point appears to lie between positions 269–281 as deduced from the alignment of the genes (not shown). Of course, the occurrence of two independent gene conversion events cannot be ruled out, but is more unlikely.

Avd seems to be well protected from homogenization with the *avrs*. As intrachromosomal recombination decreases with increasing distance between the participating repeats (37), the separation of *avd* from the *avrs* by 9 kb of apparently GC-rich DNA (unpublished observations) may represent an efficient barrier against gene conversion and crossing-over. These assumptions are, however, based on only few *avd* alleles (18, 22), and are thus speculative. On the other hand, the 3′-ends of *avrs 4* and *5* are almost identical to *avd* (Fig. 2). Considering that *avrs 4* and *5* show the 6-bp deletion and most of the nucleotide substitutions characteristic for all *avrs*, it may be that these genes have been partially converted by *avd.* This hypothesis necessitates the assumption that the conversion process exhibits polarity, so that *avd* can convert *avrs* but not

TABLE 2

Most Significant Fragments Representing Possible Gene Conversion Events in the *avd* Gene Family

Sequence	Sim P-value	Begin	End	Length	N Poly	N Dif	Tot Diff
avr3-avr7	0.0000	257	1108	852	97	0	23
avr2-avr4/5	0.0000	271	344	74	28	0	53
avr2a2-avr4/5	0.0000	271	344	74	28	0	53
avd-avr2	0.0000	1	180	180	12	0	92
avda2-avr4/5	0.0002	768	900	133	20	0	63
avda2-avr1	0.0052	126	205	81	10	0	87
avr3-avr6	0.0074	257	401	145	36	0	31
avd-avr2a2	0.0082	1	159	159	9	0	90
avr3-avr6	0.0197	838	1108	271	33	0	31
avr1-avr4/5	0.0434	271	292	22	13	0	65

Note. Sim P-value, *P* values obtained with 10.000 permutation; N Poly, number of polymorphic sites in the fragment; N Dif, number of mismatches within the fragment; Tot Difs, total number of mismatches between the two sequences.

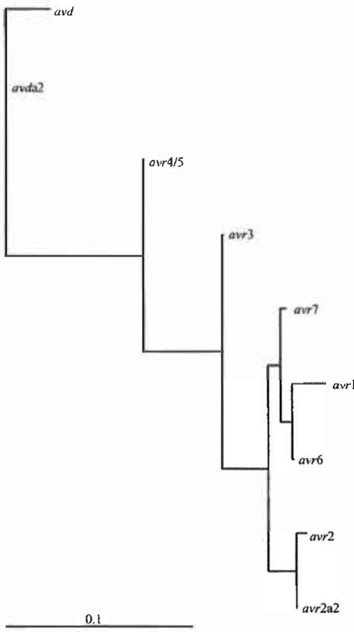

FIG. 3. Phylogenetic tree for the 1108 bp of *avd* and *avr* genes obtained with the neighbor-joining method and the GTR + G + I model of substitution using PAUP*.

vice versa. Directionality for gene conversion has been observed (37), supporting our hypothesis.

Gene conversion is believed to result from the alternative resolution of two Holliday junctions in double-strand break repair (DSBR), although other models have also been proposed (reviewed in 38). Another possible outcome in the repair process is crossing-over, as supported by the fact that as many as half of conversion events are associated with recombination (37 and references therein). Thus, recombination as well as gene conversion can also be expected to occur among the avidin gene family. As described above, *avr3* may represent a crossing-over product, supporting the alternative DSBR resolution model.

Gene conversion leads to gradual homogenization of the gene family sequences (39). In some gene families, however, certain regions in the exchanged sequences seem to be resistant to homogenization. In the human red and green color pigment genes, for example, the intron regions are more homogenous than the exon regions although the involvement of gene conversion in the evolution of the genes is evident. In this case, there has been a selective pressure toward restoring differences in the exon sequences, since homogenization of the coding sequences would reduce the spectrum of color vision (39 and references therein). In the major histocompatibility (MHC) genes, gene conversion acts to preserve the diversity between different clusters of MHC genes by favoring conversion within clusters instead of that between different clusters (37 and references therein). The killer cell inhibitory receptors (KIRs) also exhibit diversification by gene conversion or double recombination (40). On the basis of these examples and the fact that intron 2 is the most homogenized region in the *avd* and *avr* genes, it may be that restoring the differences between the exons confers selective advantage. In any case, the pattern of substitutions clearly indicates that different regions of the *avd* and *avr* genes are subject to differential models of evolution. According to Fryxell (41), genes coding proteins with related functions often coevolve. Therefore, the avidin gene family may evolve together with other immune system-related gene families such as the MHC, immunoglobulin and KIR families.

Most of the nucleotide substitutions in the putative AVRs, compared to AVD, occur in the first or second codon position (22), and about 50% of the substitutions lead to an amino acid change. The tryptophan, tyrosine, and phenylalanine residues responsible for the formation of the hydrophobic biotin-binding pocket in AVD (14) are conserved in all putative AVRs (20 and unpublished), suggesting that they are able to bind biotin. Most of the differences (encoded by exon 2) occur in a loop region located near the biotin-binding pocket in the 3-dimensional fold of the proteins (not shown). The loop confers hydrogen bonds for biotin binding in AVD, and may stabilize the binding by physically preventing dissociation of the bound molecule. Therefore, the biotin binding may exhibit lower affinity than that of AVD. The amino acids involved in inter-subunit interactions in the tetrameric protein show surprisingly numerous nonhomologous substitutions in the AVRs compared to AVD (unpublished). These substitutions, as well as the apparently lower isoelectric points (p*I*s) compared to AVD (unpublished), may distort the oligomerization of AVRs. However, the AVR sequences were used as a model for lowering the p*I* of AVD in a previous study (42). All the mutated AVDs exhibiting lower p*I*s showed high-affinity biotin binding and tetrameric structure, suggesting that the p*I* changes did not interfere with biotin binding and tetramerization. Comparably, differences in N-glycosylation patterns putatively occur between AVD and the AVRs, and these differences have also been utilized to engineer AVD (43). Again, the nonglycosylated AVD bound biotin with high affinity and showed no distortion of quaternary structure compared to native AVD. The consequences of the observed differences in primary sequence are thus not simple to infer. We are currently producing recombinant AVR proteins in order to characterize their structural and functional properties. If the AVR proteins exhibit characteristics that differ from AVD, they may indeed

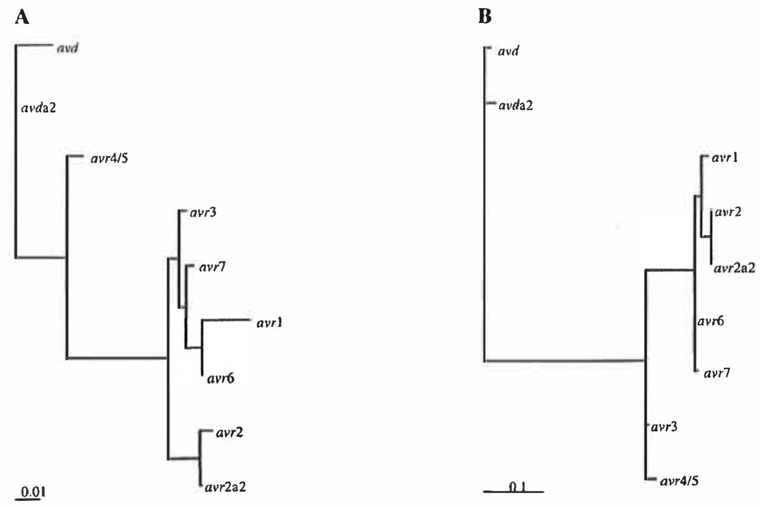

FIG. 4. Phylogenetic tree obtained from the entire sequence with the exclusion of exon 2 (A) and from exon 2 only (B). Both phylogenetic trees were constructed with the neighbor-joining method and the GTR + G + I model of substitution using PAUP*.

play a role in producing a wider variety of anti-microbial or other host defense-related functions.

ACKNOWLEDGMENT

We thank Dr. Edward Bayer for detailed information on the avidin structure.

REFERENCES

1. Wilchek, M., and Bayer, E. A. (1988) The avidin–biotin complex in bioanalytical applications. *Anal. Biochem.* **171**, 1–32.
2. Wilchek, M., and Bayer, E. A. (1990) Introduction to avidin–biotin technology. *Methods Enzymol.* **184**, 5–13.
3. Hertz, R., and Sebrell, W. H. (1942) Occurrence of avidin in the oviduct and secretions of the genital tract of several species. *Science* **96**, 257.
4. Hertz, R., Dhyse, F. G., and Tullner, W. M. W. (1949) Elevation of biotin activity in the serum of estrogen treated chicks: Relationship to hormone-induced avidin formation in the oviduct. *Endocrinology* **45**, 451–454.
5. O'Malley, B. W. (1967) *In vitro* hormonal induction of a specific protein (avidin) in chick oviduct. *Biochemistry* **6**, 2546–2551.
6. Tuohimaa, P., Joensuu, T., Isola, J., Keinänen, R., Kunnas, T., Niemelä, A., Pekki, A., Wallén, M., Ylikomi, T., and Kulomaa, M. (1989) Development of progestin-specific response in the chicken oviduct. *Int. J. Dev. Biol.* **33**, 125–134.
7. Green, N. M. (1975) Avidin. *Adv. Protein. Chem.* **29**, 85–133.
8. Stevens, L. (1991) Egg white proteins. *Comp. Biochem. Physiol.* **100B**, 1–9.
9. Elo, H. A., Kulomaa, M. S., and Tuohimaa, P. J. (1979) Avidin induction by tissue injury and inflammation in male and female chickens. *Comp. Biochem. Physiol.* **62B**, 237–240.
10. Elo, H. A., Räisänen, S., and Tuohimaa, P. J. (1980) Induction of antimicrobial biotin-binding egg white protein (avidin) in chick tissues in septic *Escherichia coli* infection. *Experientia* **36**, 312–313.
11. Botte, V., and Granata, G. (1977) Induction of avidin synthesis by RNA obtained from lizard oviducts. *J. Endocrinol.* **73**(3), 535–536.
12. Hunt, L. T., and Barker, W. C. (1989) Avidin-like domain in an epidermal growth factor homolog from a sea urchin. *FASEB J.* **3**, 1760–1764.
13. Argaraña, C. E., Kuntz, I. D., Birken, S., Axel, R., and Cantor, C. R. (1986) Molecular cloning and nucleotide sequence of the streptavidin gene. *Nucleic Acids Res.* **14**(4), 1871–1882.
14. Livnah, O., Bayer, E. A., Wilchek, M., and Sussman, J. L. (1993) Three-dimensional structures of avidin and the avidin–biotin complex. *Proc. Natl. Acad. Sci. USA* **90**, 5076–5080.
15. Bayer, E. A., Kulik, T., Adar, R., and Wilchek, M. (1995) Close similarity among streptavidin-like, biotin-binding proteins from *Streptomyces. Biochim. Biophys. Acta* **1263**, 60–66.
16. Green, N. M. (1990) Avidin and streptavidin. *Methods Enzymol.* **184**(3), 51–67.
17. Korpela, J. K., Kulomaa, M. S., Elo, H. A., and Tuohimaa, P. J. (1981) Biotin-binding proteins in eggs of oviparous vertebrates. *Experientia* **37**(10), 1065–1066.
18. Ahlroth, M. K., Kola, E. H., Ewald, D., Masabanda, J., Sazanov, A., Fries, R., and Kulomaa, M. S. (2000) Characterization and chromosomal localization of the chicken avidin gene family. *Anim. Genet.* **31**(6), 367–375.
19. Keinänen, R. A., Laukkanen, M.-L., and Kulomaa, M. S. (1988) Molecular cloning of three structurally related genes for chicken avidin. *J. Steroid Biochem.* **30**(1–6), 17–21.
20. Keinänen, R. A., Wallén, M. J., Kristo, P. A., Laukkanen, M. O., Toimela, T. A., Helenius, M. A., and Kulomaa, M. S. (1994)

Molecular cloning and nucleotide sequence of chicken avidin-related genes 1–5. *Eur. J. Biochem.* **220,** 615–621.

21. Kunnas, T. A., Wallén, M. J., and Kulomaa, M. S. (1993) Induction of chicken avidin and related mRNAs after bacterial infection. *Biochim. Biophys. Acta* **1216,** 441–445.
22. Wallén, M. J., Laukkanen, M. O., and Kulomaa, M. S. (1995) Cloning and sequencing of the chicken egg-white avidin-encoding gene and its relationship with the avidin-related genes Avr1–Avr5. *Gene* **161,** 205–209.
23. Wallén, M. J., Keinänen, R. A., and Kulomaa, M. S. (1996) Two chicken repeat one (CR1) elements lacking a silencer-like region upstream of the chicken avidin-related genes Avr4 and Avr5. *Biochim. Biophys. Acta* **1308,** 193–196.
24. Stevens, L. (1996) Avian Biochemistry and Molecular Biology, Cambridge Univ. Press, Cambridge, UK.
25. Stumph, W. E., Hodgson, C. P., Tsai, M.-J., and O'Malley, B. W. (1984) Genomic structure and possible retroviral origin of the chicken CR1 repetitive DNA sequence family. *Proc. Natl. Acad. Sci. USA* **81,** 6667–6671.
26. Wallén, M. J. (1996) The chicken avidin gene family. Ph.D. thesis. Acta universitatis Tamperensis ser A vol. 515.
27. Grassly, N. C., and Holmes, E. C. (1997) A likelihood method for detection of selection and recombination using sequence data. *Mol. Biol. Evol.* **14,** 239–247.
28. Posada, D., and Crandall, K. A. (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14,** 817–818.
29. Swofford, D. L. (1998) 'PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods), version 4, Sinauer Associates, Sunderland, MA.
30. Sawyer, S. A. (1989) Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6,** 526–538.
31. Sawyer, S. A. (1999) GENECONV: A computer package for statistical detection of gene conversion [1.81], Department of Mathematics, Washington University, St. Louis, MO.
32. Drouin, G., Prat, F., Ell, M., and Paul Clarke, G. D. (1999) Detecting and characterizing gene conversions between multigene family members. *Mol. Biol. Evol.* **16,** 1369–1390.
33. Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C., and Gelbart, W. M. (1993) An Introduction to Genetic Analysis, 5th ed., Freeman, New York.
34. Leach, D. R. F. (1994) Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *BioEssays* **16**(12), 893–900.
35. Wheeler, C. J., Maloney, D., Fogel, S., and Goodenow, R. S. (1990) Microconversion between murine H-2 genes integrated into yeast. *Nature* **347,** 192–194.
36. Geraghty, D. E., Koller, B. H., Hansen, J. A., and Orr, H. T. (1992) The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. *J. Immunol.* **149,** 1934–1946.
37. Martinsohn, J. Th., Sousa, A. B., Guethlein, L. A., and Howard, J. C. (1999) The gene conversion hypothesis of MHC evolution: A review. *Immunogenetics* **50,** 168–200.
38. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983) The double-strand-break repair model for recombination. *Cell* **33,** 25–35.
39. Li, W.-H. (1997) Molecular Evolution, Sinauer Associates, Sunderland, MA.
40. Shilling, H. G., Lienert-Weidenbach, K., Valiante, N. M., Uhrberg, M., and Parham, P. (1998) Evidence for recombination as a mechanism for KIR diversification. *Immunogenetics* **48**(6), 413–416.
41. Fryxell, K. J. (1996) The coevolution of gene family trees. *Trends Genet.* **12**(9), 364–369.
42. Marttila, A. T., Airenne, K. J., Laitinen, O. H., Kulik, T., Bayer, E. A., Wilchek, M., and Kulomaa, M. S. (1998) Engineering of chicken avidin: A progressive series of reduced charge mutants. *FEBS Lett.* **441,** 313–317.
43. Marttila, A. T., Laitinen, O. H., Airenne, K. J., Kulik, T., Bayer, E. A., Wilchek, M., and Kulomaa, M. S. (2000) Recombinant NeutraLite avidin: A non-glycosylated, acidic mutant of chicken avidin that exhibits high affinity for biotin and low non-specific binding properties. *FEBS Lett.* **467,** 31–36.

III

Copy-number fluctuation by unequal crossing-over in the chicken avidin gene family

by

Ahlroth, M. K., Ahlroth, P. & Kulomaa, M. S.

Biochem. Biophys. Res. Comm. 288: 400-406, 2001

https://doi.org/10.1006/bbrc.2001.5760

Biochemical and Biophysical Research Communications **288**, 400–406 (2001)
doi:10.1006/bbrc.2001.5760, available online at http://www.idealibrary.com on IDEAL®

Copy-Number Fluctuation by Unequal Crossing-Over in the Chicken Avidin Gene Family

Mervi K. Ahlroth,*[,1] Petri Ahlroth,†[,2] and Markku S. Kulomaa*
*Department of Biological and Environmental Science and †Jyväskylä University Museum, Section of Natural History, University of Jyväskylä, FIN-40351 Jyväskylä, Finland

Received September 18, 2001

The chicken avidin gene (*AVD*) forms a closely clustered gene family together with several avidin-related genes (*AVRs*). In this study, we used fluorescence *in situ* hybridization on extended DNA fibers (fiber-FISH) to show that the number of the *AVD* and *AVR* genes differs between individuals. Furthermore, the gene copy-number showed wide somatic variation in white blood cells of the individuals. The molecular mechanism underlying the fluctuation is most probably unequal crossing-over and/or unequal sister chromatid exchange, as judged by the Gaussian distribution of the gene counts. By definition, an increase in gene number on one locus should be accompanied by a decrease on the other locus in unequal sequence exchange. The results suggest that copy-number lability may be more common among gene families than previously thought. The chicken avidin gene family also provides an excellent model for studying the mechanisms of recombination and gene conversion. © 2001 Academic Press

Avidin, well known from the various avidin-biotin technology applications (1), is naturally found in chicken eggs (2). Its function is thought to be defending the developing chick embryo against bacterial infection by sequestering the vitamin biotin (3 and references therein, 4). Avidin also acts as a local acute defense protein in several tissues after injury or microbial infection (5–7). The gene encoding avidin has several homologues in the chicken, called avidin-related genes (*AVRs*). Seven of these homologues (*AVR1–AVR7*) have been characterized (8–10). Some *AVRs* are transcribed in chicken tissues under inflammation (11 and unpublished), but it is not currently known if the *AVR* transcripts are translated into proteins.

We recently mapped the organization of the avidin gene family (10). The gene family was found to reside in a single cluster on the chicken sex chromosome Z, close to the telomere on band Zq21. Thus, the male chicken (ZZ) harbors two avidin loci, whereas the female (ZW) has only one locus. The *AVR* genes were found to lie as a tandem array with the exception of *AVR7*, which was oriented in the opposite direction at one end of the array (10). The avidin gene was located at the other end of the array, oriented in the same direction as the majority of the *AVRs*. However, *AVD* was separated by 9 kb from the *AVRs*, whereas the distances between the *AVRs* were 2.5–2.8 kb. Evolutionary analyses indicated frequent gene conversion events between the *AVD* and *AVR* genes (12).

Interestingly, we found different sets of *AVR* genes in two genomic libraries we screened (10). Our *AVR* sets also differed from those previously observed (8, 9). None of the 11 genomic clones we analyzed contained the *AVR3* gene described by Keinänen *et al.* (8, 9), or any gene attributable as an allele of *AVR3*. The clones contained two previously uncharacterized *AVR* genes (*AVR6* and *AVR7*) instead. Furthermore, *AVRs 4* and *5* were not found together in any of the cosmid clones by Ahlroth *et al.* (10), *AVR4* being present in one library and *AVR5* in the other. In contrast, Keinänen *et al.* (9) originally cloned both *AVR4* and *AVR5* from a single genomic lambda clone of 20 kb, ruling out the possibility that the original genes were alleles to each other. The different *AVR* composition found in each library led us to a hypothesis that the number of the *AVD* and *AVR* genes may vary between different individuals.

Tandemly repeated sequences are thought to be susceptible to copy-number fluctuations and rearrangements by mechanisms such as unequal crossing-over or unequal sister-chromatid exchange (13). Telomeric position of the repeated sequences further facilitates rearrangements (14). The avidin gene family meets both of these criteria. Since not only the coding sequences

[1] To whom correspondence should be addressed at University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35 (YAB), FIN-40351 Jyväskylä, Finland. Fax: +358-14-260 2221. E-mail: mervi.ahlroth@cc.jyu.fi.
[2] Present address: Ministry of Forestry and Agriculture, P.O. Box 30, FIN-00023 Helsinki, Finland.

0006-291X/01 $35.00

but also the intergenic sequences appear to be conserved (unpublished), unequal crossing-over events among the avidin gene family are plausible. As unequal crossing-over or sister-chromatid exchange can occur in any mitotic cell division, variation is also expected to occur within individuals. Thus, each individual may be a mosaic of cells harboring different numbers of *AVD* and *AVR* genes. For clustered gene families such as the *AVD* family, unequal crossing-over events may, therefore, be considerably more frequent than previously assumed.

In this study, copy-number variation within the avidin gene family was tested using the fiber-FISH technique. The gene number was shown to differ not only between different chickens, but also between different cells within a single individual. The gene counts from each individual were ranged according to normal distribution, suggesting that the molecular mechanism underlying the fluctuation is unequal crossing-over and/or unequal sister chromatid exchange. Together with the frequent gene conversion observed within the avidin gene family (12), our results support the models that suggest a common initiation mechanism for both gene conversion and recombination, such as the double stranded DNA break repair (DSBR) model (15, 16). The results also suggest that gene copy-number fluctuation may be a relatively frequent phenomenon among gene families and, therefore, eukaryotic genomes may exhibit more lability than previously thought.

MATERIALS AND METHODS

The fiber-FISH procedure was modified from Heiskanen *et al.* (17) (see also 18–21).

The chickens. Six chickens from two different breeds (LSL and a countryside breed denoted M) were used. The LSL individuals were all females, whereas one of the M individuals was male (M5).

Preparing cell blocks. Blood from the chickens was sampled and mononuclear white blood cells were separated and collected using gradient centrifugation. For the cell line HD11, cells were trypsinized from petri dishes and washed once with 1× PBS. Cell density was adjusted to 1×10^6 cells/100 μl in PBS, and the cell suspension was mixed with an equal volume of 1.9% SeaPlaque GTG low-melting point (LMP) agarose (FMC BioProducts) and solidified in 100 μl aliquots. The solidified blocks were incubated in lysis buffer (2 mg/ml proteinase K, 1% *N*-lauroylsarcosine, 0.5 M EDTA, pH 8.0) at +37°C for 2–3 days to lyse the cells, washed with TE buffer (10 mM Tris, 1 mM EDTA) and stored in TE or 50 mM EDTA at +4°C.

For the control blocks, 1 μg of cosmid K18-233 (10 and Fig. 2J) was digested with *NotI*. After heat-inactivation of *NotI*, the DNA was mixed with LMP agarose to produce 0.8% LMP blocks. The blocks were stored as above (proteinase K treatment was omitted).

Preparing fiber slides. Microscope slides were coated with poly-L-lysine (Sigma) and stored at +4°C until used. A small piece of the cell-LMP agarose block was melted on a microscope slide in a microwave oven. The melted agarose was immediately spread over the slide by gently drawing with another glass slide. The fiber slides were either used immediately or stored overnight at +4°C.

Probe labeling. The cosmid probe (cosmid clone K18-233; 10 and Fig. 2J) was labeled with digoxigenin by nick translation. For the gene probe, either *AVD* alone or both *AVD* and all the *AVR* genes were used as a template. The gene probe was biotin labeled using the random prime method.

Hybridization. The fiber slides were denatured in 70% formamide/2× SSC at +74°C and immediately fixed in ethanol series. Probes were ethanol precipitated and dissolved in 1:1 formamide: hybridization master mix (20% dextran sulfate in 2× SSC), and denatured in a boiling water bath. Each slide was hybridized with 100 ng of cosmid probe and 50 ng of gene probe. The slides were covered with cover slips and sealed with rubber cement (Fixogum, Marabu). Hybridization was performed overnight in a moist chamber at +37°C.

Posthybridization washes and detection. The hybridized fiber slides were washed three times in 50% formamide/2× SSC and three times in 2× SSC at +44°C. The slides were then blocked at +37°C for 30 min using 5% BSA/4× SSC, 0.2% Tween. After blocking, the slides were briefly washed in 4× SSC and the labels were detected using three layers of antibody/streptavidin conjugates diluted in 0.5% BSA/4× SSC, 0.2% Tween. The detection steps were (i) streptavidin-Alexa Fluor546 (1 μg/ml, Molecular Probes) + mouse anti-DIG (0.5 μg/ml, Boehringer Mannheim), (ii) biotinylated rabbit anti-streptavidin (5 μg/ml) + goat anti-mouse-Alexa Fluor488 (10 μg/ml, Molecular Probes), and (iii) streptavidin-Alexa Fluor546 (1 μg/ml) + donkey anti-goat-Alexa Fluor488 (10 μg/ml, Molecular Probes). All incubations were performed at +37°C for 30 min. The slides were washed three times in 4× SSC at +44°C after each detection step. After the final washes, the slides were rinsed with dH_2O, air dried, and mounted with DAPI-antifade mixture (Molecular Probes).

Data analysis. The fiber slides were examined using a Leitz DM RBE fluorescence microscope equipped with a black-and-white Spot RT digital camera (Diagnostic Instruments). The green and red signals were photographed separately and the images were merged and pseudocolored using the Spot advanced software (Diagnostic Instruments). The number of red gene-specific signals overlapping green whole-locus signals was counted from 100 fibers from each individual as well as from the control and the cell line. Statistical analysis was performed using the SPSS for Windows software.

RESULTS AND DISCUSSION

Results from the extended chromatin FISH experiments confirmed the deduced organization of the avidin gene family (10). The *AVD* gene was always seen at one end of the cluster, followed by a gap and the *AVR* genes clustered close to each other. However, the total number of genes differed between individuals and even between the cells within an individual. The "standard," or most frequent, gene number was 4 or 5 *AVR*s and one *AVD*, depending on the individual (Fig. 1). Putative duplications and triplications of the *AVD* gene were occasionally observed (not shown), but it was not possible to determine if the extra signals were extra *AVD* genes or *AVR*s that had "jumped" closer to the *AVD*. Figure 2 illustrates representative fibers showing different gene numbers within and between individuals.

In some cases, the whole locus was duplicated. Duplications were observed in all possible orientations: tail-to-tail (Fig. 2F), tandem, and head-to-head (not shown). Thus, there may be several kinds of mechanisms involved in the duplications. Tandem duplications, for example, might result from long-tract conver-

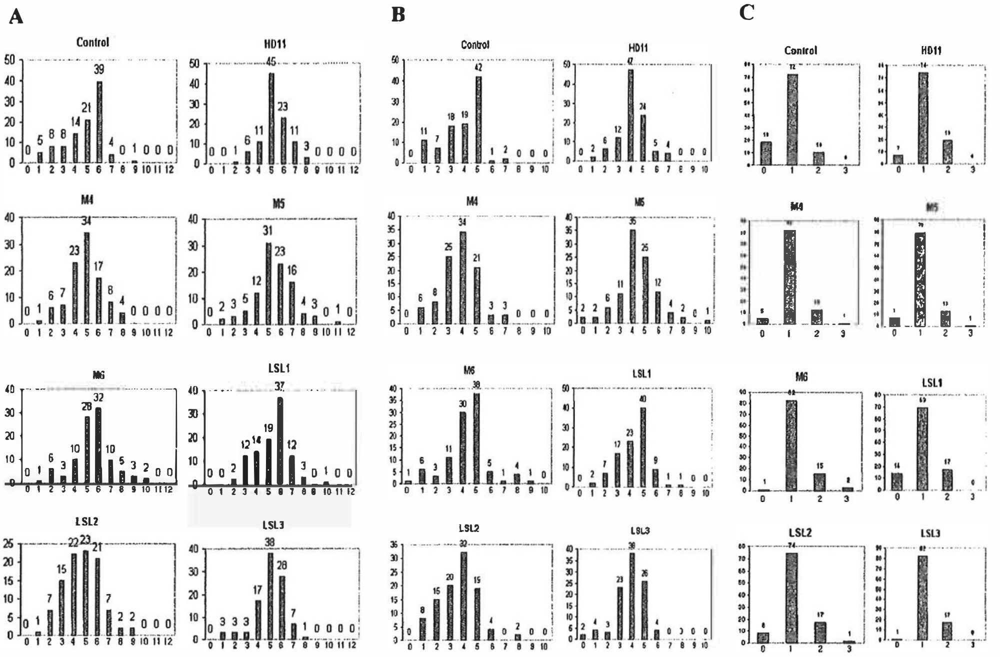

FIG. 1. Distributions of the gene counts in fiber-FISH. The x-axis shows the gene count, and the frequency of each count is shown on top of the columns. (A) Total gene counts (*AVD* + *AVR*), (B) *AVR* counts, and (C) *AVD* counts.

sion rather than crossing-over. The duplicates nearly always contained different numbers of *AVR* genes, suggesting that deletion or gain of gene copies is frequently associated with locus duplication. Furthermore, three fibers showed a duplicated *AVR* cluster connected by a single *AVD*. Another three fibers showed a single *AVR* cluster surrounded by two *AVD* genes, one on either side of the cluster (not shown). However, in these cases it was not possible to rule out the possibility that one of the putative *AVD* genes was actually an *AVR*. Altogether, 11 locus duplications were seen among the 800 fibers examined in this study, suggesting a duplication frequency of 1.4%. However, there is a source of error with respect to the duplications: they may be artificially formed by the intertwining of two DNA fibers. Nonetheless, some duplications, especially those showing a single *AVD* connecting two *AVR* clusters, were most probably authentic.

In statistical analysis, different individuals differed significantly with respect to the total gene count (ANOVA, $F_{6,698} = 4.219$, $P < 0.001$) as well as to the *AVD* count (Table 1). Different breeds (LSL and M) also differed slightly from each other (ANOVA, $F_{1,598} = 4.338$, $P = 0.038$). In contrast, there was no significant difference between the sexes within a breed (M5 male vs M4 and M6 females; *t* test, $t = -1.385$, df = 298, $P = 0.167$). Interestingly, the *AVD* count correlated with the *AVR* count (Table 1). The control (cosmid K18-233) essentially lacked gene counts greater than 6, and thus showed a highly biased distribution (Fig. 1A). Therefore, it was omitted from the statistical analysis.

The main technical problem in the fiber-FISH technique is the loss of signals due to inefficient hybridization of the probes (22, 23). One reason is that not all genes are equally accessible to the probe due to the helical conformation of DNA. As the DNA fiber is attached to the glass slide at every turn, the signal array resembles a "pearls in a chain" arrangement rather than a continuous stretch (Fig. 2). In the current application, the technique was operating near the limits of its resolution. A resolution limit of 1 kb has been proposed for fiber-FISH, determined by the optical resolution of the microscope (0.2–0.35 μm) assuming a condensation degree of 1 kb/0.34 μm for extended DNA (22 and references therein). The detection limit is lower, as sequence-tagged sites (STS) or expressed sequence tags (EST) have been mapped by fiber-FISH using probes as short as 200 bp (22, 21). Thus, it should be possible to detect signals of 200 bp separated by 1 kb under optimal conditions. However, the detection efficiency decreases along probe size, being 70–90% for probes >400 bp and 30% for probes of 200–250 bp (22). Furthermore, Horelli-Kuitunen *et al.* (21) found that detecting two or three ESTs simultaneously was even less efficient, about 15%.

In this study, each gene (1.1 kb) was represented by a single signal dot, and the intergenic distance of 2.5–2.8 kb was sufficient to separate the gene signals only on highly extended fibers (Fig. 2). A considerable amount of variation in signal counts thus represented technical errors, necessitating the analysis of large numbers of loci (100 fibers/sample). However, fibers showing gene signals exceeding the "standard," or most frequent, gene count (4 or 5 *AVR*s and one *AVD*, depending on the individual; Fig. 1) could be considered more reliable than those showing missing signals. This was inferred from the analysis of the control target, cosmid K18-233, which very seldom showed extra gene signals (5% of the cases). In contrast, signals fewer than 6 (the actual total gene count) were frequently observed (56% of the cases, Fig. 1A). Also Florijn *et al.* (22) found that nonspecific hybridization producing extra signals was rare in fiber-FISH. On the other hand, the genomic DNA samples in our study showed less technical variation than the cosmid control. This is probably due to the excess background staining of the fibers on the control slides. The degree of variation of the true gene number as opposed to technical variation in the authentic samples can be inferred from the diagrams in Fig. 1. For each case of increased gene number, there must be a corresponding count of decreased gene number, assuming that the mechanism of variation is unequal crossing-over. Thus, the column heights can be adjusted to correspond by removing those portions on the on the left hand side of the peak that exceed those on the right. For some individuals, however, the values on the right side of the peak exceed the values on the left (HD11, M5 and LSL3; Fig. 1). In these cases, genes may have been gained by long-tract gene conversion. Several putative gene conversion tracts (22–852 bp in length) have been detected within the *AVD* and *AVR* gene sequences (12). Gene conversion over tracts as long as 10 kb has been observed (16 and references therein). Clusters of *AVD* and/or *AVR* genes may therefore be gained by "copying" from one locus to another, leaving the gene number in the donor locus unchanged. The positive correlation observed for the *AVD* and *AVR* counts supports this view: misalignment of the *AVR* genes, followed by long-tract conversion, might result in copying extra *AVD* genes into the acceptor locus.

The high frequency of both gene conversion and unequal crossing-over within the avidin gene family strongly favors the view that the two events are initiated by a common mechanism. Gene conversion and recombination are perhaps best explained as alternative resolutions of two Holliday junctions in double-stranded DNA break repair (DSBR; 16), although other models have also been proposed (reviewed in 15 and 16). In any case, as many as half of conversion events are generally associated with recombination (16 and references therein).

Since the chicken chromosome Z, harboring the *AVD* gene family, is present in only one copy in the female, the recombination mechanism in the female must be gene conversion and/or unequal sister-chromatid exchange (SCE). In the male, unequal crossing-over between the two copies of chromosome Z as well as unequal SCE could occur. Surprisingly, the male did not show greater variation in gene number than the females, as might be expected because of the additional level of recombination available. It has been estimated that the overall frequency of unequal sister-chromatid exchange and gene conversion is about 1.5%, and that the frequency of crossing-over between homologous chromosomes is much higher than this (13). However, in experiments involving gene duplications in yeast, intrachromosomal recombination was observed at a rate 10–100 times higher than recombination between homologous chromosomes (16 and references therein). DNA double-strand breaks in mammalian cells are also repaired by homologous recombination using the sister chromatid as the preferred template (24 and references therein). Obviously, the frequency of the different recombination mechanisms vary between different sequence repeats and genomic contexts, as well as between cell types and organisms.

The fiber-FISH experiments in our study were limited to white blood cells (the cell line HD11 represents macrophage-type cells; 25). As already noted, considerable recombination was observed. In contrast, genomic libraries showed no evidence of recombination (8–10). This discrepancy can be explained by cell-type differences. The libraries were prepared from solid tissues (liver or muscle), in which excessive recombination frequencies have not been observed. Therefore, the libraries represented the inherent number of the genes in each individual, without interference caused by somatic rearrangements. The fact that the gene copy-number differed between the libraries (and thus also between the individuals from which the libraries were prepared) implies germ-line recombination. Another example of cell-type differences was provided by Högstrand and Böhme (26). They found elevated levels of gene conversion between the MHC II-class genes $E\beta^d$ and $A\beta^k$ in mouse sperm cells, whereas the same genes in liver cells showed no evidence for conversion.

White blood cells exhibit extraordinarily high levels of somatic recombination to bring about the rearrangements of the various antigen receptor genes, such as the immunoglobulin and T-cell receptor (TCR) genes (27). The immunoglobulin genes undergo somatic hy-

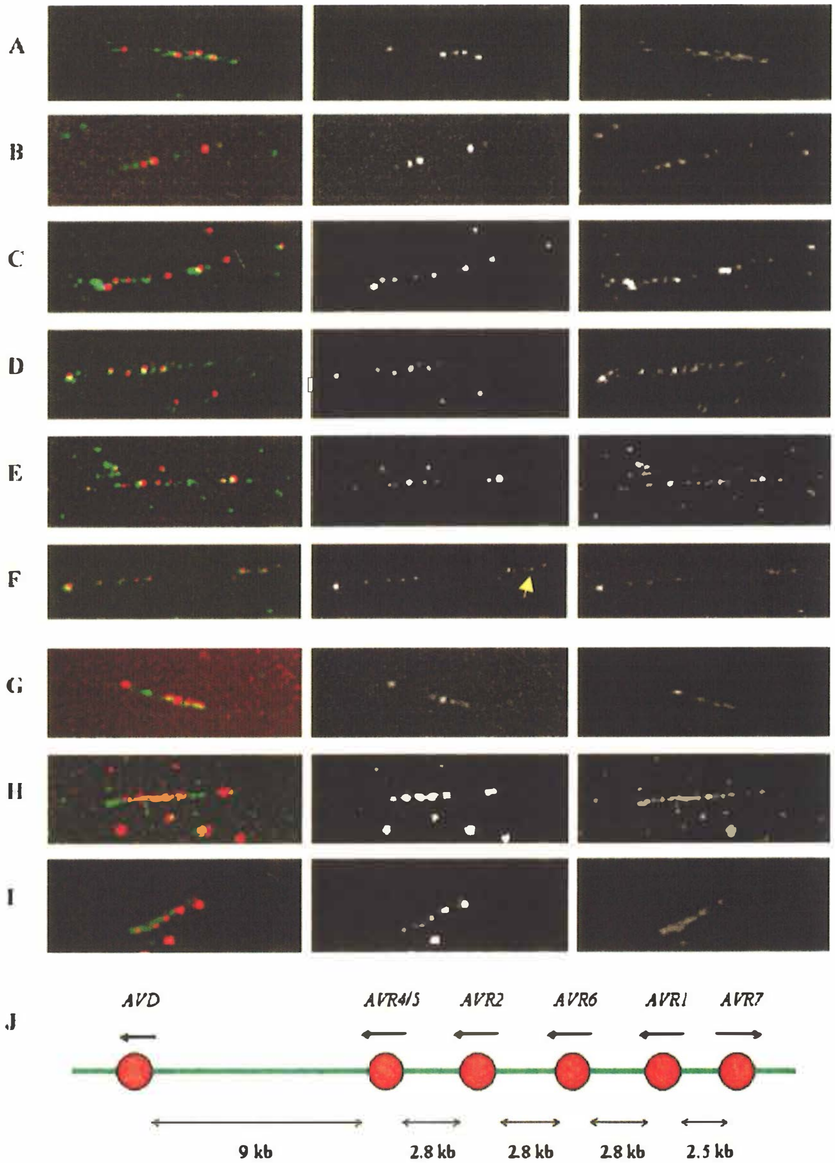

FIG. 2. Representative fiber-FISH results. On each row, separate grayscale images for the locus signal (image on the right) and the gene signals (center) are shown. The image on the left represents the two channels pseudocolored and superimposed. (A–F) Variation within a single individual (M4). (A) 5 *AVR*s + 1 *AVD* (the *AVR* signal on the far right is weak but detectable). (B) 3 *AVR*s + 1 *AVD* + 1 additional *AVR*. (C) 7 *AVR*s + 1 *AVD*. (D) 6 *AVR*s + 1 *AVD*. (E). 5 *AVR*s + 1 *AVD* (bright signal) + 1 additional *AVR* (weaker signal next to *AVD*). (F) Two gene clusters on the same DNA strand, oriented tail-to-tail. The cluster on the left shows 4 *AVR*s + 1 *AVD*, whereas the cluster on the right shows putatively 5 *AVR*s + 1 *AVD*. (The signal slightly below the righthand cluster does not coincide with the locus signal; however, examination of an enlarged view of the images shows that there are 3 closely clustered *AVR* signals where the yellow arrow points.) (G–I) Representative fibers from other individuals. (G) Cell line HD11, 4 *AVR*s + 1 *AVD*. (H) HD11, 5 *AVR*s + 1 *AVD*. (I) M6, 4 *AVR*s + 1 *AVD*. (J) A schematic presentation of the organization of the *AVD* and *AVR* genes according to the control cosmid K18-233.

TABLE 1
Covariate Analysis of Variance for the *AVR* Counts

Source	df	Mean square	F	Significance
Corrected model	7	9.197	4.981	0.000
AVD count (covariate)	1	7.959	4.311	0.038
Individuals	6	9.501	5.146	0.000
Total	700			

permutation during maturation of B-cells in the germinal centers of peripheral lymphoid organs in humans and many other animals (28; for a recent review, see 29). Hypermutation of TCR α genes in germinal center T-cells has also been reported (30). In the chicken, somatic diversification of the immunoglobulins occurs in the bursa of Fabricius by intrachromosomal gene conversion instead of hypermutation (31–33). The proximity of the chicken V genes (2.4 kb as opposed to the $\geq$10 kb in humans) probably potentiates gene conversion in *cis* (34, 33). Whatever the mechanism, the cells of the immune system clearly provide a permissive environment for recombination and mutation.

The significance of the frequent recombination of the *AVD* and *AVR* genes remains unclear. Since avidin is an anti-inflammatory agent and the AVR proteins are also potentially expressed under inflammation, it may be that the somatic recombination is targeted to produce polymorphic anti-inflammatory molecules of selective advantage. Alternatively, the genes may represent merely "passengers," recombining in white blood cells only because they are prone to do so whenever the recombination machinery is functional. Thus, the hyperconversion operating on the immunoglobulin genes in B cells (and/or TCR genes in T cells) may affect the avidin gene family as well. Recent models for somatic hypermutation have suggested that the mutations are caused by the repairing of DNA double-strand breaks by (an) error-prone DNA polymerase(s) (29, 35, 36). The double-stranded breaks (DSBs) seem to be clustered in heavily transcribed regions. According to these models, all genes expressed in maturating B cells may be susceptible to hypermutation. Evidence for this suggestion has thus far been found in only one case: the transcription factor BCL-6 has been found to exhibit elevated mutation frequency in hypermutating B cells (37, 38). The avidin gene family might represent another case. According to this model, the *AVD* and *AVR* genes would be susceptible to the same enzymes that nick the immunoglobulin genes, and the forming DSBs would resolve upon repair into either gene conversions or crossovers.

This study is one of the very few in which unequal crossing-over has been attested by directly observing gene copy number expansion and contraction. Suto *et al.* (23) studied the organization of the natural killer cell inhibitory receptor (KIR) genes using fiber-FISH. They observed anomalous signal organization in one third of the fibers they examined, as opposed to those showing the "correct" gene organization. They did not describe the nature of the anomalous arrangements but discussed possible technical grounds for them. It may be that what they observed was not technical variation, but gene rearrangements instead. Furthermore, Jeffreys *et al.* (39) showed distributions similar to ours for the minisatellite MS32 repeat numbers in human sperm cells. Other studies have shown repeat number fluctuation for large, mainly RNA-encoding gene families containing hundreds of copies (40, 41). Our study has shown frequent somatic recombination within a relatively small protein-encoding gene family, and therefore gives insight into the general degree of lability in eukaryotic genomes. Finally, the chicken avidin gene family should provide an ideal model for dissecting the mechanisms of recombination, for example by using the chicken DT40 cell line variants mutated for various recombination-associated factors (24).

ACKNOWLEDGMENTS

We wish to thank Dr. Nina Horelli-Kuitunen for her invaluable help in fiber-FISH and Juha Knuuttila for his help in data collection.

REFERENCES

1. Wilchek, M., and Bayer, E. A. (1999) Foreword and introduction to the book (strept)avidin-biotin system. *Biomol. Eng.* **16,** 1–4.
2. Green, N. M. (1975) Avidin. *Adv. Protein Chem.* **29,** 85–133.
3. Tuohimaa, P., Joensuu, T., Isola, J., Keinänen, R., Kunnas, T., Niemelä, A., Pekki, A., Wallén, M., Ylikomi, T., and Kulomaa, M. (1989) Development of progestin-specific response in the chicken oviduct. *Int. J. Dev. Biol.* **33,** 125–134.
4. Stevens, L. (1991) Egg white proteins. *Comp. Biochem. Physiol. [B]* **100,** 1–9.
5. Elo, H. A., Kulomaa, M. S., and Tuohimaa, P. J. (1979) Avidin induction by tissue injury and inflammation in male and female chickens. *Comp. Biochem. Physiol.* **62B,** 237–240.
6. Elo, H. A. (1980) Occurrence of avidin-like biotin-binding capacity in various vertebrate tissues and its induction by tissue injury. *Comp. Biochem. Physiol.* **67B,** 221–224.
7. Klasing, K. C. (1998) Avian macrophages: Regulators of local and systemic immune responses. *Poultry Sci.* **77,** 983–989.
8. Keinänen, R. A., Laukkanen, M.-L., and Kulomaa, M. S. (1988) Molecular cloning of three structurally related genes for chicken avidin. *J. Steroid Biochem.* **30,** 17–21.
9. Keinänen, R. A., Wallén, M. J., Kristo, P. A., Laukkanen, M. O., Toimela, T. A., Helenius, M. A., and Kulomaa, M. S. (1994) Molecular cloning and nucleotide sequence of chicken avidin-related genes 1–5. *Eur. J. Biochem.* **220,** 615–621.
10. Ahlroth, M. K., Kola, E. H., Ewald, D., Masabanda, J., Sazanov, A., Fries, R., and Kulomaa, M. S. (2000) Characterization and chromosomal localization of the chicken avidin gene family. *Anim. Genet.* **31,** 367–375.
11. Kunnas, T. A., Wallén, M. J., and Kulomaa, M. S. (1993) Induction of chicken avidin and related mRNAs after bacterial infection. *Biochim. Biophys. Acta* **1216,** 441–445.

12. Ahlroth, M. K., Grapputo, A., Laitinen, O. H., and Kulomaa, M. S. (2001) Sequence features and evolutionary mechanisms in the chicken avidin gene family. *Biochem. Biophys. Res. Comm.* **285,** 734–741.
13. Li, W.-H. (1997) Molecular evolution. Sinauer Associates, Inc., Sunderland, Massachusetts.
14. Perry, J., and Ashworth, A. (1999) Evolutionary rate of a gene affected by chromosomal position. *Curr. Biol.* **9,** 987–989.
15. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983) The double-strand-break repair model for recombination. *Cell* **33,** 25–35.
16. Martinsohn, J. T., Sousa, A. B., Guethlein, L. A., and Howard, J. C. (1999) The gene conversion hypothesis of MHC evolution: A review. *Immunogenetics* **50,** 168–200.
17. Heiskanen, M., Karhu, R., Hellsten, E., Peltonen, L., Kallioniemi, O. P., and Palotie, A. (1994) High resolution mapping using fluorescence *in situ* hybridization to extended DNA fibers prepared from agarose-embedded cells. *Biotechniques* **17,** 928–929, 932—933.
18. Heiskanen, M., Hellsten, E., Kallioniemi, O. P., Mäkelä, T. P., Alitalo, K., Peltonen, L., and Palotie, A. (1995) Visual mapping by fiber-FISH. *Genomics* **30,** 31–36.
19. Heiskanen, M., Kallioniemi, O., and Palotie, A. (1996) Fiber-FISH: Experiences and a refined protocol. *Genet. Anal.* **12,** 179–184.
20. Hellsten, E., Vesa, J., Heiskanen, M., Mäkelä, T. P., Järvelä, I., Cowell, J. K., Mead, S., Alitalo, K., Palotie, A., and Peltonen, L. (1995) Identification of YAC clones for human chromosome 1p32 and physical mapping of the infantile neuronal ceroid lipofuscinosis (INCL) locus. *Genomics* **25,** 404–412.
21. Horelli-Kuitunen, N., Aaltonen, J., Yaspo, M. L., Eeva, M., Wessman, M., Peltonen, L., and Palotie, A. (1999) Mapping ESTs by fiber-FISH. *Genome Res.* **9,** 62–71.
22. Florijn, R. J., van de Rijke, F. M., Vrolijk, H., Blonden, L. A., Hofker, M. H., den Dunnen, J. T., Tanke, H. J., van Ommen, G. J., and Raap, A. K. (1996) Exon mapping by fiber-FISH or LR-PCR. *Genomics* **38,** 277–282.
23. Suto, Y., Ishikawa, Y., Kasahara, M., Kasai, F., Yabe, T., Akaza, T., and Juji, T. (1998) Gene arrangement of the killer cell inhibitory receptor family on human chromosome 19q13.4 detected by fiber-FISH. *Immunogenetics* **48,** 235–241.
24. Sonoda, E., Takata, M., Yamashita, Y. M., Morrison, C., and Takeda, S. (2001) Homologous recombination in vertebrate cells. *Proc. Natl. Acad. Sci. USA* **98,** 8388–8394.
25. Beug, H., von Kirchbach, A., Döderlein, G., Conscience, J-F., and Graf, T. (1979) Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation. *Cell* **18,** 375–390.
26. Högstrand, K., and Böhme, J. (1994) A determination of the frequency of gene conversion in unmanipulated mouse sperm. *Proc. Natl. Acad. Sci. USA* **91,** 9921–9925.
27. Roitt, I. M. (1997) Essential Immunology. Blackwell Science. Oxford, UK.
28. Weigert, M. G., Cesari, I. M., Yonkovich, S. J., and Cohn, M. (1970) Variability in the lambda light chain sequences of mouse antibody. *Nature* **228,** 1045–1047.
29. Jacobs, H., and Bross, L. (2001) Toward an understanding of somatic hypermutation. *Curr. Opin. Immunol.* **13,** 208–218.
30. Zheng, B., Xue, W., and Kelsoe, G. (1994) Locus-specific somatic hypermutation in germinal centre T cells. *Nature* **372,** 556–559.
31. Reynaud, C. A., Anquez, V., Grimal, H., and Weill, J. C. (1987) A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell* **48,** 379–388.
32. Maizels, N. (1987) Diversity achieved by diverse mechanisms: Gene conversion in developing B cells of the chicken. *Cell* **48,** 359–360.
33. Carlson, L. M., McCormack, W. T., Postema, C. E., Humphries, E. H., and Thompson, C. B. (1990) Templated insertions in the rearranged chicken IgL V gene segment arise by intrachromosomal gene conversion. *Genes Dev.* **4,** 536–547.
34. Storb, U. (1996) The molecular basis of somatic hypermutation of immunoglobulin genes. *Curr. Opin. Immunol.* **8,** 206–214.
35. Dominguez, O., Ruiz, J. F., Lain de Lera, T., Garcia-Diaz, M., Gonzalez, M. A., Kirchhoff, T., Martinez, A. C., Bernad, A., and Blanco, L. (2000) DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J.* **19,** 1731–1742.
36. Tissier, A., McDonald, J. P., Frank, E. G., and Woodgate, R. (2000) Poliota, a remarkably error-prone human DNA polymerase. *Genes Dev.* **14,** 1642–1650.
37. Shen, H. M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998) Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* **280,** 1750–1752.
38. Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R. S., Klein, U., Kuppers, R., Rajewsky, K., and Dalla-Favera, R. (1998) BCL-6 mutations in normal germinal center B cells: Evidence of somatic hypermutation acting outside Ig loci. *Proc. Natl. Acad. Sci. USA* **95,** 11816–11821.
39. Jeffreys, A. J., Neil, D. L., and Neumann, R. (1998) Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J.* **17,** 4147–4157.
40. Warburton, D., Atwood, K. C., and Henderson, A. S. (1976) Variation in the number of genes for rRNA among human acrocentric chromosomes: Correlation with frequency of satellite association. *Cytogenet. Cell. Genet.* **17,** 221–230.
41. Krystal, M., D'Eustachio, P., Ruddle, F. H., and Arnheim, N. (1981) Human nucleolus organizers on nonhomologous chromosomes can share the same ribosomal gene variants. *Proc. Natl. Acad. Sci. USA* **78,** 5744–5748.

IV

Chicken avidin-related proteins (AVRs) show altered biotin-binding and physico-chemical properties as compared to avidin

by

Laitinen, O. H., Hytönen, V., Ahlroth, M. K., Pentikäinen, O. T., Gallagher, C., Nordlund, H., Ovod, V., Marttila, A., Porkka, E., Heino, S., Johnson, M. S., Airenne, K. & Kulomaa, M. S.

(submitted manuscript)

https://döi.örg/1O.1O42/O264-6O21:363O6O9