Studying Phenotypic Evolution in Domestic Animals: A Walk in the Footsteps of Charles Darwin

L. ANDERSSON

*Department of Medical Biochemistry and Microbiology, Uppsala University,
SE-75124 Uppsala, Sweden, and Department of Animal Breeding and Genetics,
Swedish University of Agricultural Sciences, SE-75123 Uppsala, Sweden*

Correspondence: leif.andersson@imbim.uu.se

Charles Darwin used domesticated plants and animals as proof of principle for his theory on phenotypic evolution by means of natural selection. Inspired by Darwin's work, we developed an intercross between the wild boar and domestic pigs to study the genetic basis for phenotypic changes during domestication. The difference in coat color is controlled by two major loci. Dominant white color is due to two consecutive mutations in the *KIT* gene: a 450-kb duplication and a splice mutation. Black spotting is caused by the combined effect of two mutations in *MC1R*: a missense mutation for dominant black color and a 2-bp insertion leading to a frameshift. A major discovery made using this pedigree is the identification of a single-nucleotide substitution in intron 3 of the gene for insulin-like growth factor 2 (*IGF2*) that is underlying a quantitative trait locus affecting muscle growth, size of the heart, and fat deposition. The mutation disrupts the interaction with a repressor and leads to threefold increased *IGF2* expression in postnatal muscle. In a recent study, we have identified the *IGF2* repressor, and this previously unknown protein, named ZBED6, is specific for placental mammals and derived from a domesticated DNA transposon.

Charles Darwin was the first to realize that phenotypic changes in domesticated animals and plants, caused by selective breeding, mimic the process of phenotypic evolution in natural populations (Darwin 1859). In fact, Darwin himself became an animal breeder and performed breeding experiments with pigeons to prove that the phenotype could be altered by selection. Nine years after the publication of *On the origins of species*, Darwin summarized his studies of domesticated species in the book *The Variation of Animals and Plants Under Domestication* (Darwin 1868). The powerful genomic tools now available allow us to further use domestic animals as models by revealing the genes and mutations that have contributed to their phenotypic evolution (Andersson 2001). Some of these mutations predate domestication (standing genetic variation), whereas others have occurred subsequent to domestication and were picked up by human selection. This screen for mutations with phenotypic effects is extremely deep; it has been going on worldwide for thousands of years and has led to genetic adaptations to various climates and production systems, as we have taken our domestic animals around the globe. These genetic changes are the reason why domesticated plants and animals are outstanding models for phenotypic evolution by means of natural selection.

In 1989, inspired by Charles Darwin's studies of phenotypic differences between wild and domesticated species, we generated an intercross between the European wild boar and domestic pigs, with the objective of using the emerging genomic tools in an attempt to reveal some of the genes that have been under selection during pig domestication. The aim of this chapter is to summarize some of the highlights from this project and their implications for evolutionary biology.

We crossed two European wild boars with eight Swedish Yorkshire (Large White) sows and generated 200 F₂ progeny. The pedigree was established with two major aims: as a resource population for developing a linkage map for the pig and for mapping trait loci of biological interest. When the project was initiated, the porcine linkage map was restricted to ~20 loci, but it grew rapidly as microsatellite markers became available for the pig. This pedigree contributed to the development of the first-generation genome-wide linkage map for the pig (Ellegren et al. 1994). We recorded a large number of phenotypes that differed markedly between wild boars and domestic pigs including coat color, growth rate, body composition, immunological traits, and some skeletal measures. Most of these traits have a complex genetic background, and we considered this a pilot experiment because it was not known whether it would be possible to detect any convincing quantitative trait loci (QTL) using a sample size of only 200 F₂ animals derived from an intercross between outbred populations.

In one of the volumes of *The Variation of Animals and Plants Under Domestication* (Darwin 1868) is an illustration of a wild boar in comparison with a pig of the Large White breed, i.e., the same breed group as used in our experiment (Fig. 1). The Large White pig was white at that time and it still is today, but there is a very striking difference in body composition because the 1850 version is fat, whereas the modern Large White pig is very lean (high muscle content, low fat content). During the 19th century, the breeding goal was to produce fat pigs because there was a high consumer demand for energy-rich food. Due to the change in our life style, the breeding goal has changed, and during the last century, there has been an intense selection to increase muscle growth and reduce

Figure 1. Illustration of a wild boar (top) and a Yorkshire Large White pig from Charles Darwin's book *The Variation of Animals and Plants Under Domestication* (Darwin 1868).

fat deposition, because consumers pay much more for meat than for lard. Our wild boar project has led to the identification of the mutations that cause the white color, a selection that primarily took place before 1850, and a mutation that has contributed very significantly to the transformation of a rather fat pig to a lean one.

SELECTION AGAINST CAMOUFLAGE ON THE FARM

There is a striking difference in coat color between the wild boar and Large White pigs and we observed a considerable amount of coat color diversity in the F_2 generation (Fig. 2). The analysis of the segregation data indicated that this diversity was controlled by two major loci: the *Dominant white* (*I*) and the *Extension* (*E*). Large White pigs were assumed to be homozygous for the dominant allele (*I*) for white color and the recessive allele (*Ep*) for black spotting, whereas wild boar were homozygous for the recessive *i* allele and the dominant *E+* allele. The segregation data for the F_2 generation did not deviate significantly from the expected 12:3:1 ratio of white (*I*⁺), wild-type (*i/i*, *E*⁺/–), and black-spotted (*i/i*, *Ep/Ep*) pigs (Johansson et al. 1992). However, a surprising finding was the observation of a partially white phenotype that we denoted Patch and that was found to be controlled by a third allele at the *Dominant white* locus, *Ip* (Fig. 2). We mapped the *Dominant white* locus to pig chromosome 8

and identified *KIT* as a strong positional candidate gene (Johansson et al. 1992). Previous studies had demonstrated that *KIT* encodes a tyrosine kinase receptor with an essential function for the migration and survival of melanoblasts and that *KIT* mutations cause pigmentation disorders in both mice (Dominant white spotting) and humans (Piebald) (Spritz 1994).

Further characterization of the *Dominant white* locus over several years revealed that the *Dominant white* allele is caused by the combined effect of two independent mutations: (1) a ~450-kb tandem duplication that encompasses the entire coding sequence of *KIT* and ~150 kb upstream of exon 1 and (2) a splice mutation at the first nucleotide of intron 17 in one of the *KIT* copies that leads to exon skipping (Johansson Moller et al. 1996; Marklund et al. 1998; Giuffra et al. 2002). The *Patch* allele possesses the duplication but not the splice mutation. The data imply an evolutionary scenario whereby the duplication first occurred and resulted in a white-spotted phenotype that was selected by humans. The splice mutation occurred subsequently and resulted in a completely white phenotype but with normally colored eyes. The explanation of why *KIT* mutants have normally colored eyes is that *KIT* has an essential function for the development of neural-crest-derived melanocytes found in hair, skin, and ear but not for the melanocytes forming the *retinal pigment epithelium* (*RPE*). Our hypothesis is that the duplication is causing white spotting because the duplicated *KIT* copy has been separated from regulatory elements located far upstream of *KIT* (Giuffra et al. 2002), and the splice mutation enhances the defect in *KIT* signaling. The skipping of exon 17 in the mature transcript removes a crucial part of the tyrosine kinase domain, and thus, one of the *KIT* copies is expected to code for a receptor with normal ligand binding but defective kinase signaling. An allele carrying only the splice mutation but not the duplication is expected to be homozygous lethal because *KIT* signaling is absolutely required for hematopoiesis. The presence of one normal *KIT* copy ensures that white pigs have a sufficient amount of *KIT* signaling to avoid severe pleiotropic effects on hematopoiesis and germ-cell development. Thus, by combining the effect of two mutations—the duplication and the splice mutation—a *KIT* allele has been “created” with more severe effect on pigmentation than any known mouse mutant, despite the fact that the allele is fully viable in the homozygous condition, whereas *KIT*-null alleles are homozygous lethal.

In his work, Charles Darwin discussed the observation that domestic animals may revert toward a wild-type phenotype “when (they) run wild” (Darwin 1859). One important explanation for this phenomenon is the altered selection pressure animals experience in a natural environment. However, the *Dominant white* allele may revert to the wild-type allele due to unequal crossing-over between the two copies of the *KIT* duplication. We revealed extensive genetic diversity at the *KIT* locus in populations of white pigs, where the copy number varied from one to three, and one or two of the copies carried the splice mutation (Pielberg et al. 2002, 2003). This genetic instability explains why breeders have never been able to

Figure 2. Segregation of coat color among F_2 animals from an intercross between European wild boar and Large White domestic pigs. Two alleles are segregating at the *Extension/MC1R* locus: the wild-type allele ($E+$) and the recessive allele for black spotting (Ep). Three alleles are segregating at the *Dominant White/KIT* locus: the recessive wild-type allele and the alleles for Patch (Ip) and Dominant white color (I). Photo by Mats Gerentz, Swedish University of Agricultural Sciences.

fix the dominant white phenotype despite it being a monogenic trait, and they have been selecting for white color for hundreds of years. Furthermore, the genetic instability implies that if white pigs were released in the wild and became feral pigs, they would soon revert to wild type due to unequal crossing-over combined with selection for sun protection and camouflage in the wild.

The second major locus explaining the difference in coat color between wild and Large White domestic pigs is the *Extension (E)* locus (Fig. 2). The *E* locus is one of the classical coat color loci in mammals, and alleles at this locus determine the relative distribution of red pheomelanin and black eumelanin. The *E* locus encodes the melanocortin 1 receptor (*MC1R*) (Robbins et al. 1993). The wild boar carries the wild-type allele $E+$, and some hairs show the classical agouti pattern with alternating black and red stripes created by the interaction among *MC1R*, its ligand melanocyte-stimulating hormone (*MSH*), and the antagonist agouti-signaling protein (*ASIP*). The Ep allele for black spotting involves two causative mutations that must have occurred consecutively (Kijas et al. 1998, 2001). First, there is a missense mutation D124N causing dominant black color found, for instance, in Hampshire pigs. This mutation is assumed to cause a constitutively active receptor that leads to the production of black pigment. Second, the Ep allele also carries a 2-bp (CC) insertion at codon 22 that expands a stretch of six cytosine nucleotides to eight. This causes a frameshift and an expected complete loss of function. The predicted phenotype of an *MC1R*-null animal is a uniform red coat color. Some Ep/Ep homozygotes are uniformly red, but most of them are red with black spots or white with black spots (Fig. 2). So why do most of these pigs exhibit black spots that is inconsistent with a complete absence of *MC1R* signaling? The explanation is that the stretch of eight C nucleotides behaves as a small microsatellite that may be affected by slippage during DNA replication in somatic cells and possibly also in

germ cells. Reverse transcriptase–polymerase chain reaction (RT-PCR) analysis using skin from black spots revealed that one copy of *MC1R* had reverted to six Cs and thereby restored the reading frame (Kijas et al. 2001). The consequence is that the missense mutation D124N becomes activated and the pigment cells produce black eumelanin.

We have not yet been able to find the genetic basis of why some Ep/Ep pigs are red with small black spots and others are white with large black spots (Fig. 2). The numbers of these two classes of pigs did not fit any Mendelian ratio, suggesting that one of the founder populations (most likely the domestic pigs) are not fixed at this locus or that the phenotype has a more complex genetic background. Interestingly, there is a very clear trend that the black spots are larger on a white background than on a red background (Fig. 2).

After describing the alleles for dominant black color (ED), black spotting (Ep), and recessive red (e) color (Kijas et al. 1998, 2001), we decided to screen for *MC1R* diversity among pigs representing 51 different Asian and European breeds of domestic pigs, as well as populations of wild boars from both Europe and China (Fang et al. 2009). The results revealed contrasting modes of evolution at the *MC1R* locus in wild and domestic pigs (Fig. 3). We have previously estimated that Asian and European wild boars diverged from a common ancestor more than 500,000 years before present on the basis of the sequence divergence of mitochondrial DNA (Kijas and Andersson 2001). We found a total of seven nucleotide substitutions in *MC1R* among European and Asian wild boars, and they were all synonymous, providing evidence for purifying selection maintaining normal *MC1R* function (Fig. 3). The most likely selection pressure is to maintain a camouflage color. In sharp contrast, we discovered a total of nine genetic changes unique to domestic pigs, and they all changed the coding sequence; eight were nonsynonymous substitutions and one was the frameshift mutation

Figure 3. Median-joining network tree of *MC1R* alleles in wild and domestic pigs from Europe and Asia. All known alleles are represented by circles and a four-digit allele designation, whereas squares represent predicted intermediate forms that have not yet been found. Thin black lines perpendicular to lines connecting alleles represent synonymous changes, whereas thicker red lines represent nonsynonymous change. Colors inside circles and squares represent observed and predicted phenotypes, respectively; question marks inside two of the squares indicate that the associated phenotype cannot be predicted because they are intermediate forms between alleles differing by two nonsynonymous substitutions. The asterisk associated with the synonymous substitutions leading to alleles 0203 and 0104 indicates the only instance of an identical mutation at different locations on the tree. (Reprinted, from Fang et al. 2009.)

described above. This highly significant excess of substitutions changing the coding sequence demonstrates that the change in coat color in domestic pigs is caused by a direct selection to alter coat color, rather than relaxed purifying selection. This provides evidence for selection against camouflage in domestic pigs, and this selection on color must have been initiated early during domestication. We have hypothesized that humans have selected on color because it distinguished the early domestic forms from their wild ancestors. Alternatively, it facilitated animal husbandry, because it is easier to keep track of a colorful pig than a camouflage-colored pig and/or simply because we were attracted by new color variants as they arose due to spontaneous mutations. Selection for black color in pigs happened independently in Asia and Europe (Fig. 3) and involved two different missense mutations (L102P and D124N) with very similar phenotypic effects (Kijas et al. 1998). The *MC1R* locus is a beautiful illustration of why domestic animals are such good models for phenotypic evolution; humans have transformed their phenotype during the last 10,000 years by cherry-picking mutations with phenotypic effects.

Both *KIT* and *MC1R* in pigs illustrate that our domestic animals have a sufficiently old history that allows the evolution of alleles differing by multiple causative mutations. We have described several additional such examples. For instance, the *Smoky* allele affecting plumage color in the chicken involves both a 9-bp insertion in the *PMEL17* gene, causing *Dominant white* color, and a 12-bp deletion in the same gene that partially restores pigmentation (Kerje et al. 2004). Similarly, alleles causing different degrees of white spotting in dogs (*Irish*, *Piebald*, and *Extreme white*) are caused by different combinations of regulatory mutations at the *MITF* locus (Karlsson et al. 2007). This is an important

lesson for studies of genotype/phenotype relationships in natural populations, including humans, because it suggests that we may often find alleles that differ by multiple functionally important mutations, in contrast to mouse mutants or monogenic disorders in humans where the paradigm has been that a phenotype or disorder almost always is caused by a single causative mutation.

GENETIC DISSECTION OF A MAJOR LOCUS CONTROLLING MUSCLE GROWTH, HEART SIZE, AND FAT DEPOSITION

The intercross between the European wild boar and Large White domestic pigs was designed to allow for the mapping of QTL controlling some of the multifactorial traits that show striking differences between the founder populations. Domestic pigs are today very lean due to the strong selection for high muscle growth and reduced fat deposition. A striking observation was that the F_1 and F_2 hybrid progeny on average became more obese than purebred domestic pigs because the wild boar founders transmitted alleles for high fat deposition that most certainly are adaptive for the survival of wild boars during periods of sparse food resources under natural conditions. This is in line with the “thrifty gene hypothesis,” which implies that alleles associated with an increased risk to develop obesity and other metabolic disorders in humans have a relatively high frequency in many populations because they were advantageous during periods of starvation (Neel 1962).

In 1994, we published the first QTL mapping paper based on the wild boar/Large White intercross (Andersson et al. 1994). This study was the first genome-wide QTL mapping study in an outbred organism, because previously published QTL mapping studies were based on intercrosses among inbred lines (see, e.g., Paterson et al. 1988). The study revealed several QTL reaching genome-wide significance, and the most prominent one was detected on chromosome 4. For each trait, the estimated QTL effects were in the expected direction, so that the QTL allele from the domestic pig increased growth, reduced fat deposition, and increased the length of the small intestine (Andersson et al. 1994). Interestingly, Charles Darwin in his book *The Variation of Animals and Plants Under Domestication* noted that one of the phenotypic changes that happened during pig domestication is an increased length of the small intestine (Darwin 1868). The assumption is that this phenotypic change is related to selection for higher feed efficiency. Subsequent studies have shown that the major QTL on pig chromosome 4 affecting multiple traits represents multiple linked loci (Berg et al. 2006).

In 1999, we (Jeon et al. 1999) and Michel Georges’ group (Nezer et al. 1999) identified a major QTL located at the distal tip of pig chromosome 2. The QTL was shown to have major effects on muscle growth, heart size, and subcutaneous fat deposition but had no significant effect on birth weight, adult body weight, or abdominal fat deposition. The QTL allele derived from the domestic Large White pig in our intercross made the pig leaner (more muscle and less fat), and this locus alone controlled ~30% of the residual phenotypic variance for muscle traits among

the F_2 animals (Jeon et al. 1999). This QTL showed paternal expression, i.e., the phenotypic effects observed in the F_2 animals were determined by the paternal allele they had received from their F_1 sire. This immediately revealed the gene coding for insulin-like growth factor 2 (*IGF2*) as the prime positional candidate gene, because this is one of the very few paternally expressed genes located in this chromosomal region and it is a well-known growth factor. We could immediately conclude that if *IGF2* was the gene, then the causative mutation must be a regulatory one because there was no difference in *IGF2* protein sequence between wild and domestic pigs. Without access to a genome assembly, we decided to sequence a porcine bacterial artificial chromosome (BAC) comprising the *IGF2* locus, and the results revealed an extensive number of evolutionarily conserved noncoding sequences consistent with the fact that *IGF2* has four different promoters and a complex regulation (Amarger et al. 2002).

In collaboration with Michel Georges' group in Liege, we collected a set of chromosomes for which we could establish the genotype at the *IGF2* QTL with confidence based on family segregation data. We named the allele associated with high muscle growth Q and the wild-type allele q . We then sequenced 28.6 kb of genomic DNA from five different Q haplotypes representing four different breeds and 10 q haplotypes derived from wild boars and domestic pigs. The region comprised the insulin (*INS*) and *IGF2* genes and their upstream regions. The resequencing data revealed that all five Q haplotypes were identical for a 20-kb region spanning from intron 1 to the 3' untranslated region (3'UTR) of *IGF2*, whereas extensive sequence diversity existed in this region between Q and q haplotypes as well as among q haplotypes (Van Laere et al. 2003). This implied that the five Q haplotypes are identical by descent (IBD) for this region and that the causative mutation(s) should be located within the 20-kb region because the sequence identity among Q chromosomes broke up outside the region.

The problem we were facing at this stage was the extensive sequence diversity, approaching 1% (!), between Q and q chromosomes, which meant that although we had achieved an exceedingly high resolution in the QTL mapping, ~150 sequence polymorphisms were still showing complete linkage disequilibrium with the two QTL alleles. At this stage, we considered the possibility that the Q hap-

lotype had an Asian origin, which would explain the large sequence diversity to the q chromosomes. In an earlier study, we had shown that Asian pigs were crossed with European pigs primarily during the 18th and 19th centuries and that many breeds of European pigs have a hybrid origin (Giuffra et al. 2000). For this reason, we incorporated into the sequence analysis a Chinese Meishan chromosome carried by one of the founder animals of a Meishan/Large White intercross developed by Alan Archibald and Chris Haley at the Roslin Institute. Segregation data proved that this chromosome should be classified as a q chromosome. Resequencing the Meishan chromosome revealed that it was identical to the Q chromosomes in the critical 20-kb region, but with one important exception: At nucleotide 3072 in *IGF2* intron 3, the Q chromosomes had an A nucleotide, whereas the Meishan chromosome as well as all other q chromosomes had a G nucleotide (Van Laere et al. 2003). A further examination showed that the G3072A substitution occurs in a CpG island that is evolutionarily conserved among placental mammals, and seven other mammals, including humans, have a G at the mutated site (Fig. 4). We concluded that this nucleotide substitution must be the quantitative trait nucleotide (QTN) causing the *IGF2* QTL in pigs (Van Laere et al. 2003).

The functional characterization of the QTN showed that it constitutes a *cis*-acting regulatory mutation that up-regulates postnatal *IGF2* expression in skeletal and cardiac muscle but not in liver (Van Laere et al. 2003). First, we showed by bisulfide sequencing that the mutation does not alter the DNA methylation pattern. Second, we used an electrophoretic mobility-shift assay (EMSA) to show that the mutation disrupts the interaction with an unknown nuclear factor and that this protein only binds the DNA sequence when it is unmethylated. Third, we used northern blot and real-time PCR analysis to show tissue-specific up-regulation of *IGF2* mRNA expression in postnatal muscle. Interestingly, the mutation had no significant effect on *IGF2* expression in prenatal muscle or postnatal liver. The latter finding was in perfect agreement with the observation that the serum *IGF2* level was unaltered between genotypes, because liver is the major source of circulating *IGF2*. Finally, transfection of a luciferase construct including the porcine QTN region and the porcine P3 promoter into mouse C2C12 myoblasts demonstrated that the unknown factor acts as a repressor at the *IGF2* locus.

Figure 4. Alignment of the nucleotide sequence for the *IGF2* QTN region in pig intron 3 with the corresponding sequences; from seven other placental mammals. The pig wild-type alleles (Pig- q) are used as the master sequence; a dot represents sequence identity to the master sequence. Pig- Q represents the mutant sequence. Red letters highlight the CG dinucleotides. A short palindrome in the near vicinity of the QTN is underlined. (Modified from Van Laere et al. 2003 [©Nature Publishing Group].)

The initial characterization of the *IGF2* QTN showed that this mutation affects the expression from the *IGF2* P2, P3, and P4 promoters (Van Laere et al. 2003). A subsequent study revealed that the mutation has a similar effect on the *IGF2* antisense transcript (Braunschweig et al. 2004). Thus, the binding of the repressor affects transcription from four promoters spread over 4 kb in the porcine genome. The highly significant effect of the *IGF2* QTN on several phenotypic traits has been confirmed in several subsequent studies (Jungerius et al. 2004; Estellé et al. 2005; Oczkowicz et al. 2009).

The very specific interaction between the wild-type sequence at the *IGF2* QTN and the unknown nuclear factor implied that it should be possible to fish out the factor using a biotin-labeled nucleotide and then determine peptide sequences by mass spectrometry. We have recently achieved this by labeling nuclear proteins from mouse C2C12 cells using the sensitive SILAC (stable isotope labeling of amino acids in culture) method (Markljung et al. 2009). Mass spectrometry analysis combined with database searches against the mouse proteome revealed that the peptides enriched using the wild-type *q* oligonucleotide but not with the mutant *Q* oligonucleotide were encoded by an open reading frame (ORF) located in intron 1 of the *Zc3h11a* gene that encodes a poorly characterized zinc finger transcription factor. Further bioinformatic analysis revealed that this ORF encodes a distinct protein comprising more than 900 amino acids with no sequence similarity to ZC3H11A. We have named this protein ZBED6 because it is the sixth mammalian protein containing the BED zinc finger domain. Furthermore, the *ZBED6* coding sequence represents a domesticated DNA transposon belonging to the hAC family that contains many active DNA transposons in fruit flies, maize, and the house fly. *ZBED6* is unique to placental mammals, but remnants of this gene are also found in platypus and opossum; however, the coding sequence is disrupted and not conserved. In contrast, the *ZBED6* coding sequence is highly conserved among all placental mammals sequenced to date. This implies that the integration occurred in a primitive mammal before the divergence of monotremes and other mammals and that it evolved an essential function in a primordial placental mammal after the split between marsupials and placental mammals but before the radiation of different families of placental mammals. Thus, *ZBED6* function is an invention shared by all placental mammals, and it may have contributed to their evolution.

The functional characterization of ZBED6 demonstrated that it is the bona fide repressor binding the QTN site in pig *IGF2* intron 3 (Markljung et al. 2009). An EMSA with recombinant ZBED6 protein confirmed the specific binding to the wild-type but not the mutant oligonucleotide, and a specific supershift of the EMSA complex was observed using an anti-ZBED6 antibody and nuclear extracts from mouse C2C12 cells. Expression analysis in mouse using northern blot analysis, real-time PCR analysis, and immunohistochemistry revealed that ZBED6 has a broad tissue distribution, both in adult animals and during development, indicating that it has a much wider function than regulating muscle growth. This was further supported by

the results of chromatin immunoprecipitation using our ZBED6 antibody and mouse C2C12 cells followed by new-generation sequencing (chromatin immunoprecipitation high-throughput sequencing [ChIP-seq]). This experiment revealed 2499 putative ZBED6 binding sites, with a minimum of 15 overlapping reads, which were considered statistically significant. The region corresponding to the pig QTN site in the mouse *Igf2* gene was one of the most highly enriched regions. A consensus binding motif of 5'-GCTCGC-3' was established based on all hits excluding *Igf2*, and this sequence is a perfect match to the pig wild-type sequence at the *IGF2* QTN, whereas the mutant pig sequence is 5'-GCTCAC-3' (Fig. 4). Thus, the results of the ChIP-seq analysis provided further support for ZBED6 being the repressor binding the pig QTN.

As many as 1200 genes in the mouse genome were associated with one or more putative ZBED6-binding sites located within 5 kb of the gene. This gene list was used to search for an enrichment of specific gene ontology (GO) classifications. This analysis revealed that the list of putative ZBED6 downstream targets was a highly nonrandom collection. Genes associated with development, regulation of biological processes, transcriptional regulation, and cell differentiation were highly enriched. More than 20% of the putative ZBED6 targets were other transcription factors. The results suggest that ZBED6 could be a master regulator of transcription in placental mammals.

The pig mutation at the QTN site in intron 3 of *IGF2* has opened up a door to start exploring the biological significance of ZBED6, and further research on this interesting transcription factor may lead to new knowledge concerning the evolution and development of placental mammals. Charles Darwin used domestic animals as a proof of principle for his theory on phenotypic evolution by means of natural selection. The IGF2/ZBED6 story now provides a proof of principle for how research on domestic animals can provide novel insight into basic biology.

ACKNOWLEDGMENTS

Sincere thanks are due to Jennifer Meadows for comments on the manuscript and for assistance in preparing the illustrations. The work summarized in this chapter has been supported by the Swedish Foundation for Strategic Research, the Knut och Alice Wallenbergs Foundation, and The Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning.

REFERENCES

- Amarger V, Nguyen M, Van Laere AS, Braunschweig M, Nezer C, Georges M, Andersson L. 2002. Comparative sequence analysis of the *Insulin-IGF2-H19* gene cluster in pigs. *Mamm Genome* **13**: 388–398.
- Andersson L. 2001. Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet* **2**: 130–138.
- Andersson L, Haley CS, Ellegren H, Knott SA, Johansson M, Andersson K, Andersson-Eklund L, Edfors-Lilja I, Fredholm M, Hansson I, et al. 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**: 1771–1774.
- Berg F, Stern S, Andersson K, Andersson L, Moller M. 2006.

- Refined localization of the FAT1 quantitative trait locus on pig chromosome 4 by marker-assisted backcrossing. *BMC Genet* **7**: 17.
- Braunschweig MH, Van Laere A-S, Buys N, Andersson L, Andersson G. 2004. IGF2 antisense transcript expression in porcine postnatal muscle is affected by a quantitative trait nucleotide in intron 3. *Genomics* **84**: 1021–1029.
- Darwin C. 1859. *On the origins of species by means of natural selection or the preservation of favoured races in the struggle for life*. Murray, London.
- Darwin C. 1868. *The variation of animals and plants under domestication*. Murray, London.
- Ellegren H, Chowdhary BP, Johansson M, Marklund L, Fredholm M, Gustavsson I, Andersson L. 1994. A primary linkage map of the porcine genome reveals a low rate of genetic recombination. *Genetics* **137**: 1089–1100.
- Estellé J, Mercadé A, Noguera JL, Pérez-Enciso M, Ovilo C, Sánchez A, Folch JM. 2005. Effect of the porcine IGF2-intron3-G3072A substitution in an outbred Large White population and in an Iberian × Landrace cross. *J Anim Sci* **83**: 2723–2728.
- Fang M, Larson G, Soares Ribeiro H, Li N, Andersson L. 2009. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet* **5**: e1000341.
- Giuffra E, Kijas JMH, Amarger V, Carlborg Ö, Jeon J-T, Andersson L. 2000. The origin of the domestic pig: Independent domestication and subsequent introgression. *Genetics* **154**: 1785–1791.
- Giuffra E, Törnsten A, Marklund S, Bongcam-Rudloff E, Chardon P, Kijas JMH, Anderson SI, Archibald AL, Andersson L. 2002. A large duplication associated with Dominant White color in pigs originated by homologous recombination between LINE elements flanking *KIT*. *Mamm Genome* **13**: 569–577.
- Jeon J-T, Carlborg Ö, Törnsten A, Giuffra E, Amarger V, Chardon P, Andersson-Eklund L, Andersson K, Hansson I, Lundström K, Andersson L. 1999. A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the *IGF2* locus. *Nat Genet* **21**: 157–158.
- 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 *PDGFRFA* on chromosome 8. *Genomics* **14**: 965–969.
- 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. *Mamm Genome* **7**: 822–830.
- Jungerius BJ, Van Laere A-S, te Pas MFW, van Oost BA, Andersson L, Groenen MAM. 2004. The IGF2-intron3-G3072A substitution explains a major imprinted QTL effect on backfat thickness in a Meishan × European white pig intercross. *Genet Res* **84**: 95–101.
- Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Andersson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ III, et al. 2007. Efficient mapping of Mendelian traits in dogs through genome-wide association analysis. *Nat Genet* **39**: 1321–1328.
- Kerje S, Sharma P, Gunnarsson U, Kim H, Bagchi S, Fredriksson R, Schütz K, Jensen P, von Heijne G, Okimoto R, Andersson L. 2004. The *Dominant white*, *Dun* and *Smoky* color variants in chicken are associated with insertion/deletion polymorphisms in the *PMEL17* gene. *Genetics* **168**: 1507–1518.
- Kijas JMH, Andersson L. 2001. A phylogenetic study of the origin of the domestic pig estimated from the near complete mtDNA genome. *J Mol Evol* **52**: 302–308.
- Kijas JMH, Wales R, Törnsten A, Chardon P, Moller M, Andersson L. 1998. Melanocortin receptor 1 (*MC1R*) mutations and coat color in pigs. *Genetics* **150**: 1177–1185.
- Kijas JMH, Moller M, Plastow G, Andersson L. 2001. A frameshift mutation in *MC1R* and a high frequency of somatic reversions cause black spotting in pigs. *Genetics* **158**: 779–785.
- Marklund S, Kijas J, Rodriguez-Martinez H, Ronnstrand L, Funa K, Moller M, Lange D, Edfors-Lilja I, Andersson L. 1998. Molecular basis for the dominant white phenotype in the domestic pig. *Genome Res* **8**: 826–833.
- Markljung E, Jiang L, Jaffe JD, Mikkelsen TS, Wallerman O, Larhammar M, Zhang X, Wang L, Saenz-Vash V, Gnirke A, et al. 2009. ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth. *PLoS Biol* **7**: e1000256.
- Neel JV. 1962. Diabetes mellitus: A “thrifty” genotype rendered detrimental by “progress”? *Am J Hum Genet* **14**: 353–362.
- Nezer C, Moreau L, Brouwers B, Coppieters W, Detilleux J, Hanset R, Karim L, Kvasz A, Leroy P, Georges M. 1999. An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nat Genet* **21**: 155–156.
- Oczkowicz M, Tyra M, Walinowicz K, Rózycki M, Rejdach B. 2009. Known mutation (A3072G) in intron 3 of the *IGF2* gene is associated with growth and carcass composition in Polish pig breeds. *J Appl Genet* **50**: 257–259.
- Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**: 721–726.
- Pielberg G, Olsson C, Syvänen A-C, Andersson L. 2002. Unexpectedly high allelic diversity at the *KIT* locus causing dominant white color in the domestic pig. *Genetics* **160**: 305–311.
- Pielberg G, Day AE, Plastow GS, Andersson L. 2003. A sensitive method for detecting variation in copy numbers of duplicated genes. *Genome Res* **13**: 2171–2177.
- Robbins LS, Nadeau JH, Johnson KR, Kelly MA, Roselli-Rehffuss L, Baack E, Mountjoy KG, Cone RD. 1993. Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* **72**: 827–834.
- Spritz RA. 1994. Molecular basis of human piebaldism. *J Invest Dermatol* **103**: 137S–140S.
- Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Andersson G, et al. 2003. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* **425**: 832–836.

Cold Spring Harbor Symposia on Quantitative Biology

Studying Phenotypic Evolution in Domestic Animals: A Walk in the Footsteps of Charles Darwin

L. Andersson

Cold Spring Harb Symp Quant Biol 2009 74: 319-325 originally published online April 7, 2010

Access the most recent version at doi:[10.1101/sqb.2009.74.039](https://doi.org/10.1101/sqb.2009.74.039)

References

This article cites 30 articles, 3 of which can be accessed free at:
<http://symposium.cshlp.org/content/74/319.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Cold Spring Harbor Symposia on Quantitative Biology* go to:
<http://symposium.cshlp.org/subscriptions>