NNo8701, 2500

Stellingen:

Behorende bij het proefschrift

Analysis of the chicken genome: mapping of monogenic traits

1. Het ontwikkelen van microsatelietmerkers uit cDNA sequenties is een goede manier om tegelijkertijd een bijdrage te leveren aan zowel de genetische koppelingskaart als de vergelijkende genenkaart.

   Dit proefschrift

2. Bulked segregant analyse met behulp van fluorescerende microsatelietmerkers, is een betrouwbare methode om de genetische locatie van enkelvoudige kenmerken efficiënt mee te bepalen.

   Dit proefschrift

3 Terwijl bulked segregant analyse met behulp van fluorescerende microsatelietmerkers tot een aanzienlijke kostenbeparing leidt in vergelijking met een studie waarbij nakomelingen individueel getypeerd worden, gaat dit niet ten koste van de genetische afstand waarop gekoppelde merkers gevonden zullen worden.

   Dit proefschrift

4. Doordat de genomen van mens en kip ondanks hun grote evolutionaire afstand zoveel geconserveerde gebieden vertonen, kunnen vergelijkende genenkaarten tussen kip en mens gebruikt worden bij het opsporen van kandidaatgenen voor bepaalde genetische kenmerken bij de kip.

   Dit proefschrift

5. Hoewel het waarschijnlijk is dat autosomale dwerggroei (*adw*) bij de kip in plaats van een recessief een incompleet recessief kenmerk is, zal dit, omdat gewichtsverschillen tussen heterozygote (*Adwadw*) en homozygote (*AdwAdw*) dieren op een leeftijd van 6 weken nog gering zijn, niet van invloed op het eventuele gebruik van het *adw* gen in slachtkippelijnen zijn.

   Dit proefschrift

6. De zogenaamde "positional candidate gene approach" zal met het vorderen van het humane genoom project in belang toenemen.

7. Afspraken over het gebruik van nomenclatuur is van essentieel belang voor het op een efficiënte manier ontwikkelen van een betrouwbare vergelijkende genenkaart.

8. Hoewel het patentrecht oorspronkelijk bedoeld was om wetenschappelijke vooruitgang te bevorderen werkt het op dit moment in veel gevallen averechts.

9. Met de uitspraak van minister Borst om uitsluitend diagnostische DNA-tests voor behandelbare erfelijke ziekten op de markt toe te laten (*VSN contactmagazine augustus 1998*), houdt zij geen rekening met het feit dat vroegtijdige kennis omtrent het al dan niet hebben van een onbehandelbare erfelijke ziekte vaak tot andere en beter overwogen keuzes kan leiden.

10. In de grondwet van een ware democratie behoort de mogelijkheid tot het houden van een referendum opgenomen te zijn.

11. Door het antivaccinatie beleid wat door de Europese Unie bij varkenspest gehanteerd wordt, krijgt het begrip varkensstapel een geheel andere betekenis.

12. Mensen met een GSM telefoon zijn niet mobiel maar aangelijnd.

13. De uitdrukking "zinloos geweld" impliceert ten onrechte dat geweld ook zinvol kan zijn.

14. Helaas wordt de waarde van persoonlijke gezondheid pas juist ingeschat als deze niet geheel tot expressie komt.

Carolien Ruyter-Spira
Wageningen, 13 oktober 1998

Analysis of the Chicken Genome:

Mapping of Monogenic Traits.

Promotor:	dr. ir. E.W. Brascamp
	hoogleraar in de veefokkerij
Co-promotoren:	dr. ir. M.A.M. Groenen
	universitair docent bij de leerstoelgroep Fokkerij en Genetica
	dr. M. Georges
	hoogleraar aan de Université de Liegè, België

Analysis of the Chicken Genome:

Mapping of Monogenic Traits.

Carolien P. Ruyter-Spira

Proefschrift

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van de Landbouwuniversiteit Wageningen,
dr.C.M. Karssen,
in het openbaar te verdedigen
op dinsdag 13 oktober 1998
des namiddags te vier uur in de Aula.

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Ruyter-Spira, Carolien P.

Analysis of the chicken genome: mapping of monogenic traits / Carolien P. Ruyter-Spira

Doctoral thesis, Wageningen Agricultural University. - With ref. - With summary in Dutch.

ISBN: 90 - 5485 - 959 - 8

Subject headings: genetic linkage map; comparative mapping; chicken; monogenic traits

Cover page: "Dominant White" designed by Dirk Ruyter.

Abstract

Ruyter-Spira, C.P., 1998. Analysis of the chicken genome: mapping of monogenic traits. Doctoral thesis, Wageningen Agricultural University, P.O. Box 338, 6700 AH Wageningen, The Netherlands.

The development of genetic linkage maps in farm animals is progressing rapidly. Linkage maps can be used to identify genetic loci responsible for genetic variation in traits of economic importance. The ultimate goal is to find the underlying genes involved in these traits. To achieve this, the so called positional candidate gene approach is gaining in importance. This approach is based on the genetic localization of a trait using genetic linkage analysis in livestock species. Subsequent comparative mapping of the trait locus with the gene-rich maps of the human and the mouse may reveal candidate genes for the trait in question. For the construction of comparative maps the genetic localization of many genes needs to be determined. In this thesis, the development of highly informative microsatellite markers from expressed sequences, which will contribute to the (comparative) genetic linkage map of the chicken, is described. In addition to this, the genomic localization of two monogenic traits are determined using a technique called bulked segregant analysis. Finally, candidate genes for both traits are evaluated.

Ter nagedachtenis aan mijn moeder

"Science is a match that man has just gotta light. He thought he was in a room-in moments of devotion, a temple-and that his light would be reflected from and display walls inscribed with wonderful secrets and pillars carved with philosophical systems wrought into harmony. It is a curious sensation, now that the preliminary sputter is over and the flame burns up clear, to see his hands lit and just a glimpse of himself and the patch he stands on visible, and around him, in place of all that human comfort and beauty he anticipated-darkness still."

H.G.Wells, 1891

Voorwoord.

In het boekje dat nu voor u ligt staat het onderzoek beschreven dat ik in de periode van september 1994 tot september 1997 bij de leerstoelgroep fokkerij en genetica heb uitgevoerd. Door deze periode chronologisch te doorlopen krijg ik de kans om de vele mensen die mij gesteund en geholpen hebben te bedanken.

Allereerst mijn Dirk, die natuurlijk niet alleen in deze periode mijn steun en toeverlaat is, maar op wie ik altijd kan vertrouwen. Ik vind het moeilijk de juiste woorden te vinden om mijn dankbaarheid in uit te drukken en geloof ook niet dat dat ooit mogelijk is. Jij hebt zoveel voor me gedaan, als collega op het lab, als mijn beste vriend op het thuisfront en nu ook als grafisch vormgever van dit proefschrift. Fantastisch! Vervolgens wil ik Martien Groenen bedanken voor het geven van zowel de mogelijkheid als de begeleiding om dit onderzoek uit te kunnen voeren. Wat was ik in de wolken toen je mij een baan binnen het caseïne onderzoek aanbood! Ook gaat mijn dank uit naar alle mensen die het onderzoek praktisch ondersteund hebben en als co-auteur op de publikaties vermeld staan.

Dan nu weer terug naar 1994. Het screenen van de cDNA bank wordt gestart en leidt al snel tot een publikatie. Helaas gaat het met de GDRDA methode minder voorspoedig, maar door snel een alternatieve strategie te kiezen komt het Eureka moment, waar iedere onderzoeker zijn voldoening en motivatie uit haalt, toch nog, wanneer we met succes het kenmerk "dominant wit" weten te mappen. Ik wil de leden van mijn begeleidingscommissie; Pim Zabel, mijn promotor Pim Brascamp en co-promotoren Michel Georges en Martien Goenen bedanken voor hun steun en adviezen bij het nemen van dergelijke besluiten en het kritisch lezen van alle concept publicaties. Vervolgens breekt in november 1995 mijn minder mobiele "kruk" periode aan. In zo'n tijd leer je je collega's pas echt goed kennen. Richard, Maud, Rosilde, Tino, 2x Jan, Frits, Tineke, Beja, Pieter, Saskia, Jos, Esther, Paul, Pim, Antoine, Peter, Annelies, Jos, Martien, hartstikke bedankt voor het dragen van al die kopjes koffie en voor het zorgen voor een fantastische sfeer op het lab. Jan van der Poel, het was echt geweldig dat je, samen met Dirk-Jan en Ad, het onderzoek tijdens mijn verlofperiode voort hebt gezet. Eénmaal teruggekomen konden de krukken gelukkig snel de deur weer uit, en was het weer een heerlijke "labtijd" waarin de nadruk op het kenmerk autosomale dwerggroei kwam te liggen. Aliene, kleine meid, het was weleens afzien als ik weer

eens te laat thuis kwam voor een voeding doordat een experiment was uitgelopen. Bedankt voor het wachten. Nienke, bedankt voor het merken en draaien van de eitjes. Best leuk op mama's werk hè? Henk Schipper, bedankt voor je technische adviezen bij het isoleren van de embryo's en Torik Ayoubi voor de inspirerende samenwerking op afstand. In deze tijd werd ook het ABC team gevormd. Ant en Birgitte, jullie chocola en gevleugelde woorden werkten zeer stimulerend. Tenslotte zat het praktische werk erop en was het tijd om dit boekje te schrijven. Richard Spelman en Henk Bovenhuis, bedankt voor jullie bruikbare suggesties bij het schrijven van de discussie.
En toen werd jij ernstig ziek Mam. We deden beiden ons best om januari te halen; ik zou dan promoveren en jij zou mij vanuit de zaal aanmoedigen. We hebben het beiden niet gered. Ik mis je elke dag! Jij blijft mijn grote voorbeeld, en hebt mij tot het eind aan toe gestimuleerd om dit boekje toch af te maken.....En hier ligt het dan.

Carolien

Contents

Chapter 1	General introduction	9
	1.1 Genome analysis in farm animals	11
	1.2 Analysis of the chicken genome	12
	1.3 Mapping of monogenic traits in farm animals	14
	1.4 Comparative mapping in the chicken	15
	1.5 Scope of this thesis	21
Chapter 2	Development and mapping of polymorphic microsatellite markers derived from a chicken brain cDNA library.	29
Chapter 3	Developing microsatellite markers from cDNA: a tool for adding expressed sequences to the genetic linkage map of the chicken.	43
Chapter 4	Bulked segregant analysis using microsatellites: mapping of the dominant white locus in the chicken.	55
Chapter 5	The *HMGI-C* gene is a likely candidate for the autosomal dwarf locus in the chicken.	67
Chapter 6	Chromosomal assignment of the gene encoding *c-KIT* in the chicken.	85
Chapter 7	Nucleotide sequence of the chicken *HMGI-C* cDNA and expression of the *HMGI-C* and *IGF1* genes in autosomal dwarf chicken embryos.	93
Chapter 8	General discussion	103
Summary		
Samenvatting		
Curriculum vitae		

Chapter 1

General introduction.

General introduction.

1.1 Genome analysis in farm animals.

Since the early domestication of animals, man has been exploiting the fact that many traits in livestock are genetic in nature. Selection was direct and straightforward and without being aware of the underlying mechanisms, experience has provided the knowledge that selecting superior animals as parents for the next generation can lead to an overall better herd.

Only in the beginning of the twentieth century, Morgan and colleagues demonstrated in the fruitfly that some traits or phenotypes were inherited together reflecting the concept of linkage (Sturtevant 1913, Morgan *et al.* 1915). Linkage describes the tendency of genes to be inherited together as a result of their location on the same chromosome. They also showed that, due to recombination during meiosis, certain groups of traits originating from a single chromosome are not always inherited together. The closer, or more linked, genes are on a particular chromosome, the smaller the probability they will separate during meiosis. This knowledge has led to the definition of genetic distance between traits on chromosomes and resulted in the development of the first genetic linkage maps. A genetic linkage map shows the arrangement of genes, phenotypes and markers along the length of the chromosomes and is based on the analysis of the frequency of co-inheritance of genetic markers. The concept of linkage also initiated the search for markers that could be applied for predicting certain production traits, and thus could serve in marker assisted selection. In the 1950s and 1960s quite some effort was focussed on exploiting blood groups of livestock species as genetic markers in segregation analysis experiments, however, due to the limited number of different blood groups, this only allows the screening of 10% of the total genome.

The advent of recombinant DNA technology in the 1970s, and the development of highly abundant polymorphic microsatellite markers (Weber & May 1989; Smeets *et al.* 1989) that cover the whole genome and permit segregation analysis, brought about a novel and powerful approach for creating advanced genetic linkage maps. At this moment, dense genome maps have been constructed for a wide variety of species,

including farm animals, the human map being the most advanced. While the human genome project is mainly focussed on the identification of the complete set of human genes and their function, the agricultural industry is basically interested in identifying genes underlying economically important traits in plant and livestock species. As already mentioned, traits can be localized in the genome through associations between performance and the inheritance of marker genes in a suitable pedigree. One of the advantages in animal genetics is that many years of selection pressure has led to the existence of divergent lines which can be applied for constructing large families in which these interesting traits are segregating. Thus, the genetic variation underlying monogenic traits as well as polygenic traits can be unravelled leading to a better understanding of the mechanisms of various physiological processes.

1.2 Analysis in the chicken genome.

Karyotype and cytogenetic map

The chicken genome consists of 38 pairs of autosomes and two sex chromosomes, Z and W, the female being the hetero gametic sex. The autosomes can be divided in eight macrochromosomes and 30 microchromosomes (Yamashina 1944). The physical size of the chicken genome is small relative to other domesticated animals, approximately 1.2×10^9 bases (Olofsson & Bernardi 1983). However, the genetic length is estimated at least 3750 cM (Groenen *et al.* 1998) which resembles the length of the human genome map. This is in agreement with the findings of Rodionov *et al.* (1992) who suggested that recombination rates in chickens are approximately 3 times higher than in mammals. This implies that once a quantitative trait locus (QTL) has been mapped to a certain chromosomal region, the actual size in base pairs that has to be examined to identify the gene itself will be (on average) threefold smaller than in mammals.

Cytogenetic maps show the locations of genes on individual chromosomes and provide landmarks for the assignment and orientation of linkage groups to specific chromosomes. Until now, ten linkage groups have been assigned to the first eight macro chromosomes, chromosome 16, and the Z chromosome by *in situ* hybridization. The microchromosomes are cytogenetically indistinguishable using current banding techniques and some are not expected to exhibit recombination over most of their

length and will be detected as single point linkage groups (Rodionov *et al.* 1992). Future development of microchromosome specific probes, which can be used in fluorescent *in situ* cohybridization experiments, will further facilitate the alignment of both the physical and genetic map of the chicken.

Genetic linkage maps

The first genetic linkage map of the chicken, showing the locations of classical mutations leading to abnormal phenotypes, was published in 1936 by Hutt. The last update of this classical map, showing 43 loci on five linkage groups is by Bitgood & Somes (1993). Two of the linkage groups have been assigned to chromosome 1 and the Z chromosome respectively.

Besides the classical genetic linkage map, three different molecular genetic linkage maps containing molecular polymorphic markers have been developed independently, each map being based on a different experimental cross.

The Compton reference family (Bumstead & Palyga 1992) is a backcross (BC) between two individual inbred White Leghorn chickens which differ in their susceptibility to a number of diseases. A total number of 445 loci have been mapped on 38 linkage groups and 95% of the loci are linked to at least one other locus. The total length is 3800 cM with an average mapping interval of 9 cM. Because the BC was made by mating two F1 females to a single male, no Z chromosome-linked markers could be mapped.

Crittenden *et al.*(1993) also developed a BC population called the East Lansing (EL) reference family. In this cross the F1 is the male. To produce the F1, an inbred Red Jungle Fowl was crossed with a highly inbred White Leghorn, on the presumption that this interspecific cross would yield more polymorphisms than crosses of two chicken lines. Until now, 728 loci have been mapped on 38 linkage groups with an average mapping interval of 5 cM. Both the Compton and the EL reference population have been used as the two standard panels for developing the chicken molecular linkage map.

A third comprehensive genetic linkage map, solely based on microsatellite markers, has been developed by Groenen *et al.*(1998). Because a large population consisting of 10 full sib families of a cross between two extreme commercial broiler lines, with a

total of 476 individuals (F1 and F2) has been used, genetic distances can be accurately estimated. The map comprises 480 microsatellite markers, of which 472 have been assigned to 30 different linkage groups and 8 were found to be unlinked. The markers cover in total 3145 cM with an estimated genome coverage of 90-95%. Although the average marker spacing of this map is only 7 cM, there are still several regions, where the distance between two adjacent markers is considerably larger than the preferred maximum distance of 20 cM necessary for QTL detection. Some of the microchromosomes are not represented in the current map. A possible explanation for the poor coverage of the microchromosomes with microsatellites comes from the observation of Primmer *et al.* (1997) who, by *in situ* labelling, showed that these chromosomes have a relatively low concentration of CA microsatellites.

1.3 Mapping of monogenic traits in farm animals.

Once developed, saturated genetic linkage maps will greatly facilitate the search for disease and trait loci via linkage analysis approaches, that test for cosegregation within families with a random marker locus. In farm animals, the number of traits that have been mapped is increasing rapidly. Because multifactorial traits are more difficult to analyse, the majority of traits localized so far are involved in monogenic Mendelian disorders or phenotypes with simple patterns of inheritance. Examples of traits of economic importance are the malignant hyperthermia locus in the pig (Davies *et al.* 1988), the Booroola fecundity gene in sheep (Montgomery *et al.* 1994), the double muscling locus in cattle (Charlier *et al.* 1995) and several genes controlling coat colour in various species (Johansson *et al.* 1992, Johansson *et al.* 1994, Klungland *et al.* 1995, Charlier *et al.* 1996, Mariani *et al.* 1996).

As already mentioned in the previous paragraph, in the chicken 43 mutant phenotypes based on single loci have been assigned to the classical genetic linkage map (Bitgood & Somes 1993). Because most of these mutations do not segregate in the International Reference families, these maps can not be aligned without independent segregation analysis of the mutant phenotypes using the markers present on the molecular genetic linkage map. Until now, four monogenic traits have been mapped on the molecular genetic linkage map of the chicken: sex-linked dwarfism (Levin *et al.* 1993), dermal melanin inhibitor (Levin *et al.* 1994), dominant white

plumage (Ruyter-Spira *et al.* 1997) and autosomal dwarfism (Ruyter-Spira *et al.* 1998).

1.4 Comparative mapping in the chicken.

Positional candidate gene approach

It has been acknowledged for a long time that homologous genes between different vertebrates and invertebrates have been conserved to a large extent over many million years of evolution (O'Brien *et al.* 1993). Not only do individual gene sequences show conservation between species, also large chromosomal segments harbor the same homologous genes. This information can be transformed into comparative maps, which display the chromosomal location of homologous genes in different species and highlight genetic segments that are conserved during evolution (Eppig 1996). Comparative maps enable animal geneticists to exploit the wealth of information being offered by the human genome mapping community in the search for candidate genes for disease phenotypes and other (polygenic) traits. An attractive approach to cloning genes underlying traits from these comparative maps, is the positional candidate gene approach (Collins 1995). This strategy relies on a combination of mapping a trait to a chromosomal subregion using linkage analysis, followed by a survey of the conserved interval in the human or mouse so as to reveal the location of attractive candidates (Figure 1).

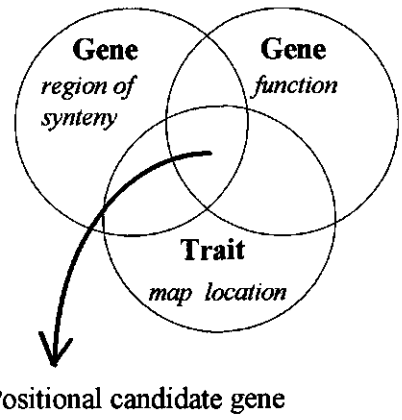


Figure 1. Schematic representation of the positional candidate gene approach. A positional candidate gene for a certain trait is pinpointed based on its genetic location, comparative mapping and the knowledge about physiological functions of the genes present in the syntenic region. (Adapted from Sheffield *et al.* 1995)

As the human transcript map is improving dramatically, the positional candidate approach will become an alternative for blindly subcloning tremendously large regions of the animal genome in search for candidate genes as is being done using positional cloning, or the testing of numerous candidate genes as is being done using the candidate gene approach.

Examples of successful applications of both the positional candidate gene approach and the candidate gene approach in livestock species are depicted in table 1.

Additionally, comparative maps between livestock species and man, can also be in favour of the human geneticist. Because of the possibility to produce large segregating families, animal models open the door to dissecting more complex traits like growth, weight and fat distribution (Andersson *et al.* 1994). Also the availability of many well described mutant phenotypes in agricultural species, especially chicken (Crawford 1990), may provide insight in various human diseases or developmental disorders. Because of the easy access to early embryo's, the chicken provides a very good model in this respect.

Table 1. Monogenic traits with known molecular basis in farm animals.

Species	Trait	Gene coding for	Reference
Chicken	Sex linked dwarfism+	growth hormone receptor	Huang *et al.* 1993
	Restricted ovulator+	oocyte vitellogenesis receptor	Bujo *et al.* 1995
	Nanomelia+	aggrecan	Li *et al.* 1993
	Henny feathering+	aromatase	Matsumine *et al.* 1991
	Riboflavinuria+	riboflavin-binding protein	Maclachlan *et al.* 1993
Cow	Extension coat colour*	melanocyte stimulating hormone receptor	Klungland *et al.* 1995
	Double muscling*	myostatin gene	Grobet *at al.*1997 Kambadur *et al.* 1997
	Glycogen storage disease+	myophosphorylase	Tsujino *et al.* 1996
Pig	Malignant hyperthermia*	calcium release channel gene	Fujii *et al.* 1991
	Dominant white*	mast/stem cell growth factor receptor (KIT)	Johansson-Moller *et al.* 1996

* Causative gene identified using positional candidate gene approach.
\+ Causative gene identified using candidate gene approach.

Comparative map of the chicken

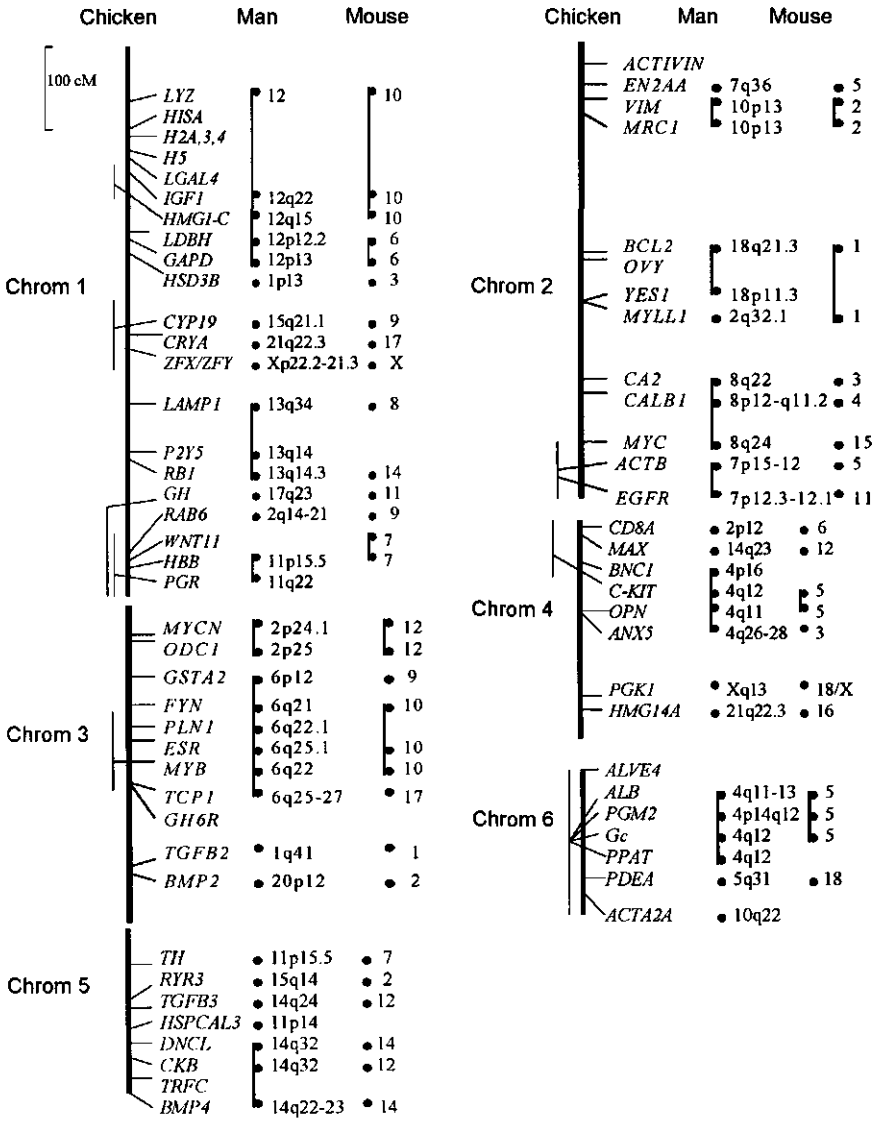
Despite the divergence of birds and mammals by over 300 million years of evolution, evidence for the existence of conserved chromosomal segments has been provided by various groups (Palmer & Jones 1986; Burt *et al.* 1995; Klein *et al.* 1996; Smith *et al.* 1996; Heltemes *et al.* 1997; Jones *et al.* 1997; Smith *et al.* 1997).

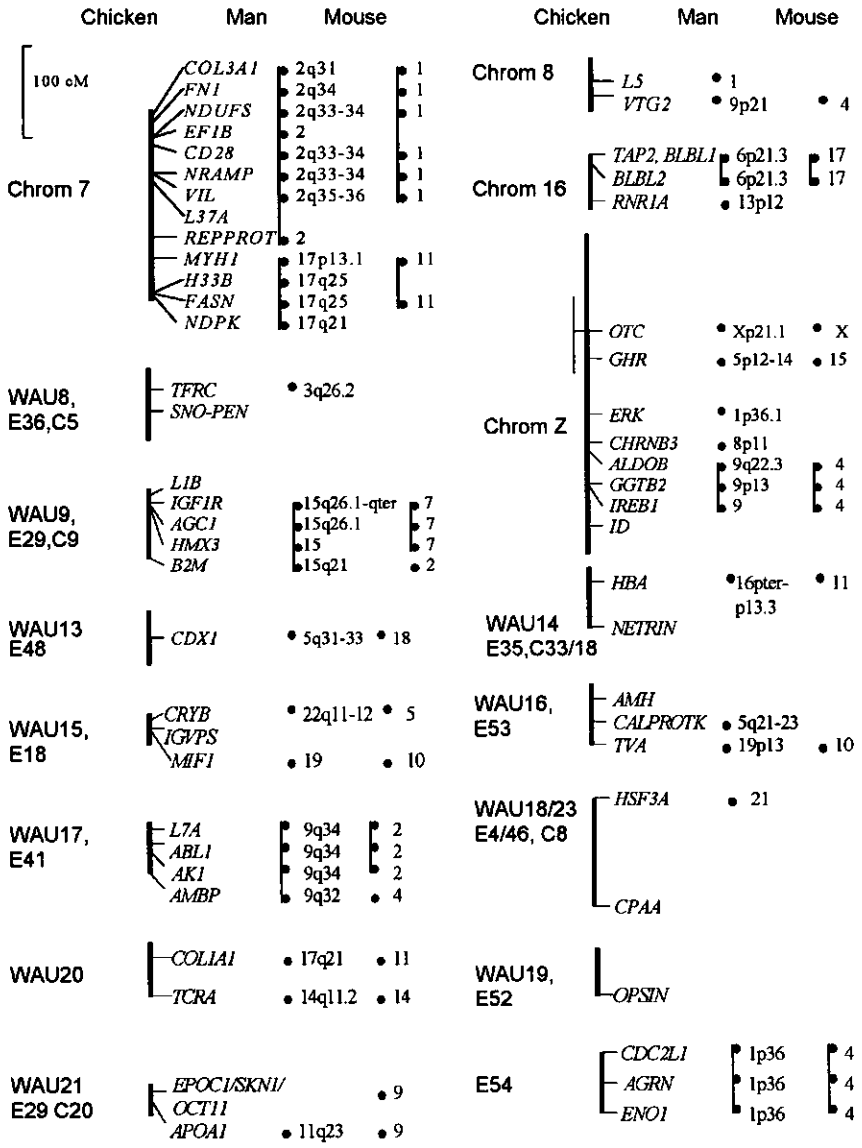
In figure 2 a preliminary comparative map of the chicken is presented. This map is based on 128 genes, the so-called type I markers as defined by O'Brien & Graves (1991), which have been mapped on the East Lansing and Compton reference families and the Wageningen resource population. Some assignments have also been made by *in situ* hybridization. Because the comparative map is the result of the alignment of the three different linkage maps mentioned above, distances shown may be inaccurate. In total, 105 human homologues and 87 murine homologues have been mapped resulting in 19 human-chicken and 16 mouse-chicken conserved segments, some of them showing rearrangements of gene order. Most striking in the chicken is the extended length of some of the conserved segments. In the pig, Johansson *et al.* (1995) report a similar number of conserved segments when comparing 83 coding loci with the human and mouse maps, some of them also showing intra chromosomal rearrangements. Also in the pig-human comparative map, the average length of the conserved segments is quite large, however these species have 'only' been separated by 65 million years of evolution.

One example of a large conserved chromosomal segment covering approximately 53-88 cM, is located on chicken microchromosome WAU9/E29/C9. This segment is showing conserved synteny with human chromosome 15 (Jones *et al.* 1997). Their results suggest that the region of conserved synteny extends to the entire length of this microchromosome. This would support the hypothesis of Takagi & Sasaki (1974), that the increase in the number of chicken chromosomes during evolution has occurred mostly through a process of chromosome fission. Therefore, some microchromosomes may represent conserved segments consisting of genes kept together during evolution because of functional reasons. This makes the chicken an interesting model to elucidate functional relationships between different genes. Another example supporting this hypothesis is the conserved gene cluster present on chicken microchromosome WAU17/E41 and human chromosome 9. Still, conclusions about conserved clusters

between birds and mammals must be drawn with care. Some syntenic regions on the present comparative map of the chicken are only based on two homologous genes and in other cases homologues are only found in one other species. In addition, because of the existence of highly conserved gene families, sequence similarity does not provide indisputable evidence of homology (Andersson *et al.* 1996). However, since comparative mapping data are accumulating rapidly, more evidence will be available in the near future.

Figure 2. Preliminary comparative genome map of the chicken. Only linkage groups containing mapped genes are shown. Information on the position of the chicken type I loci has been derived from the world wide web URL: http://ri.bbsrc.ac.uk/cgi-bin/anubis, of the Roslin Institute. The human and murine gene localizations were obtained from the world wide web URL: http://ncbi.nlm.nih.gov/repository/OMIM/genemap, Online Mendelian Inheritance in Man (OMIM) of the National Institute of Health (NIH). (⊢) indicates the location of genes which have been assigned by *in situ* hybridization. (●) indicates the gene homolog in human or mouse. (|) indicates conserved chromosomal segments. Abbreviations belonging to the mapped chicken genes are in appendix 1.




Chicken
Man
Mouse
100 cM
Chrom 7
COL3A1
FN1
NDUFS
EF1B
CD28
NRAMP
VIL
L37A
REPPROT
MYH1
H33B
FASN
NDPK
2q31
2q34
2q33-34
2
2q33-34
2q33-34
2q35-36
2
17p13.1
17q25
17q25
17q21
1
1
1
1
1
1
11
11
WAU8, E36,C5
TFRC
SNO-PEN
3q26.2
WAU9, E29,C9
L1B
IGF1R
AGC1
HMX3
B2M
15q26.1-qter
15q26.1
15
15q21
7
7
7
2
WAU13 E48
CDX1
5q31-33
18
WAU15, E18
CRYB
IGVPS
MIF1
22q11-12
19
5
10
WAU17, E41
L7A
ABL1
AK1
AMBP
9q34
9q34
9q34
9q32
2
2
2
4
WAU20
COL1A1
TCRA
17q21
14q11.2
11
14
WAU21 E29 C20
EPOC1/SKN1/ OCT11
APOA1
11q23
9
9
Chrom 8
L5
VTG2
1
9p21
4
Chrom 16
TAP2, BLBL1
BLBL2
RNR1A
6p21.3
6p21.3
13p12
17
17
Chrom Z
OTC
GHR
ERK
CHRNB3
ALDOB
GGTB2
IREB1
ID
Xp21.1
5p12-14
1p36.1
8p11
9q22.3
9p13
9
X
15
4
4
4
WAU14 E35,C33/18
HBA
NETRIN
16pter-p13.3
11
WAU16, E53
AMH
CALPROTK
TVA
5q21-23
19p13
10
WAU18/23 E4/46, C8
HSF3A
CPAA
21
WAU19, E52
OPSIN
E54
CDC2L1
AGRN
ENO1
1p36
1p36
1p36
4
4
4

1.5 Scope of this thesis

The aim of the investigations described in this thesis can be divided in three subthemes, all related to the positional candidate gene approach.

The development and mapping of type I genetic markers

Because the positional candidate gene approach has a high potential in cloning disease or trait genes, the development of markers representing genes is needed to improve the comparative map of the chicken. The reason for the limited number of type I markers at this time is that, due to functional sequence conservation, genes exhibit less DNA polymorphisms than noncoding regions of the genome. Another reason is that most type I markers which are developed until now are based on labour intensive restrictive fragment length polymorphism (RFLP) markers. Therefore, the first goal of this thesis was to add a number of type I loci, based on microsatellite sequences, to the genetic linkage map of the chicken. This is described in chapter two and three.

The development of an efficient method for mapping monogenic traits.

As segregation analysis experiments are extremely time consuming due to the large amount of genotypings needed, Michelmore *et al.* (1991) developed an alternative approach, called bulked segregant analysis (BSA), that circumvents this problem. The method involves the comparison of two pooled DNA samples of individuals from a segregating population originating from a single cross. Within each pool the individuals are identical for the trait, locus or gene of interest, but are arbitrary for all other genes (Figure 3). Both pools are analysed to identify markers that show differences in allele frequency, indicating that these markers are linked to the locus determining the trait used to construct the pools. Since the analysis of pools can replace the analysis of large numbers of individual segregating progeny, considerable savings are possible. RFLP and RAPD markers closer than 15 cM to the trait of interest are likely to be detected, whereas the limit of detection was estimated at 25 cM on either side of the target locus.

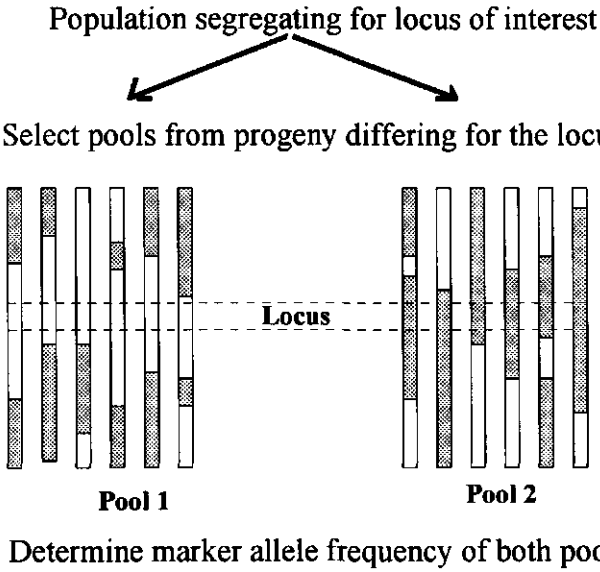


Figure 3. Schematic representation of bulked segregant analysis. A population segregating for a certain trait or locus is created. Individuals from the F1 progeny showing opposite alleles for the locus are selected to prepare DNA-pool 1 and 2. Finally a total genome scan using molecular markers is performed on both DNA-pools. A difference in marker allele frequency between both pools is an indicative of linkage between locus and marker.

Because the use of fluorescently labelled microsatellite primers can enhance the accuracy of estimated allelic frequencies of pooled DNA samples, and because the majority of the markers present on the genetic linkage map of the chicken are microsatellite markers, the second goal of this study was to investigate the use of these type of markers in combination with the BSA approach in order to map monogenic traits in the chicken. The localization of the dominant white locus is described in chapter four, whereas chapter five deals with the genetic localization of the autosomal dwarf locus.

The identification of candidate genes

The next step to be taken is the identification and verification of candidate genes for both traits. Especially autosomal dwarfism displays a very interesting phenotype that is highly suited for further exploration, because the underlying gene carrying the causative mutation is involved in the genetic pathway regulating growth and development. These results are discussed in chapter five and seven. In addition to this, a candidate gene for the dominant white phenotype is evaluated in chapter six. Finally,

the thesis ends in chapter eight in a general discussion to evaluate the results that have been achieved in this study. In this chapter more potential candidate genes for both traits are proposed, and alternative approaches for the identification of genes underlying monogenic traits are discussed.

References.

Andersson, L., Haley, C.S., Ellegren, H., Knott, S.A., Johansson, M., Andersson, K., Andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Hakansson, J. & Lundstrom, K. 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**, pp. 1771-1774.

Andersson *et al.* 1996. Comparative Genome organization of vertebrates. *Mammalian Genome* **7**, pp. 717-734.

Bitgood, J.J. & Somes, R.G. 1993. Gene map of the chicken (Gallus gallus or G. domesticus) pp. 4332-4342 in: *Genetic maps*, 6th ed. S.J. O'Brien, ed. Cold Spring Harbor Laboratory Press, Plainview, NY.

Bujo, H., Yamamoto, T., Hayashi, K., Hermann, M., Nipf, J. & Schneider, W.J. 1995. Mutant oocytic low density lipoprotein receptor gene family member causes atherosclerosis and female sterility. *Proc. Nat. Acad. Sci. USA* **92**, pp. 9905-9909.

Burnstead, N., & Palyga, J. 1992. A preliminary linkage map of the chicken genome. *Genomics* **13**, pp. 690-697.

Burt, D.W., Burnstead, N., Bitgood, J.J. Ponce de Leon, F.A. & Crittenden, L.B. 1995. Chicken genome mapping: a new era in avian genetics. *Trends in Genetics* **11,** pp. 190-194.

Charlier, C., Coppieters, W., Farnir, F., Grobet, L., Leroy, P., Michaux, C., Mni, M., Schwers, A., Vanmanshoven, P., Hanset, R. & Georges, M. 1995. The mh gene causing double-muscling in cattle maps to bovine chromosome 2. *Mammalian Genome* **6**, pp. 788-792.

Charlier, C., Denys, B., Belanche, J.I., Coppieters, W., Grobet., L., Mni, M., Womack, J., Hanset, R. & Georges, M. 1996. Microsatellite mapping of a major determinant of White Heifer Disease: The bovine roan locus. *Mammalian Genome* **7**, pp. 138-142.

Collins, F.S. 1995. Positional cloning moves from perditional to traditional. *Nature Genetics* **9**, pp. 347-350

Crawford, R.D. 1990. Poultry Breeding and Genetics, ed. R.D. Crawford, Amsterdam, Elsevier.

Crittenden, L.B., Provencher, L., Santangello, L., Levin, I., Abplanalp, H., Briles, & Dodgson, R.W. 1993. Characterization of a Red Jungle Fowl by White Leghorn backcross reference population for molecular mapping of the chicken genome. *Poultry Science* **73**, pp. 334-348

Davies, W., Harbitz, I., Fries, R., Stranzinger, G. & Hauge, J.G. 1988. Porcine malignant hyperthermia carrier detection and chromosomal assignment using a linked probe, *Animal Genetics* **19**, **3**, pp. 203-212.

Eppig, J.T. 1996. Comparative maps: adding pieces to the mammalian jigsaw puzzle. *Curr. Opi. Genet. Dev.* **6**, pp. 723-730.

Fujii, J., Otsu, K., Zorzato, F., Deleon, S., Khanna, V.K., Weiler, J.E., O'Brien, P.J. & MacLennan, D.H. 1991. Identification of a mutation in the porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253**, pp. 448-451.

Grobet, L., Martin, L.J., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Menissier, F., Masabanda, J., Fries, R., Hanset, R. & Georges, M. 1997. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nature Genetics* **17**, 1, pp. 17-74.

Groenen, M.A.M., Crooijmans, R.P.M.A.,Veenendaal, A., Cheng, H.H., Siwek, M. & van der Poel, J.J. 1998. A comprehensive microsatellite linkage map of the chicken genome, *Genomics* **49,** pp.265-274.

Heltemes, L.M., Tuggle, C.K. & Lamont, S.J. 1997. Isolation and mapping of two chicken POU family genes and the identification of a syntenic group with human and mouse. *Animal Genetics* **28**, pp. 346-350.

Huang, N., Cogburn, L.A., Agarwal, S.K., Marks, H.L. & Burnside, J. 1993. Overexpression of a truncated growth hormone receptor in the sex-linkeddwarf chicken: evidence for a splice mutation. *Molecular Endocrinology* **7**, 11, pp. 1391-1398.

Hutt, F.B. 1936. Genetics of the fowl. A tentative chromosome map. Neue Forsch.

Tierz. Abstammungsl. (Duerst Festschr.):105-112, Verbandsdruckerei AG, Bern.

Johansson, M., Ellegren, H., Marklund, L., Gustavsson, U., Ringmar-Cederberg, E., Andersson, K., Edfors-Lilja, I. & Andersson, L. 1992. The gene for dominant white color in the pig is closely linked to ALB and PDGFRA on chromosome 8. *Genomics* **14**, pp. 965-969.

Johansson, M., Marklund, L., Sandberg, K. & Andersson, L. 1994. Cosegregation between the chestnut coat color in horses and polymorphisms at the melanocyte stimulating hormone (MSH) receptor locus. *Animal Genetics* **25**, **2**, pp. 35.

Johansson, M., Ellegren, H. & Andersson, L. 1995. Comparative mapping reveals extensive linkage conservation-but with gene order rearrangements-between the pig and the human genomes. *Genomics* **25,** pp. 682-690.

Johansson-Moller, M., Chaudhary, R., Hellmen, E., Hoyheim, B., Chowdhary, B. & Andersson, L. 1996. Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome* **7**,11, pp. 822-830.

Jones, C.T, Morrice, D.R., Paton, I.R. & Burt, D.W. 1997. Gene homologs on human chromosome 15q21-q26 and a chicken microchromosome identify a new conserved segment. *Mammalian Genome* **8**, 436-440.

Kambadur, R., Sharma, M., Smith, T.P.L. & Bass, J.J. 1997. Mutations in myostatin (GDF8) in Double-Muscled Belgian Blue and Piedmontese cattle. *PCR Methods and Applications* 1;7(9), pp. 910-915.

Klein, S., Morrice, D.R., Sang, H., Crittenden, L.B., & Burt, D.W. 1996. Genetic and physical mapping of the chicken *IGF1* gene to chromosome 1 and conservation of synteny with other vertebrate genomes. *J. of Heredity* **87**; 1, pp. 10-14.

Klungland, H., Vage, D.I., Gomez-Raya, L., Adalsteinsson, S. & Lien, S. 1995. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mammalian Genome* **6**, pp. 636-639.

Levin, I., Crittenden, L.B. & Dodgson, J.B. 1993. Genetic map of the chicken Z chromosome using random amplified polymorphic DNA (RAPD) markers. *Genomics* **16**, pp. 224-230.

Levin, I., Santangello, L., Cheng, H., Crittenden, L.B. & Dodgson, J.B. 1994. An

autosomal genetic linkage map of the chicken. *J.of Heredity* **85**, pp. 79-85.

Li, H., Schwartz, B. & Vertel, B.M. 1993. cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia. *J. of Biol. Chem.* **268**, pp. 23504-23511.

Maclachlan, I., Nimpf, J., White, H.B. & Scneider, W.J. 1993. Riboflavinuria in the rd chicken- 5' splice mutation in the gene for riboflavin-binding protein. *J. of Biol Chem.* **268**, pp. 23222-23226.

Mariani, P., Johansson, M., Hoyheim, B., Marklund, L., Davies, W., Ellegren, H. & Andersson, L. 1996. The extension coat color locus and the loci for blood group O and tyrosine aminotransferase are on pig chromosome 6. *J. of Heredity* **87**, 4, pp. 272-276.

Matsumine, H., Herbst, M.A., Ou, S.H.I., Wilson, J.D. & Mcphaul, M.J. 1991. Aromatase messenger RNA in the extragonadal tissues of chickens with the henny feathering trait is derived from a distinctive promoter structure that contains a segment of a retroviral long terminal repeat- functional organization of the Sebright, Leghorn, and campine aromatase genes. *J. of Biol. Chem.* **266**, pp. 19900-19907.

Michelmore, R.W., Paran, I. & Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Nat. Acad. Sci. USA* **88.** pp. 9828-9832.

Montgomery, G.W., Crawford, A.M., Penty, J.M., Dodds, K.G., Ede, A.J., Henry, H.M., Pierson, C.A., Lord, E.A., Galloway, S.M., Schmack, A.E. *et al.* 1994. The ovine Booroola fecundity gene (FecB) is linked to markers from a region of human chromosome 4q. *Nature Genetics* **4**, pp. 410-414.

Morgan, T.H., Sturtevant, A.H., Muller, H.J. & Bridges, C.B. 1915. The mechanism of Mendelian Heredity. Holt, Reinhart & Winston, New York.

O'Brien, S.J., Graves, J.A.M. 1991. Report on the committee on comparative genome mapping. *Cytogenet. Cell Genet.* **58**, pp. 1124-1151.

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A. & Copeland, N.G. 1993. Anchored reference markers for comparative genome mapping in mammals.

Nature Genetics **3**, pp. 103-112.

Olofsson, B. & Bernardi, G. 1983. Organization of nucleotide sequences in the chicken genome. *Eur. J. Biochem.* **130**, pp. 241-245.

Palmer, D.K. & Jones, C.J. 1986. Gene mapping in chicken-Chinese hamster somatic cell hybrids. *J. of Heredity* **77**, pp. 106-108.

Primmer, C.R., Raudsepp, T., Chowdhary, B.P., Moller, A.P. & Ellegren, H. 1997. Low frequency of microsatellites in the avian genome. *Genome Research* **7**, pp. 471-482.

Rodionov, A.V., Yu. A., Myakoshina, L.A., Chelysheva, I.V., Solovei, B. & Gaginskaya, E.R. 1992. Chiasmata in the lambrush chromosomes of Gallus Gallus Domesticus: The cytogenetic study of recombination frequency and linkage map lengths. *Genetika* **28**, pp. 53-63

Ruyter-Spira, C.P., Liang Gu, Z., van der Poel, J.J. & Groenen, M.A.M. Groenen, 1997. Bulked segregant analysis using microsatellites: Mapping of the Dominantwhite locus in the chicken. *Poultry Science* **76**, pp. 386-391.

Ruyter-Spira, C.P., de Groof, A.J.C., van der Poel, J.J., Herbergs, J., Masabanda, J., Fries, R. & Groenen, M.A.M. 1998. The *HMGI-C* gene is a likely candidate for the autosomal dwarf locus in the chicken. *J. of Heredity,* **89, 4,** pp. 295-300

Sheffield, V.C., Nishimura, D.Y. & Stone, E.M. 1995. Novel approaches to linkage mapping. *Curr. Opin. Genet. Dev.* **5**(3) pp. : 335-341

Smeets, H.J.M. Brunner, H.G., Ropers, H.H. & Wieringa, B. 1989. Use of variable simple sequence motifs as genetic markers: Application to study of myotonic dystrophy. *Hum. Genet.* **83**, pp. 245.

Smith, E.J., Cheng,H. & Vallejo, R. 1996. Mapping candidate chicken genes: An alternate approach. *Poultry Science* **75** pp. 642-647

Smith, E.J., Lyons, L.A., Cheng, H.H. & Suchyta, S.P. 1997. Comparative mapping of the chicken genome using the East Lansing Reference population. *Poultry Science* **76**, pp. 743-747.

Sturtevant, A.H. 1913. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *J.Exp.Zool.* **14**, pp. 43-59.

Takagi, N. & Sasaki, M. 1974. A phylogenetic study of bird karyotypes. *Chromosoma*

46, pp. 91-120.

Tsujino, S., Shanske, S., Valberg, S.J., Cardinet, G.H., Smith, B.P. & DiMauro, S. 1996. Cloning of bovine muscle glycogen phosphorylase cDNA and identification of a mutation in cattle with myophosphorylase deficiency, an animal model for McArdle's disease. *Neuromuscul. Disord.* **6**(1), pp. 19-26.

Weber, J.L. & May, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am.J.Hum.Genet.* **44**, pp. 388

Yamashina, M.Y. 1944. Karyotype in birds. I. Comparative morphology of chromosomes in seventeen races of domestic fowl. *Cytologia* **13**, pp. 270-296.

Chapter 2

Development and mapping of polymorphic microsatellite markers derived from a chicken brain cDNA library.

C.P. Ruyter-Spira, R.P.M.A. Crooijmans, R.J.M. Dijkhof, P.A.M. van Oers, J.A. Strijk, J.J. van der Poel and M.A.M. Groenen.

Department of Animal Breeding, Wageningen Institute of Animal Sciences (WIAS), Wageningen Agricultural University, Postbox 338, 6700 AH, Wageningen, The Netherlands.

published in: *Animal Genetics* 1996, **27(4)**, pp. 229-234

Development and mapping of polymorphic microsatellite markers derived from a chicken brain cDNA library.

C.P. Ruyter-Spira, R.P.M.A. Crooijmans, R.J.M. Dijkhof, P.A.M. van Oers, J.A. Strijk, J.J. van der Poel and M.A.M. Groenen

Summary.

Until now the genetic linkage map in chicken has mainly been based on random genomic markers. Adding expressed sequence tags (EST's) to the genetic linkage maps is becoming of more importance because ESTs can form the basis for comparative mapping studies. This may be helpful for the detection of candidate genes for quantitative trait loci (QTL's).

In our study we used a $(TG)_{13}$ repeat as probe for the detection of microsatellites in a chicken brain cDNA library. After hybridization, 0.15% of the cDNA clones gave a positive signal. The cDNA complexity of the library was high; of the 90 cDNA clones that were sequenced 60 occurred only once. For 29 clones primer sets for the polymerase chain reaction could be developed. Twenty one microsatellites were polymorphic on one or more of the test panels and 15 markers could be mapped on either or both of the international reference families. Because sequence homology between chicken and mammalian cDNAs is sometimes low it appeared difficult to assess the level of sequence homology that indicated a true homologous transcript. In our study seven cDNA clones, of which three could be mapped, showed a relatively high percentage of sequence homology with sequences found in other species. Because sequencing and mapping of expressed sequence tags in human and mouse is progressing very rapidly, it is predicted that further information will soon be readily available. Therefore, increasing the number of expressed sequences on the chicken genetic linkage map will be of value for comparative mapping studies in the near future.

Introduction.

The development of genetic linkage maps in farm animals is progressing rapidly. The primary goal is the identification of genetic loci responsible for genetic variation in

traits of economic importance (Soller 1994). Markers for these loci can then be used in breeding programmes.

In chicken more than 460 loci have been mapped; these loci consist mainly of restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNAs (RAPD), minisatellites and microsatellites (for a detailed review see Burt *et al.* 1995). Until now a few expressed sequences have been mapped (Khatib *et al.* 1993, Bumstead *et al.* 1994, Crooymans *et al.* 1995). In particular type I loci are of great importance as they can be used in comparative mapping studies to determine evolutionary relationships between different species. Linkage associations between homologous genes of different species may be of functional significance (O'Brien *et al.* 1993). In animal breeding the genetic maps of livestock species can be compared with the much more detailed human and mouse map. Once a quantitative trait has been located, synteny of genes may predict possible candidate genes that affect the quantitative trait (Soller, 1994). An example of a syntenic group in birds and mammals is the genecluster GC, ALB, PPAT and PGM2 (Palmer & Jones 1986). This shows that even between birds and mammals, species separated by over 300 million years of evolution, a certain extent of synteny has been maintained.

One way to add expressed sequences to a linkage map is to define DNA polymorphism present in cDNA clones and to subsequently follow the segregation of these polymorphism in a reference family. In the human, Hearne *et al.* (1992) found that microsatellites were evenly distributed and equally variant among genomic and cDNA sequences. Because microsatellites are known to be highly polymorphic and polymorphism are easy to detect with PCR they are ideal markers for constructing genetic maps.

In our study we attempt to map microsatellite markers, derived from chicken brain cDNA clones, on either or both of the two international reference families (Bumstead & Barrow 1987, Crittenden *et al.* 1993). cDNA sequences were compared with sequences present in the Genbank/EMBL database and the human cDNA database of The Institute for Genome Research (TIGR).

Materials and methods.

Screening of the chicken brain cDNA library.

A chicken brain cDNA library of a female White Leghorn was obtained from Clontech (Palo Alto, CA, USA). This library contained cDNA fragments, with a mean size of 1.7 kb, cloned into the *Eco*RI site of λgt10. Approximately 210.000 plaques were filter hybridized to a radioactively endlabeled $(TG)_{13}$ probe according to Crooijmans *et al.* (1993). Positive plaques isolated from the first round of hybridization were purified in two subsequent rounds of hybridization.

Subcloning of λgt10 inserts into plasmid PTZ18R.

The cDNA insert in λgt10 was isolated using PCR with primer pairs located on both sites of the *Eco*R1 cloning site. PCR was performed in 96 well microtiter plates (Costar, Cambridge, MA, USA) in a 40 µl volume. The PCR conditions were 1.5 mM $MgCl_2$, 50 mM KCL, 10 mM Tris.HCL pH=8.3, 1 mM tetra methyl ammonium chloride (TMAC), 0.1% Triton X-100, 0.01% gelatine, 200 mM dNTP, 0.5 unit super Tth (HT Biotechnology Ltd, Cambridge, UK) ,30 ng of each primer and 10 µl denatured high titer λgt10 solution.

PCR was performed with 35 rounds of 30 s at 94°C, 45 s 50°C and 1.45 min 72°C, followed by a final elongation step of 10 min 72 °C.

PCR products were purified on PCR purification columns (QIAquick Spin PCR Purification Kit, Quiagen Inc., Chatsworth, CA), restricted with *Eco*R1 and subcloned into the *Eco*RI site of plasmid pTZ18R.

In cases where no PCR product could be obtained, λgt10 DNA was isolated. Phages were plated out overnight at confluency on a 10 cm diameter petri dish. The phages were eluted with 5 ml SM buffer (100 mM NaCl, 10 mM Tris pH 7.9, 10 mM $MgSO_4$). A 500 µl aliquot of the eluate was lysed with 10 µl 0.25 M EDTA and 15 µl 10% SDS for 10 min at 68°C. After phenol/chloroform purification the DNA was precipitated, dried and finally dissolved in 50 µl of TE. DNA was restricted with *Eco*R1 and the inserts were subcloned into pTZ18R.

Sequence analysis.

Plasmid DNA was prepared by the alkaline lysis method using the Quiaprep plasmid

isolation kit (Qiagen Inc.). DNA sequencing was performed using the autoread sequence kit (Pharmacia, Uppsala, Sweden). cDNA clones were sequenced from both termini. Sequences were determined on the Automated Laser Fluorescent (ALF) sequencer (Pharmacia) using a 6% denaturing polyacrylamide gel (sequagel-6, National Diagnostics, Atlanta, G) and further analysed using the Microgenie program of Beckman. Comparison of sequences with sequences in the Genbank/EMBL databases (release 86) was performed using GCG FASTA software (Genetics Computer Group 1991) using a word size of 6, a gap creation penalty of 12 and a gap extension penalty of 4. Nucleotide sequences of the cDNA clones for which primer pairs were developed were also compared with sequences present in the Human cDNA Database (HCD) version 2.0.1 of The Institute for Genomic Research (TIGR) using the software available through HCD at level 1 (Adams *et al.* 1995).

Polymerase chain reaction and product analysis

PCR primers of 20 nucleotides on both sides of the repeats were designed. One primer from each locus was fluorescently labelled with HEX, TET or FAM phosphoramidites (Applied Biosystems, Waterstadt, Germany). PCR reactions were performed as described above with minor modifications; 10 - 100 ng genomic DNA was used at an annealing temperature of 50 or 55°C in a 20 µl reaction volume. PCR products were analysed using 6% denaturing polyacrylamid gels on an ABI automated sequencer (Applied Biosystems Inc., Foster City, CA). Fragment sizes were calculated relative to the Genescan-350 TAMRA standard with the GENESCAN software version 1.2.2-1 (Applied Biosystems Inc., Foster City, CA).

Reference families and linkage analysis

The markers were tested for polymorphism on a panel that comprised the parents of the reference families of East Lansing (Crittenden *et al.* 1993) and Compton (Bumstead & Palyga 1992). In addition, a Wageningen testpanel of three animals of different layer lines and two animals of different broiler lines was analysed. When polymorphism was detected in one of the reference families, the respective family was genotyped for the microsatellite. Linkage analyses were performed with MAP

MANAGER software (version 2.6 for Macintosh; K. Manly and R. Cudmore, Roswell Park Institute, Buffalo, NY). The best order of markers was performed manually by minimizing double recombinants.

Results and discussion.

Development and sequencing of microsatellites

In our study we isolated microsatellites from a chicken brain cDNA library. Because the brain is composed of a large variety of different cell types that exhibit a complex array of morphological and physiologic properties, RNA complexity is higher than in other organs (Morrison & Griffin 1985). Therefore the chance of finding a large variety of different cDNA's is high (Adams *et al.* 1992).

After hybridizing 210.000 phages to a $(TG)_{13}$ probe, 0.15% of the plaques gave a positive signal. The 120 most strongly hybridizing plaques were isolated and purified. Inserts of 90 cDNA clones were isolated, subcloned into pTZ18R and partially sequenced. As expected the variety of different cDNA clones in this library was high. Of the 90 cDNA clones that were sequenced 60 occurred only once.

Linkage analysis

For 29 of the 68 different cDNA clones primer pairs could be developed. The remaining clones showed either short repeats, repeats located at the end of the inserts or no repeats at all because the clones were only partially sequenced.

The 29 microsatellites were tested for polymorphism on the different test panels (see *Material and methods*). This resulted in 21 polymorphic markers (Table 1) of which 15 could be mapped on the Compton- and/or East Lansing reference family. The map locations of these markers as well as the most likely gene orders are shown in Table 2. Because four microsatellites could be mapped in both reference families, it became clear that linkage groups E1 and C15, E2 and C1, E36 and C10 and finally E51 and C28 were corresponding linkage groups in both reference families. The alignment of linkage groups E1/C15 and E2/C1 was also reported by Crooijmans *et al.* (1995b).

Table 1. cDNA clones containing microsatellites for which PCR primer pairs were developed.

cDNA No.	Microsat. sequence	MCW No.	Primer sequences	Polym.[1]	Length	Genbank/ EMBL
1A	$(CA)_{12}$	144	TGGCGTAACCTCCTGCTGCT GCAGCCGAACTGTTTTGGCA	-	87	L48877
4A	$(TG)_{20}$	191	TGTAATCAGCATTTAATAGA AGGCAAACAGTTGTGAAACT	W	202-232	L48878
10A	$(TG)_4n_2$- $(TG)_4n_4(TG)_5$	143	ATCTGTTGCACACTCATTGC TACTGAGCATTGTGCATGCG	C	100-105	L48880
13A	$(TG)_{24}$	190	GTGATCATTTCTACATGCAG ACAACAGAACTAAACAAATA	E,C	60-105	L48881
14A	$(TG)_{21}$	142	GTCTAAAGAAATACACATAC CTGAAAGATCTAAGCTTTGT	W	68-91	L48882
18A	$(CA)_{13}(TA)_6$	141	GTATGAGTATAGCTGTATTG CTGAAAGATCTAAGCTTTGT	E,W	249-257	L48883
22A	$(CA)_{14}$	153	ACTGCCTGATGTAAACAAGT CATATGGAAATGGCGCAGCT	W	73-75	L48885
23A	$(CA)_6n_2(CA)_5n_8$- $(CAAA)_5$	155	GGTTAGTAATGTTCCTCATC AGACATCAATGAGTCAGTCA	W	130-138	L48886
35A	$(AT)_6(GT)_{10}$	163	GAAGTGTGTTCTACAGTCTG ATAGGCATTGTTAATGTACC	C	100-102	L48890
37A	$(TG)_{13}$	162	TCATCTCCAGACCTGGCCTG GCATTTACATTGTAACAACT	E,C	75-85	L48891
38A	$(GT)_{10}n_2(TG)_5$	186	AGCCAATCACCTGCGATCAC TAATCCTATCTTTATACAGC	-	122	L48892
42A2	$(AC)_{15}$	189	CACGACCTTTGCACGTCTAT CCGAGCGCGTGTGGTGCATG	C	110-124	L48893
44A	$(TG)_8$	149	ACTCCTACAACAGCATACAT TGCAATTAAAGGAGTAACCT	E,W	76-94	L48895
48A	$(CA)_9$	188	GTGACAGCGGCAGAGATGGA CGCACAGCCCCACTCGCACA	E,W	177-183	L48897
51A	$(AC)_4n(AC)_4n(AC)6$	--	GTGCAATGGGCGTGTTAG TCCCAACACACCCTACGT	-	110	L48898
56A	$(CA)_{10}$	187	ATCTAGTTTGGACAAGTTAC CTAACTTATTTCAGTCAAAT	E,W	117-132	L48899
61A	$(CA)_{12}$	197	GTGCTGCTGGGTTTAACCTA CTCACACGCGCACATACTTA	E,C,W	95-112	L48901
2	$(TG)_{10}$	106	GGCAACTAAGTTGTGGACTG GCAGCATTCAGTGGGATAAT	E,C,W	125-129	L48902
3	$(TG)_{17}$	108	GTCTCGTGGAGATGATCTAT ATAATGAAGACACCGACATT	W	97-101	L48903
4	$(CA)_{17}$	109	TGATTCACTTGATGGTCGAG CTATCAATTTATTCTGCCTT	E,W	131-154	L48904
11	$(CA)_2n_2(CA)_6A_8$	113	AGTGTATCCAGCCCACACTT GAAGTGGTGTCATCAAGGAC	W	100-102	L48905
31	$(CA)_{11}$	107	GAACAGAACTCTGTTTACTG TCTGCTTACCTCAACTGACA	E,W	114-120	L48906
37	$(GT)_{11}$	110	CATCTGTGTTACTGTCACAG TCAGAGCAGTACGCCGTGGT	E,W	102-114	L48908
42	$(AC)_8$	111	GCTCCATGTGAAGTGGTTTA ATGTCCACTTGTCAATGATG	E,W	104-110	L48909

(1) C, polymorphic on Compton reference family; E, polymorphic on East Lansing reference family;W, polymorphic on Wageningen test panel; -, monomorphic on all test panels.

Although six markers were not informative on the international reference families, they were polymorphic in the resource family currently under investigation in our laboratory and will be mapped in the future.

Database search for sequence homology

Because birds and mammals are evolutionary diverse we investigated the expected range of sequence homology between chicken and human gene homologues. The coding- and 3' untranslated regions (UTR) of six different chicken cDNA sequences from the Genbank/EMBL databases were compared with their human homologues. This resulted in sequence homologies varying from 54.2% to 76.9%. In four cases homology was only found in the coding region of the genes. Considering the wide range and low percentages of sequence homology even in coding regions, it would be difficult to determine if low percentages of homology indicated true homologues in our study.

From the 68 different chicken brain cDNA clones six clones showed a relatively high percentage of sequence homology with sequences from other species present in the Genbank/EMBL database (Table 3). Sequences of the clones for which primer pairs were developed were also compared with the human cDNA database. This resulted in the additional sequence homology between cDNA31 and a tentative human consensus sequence (THC108215). This THC was the result of the alignment of two expressed sequence tags both derived from a human brain cDNA library.

Three of the seven cDNA clones that showed sequence homology with sequences from other species (cDNA18A, cDNA31 and cDNA42A2) were mapped. Alignment of three different cDNA clones overlapping clone cDNA18A yielded a 1050 bp sequence which was 72.9% homologous to THC102609. The putative identity of this tentative human consensus sequence was the isolog of neuroendocrine-specific protein C. Although neuroendocrine-specific protein is a tissue specific protein, its isolog has been found in many other different tissues. Clone cDNA42A2 showed homology with the human gene encoding leukocyte adhesion glycoprotein Mac-1 which is located on chromosome 16p11.2 (Callen *et al.* 1991). The chicken cDNA clone only covered the 3' UTR of the gene resulting in a sequence homology percentage of 66.8%. Because lower sequence homologies are expected between the 3' UTRs of different species it is possible that the human gene is the true homologue of the chicken cDNA. Although

the chromosomal location of the human Mac-1 gene is known, conclusions about possible syntenic groups in chicken and man can not be drawn for this locus at present.

Table 2. Location of the microsatellite markers mapped on the East Lansing (E) and Compton (C) reference families.

cDNA	Marker	Linkage group	Most likely gene order[a]
2	MCW106	E1	JFEVB- 2.3- **MCW106** -4.1-HUJ3
		C15	ADL43- 14.9- **MCW106** -8.2-LYZ
4	MCW109	E1	MSU28- 6.3- **MCW109** -6.2-MSU78
31	MCW107	E15	HGB- 13.9- **MCW107** -11.5-LEI7/LEI18
37	MCW110	E51	MCW197- 21.3- **MCW110** -6.1-MSU10/MCW104
42	MCW111	E1	EV2- 2.1- **MCW111** /MSU79-5.8-ADL19
10A	MCW143	C33	MCW157- 26.4- **MCW143** /MCW189-9.8-HSPA3
13A	MCW190	E36	MCW135- 8.3- **MCW190** -19.6-EIF4A2
		C10	MCW135- 7.9- **MCW190** -1.9-COM51/ADL21
18A	MCW141	E2	MSU8- 6.4- **MCW141** -34.4-MCW169
35A	MCW163	C3	COM150- 9.6- **MCW163** -15.7-CALB4
37A	MCW162	E2	MSU12- 14.9- **MCW162** - 2.2-COM95
		C1	MCW139- 24.4- **MCW162** - 1.9-COM5
42A2	MCW189	C33	MCW157- 26.4- **MCW189** /MCW143-9.8-HSPA3
44A	MCW149	unlinked	**MCW149**
48A	MCW188	E22	I- 16.3- **MCW188** -14.3-LEI3
56A	MCW187	E2	MCW4- 7.3- **MCW187** -4.1-ADL7
61A	MCW197	E51	**MCW197** -21.3-MCW110
		C28	COM114- 2.2- **MCW197** -17.8-HP103E4

(a) Distance between markers is given in cM.

The high complexity of cDNA in the chicken brain cDNA library and the presence of $(TG)_n$ repeats in 0.15% of the clones, were found to be an appropriate combination for the inclusion of expressed sequence tags (ESTs) to the genetic linkage map; consequently we were able to add 15 additional type 1 loci to the genetic linkage map of the chicken. The development of such ESTs is becoming of more importance because of its potential to yield further information about relationships between the chicken genome and genomes of other species. In our study three of the 15 mapped chicken brain cDNA clones showed sequence homology with sequences from other species. This number will probably soon increase since sequencing and mapping of EST's in man and mouse is also progressing very rapidly. Therefore a wealth of information for comparative mapping studies will soon become available. However, considering the wide range of sequence homologies between known human and chicken homologues and the fact that homology between relatively short stretches of sequence does not necessarily mean that the sequences represent identical transcripts, it is apparent that any conclusion on possible syntenic groups must be

drawn with caution.

Table 3. Sequence homology between chicken brain cDNA clones and sequences present in the Genbank/EMBL database. Clones cDNA 1A, cDNA7A, cDNA18A, cDNA19A, cDNA34A and cDNA38A also showed homology with other database entries containing overlapping sequences or gene homologues from different species.

cDNA nr.	% homology (bp)	Sequence homology with:	Genbank/ EMBL[a]	Species
1A	81.1% 259 bp	EST29139 sim.to microtubule associated protein	T31248	Human[1]
	76.3% 228 bp	microtubule associated protein 1A+1B, light chain3	U05784	Rat[2]
	69.5% 132 bp	EST Tsg125t3	X61837	Mouse[3]
3	77.6% 116 bp	mRNA for GTP binding protein (RAB6)	M28212	Human[4]
4	100 % 175 bp	repeat region	X82816	Chicken[5]
7A	91.8% 171 bp	mitochondrial genome	X52392	Chicken[6]
18A	72.9% 1050 bp	*THC sim.to neuroendocrine specific protein C*	THC102609[b]	*Human*[1]
31	78.3% 170 bp	tentative human consensus sequence	THC108215[b]	Human[1]
34A	88.8% 241 bp	18s ribosomal DNA	U13369	Human[7]
38A	86.4% 490 bp	ZFX, mRNA for putative transcription activator	X59740	Human[8]
42A1	97.3% 812 bp	LIMK, a lung cDNA encoding a protein kinase	D26310	Chicken[9]
42A2	66.8% 292 bp	Leukocyte adhesion glycoprotein Mac-1	J03925	Human[10]

(a) Accession number to Genbank/EMBL database
(b) accession number to Human cDNA Database (HCD)
The tentative human consensus sequence THC108215 is the assembly of the expressed sequence tags R51688 and T78222.
(1) Adams *et al.* (1995)
(2) Mann & Hammarback (1994)
(3) Hoog (1991)
(4) Zahroui *et al.* (1989)
(5) Gibbs *et al.* (1994)
(6) Lesjardins & Morais (1990)
(7) Financsek *et al.* (1982)
(8) Schneider-Gädicke *et al.* (1989)
(9) Ohashi *et al.* (1994)
(10)Corbi *et al.* (1988)

References.

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. & Venter, C. 1992. Sequence identification of 2375 human brain genes. *Nature* **355**, pp. 632-634.

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D *et al.* 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million basepairs of cDNA sequence. *Nature* **377**, suppl., pp. 173-174.

Bumstead, N. & Barrow, P.A. 1987. Genetics of resistance to Salmonella typhimurium in newly hatched chicks. *British Poultry Science* **29**, pp. 521-530.

Bumstead, N., Young, J.R., Tregaskes, C., Palyga, J. & Dunn, P.P.J. 1994. Linkage

mapping and partial sequencing of 10 cDNA loci in the chicken. *Animal genetics* **25**, pp. 337-341.

Burt, D.W., Bumstead, N., Bitgood, J.J., Ponce de Leon, F.A. & Crittenden, L.B. 1995. Chicken genome mapping: a new era in avian genetics. *Trends in Genetics*. **11**, pp. 190-194.

Callen, D.F., Chen, L.Z., Nancarrow, J., Whitmore, S.A., Apostolou, S., Thompson, A.D., Lane, S.A., Stallings, R.L., Hildebrand, C.E., Harris, P.G. & Sutherland, G.R. 1991. Current state of the physical map of human chromosome 16. *Cytogenet. Cell Genet*. **58**, 1998.

Corbi, A.L., Kishimoto, T.K., Miller, L.J. & Springer, T.A. 1988.The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b) alpha subunit. Cloning, primary structure, and relation to the integrins, von Willibrand factor and factor B. *J. of Biol. Chem.* **263**, pp. 12403-12411.

Crittenden, L.B., Provencher, L., Santangello, L., Levin, I., Abplanalp, H., Briles, R.W., Briles, W.E. & Dodgson, J.B. 1993. Characterization of a Red Jungle Fowl by White Leghorn backcross reference population for molecular mapping of the chicken genome. *Poultry Science* **72**, pp. 334-348.

Crooijmans, R.P.M.A., van Kampen, A.J.A., van der Poel, J.J. & Groenen, M.A.M. 1993. Highly polymorphic microsatellite markers in poultry. *Animal Genetics* **34**, pp. 441-443.

Crooijmans, R.P.M.A., van der Poel, J.J. & Groenen, M.A.M. 1995a. Functional genes mapped on the chicken genome. *Animal Genetics* **26**, pp. 72-78.

Crooijmans, R.P.M.A., van Oers, P.A.M., Strijk, J.A., van der Poel, J.J. & Groenen, M.A.M. 1995b. Preliminary linkage map of the chicken (Gallus domesticus) genome based on microsatellite markers: 77 new markers mapped. *Poultry Science* **75**, pp. 746-754.

Desjardins, P. & Morais, R. 1990. Sequence and gene organization of the chicken mitochondrial genome: A novel gene order in higher vertebrates. *J. of Mol. Biol.* **212**, pp. 599-634.

Financsek, I., Mizumoto, K., Mishima, Y. & Muramatsu, M. 1982. Human ribosomal RNA gene: nucleotide sequence of the transcription initiation region and comparison of

three mammalian genes. *Proc. Nat. Acad.Sci. U.S.A.* **79**, pp. 3092-3096.

Hearne, C.M., Ghosh, S. & Todd, J.A. 1992. Microsatellites for linkage analysis of genetic traits. *Trends in Genetics* **8**, pp. 288-293.

Hoog, C. 1991. Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy. *Nucleic Acids Research* **19**, pp. 6123-6127.

Kahtib, H., Gelislav, E., Crittenden, L.B., Bumstead, N. & Soller, M. 1993. Sequence-tagged microsatellite sites as markers in chicken reference and resource populations. *Animal Genetics* **23**, pp. 355-362.

Mann, S.S. & Hammarback, J.A. 1994. Molecular characterization of light chain 3. A microtubule binding subunit of MAP1A and MAP1B. *J of Biol. Chem.* **269**, pp. 11492-11497.

Morrison, M. & Griffin, W. 1985. Molecular biology of the of the mammalian brain. In C. Zomzely-Neurath and W. Walker (Eds.), *Gene expression in brain*, Wiley, New York, pp. 57-98.

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A. & Copeland, N.G. 1993. Anchored reference loci for comparative genome mapping in mammals. *Nature genetics* **3**, pp. 103-112.

Ohashi, K., Toshima, J., Tajinda, K., Nakamura, T. & Mizuno, K. 1994. Molecular cloning of a chicken lung cDNA encoding a novel protein kinase with N-terminal two LIM/double zinc finger motifs. *J. of Bioch.* **116**, pp. 636-642.

Palmer, D.K. & Jones, C. 1986. Gene mapping in chicken-Chinese hamster somatic cell hybrids. *J. of Heredity* **77**, pp. 106-108.

Schneider-Gädicke, S., Beer-Romero, P., Brown, L., Mardon, G., Luoh, S. & Page, D.C. 1989. Putative transcription activator with alternative isoforms encoded by human ZFX gene. *Nature* **342**, pp. 708-711.

Soller, M. 1994. Marker assisted selection-an overview. *Animal Biotechnology* **5**, no.2, pp. 193-207.

Zahroui, A., Touchot, N., Chardin, P. & Tavitian, A. 1989. The human Rab genes encode a family of GTP-binding proteins related to yeast YPT1 and SEC4 products involved in secretion. *J. of Biol. Chem.* **264**, pp. 12394-12401.

Chapter 3

Developing microsatellite markers from cDNA; A tool for adding expressed sequence tags to the genetic linkage map of the chicken.

C.P. Ruyter-Spira, D.J. de Koning, J.J. van der Poel, R.P.M.A. Crooijmans, R. J.M. Dijkhof and M.A.M. Groenen.

Department of Animal Breeding and Genetics, Wageningen Institute for Animal Sciences (WIAS), Postbox 338, 6700 AH Wageningen, The Netherlands.

Published in: *Animal genetics* 1998, **29(2)**, pp. 85-90

Developing microsatellite markers from cDNA; A tool for adding expressed sequence tags to the genetic linkage map of the chicken.

C. P. Ruyter-Spira, D. J. de Koning, J. J. van der Poel, R. P. M. A. Crooijmans, R. J. M. Dijkhof and M. A. M. Groenen.

Summary.

A chicken embryonic cDNA library was screened with a $(TG)_{13}$ probe in order to develop polymorphic microsatellite markers. The redundancy of the embryonic cDNA library with a chicken brain cDNA library, which was used for microsatellite development in a previous study, was extremely high. Of the 300 $(TG)_{13}$ positive clones, only 80 were unique for the embryonic cDNA library. Still, nine expressed sequences derived from the embryonic cDNA library were mapped in the Wageningen (WAU) resource population. In addition seven microsatellite markers from the chicken brain cDNA library, which were monomorphic or unlinked in the two international reference families in the previous study, were also mapped in the WAU population. Three of the 16 mapped chicken ESTs showed relatively high percentages of sequence similarity to sequences found in other species. As two of these genes, RAB6 and ZFX/ZFY, have been mapped in humans, they contribute to the comparative map of the chicken.

Introduction.

One of the principal reasons for the development of genetic maps in agricultural species is to locate and identify genes underlying diseases and economically important traits. Exploitation of the more advanced human and murine transcript maps by means of comparative mapping, will be important for the positional candidate approach to trait gene cloning (Collins 1995). Although the extent of sequence conservation in coding regions between birds and mammals is quite variable (Ruyter-Spira *et al.* 1996), the localization of approximately 125 chicken genes and expressed sequence tags (ESTs) has led to the identification of at least 11 groups of conserved synteny between chickens and humans (Palmer & Jones 1986, Burt *et al.* 1995, Klein *et al.* 1996). One way to increase the number of expressed sequences on the genetic linkage map is to

develop polymorphic markers from cDNA clones. Microsatellites of the $(TG)_n$ type have been shown to be present in 0.15% of the clones of a chicken brain cDNA library, predominantly in the 3' untranslated region (Ruyter-Spira *et al.* 1996). By mapping these microsatellite markers 15 additional ESTs have been localized on the chicken genetic linkage map.

Here, the authors have used the same approach to develop polymorphic microsatellite markers from a chicken embryonic cDNA library. Considering the homologous developmental patterns shared by individual species, genes expressed during early development were expected to show higher rates of sequence conservation. Nine embryonic cDNAs were placed on the WAU genetic linkage map of the chicken, of which one showed homology with a human EST. Seven ESTs derived from a chicken brain cDNA library, which could not be mapped in the international reference families in a previous study (Ruyter-Spira *et al.* 1996), were also mapped in the WAU reference family.

Materials and Methods.

Screening of the chicken embryonic cDNA library

A chicken Uni-ZAP™XR cDNA library from five day old embryos, was obtained from Stratagene (La Jolla, CA, USA). Approximately 300,000 clones were filter hybridized to a radioactively end-labelled $(TG)_{13}$ probe according to Crooijmans *et al.* (1993). In order to decrease redundancy with microsatellites developed from a chicken brain cDNA library (Ruyter-Spira *et al.* 1996), a second set of filters from the same library was also hybridized with radioactively end-labelled PCR primers developed from $(TG)_{13}$ positive clones in this previous study. Plaques that were only positive for the $(TG)_{13}$ probe in the embryonic cDNA library were purified in two subsequent rounds of hybridization.

Sequence analysis

Plasmid preparation, DNA sequencing and sequence analysis were performed according to Ruyter-Spira *et al.* (1996). Sequences were compared with sequences present in the GenBank/EMBL databases (release 99.0) using GCG FASTA software (Genetics Computer Group 1991). A homology search was also performed using the

Human cDNA Database (HCD) (dataversion 2.1) of the Institute for Genomic Research (TIGR) (Adams *et al.* 1995). The software (version 2.0.1) available through HCD, aligns nucleotide sequences both against the nucleotide database and a six-frame translation of the nucleotide database.

Polymerase chain reactions and product analysis

PCR primers on both sides of the repeats were designed, resulting in PCR products varying from 90 until 250 bp in size. One primer from each pair was fluorescently labelled with HEX, TET or FAM phosphoramidites (Applied Biosystems, Waterstadt, Germany). PCR reactions were performed and analysed according to Crooijmans *et al.* (1996) using the ABI 373A DNA sequencer (Perkin Elmer, Applied Biosystems Division, Foster City, CA, USA).

Reference family and Linkage analysis

Polymorphic microsatellite markers were mapped on the Wageningen resource population (Groenen *et al.* in preparation). This experimental population comprises an F2 of 10 full sib families, with an average of 46 animals per family, derived from an intercross between F1 animals resulting from a cross between two extreme commercial broiler lines differing in both reproductive and growth performance. Linkage analysis was performed using CRIMAP version 2.4 (Green *et al.* 1990).

Results and discussion.

Screening of the embryonic cDNA library

The screening of 300,000 plaques from a chicken embryonic cDNA library resulted in approximately 300 $(TG)_{13}$ positive clones (0.1%), which is comparable to the number of $(TG)_{13}$ positive clones found in the chicken brain cDNA library (0.15%). To decrease the redundancy of $(TG)_{13}$ positive clones also present in the chicken brain cDNA library, a subtraction was carried out (see *Materials and methods*), resulting in only 80 $(TG)_{13}$ positive cDNA clones unique for the embryonic cDNA library. The high redundancy (77%) is probably caused by the presence of cDNA clones derived from embryonic brain tissue which are also present in the embryonic cDNA library. The subtraction was worthwhile as only two clones previously isolated from the brain library were sequenced

again.

Table 1. Polymorphic microsatellite markers and their PCR primer sequences, which were developed from clones derived from a chicken embryonic cDNA library.

cDNA	Marker[1]	Microsatellite	Length (bp)	Primer sequences	Genbank/ EMBL[2]
14	MCW271	$(AC)_7$--$(CA)_{10}$	200	GTTGCTTAAAAGCATTCTGT ACACTAACAACTCGTTTATG	AF030577
19	MCW204	$(GT)_{10}$	96	AGATTCGTGCAAGCGCTGCTG TCAACTAATGGCACAACCGTG	AF030578
21	MCW206	$(TG)_9$	232	CTTGACAGTGATGCATTAAAT ACATCTAGAATTGACTGTTC	AF030579
33	MCW272	$(AC)_7CC(AC)_8$	135	CTCAGGTTCCAGCTGATGC ACTCAGAGCCACTTAGCTGC	AF030580
43	MCW203*	$(GT)_2TG(GT)_7$	183	CTGTTGTAATGATAGAAGAGC CTGTTCACGTCTAGTCCCCG	AF030581
104	MCW273	$(TG)_9$	165	TGTTATGCCATTGCTTTTAG CAGCACCCACATCAAGATCC	AF030582
230	MCW299*	$(TG)_8$	98	TGTTTGGAATTGTAAACTGC TGAAGGAAAGGTGAGTGAGAC	AF030583
266	MCW275	$(TG)_8$	140	TTTTTTTCGAGTTTCTGCAG AAACCTGACTTCGATACC	AF030584
320	MCW276	$(TG)_8(AG)_5$	205	ACTCTGAGTGGAATTACCTG ATTTCTGTTAGAAGCAGCTGC	AF030585
329	MCW216	$(GT)_9$	145	GGGTTTTACAGGATGGGACG AGTTTCACTCCCAGGGCTCG	AF030586
439	MCW274	$(CA)_9$	155	TCATGTAATGTAACCTCTGC GATGTACCAGTGCCTCAGCG	AF030587

1, Microsatellites denoted with * were monomorphic on the parents of the WAU recourse population
2 Accession number of the Genbank/EMBL database

Development of microsatellites and linkage analysis

Because DNA polymorphism and repeats are primarily located in the 3' untranslated region (Levit 1991, Hearne *at el.* 1992), all 80 clones were partially sequenced from both termini in order to maximize the chance to find the protein coding parts of the cDNA which are more informative for interspecies comparison of nucleotide sequences. Of the 80 cDNA clones sequenced, 40 clones were unique, whereas 38 other clones, representing nine different genes, occurred at a higher frequency. The redundancy within the embryonic cDNA library was somewhat higher than the redundancy found in the chicken brain cDNA library (Ruyter-Spira *et al.* 1996) in which 66% of the clones occurred only once. This could be expected since the brain is composed of a larger variety of different cell types, leading to a higher RNA complexity (Morrison and Griffin 1985). Finally, one clone turned out to be homologous to a previously developed microsatellite marker (MCW301) from a genomic chicken library (Crooijmans *et al.* 1997).

For 29 of the 49 different cDNA clones it was not possible to develop PCR primers because these clones contained either short repeats, repeats located at the end of the inserts, or the repeats were not detected after partial sequence analysis. Of the 20 remaining microsatellites tested for polymorphism on the parents of the WAU resource population (Groenen *et al.* 1998), nine were polymorphic microsatellites markers (Table 1), seven were monomorphic, and for four no product was detected. The high proportion of monomorphic microsatellites is probably due to the isolation of clones containing predominantly shorter repeats. Indeed, the redundant clones that were eliminated showed the strongest hybridization signals with the $(TG)_{13}$ probe thereby probably representing longer repeats.

Table 2. Location of the microsatellite markers mapped on the WAU resource population.

cDNA	Marker	Linkage group	Most likely gene order[a]
Embryonic cDNA library			
14	MCW271	WAU11	ADL301-13- **MCW271** - 3-LMU15
19	MCW204	WAU16	MCW227- 1- **MCW204** - 13-ADL299
21	MCW206	WAU2	ADL185- 4- **MCW206** - 10-LEI86
33	MCW272	WAU2	MCW96 - 1- **MCW272** - 14-MCW234
104	MCW273	WAU1	LEI139- 1- **MCW273** - 14-LEI169
266	MCW275	WAU11	**MCW275** - 1-ADL322
320	MCW276	WAU4	LEI76 - 9- **MCW276** - 9-MCW284
329	MCW216	WAU13	MCW197- 6- **MCW216** - 5-ADL147
439	MCW274	WAU2	MCW239-12- **MCW274** - 7-MCW65
Brain cDNA library			
4A	MCW191	WAU4	MCW240- 1- **MCW191** - 15-ADL331
14A	MCW142	WAU2	MCW247-13- **MCW142** - 4-ADL185
22A	MCW153	WAU2	MCW173- 7- **MCW153** - 1-MCW87
38A	MCW186	WAU1	MCW46 -12- **MCW186** - 13-UMA356
44A	MCW149	WAU8	ADL259- 4- **MCW149** - 6-MCW134
3	MCW108	WAU1	MCW107- 1- **MCW108**
11	MCW113	WAU5	MCW32 - 2- **MCW113** - 3-LEI149

[a] Distance between markers is given in cM.

The map locations of all nine polymorphic microsatellites as well as the most likely gene orders are shown in Table 2. In addition to these markers, six microsatellites developed from the chicken brain cDNA library in the previous study, which could not be mapped in the East Lansing (Crittenden *et al.* 1993) and Compton (Bumstead & Palyga 1992) international reference families, and one marker (MCW149) which was found to be unlinked in these populations, were mapped in the WAU resource population in the present study (Table 2). All 16 markers could be placed in one of the 26 autosomal linkage groups currently comprising the WAU genetic linkage map.

Markers MCW186, MCW274 and MCW276 are of particular value since they are located in genomic regions which were sparsely covered with other markers.

In this study, 12 of the 16 ESTs map to one of the macrochromosomes. Taking into account all 40 ESTs from which microsatellites markers have been developed (Crooijmans *et al.* 1995, Ruyter-Spira *et al.* 1996, the present study), no bias towards an over representation of expressed sequences localized on microchromosomes was observed as was suggested by McQueen *et al.* (1996).

Database search for homology

Seventeen of the 49 different TG-positive clones were similar to sequences from the Genbank/EMBL and HCD databases (Table 3)..

Table 3. Sequence homology between clones from both the chicken embryonic and the brain cDNA library and sequences present in the Genbank/EMBL database.

cDNA	%hom.[1]	bp.	Sequence homology with:	Acc.nr.[2]	From[3]
14	70.5	217	h.EST	T89971	Lung
86	87.2	288	h.cDNA sim.to glia activating factor precursor	H62672	Fetal liver/spleen
196	65.2	296	h.serum response factor	J03161	HELA celline
239	73.4	128	h.EST	R97363	Fetal liver/spleen
285	63.8	365	xenopus laevis mRNA for alpha-T3-globin	X02796	Tadpole
349	82.6	322	h.EST	H15433	Infant brain
367	74.4	246	h.EST	H10275	Infant brain
402	71.7	244	rat proto-oncogene (Ets-1) mRNA	L20681	Lymphoma
418	73.9	399	h.transcription factor BTF3a	X53280	
420	62.0	255	h.betaine/GABA transporter	L42300	Brain
435	74.6	173	h.EST	D54048	Fetal brain
468			chicken smooth muscle myosin light chain	X06387	
481			chicken tyrosine kinase 1 mRNA	M69243	
489	87.6	226	h.extracellular signal-related kinase 3	X80692	Fetal muscle
506			chicken mRNA for agrin	M94271	Embryonic spinal cord
515	96.0	200	h.EST similar to translation initiation factor	T30674	Spleen
526	75.7	382	h.mRNA for p57	D44497	Blood leukocytes

[1] Percentage of homology observed in a stretch of x basepairs (bp).
[2] Accession number to Genbank/EMBL database.
[3] Tissue or cell type from which the homologous clones were isolated.

As expected, this number is substantially higher than the number of clones from the chicken brain cDNA library showing sequence homology with sequences from the databases (15%). At least ten clones show sequence similarity to mRNA sequences derived from clones originating from rapidly proliferating fetal, juvenile or tumorigenic tissue types. Most of these genes are involved in the regulation of gene expression and signal transduction, and are known to be highly conserved across species during evolution explaining the high number of clones for which sequence homology was

found. Unfortunately, only one clone from the embryonic cDNA library (DJ14/MCW271) for which sequence homology was found could be placed on the genetic linkage map of the chicken, however, the map location of its human counterpart remains to be determined. Alternatively, the remaining clones for which homology was found, could be mapped using either fluorescence *in situ* hybridization (FISH) or radiation hybrids in the future.

```
MCW108
CCTCAAGAGCAGCCTGTCAGTGAAGGAGGCTG-CCCTGCTAATACCACGTGGCTTCT--G
|||||  |||||  ||  ||||||||||||||||||  |||||||||  | |  | | |||
CCTCAGGAGCAACCAGTCAGTGAAGGAGGCTGTTCCTGCTAATGTCCCTAGTCATCTTCA
hsrab6[1]

ACTTCCTTCCAGAACATCACTGCTTTCCCTTCCCTTTACTCTTCATTGACTGCAGTGTGA
||  |  |||  |||||   ||||||||||
ACCTTCTT-CAGAAGCTCACTGCTTT


MCW188
AACCAGCACATTATGCGGCATCATAAAGATGTTGGGCTGCCTTAAAGTCTCTTTCACAGA
||||||||||||  |||||  ||||||||||  |||||  |||||  |||      |||   |||||
AACCAGCACATAATGCGACATCATAAAGAAGTTGGCCTGCCCTAACAATACTTCTACAGA
ZFX[2]

CTTACGGCGGGAATTGCCAAGGAATTTGGCCTTTCGGGCAGTTAGCTTATTTTAAAGCCA
   |||        ||    |  ||  ||||||  ||    ||||  |  | ||||||||||||
---ACG-------TTTGTAGAGATATTGGCC-TTGAAGCAGAAAATTCATTTTAAAGCCA

ATCACCTGCGATCACACACACAAACACACACACACACACACGAAAAAAAAAAACCAAC
||||     |  |||||  |||
ATCAGTCTCATTCACATACAATACTGTATATTGATTTATGCTGTGTACAAATAGAATTAT
```

Figure 1. Nucleotide alignments of chicken ESTs MCW108 and MCW188 with the human RAB6 and ZFX genes. Only the parts representing the transition from the coding region (bold) into the 3' untranslated region are shown.
[1]. Genbank accession number M28212 [2]. Genbank accession number M30608

Interestingly, two microsatellite markers (MCW108 and MCW186) derived from the chicken brain cDNA library, showing sequence similarity to genes from other species, were both mapped in the present study. MCW108 shares 77.6% nucleotide identity in a 116 basepair (bp) overlap with the human gene encoding GTP binding protein (RAB6) (Figure 1). Although the RAB-genes encode a large family of GTP-binding

proteins, overlap between the chicken and human genes is also found in the 3' untranslated region suggesting that the chicken EST is the true homologue of the human RAB6 gene. Because RAB6 is located on human chromosome 2q14-q22, this information contributes to the comparative map of the chicken. MCW186 shows a sequence similarity of 86.4% with the last exon and part of the 3' untranslated region (UTR) of the gene encoding the human zincfinger ZFX or ZFY protein (Figure.1). Because both the ZFX and ZFY genes have diverged from a common ancestral gene and have a similar genomic organization (Schneider-Gädicke *et al.* 1989), it could not be determined whether MCW186 is the homologue of ZFX or ZFY. However, also in this case the high percentage of similarity in the 3' UTR of the ZFX/ZFY gene, even continuing on the 3' side of the CA-repeat (data not shown), makes it very likely that one is not dealing with other zinc finger proteins closely related to ZFX/ZFY. Although ZFX/ZFY is located on the short arm of the sex chromosomes in eutherian mammalian species, it is located on chromosome 1 in the chicken. Also in marsupials (Sinclair *et al.* 1988), monotremes (Graves 1991) and reptiles (Bull *et al.* 1988) ZFX/ZFY is located on one of the autosomes, suggesting that the pseudo autosomal regions of the eutherian sex chromosomes have been acquired from autosomes early in eutherian radiation (Graves 1991, 1995).

Conclusion.

Although the isolation and development of microsatellites from a cDNA library has proven to be a rapid method in order to add genes to the genetic linkage map of the chicken (Ruyter-Spira *et al*, 1996), the high redundancy of $(TG)_{13}$ positive clones observed between the embryonic and brain cDNA library, resulted in only nine new markers derived from the embryonic cDNA library. Because this artefact is probably caused by the use of total embryonic cDNA, including cDNA from the embryonic brain, the screening of another cDNA library from a different tissue type for the same purpose is recommended.

References

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D. *et al.* .1995. Initial assessment of human gene diversity and expression patterns based upon 83 million basepairs of cDNA sequence. *Nature* **377** (suppl.), pp. 173-174.

Bull, J.J., Hillis, D.M. & O'Steen, S. 1988. Mammalian ZFY sequences exist in reptiles regardless of sex-determining mechanism. *Science* **242**, pp. 567-569.

Bumstead, N. & Palyga, J. 1992. A preliminary linkage map of the chicken genome. *Genomics* **13**, pp. 690-697.

Burt, D.W., Bumstead, N., Bitgood, J.J., Ponce de Leon, F.A. & Crittenden, L.B. 1995. Chicken genome mapping: a new era in avian genetics. *Trends in Genetics*. **11**, pp. 190-194.

Collins, S. 1995. Positional cloning moves from perditional to traditional. *Nature Genetics* **9**, pp. 347-350.

Crittenden, L.B., Provencher, L., Santangello, L. *et al.* 1993. Characterization of a Red Jungle Fowl by White Leghorn backcross reference population for molecular mapping of the chicken genome. *Poultry Science* **72**, pp. 334-348.

Crooijmans, R.P.M.A., van Kampen, A.J.A., van der Poel, J.J. & Groenen, M.A.M. 1993. Highly polymorphic microsatellites in poultry. *Animal Genetics* **34**, pp. 441-433.

Crooijmans, R.P.M.A., van der Poel, J.J. & Groenen, M.A.M. 1995. Functional genes mapped on the chicken genome. *Animal Genetics* **26**, pp. 72-78.

Crooijmans, R.P.M.A., van Oers, P.A.M., Strijk, J.A., van der Poel, J.J. & Groenen, M.A.M. 1996. Preliminary linkage map of the chicken (Gallus Domesticus) genome based on microsatellite markers: 77 new markers mapped. *Poultry Science* **75**, pp. 746-754.

Crooijmans, R.P.M.A., Dijkhof, R.J.M., van der Poel, J.J. & Groenen, M.A.M. 1997. New microsatellite markers in chicken optimized for automated fluorescent genotyping. *Animal Genetics* **28**, pp. 427-437.

Graves, J.A. 1991. Mammalian genome evolution: new clues from comparisons of eutherians, marsupials and monotremes. *Comparative Biochemistry and Physiology A.* **99** (1-2), pp. 5-11.

Graves, J.A. 1995. The evolution of mammalian sex chromosomes and the origin of

the sex determining genes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences.* **350** (1333), pp. 305-311.

Green, P., Falls, K. & Crooks, D. 1990. Documentation for CRI-MAP version 2.4. Washington University School of Medicine, St.Louis, MO, USA.

Hearne, C.M., Ghosh, S. & Todd, J.A. 1992. Microsatellites for linkage analysis of genetic traits. *Trends in Genetics* **8**(8), pp. 288-293.

Klein, S., Morrice, D.R., Sang, H., Crittenden, L.B. & Burt, D.W. 1996. Genetic and physical mapping of the chicken *IGF1* gene to chromosome 1 and conservation of synteny with other vertebrate genomes. *J. of Heredity* **87**, pp. 10-14.

Levitt, R.C. 1991. Polymorphisms in the transcribed 3' untranslated region of eukaryotic genes. *Genomics* **11**, pp. 484-489

McQueen, H.A., Fantes, J., Cross, S.H., Clark, V.H., Archibald, A.L. & Bird, A.P. 1996. CpG islands of chicken are concentrated on microchromosomes. *Nature Genetics.* **12**, pp. 321-324.

Morrison, M. & Griffin, W. 1985. Molecular biology of the mammalian brain. In: *Gene expression in brain.* (ed. by C. Zomzely-Neurath & W. Walker), Wiley, New York. pp. 57-98

Palmer, D.K. & Jones, C. 1986. Gene mapping in chicken-chinese hamster somatic cell hybrids. *J. of Heredity* **77**, pp. 106-108.

Ruyter-Spira, C.P., Crooijmans, R.P.M.A., Dijkhof, R.J.M., van Oers, P.A.M., Strijk, J.A., van der Poel, J.J. & Groenen, M.A.M. 1996. Development and mapping of polymorphic microsatellite markers derived from a chicken brain cDNA library. *Animal Genetics* **27**, pp. 229-234.

Schneider-Gädicke, A., Beer-Romero, P., Brown, L.G., Nussbaum, R. & Page, D.C. 1989. ZFX has a gene structure similar to ZFY, the putative human sex determinant, and escapes X inactivation. *Cell* **57**, pp. 1247-1258.

Sinclair, A.H., Foster, J.W., Spencer, J.A., Page, D.C., Palmer, M.S., Goodfellow, P.N. & Graves M. 1988. Sequences homologous to ZFY, a candidate human sex-determining gene, are autosomal in marsupials. *Nature* **336**, pp. 780-783.

Chapter 4

Bulked Segregant Analysis using Microsatellites: mapping of the Dominant White Locus in Chicken.

CAROLIEN P. RUYTER-SPIRA[1], ZHI LIANG GU[2], JAN J. VAN DER POEL[1] AND MARTIEN A.M. GROENEN[1]

[1]Department of Animal Breeding, Wageningen Institute of Animal Sciences (WIAS), Wageningen Agricultural University, Postbox 338, 6700 AH Wageningen, The Netherlands

[2]Laboratory for Poultry Genetics and Breeding, Jiangsu Provincial Poultry Science Research Institute, Sangyuan Road 46, Yangzhou, Jiangsu 225003, P.R. China

Published in; *Poultry Science* 1997, **76** pp. 386-391

Bulked Segregant Analysis using Microsatellites: mapping of the Dominant White Locus in Chicken.

CAROLIEN P. RUYTER-SPIRA, ZHI LIANG GU, JAN J. VAN DER POEL AND MARTIEN A.M. GROENEN.

Abstract.

In order to perform a linkage study, the genotypes of a large number of individuals from a segregating population need to be determined. In case the phenotype to be mapped is influenced by a single locus or a major gene, sampling of the DNA from individual animals with the same phenotype into a single pool (bulked segregant) can reduce the number of typings. In this study we used bulked segregant analysis in order to map the dominant white locus in chicken. In a pilot experiment, we showed that allele frequencies can be accurately estimated from pooled samples using fluorescently labelled microsatellite markers. A segregating population for the dominant white locus was obtained by performing a cross between a white male chicken (genotype *Ii* for dominant white) and a black female chicken (*ii*). The resulting progeny of 21 white and 18 black chickens was divided in two pools. Genotypes for both the parents and the pools were determined using 168 fluorescently labelled microsatellite markers, of which 68 were informative. The relative allele frequencies between the pools were estimated for these 68 informative markers. One marker (MCW188) was found to segregate with the dominant white locus. Subsequent typing of all individuals from this cross and an additional 148 animals from five different families showed only two recombinants between the marker and the dominant white locus resulting in a $LOD_{linkage}$ score (log_{10} of odds) of 36. Using the pooled DNA approach, the dominant white locus was successfully mapped on linkage group 22 of the East Lansing reference family at a distance of 2 cM from MCW188.

Introduction.

In animal genetics, the generation of genetic maps for livestock is progressing very rapidly. In linkage studies, genetic markers are used for detecting genetic variation

among different genotypes. Particularly, the genetic variation underlying production traits and heritable diseases is of high importance. In order to identify markers that are tightly linked to these loci, large segregating populations need to be analysed with molecular markers covering the whole genome. For certain traits or loci the large number of DNA samples that need to be analysed can be reduced using a DNA pooling strategy (Darvasi and Soller, 1994; Tinker *et al.*, 1994; Wang and Patterson, 1994). This procedure involves the construction of two pools based on DNA from individuals derived from a segregating population differing for a phenotypic trait or a specific region of the genome. The test for linkage is based on the estimation of differences in marker allele frequencies between the alternative pools. Michelmore *et al.*(1991) successfully used a pooling strategy (bulked segregant analysis) to develop markers linked to disease-resistance genes. In that study they used restriction length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD) as markers. Giovannoni *et al.*(1991) used the same technique for isolating markers from a specific chromosomal interval.

Microsatellite markers are very useful markers in linkage studies because they can detect a large number of alleles. Therefore, the probability that they are polymorphic in a family is high. Furthermore, the analysis is based upon the polymerase chain reaction (PCR), which makes the determination of genotypes easy to perform, the genotypes easy to score and requires a minimum amount of genomic DNA. Application of fluorescently labelled PCR primers allows the PCR products to be analysed on an automated DNA sequencer.

Pacek *et al.* (1993) used this method to amplify and analyze short tandem repeat and variable number of tandem repeat regions from pooled samples. They found that the allele frequencies from the pooled samples determined by quantitative analysis were in good agreement with the allele frequencies calculated from individual samples. Khatib *et al.* (1994) and Crooijmans *et al.* (1996b) showed that differences in allele frequency of microsatellites between DNA pools can be successfully identified. However, significant differences are rarely obtained when differences in marker allele frequency are small (< 0.10).

The aim of our study was to investigate the applicability of bulked segregant analysis in combination with the use of microsatellite markers in a linkage study. We chose to

study the dominant white locus in the chicken as a test case because this is a single gene trait with full penetrance and the phenotypes are easy to follow in a segregating population. Development of closely linked markers to this trait is also of economic importance, since in the chicken meat industry black broilers are not favoured. The presence of the recessive allele for dominant white in the population makes it necessary to screen male chickens before they can be used in production lines. A genetic marker for this trait would shorten the generation interval by making a test cross superfluous. In a pilot experiment, reproducibility and reliability of the estimation of allele frequencies was studied. Secondly, two different DNA pools consisting of either black or white chickens, derived from a single cross, were screened with 168 microsatellite markers. One of these markers was closely linked to the dominant white locus.

MATERIALS AND METHODS.

Source of DNA

The DNA samples used for the initial experiment, in which reproducibility and reliability of the estimation of allele frequency was tested, were derived from the East Lansing reference population (Crittenden *et al.*, 1993). The DNA samples used for the linkage experiment were obtained from chicken families segregating for the dominant white allele. The DNA from the black and white offspring of this family was divided in two pools.

DNA Extraction and Pool Preparation

DNA was extracted from 25 µL blood samples using a genomic DNA isolation kit (Puregene, Gentra Systems Inc., Research Triangle Park, NC 27709). DNA concentration was determined spectrophotometrically, and DNA was diluted to a final concentration of 10 ng/µL. Because an equal DNA concentration for all individuals present in the pools is crucial for accurately estimating allelic frequencies, the final concentration was checked and adjusted if necessary . Finally, from all individuals an equal aliquot was added to the corresponding pool.

Polymerase Chain Reactions

The PCR was performed in a total volume of 20 µL containing 100 ng genomic DNA, 1.5 mmol $MgCl_2$, 50 mmol KCl, 10 mmol Tris.HCl pH=8.3, 1 mmol tetramethyl-ammoniumchloride (TMAC), 0.1% Triton X-100, 0.01% gelatin, 200 mmol dNTP, 0.25 unit Goldstar polymerase (Eurogentech S.A., Ougree, 4102, Belgium.) and 30 ng of each primer. One of the PCR primers was fluorescently labelled with HEX, TET or FAM phosphoramidites (Perkin Elmer, Applied Biosystems Division, Foster City, CA 94404). The PCR was performed for 5 min at 94°C followed by 35 cycles of 30 s at 94°C, 45 s at 55°C, 90 s at 72°C, with a final elongation step of 10 min at 72°C. For some microsatellite markers it was necessary to perform the annealing step at 50°C. A total of 168 Microsatellite markers (Crooijmans *et al.* 1993, 1994, 1995, 1996a) and Ruyter-Spira *et al.*, 1996) were tested on both pools and parents.

Product Analysis

The fluorescent PCR products were separated on 6% denaturing polyacrylamide gels using an Automated Laser Fluorescent (ALF) DNA sequencer (Pharmacia LKB Biotechnology, Uppsala, Sweden). Between 0.2 and 5 µL of the PCR products were analysed after 5 min denaturation in a 50% formamide solution containing blue dextran. In each lane, three PCR products differing in size range were loaded, together with a standard molecular size marker. Allelic frequencies in the pools were estimated using the A.L.F. Fragment Manager program (Pharmacia LKB Biotechnology, Uppsala, Sweden). For markers that showed a marked difference in allelic frequencies between the pools, all individuals were typed separately.

Linkage Analysis

The $LOD_{linkage}$ scores (log_{10} of odds) and distances were calculated using the linkage analysis software program CRI-MAP version 2.4 (Green *et al.*, 1990). A minimum $LOD_{linkage}$ score of three was taken to indicate linkage.

Results and discussion.

Reproducibility and Reliability of the Estimation of Allele Frequencies in DNA Pools.

In a pilot experiment, we tested whether the pooled sample approach was suitable

for distinguishing differences in allele frequencies. DNA from animals of the East Lansing reference family (Crittenden *et al.*, 1993) was divided into four different pools based on their genotype for microsatellite ADL20 and MCW112 respectively (Figure 1). The first two pools were typed for microsatellites MCW61 and MCW68, which are located on either side of ADL20 on the genetic linkage map. Pools 3 and 4 were genotyped for microsatellite MCW58. All amplifications were performed twice. A third amplification was performed in a replicate pool.

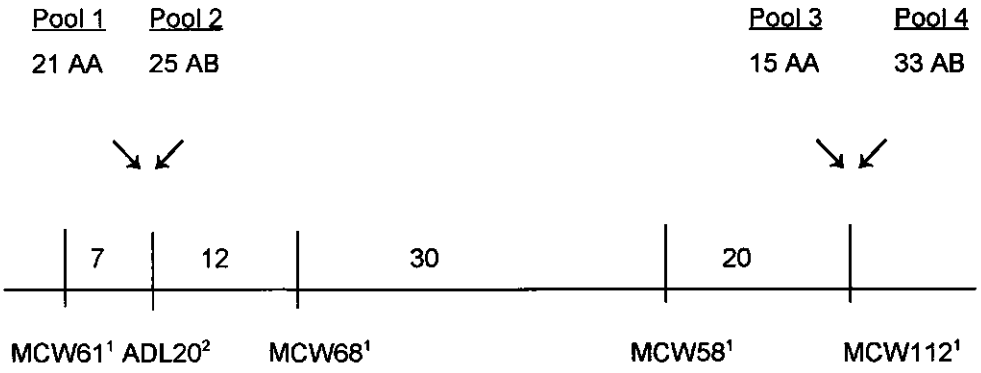


Figure 1. Part of linkage group E1 of the East Lansing reference family (Crittenden *et al.* 1993). Distances between markers are indicated in centiMorgans. Pools are based on their genotype (AA or AB) for marker ADL20 (Pool1 and Pool2) and MCW112 (Pool3 and Pool4). 1. Crooijmans *et al.* 1996a. 2. Cheng & Crittenden 1994.

For all pools, allele frequencies were estimated and compared with real allele frequencies based on individual typings (Figure 2A). The results show that estimated allele frequencies are in good agreement with real allele frequencies, and the results are both reproducible within pools and between replicate pools. In the curves representing the allele frequencies of the pools (Figure 2B), differences can be clearly distinguished, even for pools 3 and 4, where the marker is located at a genetic distance of 20 cM. For MCW58 differential amplification of the A and B alleles was observed in individual animals (data not shown). Therefore, a correction based on this difference was made for the estimation of allele frequencies in the pools. When differential

amplification for three different alleles was observed, an appropriate correction factor was difficult to estimate (data not shown). Because in this experiment values for allele frequencies were not used for the calculation of real allele frequencies, this was not a problem.

2A

		MCW 68 Pool 1	Pool 2	MCW 61 Pool 1	Pool 2	MCW 58 Pool 3	Pool 4
REAL	indiv.typ	20BB/1AB	3BB/22AB	20AA/1AB	2AA/23AB	13AA/2AB	9AA/24AB
	all.freq	2:98	44:56	98:2	54:46	93:7	64:36
ESTIMATED	all. Freq I	5:95	40:60	100:0	58:42	90:10	65:35
	II	5:95	40:60	100:0	55:45		
	III	5:95	40:60	100:0	54:46		

2B

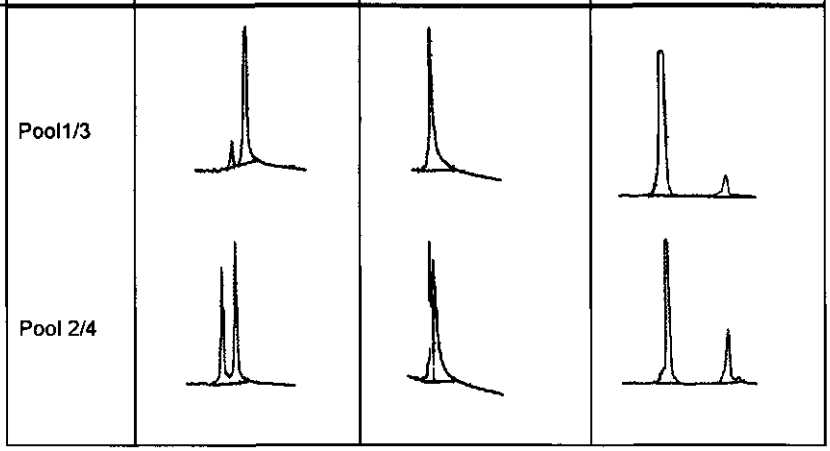


Figure 2. Comparison of real- and estimated allele frequencies. A) Allele frequencies are the relative proportion of the A and B alleles present within the pools given in percentages. Real allele frequencies (all. freq.) are based on individual genotypes (indiv. typ.) present within a pool. Estimated allele frequencies are derived from the estimations of the surface below the curves representing the alleles. Estimated allele frequencies I and II are derived from the estimations within the pool, allele frequency III is derived from amplifications in a replicate pool. B) The separation of the alleles for the microsatellites on the automatic DNA sequencer

The dominant white locus.

The aim of this experiment was to find a microsatellite marker linked to the dominant white locus in chicken using bulked segregant analysis. In order to obtain a family segregating for the dominant white locus, a white male chicken (genotype *Ii* for dominant white) was crossed with a black female chicken (*ii*). This resulted in a progeny of 21 white and 18 black chickens which were divided in two pools. Both parents and

pools were typed for 168 fluorescently labelled microsatellite markers. Because the segregation of the alternative dominant white alleles (*I/i*) in the father had to be followed, only the 68 markers that were heterozygous in this parent were informative. In order to detect linkage, we looked at differences in allele frequencies between the pools and compared them with the expected allele frequencies in an unlinked situation based on the genotype of the parents for the respective marker. The analysis resulted in one strong candidate, microsatellite MCW188 (Ruyter-Spira *et al.*, 1996), which showed no recombinants in the pooled samples. This was confirmed by individually typing the progeny. In Figure 3, both genotypes for dominant white and MCW188 are shown for the parents and the pooled samples. The allele frequencies for the remaining informative markers did not differ more than 13 % from the expected allele frequencies per pool present in an unlinked situation. No false positives were detected. Because MCW188 is located on linkage group 22 (E22) in the East Lansing reference family (Ruyter-Spira *et al.*, 1996), the dominant white locus can also be assigned to this linkage group. In this population a linkage distance of 16.3 cM was calculated between both loci. No other microsatellites are present in this region.

In order to get a better estimation of the position of the dominant white locus with regard to the position of MCW188 a total of 18 crosses between white commercial broiler males (*I?*) and black test females (*ii*) were performed. The offspring of nine families contained black chickens indicating that the males carried the *Ii* alleles for dominant white. In four of these nine families, the genotype of the fathers for MCW188 was either BB or AA. The segregation of the dominant white alleles was followed in the remaining five families, for which the fathers were heterozygous (AB) for MCW188. The offspring consisted of a total of 148 animals. Only two recombinants were detected resulting in a LOD score of 36 and an estimated distance between MCW188 and the dominant white locus of 2 cM.

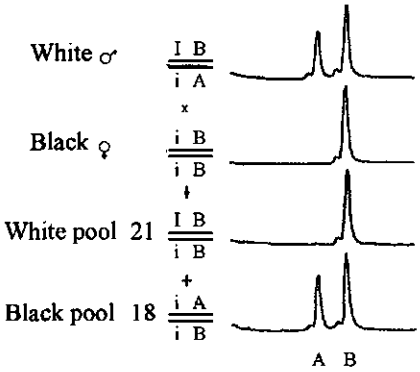


Figure 3. Genotypes for the dominant white locus (*I/i*)and microsatellite MCW88(A/B) for the parents and pools in the family segregating for the dominant white colour. All animals in the white pool turned out to be BB for MCW188 and all animals in the black pool turned out to be AB for this marker

Conclusions.

In this study we showed that the pooled DNA approach is a reliable method to perform a linkage experiment to locate a single gene trait. The combination with fluorescently labelled microsatellites makes this method very sensitive in detecting linked markers at greater genetic distances, reducing the amount of false negatives. When only two alleles are involved, absolute allele frequencies can be estimated accurately. Therefore, the number of recombinants within a pool can be calculated, which makes it possible to directly estimate genetic distance between marker and trait. Because the correction factor for differential amplification of alleles is difficult to estimate if more than three alleles are involved, absolute allele frequencies and number of recombinants can not be accurately estimated. Still, comparing differences in relative allele frequencies between the pools can be used to determine linkage. This can be followed by individual typing to determine exact recombination frequencies. Following this procedure a huge reduction in the amount of typings can be achieved.

ACKNOWLEDGMENTS.

The authors wish to thank Euribrid B.V. for supplying the blood samples from the chickens segregating for the dominant white locus.

References.

Cheng, H.H. & Crittenden, L.B. 1994. Microsatellite markers for genetic mapping in the chicken. *Poultry Science* **73.** pp. 539-546.

Crooijmans, R.P.M.A.,van Kampen, A.J.A., van der Poel, J.J. & Groenen, M.A.M. 1993. Highly polymorphic microsatellite markers in poultry. *Animal Genetics* **3.** pp. 441-443.

Crooijmans, R.P.M.A., van Kampen, A.J.A., van der Poel, J.J. & Groenen, M.A.M. 1994. New microsatellite markers on the linkage map of the chicken genome. *J. of Heredity* **85.** pp. 410-413.

Crooijmans, R.P.M.A., van Kampen, A.J.A., an der Poel, J.J. & Groenen, M.A.M. 1995. Functional genes mapped on the chicken genome. *Animal Genetics* **26.** pp. 73-78.

Crooijmans, R.P.M.A., van Oers, P.A.M., Strijk, J.A., van der Poel, J.J. & Groenen, M.A.M. 1996a. Preliminary linkage map of the chicken (Gallus domesticus) genome based on microsatellite markers: 77 new markers mapped. *Poultry Science* **75.** pp. 746-754.

Crooijmans, R.P.M.A., Groen, A.F., van Kampen, A.J.A., van der Beek, S., van der Poel, J.J. & Groenen, M.A.M. 1996b. Microsatellite polymorphism in commercial broiler and layer lines estimated using pooled blood samples. *Poultry Science* **75.** pp. 904-909.

Crittenden, L.B., Provencher, L., Santangelo, L., Levin, I., Abplanalp, H., Briles, R.W., Briles, J.B. & Dodgson, J.B. 1993. Characterization of a red jungle fowl by white leghorn backcross reference population for molecular mapping of the chicken genome. *Poultry Science* **72** pp. 334-348.

Darvasi, A. & Soller, M. 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics* **138.** pp. 1365-1373.

Giovannoni, J.J., Wing, R.A., Ganal, M.W. & Tanksley, S.D. 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Research* **19.** No. 23 pp. 6553-6558.

Green, P., Falls, K. & Crooks, D. 1990. Documentation for CRI-MAP, version 2.4. St.Louis: Washington University School of Medicine.

Khatib, H., Darvasi, A., Plotski, Y. & Soller, M. 1994. Determining relative microsatellite allele frequencies in pooled DNA samples. *PCR methods Applications* **4.** pp. 13-18.

Michelmore, R.W., Paran, I. & Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Nat. Acad. Sci. USA* **88.** pp. 9828-9832.

Pacek, P., Sajantila, A. & Syvänen, A.1993. Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. *PCR methods Applications* **2.** pp. 313-317.

Ruyter-Spira, C.P., Crooijmans, R.P.M.A., Dijkhof, R.J. van Oers, P.A.M., Strijk, J.A. van der Poel, J.J. & Groenen, M.A.M. 1996. Development and mapping of polymorphic microsatellite markers derived from a chicken brain cDNA library. *Animal Genetics* **27.** pp 229-234.

Tinker, N.A., Mather, D.E. & Fortin, M.G. 1994. Pooled DNA for linkage analysis: practical & statistical considerations. *Genome* **37.** pp. 999- 1004.

Wang, G.L., & Paterson, A.H. 1994. Assessment of DNA pooling strategies for mapping of QTLs. *Theor. Appl. Genet.* **88** pp. 355- 361.

Chapter 5

The HMGI-C Gene is a Likely Candidate for the Autosomal Dwarf Locus in the Chicken.

Carolien P. Ruyter-Spira[1], Ad J.C. de Groof[1], Jan J. van der Poel[1], Jos Herbergs[1], Julio Masabanda[2], Ruedi Fries[2] and Martien A.M. Groenen[1]

[1] *Department of Animal Breeding, Wageningen Institute of Animal Sciences (WIAS), Wageningen Agricultural University, Postbox 338, 6700 AH, Wageningen, The Netherlands.*

[2]*Technische Universität München, Lehrstuhl für Tierzucht, 85350 Freisins-Weihenstephan, Germany.*

Published in: *Journal of Heredity*, 1998, **89(4)** pp. 295-300

The *HMGI-C* Gene is a Likely Candidate for the Autosomal Dwarf Locus in the Chicken.

Carolien P. Ruyter-Spira, Ad J.C. de Groof, Jan J. van der Poel, Jos Herbergs, Julio Masabanda, Ruedi Fries and Martien A.M. Groenen.

Abstract.

In order to map the autosomal dwarf (*adw*) locus in the chicken, 11 segregating families were created. Initially, five of these families were used for a linkage experiment in which the genome was scanned with microsatellites using a technique called bulked segregant analysis. Subsequently, animals from 11 families were typed individually for microsatellites that appeared to be linked. We were able to detect genetic linkage of the *adw* locus to five different microsatellite markers on chromosome 1, the closest showing a recombination fraction of only 0.03. (LOD-score 32.12). In mice, the phenotype "pygmy" shows a striking similarity to the autosomal dwarf phenotype in chickens, both having a disproportionately large head. The pygmy locus has been mapped on mouse chromosome 10 and found to represent a mutation in the gene coding for High Mobility Group Protein I-C (*HMGI-C*). Considering the synteny between regions of chicken chromosome 1, mouse chromosome 10 and human chromosome 12, and taking into account both the phenotypic characteristics and the mode of inheritance of the chicken *adw* and the mouse pygmy loci, the *HMGI-C* gene is a major candidate gene for the *adw* locus in the chicken. Fluorescence *in situ* hybridization of metaphase chromosomes with the chicken *HMGI-C* gene as a probe, showed that the chicken *HMGI-C* gene is indeed closely linked to marker *LEI146* on chromosome 1.

Introduction.

Growth is the result of complex interactions between multiple genetic and exogenous factors. For a better understanding of this complex trait the detailed analysis of existing growth disorders that are based on single gene mutations can be very informative. A spontaneous autosomal recessive mutation in Leghorn chickens affecting overall body size (except the head) was reported by Cole (1973). Adult birds

were reduced in body weight by 30 percent, and were distinguishable from non-dwarf chickens by 6-8 weeks of age. The endocrine basis underlying the reduced growth in autosomal recessive dwarf (*adw*) chickens has been studied in the past (Scanes *et al.* 1983; Huybrechts *et al.* 1984). These studies showed that plasma concentrations of immunoreactive growth hormone and insulin-like growth factor I were unaffected. Furthermore, the so-called autosomal dwarf (*adw*) chickens appeared to be slightly hypothyroid having plasma concentrations of both T4 and T3 somewhat lower than in the control strain. However, these minor changes did not suggest that this was the major cause of the reduced growth rate.

In mice, a dwarf phenotype called pygmy, showing a striking resemblance with the phenotypic characteristics of *adw* chickens, was described by MacArthur (1944). A transgenic insertional mutant (Xiang *et al.* 1990), allelic to pygmy mice, was characterized (Benson & Chada, 1994). Adult mice, homozygous for the mutation, showed a 40% reduction in body size, the head being disproportionately larger. As in *adw* chickens, the small size in pygmy mice could not be explained by aberrations in the growth hormone-insulin-like growth factor endocrine pathway. Recently, Zhou *et al.* (1995) showed that the pygmy phenotype arises from the inactivation of High Mobility Group Protein I-C (*HMGI-C*). *HMGI-C* has been shown to be involved in the regulation of cell proliferation (Zhou *et al.* 1995). In wild type mice the *HMGI-C* gene was found to be expressed only during embryogenesis in most tissues and organs. Remarkably, HMGI-C expression was not seen in the embryonic brain except in a small section of the forebrain, coinciding with the observation that most tissues in pygmy mice are 40-50% smaller than wild-type tissues, the only tissue of normal size being the brain (Benson and Chada, 1994).

As an initial step towards dissecting the genetic basis of autosomal dwarfism in the chicken, a linkage study was performed using families segregating for the *adw* locus. Because autosomal dwarfism is influenced by a single gene with complete penetrance, this trait is ideally suited to be analysed using bulked segregant analysis (Ruyter-Spira *et al.* 1997). This method involves the construction of two pools of DNA from animals derived from a segregating population differing for the phenotypic trait of interest. These pools are genetically dissimilar in the selected region but heterozygous at all

other regions (Michelmore *et al.* 1991). The test for linkage is based on the estimation of differences in marker allele frequencies between the alternate pools. Because only the genotypes of both pools and parents need to be determined a huge reduction in the required number of typings can be achieved.

Applying this method to a chicken population in which the *adw* locus is segregating, we were able to map the *adw* locus to chromosome 1. As a second step we determined the chromosomal location of candidate gene *HMGI-C* using fluorescence *in situ* hybridization on metaphase chromosomes.

Materials and Methods.

Experimental Cross.

A population, segregating for the *adw* locus, was obtained by crossing White Leghorn cocks, homozygous recessive at the putative autosomal dwarfism locus (*adw/adw*), with Cornish hens carrying the wild type allele (*ADW/ADW*). Eleven different cocks and hens of the F1 (*ADW/adw*) were crossed resulting in an offspring of 225 animals. At 6 weeks of age the F2 animals were phenotyped based on their bodyweights and the size and form of the heads.

Animals used for Linkage Analysis.

Due to the large variance in bodyweights it was sometimes difficult to distinguish *adw* chickens from non-*adw* chickens. Therefore, only animals for which the phenotypes could clearly be classified were included in the linkage experiments. Five families with the largest number of offspring were selected for bulked segregant analysis. The offspring was divided in dwarf and non-dwarf pools, each pool containing six to eleven individuals. For the linkage experiment based on individual genotypings, 145 animals with clear phenotypes were used.

DNA Isolation and Pool Preparation.

Blood samples in EDTA were taken from chickens of all F2 families and their parents, and frozen at -80°C until analysis. For each of the five families selected for bulked segregant analysis, blood samples from dwarf, respectively non-dwarf F2 animals, were pooled. DNA from pools and individuals was isolated using a genomic

DNA isolation kit (Puregene, Gentra Systems Inc., Minneapolis, MA, USA) and diluted in Tris-EDTA (pH=8) to a final concentration of 10 µg/ml.

Polymerase Chain Reactions and Product Analysis.

A set of 111 microsatellite markers (Crooijmans *et al.* 1993, 1994, 1995, 1996; Ruyter-Spira *et al.* 1996), covering about 80 percent of the chicken genome, was selected for a total genome scan. PCR was performed according to Ruyter-Spira *et al.* (1997). The amplification products were separated on a 6% denaturing polyacrylamide gel (Sequagel-6, National Diagnostics, Atlanta, G, USA) using an automated laser sequencer (Automated Laser Fragment sequencer, Pharmacia LKB Biotechnology, Uppsala, Sweden and ABI 373A, Perkin Elmer, Applied Biosystems Division, Foster City, CA). The curves representing the allelic frequencies of the pooled samples were visually inspected. In case a marker showed large differences between the allelic frequencies of both pools, allelic frequencies were calculated using the A.L.F. fragment manager program (Pharmacia LKB Biotechnology, Uppsala, Sweden) and if possible, corrected for differential amplification. Subsequently, individual progeny samples from all 11 families (145 animals) were tested for these markers to determine exact allelic frequencies in order to estimate genetic distances.

Linkage Analyses.

Twopoint and multipoint linkage analysis were performed using the computer program LINKAGE (Lathrop *et al.* 1984). A minimum $LOD_{linkage}$ score of three was taken to indicate linkage.

Isolation of Chicken HMGI-C and Microsatellite LEI146 Genomic Clones.

In order to isolate genomic clones corresponding to the *HMGI-C* gene, 2.6×10^5 plaques of a LambdaEMBL3 chicken genomic library with an average insert size of 15 kb (Clontech Inc., Palo Alto, CA, USA), were screened by hybridization using a cDNA probe representing the entire coding region of the murine *HMGI-C* cDNA (Manfioletti *et al.* 1991). The complete insert of one of the five clones which remained positive after tertiary screening, was subcloned into the *Sal*I site of the pBluescript II (KS+) vector,

transformed into Escherichia coli XLI-blue, and further characterized by restriction mapping and Southern blot analysis. Plasmid DNA was prepared by the alkaline lysis method using the Quiaprep plasmid isolation kit (Quiagen Inc., Chatsworth, CA). Two adjacent *PstI* fragments, both containing a coding part of the chicken *HMGI-C* gene, were subcloned and partially sequenced using the autoread sequence kit (Pharmacia, Uppsala, Sweden). Sequences were determined on the Automated Laser Fragment sequencer and further analysed using the Microgenie program (Beckman Instruments, Palo Alto, CA, USA). To obtain genomic clones containing the *LEI146* marker sequence, the same library was screened using a *LEI146* PCR fragment as the probe.

Fluorescence in situ Hybridization.

Metaphase chromosomes were obtained from a culture of chicken embryo fibroblasts after treatment with a hypotonic solution (75mM KCl) and fixed in methanol-acetic acid (3:1).The spreads were stained with quinacrine (Caspersson *et al.* 1970), analysed by fluorescence microscopy (Zeiss Axiophot-2) and photographed using a cooled CCD camera. For fluorescence *in situ* hybridization the *LEI146* probe was labelled with biotin-16-dUTP (Boehringer, Mannheim) and the *HMGI-C* probe was labelled with digoxigenin-11-dUTP (Boehringer, Mannheim). The biotinylated probe was detected by subsequent incubation with AvFitc, BIOGAA and AvFitc (Vektor Brunschwig Chemie, Amsterdam, The Netherlands). The digoxigenin labelled probe was detected by MAdig (Sigma, St.Louis, MO, USA), RAMTRITC and SWARTRITC (both from Vektor Brunschwig Chemie, Amsterdam, The Netherlands). The DNA was counterstained with DAPI. In the pre-photographed spreads the signals were detected by fluorescence microscopy and pictures were taken by a cooled CCD camera.

Results.

Bulked Segregant Analysis

Of the 111 microsatellites selected for the total genome scan, 31 were tested using bulked segregant analysis. During visual inspection of the allelic frequencies, a marked difference in allelic frequencies between the dwarf and non-dwarf pools was observed for markers, *MCW18, MCW43* (Figure 1) and *MCW112* in all five families. Markers *MCW43, MCW18* and *MCW112* map to the same region on chromosome 1. Two

markers, not linked to those mentioned above, also showed differences in allelic frequency between the dwarf and non-dwarf pools in one and two families. However, because these differences were not consistent in other families, and considering the fact that both markers mapped to a different chromosomal region, these markers were dismissed for further analysis.

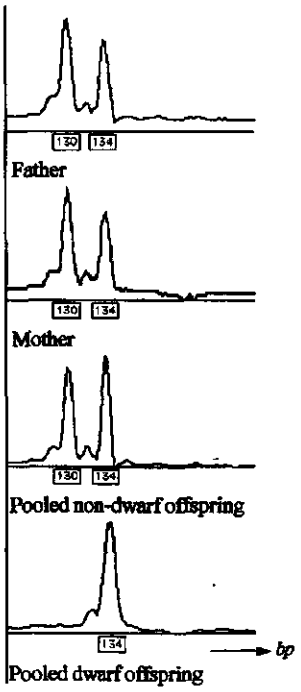


Figure 1. Genotypes for microsatellite *MCW43* for both parents, pooled dwarf and non-dwarf offspring in one of the families in which the *adw* locus is segregating. A large difference in allelic frequency for both pools is observed. Allele sizes are shown in basepairs (bp).

Individual Linkage Analysis.

In order to verify the results from the bulked segregant analysis experiment and in order to obtain a more accurate estimation of the location of the *adw* locus on the genetic linkage map, DNA samples from 145 dwarf and non-dwarf chickens were subjected to individual linkage analysis. Five microsatellite markers (*MCW43, MCW18, ADL307, LEI71, LEI146*), all being located in the region of the *adw* locus, were analysed. One of these markers (*LEI146*) was located in the region between *MCW43* and *MCW18*. Using two point analysis, positive linkage with the *adw* locus was found for all markers except *ADL307* and *LEI71*, the recombination fraction between *LEI146* and the *adw* locus being as low as 0.03 ($LOD_{linkage}$ score= 31.98). A multipoint analysis revealed the most likely position of the *adw* locus with respect to the markers present on linkage group WAU1 (Figure 2).

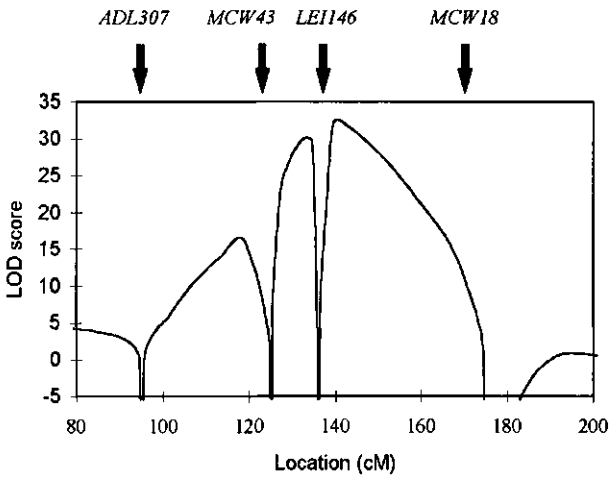


Figure 2. Multipoint lod score curves obtained for the *adw* locus with respect to microsatellite markers on chicken chromosome 1 p, close to the centromere, using the LINKMAP program of the linkage package.

Cloning, Sequencing and Fluorescence in situ Hybridization of the Chicken HMGI-C Gene

Because the *HMGI-C* gene is a likely candidate for the *adw* locus in the chicken (see discussion) a genomic chicken library was screened using a murine *HMGI-C* cDNA probe (Manfioletti *et al.* 1991). Comparison of murine with human *HMGI-C* cDNA sequences revealed that *HMGI-C* encodes a protein which is highly conserved across species (Patel *et al.* 1994). The same library was also screened with a PCR fragment obtained with microsatellite *LEI146*. For both probes positive clones, approximately 15 kb in length, were isolated and subcloned. Two adjacent *Pst*I fragments, containing coding sequences of the chicken *HMGI-C* gene, were subcloned and partially sequenced. Parts of these clones showed a 100% match with sequences of exon 5 derived from a chicken *HMGI-C* cDNA clone which contains the complete coding part for *HMGI-C*, including a region which is *HMGI-C* specific (Ruyter-Spira *et al.* in prep.). Comparing the deduced amino acid sequence from chicken *HMGI-C* exon 5 with the human *HMGI-C* and *HMGY* (another member of the highly conserved *HMGI* class of proteins) sequences, confirms that we indeed cloned *HMGI-C* (Figure 3). The *HMGI-C* genomic clone was either separately, or simultaneously with the clone containing the *LEI146* sequence, used for fluorescence *in situ* hybridization of metaphase chicken

chromosomes using different fluorescent dyes for each probe. As a result a specific signal for both probes was found on chromosome 1p proximal to the centromere (Figure 4).

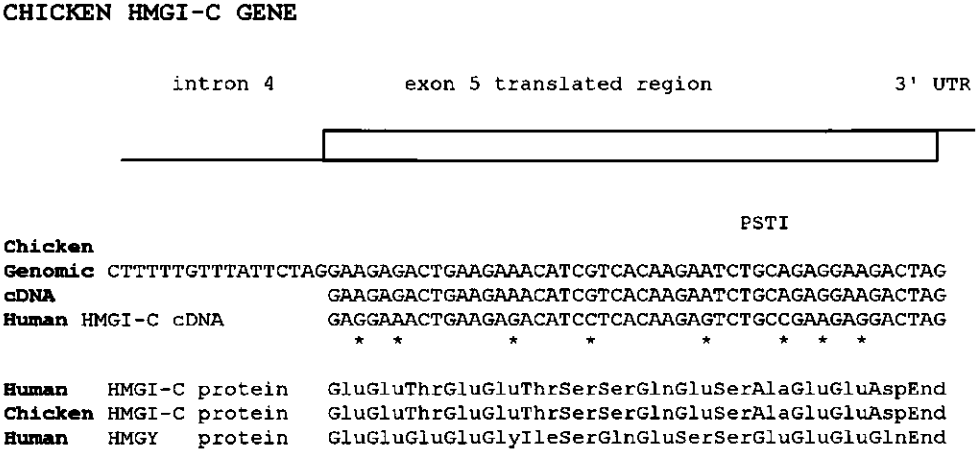


Figure 3. Nucleotide sequence of a part of exon 5 of the chicken *HMGI-C* gene derived from a cDNA clone and two genomic subclones. Sequences are compared with the nucleotide respectively protein sequence of human *HMGI-C* and *HMGY* (Patel *et al.*, 1994). Differences are indicated with *.

Discussion.

In this study we successfully mapped the autosomal dwarf (*adw*) locus in the chicken to chromosome 1 using bulked segregant analysis. This technique resulted in the identification of positive linkage between the *adw* locus and three microsatellite markers, and even linkage of the *adw* locus to a marker at a genetic distance of approximately 36 cM (*MCW18*) could still be clearly detected. Like in a previous experiment, in which bulked segregant analysis was used to map the dominant white locus in the chicken (Ruyter-Spira *et al.* 1997), the estimated allelic frequencies of the pools in the present experiment were in good agreement with allelic frequencies calculated from individual genotypes for these markers (data not shown). This makes the pooled DNA approach a reliable method to perform a linkage experiment in order to locate single gene traits. Because only 25% of the offspring consisted of dwarf chickens, and due to the difficulties to distinguish dwarf animals from non- dwarf animals, DNA pools of relatively small size were used. This increases the probability that the pools would share areas of homozygosity other than the targeted interval,

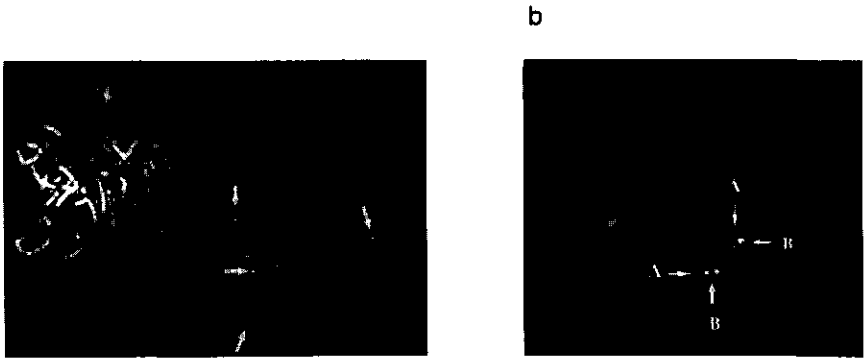


Figure 4. Fluorescence *in situ* hybridization of chicken metaphase chromosomes with the chicken *HMGI-C* probe (Figure 4a), and *HMGI-C* (A) and *LEI146* (B) as cohybridizing probes (Figure 4b). A specific signal for both probes can be seen on chromosome 1p proximal to the centromere.

resulting in increased numbers of false positively linked markers (Giovannoni *et al.* 1991). Indeed, false positive linkage was found for two microsatellite markers located at two different positions on the genetic linkage map. However, the use of five different mapping families was effective to recognize this phenomenon. Individual linkage analysis of 145 dwarf and non-dwarf chickens with additional markers, showed that the *adw* locus in the chicken is closely linked to microsatellite marker *LEI146* on the short arm of chromosome 1 (Figure 5). This region also harbors the genes coding for lysozyme, histone 5 and Insulin Like Growth Factor 1 (*IGF1*) (Klein *et al.* 1996). Although *IGF1* has an important role in the control of growth and metabolism, plasma concentrations of immunoreactive *IGF1*, thyroxine, tri-iodothyronine and growth hormone were shown to be unaffected in autosomal dwarf chickens (Scanes *et al.* 1983; Huybrechts *et al.* 1984). Therefore, reduced growth in *adw* chickens is unlikely to be related to the somatotroph axis and *IGF-I* was not considered as a major candidate gene for the dwarf phenotype. However, considering the fact that only immunoreactive levels in stead of biologically active levels of *IGF1* in *adw* chickens have been measured, we can not completely rule out the possibility that *IGF1* is playing a role in the reduced growth of *adw* chickens. Both the Lysozyme and IGFI gene have been mapped in human and mouse, and shown to be part of a conserved syntenic

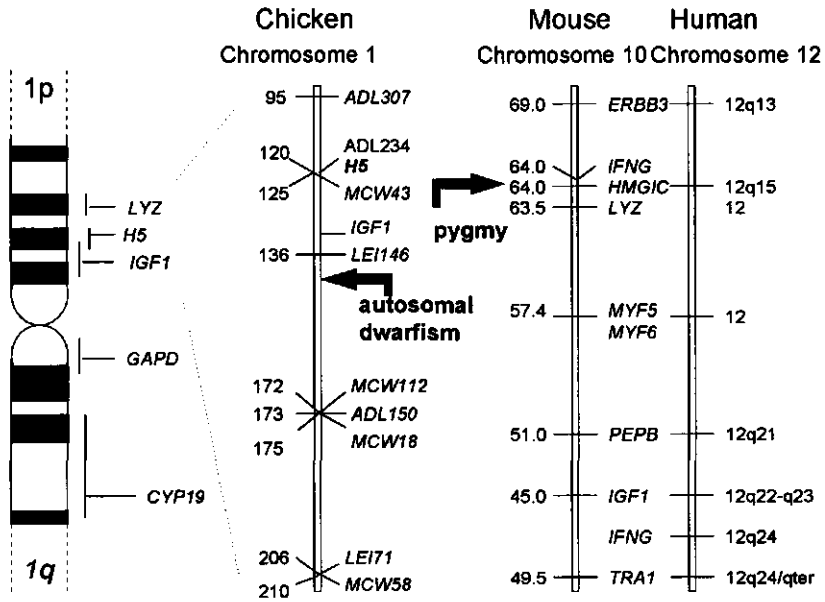


Figure 5. The autosomal dwarf locus in the chicken is located close to marker *LEI146* on chromosome 1p in a segment which shows conserved synteny with parts of mouse chromosome 10 and human chromosome 12. A part of linkage group WAU1 (Groenen *et al.*, in prep.) surrounding the *adw* locus, is shown. Because *H5* and *IGF1* have not originally been mapped on the WAU family, but do show linkage to markers that have been mapped in both the East Lansing and WAU family, their location is shown in italics. A selection of genes which are present on both human and mouse chromosomes are indicated using human nomenclature.

group on mouse chromosome 10 and human chromosome 12 (Figure 5)(Klein *et al.* 1996). Interestingly, in mice this region also harbors the pygmy locus. Pygmy mice display a dwarf phenotype which, like autosomal dwarfism in the chicken, can not be explained by aberrations in the growth hormone-*IGF1* endocrine pathway (Benson & Chada, 1994). Zhou *et al.* (1995) showed that a mutation in the gene encoding the *HMGI-C* Protein is responsible for the pygmy phenotype in mice. They demonstrate that *HMGI-C* expression is restricted to the embryonic stage and that it serves as a regulator of cell proliferation. In the human, *HMGI-C* is also part of the conserved chromosomal gene cluster on chromosome 12. Like *adw* chickens, pygmy mice also show craniofacial defects, however the mutation segregates as an incomplete recessive trait whereas *adw* in the chicken has been reported to be a complete recessive trait. The large variance in bodyweights observed in the present experiment, especially for the non-dwarf chickens, suggested that *adw* might also be inherited in

an incomplete recessive manner. To test this hypothesis, we typed the entire offspring (225 animals) from all 11 families in which *adw* was segregating for the most nearby marker (LEI146). Homozygous non-dwarf animals showed a tendency to have slightly increased body weights compared to heterozygous non-dwarf animals (Figure 6), however the difference between the bodyweights of heterozygous and homozygous non-dwarf animals was not significant. This might be due to the fact that bodyweights were measured at 6 weeks of age whereas chicken only reach their adult bodyweights after a period of 18 weeks. As shown in a developmental analysis in pygmy mice, homozygous dwarfs can first be identified on the basis of reduced body weight at day 15.5 of gestation, whereas heterozygotes can only be distinguished from wild type mice at the age of 10 weeks when growth is completed (Benson & Chada, 1994).

Considering the conserved synteny between the region of the *adw* locus on chicken chromosome 1, mouse chromosome 10 and human chromosome 12, and taking into account both the phenotypic characteristics and the mode of inheritance of the *adw* locus in chickens and the pygmy locus in mice, the *HMGI-C* gene was considered to be a major candidate gene for the *adw* locus in the chicken. Nucleotide sequence alignment of a part of the isolated genomic chicken *HMGI-C* clone and human *HMGI-C* exon 5 showed eight mismatches none of which resulted in amino acid substitutions (Figure 3). *HMGI-C* exon 5 codes for one of the four highly conserved domains which are present in all members of the High Mobility Group Protein family (Manfioletti *et al.* 1991; 1995; Chau *et al.* 1995). Comparing the amino acid sequence of this domain with the same region present in the human *HMGY* gene, another member of the highly conserved *HMGI* Protein family, shows that we indeed cloned the chicken counterpart of *HMGI-C* (Figure 3). Manfioletti *et al.* (1995) report the isolation of intronless *HMGI-C* pseudogenes sharing a sequence homology of 92% with *HMGI-C* cDNA. Also other studies report the presence of pseudogenes in other members of the group of high mobility proteins (Johnson *et al.* 1989, Srikantha *et al.* 1987). In our study however, comparison of sequences from the chicken *HMGI-C* cDNA with genomic clone 1, shows the presence of an intron (Figure 3), showing that it is not a processed pseudogene.

We used the genomic chicken *HMGI-C* clone, and a clone containing the sequence of the marker (*LEI146*) which was found to be closely linked to the *adw* locus, as cohybridizing probes in a fluorescence *in situ* hybridization experiment of chicken

metaphase chromosomes. Indeed, *HMGI-C* was found to hybridize to the same site as *LEI146*, though the limited resolution of the metaphases did not allow a centromere-telomere ordering of the probes (Figure 4).

In conclusion, in this study we applied the positional candidate gene approach in order to identify a likely candidate for the *adw* phenotype in the chicken. Functional information about a gene or phenotype is combined with its chromosomal localization and the existence of conserved syntenic groups in other species. This way the increasingly dense human and mouse transcript maps can be exploited (Collins, 1995). The use of this approach also emphasizes the need for the localization of more expressed sequences on the chicken genetic linkage map in order to further explore the extent of synteny conservation between the chicken and other species.

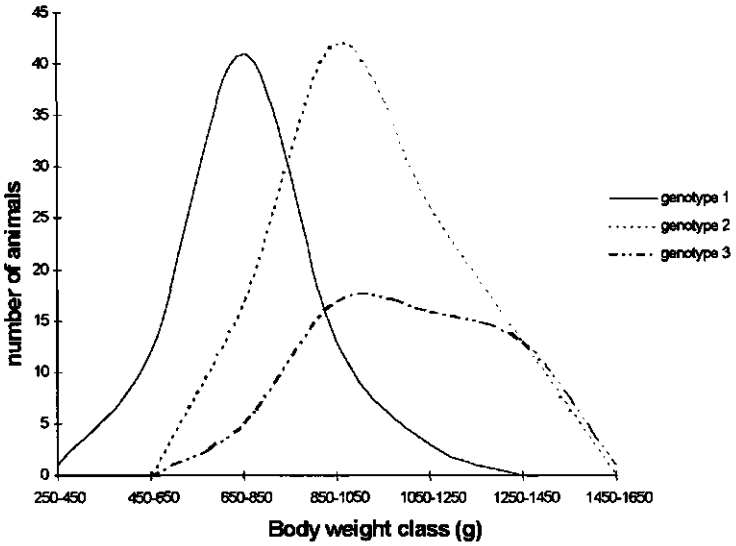


Figure 6. Body weight distribution of animals derived from the families segregating for the *adw* locus, according to their genotype for microsatellite marker *LEI146*. In general, *adw/adw* homozygous dwarf animals are represented by marker genotype 1, *ADW/adw* heterozygous non-dwarf animals by genotype 2, and *ADW/ADW* homozygous non-dwarf animals by genotype 3. However, it must be taken into account that 3 percent of all animals are recombinants.

Acknowledgments.

We thank Dr. W. J. T. Zabel, Prof.M.Georges and Prof.E.W.Brascamp for critically

reading the manuscript, Euribrid b.v. for supplying the blood samples of the families segregating for the autosomal dwarf locus. We are also grateful to Dr.G.Manfioletti for supplying the murine *HMGI-C* cDNA clone. The development of a diagnostic test for the autosomal dwarf mutation is the subject of a patent application supported by Euribrid b.v..

References.

Benson, K..F & Chada, K. 1994. Mini-mouse: phenotypic characterization of a transgenic insertional mutant allelic to pygmy. *Genet Research* **64**, pp 27-33.

Bumstead, N., Young, J.R., Tregaskes, C., Palyga, J.& Dunn, P.P.J. 1994. Linkage mapping and partial sequencing of ten cDNA loci in the chicken. *Animal Genetics* **25**, pp. 337-341.

Caspersson, T., Zech, L. & Johansson. C. 1970. Differential binding of alkylating fluorochromes in human chromosomes. *Experimental Cell Research* **60**, pp. 315-319.

Chau, K.Y., Patel, U.A., Lee, K.D., Lam, H.P.& Crane-Robinson, C. 1995. The gene for the human architectural transcription factor HMGI-C consists of five exons each coding for a distinct functional element. *Nucleic Acids Research* **23, 21**, pp. 4262-4266.

Cole, R.K. 1973. An autosomal dwarfism in the fowl. *Poultry Science* **52**, pp. 2012-2013 (Abstract).

Collins, S. 1995. Positional cloning moves from perditional to traditional. *Nature Genetics* **9**, pp. 347-350.

Crooijmans, R.P.M.A., van Kampen, A.J.A., van der Poel, J.J.& Groenen, M.A.M. 1993. Highly polymorphic microsatellite markers in poultry. *Animal Genetics* **34**, pp. 441-443.

Crooijmans, R.P.M.A., van Kampen, A.J.A., van der Poel, J.J. & Groenen, M.A.M. 1994. New microsatellite markers on the linkage map of the chicken genome. *J. of Heredity* **85**, pp. 410-413.

Crooijmans, R..P.M.A., van Kampen, A.J.A., van der Poel, J.J. & Groenen, M.A.M. 1995. Functional genes mapped on the chicken genome. *Animal Genetics* **26**, pp.

73-78.

Crooijmans, R.P.M.A., van Oers, P.A.M., Strijk, J.A., van der Poel, J.J. & Groenen, M.A.M. 1996. Preliminary map of the chicken (Gallus Domesticus) genome based on microsatellite markers : new markers mapped. *Poultry Science* **75**, pp. 746-754.

Giovannoni, J.J., Wing, R.A., Ganal, M.W. & Tanksley, S.D. 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Research* **19, 23**, pp. 6553-6558.

Huybrechts, L.M., King, D.B., Lauterio, T.J., Marsh, J. & Scanes, C.G. 1984. Plasma concentrations of somatomedine-C in hypophysectomized, dwarf and intact growing domestic fowl as determined by heterologous radioimmunoassay. *J. of Endocrinology* **104**, pp. 233-239.

Hutchison, N.J. & Le Cie,l C. 1991. Gene mapping in chickens via fluorescent *in situ* hybridisation to mitotic and meiotic chromosomes. *J. of Cell Biochem Suppl* **15E**, pp. 205.

Johnson, K.R., Lehn, D.A. & Reeves, R. 1989. Alternative processing of mRNAs encoding mammalian chromosomal high-mobility- group proteins *HMGI* and *HMGY*. *Mol Cell Biol* **9**, pp. 2114-2123.

Klein, S., Morrice, D.R., Sang, H., Crittenden, C.B. & Burt, D.W. 1996. Genetical and physical mapping of the chicken *IGF1* gene to chromosome 1 and conservation of synteny with other vertebrate genomes. *J. of Heredity* **87**, pp. 10-14.

Lathrop, G.M., Lalouel, J.M. & Ott, J. 1984. Strategies for multilocus linkage analysis in humans. *Proc. Nat. Acad. Sci. USA* **81**, pp. 3443-3446.

MacArthur, J.W. 1944. Genetics of body size and related characters. *American Naturalist* **78**, pp. 142-157.

Manfioletti, G., Giancotti, V., Bandiera, A., Buratti, E., Sautiere, P., Cary, P., Crane-Robinson, C., Coles, & Goodwin, G. 1991. cDNA cloning of the *HMGI-C* phosphoprotein, a nuclear protein with neoplastic and undifferentiated phenotypes. *Nucleic Acids Research* **19, 24**, pp. 6793-6797.

Manfioletti, G., Rustighi, A., Mantovani, F., Goodwin, G.H. & Giancott,i V., 1995. Isolation and characterization of the gene coding for murine high-mobility-group protein *HMGI-C*. *Gene* **167**, pp. 249-253.

Michelmore, R.W., Paran, I. & Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregation analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Nat. Acad. Sci. USA* **88**, pp. 9828-9832.

Patel, U.A., Bandiera, A., Manfioletti, G., Giancotti, V., Chau, K.Y. & Crane-Robinson, C. 1994. Expression and cDNA cloning of human HMGI-C phosphoprotein. *Biochem and Biophys Res Commun* **201**, pp. 163-70.

Ruyter-Spira, C.P., Crooijmans, R.P.M.A., Dijkhof, R., van Oers, P.A.M., Strijk, J.A., van der Poel, J.J. & Groenen, M.A.M. 1996. Development and mapping of polymorphic microsatellite markers derived from a chicken brain cDNA library. *Animal Genetics* **27**, pp. 229-234.

Ruyter-Spira, C.P., Gu, Z.L., van der Poel, J.J. & Groenen, M.A.M. 1997. Bulked segregant analysis using microsatellites: mapping of the dominant white locus in chicken. *Poultry Science* **76**, pp. 386-391.

Scanes, C.G., Marsh, J., Decuypere, E., & Rudas, P. 1983. Abnormalities in the plasma concentrations of thyroxine, tri-iodothyronine and growth hormone in sex linked dwarf and autosomal dwarf White Leghorn domestic fowl (Gallus domesticus). *J. of Endocrinology* **97**, pp. 127-135.

Srikantha, T., Landsman, D. & Bustin, M. 1987. Retropseudogenes for human chromosomal protein *HMG-17. J of Mol Biol* **197**, pp. 405-413.

Xiang, X., Benson, K.F. & Chada, K. 1990. Mini-Mouse: Disruption of the pygmy Locus in a Transgenic Insertional Mutant. *Science* **247**, pp. 967-969.

Zhou, X., Benson, K.F., Ashar, H.R. & Chada, K. 1995. Mutation responsible for the mouse phenotype in the developmentally regulated factor *HMGI-C. Nature* **376**, pp. 771-774.

Chapter 6

Chromosomal assignment of the gene encoding *c-KIT* in the chicken.

Chromosomal assignment of the gene encoding *c-KIT* in the chicken.

Plumage colour is of economical significance in the chicken industry, simply because consumers prefer to eat meat surrounded by a white skin. This has been accomplished by incorporating the dominant white mutation in most current commercial meat stocks. Although dominant white has been given the symbol *I* for inhibitor of colour (black pigment), *I* is incompletely dominant to its recessive allele *i*, which is associated with the absence of the *I*-effect on pigment. However, the phenotypic effects of *I* depend on the genetic background. For example, on a self-black (*E*) background chickens are black-flecked white, while the amount of black spotting is reduced when sex-linked barring (*B*) is present. In contrast, *I* is a relatively ineffective inhibitor of red pigment, pheomelanin.

On the classical linkage map published by Hutt (1964), *I* is located on linkage group II. In chapter four of this thesis, the genetic localization of the dominant white locus, close to marker *MCW188* on linkage group EL22 in the East Lansing reference family is described. From the classical linkage map it is clear that EL22 extends beyond the *I* locus (Figure 1).

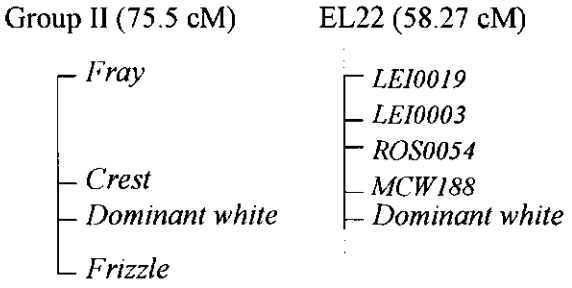


Figure 1. Location of the dominant white locus on the classical (left) and the East Lansing (right) genetic linkage maps. Genetic distances are given in centimorgan (cM)

In the Wageningen resource population *MCW188* was found to be unlinked. Thus far, no chromosomal assignments of classical linkage group II, or East Lansing linkage group 22 have been made. As no genes localised on EL22 have been reported as yet,

no comparative mapping data are available. Therefore, the positional candidate gene approach can not be applied, and candidate genes for *I* can only be inferred based on knowledge of the function of known genes and/or phenotypic similarities in other species for which the causative gene has been identified.

A genetic disorder of pigmentation in the human, called piebaldism, is caused by inactivation of the proto-oncogene *c-KIT* (also known as steel factor receptor or receptor tyrosine kinase)(Fleischman *et al.* 1991). In the mouse, *c-KIT* mutations have been identified at the dominant white spotting (*W*) locus, leading to phenotypes with a varying degree of severity characterised by deficiencies in gametopoiesis, melanogenisis and hematopoiesis (Geissler *et al.* 1988, Nocka *et al.* 1989, Tan *et al.* 1990). One of the mutants, called W^{sh}, only affects mast cells and melanogenisis because of an expression mutation leading to ectopic expression of *c-KIT* (Besmer *et al.* 1993, Duttlinger *et al.* 1993). Johansso-Moller *et al.* (1996) who studied the dominant white phenotype in the pig, which phenotypically resembles the *I* phenotype in the chicken, found a duplication and a splice mutation in the pig *c-KIT* gene (L. Andersson, pers. comm.) probably giving rise to a non-functional form of *c-KIT*. Mackenzie *et al.* (1997) have shown that the survival of immature melanoblasts and subsequent melanoblast proliferation depends crucially upon *c-KIT* signalling. This explains the absence of melanocytes in the unpigmented skin from *W* mutant mice (Silvers 1979) and *I* mutant pigs (Johansson-Moller *et al.* 1996). Although melanocytes containing pigment have been found in developing feathers from *I* chickens (Dorris 1938, Hamilton 1940), they do undergo premature cell death prior to the deposition of melanin, resulting in hypomelanotic feathers (Willier and Rawles, 1940, Bowers *et al.* 1992).

We looked into the possibility of *c-KIT* being a candidate gene for the *I* locus by determining the chicken chromosomal localisation of both *c-KIT* and the marker to which the *I* locus was found to be linked using fluorescence *in situ* hybridization. The results demonstrated that the chicken *c-KIT* mapped to chromosome 4 (Figure 2). Although from the present data it is difficult to extrapolate the localisation of *c-KIT* on the genetic linkage map, it is likely that *c-KIT* maps to a chromosomal segment which is conserved between chicken, human and mouse (Figure 3). Another part of this conserved chromosomal region in human and mouse, is also found on chicken

chromosome 6 indicating a chromosomal rearrangement during avian evolution.

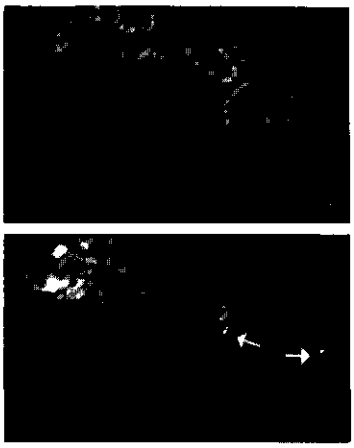


Figure 2. Fluorescence *in situ* hybridization (FISH) of chicken metaphase chromosomes using a chicken *c-KIT* probe. A 5.1 kb chicken cDNA clone containing the entire coding region of the chicken *c-KIT* cDNA (Sasaki *et al.* 1993), kindly provided by Dr. M. Sakurai from the National Institute of Animal Health in Tsukuba, Japan, was used as a probe to screen a chicken genomic lambda library. This resulted in the isolation of a 15 kb chicken lambda clone containing *c-KIT* sequences which was subsequently used as a probe in the FISH experiment using methods described in chapter seven. Arrows indicate the localization of the *c-KIT* probe on chromosome 4.

Currently, the chromosomal position of the marker sequence remains to be determined. However, from the present results it can be concluded that *c-KIT* is not a likely candidate for the *I* phenotype in the chicken, because mapping data indicated that the *I* locus is present on linkage group EL22 instead of chromosome 4. However, we can not completely rule out *c-KIT* as a candidate because the genetic linkage map of chromosome 4 might not fully represent the entire length of this chromosome, leaving the possibility that EL22 maps to its telomere.

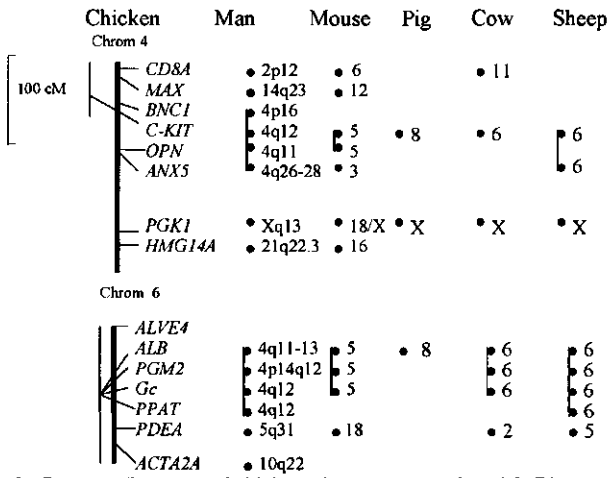


Figure 3. Comparative map of chicken chromosomes 4 and 6. Distances are given in centimorgan (cM). ├ Genes assigned by *in situ* hybridisation.

As mentioned, it is obvious from the classical linkage group II that a part of linkage group EL22 is still missing. Although very unlikely, this might be the reason that no linkage between EL22 and chromosome 4 has been detected sofar. Future *in situ* hybridisation of the marker sequences present on EL22 will help to resolve this puzzle.

References.

Besmer, P., Manova, K., Duttlinger, R., Huang, E.J., Packer, A., Gyssler, C. & Bachvarova, R.F. 1993. The KIT-ligand (steel factor) and its receptor *c-KIT/W*: pleiotropic roles in gametogenesis and melanogenesis. *Development Supplement,* pp.125-137.

Bowers, R.R., Harmon, J., Prescott, S., Asano, J. & Wynne, S. 1992. Fowl model for vitiligo: genetic regulation on the fate of the melanocytes. *Pigment Cell Research*, **Suppl. 2**, pp. 242-248.

Dorris, F. 1938. The production of pigment in vitro by chick nearal crest. *Arch. Entwick. Organ.* **138**, pp. 323-345.

Duttlinger, R., Manova, K., Chu, T.Y., Gyssler, C., Zelenetz, A.D., Bacharova, R.F. & Besmer, P. 1993. *W-sash* affects positive and negative elements controlling *c-KIT* expression: ectopic *c-KIT* expression at sites of *KIT*-ligand expression affects melanogenesis. *Development* **118**, 3, pp. 705-717.

Fleischman, R.A., Saltman, D.L., Stastny, V. & Zneimer, S. 1991. Deletion of the *c-KIT* proto-oncogene in the human developmental defect piebald trait. *Proc. Nat. Acad. Sci. USA* **88**, pp. 10885-10889.

Geissler, E.N., Ryan, M.A. & Housma, D.E. 1988. The dominant white spotting (*W*) locus of the mouse encodes the *c-KIT* proto-oncogene. *Cell* **55**, pp.185-192.

Hamilton, H.L. 1940. A study of the physiological properties of melanophores with special reference to their feather coloration. *Anat. Rec.* **78**, pp. 525-547.

Hutt, F.B. 1964. Animal Genetics. The Ronald Press, Co., New York.

Johansson-Moller, M., Chaudhary, R., Hellmen, E., Hoyheim, B., Chowdhary, B. & Andersson, L. 1996. Pigs with the dominant white coat color phenotype carry a duplication of the *KIT* gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome* **7**, 11, pp. 822-830.

Mackenzie, M.A., Jordan, S.A., Budd, P.S. & Jackson, I.J. 1997. Activation of the receptor tyrosine kinase *KIT* is required for the proliferation of melanoblasts in the mouse embryo. *Dev. Biol.* **192**, 1, pp. 99-107.

Nocka, K., Majumder, S., Chabot, B., Ray, P., Cervonne, M., Bernstein, A. & Besmer, P. 1989. Expression of *c-KIT* gene products in known cellular targets of *W* mutations in normal and *W* mutant mice - Evidence for impaired *c-KIT* kinase in mutant mice. *Genes Dev.* **3**, pp. 816-826.

Sasaki, E., Okamura, H., Chikamune, T., Kanai, Y, Watanabe, M., Naito, M & Sakurai, M. 1993. Cloning and expression of the chicken *c-KIT* proto-oncogene. *Gene* **128**, pp. 257-261.

Silvers, W.K. 1979. The coat colors of mice. New York: Springer Verlag.

Tan, J.C., Nocka, K., Ray, P., Traktman, P. & Besmer, P. 1990. The dominant W^{42} phenotype results from a missense mutation in the *c-KIT* receptor kinase. *Science* **247**, pp. 209-212.

Willier, B.H. & Rawles, M.E. 1940. The control of feather color pattern produced by grafting melanophores during embryonic development. *Am. Nat.* **75**, pp. 136-146.

Chapter 7

Nucleotide sequence of the chicken *HMGI-C* cDNA and expression of the *HMGI-C* and *IGF1* genes in autosomal dwarf chicken embryos.

C.P. Ruyter-Spira[1], J. Herbergs[1], E. Limpens[1], J.A. Marsh[2], J.J. van der Poel[1], T.A.Y. Ayoubi[3] and M.A.M. Groenen[1].

1 Department of Animal Breeding, Wageningen Institute of Animal Sciences (WIAS), Wageningen Agricultural University, Postbox 338, 6700 AH, Wageningen, The Netherlands.

[2] Department of microbiology and Immunology, C5-103 VMC, Cornell University, Ithaca, NYT 14853, USA

[3] Center for Human Genetics and Flaanders Inter University Institute for Biotechnology, University of Leuven, Herestraat 49, B-3000, Leuven, Belgium.

Accepted for publication in *Biochimica et Biophysica Acta*, Elsevier Science bv.

Nucleotide sequence of the chicken *HMGI-C* cDNA and expression of the *HMGI-C* and *IGF1* genes in autosomal dwarf chicken embryos.

C.P. Ruyter-Spira, J. Herbergs, E. Limpens, J.A. Marsh, J.J. van der Poel, T.A.Y. Ayoubi and M.A.M. Groenen.

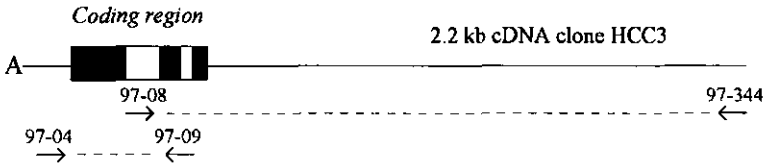
Abstract.

Mutations in the genes for High Mobility Group Protein I-C (*HMGI-C*) and Insulin-like Growth Factor-1 (*IGF1*) are known to be responsible for dwarf phenotypes in the mouse. Because the locus for autosomal dwarfism (*adw*) in the chicken maps to a region which is syntenic to a region in the human and mouse in which the *HMGI-C* and *IGF1* genes are located, *HMGI-C* and *IGF1* are likely candidate genes for *adw* in the chicken. In this study their possible role in the establishment of this phenotype has been investigated. We have cloned and sequenced the complete coding region of the chicken *HMGI-C* cDNA. Comparison with its human counterpart revealed a nucleotide sequence conservation of 84%. Only nine amino acid substitutions occurred, none of which were located in the four highly conserved domains. Northern blot analysis showed no difference in the expression of the *HMGI-C* gene between *adw* and wild type chicken embryos. Also no mutations in both the *HMGI-C* and the *IGF1* mRNA nucleotide sequence were detected *in adw* chicken embryos.

Introduction.

The identification of mutations in single genes, that result in aberrant developmental phenotypes, will lead to the unravelling of the complicated interactive processes involved in growth. For this purpose, dwarf strains of various animals have been proven to be useful models. In the chicken, several forms of dwarfism have been described (May 1935, Godfrey 1953, Hutt 1959, Cole 1973), but the causative gene for recessive autosomal dwarfism (*adw*) has not been identified as yet. Recently, the autosomal dwarf locus has been mapped to chicken chromosome 1 in a region which is syntenic to a region of chromosome 10 in the mouse and chromosome 12 in the human (Ruyter-Spira *et al.* 1998) . This region also harbors the genes encoding insulin-like growth factor I (*IGF1*) and high mobility group protein I-C (*HMGI-C*), which are potential

candidates for *adw*. Zhou *et al.* (1995) showed that *HMGI-C* plays a major role in the regulation of cell proliferation during embryonic development. A dwarf phenotype in the mouse, called pygmy, arises from the inactivation of the *HMGI-C* gene during embryonic development. Pygmy mice show a striking resemblance to the *IGF1* null mouse mutants created by Jeh-Ping *et al.* (1993). In both types of dwarfism, craniofacial appearance is affected in a similar way, which is also seen in *adw* chickens.



B

```
-66
CGTGTGCTTCCCCCCGTCGCGAGGTGCCGCGCAGAGCCGGGCGGAGGGCTGAGCTCTCGGCTCGCC

+1
ATG AGC GCC CAA GGC GAG GGA CCC GGC CAG TCT TCC ACC GCC GCC CCG GAG 51
Met Ser Ala Gln Gly Glu Gly Pro Gly Gln Ser Ser Thr Ala Ala Pro Glu 17

CAA CCT GCC GCC GCC GAG CCG CAG AAG CGA GGA CGA GGC AGA CCC AGG AAG 102
Gln Pro Ala Ala Ala Glu Pro Gln Lys Arg Gly Arg Gly Arg Pro Arg Lys 34

CAG CCC CAA GAA CCA ACT GGT GAA CCA TCT CCT AAA AGA CCA AGA GGA AGA 153
Gln Pro Gln Glu Pro Thr Gly Glu Pro Ser Pro Lys Arg Pro Arg Gly Arg 51

CCC AAG GGA AGC AAA AAC AAG AGT CCC TCT AAA GCA GCT CAG AAG AAA GCA 204
Pro Lys Gly Ser Lys Asn Lys Ser Pro Ser Lys Ala Ala Gln Lys Lys Ala 68

GAA GCC ACT GGT GAA AAG CGA CCC CGA GGG CGG CCC AGA AAA TGG CCT CAA 255
Glu Ala Thr Gly Glu Lys Arg Pro Arg Gly Arg Pro Arg Lys Trp Pro Gln 85

CAA GTG GTT CAA AAG AAG CCT GCT CAG GAA GAG ACT GAA GAA ACA TCG TCA 306
Gln Val Val Gln Lys Lys Pro Ala Gln Glu Glu Thr Glu Glu Thr Ser Ser 102

                          330
CAA GAA TCT GCA GAG GAA GAC TAG GAGACCGCACCATGCAATTTCTACCTCATCAGCAG 365
Gln Glu Ser Ala Glu Glu Asp *** 110

TTGGGTCTTTTGAAGGGAAGAAGACACTGCCTTGACCACTTATTTTCTA 414
```

Figure 1. (A) Chicken *HMGI-C* clone HCC3. Primers used for RT-PCR are indicated. First strand synthesis was performed using 97-344 5'GATTAGTTGCAGTCCTAACTG 3' or 97-09 5' CGCTTTTCACCAGTGGCTTC 3'. Subsequent PCR amplification of the first-strand product utilized the same 3' specific primers in combination with primers 97-08 5'GAACCAACTGGTGAACCATC 3' or 97-04 5'CGCGAGGTGCCGCGCAGAGC 3'.
(B) Nucleotide and deduced aminoacid sequence of the chicken *HMGI-C* cDNA coding region together with parts of its 3' and 5' UTR. The coding region starts at position +1. The stopcodon for translation is designated by *. The sequence data have been deposited in the Genbank/EMBL under accession number AF058287.

As endocrinological studies provided evidance showing that *IGF1* levels are unaffected in *adw* chicken (Huybrechts *et al.* 1984), the *HMGI-C* gene is considered the most likely candidate gene. However, based on the map location and the fact that only immunoreactive rather than of biologically active levels of *IGF1* have been measured, involvement of *IGF1* in *adw* can not be excluded. In the present study, the possible role of *HMGI-C* and *IGF1* in the establishment of autosomal dwarfism in chickens was investigated in more detail. *HMGI-C* expression patterns in *adw* and wild-type chicken embryos were compared, and *IGF1* and *HMGI-C* cDNA sequences from *adw* chickens were scanned for mutations.

```
Human    M S A R G E G A G Q P S T S A Q G Q P A A P A P Q K R G
Chicken  M S A Q G E G P G Q S S T A A P E Q P A A A E P Q K R G
Mouse    M S A R G E G A G Q P S T S A Q G Q P A A P V P Q K R G

         Domain I                          Domain II
Human    R G R P R K Q Q Q E P T G E P S P K R P R G R P K G S K
Chicken  R G R P R K Q P Q E P T G E P S P K R P R G R P K G S K
Mouse    R G R P R K Q Q Q E P T C E P S P K R P R G R P K G S K

                                                Domain III
Human    N K S P S K A A Q K K A E A T G E K R P R G R P R K W P
Chicken  N K S P S K A A Q K K A E A T G E K R P R G R P R K W P
Mouse    N K S P S K A A Q K K A E T I G E K R P R G R P R K W P

                                                Domain IV
Human    Q Q V V Q K K P A Q E E T E E T S S Q E S A E E D -
Chicken  Q Q V V Q K K P A Q E E T E E T S S Q E S A E E D -
Mouse    Q Q V V Q K K P A Q - E T E E T S S Q E S A E E D -
```

Figure. 2. Alignment of the chicken *HMGI-C* protein with the human and murine sequences. Mismatches between the chicken and murine/human sequences are shown in bold. Domains I-III represent the 3 conserved basic DNA binding segments. Domain IV represents the c- terminal acidic tail.

Because no chicken *HMGI-C* cDNA sequences have been reported, a 2.2 kb *HMGI-C* cDNA clone was isolated from a Uni-ZAP^{TM}XR cDNA library from 5 day old wild type chicken embryos (Stratagene), using a cDNA fragment representing the entire coding region of the murine *HMGI-C* cDNA (Manfioletti *et al.* 1991) as probe. Sequencing was performed using the autoread sequence kit (Pharmacia). Sequences were determined on an Automated Laser Fragment sequencer and further analysed using the Microgenie program of Beckman. Sequence analysis revealed that the chicken *HMGI-C* clone contained 66 bp of the 5' untranslated region (UTR), the entire coding region of 330 bp and approximately 1.8 kb of the 3'-UTR (Fig.1A). The nucleic acid sequence of the

coding region of the chicken *HMGI-C* (Fig.1B) was compared with its murine (Manfioletti *et al.* 1991) and human (Patel *et al.* 1994) homologues, and was found to be highly conserved (84% and 82%). Unlike human and mouse(Manfioletti *et al.* 1995), no homology between chicken and mouse, and chicken and human, was detected in the 5'-UTR. Instead, the sequenced parts of the 3'-UTR were highly conserved. Birds and mammals are evolutionary diverse and the extent and length to which sequence conservation is found in this part of the gene is uncommon (Ruyter-Spira *et al.* 1998), suggesting the 3' UTR to play a yet to define functional role. Translation of the coding region shows, in comparison to the human and mouse *HMGI-C* protein, nine and 13 changes of amino acids respectively (Figure 2). None of these occurred in the highly conserved basic DNA binding domains I-III, the so called A-T hooks (Reeves & Nissen, 1990) or the C-terminal acidic domain IV (Manfioletti *et al.*1991, Patel *et al.* 1994).

Subsequently, *HMGI-C* expression patterns were studied during embryonic development in both wild-type and *adw* chickens from the Cornell K strain of the White Leghorn. Total RNA was isolated from 3 or 4, 6 and 8 day old chicken embryo's, using the guanidinium thiocyanate procedure (Chromczynski & Sacchi, 1987). RNA (15 µg) was glyoxylated, size fractionated through a 1% agarose gel and transferred to Hybond-N using standard procedures (Ausubel *et al.* 1995). Hybridization of Northern blots was performed according to the procedure of Church & Gilbert (1984) using the chicken *HMGI-C* cDNA clone as a probe. Northern blot analysis using total RNA prepared from wild-type and *adw* chicken embryos isolated at various days, showed two clear transcripts of approximately 3.7 and 4.3 kb at comparable levels (Figure 3). Manfioletti *et al.* (1995) and Chau *et al.* (1995) also report differential *HMGI-C* transcripts in the mouse and human, although they observed a difference in abundancy between the products. Temporal expression patterns of *HMGI-C* in 3 to 8 day old chicken embryos showed a decrease with age, which is in agreement with the expression pattern observed in the mouse (Zhou *et al.* 1995). In pygmy mice no *HMGI-C* expression was detected at all, whereas in the present study no difference in *HMGI-C* expression was observed between wild type and *adw* chicken embryos of the same age. This indicates that autosomal dwarfism is not a consequence of altered levels or

major mutations in the *HMGI-C* mRNA.

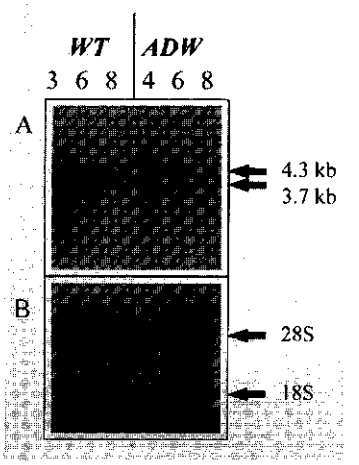


Figure 3. Northern blot analysis of *wildtype (wt)* and *adw* chicken *HMGI-C* mRNA expression in 3 resp. 4, 6 and 8 day old embryos. (A) 3.7 and 4.3 kb *HMGI-C* RNA fragments. (B) Ethidium bromide stained agarose gel (used in A) showing the relative loading of RNA samples.

Finally, the *HMGI-C* mRNA sequence was determined in homozygous dwarf animals in order to scan for point mutations possibly leading to a loss of biological function of the proteins. RT-PCR (Ready To Go kit, Pharmacia Biotech) was used to obtain *HMGI-C* cDNA fragments, which were subsequently ligated in plasmid *PUC18* using the Sureclone Ligation kit (Pharmacia Biotech) and sequenced using the autoread sequence kit (Pharmacia Biotech). The location and sequence of the PCR primers, are depicted in figure 1A. No mutations in the coding part of the *HMGI-C* cDNA were found. Therefore, the possible role of the second candidate gene, *IGF1*, was investigated. Because previous studies have already shown that *IGF1* plasma concentrations in growing *adw* chickens are unaffected (Huybrechts *et al.* 1984), we only scanned the *IGF1* mRNA of *adw* chicken embryos for mutations. The coding region of the *IGF1* mRNA sequence was amplified performing RT-PCR using the following primers: 5' TCGTTTCATGAGACGTCTACC 3' and 5'TGACATTGCTCTCAACATCTC 3'. Also for *IGF1*, sequence analysis of the resulting 615 bp fragment revealed no mutations.

In conclusion, in this study we did not detect any changes in the expression pattern of the *HMGI-C* gene, nor any mutations in *IGF1* or *HMGI-C* mRNA of *adw* chickens. Also, *HMGI-C* protein levels in *adw* chicken embryos were normal (data not shown). Therefore, both the *HMGI-C* and the *IGF1* genes are unlikely to represent candidate genes for autosomal dwarfism in the chicken. Another potential candidate gene is collagen 2A1 (*COL2A1*) which is a major protein component of the extracellular matrix of cartilage. Various mutations in the *COL2A1* gene are associated with a broad

spectrum of disorders characterized by abnormal growth or development of cartilage (Tiller *et al.* 1995). *COL2A1* has been mapped to human chromosome 12q13.11-q13.12 (Takahashi *et al.* 1990) which might be part of the syntenic region between chicken and man in which the *adw* locus resides. However, also other genes present in the syntenic region on human chromosome 12q are known to be involved in growth regulation. Therefore, future development of a chicken transcript map of higher density, which will result in a high resolution human-chicken comparative map, is needed. This will be of further assistance in determining the boundaries of the syntenic regions and will, consequently, narrow down the region in which candidate genes might reside.

Acknowledgments.

We are very grateful to Prof. G. Manfioletti for supplying the murine *HMGI-C* cDNA clone and to A. Nunen for performing the northern blot analysis. We also thank Dr. P. Zabel for critically reading the manuscript.

References.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. (Eds) 1995. Current Protocols in molecular biology, Wiley, New York.

Chau, K. Patel, U. A. Lee, K.D. Lam, H.P. & C. Crane-Robinson, 1995. The gene for the human architectural transcription factor *HMGI-C* consists of five exons each coding for a distinct functional element. *Nucleic Acids Research* **23, 21**pp. 4262-4266.

Chromczynski, P. & Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, pp. 156-159.

Curch, G.M. & Gilbert, W. 1984. Genomic sequencing. *Proc. Nat. Acad. Sci.* **81**, pp. 1991-1995.

Godfrey, E.F. 1953. THe genetic control of growth and adult body weight in the domestic fowl. *Poultry Science* **32**, pp. 248-259.

Hutt, F.B. 1959. Sex-linked dwarfism in the fowl. *J of Heredity* **50**, pp. 209-221.

Cole, R.K. 1973. An autosomal dwarfism in the fowl. *Poultry Science* **52**, pp.

2012-2013 (Abstract).

Huybrechts, L.M., King, D.B., Lauterio, T.J., Marsh, J. & Scanes, C.G. 1984. Plasma concentrations of somatomedine-C in hypophysectomized, dwarf and intact growing domestic fowl as determined by heterologous radioimmunoassay. *J. of Endocrinology* **104**, pp. 233-239.

Jeh-Ping, L., Baker, J., Perkins, A.S., Robertson, E.J. & Efstratiadis, A. 1993. Mice carrying null-mutations of the genes encoding insulin-like growth factor 1 (Igf-1) and type 1 IGF receptor (Igf1r). *Cell* **75**, 59-72.

Manfioletti, G., Giancotti, V., Bandiera, A., Buratti, E., Sautiere, P., Cary, P., Crane-Robinson, C., Coles, & Goodwin, G.H. 1991. cDNA cloning of the *HMGI-C* phosphoprotein, a nuclear protein with neoplastic and undifferentiated phenotypes. *Nucleic Acids Research* **19, 24,** *pp.* 6793-6797.

Manfioletti, G., Rustighi, A., Mantovani, F., Goodwin, G.H. & Giancotti, V. 1995. Isolation and characterization of the gene coding for murine high-mobility-group protein *HMGI-C. Gene* **167**, pp. 249-253.

Maw, A.J.G. 1935. The inheritance of skeletal dimensions in the domestic fowl. *Sci. Agric.* **16**, pp. 85-112.

Patel, U.A., Bandiera, A., Manfioletti, G., Giancotti, V., Chau, K. & Crane-Robinson. 1994. Expression and cDNA cloning of human *HMGI-C* phosphoprotein. *Biochem. Biophys. Res. Comm.* **201, 1;** 63-70.

Reeves, R. & Nissen, M.S. 1990. The AT-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J. of Biol. Chem.* **265** pp. 8573-8582.

Ruyter-Spira, C.P., de Groof, A.J.C., van der Poel, J.J., Herbergs, J., Masabanda, J., Fries, R. & Groenen, M.A.M. 1998. The HMGI-C Gene is a Likely Candidate for the autosomal dwarf Locus in the Chicken. *J.of Heredity* **89** pp. 295-300.

Takahashi, E., Hori, T. & Sutherland, G.R. 1990. Mapping of the huan type II collagen gene (COL2A1) proximal to fra(12) (q13.1) by nonisotopic *in situ* hybridization. *Cytogenet. Cell Genet.* **54**, pp. 84-85.

Tiller, G.E., Polumbo, P.A., Weis, M.A., Bogeart, R., Lachman, R.S., Cohn, D.H., Rimoin, D.L. & Eyre, D.R. 1995. Dominant mutations in the type II collagen gene,

COL2A1, produce sponyloepimetaphyseal dysplasia, Strudwick type. *Nat. Genet.* **11, 1** pp. 87-89.

Zhou, X., Benson, .K.F, Ashar, H.R. & Chada, K. 1995. Mutation responsible for the mouse phenotype in the developmentally regulated factor *HMGI-C*. *Nature* **376**, pp. 771-774.

Chapter 8

General discussion.

General discussion

Animal genomics is a rapidly expanding field. Techniques are becoming more and more advanced and the information present in various databases is accumulating rapidly. In this final chapter, I will evaluate the results that have been achieved in this study in the light of these new developments, and discuss them in retrospect with the original objectives, and finally add future directions.

1. The development and mapping of type I genetic markers as a contribution to the comparative map of the chicken.

In 1994, when the work of this thesis was started, about 60 genes had been mapped using the two international reference families. In view of the accumulating evidence showing that even between birds and mammals a large proportion of the genome had been conserved during evolution, the animal genome research community set out to increase the number of genes on the chicken genetic linkage map. Ultimately, a "gene-rich" comparative map would be of assistance in identifying regions of chromosomal homology between different species and would furthermore provide candidate genes for many animal QTLs of agricultural importance. In this respect, one of the objectives for this thesis was to develop microsatellite markers from cDNA sequences in order to be added as type I loci to the genetic linkage map of the chicken. The major advantage of this approach lies in the simultaneous mapping of highly polymorphic markers and genes, thus contributing to the development of both the genetic and the comparative maps. In total, 15 cDNA clones from a chicken brain cDNA library and nine cDNA clones derived from a chicken embryonic cDNA library have been mapped (Ruyter-Spira *et al.* 1996, Ruyter-Spira *et al.* 1998). Unfortunately, only six of these ESTs showed a relatively high sequence homology with sequences found in other species, of which only two had been mapped. Of course, with the expectation that the human genome sequencing project will be completed within 5 years from now, the number of chicken ESTs finally contributing to the comparative map will increase. However, to monitor this progress, a database search has to be performed on a regular basis. Another point of consideration is that, in this case, sequence similarity is the only

criterium for gene homology. Since many genes belong to large gene families that resulted from gene duplications during evolution, it will not always be possible to ascertain which one of the paralogues genes represents the homologous gene. Also, when genes between distantly related species like chicken and man are compared, a wide range in the extent to which these sequences are conserved is observed, and sometimes there might not even be a homologue existing.

Another approach to efficiently add expressed sequences to the chicken genetic linkage map was followed by Crooijmans *et al.* (1995) who screened sequences belonging to identified genes present in the Genbank and EMBL databases for microsatellite repeats. Similarly, Smith *et al.* (1996 and 1997) exploited this sequence information to develop PCR primers to amplify and sequence the parents of the East Lansing reference family in either an intron or the 3' region of the gene of interest. Because these noncoding regions are less conserved, a point mutation is often revealed for which allele specific mismatched primers can be developed to follow its segregation and map the gene. One drawback from this approach is that the PCR primers are tailored for the polymorphisms found in the East Lansing reference family, and are unlikely to be applicable to other families. The advantage of both methods is that they are PCR-based and that the identity of the gene has already been determined.

A third approach to increase the number of gene loci suitable for comparative genetic analysis in the linkage maps of vertebrate species has been proposed by O'Brien *et al.* (1993). The method involves the use of so called comparative anchor tagged sequences (CATS), which are based on well characterized coding genes spaced across the human genome. To develop CATS, primers are developed from exon sequences flanking introns, which are conserved between at least two eutherian species, and are subsequently used for mapping purposes in different species. This approach should further facilitate direct comparisons between divergent species' gene maps. However, Lyons *et al.* (1997) demonstrated that only 25% of the designed CATS result in an optimal amplification product in the chicken compared to 50% in eutherian mammals. Even if single PCR products are yielded, the polymorphic nature of these CATS still remains to be investigated and also the risk of amplification of paralogs

should not be overlooked.

Due to the joint efforts of various groups to add expressed sequences to the genetic linkage map of the chicken, the number of mapped genes has more than doubled since 1994 and has led to the recognition of at least 19 conserved chromosomal segments between human and chicken. A preliminary comparative map of the chicken, vis-à-vis the human and the mouse, is presented in the introduction of this thesis. Distances and gene orders on this map may be inaccurate because of the alignment of three different linkage maps together with genes assigned by *in situ* hybridization. Therefore, the present comparative map is not suitable for making valid estimations of the average length of the conserved fragments and the number of chromosomal rearrangements that have occurred since the split of different lineages. However, the present comparative map gives a good indication as to what extent avian and mammalian genomes have been conserved during evolution, and it appears that the degree of conservation observed between human and chicken is higher than the degree of conservation between chicken and mouse. This was also observed for distantly related eutherians, marsupials and monotremes when compared with the human and mouse. This indicates that the observed differences in conservation do not reflect a greater evolutionary distance but a more rapid rate of evolution in rodents instead (Andersson *et al.* 1996). Currently, a more extensive consensus chicken genetic linkage map is under development which will be used as the basis for a more accurate comparative map (Burt pers. comm.).

Specific attention should be paid to the contribution of the microchromosomes to the present comparative map. In the past, several genes have been assigned to the microchromosomes by *in situ* hybridization thus providing evidence that they do not represent genetically inert reserves of DNA as was suggested by Bloom and Bacon (1985). Dominguez-Steglich *et al.* (1993) even suggested that there may exist an imbalance in the chromosomal distribution of genes favouring microchromosomal localization. This hypothesis was strengthened by the observations of McQueen *et al.* (1996), who have shown that there is a disproportionally high number of CpG islands on the microchromsomes which is indicative of high gene content. Also, replication banding studies proved that chicken microchromosomses in general replicate their DNA earlier than macrochromsomes indicating the presence of coding sequences and

enhanced genetic activity (Schmid *et al.* 1989). Should microchromosomes actually contain an exceptionally high gene content as suggested, the probability of QTLs located on microchromsomes is also high. It is, therefore, unfortunate that particularly these microchromosomes appear to contain an extremely low microsatellite frequency (Primmer *et al.* 1997). Still, there is currently no indisputable statistical evidence based on large numbers of gene assignments, in favour of this suggested imbalance. Also, when looking at the comparative map of the chicken presented in this thesis, no bias towards an over representation of coding loci on microchromosomes was observed. When correcting for the genes that have been mapped by microsatellite $(CA)_n$ markers, that are known to be under represented on the microchromosomes (Primmer *et al.* 1997), 25% of the genes are located on the microchromosomes, which in turn represent approximately 25% of the total chicken genome (Bloom *et al.* 1993).

An improvement in the comparative map is to be expected from the application of microchromosome specific physical clones (BACS, YACS, cosmids)(Crooijmans pers. comm.) in two-colour fluorescent *in situ* hybridization experiments (Vignal, pers. comm.). This will lead to the discrimination of individual microchromosomes and the integration of both the cytological and genetic linkage maps of the chicken (Morisson *et al.* 1996). Another development involves the exploitation of chromosome specific chicken libraries (Ponce de Leon *et al.* 1998), or micro dissected parts of chicken chromosomes (Masabanda *et al.* 1998), as chromosome paints to identify homologous human chromosomal regions on human chromosome spreads. Finally, radiation hybrids could also be of assistance in the development of ordered comparative maps as a polymorphism per se is not a requirement for mapping.

2. Candidate genes for the dominant white locus in the chicken.

In chapter four of this thesis the mapping of the dominant white locus to East Lansing linkage group 22 is described. Subsequent chromosomal assignment of candidate gene *c-KIT* to chicken chromosome four, which is being described in chapter six, makes it unlikely that this is the causative gene. Another candidate gene, which, like *c-KIT*, is involved in the maintenance of the pigment producing melanocytes, is a growth factor called steel cell factor (*SCF*) (Murphy *et al.* 1992). Because *SCF* is the

ligand of *c-KIT* it therefore functions in the same biochemical pathway and mutations have been shown to give rise to a similar phenotype. This makes *SCF* an attractive candidate for the *I* locus in the chicken.

More candidate genes may be deduced from structural and physiological studies investigating melanocytes. Before going into detail on this, I will first explain the basics of pigment physiology.

Most of the feather colours in fowl result from the presence of the black and red pigments, the so called melanins. The melanins are synthesised in melanocytes which develop from the neural crest during embryonal development and are finally localised in the dermis and epidermis. It must be noted that eumelanin and pheomelanin are produced by two different types of melanocytes. In the melanocytes, melanin biosynthesis from tyrosine occurs in the Golgi system. The matrix, which is the substructure to which the melanin is attached, is produced in the smooth endoplasmatic reticulum. After fusion of the matrix with Golgi associated vesicles, intercellular granules called melanosomes are formed and melanin is synthesised. The melanosomes are translocated to the tips of the melanocyte dendrites after which they are pinched off and phagocytosed by the adjacent keratinocytes, with the melanosomes finally being transferred to the developing feather.

In the developing feathers from White Leghorns (WL) carrying the *I* gene, pigment containing melanocytes were shown to be present (Dorris 1938, Hamilton 1940). However, prior to the deposition of melanin to the feather, they undergo premature cell death, resulting in hypomelanotic feathers (Willier and Rawles, 1940, Bowers *et al.* 1992). Ultrastructural examination of WL melanocytes showed irregularly shaped melanosomes (Brumbaugh 1971), which appeared to be acid phosphatase-containing autophagosomes, which after fusing into larger bodies led to early death of the pigment cell (Jimbow *et al.* 1974). Apparently, the melanocyte attempts to protect itself from these abnormal, and likely cytotoxic, organelles by packaging them for degradation in autophagosomes, ultimately leading to cell necrosis. Melanogenisis produces significant levels of oxygen radicals (Thody *et al.* 1991) and the accumulation as a result of overproduction (Lerner 1971) or as a result of a defect in the protection mechanism against these toxic products (Bowers *et al.* 1992), may be causing premature death of melanocytes. Bowers *et al.* (1994) measured in vivo concentrations

of the antioxidants gluthation (*GSH*), catalase and superoxide dismutase (*SOD*) in Barred Plymouth Rock (BPR), White Leghorn (WL) and wild type Jungle Fowl (JF) feathers. BPR chickens carry the barring (*B*) gene, which is another pigment mutant in which premature death of melanocytes occurs. WL chickens also carry this gene besides the *I* gene. They showed a significant reduction of GSH activity of 66% in both the WL and BPR chickens in comparison to JF chickens, while catalase activity was not affected. The WL SOD activity is reduced to 50% of the BPR SOD activity, whereas the BPR *SOD* activity is reduced to 50% of the JF *SOD* activity. They also showed that addition of these antioxidants can rescue in vitro cultured mutant melanocytes. Therefore, it is suggested that the higher level of oxygen radicals, leading to premature melanocyte death in the WL, is caused by a reduction in *GSH* and *SOD* concentration due to the *B* gene, and that a further decrease in the level of SOD is directed by the *I* gene. Also in vitiligo, a human cutaneous hypopigmentation disorder characterized by premature death of melanocytes, an imbalance in the anti-oxidant system in the melanocyte is observed. In this case, the imbalance arises from a lower catalase activity (Maresca *et al.* 1997). Further research is needed to resolve the cause of these lowered SOD or catalase activities, which could lead to identification of potential candidate genes for *I* in the chicken.

3. Candidate genes for the autosomal dwarf locus in the chicken.

In chapter five of this thesis, the genetic localisation of the autosomal dwarf (*adw*) locus to a syntenic region between chicken, mouse and man is described. Comparative mapping revealed two candidate genes, insulin-like growth factor 1 (*IGF1*) and high mobility group protein I-C (*HMGI-C*), in which mutations are known to cause dwarf phenotypes in the mouse resembling the phenotypic characteristics of *adw* in the chicken. Although immunoreactive *IGF1*-plasma concentrations in *adw* chickens were found to be normal, there were neither data on biological activity nor complementation studies of this hormone. In addition, based on our mapping data we could not exclude *IGF1* as a candidate gene. While the chromosomal localisation of *IGF1* was already known, we provided evidence showing that *HMGI-C* is located in the same conserved region of chicken chromosome 1. In chapter six, however, we demonstrate that the size

and level of *HMGI-C* mRNA from dwarf embryos is not different from the size and level of *HMGI-C* mRNA found in wild type chickens. Moreover, no mutations in the coding regions of *IGF1* and *HMGI-C* mRNA have been detected. Therefore, it is not likely that mutations in these genes are involved in the establishment of autosomal dwarfism in the chicken. However, because we measured *HMGI-C* mRNA levels isolated from whole embryos, we can not exclude the possibility of ectopic or differential *HMGI-C* expression in the various target tissues and organs as a consequence of potential mutations in the regulatory sequences in the 5' or 3' untranslated regions of the *HMGI-C* gene.

It might also be that other, still unknown, genes involved in the *HMGI-C* pathway are responsible for autosomal dwarfism. Preliminary results from the screenings of pygmy mice, that aimed at identifying target genes of *HMGI-C*, point towards the direction of elevated levels of *COL2A1* (Ayoubi, pers. comm.) Interestingly, *COL2A1* is also located in the conserved chromosomal cluster on human chromosome 12 and mouse chromosome 10, however, its position in the chicken remains to be determined. *COL2A1* is the major protein component of the extracellular matrix of cartilage, and expression is observed at the onset of embryonic development. Mutations in *COL2A1* have been shown to be responsible for a broad spectrum of disorders in human, characterised by abnormal growth or development of cartilage. This might explain the leg problems that occurred when the *adw* gene was introduced in a broiler strain of chickens (Leenstra, pers. comm.). Although, like in *adw* chickens, macrocephaly is one of the characteristics observed in human syndromes that are caused by mutations in the *COL2A1* gene, in most cases, affected individuals also suffer from ocular abnormalities and hearing loss which is not reported for *adw* chickens. Still, *COL2A1* should be considered as a candidate gene for this dwarf phenotype in the chicken.

Another gene that deserves special attention, is the gene encoding insulin like growth factor binding protein 6 (*IGFBP6*). Because its precise location on human chromosome 12 remains to be determined, it is not yet known if *IGFBP6* maps within the conserved region in which we are interested. Still, the action of *IGFBPs* within the *IGF1* and-2 pathway is very interesting, since *IGF*-binding proteins have been found to prolong the half-life of the *IGFs* and have been shown to either inhibit or stimulate the growth promoting effects of the *IGFs* on cell culture. Ubiquitous *IGFBP6* expression

is initiated during embryonal development (Shimasaki *et al.* 1991) and has a marked preferential binding affinity for *IGF2* (Bach *et al.* 1993). Because *IGF2* is known to possess growth promoting activity, and the disruption of the *IGF2* gene in the mouse has been shown to cause a 40% reduction in size of heterozygous progeny (DeChiara *et al.* 1990), *IGFBP6* can also be considered as a candidate gene for *adw* in the chicken.

Finally, new developments in the attempts for the identification of candidate genes for yet another locus in the mouse, affecting growth regulation, must be followed carefully. In this case it concerns the high growth (*hg*) locus which causes a 30-50% increase in weight of homozygous individuals. The *hg* locus has been shown to be closely linked to the *IGF1* (Medrano *et al.* 1992) and decorin (*dcn*) (Horvat & Medrano 1995) genes, on mouse chromosome 10, in the conserved region in which we are interested. In the chicken, Masabanda *et al.* (1998) demonstrated by *in situ* hybridisation that the *dcn* gene is also located in the conserved chromosomal region on chromosome 1 (p16-p21) of the chicken, close to *HMGI-C* (p16-p21), in the region of the *adw* locus. Therefore, the possibility should be considered that the growth regulating gene, responsible for the *hg* mutation in the mouse, might also be involved in the establishment of *adw* in the chicken. *IGF1* and *dcn* were not found to be allelic with *hg*, and subsequent results from Horvat and Medrano (1996) suggest that the *hg* locus maps to a deletion in mouse chromosome 10, from which a marker will be used as an entry point to physical cloning of the *hg*-containing segment in the mouse.

4. Other strategies leading to the identification of the genes underlying dominant white and autosomal dwarfism, and monogenic traits in general.

In this thesis, emphasis is on the application of comparative mapping in the positional candidate gene approach which can be used for the identification of genes that are involved in the establishment of monogenic traits. Although the chromosomal regions controlling dominant white and autosomal dwarfism were successfully mapped using bulked segregant analysis, we were not able to identify the genes underlying these traits using the positional candidate gene approach. As dominant white is located

on a small linkage group to which no other known genes have been mapped, no comparative mapping could be performed. The most likely candidate gene for the autosomal dwarf locus, which was selected using the positional candidate gene approach, was not involved in this dwarf syndrome.

Another strategy that might lead to the isolation of the causative genes is map based cloning. In this approach the marker most significantly linked to the trait of interest is used to initiate a chromosome walk towards the underlying gene using overlapping large insert clones. The major drawbacks from this approach is the large area of DNA to be traversed because of the low mapping resolution in the initial scan, and the complications caused by the high proportion of repetitive DNA present in species with large genomes. Recently, in many species technological and analytical developments have enabled us to selectively identify molecular markers that are very close to the target gene, resulting in little chromosome walking over a small region (Tanksley et al. 1995). In this chapter, these new strategies involved in map based cloning will be discussed.

Region specific development of a high resolution map

A first requirement to be met identifying markers closely linked to the loci affecting the trait of interest, is a region specific high density marker map. Chromosome flow sorting and chromosomal microdissection (Ponce de Leon 1998), resulting in chromosome specific libraries from which microsatellite markers can be developed, are useful tools to meet this requirement. However, the small size of the chicken microchromosomes complicates the application of these techniques for all loci.

Giovannoni *et al.* (1991) showed that bulked segregant analysis provides an effective approach for isolating molecular markers, that are specific to any discrete genomic interval. This approach, which is based on the detection of differences in marker allele frequencies between two pools of DNA from segregating individuals, homozygous for opposing alleles for a targeted chromosome interval, is ideally suited for filling in gaps in genetic maps, with the particular advantage that DNA samples from the same segregating resource population can be pooled in different combinations to target any interval in the genome.

The high resolution marker order along the chromosomes can subsequently be

determined using large insert clones, containing the marker sequences, as probes in fluorescence *in situ* hybridization (FISH) experiments. Laan *et al.* (1996) demonstrated that FISH on mechanically stretched chromosomes (Laan *et al.* 1995), which has a mapping resolution of 100 kb or more, is an appropriate method. An even higher resolution of 1-5 kb until 500 kb can be obtained by fiber-FISH (Heiskanen *et al.* 1994), though such a resolution is not needed when mapping markers with an interval of approximately 1 cM. Finally, radiation hybrids can be particularly useful for the physical ordering of large insert contigs.

Fine mapping of the trait

Once the chromosomal region of interest is saturated with markers, high resolution mapping of the trait can be undertaken. High resolution mapping requires a large number of animals (progeny) to actually find recombinants. A number of approaches, including introgression lines (Eshed and Zamir 1995), advanced inbred lines (Darvasi and Soller 1995), advanced backcrosses (Tanksley and Nelson 1996) and interval specific congenic strains (ISCS)(Darvasi 1997), have been developed to minimize the number of progeny required. These approaches are based on a procedure called genetic chromosome dissection (GCD), in which genetic markers are used to construct chromosomes recombinant for known specific regions. These methods should not only be applicable to monogenic traits but also to quantitative trait loci (QTL). Darvasi (1998) calculated that six generations are required in the establishment of an interval specific congenic strain to fine map one QTL. A reduction in the QTL mapping interval from 5 to 1 cM could be achieved with 380 individuals. This is much lower than the required 7200 individuals for a backcross, or 19200 individuals required in a F_2-population, to map to the same precision. Since the chicken has a relatively short generation interval, this method might be suited for fine-mapping monogenic traits or QTL in poultry genomics. However, for QTL it remains to be seen whether the phenotypic effects of single isolated QTL for production traits can easily be categorised. This will mainly depend upon the complexity of the trait, and the proportion of phenotypic variation explained by the separate isolated QTL. In addition, the phenotypic effects of some traits, like feed conversion or egg production, can not be measured immediately. This

will slow down the identification of recombinants that will be used for the next round of backcrossing in order to establish the ISCS.

Another method feasible for fine mapping of some traits is linkage disequilibrium mapping (LDM)(Hästbacka *et al.* 1992). Linkage disequilibrium occurs when haplotype combinations of alleles at different loci occur more frequently than would be expected from random association, and can be detected performing an association study in populations that are derived from one founder population in which the mutation originally occurred. Therefore, fine mapping using this procedure does not require the production of a large number of animals in a new experimental population. LDM could be used for the fine mapping of the dominant white mutation which is segregating in several populations.

Another method used for fine mapping, which is closely related to LDM, is the identity by descent (IBD) approach. Like LDM, IBD also exploits the fact that most inherited disorders are genetically homogeneous. Dunner *et al.* (1997) demonstrate that commonly shared chromosomal segments not only occur within a given breed (Charlier *et al.* 1996) but also between different breeds. Because some mutations trace back many generations, advantage is taken of many intervening recombination events, leading to the identification of very small IBD regions. For the *adw* phenotype, a few small isolated families descending from the Cornell K strain of the White Leghorn in which the mutation originally occurred in 1969, can be used for the fine mapping of this trait using the IBD approach. As for the dominant white mutation, if historical records can prove that the dominant white mutation present in the White Leghorn, White Laced Red Cornish, Buff Laced Polish and both the Old English and Modern Red Pyle Games, are obtained from a common ancestor, this approach is also highly suitable for the fine mapping of this trait.

Isolation and evaluation of candidate genes from the target locus

The subsequent challenge to be resolved in high resolution mapping, is the identification and subsequent verification of the mutation responsible for the observed genetic variation. Fine-mapping of a trait to a 1 cM interval means, that in case of the chicken, candidate genes reside within a physical distance of approximately 500 kb. In centromeric and telomeric regions in which meiotic recombination is suppressed,

physical distances are even larger. A contig map of overlapping large insert fragments covering the region in which the trait has been mapped, will have to be developed. Recently, a chicken BAC library with an average insert size of 130 kb has been prepared (Crooijmans, unpubl. res.) which can be used for this purpose. Although the density of genes will vary throughout the genome, a 500 kb region can easily encompass more than 50 genes. Therefore, cDNA selection systems will result in the identification of numerous cDNA clones. Subsequent determination of the cDNA sequences, and their comparison with those in sequence databases in order to identify any homology with genes of known function, might give some evidence whether the cDNA corresponds to the target gene. Further information might be gained from studying expression patterns or analysing mRNA sequences of the genes in question in animals with alternate trait phenotypes.

A recently developed high throughput hybridization technique, which can be used for the detection of DNA polymorphisms or magnitude of mRNA expression, is the high-density DNA-micro array system (Schena *et al.* 1995, Lashkari *et al.* 1997). Short stretches of cDNA sequences can be arrayed on glass slides and subsequently be hybridized to fluorescently labelled cDNA from a large number of animals. Scanning with a laser microscope will provide quantitative hybridization signals. Using this method, the expression pattern of many candidate genes can be studied in parallel.

For some traits, however, it is possible to make a preselection leading to a reduction in the number of candidate genes that need to be evaluated. For example, because chickens that are homozygous for the *adw* gene, can be distinguished from wild type chickens at birth (although very difficult), we can conclude that the causative gene must be expressed during embryonic development. Therefore, the large insert clones covering the *adw* locus can be used to screen embryonic cDNA libraries of different developmental stages, resulting in a preselection of potential candidate genes.

Development of a region specific comparative map

In cases where the trait of interest has been mapped to an area of the genome for which no synteny with other species has been identified as yet, one can also first aim at a region specific development of the comparative map. Because the vast majority of

the human genome is now contained in YAC contigs, chicken cDNA clones isolated from the region of interest, can be used to probe human YACS (Georges & Andersson 1996) directly leading to the human map position of the homologous gene. Alternatively, amplified DNA from micro dissected chromosomal material of the region to which the trait has been mapped, can be used as a heterologous probe in FISH on human chromosomes (zoo-FISH), thereby identifying the homologous human map position. Masabanda *et al.* (1998) have demonstrated that, even for two distantly related species like man and chicken, zoo-FISH can be performed.

Following these approaches, positional candidate genes for the dominant white locus may be identified. In case of the *adw* locus, a region specific comparative map of higher density may be obtained, identifying more precisely the boundaries of the conserved chromosomal segments. This would narrow down the region of interest which should be explored for candidate genes in the comparative species.

Concluding remarks

It must be acknowledged that no single method will work for all traits. Choosing the most appropriate strategy for an experiment mapping a particular trait, will depend on a number of factors. For instance, knowledge of the genetic history of the population is a critical requirement in linkage disequilibrium and IBD mapping, because the timing of the gametic disequilibrium event need to be ascertained. In addition to this, the effective population size, the mode of inheritance and penetrance of the trait, laboratory equipment, expertise and available resources are determining factors in the choice of the approach to be followed to isolate genes underlying traits.

The development of new techniques, like the micro-array system, will open the door to new strategies or will enable us to efficiently improve old strategies, like the candidate gene approach. Finally, within the next 5-10 years the entire human genome sequence will be completed and the positional candidate gene approach will gain importance.

References.

Andersson, L., *et al.* 1996. Comparative Genome organization of vertebrates.

Mammalian Genome **7**, pp. 717-734.

Bach, L.A., Hsieh, S., Sakano, K., Fujiwara, H., Perdue, J.F. & Rechler, M.M. 1993. Binding of mutants of human insulin-like growth factor II to insulin-like growth factor binding proteins 1-6. *J. of Biol. Chem.* **268**, 13, pp. 9246-9254.

Bloom, S.E. & Bacon, L.D. 1985. Linkage of the major histocompatibility (B) complex and the nucleolar organizer in the chicken. Assignment to a microchromosome. *J. of Heredity* **76** (3), pp. 146-154.

Bloom, S.E., Delany, M.E. & Muscarella, D.E. 1993. Constant and variable features of avian chromosomes. In: manipulation of the avian genome, eds. Etches, R.J. and Gibbons, A.M.V. pp.39-59. CRC press, Florida.

Bowers, R.R, Harmon, J., Prescott, S., Asano, J. & Wynne, S. 1992. Fowl model for vitiligo: genetic regulation on the fate of the melanocytes. *Pigment Cell Research Suppl* **2, pp.** 242-248.

Bowers, R.R., Lujan, J., Biboso, A., Kridel, S. & Varkey, C. 1994. Premature avian melanocyte death due to low antioxidant levels of protection: fowl model for vitiligo. *Pigment Cell Research* **7**, pp. 409-418.

Brumbaugh, J.A. 1971. The ultrastructural effects of the *I* and *S* loci upon black-red melanin differentiation in the fowl. *Devel. Biol.* **24**, pp. 392-412.

Charlier, C., Farnir, F., Berzi, P., Vanmanshoven, P., Brouwers, B., Vromans, H. & Georges, M. 1996. Identity-by-descent mapping of recessive traits in livestock: Application to map the bovine *syndactyly* locus to chromosome 15. *Genome Research* **6**, pp. 580-589.

Crooijmans, R.P.M.A., van Kampen, A.J.A., van der Poel, J.J. & Groenen, M.A.M. 1995. Functional genes mapped on the chicken genome. *Animal Genetics* **26**, pp. 73-78

Darvasi, A. & Soller, M. 1995. Advanced inbred lines, an experimental population for fine genetic mapping. *Genetics* **141**, pp. 1199-1207.

Darvasi, A. 1997. Interval-specific congenic strains (ISCS): an experimental design for mapping a QTL into a 1-centimorgan interval. *Mammalian Genome* **8**, pp. 163-167.

Darvasi, A. 1998. Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genetics* **18**, pp. 19-24.

DeChiara, T.M., Robertson, E.J. & Efstratiadis, A. 1990. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* **345**, 6270, pp. 78-80.

Dominguez-Steglich, M., Robbins, J. & Schmid, M. 1993. Mapping of the chicken N-CAM gene and a myosin heavy chain gene: avian microchromosomes are not genetically inert reserves of DNA. *J. of Experimental Zoology* **265**, pp. 295-300.

Dorris, F. 1938. The production of pigment in vitro by chick neural crest. *Arch. Entwick. Organ.* **138**, pp. 323-345.

Dunner, S., Charlier, C., Farnir, F., Brouwers, B., Canon, J. & Georges, M. 1997. Towards interbreed IBD fine mapping of the mh locus: double-muscling in the Asturiana de los Valles breed involves the same locus as in the Belgian Blue cattle breed. *Mammalian Genome* **8**(6), pp. 430-435.

Eshed, Y. & Zamir, D. 1995. An introgression line population of lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**, pp. 1147-1162.

Georges, M. & Andersson, L. 1996. Livestock genomics comes of age. *Genome Research* **6**, pp. 907-921.

Giovannoni, J.J., Wing, R.A., Ganal, M.W. & Tanksley, S.D. 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Research* **11**, 19(23), pp. 6553-6558

Hamilton, H.L. 1940. A study of the physiological properties of melanophores with special reference to their feather coloration. *Anat. Rec.* **78**, pp. 525-547.

Hästbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. & Lander, E. 1992. Linkage disequilibrium mapping in isolated founder populations, diastrophic dysplasia in Finland. *Nature Genetics* **2**, pp. 204-211.

Jimbow, K., Szabo, G. & Fitzpatrick, T.B. 1974. Ultrastructural investigation of autophagocytosis of melanosomes and programmed death of melanocytes in White Leghorn feathers: a study of morphological events leading to hypomelanosis. *Devel. Biol.* **36**, pp. 8-23.

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**, pp. 928-933.

Horvat, S. & Medrano, J.F. 1995. Interval mapping of high growth (*hg*), a major locus that increases weight gain in mice. *Genetics* **139**, pp. 1737-1748.

Horvat, S. & Medrano, J.F. 1996. The high growth (*hg*) locus maps to a deletion in mouse chromosome 10. *Genomics* **36**, 3, pp. 546-549.

Laan, M., Kallioniemi, O.P., Hellsten, E., Alitalo, K., Peltonen, L. & Palotie, A. 1995. Mechanically stretched chromosomes as a target for FISH mapping. *Genome Research* **5**, pp. 13-20.

Laan, M., Isosomppi, J., Klockars, T., Peltonen, L. & Palotie, A. 1996. Utilization of FISH in positional cloning: an example on 13q22. *Genome Research* **6**, pp. 1002-1012.

Lashkari, D.A., DeRisi, J.L., Mc Cusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O. & Davis, R.W. 1997. Yeast microarrays for genome wide parallel genetic and gene exression analysis. *Proc. Nat. Acad. Sci. USA* **94**, 24, pp. 13057-13062.

Lerner, A.B. 1971. On the etiology of vitiligo and grey hair. *Am. J. of Med.* **51**, pp. 141-147.

Lyons, L.A., Laughlin, T.F., Copeland, N.G., Jenkins, N.A., Womack, J.E. & O'Brien, S.J. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics* **15**, pp. 47-56.

Maresca, V., Roccella, M., Roccella, F., Camera, E., Del Porto, G., Passi, S., Grammatico, P. & Picardo, M. 1997. Increased sensitivity to peroxidarive agents as a possible pathogenic factor of melanocyte damage in vitiligo. *J. of Invest. Dermatol.* **109**, 3, pp. 310-313.

Masabanda, J., Sazanov, A.A. & Fries, R. 1998. Mapping of individual loci and zoo-FISH indicates extended synteny conservation of human chromosome 12 and chicken chromosome 1. Abstract in: Plant & Animal Genome IV Conference, San Diego, Januari 18-22.

McQueen, H.A.M., Fantes, J., Cross, S.H., Clark, V.H., Archibald, A.L. & Bird, A.P. 1996. CpG islands of chicken are concentrated on microchromosomes. *Nature Genetics* **12**, pp. 321-324.

Medrano, J.F., Pomp, D., Taylor, B.A. & Bradford, G.E. 1992. The *high growth* gene (*hg*) is located on chromosome 10 linked to *IGF1*. Advances in gene technology: Feeding the world in the 21st century, ed.: W.J. Whelan *et al.* The 1992 Miami Bio/ Technology Winter Symposium 1, pp. 12.

Morisson M., Fillon, V., Pouzadoux, A., Berge, R., Zoorob, R., Auffray, C. & Vignal, A. 1996. Towards the integration of the chicken genetic and cytogenetic maps. Abstract in: XXVth International Conference on Animal Genetics. 21-25 July. pp.107.

Murphy, M., Reid, K., Williams, D.E., Lyman, S.D. & Bartlett, P.F. 1992. Steel factor is required for maintenance, but not differentiation, of melanocyte precursors in the neural crest. *Dev. Biol.* **153**, 2, pp. 396-401.

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A. & Copeland, N.G. 1993. Anchored reference markers for comparative genome mapping in mammals. *Nature Genetics* **3**, pp. 103-112.

Ponce de Leon, F.A., Ambady, S. and Cheng, H.H. 1998. Development of chromosome-specific libraries for chicken chromosomes 1, 2, 3 and 4. Abstract in: XXVth International Conference on Animal Genetics. 21-25 July. pp.107.

Primmer, C.R., Raudsepp, T., Chowdhary, B.P., Moller, A.P. and Ellegren, H. 1997. Low frequency of microsatellites in the avian genome. *Genome Research* **7**, pp. 471-482.

Ruyter-Spira, C.P., Crooijmans, R.P.M.A., van Oers, P.A.M., Strijk, J.A., van der Poel, J.J. & Groenen, M.A.M. 1996. Development and mapping of polymorphic microsatellite markers derived from a chicken brain cDNA library. *Animal Genetics* **27**, pp. 229-234.

Ruyter-Spira, C.P., de Koning, D.J., van der Poel, J.J., Dijkhof, R.J.M. & Groenen, M.A.M. 1998. Developing microsatellite markers from cDNA; A tool for adding expressed sequence tags to the genetic linkage map of the chicken. *Animal Genetics* **29.** pp. 85-90.

Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 5235, pp. 467-470.

Schmid, M., Enderle, E., Schindler, D. & Schempp, W. 1989. Chromosome banding and

DNA replication patterns in bird karyotypes. *Cytogenet. Cell Genet.* **52**, pp. 139-146.

Shimasaki, S., Gao, L., Shimonaka, M & Ling, N. 1991. Isolation and molecular cloning of insulin-like growth factor-binding protein-6. *Mol. Endocrinol.* **5**, 7, pp. 938-948.

Smith, E.J., Cheng, H. & Vallejo, R. 1996. Mapping candidate chicken genes: An alternate approach. *Poultry Sciience* **75**, pp. 642-647

Smith, E.J., Lyons, L.A., Cheng, H.H. & Suchyta, S.P. 1997. Comparative mapping of the chicken genome using the East Lansing Reference population. *Poultry Science* **76**, pp. 743-747.

Tanksley, S.D., Ganal, M. & Martin, G.B. 1995. Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends in Genetics* **11**(2), pp. 63-68

Tanksley, S.D. & Nelson, J.C. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.* **92**, pp. 191-203.

Thody, A, J. Higgins, E., Wakamatsu, K., Ito, S., Burchill, S. & Marks, J. 1991. Pheomelanin as well as eumelanin is present in human epidermis. *J. of Invest. Dermatol.* **97**, pp. 340-344.

Willier, B.H. & Rawles, M.E. 1940. The control of feather color pattern produced by grafting melanophores during embryonic development. *Am. Nat.* **75,** pp. 136-146.

The development of genetic linkage maps in farm animals is progressing rapidly. Linkage maps can be used to identify genetic loci responsible for genetic variation in traits of economic importance. The ultimate goal is to find the underlying genes involved in these traits. To achieve this, the so called positional candidate gene approach is gaining in importance. This approach is based on the genetic localization of a trait using genetic linkage analysis in livestock species. Subsequent comparative mapping of the trait locus with the gene-rich maps of the human and mouse may reveal likely candidate genes for the trait in question. In the introduction of this thesis, a preliminary comparative map of the chicken is presented.

For the construction of comparative maps the genetic localisation of many genes needs to be determined. Therefore, the first aim of the study was to add an extra number of expressed sequences to the genetic linkage map of the chicken. In order to develop expressed sequence tags, microsatellite markers, containing a $(TG)_n$ repeat, have been developed from a brain and an embryonic cDNA library, and mapped on the international reference families and/or the Wageningen resource population. Initially, approximately 0.10%-0.15% of the cDNA clones of both libraries gave a positive hybridisation signal with a $(TG)_{13}$ probe. The brain cDNA library was selected because of its expected high RNA complexity due to the large variety of different cell types present in this tissue. This would increase the chance of finding a large variety of different cDNA's. Indeed, 60 of the 90 $(TG)_{13}$ positive cDNA clones that were sequenced turned out to be unique. Of these 60 clones, 22 expressed sequence tags were finally mapped, of which 5 showed sequence homology with known genes or expressed sequence tags (EST) in the mouse or human. Because of the high redundancy of $(TG)_{13}$ positive clones between the embryonic and brain cDNA library, only nine new markers could be developed from the embryonic cDNA library. The embryonic cDNA library was chosen because developmental patterns shared by individual species are highly conserved. Therefore, genes expressed during early development were expected to show higher rates of sequence conservation resulting in a better identification of homologous genes between the chicken and other species. This appeared to be true, since 35% of the unique $(TG)_{13}$ positive clones showed similarity to sequences from the Genbank/EMBL and HCD databases, instead of 15% for the brain cDNA library. Unfortunately, only one of the clones for which homology

was detected could be mapped. Still, the isolation and development of microsatellites from cDNA libraries has proven to be a rapid method in order to add genes to the genetic linkage map of the chicken. However, considering the wide range of sequence homologies between known human and chicken homologues, and the fact that homology between relatively short stretches of sequence does not necessarily mean that the sequences represent identical transcripts, it is apparent that any conclusion on possible syntenic groups must be drawn with caution.

The second objective of this study was to develop a quick and reliable method in order to localize monogenic traits. To achieve this, an already existing technique called bulked segregant analysis (BSA), which has originally been used for the localization of monogenic traits in plants using RAPD and RFLP markers, was combined with the use of fluorescently labelled microsatellite markers. BSA involves comparing two pooled DNA samples of individuals from a segregating population originating from a single cross. Within each pool the individuals are identical for the trait, locus or gene of interest, but are arbitrary for all other genes. Both pools are analysed to identify markers that show differences in allele frequency, which will indicate that these markers are linked to the loci determining the trait used to construct the pools. Since the analysis of pools can replace the analysis of large numbers of individual segregating progeny, considerable savings are possible.

In a pilot experiment and in a test case for which the dominant white (*I*) locus was selected because of its full penetrance and its phenotype which can be easily categorized, we found that the combination of BSA with fluorescently labelled microsatellites is very sensitive in detecting linked markers at greater genetic distances. In addition to this, in case only two alleles are involved, absolute allele frequencies can be estimated accurately. This way, the number of recombinants within a pool can be calculated, which makes it possible to directly estimate genetic distance between marker and trait.

The dominant white locus was mapped on the small linkage group EL22 of the East Lansing reference family at a distance of 2 cM from *MCW188*. As no genes localised on EL22 have been reported as yet, no comparative mapping data are available. Therefore, the positional candidate gene approach can not be applied, and candidate genes for *I* can only be inferred based on knowledge of the function of known genes

and/or phenotypic similarities in other species for which the causative gene has been identified. One such gene is *c-KIT*, which is known to be involved in the dominant white phenotype in the pig, and pigmentation disorders in human and mouse. Fluorescent *in situ* hybridization (FISH) of a 15 kb lambda clone on chicken metaphase chromosomes revealed that the chicken *c-KIT* gene is localized on macro chromosome 4 making it unlikely that this is the causative gene for *I* in the chicken.

The second trait we analysed using BSA was recessive autosomal dwarfism (*adw*) in the chicken. Because this monogenic trait is characterized by reduced growth, this model offers the opportunity to isolate and identify one of the components involved in a complex trait like growth. We were able to map the *adw* locus to chromosome 1 in the region of the insulin-like growth factor 1 (*IGF1*) locus. Although *IGF1* is known to play a role in the regulation of growth, endocrinological studies in the past have shown that plasma concentrations of immunoreactive *IGF1* were unaffected in *adw* chickens. Another gene located in this area is the lysozyme (*LYZ*) gene. In the human and mouse, the *IGF1* and *LYZ* genes are both located in a conserved chromosomal segment. Interestingly, in the mouse the phenotype "pygmy" which shows a striking similarity to the autosomal dwarf phenotype in chickens, is also located in this region. The pygmy phenotype arises from the inactivation of the High Mobility Group I-C gene (*HMGI-C*). *HMGI-C* has been shown to be involved in the regulation of cell proliferation and is being expressed in the embryonic stage in all organs except a small section of the forebrain. This coincides with the observation that most tissues in pygmy mice are 40-50% smaller than wild-type tissues, the only tissue of normal size being the brain. Considering the synteny between regions of chicken chromosome 1, mouse chromosome 10 and human chromosome 12, and taking into account both the phenotypic characteristics and the mode of inheritance of the chicken *adw* and the mouse pygmy loci, the *HMGI-C* gene is a major candidate gene for the *adw* locus in the chicken. Finally, fluorescence *in situ* hybridization of chicken metaphase chromosomes using the chicken *HMGI-C* gene as a probe, showed that the chicken *HMGI-C* gene is indeed located on chromosome 1 in the region of the *adw* locus. However, northern blot analysis showed no difference in the expression of the *HMGI-C* gene between *adw* and wild type chicken embryos. Also no mutations in both the *HMGI-C* and the *IGF1* in RNA nucleotide sequence were detected in *adw* chicken embryos.

Finally, in the discussion of this thesis, other candidate genes for both autosomal dwarfism and dominant white are proposed. In addition to this, alternative procedures that may lead to the isolation of the causative genes, are evaluated.

Samenvatting

Ontwikkelingen op het gebied van genetische merker kaarten bij landbouwhuisdieren gaan zeer snel. Deze kaarten kunnen gebruikt worden bij het identificeren en analyseren van bepaalde plaatsen in het genoom waarop genen liggen die verantwoordelijk zijn voor de genetische variatie van economisch belangrijke kenmerken. Om deze genen te kunnen isoleren is de zogenaamde positionele kandidaatgen klonerings methode een belangrijk hulpmiddel. Deze methode is gebaseerd op het lokaliseren van een kenmerk op het genoom, waarna via een vergelijking van de genenkaart met die van andere diersoorten kandidaatgenen uit deze regio kunnen worden aangewezen. Op deze manier kan gebruik worden gemaakt van de "genrijke" genoomkaarten van mens en muis. In de inleiding van dit proefschrift is een eerste vergelijkende genenkaart tussen kip, mens en muis gepresenteerd.

Om een goede vergelijkende genenkaart te ontwikkelen moet de genomische lokalisatie van vele genen bepaald worden. Eén van de doelstellingen van dit onderzoek was dan ook het ontwikkelen van genetische merkers uit cDNA sequenties afkomstig uit een embryonale en hersen cDNA bank. Deze cDNA banken zijn met een $(TG)_{13}$ probe gescreened, waarna de positieve kloons gebruikt werden om microsatelliet merkers te ontwikkelen. Deze werden vervolgens op de genetische koppelings kaarten van internationale referentie families, en/of de Wageningse kippen populatie waarmee associatie studies verricht wordt, met behulp van een koppelings studie geplaatst. Er was specifiek voor een cDNA bank van hersenweefsel gekozen, omdat het bekend is dat in dit weefsel zeer veel verschillende genen tot expressie komen, waardoor de kans op het vinden van een grote variëteit aan cDNAs vergroot wordt. Dit bleek ook het geval te zijn; 60 van de 90 $(TG)_{13}$ positieve cDNA kloons waren uniek. Uiteindelijk konden er hiervan 22 op de kaart geplaatst worden, waarvan 5 kloons homologie met sequenties uit de databank vertoonden. Omdat de embryonale ontwikkelingsstadia van verschillende diersoorten grote overeenkomsten vertonen, lag het in de verwachting dat de genen die hierbij betrokken zijn ook een hogere mate van conservering tussen (dier)soorten vertonen. Inderdaad werd voor 35% van de unieke kloons afkomstig uit de embryonale cDNA bank een hoge mate van sequentie homologie met genen uit de databanken gevonden, terwijl dit percentage voor de hersenbank maar 15% bedroeg. Helaas konden er uit de embryonale cDNA bank maar 9 merkers op de kaart geplaatst worden doordat de $(TG)_{13}$ positieve kloons uit de

hersen en embryonale cDNA banken een grote overlap vertoonden. Toch is het ontwikkelen van microsatelliet merkers uit cDNA kloons een snelle en efficiënte methode gebleken om genen op de kaart te plaatsen. Wanneer deze cDNA sequenties echter voor het ontwikkelen van de vergelijkende genenkaart gebruikt worden, dient ermee rekening te worden gehouden dat het criterium van homologie alleen gebaseerd is op sequentie informatie, waardoor conclusies over geconserveerde regio's tussen 2 genomen zeer voorzichtig getrokken dienen te worden.

Het tweede doel van dit promotieonderzoek was het uittesten van de zogenaamde 'bulked segregant' analyse methode (BSA), in combinatie met fluoriserende microsateliet merkers, om monogene kenmerken te lokaliseren. De BSA methode is gebaseerd op het vergelijken van merker allelfrequenties afkomstig van 2 verschillende mengmonsters, welke zijn samengesteld uit DNA van nakomelingen uit een familie waarin het te onderzoeken kenmerk segregeert. Binnen elk mengmonster is DNA opgenomen afkomstig van individuen die identiek zijn voor het kenmerk, terwijl voor de rest van het genoom de allelen volgens toeval verdeeld zijn. Daardoor zal een gevonden verschil in allelfrequentie van de merkers tussen beide monsters een indicatie zijn dat de merker zich in de buurt van het gen bevindt dat het kenmerk veroorzaakt. Doordat het analyseren van DNA mengmonsters het analyseren van grote aantalen individuele DNA monsters overbodig maakt, is de BSA methode zeer efficient.

In dit promotieonderzoek hebben we met behulp van de BSA methode de genomische lokalisatie van de monogene kenmerken dominant wit (*I*) en autosomale dwerggroei (*adw*) bepaald. We hebben aangetoond dat deze methode zeer gevoelig is, en dat merkers die zich op een afstand van 30cM van een kenmerk, of andere merker bevinden, goed gedetecteerd kunnen worden. In gevallen waarin maar 2 merker allelen betrokken waren, was het zelfs mogelijk om de genetische afstand zeer nauwkeurig te schatten.

We hebben bepaald dat het kenmerk dominant wit op de kleine koppelingsgroep EL22 van de East Lansing referentiefamilie gelokaliseerd is. Doordat op deze koppelingsgroep nog geen andere genen geïdentificeerd zijn, en een vergelijkende genenkaart voor deze regio dus niet voorhanden is, kan geen gebruik worden gemaakt van de positionele kandidaatgen klonerings methode. Kandidaatgenen kunnen alleen aangewezen worden aan de hand van fysiologische kennis van bekende genen die bij

het pigmentatie proces betrokken zijn en/of fenotypische overeenkomsten bij andere (dier)soorten waarvan de onderliggende genen al geïdentificeerd zijn. Een gen waarvan bekend is dat deze betrokken is bij het dominant wit fenotype bij het varken, en pigmentatie afwijkingen bij mens en muis, is *c-KIT*. Fluoriserende *in situ* hybridisatie (FISH) op metafase chromosomen met het *c-KIT* gen van de kip als probe, wees echter uit dat *c-KIT* op macrochromosoom 4 gelokaliseerd is, waardoor het onwaarschijnlijk wordt dat *c-KIT* het verantwoordelijke gen voor *I* bij de kip is.

In hoofdstuk 5 van dit proefschrift staat de genetische lokalisatie van autosomal e Dwerggroei beschreven. Doordat ook *adw* een monogeen kenmerk is, is dit een ideaal modelsysteem om één van de componenten die betrokken zijn bij een complex kenmerk als groei, te kunnen isoleren. Het blijkt dat het *adw* locus op chromosoom 1 gelegen is, in de buurt van het *IGF1* gen, waarvan bekend is dat deze een rol speelt bij de regulatie van groei. Endocrinologisch onderzoek heeft echter uitgewezen dat de *IGF1* plasmaconcentratie van *adw* dieren niet verschilt van die van wild type dieren. Een ander gen, in dezelfde regio, is het lysozym (*LYZ*) gen. Zowel bij de mens als bij de muis zijn het *IGF1* en het *LYZ* gen beide gelokaliseerd in een geconserveerde chromosomale regio. In dezelfde geconserveerde regio is bij de muis ook het "pygmy" fenotype gelokaliseerd. Karakteristieken van dit fenotype lijken zeer veel op die we aantreffen bij *adw* kippen. Het is bekend dat het pygmy fenotype veroorzaakt wordt door inactivatie van het *HMGI-C* gen. Dit gen is betrokken bij de regulatie van celproliferatie en komt embryonaal tot expressie in bijna alle weefsels behalve een klein gedeelte van de hersenen. Gezien deze feiten lijkt het *HMGI-C* gen een zeer aantrekkelijke kandidaat voor het *adw* fenotype bij de kip. *in situ* hybridisatie met zowel een kloon waarop de aan *adw* gekoppelde merker gelegen is, en een kippe *HMGI-C* kloon als probes, laat ook zien dat dit gen op chromosoom 1 in de *adw* regio gelegen is. In een vervolg experiment hebben we met behulp van een northern blot echter aangetoond dat het *HMGI-C* gen normaal tot expressie komt. Sequentie analyse van het coderende gedeelte van het *HMGI-C* en *IGF1* mRNA liet ook geen verschillen tussen *adw* en wild type embryo's zien.

Tot slot worden in de discussie van dit proefschrift enkele andere potentiële kandidaat genen voor *I* en *adw* besproken. De discussie eindigt met een overzicht van alternatieve strategieën en recente methodische ontwikkelingen die tot het isoleren van

de onderliggende genen zouden kunnen leiden.

Curriculum Vitae

Carolina Petronella Spira was born on June 11, 1964 in Alkmaar, the Netherlands. She obtained her BSc in Animal Science in 1986 at the Agricultural High School "De Drieslag" in Dronten, after which she started her MSc in Animal Nutrition and Genetechnology at the Agricultural University in Wageningen. In 1988, she spent 6 months in Edmonton, Canada, where she studied the physiological effects of a daily administration of bovine somatotropin on gastro-intestinal parameters in the cow. Subsequently she made a switch to the field of molecular biology, and spent 6 months at the company of Genepharming, Leiden, in order to study the expression of a bovine αs1-casein promoter construct in a murine mammary gland epithelial celline. In 1990 she participated in a research project at the department of Animal Genetics in Uppsala, Sweden, in order to develop a method to discriminate different alleles of the BOLA complex in the cow. She finished her MSc in 1991 after which she worked as a PhD research fellow at the Department of Animal Breeding and Genetics, initially studying the regulation of casein expression in the cow. In 1994 her first daughter Nienke was born. At this time she also started to work on a research project titled: genome analysis of the chicken, of which the results are described in this thesis. In 1996 her second daughter Aliene was born.

From 1 May 1998 she is employed at the Centre for Plant Breeding and Reproduction Research, Wageningen.

Appendix 1

Abbreviations belonging to the mapped loci on the comparative map of the chicken.

ABL1	Abelson viral oncogene homologue
ACTA2A	Smooth muscle actin α2
ACTB	ß actin
ACTIVIN	Activin 2B
AGC1	Aggrecan
AGRN	Agrin
AK1	Adenylate kinase 1
ALB	Serum albumin
ALDOB	Aldolase B, fructose biphosphate
ALVE4	Endogenous retrovirus 4
AMBP	α microglobulin
AMH	Anti Mullerian hormone
ANX5	Annexin 5
APOA1	Apolipoprotein A1
B2M	ß2 microglobulin
BCL2	B-cell lymphoma 2
BLBL1	MHC B complex Class II
BLBL2	MHC B complex Class II
BMP2	Bone morphogenic protein
BMP4	Bone morphogenic protein 4
BNC1	Brain neuron cytoplasm.prot.1
CA2	Carbonic anhydrase II
CALB1	Vitamin D induced Calbinding protein
CALPROTK	calcium/calmodulin dependent proteinase K
CD8A	Lymphocyte surface marker
CD28	CD28 antigen
CDC2L1	Cell division cycle 2 protein kinase
CDX1	Caudal type homeobox transcription factor 1
CHRNB3	Nicotinicacetylcholine receptor
CKB	Creatine kinase
c-KIT	Tyrosine kinase receptor
COL1A1	Collagen type 1α1 chain
COL3A1	Collagen type 3α1 chain
CPAA	Erythrocyte alloantigen A
CRYA	α A Crystallin
CRYB	ß B1 Crystallin
CYP19	Cytochrome P450
DNCL	Dynein heavy chain
EF1B	Eukaryotic translation elongation factor 1B

EGFR	Epidermal growth factor receptor
EN2AA	Engrailed protein en-2
ENO1	Enolase A
EPOC	1/SKN1/OCT1 Pou domain proteins
ERK	Extracellular signal related kinase
ESR	Estrogen receptor
FASN	Fatty accid synthase
FN1	Fibronectin 1
FYN	Fyn oncogene related protein
GAPD	Glyceraldehyde 3P dehydrogenase
GC	Vitamin D binding protein
GGTB2	UDP4 ß-galactosyl transferase
GH	Growth hormone
GHR	Growth hormone receptor
GH6R	Growth hormone 6 receptor
GSTA2	Glutathione-S-transferase
H2A,3,4	Duplicated genes from H2A,3,4
H5	Histon 5
H33B	Histone, replacement H3.3B variant
HBA	α globin gene cluster
HBB	B-Haemoglobin
HISA	Histone A gene cluster
HMGI-C	High mobility group protein I-C
HMX3	H6 related gene
HMG14A	High mobility group protein 14A
HSD3B	3 ß hydroxyst. dehydrogenase
HSF3A	Heat shock factor 3
HSPCAL3	Heat shock protein 90kD, αlike3
ID	Dermal melanin inhibitor
IGF1	Insulin like growth factor
*IGF1*R	Insulin like growth factor 1 receptor
IGVPS	Immunoglobulin pseudo V26 and V6
IREB1	Iron response element binding protein
L1B	Ribosomal protein 1B
L5	Ribosomal protein L5
L7A	Ribosomal protein L7A
L37A	Ribosomal protein L37A
LAMP1	Lysosomal associated membrane protein
LDBH	Lactate dehydrogenase B4
LGAL4	Lectin, ß-galactoside binding
LYZ	Lysozyme
MAX	Max gene
MIF1	Macrophage migration inhibitor factor 1
MRC1	Human macrophage mannose rececptor
MYB	Avian myoblast viral oncogene homologue

MYC	C-myc oncogene
MYCN	N-Myc oncogene
MYH1	Myosin heavy chain polypeptide
MYLL1	Myosin light polypeptide
NDPK	Norrie disease (pseudoglioma)
NDUFS	Mitochondrial NADH coenzyme Q reductase
NETRIN	Netrin 2
NRAMP	Natural resistance-associated macrophage protein 1
ODC1	Ornithine decarboxylase 1
OPN	Osteopontin
OPSIN	Opsin
OTC	Ornithine transcarbamylase
OVY	Ovalbumin Y
P2Y5	G protein coupled purino receptor
PDEA	cGMP α phophodiesterase
PGK1	Phosphoglycerate kinase
PGM2	Phosphoglucomutase 2
PPAT	Phosphoribosylpyrophosphate amidotransferase
PGR	Progesteron receptor
PLN1	Caudiae cardiac phospholamban
RAB6	GTP binding protein
RB1	Retinoblastoma oncogene
REPPROT	Human intestinal DNA replication protein
RNR1A	Ribosomal RNA 1A
RYR3	Ryanodine receptor
SNO-PEN	Ski novel overexpressed N
TAP2	Transporter 2, ABC(ATP binding cassette)
TCP1	T-complex 1
TCRA	T-cell receptor α chain
TFRC	Transferrin receptor
TGFB2	Transforming growth factor B2
TGFB3	Transforming growth factor B3
TRF-C1	Murine transcription factor C1
TH	Tyrosine hydroxilase
TVA	Avian leukosis virus subgroup A receptor
VIL	Villin 1
VIM	Vimentin
VTG2	Vitellogenin 2
WNT11	Wingless related MMTV integration site
YES1	Yamaguchi sarcoma viral oncogene homologue
ZFX/ZFY	X/Y linked zinc finger transcriptional activator