Linköping Studies in Science and Technology
Dissertation No. 2073

Genomics and Transcriptomics of Behaviour and Plumage Colouration

Jesper Fogelholm

Linköping Studies in Science and Technology
Dissertation No. 2073

Genomics and Transcriptomics of Behaviour and Plumage Colouration

Jesper Fogelholm

IFM Biology
Department of Physics, Chemistry and Biology
Linköpings universitet, SE-581 83 Linköping, Sweden
Linköping 2020

During the course of the research underlying this thesis, Jesper Fogelholm was enrolled in Forum Scientium, a multidisciplinary doctoral program at Linköping University, Sweden.

© [Jesper Fogelholm 2020]

Printed in Sweden by LiU-Tryck, Linköping, Sweden, 2020

ISSN 0345-7524

ISBN 978-91-7929-848-7

Genomics and Transcriptomics of Behaviour and Plumage Colouration

By

Jesper Fogelholm

April 2020

ISBN 978-91-7929-848-7

Linköping Studies in Science and Technology

Dissertation No. 2073

ISSN 0345-7524

Abstract

The aim throughout this thesis has been to investigate the underlying genetics of behaviours and feather colour and plumage patterns by using chickens as a model organism. Chickens are extremely important as a food source, both in terms of egg, as well as meat production. As such there is a large research interest for them, and they provide an excellent model to study the effects of domestication and evolution, since the ancestor to our domestic breeds the Red Junglefowl can still be found living freely in the wild. This allows us to set up long term crossing experiments where we can harness the power of recombination events and genome wide sequencing to perform genome wide mapping studies. I also want to take the opportunity to integrate the results from all of my work and consider it in perspective of the domestication syndrome.

In **Paper I** we investigated the **Social Reinstatement** behaviour which combines aspects of sociality and anxiousness. We detected several QTL and some overlap with Open Field behaviour from previous work within the group. By combining genomic and transcriptomic methods three strong candidate genes were found: ***TTRAP***, ***ACOT9*** and ***PRDX4***.

In **Paper II Tonic Immobility**, another classic behaviour was examined. Once more there was some overlap with the QTL regions discovered in earlier work, and it turns out that two of the most well supported candidate genes for Tonic Immobility is ***ACOT9*** and ***PRDX4***. These two genes had also been implicated with a pH dependent meat quality trait. Therefore, we conducted experiments in an additional smaller scale test cohort to investigate any potential link between the two traits. Following statistical multiple testing corrections, no significant association was found.

The remaining papers in the thesis investigated different types of feather patterning and colour. In **Paper III** we determined that the underlying genetic mechanism behind the striped appearance of the sex-linked barring feathers is likely caused by cyclic depletion and renewal of the pigment producing melanocyte cells during feather growth, which is a consequence of specific mutations in the gene ***CDKN2A***.

Paper IV took a quantitative approach to colour by measuring and quantifying the pheomelanin colour ranging from dark red to yellow. We identified five main candidate genes for the intensity of red colouration, ***CREBBP***, ***WDR24***, ***ARL8A***, ***PHLDA3*** and ***LAD1***. They are all regulated by a trans-acting eQTL located within the QTL region previously associated with behaviours in Paper I and Paper II.

Finally, in **Paper V** we turned our attention from pigment-based colour traits to an iridescent structural colour. Here we followed up the QTL mapping performed in our F₈ lab intercross with a Genome Wide Association Study in two feral populations from the islands of Kauai and Bermuda. RNA-sequencing was then performed in selected individuals from both feral populations in addition to individuals from the F₃ generation of our domestic x wild intercross. The main region of interest is located between 17.4 -17.5Mb on chromosome Z, with the main candidate genes being ***MAP3K1***, ***Zinc finger RNA binding protein 2***, and ***Zinc finger protein***.

After integrating and viewing the results from the work conducted as a part of this thesis from the perspective of the **Domestication Syndrome**, I have found that there are a lot of potential connections between the traits that I have studied. For instance, the same **QTL region on chromosome 10** is detected in association with the **behaviour traits** in Paper I and Paper II and the **quantitative colour trait** in Paper IV. I believe that the domestication syndrome is caused by the underlying functional arrangement of the genome, which causes correlated responses

in nearby genes and their associated traits, when selective forces such as domestication are applied on the primary trait.

Keywords: Genomics, Transcriptomics, Behaviour, Colour, Gene Expression, Domestication.

IFM Biology
Linköpings universitet
SE-581 83 Linköping, Sweden

Acknowledgement

First of all, I would like to thank my family, Camilla and Isabella, for more things than I have space to list here, and for listening to strange stories about tiny almost invisible blueprints that live inside each cell, and making it possible for me to go on conferences and spend a month away doing fieldwork on Hawaii.

I would also like to thank my supervisor Dominic Wright for the opportunity to conduct this work as a PhD student in his lab, as well as plenty of close fought games of Infinity over the last couple of years.

I would also like to thank all of my past and present colleagues at the Biology department that have helped me with all sorts of things over the years.

I especially want to thank my colleagues Rebecca, Mia and Andrey from my old office for their lovely company, support and thoughtful discussions whilst away on coffee supply runs, or in the immediately following Fika. The same applies for my colleagues in the new office, and anyone else (Truls) unfortunate enough to get caught by the coffee train.

I also want to add a special thanks to Robin and Marisa who have helped me in the lab, and on the computer and for proofreading and shaping of this thesis. And to “big-brother” Martin for help and guidance in the forest of Microarray analyses and coding in R.

And last but not least I want to thank everyone that contributed with chicken art for the cover of this thesis.

Linköping April 2020

Jesper Fogelholm

List of publications included in this thesis

Paper I

Genetics and genomics of social behavior in a chicken model.

Martin Johnsson, Rie Henrikssen, Jesper Fogelholm, Andrey Höglund, Per Jensen, Dominic Wright. *Genetics*. **209**, 209-221. (2018)

Paper II

Genetical genomics of Tonic Immobility in the chicken.

Jesper Fogelholm, Samuel Inkabi, Andrey Höglund, Robin Abbey-Lee, Martin Johnsson, Per Jensen, Rie Henrikssen, Dominic Wright. *Genes*, **10**, 341. (2019)

Paper III

The evolution of sex-linked barring alleles in chickens involves both regulatory and coding changes in CDKN2A.

Doreen Schwochow Thalmann, Henrik Ring, Elisabeth Sundström, Xiaofang Cao, Mårten Larsson, Susanne Kerje, Andrey Höglund, Jesper Fogelholm, Dominic Wright, Per Jemth, Finn Hallböök, Bertrand Bed'Hom, Ben Dorshorst, Michèle Tixier-Boichard, Leif Andersson. *PLOS Genetics*. **13**, 4. (2017)

Paper IV

CREBBP and WDR 24 identified as candidate genes for quantitative variation in red-brown plumage colouration in the chicken.

Jesper Fogelholm, Rie Henrikssen, Andrey Höglund, Nazmul Huq, Martin Johnsson, Reiner Lenz, Per Jensen, Dominic Wright. *Scientific Reports*. **10**, 1161. (2020)

Paper V

Identifying candidate genes for structural iridescent plumage colouration in the chicken.

Jesper Fogelholm, Maria Luisa Martin Cerezo, Andrey Höglund, Robin Abbey-Lee, Rie Henrikssen, Per Jensen, Dominic Wright.
(included in the thesis as a manuscript)

All papers included in this thesis, except for Paper II are reprinted under the Creative Commons CC BY 4.0 License.

[Creative Commons Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

No permission is needed for Paper II Genetical Genomics of tonic Immobility in the Chicken since it is reproduced as a part of a thesis.

“Permission from the GSA is not needed if you will use the material in an article published in GENETICS or if you are reproducing an article (on which you are an author) for your dissertation.”

List of publications not included in this thesis

Patterns of exchange of Multiplying Onion (*Allium cepa* L. *Aggregatum*-Group) in fennoscandian home gardens.

Matti W. Leino, Svein Ø Solberg, Hanna Maja Tunset, Jesper Fogelholm, Else-Marie Karlsson Strese, Jenny Hagenblad. *Economic Botany*. **72**, 346-356. (2018)

Genetical genomics of growth in a chicken model.

Martin Johnsson, Rie Henriksen, Andrey Höglund, Jesper Fogelholm, Per Jensen, Dominic Wright. *BMC Genomics*. **19**, 72. (2018)

Mating induces the expression of immune- and pH-regulatory genes in the utero-vaginal junction containing mucosal sperm-storage tubuli of hens.

Atikuzzaman M, Mehta Bhai R, Fogelholm J, Wright D, Rodriguez-Martinez H. *Reproduction*. **150**(6), 473-483. (2015)

The following papers are submitted as preprint versions:

Intra-individual behavioural variability: a trait under genetic control.

Rie Henriksen, Andrey Höglund, Jesper Fogelholm, Robin Abbey-Lee, Martin Johnsson, Niels Dingemanse, Dominic Wright
bioRxiv 795864; doi: <https://doi.org/10.1101/795864>

The genetic regulation of size variation in the transcriptome of the cerebrum in the chicken and its role in domestication and brain size evolution.

Andrey Höglund, Katharina Strempfl, Jesper Fogelholm, Dominic Wright, Rie Henriksen.
doi: 10.21203/rs.3.rs-20989/v1

Contents

Abstract	3
Acknowledgement	6
List of publications included in this thesis	7
List of publications not included in this thesis	9
Contents	11
Populärvetenskaplig sammanfattning	13
Paper summaries	17
Paper I	17
Paper II	18
Paper III	19
Paper IV	20
Paper V	22
Introduction	24
Advanced Intercross lines and genetic mapping	25
Genomics: Genome wide mapping analyses	26
Transcriptomics: gene expression analysis	28
Domestication and the domestication syndrome	30
Behaviour	32
Plumage colour	33
Integration	36
Feral birds on Kauai and Bermuda	37
Conclusion	42
List of references	43
Paper I	47
Paper	61
Paper III	79
Paper IV	103
Paper V	115
Avhandlingar från avdelningarna Biologi och Teoretisk Biologi	136

Populärvetenskaplig sammanfattning

Domesticering är en kraftfull process som kan åstadkomma enorma förändringar hos växter och djur tack vare artificiell selektion. Vår förmåga att domesticera djur och växter har varit extremt viktig för människors överlevnad under tusentals år. Tack vare denna process har vi bland annat kunnat samarbeta med hundar vid jakt, och vi har kunnat övergå från ett nomadsamhälle till en mer bofast livsstil med jordbruk som huvudsaklig födokälla. Domesticering kommer med stor sannolikhet även fortsatt vara viktigt då jordbruk och djurhållning är nödvändiga för att kunna säkerställa en tillräcklig matproduktion för en växande global population. Samtidigt har jordbruk och djurhållning stor miljöpåverkan och de måste de kunna samexistera med naturen.

När det kommer till djur talar man ofta om ett så kallat domesticeringsyndrom. Detta innefattar bland annat förändringar i fysiologi, beteende samt färg. Det som gör detta fenomen extra intressant att studera är att samma eller liknande förändringar återfinns hos olika djurslag. Det finns ett flertal olika teorier kring varför dessa länkar mellan olika fenotyper (egenskaper) uppkommer. För att kunna studera detta närmare använder vi oss utav korsningar mellan vilda Röda Djungelhöns och tama White Leghorns för att undersöka den underliggande genetiska arkitekturen som reglerar och kontrollerar fenotyper som förändrats under domesticeringsprocessen.

En mekanism som skulle kunna ge upphov till de länkade fenotyperna i domesticeringssyndromet är pleiotropi, det vill säga att en gen påverkar mer än en fenotyp. För att kunna länka samman gener och fenotyper använder vi oss utav en trestegsmetod som först identifierar ett genomiskt område med QTL-kartläggning. Detta följs upp med en genuttrycksanalys som inkluderar alla gener som ligger inom det identifierade QTL-intervallet. Slutligen kombinerar vi genetiska markörer (SNPs) med genuttrycksdata i en eQTL analys för att på så vis kunna identifiera de alleler som påverkar genuttrycket. Vilket i sin tur ligger till grund för de fenotypförändringar som observerats i individerna med just den specifika alleluppsättningen. I artikel I identifierade vi 24 QTL-regioner

som påverkar beteendet ”Social Reinstatement”. I korthet kan det beteendet beskrivas som hur benägen en individ är att söka upp sina artfränder när den släppts ut i en testarena. Detta beteende involverar dels komponenter utav socialt samspel, men även rädsla inför att vara isolerad. Vi identifierade fem kandidatgener som med stor sannolikhet påverkar detta beteende. Två utav dessa gener *ACOT9* och *PRDX4* har även förekommit i andra sammanhang, dels som kandidatgener för beteendet ”Tonic Immobility”, som är ett försvarsbeteende mot rovdjur, men generna har även visat sig påverka köttkvalitet via pH-förändringar. I artikel II undersökte vi denna koppling närmare för att se om någon utav generna har pleiotropa effekter. QTL och eQTL-kartläggningen stödde återigen *ACOT9* och *PRDX4* som kandidatgener, däremot såg vi ingen korrelation mellan de beteenden vi testade eller pH-värdet som påverkar köttkvaliteten.

Domesticeringsprocessen brukar ge upphov till nya färgkombinationer jämfört med icke-domesticerade vilda individer. Detta förekommer hos de flesta domesticerade djur och hos tamhöns finns mängder med färgvarianter som inte existerar i det vilda. Våra labbpopulationer är en korsning utav vilda Röda Djungelhöns samt tama White Leghorns. De uppvisar i tidiga generationer en uppsjö utav färgvarianter som gradvis övergår till att inkludera mer vitt. Tack vare att vi även har tillgång till två populationer med förvildade höns på öarna Kauai (som härstammar från korsningar mellan vilda Djungelhöns och tama Leghorns) samt Bermuda (härstammar från tamhöns) har vi en unik möjlighet att kunna kartlägga den underliggande genetiken som åstadkommer de olika färgerna vi kan se i fjäderdräktena.

Inom gruppen fåglar finns det ett flertal olika slags pigment. Hos höns färgas fjäderdräkten av två pigmenttyper. Dels pheomelanin som ger upphov till olika nyanser utav gult, orangt och rött, dels eumelanin som ger upphov till grått, svart och brunt. De mönster som kan ses i fjäderdräkten är extremt komplexa, då det finns mönster som innefattar hela eller mindre regioner utav fågeln, till exempel på vingarna eller ryggen. Men det går även hitta mönster på enstaka fjädrar i form utav ränder och olikfärgade zoner. Den reglering som krävs för att skapa dessa mönster behöver vara väldigt exakt, vilket innefattar både tidsbestämd och rumslig kontroll över genuttrycket. Det finns till och med fjädrar som har olika mönster på höger respektive vänster halva. Artikel III

identifierar hur två ”non-coding-mutationer” och två ”missense-mutationer” tillsammans leder till olika varianter utav ”barring”-fenotypen som ger upphov till tydliga ränder på fjädrarna. I artikel IV identifierar vi kandidatgener som ger upphov till färgnyanser i spannet mellan mörkrött och ljusgult. Genom att använda samma ”top-down”-metod, som vi tillämpar för beteendena och med kvantifiering av de olika nyanserna på en stigande skala, har vi identifierat fem kandidatgener. Genom att mäta genuttrycksnivåerna för alla gener som är aktiva när fjädrarna växer och färgen bildas, kan vi se att ett ökat uttryck utav generna *CREBBP* (som sedan tidigare är känd som en transkriptionsfaktor, samt är involverad i melaninsyntesen) och *WDR24* (som bland annat reglerar lysosomfunktionen) minskar mängden pheomelanin vilket innebär en reducering utav den röda färgen i fjädrarna.

Pigment som ger upphov till färg fungerar genom att selektivt absorbera och reflektera olika delar utav ljusspektrumet. I fallet med färgspannet som beskrivits ovan, absorberas allt ljus förutom det i intervallet mellan ~550-750nm, som istället reflekteras och därför upplevs föremålet som gult eller rödaktigt. Vit färg åstadkoms när allt ljus reflekteras, medan absorption av ljus istället resulterar i svart. Det går dock att åstadkomma färg genom att selektivt ”bryta isär” vitt ljus precis som i ett prisma, så kallad strukturell färg. Denna färg finns i fågelfjädrar, exempelvis den gröna färgen som kan ses hos skator och gräsänder. Den åstadkommes när ljus med rätt infallsvinkel träffar en speciell nanostruktur inuti fjädern. På grund utav de olika refraktiva egenskaperna hos de olika lagren inuti fjädern bryts ljuset och skapar en metallic-skimrande effekt. Detta fenomen är vanligt förekommande hos fåglar och den här typen utav färg verkar ha uppkommit för miljontals år sedan. Färgen kan till och med återfinnas hos dinosaurier utav släktet *Theropoda*, som är förfäder till dagens fåglar. Fenomenet och strukturen som ger upphov till denna speciella form utav färg är väldigt iögonfallande och har studerats noggrant under lång tid, och de fysiska aspekterna som krävs för att skapa färgen är välkända. Däremot är det, så vitt vi vet, ingen som känner till vilka de bakomliggande genetiska mekanismerna är. I labbpopulationer blir ofta den här färgen ovanlig efter ett par generationer vilket gör den svår att studera. Ytterligare en faktor som komplicerar är faktumet att fenotypen som ger upphov till färgen blir osynlig om fjädrarna inte har ett underliggande lager utav svart eumelanin. För att kunna identifiera de genetiska mekanismerna bakom färgen använder vi oss i artikel V utav dels

en labbpopulation för QTL-kartläggning, men även de förvildade hönspopulationerna på Kauai och Bermuda där den gröna strukturella färgen är vanligt förekommande. Genom att kombinera helgenomssekvensering och RNA-sekvensering kan vi identifiera enskilda punktmutationer (SNPS) i genomet som associerar med den strukturella färgen i en så kallad GWAS-analys. Hittills har vi observerat starka signaler från Z-kromosomen i närheten utav genen *MAP3K1* som sedan tidigare är känd för att vara involverad i regleringen utav signaleringskaskader. Vi har även hittat associationer med ett flertal *Keratin*-gener på kromosom 25. Fjädrarnas yttre lager är uppbyggt utav just detta protein, och man vet att det påverkar hur ljusets bryts, så dessa gener verkar väldigt lovande som kandidatgener för denna strukturella färg.

Avslutningsvis har jag integrerat och kombinerat resultaten från artiklarna i denna avhandling och analyserat dem utifrån ett domesticerings-syndroms-perspektiv. Jag har hittat åtskilliga länkar mellan resultaten från de olika artiklarna, exempelvis finner vi att QTL-regionen på kromosom 10 är underliggande både för beteendena i artikel I och II, men den är även associerad med den röd-gula färgen på fjädrarna i artikel IV. Enligt mig beror dessa återkommande kopplingar, som utgör domesticerings-syndromet, på den underliggande genetiska strukturen där generna är "sorterade" efter funktion längs med kromosomerna. När man selekterar för en egenskap via en eller flera gener, kommer det även att påverka den närmast intilliggande genomiska regionen, vilket påverkar de gener som befinner sig där. Om dessa närliggande gener påverkar andra egenskaper innebär detta då att selektionen för en egenskap även sammanfaller med ytterligare fenotyper. Om egenskapen i fråga är evolutionärt gammal, och evolutionärt konserverad, kan vi även förvänta oss att den intilliggande genomiska regionen är likartad i olika arter. Detta får då till följd att selektion för en egenskap får sammanfallande effekter i flera arter. Därför antar jag att domesticerings-syndromet är en följd utav den underliggande genetiska strukturen som skapats under årmiljoner utav evolution.

Paper summaries

Paper I

Genetics and genomics of social behaviour in a chicken model

In Paper I we used the eighth generation of an Advanced Intercross Line (F_8 AIL) and a three-stage method to identify candidate genes for a social behaviour, known as Social Reinstatement. This behaviour combines aspects of an anxiety/fear response with social motivation. The test is conducted in a runway arena with unknown conspecifics located at the far end from where the tested individual is released. An individual that spends more time in the vicinity of its conspecifics, and not exploring the arena is considered to be more anxious and has a higher social motivation.

The initial genomic mapping step was to perform a Quantitative Trait Loci (QTL) analysis for each of the 22 different measures obtained from the behaviour test. This identified 24 genome wide significant QTL on 16 chromosomes, there is a notable overlap on chromosome 2 and 10 between the Social Reinstatement QTL and Open Field behaviour from a previous study (M. Johnsson, Williams, Jensen, & Wright, 2016). This is followed up with a transcriptomic analysis using microarrays to measure gene expression levels for all genes located within the confidence interval of the detected QTL regions. The expression levels are combined with genomic marker data in an expression QTL analysis (eQTL). A total of five candidate genes show a correlation between gene expression and behaviour and have a regulatory eQTL present within the confidence interval of the QTL region. The gene *TTRAP* showed the highest support, but two other genes, *ACOT9* and *PRDX4* were also well supported see figure 1.

Figure 1. LOD score profiles along chromosome 10 and 2 for behavioural QTL in orange, as well as eQTL for *ACOT9* in black and *PRDX4* in light blue.

Paper II

Genetical genomics of tonic immobility in the chicken

In Paper II, we again used the F_8 AIL and a very similar approach to that used in Paper I in order to find candidate genes for another classic behavioural trait, Tonic Immobility. This is another fear related behaviour that is believed to be a form of predation defence and it is described in numerous species ranging from sharks to chickens. The test subject is placed on its back, and the time until the animal rights itself is then recorded, a longer duration is believed to represent a more fearful individual.

The initial QTL mapping analysis revealed that there was overlap between the QTL regions for Tonic Immobility and Social Reinstatement (see table 1). Following the same procedure as in Paper I, we identified seven candidate genes. Two of them were *ACOT9* and *PRDX4*, which indeed are the same as in Paper I. *PRDX4*, shows the highest correlation with tonic immobility. Another group had also identified *PRDX4* and *ACOT9* as strong candidate genes for a meat quality trait influenced by muscle pH. This prompted us to investigate any potential links between the Tonic immobility behaviour and pH in an additional F_{13} test cohort. Unfortunately, this endeavour proved unsuccessful and no significant correlation between the two traits was found.

Trait	Chr	Position	LOD	Add $\pm$ S.E.	Dom $\pm$ S.E.	Lower CI	Upper CI	Lower Marker	Position (Chr:bp)	Upper Marker	Position (Chr:bp)	Covariates	Interaction	R^2
Ti maximum d.	1	2014	5.8	-19.9 $\pm$ 21.6	-25.7 $\pm$ 42.1	1991	2038	1_190334672	1:190334672	1_195271649	1:195271649	sex, batch, PC2,10	182014.0:280485.0	3.7
Ti duration 2	1	1750	7.3	-30.8 $\pm$ 11.7	-2.0 $\pm$ 15.4	1735	1756	Gg_rs10728648	1:144081810	oup-23-342-18608-S-2	1:151495099	sex, batch, PC2,6	181748.0:7803.0	5
Ti average d.	2	774	6.5	-10.8 $\pm$ 9.2	-28.7 $\pm$ 14	764	795	RRL1120	2:116617839	Gg_rs15146557	2:117659953	sex, batch, PC2,6	26074.0:24860.7	4.6
Ti maximum d.	2	485	7.3	27.9 $\pm$ 13.1	-3.3 $\pm$ 17	474	495	Gg_rs14190859	2:60361704	Gg_rs15110213	2:65012153	sex, batch, PC2,10	182014.0:280485.0	4.7
Ti maximum d.	2	775	7.3	-9.9 $\pm$ 12.8	-18.1 $\pm$ 19.3	767	798	RRL1120	2:116617839	Gg_rs15146557	2:117659953	sex, batch, PC2,10	26075.0:24861.0	4.7
Ti duration 2	2	181	7.9	-63.0 $\pm$ 14.9	-35.7 $\pm$ 18.3	170	195	Gg_rs15070032	2:19473457	2_2309794	2_2309794	sex, batch, PC2,6	24860.7:28181.0	5.2
Ti duration 1	4	283	9.7	33.4 $\pm$ 10.5	11.4 $\pm$ 16.2	272	292	Gg_rs14446625	4:31963411	4_37860292	4_37860292	sex, batch, PC1,4	48283.0:158187.9	7.8
Ti duration 1	6	259	6.9	5.6 $\pm$ 9.1	10.7 $\pm$ 12.2	247	270	6_25762392	6:25762392	Gg_rs14592224	6:30433395	sex, batch, PC1,4	68258.7:108185.0	5.5
Ti duration 2	7	4	11.5	11.2 $\pm$ 13.5	48.5 $\pm$ 17.7	0	8	Gg_rs15826188	7:1654910	Gg_rs15828492	7:2444843	sex, batch, PC2,6	181748.0:7803.0	7.8
Ti average d.	10	99	5.9	-15.1 $\pm$ 7.7	14.9 $\pm$ 10.1	86	109	Gg_rs14941656	10:2678686	Gg_rs14030314	10:5805005	sex, batch, PC2,6	10899.0:208247.7	4.1
Ti maximum d.	10	139	8.4	-24 $\pm$ 11.4	-35.5 $\pm$ 15.1	133	144	10_9525779	10:9525779	Gg_rs14947769	10:11621480	sex, batch, PC2,10	108139.0:12885.0	5.4
Ti duration 1	10	185	8.5	2.4 $\pm$ 9.5	-34.6 $\pm$ 12.2	176	198	Gg_rs14008254	10:12705455	GC_rs14051952	10:15605204	sex, batch, PC1,4	68258.7:108185.0	6.8
Ti maximum d.	12	85	7.2	-24.2 $\pm$ 11.5	8.7 $\pm$ 14.7	77	190	Gg_rs1369494	12:5140826	12_14051161	12:14051161	sex, batch, PC2,10	108139.0:12885.0	4.7
Ti duration 1	15	188	7.9	-18.0 $\pm$ 9.6	-19.5 $\pm$ 12.3	177	189	Gg_rs14095161	15:10407522	Gg_rs14095923	15:11498734	sex, batch, PC2,6	48283.0:158187.9	6.3
Ti average d.	20	247.7	6.4	-9.4 $\pm$ 7.1	24 $\pm$ 10.1	237	252	Gg_rs15177950	20:10931547	Gg_rs14280872	20:13367853	sex, batch, PC1,4	10899.0:208247.7	4.8
Ti average d.	24	60.7	7	-20.9 $\pm$ 8.5	-21.0 $\pm$ 11	53	67	GC_rs161694400	24:1782119	Gg_rs14294768	24:3030659	sex, batch, PC2,6	26074.0:24860.7	4.9
Ti maximum d.	24	61	8.7	-26.4 $\pm$ 11.9	-25.0 $\pm$ 15.4	54	66	GC_rs161694400	24:1782119	Gg_rs14294768	24:3030659	sex, batch, PC2,10	26075.0:24861.0	5.6
Ti duration 2	24	61	16.5	21.5 $\pm$ 15.3	-40.9 $\pm$ 38.7	54	65	GC_rs161694400	24:1782119	Gg_rs14294768	24:3030659	sex, batch, PC2,6	7803.0:24860.7	11.5

Table 1. Results from the QTL mapping performed in the F_8 AIL population. Note that the region on chromosome 10 is overlapping with the Social Reinstatement QTL from Paper I, and the QTL on chromosome 1 is approximately 500cM upstream from the Social Reinstatement QTL.

Paper III

The evolution of *sex-linked barring* alleles in chickens involves both regulatory and coding changes in *CDKN2A*

In Paper III the candidate gene *CDKN2A* had already been identified in a previous study, the objective here was to investigate the underlying molecular mechanism for *sex-linked barring* which is a characteristic striped feather pattern with three known alleles B₀, B₁ and B₂.

Figure 2 depicts an overview of the results from this study. Upregulation of *CDKN2A* gene expression by is caused by non-coding mutations (part A in figure). Missense mutations in the B₁ and B₂ alleles disrupts the interaction between the protein and the next step in the signalling pathway (part B), thus counteracting the effects of the increased gene expression. Part C represents a wild-type single colour feather. In part D we can see an illustration of the proposed mechanism underlying *sex-linked barring*. The pool of pigment producing melanocyte cells is depleted and then replenished in a temporally cyclic fashion which causes the distinctive banding seen in the fully developed feather.

Figure 2. An overview of the mechanisms underlying the plumage pattern *sex-linked barring*

Paper IV

***CREBBP* and *WDR24* identified as candidate genes for quantitative variation in red-brown plumage colouration in the chicken**

In Paper IV we are investigating another plumage colour trait, this time we were interested in a quantitative trait. There are two main types of pigment in chicken feathers, eumelanin which produces black and brown colour and pheomelanin which corresponds to the yellow, orange and red end of the spectrum.

Just like in Paper I and II we start with large scale QTL mapping in the F₈ AIL, which is then followed up with eQTL fine mapping in two later generations (F₁₀ and F₁₂). Once again, the QTL region on chromosome 10 makes an appearance in addition to 6 more regions. Fine mapping and gene expression correlation with colour score revealed five main candidate genes, *CREBBP* and *WDR 24* on chromosome 14 and *ARL8A*, *PHLDA3* and *LAD1* on chromosome 26. Interestingly they are all regulated via a trans-acting eQTL located within the reoccurring chromosome 10 QTL region, and as we can see in figure 3b and 3c increased expression results in a reduction of red colour intensity.

Figure 3. An overview of the main results, in a we can see representative males from the two lines used for the intercross. b and c show the correlation between gene expression and colour score. d-f shows the overview and details of the eQTL analysis.

Paper V

Identification of candidate genes for a structural iridescent colour in chicken feathers

In Paper V we explore another type of plumage colouration, iridescent structural colour. Unlike the pigment-based colours, the iridescent type of structural colour changes appearance with the viewing angle. The reason for this phenomenon is that the colour is caused by a physical interaction between the incoming light and a specialised nanostructure within the feather barbules. This structure could be likened to a prism that breaks the incoming light into its component form and then selectively reflects specific wavelengths. The physical mechanism that causes this is well studied. It is the difference in refractive indices between the components of the nanostructure inside the feather keratin, melanin and air that produces these vivid metallic colours, that have an evolutionary origin reaching all the way back to the paravian dinosaurs. Despite this very little is known about the underlying genetics of structural iridescence.

The presence of Dominant White within our lab intercross makes this colour very rare and thus difficult to study. We performed initial QTL mapping within a subset of the F₈ AIL, which highlighted a region on the Z chromosome. This was followed up with a Genome Wide Association Study (GWAS) in two feral populations from the islands of Kauai and Bermuda where this colour is commonly found in both male and female chickens. The final step involves RNA sequencing to investigate gene expression levels in developing feathers. By combining these results, we identified ~130 candidate genes with some of our best candidate genes located in close proximity to the peaks of the GWAS analysis such as *MAP3K1* and *Zinc finger RNA binding protein 2*. An overview of the results from chromosome Z can be seen in figure 4.

Figure 4. This represents an overall view of the results connecting the structural indelcent colour with chromosome 2. We can see that several of the differentially expressed transcripts are located in the vicinity of the Fst -peaks and nearby the most significant SNPs from the GWAS.

Introduction

Chickens are the most widespread livestock in the world and outnumber humans by approximately 3 to 1. Humans have kept chickens for millennia and it is believed that they were domesticated around 8 000 years ago(Fumihito et al., 1994). But where did the first chicken come from? It is believed that the ancestor to the modern day production chicken is the Red Junglefowl native to the jungles of South East Asia(Fumihito et al., 1996). From there they have spread around the world and can now be found nearly everywhere there are humans. If we go back even further in time, to almost 67 million years ago, we could have seen the last common ancestor of what was going to become ducks and chickens walking around near the shoreline. The bird in question is called *Asterionis maastrichtensis* and a fossil was recently found in a quarry near Eben-Emael, Liège in Belgium(Field, Benito, Chen, Jagt, & Ksepka, 2020). If we delve even deeper into the evolutionary history of the chicken, and avian species as a whole, we arrive at the paravian dinosaurs which represent a middle ground between birds and dinosaurs around 120 million years ago. During this era, we can find giants like the *Tyrannosaurus rex* as well as smaller creatures like the velociraptor which both belong to the theropod group of dinosaurs which are the ancestors of all paravian dinosaurs. In this group of dinosaurs you can also find the first evidence of one of the most defining features of modern day birds: feathers(Xu et al., 2004).

Now that we have covered the evolutionary origins of the Red Junglefowl we can turn our attention to the process that turned them into our modern-day layers (egg-producing breeds) and broilers (meat production breeds): domestication. In some ways domestication can be seen as a sped-up version of evolution. Where natural selection has been replaced with artificial selection orchestrated by humans. Both processes have the power to change an animal to better fit in with its surroundings by altering their genome, which will leave tell-tale signs for scientists to discover.

Throughout my research, I have mainly used a domestic x wild intercross population, and a staged top down approach for the identification of candidate genes. First the genomic region associated with a trait is

identified by performing genome wide genetic mapping, which is then followed up with a more detailed examination of the region, for instance by performing gene expression analyses and fine mapping using additional marker data.

The aims of the papers I have included in this thesis revolves around the identification of candidate genes that regulate or control specific phenotypes. The first two papers are mostly related with domestication and candidate genes for fear related behaviours, as such they can have large impacts upon animal welfare. The subsequent papers have a more evolutionary perspective; focusing on different plumage phenotypes. Since both behaviour and colour are some of the core features of the widely discussed domestication syndrome it also features as a part of this thesis.

Advanced Intercross lines and genetic mapping

A common test design for genomic studies using QTL mapping is to perform an F_2 cross, however this design has low resolution caused by large linkage blocks, similar to a recent selective sweep, though even lower resolution. A standard F_2 cross will have an average confidence interval of 20 centi Morgans (see later), which depending on the species can easily represent around 20 Megabases in size. By expanding the test population beyond the F_2 generation and thereby turning them into an Advanced Intercross Line (AIL), it is possible to increase the resolution of genetic mapping, since each breeding event allows for recombination events to occur. The recombination events will fragment the large linkage blocks and thus yield a higher resolution during mapping. Most of the work in this thesis is founded upon the genetic mapping performed in an eighth generation (F_8) Advanced Intercross Line.

This thesis includes genomics as well as transcriptomics. The field of genomics involves research into the genomic structure and the DNA molecule which is the most fundamental unit in any living creature or plant. Transcriptomics has the RNA molecule at its core, which is the first level of output from the genome/DNA molecule. By combining these two research fields we can see how selection and mutations act upon the DNA and ultimately alter phenotypes that are mediated through alterations in gene expression.

As already mentioned, chickens are extremely important as a livestock animal, as such there is a great interest surrounding the genetics of chicken from an industrial standpoint. It is no surprise then that the chicken genome was one of the first genomes to be sequenced and made available. The draft version of the chicken genome was made available in December 2004 (International Chicken Genome Sequencing Consortium, 2004), which is a little more than one year after the first human reference genome was completed. The chicken genome is curated by the International Chicken Genome Consortium and the current version (at the time of writing) GRCg6a GCA000002315.5 was released in March 2018. A well-developed genetic toolkit and a thoroughly annotated reference genome are two of the cornerstones that are essential when it comes to identifying candidate genes.

Genomics: Genome wide mapping analyses

By combining phenotypic information with genetic markers such as Single Nucleotide Polymorphisms (SNPs) it is possible to statistically associate the phenotypic trait to a genomic location. Two common methods are Quantitative Trait Loci (QTL) mapping or Genome Wide Association Studies (GWAS). The principle behind them is that we have known genetic differences i.e. markers, between our two populations (domestic and wild) spaced out as evenly as possible across the genome. If all individuals with the trait of interest share the same version of the marker at a certain location, that location is potentially causative to the phenotype. The resolution i.e. the size of the detected region is dependent on the total number of markers but also on the genetic architecture of the study population itself. The association between the phenotypic effect and the marker is based upon the concept of genetic linkage (in the case of an F_2 or advanced intercross) or linkage disequilibrium (in the case of a Genome Wide Association Study). In the case of linkage, the underlying principle is that any two regions on a chromosome that are located close to one another will be inherited together. As the distance between the two points/markers increase, the likelihood of a recombination event happening at any point between the two markers also increases. The likelihood of a recombination event occurring between two points in the genome has been turned into a relative distance measurement, which is

measured in centi-Morgan (cM) after the geneticist Thomas Hunt Morgan. The distance of 1cM is equal to a recombination likelihood of 1%, i.e. in 100 offspring we expect 1 recombination event to occur. This in turn means that in one offspring we expect to find that the allele for marker A is shared with one parent and that the allele for the nearby (1cM away) marker B is shared with the other parent. In the other 99 cases we expect the offspring to share both the A and B marker allele with one of its two parents, i.e. there is a 1 in 100 chance that a recombination event (cross-over) in the offspring disrupts the allelic combination seen in the parents.

A QTL mapping study statistically tests whether the trait of interest co-segregates with any of the known marker alleles, and the result is a genomic range based in cM. This means that we can only determine the physical location of the QTL to its nearest marker. A QTL study uses relatively few but carefully selected markers to test for associations with the phenotype, which might impact resolution negatively, but it is very useful if sequencing power is limited and is still useful as a first step towards identification of candidate genes. Another approach is the Genome Wide Association Study or GWAS for short, here we harness the power of modern genomics using either high density SNP chips or high-throughput sequencing to generate thousands or even hundreds of thousands of SNP genotypes. We are still using statistics to test for associations between our selected phenotype and the different alleles of our SNP genetic markers, but the key difference is found in the amount of SNP marker data available. The classic QTL study maps recombination breakpoints, and therefore needs a far less dense marker map to cover the genome. In contrast, the GWAS uses the huge number of recombinations that have built-up over many years in a natural population. Whilst the F_2 intercross QTL study requires a strict breeding design created by crossing two parental populations, before then inter-crossing the F_1 generation, the GWAS uses any available natural population, or subset of several populations. These then generate far smaller confidence intervals for the detected loci. This also comes at a price, however, as the effect sizes of detected alleles (the amount of phenotypic variation that the allele explains) in such natural populations is often far smaller than that of a QTL detected in a linkage study. The GWAS also requires a much larger number of markers to cover the genome.

Whole genome sequencing data also allows for identification of regions under selection in the genome of a population. Tajima's D is a measurement that compares the expected mutational frequency under neutral selection with the observed frequency. By doing this it is possible to find genomic regions that are under negative or positive selection. Selective sweeps are phenomena that occur in the genome whenever there is positive selection on an allele. Just as we see linkage between our markers and the phenotype in the QTL mapping we also expect to find linkage between the selected allele and the neighbouring genomic region surrounding it, often referred to as the hitchhiking effect. The more recent the sweep is, the larger these regions are expected to be since recombination events haven't started to break them apart (more on recombination in the AIL section). Another available method is Fixation index scanning (F_{ST} -scanning) where variants of SNPs in two or more groups are compared against each other as seen in Paper V. It measures between 1 and 0, where 0 indicates a shared allele and 1 indicates that the groups being compared have alternative alleles at that position. In essence the F_{ST} describes if a genomic region is from a distinct subgroup or instead if it represents a single admixed population, with a greater than expected number of homozygote genotypes (an excess of homozygosity) if subgroups are present. This can easily be visualised and allows easy identification of regions that are differentiated between the two groups across the genome.

Transcriptomics: gene expression analysis

There are several methods for performing gene expression analyses, the one thing they have in common is that they will all provide a snapshot of the ongoing activity in the cells of the selected tissue. If the tissue type under investigation is an amalgamation of different cell types, we will see the average gene expression pattern across those cell types unless we are using single cell RNA-sequencing. This has a lot of implications for the process of identifying candidate genes for a trait. The genomic analyses will often point towards a region that contains large numbers of different genes, but it does not provide any information about which type of tissue or in which time window the gene(s) in those regions exert their effect upon the phenotype.

If you are interested in the expression levels of a gene within a tissue, performing a Quantitative Polymerase Chain Reaction (qPCR) is the most straight forward approach. It is a robust method that has been used countless times and is particularly suited towards analysing a smaller number (<10) of candidate genes. For Paper II where we had a very reduced list of candidate genes, due to previously completed work we used this targeted approach to investigate the gene expression levels in two tissue types. Despite this reduced number of candidate genes when we applied multiple testing correction to the results they were no longer significant at the adjusted $\alpha < 0.05$ level. This highlights another problematic area in candidate gene discovery, and considering that our other two favourite methods for measuring gene expression measures the expression levels of 16 878 genes at the same time, this problem is even more pronounced. In Paper I, II and IV we have used Microarrays to measure the level of gene expression across the whole chicken genome, this is accomplished by measuring the expression levels of 32 785 probesets, this means that every gene is on average measured with 1.95 probesets. In order to restrict the amount of multiple testing corrections we have to apply, we only consider the genes found within the confidence intervals of the detected QTLs. By correlating the expression levels of those genes, with the quantitative phenotype values, we can filter out any genes that are not statistically associated with the phenotype. The last step is to perform a causality analysis where we statistically test whether the observed gene expression levels of the remaining genes are dependent on the allelic versions of our selected genetic markers. If the gene expression levels are different for the two alleles of our genetic marker, we now know that something in linkage with this marker is altering the expression level of that gene. We can also tell from which genetic background the two alleles come from, i.e. the wild Red Junglefowl or the domestic White Leghorn. At this stage, we can tell that something in linkage with the marker is altering the gene expression levels of specific genes that are correlated with the phenotype in question. For Paper V we have followed the same overarching steps, but we decide to use whole genome data in addition to the QTL mapping which was combined with RNA sequencing. Compared to the microarrays of the previous papers this now allows us to test all 39 288 known transcripts within the genome, which also includes alternatively spliced transcripts from the same gene. This means we get an even more detailed picture, and that we have an even more in-depth knowledge about the underlying

genetic mechanism regulating the trait. However, in Paper V, we were essentially looking at a binary trait (presence or absence of structural iridescent green), so were therefore unable to correlate strength of phenotypic expression with strength of gene expression. Therefore, both strengths and weaknesses are present in both methods.

Domestication and the domestication syndrome

The modern day chicken is thought to have been domesticated around 8000 years ago somewhere in central Asia (Fumihito et al., 1994; Miao et al., 2013; Tixier-Boichard, Bed'hom, & Rognon, 2011). The behaviour of a domestic animal is markedly different to that of its wild counterpart. Since the ancestor of our domestic breeds can still be found in the wild we have a unique opportunity to study the effects of domestication in the chicken. The first two papers in this thesis investigate the behavioural response to potentially stressful and or fearful situations. Selection on an altered fear response (especially towards humans) is most likely a key factor in the domestication process. Selection on an altered fear response is quite likely to have happened in all domestic animals since they need to survive and breed in close contact with humans. After a number of generations of selective breeding, the animals can now live and breed under the care of humans. This domestication process usually results in animals that are markedly different in both behaviour and appearance compared to their wild counterparts. This is perhaps not a huge surprise given that the selective pressure caused by the artificial selection is extremely high and that the living environment of the animal is also extremely different. There is one particular aspect of domestication that has intrigued scientists for ages, most domestic animals resemble one another and seem to share key characteristics, such as coat colour, increased tameness and altered cranial morphology to name a few. This is known as the domestication syndrome, and it has also divided the opinion of researchers for just as long, see for instance (Lord, Larson, Coppinger, & Karlsson, 2020). I would like to take this opportunity to share my views on the matter. The question is whether the domestication process in itself leads to the same type of correlated phenotypic changes in all domesticated animals, or if there has been independent selection for all of the traits?

Before we go any further, there is one thing in particular we should address, my definition of a domestic animal because that will have fundamental implications for this question as a whole. I think that most would agree that evolution is an ongoing process, as such all living things are continually adapting and evolving. In my opinion domestication and evolution are fundamentally the same, the difference is the origin of the selective pressures changing the organism in question. This means that I also consider domestication an ongoing process. Animals are not simply domesticated or not, there is a sliding scale with different degrees of domestication possible, i.e. the more generations a species has been subject to the selective pressures caused by domestication the more domesticated it will be. This means that there is no upper limit since it is a continuing process. That has major implications for the definition of the domestication syndrome as well. Some hold the view that in order for the domestication syndrome to be true all domestic animals must display all of the suggested traits in order for the syndrome to be valid. My definition would be that the same traits appear across different species without the specific selection upon those traits. Potentially any given species can be domesticated to a varying degree, thus they can display more or less of the different aspects of the domestication syndrome.

I believe that the most fundamental question that needs to be asked in relation to the domestication syndrome is; what is selection actually acting upon, and what is required in order to alter a phenotype? If we consider a theoretical scenario for a growth trait. Our aim is to select for a larger chicken. In a single gene, single phenotype scenario we could have selection on a gene encoding a glucose receptor that acts as the first input to an energy homeostasis pathway. By selecting for larger chickens in our population we are lowering the activation threshold for this pathway which changes the food search behaviour of the chickens with the alternative allele and they will start searching for food at an earlier point compared to the others, as a result they will also grow to a bigger size because they also behave differently. This has now also created a scenario were this gene could be classified as being pleiotropic since it alters both behaviour and physical size.

Based upon the research I have conducted as a part of this thesis I believe that single gene traits are quite a bit rarer compared to complex

traits that are regulated by many genes. It is a well-known fact that genes are not randomly distributed across the genome. It is also known that genes which are clustered together, tend to be expressed together (which is something I have observed in our gene expression data as well) and usually take part in the same or similar functional networks (Hershberg, Yegerlotem, & Margalit, 2005; Williams & Bowles, 2004; Woo, Walker, & Churchill, 2010). There is also evidence for spatial arrangement within the cell nuclei (Thévenin, Ein-Dor, Ozery-Flato, & Shamir, 2014), and genomic regions that are found to be in contact with each other are known as Topologically Associated Domains (TADs). The implication for these regions are currently being investigated see (Beagan & Phillips-Cremens, 2020) for a review. Another question is whether genes can perform different functions in different tissues depending on the availability of other genes in the network (McCole, Erceg, Saylor, & Wu, 2018). I think it could be argued that a complex trait with input from multiple genes can be regulated with higher precision/resolution compared to a single gene case since there will be more opportunities for interactions with the network and more targets for the selective forces to act upon. I also believe that a network is more resilient towards perturbations since there are more opportunities to rescue the phenotype, the missense mutations in the B₁ and B₂ alleles in Paper III could be considered an example of this.

Behaviour

Fear related behaviours such as Social reinstatement and Tonic immobility can have large impacts on the welfare of animals in a production setting. Not only as direct short-term effects but also in the form of a changed overall stress level caused by the altered fear response which could lead to chronic stress issues. As with any genomic mapping study it is extremely important that you are aware of exactly what you are measuring, and the consequences that has for the genetic mapping. For example, in Paper IV where we investigate a colour trait we find different genomic regions (QTLs) for the peak colour intensity in single feather and the average value measured across the wing. In this case it is not unlikely that this is because the measure of average colour across the wing could be influenced by certain aspects of patterning, which then most likely have a separate architecture from the colour. Behaviours

such as Social reinstatement and Tonic immobility are complex traits to analyse, and just like with the colour phenotypes the interpretation of the result is also linked with the interpretation of the test. I would consider myself more of a geneticist than an ethologist, but my interpretation is that some aspects of the two behavioural tests in Paper I and Paper II are overlapping and it might be related with a more general fear response. An overlapping QTL for open field behaviour has also been detected in this intercross (M. Johnsson, Williams, et al., 2016). The candidate genes for the behaviours investigated in Papers I and II, *ACOT9* and *PRDX4* also appear as candidates in a set of articles on meat quality traits (X. Li et al., 2015; J. Nadaf et al., 2014; Javad Nadaf et al., 2007). As such, there might be a common pleiotropic core that regulates and modulates certain aspects of all these traits, which incidentally also fits well with the appearance of a domestication syndrome. Another aspect of domestication is the appearance of novel coat or plumage colours and patterns not seen in the wild (Cieslak, Reissmann, Hofreiter, & Ludwig, 2011).

Plumage colour

The remaining three papers in this thesis are all investigations of different colour phenotypes. There are two main types of colour in feathers, pigment based and structural colours. The main difference between them is that pigments absorb certain wavelengths and reflects others, and they are also angle independent i.e. they always look like they have the same colour. Whereas structural colours, as the name suggest depend on nanoscale structures within the feather to interfere with the light and only reflect specific wavelengths. In the chicken plumage we find two types of pigment, either eumelanin which is responsible for creating blacks and browns, or pheomelanin which produces colours in the yellow to red spectrum.

Structural colours can also be divided into two main types, the angle independent, or the angle dependent iridescent colours (see figure 5 for an example of an angle dependent colour). In the *Galliformes* order iridescent structural colours are common, they will usually appear black, unless they are viewed from the correct angle where they will sparkle in metallic shades of blues, greens and in some cases even gold. In the chicken, we primarily find iridescence on the wings and in the tail feathers and they are usually green or purple in appearance.

Figure 5. Examples of iridescent feathers from magpie and chicken. On the left-hand side, we can see what appears to only be a black feather from a magpie, when viewed from a slightly different angle its iridescent nature becomes apparent.

In the lower right corner, we can see that the iridescent pattern is continuous over several feather. In the chicken feather in the top right corner we can see a faint horizontal banded effect with purple and golden regions offset from the main green hue.

The colour producing cells are known as melanosomes and they originate from the same stem cells as the nervous system, which has led to the rise of the neural crest hypothesis, which suggests that behaviour and colour are linked by the development and distribution of these neural crest cells and by the common cell lineage of the nervous system and melanocytes (Wilkins, Wrangham, & Fitch, 2014). Feathers are a key characteristic of birds and are usually associated with flight, however they appeared in the evolutionary lineage of birds long before they were capable of flight. What is believed to be structurally coloured iridescent feathers have been found in fossils of paravian dinosaurs such as the *Microraptor* (Q. Li et al., 2012). This suggests that the genetic architecture of plumage colouration is evolutionary ancient and is very likely to be well conserved between different species of birds. There is evidence that some colour phenotypes are used as honest signalling of individual fitness, see for instance (Leclaire, Perret, Galván, & Bonadonna, 2019), and that sexual selection seems to be a large driving force behind the divergence of colour in birds (Cooney et al., 2019). As such feathers and their colours are important from an evolutionary perspective. The connection between plumage colour, sexual selection and honest signalling might also have implications for the domestication syndrome.

If the genetic architecture of colour phenotypes is evolutionarily ancient, and they represent honest signals of individual fitness, we should expect to find traits that are linked with the colour trait by for instance pleiotropy. If the selection caused by domestication acts upon these evolutionarily old architectures that are already linking traits together we should see changes in multiple traits. This could neatly explain how selection upon behaviour alters the plumage colour in the chicken or vice versa. By integrating the results from the papers in this thesis, I believe we can provide some insight into the underlying cause of the domestication syndrome

Integration

In Figure 6 I have summarised the overlap between candidate genes and QTL regions for all five papers included in this thesis. In Paper III a sex-linked colour phenotype that causes striped feathers was investigated, with the conclusion being that a gene known as *CDKN2A* regulates the appearance of the feather by temporal management of the melanocytes. I have not found any direct links between this gene and behaviour, however the neighbouring gene on the Z chromosome *MTAP* is actually one of the candidate genes for the Social Reinstatement behaviours investigated in Paper I. In Paper IV I have studied a quantitative colour trait namely the intensity of red-brown colouration. In this case, we can see that the trans-acting eQTL that regulates the expression of these candidate genes is located within the QTL region associated with the behaviours in Paper I and II. Another interesting fact is that there is a selective sweep with a domestic origin within this region. So, there is evidence that there has been recent selection acting upon this genomic region.

Figure 6. Overlapping results from the five papers that are included in this thesis. Arrows indicate connections between candidate genes and traits. The grey boxes represent additional phenotypes that might be under selection during domestication.

The white colour seen in the White Leghorns is known as Dominant White, and for a good reason. After a number of generations, the number of non-white individuals in the population starts to decrease rapidly, since this locus is dominant it will override the coloured phenotypes and the proportion of white individuals increases dramatically. This might of course happen with other phenotypes that are less conspicuous, in those cases it will be harder to detect that there is a population bias. It would be ideal to have another population of chickens that share a similar genetic background with our lab population to reduce the risk of genetic drift and fixation influencing our results. If that population was living freely on an island, external gene flow would be virtually non-existent, and would also mean that sexual selection is reinstated. Phenotypes such as plumage colour, which might be under sexual selection would be ideal to study in such a population.

Feral birds on Kauai and Bermuda

It just so happens that at least two such populations actually exist on Kauai and Bermuda, you might even have witnessed the powerful event that created one of these populations without realising. If you have seen the movie Jurassic Park from 1993 you might recall that a tropical storm featured in the plot of the movie. This movie was made on the Hawaiian island of Kauai, and the storm in the movie is actually hurricane Iniki, which is one of the largest hurricanes to hit the islands. The hurricane released a lot of domestic chickens on the island, and since then they are roaming freely. This process is known as feralization, where a once tame animal is released into the wild, it can be likened to the opposite of domestication, although not genetically see (M. Johnsson, Gering, et al., 2016) for details. From a selection perspective this means that both natural and sexual selection is reinstated. However, it could probably be argued that one of the primary reasons for why there are so many chickens on Kauai is that they have very few (if any) natural predators on the island. As such the total load of selective pressure is quite likely significantly lower compared to the jungles of South East Asia. This would also mean that the sexual selection which is still present becomes relatively higher. The situation on Bermuda is very similar to the one on Kauai, but there is one key difference between the two populations. The

chickens on Bermuda all originate from modern domestic breeds, which is not the case on Kauai. Evidence suggest that the domestic breeds that were released in the 90s have interbred with a much older population of chickens, most likely Red Junglefowl brought by the Polynesian settlers around 800 years ago (Gering, Johnsson, Willis, Getty, & Wright, 2015; M. Johnsson, Gering, et al., 2016). In these populations we should expect to find signals of recent selection such as selective sweeps, especially in the case with traits that under sexual selection. Since the birds on Kauai have a larger and more diverse gene pool, we would expect to find a larger variation in phenotypes here unless they are being actively selected on by directional selection.

In figure 7 we can see a comparison of the colour distribution across the two islands. Here we can see that the purple version of the structural iridescence is more common on Bermuda, in both males and females. The instances of white patches in the plumage is also much higher on Bermuda relative to Hawaii (the size of the patches is also generally much larger on Bermuda), but compared to our intercross populations there is a decreased amount of white in both populations. This seems to fit well with the idea that there is a higher number of Red Junglefowl alleles available on Hawaii, it also seems as if there is active (sexual) selection upon the iridescent colours in both feral populations.

Figure 7. In this figure we can see how common each colour is in the plumage of males and females that have been sampled on the two islands. The numbers above each bar represent fractions relative to the total number sampled for each sex on the island.

The list of candidate genes for the structural iridescent colour in Paper V is currently quite large at 126 genes total for the complete analysis and around 180 genes for the single feather comparison. It seems quite likely that the structural iridescent colour is a complex trait which is regulated by more than one gene, however I do not think that all of them are directly linked with the regulation of the iridescent colour, although theoretically, they could be. Some candidates are of course more interesting than others, especially if the gene is located in close proximity with a significant SNP from the GWAS. However, we could also look from another perspective at these candidate gene lists. We have reason to believe that there has been recent selection for the iridescent phenotype on both Bermuda and Hawaii since it is much more common than in our outbred intercross. If selection is acting at the gene expression level, we might be seeing the effect of this selection. Since the gene expression levels between neighbouring genes are correlated, and there seems to be active selection for altered expression levels, we might be seeing the correlated response in the genomic area surrounding the as-of-yet-unknown causative genes. This is then the transcriptomic equivalent of a genomic selective sweep, just like its genomic counterpart the length of the sweep region will be reduced over time as recombination or following selection fine tunes the expression levels of the affected genes. If such transcriptomic sweeps exist, they could also be a part of the explanation for the domestication syndrome. If there is selection acting upon a domestication trait gene network, it could also affect the neighbouring gene networks on the genome. This could for instance be mediated through transcription factors, or increased recruitment of transcriptional machinery that in turn could be affected by changes in the DNA methylation pattern. Another potential mechanism is the spatial structuring of the genome within the cells i.e. TADs. Either of these scenarios could lead to a higher number of genes showing expression level correlations with the trait in question.

Among the genes that are statistically correlated with the structural iridescent colour phenotype on Hawaii and Bermuda (which might have been under strong selection for around 30 years) we find quite a few that have been established as candidate genes for other domestication traits such as behaviour. The gene *ANKRD29* is included as a candidate gene both for structural iridescence in Paper V as well as Social Reinstatement in Paper I. It has also very recently been associated with plumage

colour scored using a colorimeter (Huang, Pu, Song, Sheng, & Hu, 2020). Another gene of interest is *SMAD6* which is an integral part of the *TGF- β / BMP* signalling pathway, *SMAD6* inhibits *BMP2* (Imamura et al., 1997; Lin et al., 2003; Tsubakihara et al., 2015) *BMP2* in turn is important for egg laying and bone strength characteristics, as well being a candidate (along with the adjacent gene *HAOI*) for comb size (Martin Johnsson et al., 2012). Comb size in chickens is a sexually selected ornament, used by both males and females for mate choice purposes. It was also found to be an honest and reliable indicator of fecundity investment and bone density in females, and bone density in males. This is not the only colour gene with a link with comb size, with the gene *PIK3R1* also having prior links to comb size in chickens (Y. Liu et al., 2018).

Two of the candidate genes identified in Paper IV *CREBBP* and *WDR24* were very recently found to be down regulated in the pituitary gland of chicken with high egg production measured over 300 days (Mishra et al., 2020). *CREBBP* and *WDR24* belong to the Jak-STAT and m-TOR signalling pathways respectively. This represents another case where a trait that is very likely to be under selection during domestication (egg laying and altered reproduction cycle) shares candidate genes with a trait (colour) that is included in the domestication syndrome.

ARVCF is another gene that has been identified in relation with iridescent colouration in Paper V. In the scientific literature it is mostly known for its relation with schizophrenia (C. Liu et al., 2019; Mas et al., 2010, 2009) but it has also been associated with working memory in mice (Suzuki et al., 2009). In a study investigating the genetic basis of beagle-human communication *ARVCF* was found to be associated with human directed help-seeking behaviour (Persson, Wright, Roth, Batakis, & Jensen, 2016). Once again, we find a potential connection between colour and a domestication related trait. Which highlights the potential for common gene networks and preserved pathways across species.

Conclusion

It seems then, that there are multiple cases where genes are being identified as candidate genes for more than one of the traits that are suggested to form the domestication syndrome. Based on my own research and observations that are included in this thesis I would suggest that this is caused by the underlying genetic architecture.

Selection for one trait alters the gene expression pattern of multiple surrounding genes as well as for the true causative gene(s), this will cause secondary phenotypic alterations that appear together with the first trait. If the underlying genetic architecture is evolutionary conserved, it is possible that similar selection across different domesticated species will cause similar secondary phenotypic changes since they affect the same gene networks that are performing similar functions across different species.

List of references

- Beagan, J. A., & Phillips-Cremins, J. E. (2020). On the existence and functionality of topologically associating domains. *Nature Genetics*, 52(1), 8–16. <https://doi.org/10.1038/s41588-019-0561-1>
- Cieslak, M., Reissmann, M., Hofreiter, M., & Ludwig, A. (2011). Colours of domestication. *Biological Reviews*, 86(4), 885–899. <https://doi.org/10.1111/j.1469-185X.2011.00177.x>
- Cooney, C. R., Varley, Z. K., Nouri, L. O., Moody, C. J. A., Jardine, M. D., & Thomas, G. H. (2019). Sexual selection predicts the rate and direction of colour divergence in a large avian radiation. *Nature Communications*, 10(1), 1773. <https://doi.org/10.1038/s41467-019-09859-7>
- Field, D. J., Benito, J., Chen, A., Jagt, J. W. M., & Ksepka, D. T. (2020). Late Cretaceous neornithine from Europe illuminates the origins of crown birds. *Nature*, 579(7799), 397–401. <https://doi.org/10.1038/s41586-020-2096-0>
- Fumihito, A., Miyake, T., Sumi, S., Takada, M., Ohno, S., & Kondo, N. (1994). One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. *Proceedings of the National Academy of Sciences of the United States of America*, 91(26), 12505–12509. <https://doi.org/10.1073/pnas.91.26.12505>
- Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., ... Ohno, S. (1996). Monophyletic origin and unique dispersal patterns of domestic fowls. *Proceedings of the National Academy of Sciences of the United States of America*, 93(13), 6792–6795. <https://doi.org/10.1073/pnas.93.13.6792>
- Gering, E., Johnsson, M., Willis, P., Getty, T., & Wright, D. (2015). Mixed ancestry and admixture in Kauai’s feral chickens: invasion of domestic genes into ancient Red Junglefowl reservoirs. *Molecular Ecology*, 24(9), 2112–2124. <https://doi.org/10.1111/mec.13096>
- Hershberg, R., Yegerlotem, E., & Margalit, H. (2005). Chromosomal organization is shaped by the transcription regulatory network. *Trends in Genetics*, 21(3), 138–142. <https://doi.org/10.1016/j.tig.2005.01.003>
- Huang, T., Pu, Y., Song, C., Sheng, Z., & Hu, X. (2020). A quantitative trait locus on chromosome 2 was identified that accounts for a substantial proportion of phenotypic variance of the yellow plumage color in chicken. *Poultry Science*. <https://doi.org/10.1016/J.PSJ.2020.01.030>
- Imamura, T., Takase, M., Nishihara, A., Oeda, E., Hanai, J., Kawabata, M., & Miyazono, K. (1997). Smad6 inhibits signalling by the TGF- β superfamily. *Nature*, 389(6651), 622–626. <https://doi.org/10.1038/39355>

- International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432(7018), 695–716. <https://doi.org/10.1038/nature03154>
- Johnsson, M., Gering, E., Willis, P., Lopez, S., Van Dorp, L., Hellenthal, G., ... Wright, D. (2016). Feralisation targets different genomic loci to domestication in the chicken. *Nature Communications*, 7(1), 12950. <https://doi.org/10.1038/ncomms12950>
- Johnsson, M., Williams, M. J., Jensen, P., & Wright, D. (2016). Genetical Genomics of Behavior: A Novel Chicken Genomic Model for Anxiety Behavior. *Genetics*, 202(1), 327–340. <https://doi.org/10.1534/genetics.115.179010>
- Johnsson, Martin, Gustafson, I., Rubin, C.-J., Sahlqvist, A.-S., Jonsson, K. B., Kerje, S., ... Wright, D. (2012). A Sexual Ornament in Chickens Is Affected by Pleiotropic Alleles at HAO1 and BMP2, Selected during Domestication. *PLoS Genetics*, 8(8), e1002914. <https://doi.org/10.1371/journal.pgen.1002914>
- Leclaire, S., Perret, S., Galván, I., & Bonadonna, F. (2019). Pheomelanin-based coloration is related to individual quality and oxidative stress in blue petrels. *Evolutionary Ecology*, 33(6), 873–887. <https://doi.org/10.1007/s10682-019-10010-7>
- Li, Q., Gao, K.-Q., Meng, Q., Clarke, J. A., Shawkey, M. D., D’Alba, L., ... Vinther, J. (2012). Reconstruction of Microraptor and the evolution of iridescent plumage. *Science (New York, N.Y.)*, 335(6073), 1215–1219. <https://doi.org/10.1126/science.1213780>
- Li, X., Liu, X., Nadaf, J., Le Bihan-Duval, E., Berri, C., Dunn, I., ... De Koning, D.-J. (2015). Using Targeted Resequencing for Identification of Candidate Genes and SNPs for a QTL Affecting the pH Value of Chicken Meat. *G3 & Genes | Genomes | Genetics*, 5(10), 2085–2089. <https://doi.org/10.1534/g3.115.020552>
- Lin, X., Liang, Y.-Y., Sun, B., Liang, M., Shi, Y., Brunicardi, F. C., ... Feng, X.-H. (2003). Smad6 Recruits Transcription Corepressor CtBP to Repress Bone Morphogenetic Protein-Induced Transcription. *Molecular and Cellular Biology*, 23(24), 9081–9093. <https://doi.org/10.1128/MCB.23.24.9081-9093.2003>
- Liu, C., Kanazawa, T., Tian, Y., Mohamed Saini, S., Mancuso, S., Mostaid, M. S., ... Bousman, C. (2019). The schizophrenia genetics knowledgebase: a comprehensive update of findings from candidate gene studies. *Translational Psychiatry*, 9(1). <https://doi.org/10.1038/s41398-019-0532-4>
- Liu, Y., Tu, Y., Zhang, M., Ji, G., Wang, K., Shan, Y., ... Zou, J. (2018). Identification of molecular pathways and candidate genes associated with cocks’ comb size trait by genome-wide transcriptome analysis. *Scientific Reports*, 8(1), 2015. <https://doi.org/10.1038/s41598-018-20373-6>

- Lord, K. A., Larson, G., Coppinger, R. P., & Karlsson, E. K. (2020). The History of Farm Foxes Undermines the Animal Domestication Syndrome. *Trends in Ecology & Evolution*, 35(2), 125–136. <https://doi.org/10.1016/J.TREE.2019.10.011>
- Mas, S., Bernardo, M., Gassó, P., Álvarez, S., Garcia-Rizo, C., Bioque, M., ... Lafuente, A. (2010). A functional variant provided further evidence for the association of ARVCF with schizophrenia. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, 153(5), 1052–1059. <https://doi.org/10.1002/ajmg.b.31073>
- Mas, S., Bernardo, M., Parellada, E., Garcia-Rizo, C., Gassó, P., Álvarez, S., & Lafuente, A. (2009). ARVCF single marker and haplotypic association with schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 33(6), 1064–1069. <https://doi.org/10.1016/j.pnpbp.2009.06.001>
- McCole, R. B., Erceg, J., Saylor, W., & Wu, C. (2018). Ultraconserved Elements Occupy Specific Arenas of Three-Dimensional Mammalian Genome Organization. *Cell Reports*, 24(2), 479–488. <https://doi.org/https://doi.org/10.1016/j.celrep.2018.06.031>
- Miao, Y.-W., Peng, M.-S., Wu, G.-S., Ouyang, Y.-N., Yang, Z.-Y., Yu, N., ... Zhang, Y.-P. (2013). Chicken domestication: an updated perspective based on mitochondrial genomes. *Heredity*, 110(3), 277–282. <https://doi.org/10.1038/hdy.2012.83>
- Mishra, S. K., Chen, B., Zhu, Q., Xu, Z., Ning, C., Yin, H., ... Li, D. (2020). Transcriptome analysis reveals differentially expressed genes associated with high rates of egg production in chicken hypothalamic-pituitary-ovarian axis. *Scientific Reports*, 10(1), 5976. <https://doi.org/10.1038/s41598-020-62886-z>
- Nadaf, J., Berri, C., Dunn, I., Godet, E., Le Bihan-Duval, E., & De Koning, D. J. (2014). An Expression QTL of Closely Linked Candidate Genes Affects pH of Meat in Chickens. *Genetics*, 196(3), 867–874. <https://doi.org/10.1534/genetics.113.160440>
- Nadaf, Javad, Gilbert, H., Pitel, F., Berri, C. M., Feve, K., Beaumont, C., ... Le Bihan-Duval, E. (2007). Identification of QTL controlling meat quality traits in an F2 cross between two chicken lines selected for either low or high growth rate. *BMC Genomics*, 8(1), 155. <https://doi.org/10.1186/1471-2164-8-155>
- Persson, M. E., Wright, D., Roth, L. S. V, Batakis, P., & Jensen, P. (2016). Genomic regions associated with interspecies communication in dogs contain genes related to human social disorders. *Scientific Reports*, 6. <https://doi.org/10.1038/srep33439>

- Suzuki, G., Harper, K. M., Hiramoto, T., Funke, B., Lee, M., Kang, G., ... Hiroi, N. (2009). Over-expression of a human chromosome 22q11.2 segment including TXNRD2, COMT and ARVCF developmentally affects incentive learning and working memory in mice. *Human Molecular Genetics*, 18(20), 3914–3925. <https://doi.org/10.1093/hmg/ddp334>
- Thévenin, A., Ein-Dor, L., Ozery-Flato, M., & Shamir, R. (2014). Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. *Nucleic Acids Research*, 42(15), 9854. <https://doi.org/10.1093/NAR/GKU667>
- Tixier-Boichard, M., Bed'hom, B., & Rognon, X. (2011). Chicken domestication: From archeology to genomics. *Comptes Rendus Biologies*, 334(3), 197–204. <https://doi.org/10.1016/J.CRVI.2010.12.012>
- Tsubakihara, Y., Hikita, A., Yamamoto, S., Matsushita, S., Matsushita, N., Oshima, Y., ... Imamura, T. (2015). Arkadia enhances BMP signalling through ubiquitylation and degradation of Smad6. *Journal of Biochemistry*, 158(1), 61–71. <https://doi.org/10.1093/jb/mvv024>
- Wilkins, A. S., Wrangham, R. W., & Fitch, W. T. (2014). The “Domestication Syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, 197(3), 795–808. <https://doi.org/10.1534/genetics.114.165423>
- Williams, E. J. B., & Bowles, D. J. (2004). Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Research*, 14(6), 1060–1067. <https://doi.org/10.1101/gr.2131104>
- Woo, Y. H., Walker, M., & Churchill, G. A. (2010). Coordinated expression domains in mammalian genomes. *PloS One*, 5(8), e12158. <https://doi.org/10.1371/journal.pone.0012158>
- Xu, X., Norell, M. A., Kuang, X., Wang, X., Zhao, Q., & Jia, C. (2004). Basal tyrannosauroids from China and evidence for protofeathers in tyrannosauroids. *Nature*, 431(7009), 680–684. <https://doi.org/10.1038/nature02855>

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-165249>

FACULTY OF SCIENCE AND ENGINEERING

Linköping Studies in Science and Technology, Dissertation No. 2073, 2020
Department of Physics, Chemistry and Biology

Linköping University
SE-581 83 Linköping, Sweden

www.liu.se