G3: Genes|Genomes|Genetics Early Online, published on February 8, 2019 as doi:10.1534/g3.119.400038

Bidirectional selection for body weight on standing genetic variation in a chicken model

Mette Lillie[*1], Christa F. Honaker[†], Paul B. Siegel[†], Örjan Carlborg[*]

[*] Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

[†] Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

Corresponding author: Mette Lillie, mette.lillie@bioenv.gu.se

[1] Current affiliations: Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden and Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

Data Availability Statement:

Pooled genome data generated for this study is available from the Short Read Archive (SRA) BioProject: PRJNA516366; BioSample: SAMN10787895; Accessions: SRR8480632-SRR8480641 (Release date: 2019-04-01).

Supplementary Material (Figures S1, S2, S3 and S4 and Tables S1, and S2) has been deposited on figshare.

Short running title:

Sweeps in the Virginia body weight lines

Keywords:

Chicken, White Plymouth Rock, selective sweeps, body weight, quantitative trait

Corresponding author information:

Name: Mette Lillie

Email address: mette.lillie@bioenv.gu.se

Mailing address: University of Gothenburg

Dept for Biological and Environmental Sciences

Mette Lillie

PO Box 463

SE 405 30 Göteborg

Phone number: +46 (0)31-786 0000

ABSTRACT

Experimental populations of model organisms provide valuable opportunities to unravel the genomic impact of selection in a controlled system. The Virginia body weight chicken lines represent a unique resource to investigate signatures of selection in a system where long-term, single-trait, bidirectional selection has been carried out for more than 60 generations. At 55 generations of divergent selection, earlier analyses of pooled genome resequencing data from these lines revealed that 14.2% of the genome showed extreme differentiation between the selected lines, contained within 395 genomic regions. Here, we report more detailed analyses of these data exploring the regions displaying within- and between-line genomic signatures of the bidirectional selection applied in these lines. Despite the strict selection regime for opposite extremes in body weight, this did not result in opposite genomic signatures between the lines. The lines often displayed a duality of the sweep signatures, where an extended region of homozygosity in one line, in contrast to mosaic pattern of heterozygosity in the other line. These haplotype mosaics consisted of short, distinct haploblocks of variable between-line divergence, likely the results of a complex demographic history involving bottlenecks, introgressions and moderate inbreeding. We demonstrate this using the example of complex haplotype mosaicism in the *growth1* QTL. These mosaics represent the standing genetic variation available at the onset of selection in the founder population. Selection on standing genetic variation can thus result in different signatures depending on the intensity and direction of selection.

INTRODUCTION

Initial thinking on how adaptive processes shape the genome was modeled by Maynard Smith and Haigh (1974), who demonstrated that a beneficial mutation favored by natural selection will increase in frequency within a population. Linkage disequilibrium in the flanking region of a selected allele will result in a characteristic valley of diversity around the selected variant,

known as a hard selective sweep (Kaplan et al. 1989; Stephan et al. 1992). Aside from hard selective sweeps, selection on recessive variants, variants contributing to the standing genetic variation in a population, and partial sweeps (Hermisson and Pennings 2005; Teshima and Przeworski 2006) typically have weaker effects. Empirical studies, however, suggest that this type of selection is the most abundant mode of adaptation in recent evolution in both *Drosophila melanogaster* (Garud et al. 2015) and humans (Pritchard et al. 2010; Hernandez et al. 2011; Schrider and Kern 2017). Furthermore, polygenic adaptation describes how selection acts on standing genetic variation across the many loci contributing to a quantitative polygenic trait leading to a new phenotypic optimum by way of modest allele frequency changes across these loci (Pritchard and Di Rienzo 2010; Pritchard et al. 2010). Although this mode of adaptation would respond very rapidly to changes in the selective environment, it would not necessarily lead to fixation for any one variant (Pritchard et al. 2010).

Modes of adaptation are not mutually exclusive, and the genomic signature that results will be dependent upon factors such as the genetic architecture of the trait, standing genetic variation available within this architecture, effective sizes of standing haplotypes, and population demography. By exploiting population genetic signals, researchers are increasingly able to detect the underlying modes of selection, from initial sweep scans that identify valleys of low diversity resulting from hard sweeps, to various recent developments to detect and differentiate between soft and hard sweeps (Berg and Coop 2014; Garud et al. 2015; Schrider and Kern 2017). Whereas previous studies have attempted to uncover selection throughout human history (Coop et al. 2009; Schrider and Kern 2017), much can be learned from research with model organisms, such as selection in experimental populations of *Drosophila* (Burke et al. 2010) or mice (Chan et al. 2012). In particular, long-term selection experiments have well-defined population histories, likely have stronger selection signatures in the genome due to an isolation of the trait under selection, and allow breeding of crosses to test for adaptive trait associations to candidate sweeps. The Virginia body weight chicken lines,

whose history of long-term, single-trait, bi-directional selection from a common founder White Plymouth Rock population has been well-characterized, affords us the opportunity to dissect the genomic selective-sweep signatures of strongly selected loci (Figure 1). Founders for the Virginia lines were generated via crossing of seven partially inbred White Plymouth Rock lines, in effect constraining the standing diversity. From this base population in 1957, bidirectional selection for 8-week body weight was carried out to form the high (HWS) and low (LWS) Virginia body weight lines. The breeding regime structured to minimize inbreeding and minimize the stochastic fixation of alleles that could affect small breeding populations (Siegel 1962; Marquez et al. 2010). Relaxed sublines for both HWS and LWS were produced from selected generation 44 (high weight relaxed: HWR; low weight relaxed: LWR) (Dunnington et al. 2013). After 55 generations of divergent selection, a 15-fold difference in body weight exists between the lines (Jambui et al. 2017b). With its well-defined population history, bi-directional single trait selection regime, and well-defined polygenic architecture of the adaptive trait, the Virginia body weight lines represent an invaluable resource to investigate the genomic signatures of selection.

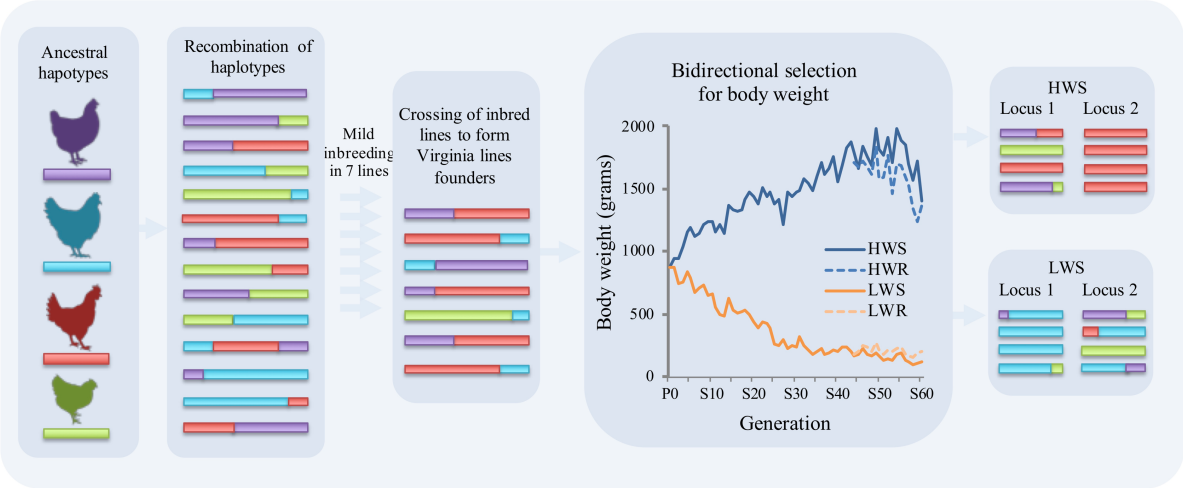


Figure 1. Representation of the formation of genomic signatures in the Virginia body weight lines. Ancestral haplotypes that contributed to the formation of the White Plymouth Rock breed would have recombined over time. Haplotype variation was likely constrained in the

111 seven partially inbred lines, which were crossed to form the founder population of the
112 Virginia body weight lines. The bidirectional change in eight-week body weight in the
113 Virginia body weight lines is presented in the graph from the parental (P0) across the selected
114 generations (S1-S60) for the high weight selected (HWS; blue unbroken line) and low weight
115 selected (LWS; orange unbroken line) body weight lines. Body weight within relaxed
116 sublines are also shown for the high weight relaxed line (HWR; blue broken line) and low
117 weight relaxed line (LWR; orange broken line) at corresponding selected generation.
118 Haplotype frequency and ancestral haplotype origin are depicted at two hypothetical loci to
119 the right. At locus 1, HWS continues to segregate with multiple haplotypes, whereas LWS is
120 fixed for an extended blue haplotype. At locus 2, HWS is fixed for an extended red haplotype,
121 whereas LWS continues to segregate with multiple haplotypes.
122
123 Previous research has demonstrated that standing genetic variations in many loci from the
124 founder population of the Virginia lines contribute to the observable difference in body
125 weight. Sweep scans using individual genotypes from a 60k SNP-chip have revealed
126 numerous regions of differentiation between the selected lines (Johansson et al. 2010;
127 Pettersson et al. 2013). These were explored for associations with body weight and a large
128 number of them were found to have contributed to selection response using available standing
129 genetic variation (Sheng et al. 2015). Using genotype data from these sweeps and earlier
130 identified QTL, together with phenotypic data from an F_{15} intercross between the lines, 20
131 independent associations to 8-week body weight have been confirmed (Zan et al. 2017).
132 Recently, pooled genome resequencing was applied to several generations from these lines to
133 investigate whether these data could reveal any additional features of the genomic impacts of
134 bidirectional selection on body weight to those previously found using the 60k SNP-chip data.
135 An initial evaluation of these data was integrated into a review of the knowledge gained from
136 multiple lines of inquiry in the Virginia body weight lines, with a particular focus on

identified QTL regions (Lillie et al. 2017a). Here, performed is a more in-depth analysis of this pooled genome resequencing dataset to characterize the genomic signatures of highly differentiated regions across the whole genome between the bidirectionally-selected lines: the putative selective sweeps. As earlier shown for the QTL regions (Lillie et al. 2017a), between-line differentiation was seldom due to complete fixation of different haplotypes in the lines. Instead, most selective sweeps resulted from extended runs of homozygosity in one line, contrasting to persistence of heterozygosity in the other. This duality of selection signatures and haplotype structures was illustrated by dissecting the complex differentiation in an earlier identified QTL on chromosome 1. Typically, this heterozygous region was comprised of multiple, distinct regions, with variable diversity, and between-line divergence. The directions of these relationships do not suggest any line bias, suggesting that the physiological plateau since generation 35 in the LWS due to a disruption of food-consumption and an inability to enter egg production at less than 1000 g (Siegel and Dunnington 1985; Siegel and Dunnington 1987; Jambui et al. 2017a) has not had a major influence on these patterns.

These empirical observations and our knowledge about the history of these populations imply that the duality of these signatures genome-wide reflect positive selection for one large-effect haplotype in one line while in the other line negative selection would remove this haplotype, allowing other haplotypes to continue to segregate. Mosaic haplotype structures within these segregating regions reflect standing variation in the founder population, likely resulting from ancestral haplotype recombination along a history of bottlenecking, inbreeding, and crossbreeding.

MATERIALS AND METHODS

Virginia body weight chicken lines

All animal procedures were carried out by experienced handlers and in accordance with the Virginia Tech Animal Care Committee animal use protocols (IACUC-15-136). The Virginia body weight lines were formed from a founder population resulting from crossing seven lines originating in 1949 that had undergone mild inbreeding. Established in 1957, bidirectional selection for body weight at 8 weeks of age was initiated to produce the closed selected lines: high weight selected (HWS) and low weight selected (LWS) (Siegel 1962). Breeding focused on a response to selection, while attempting to minimize inbreeding (Marquez et al. 2010). Effective population sizes in the LWS and HWS lines have been estimated as 38.3 and 32.1, respectively (Marquez et al. 2010). Relaxed sublines for both HWS and LWS were produced from selected generation 44, and are referred to as high weight relaxed (HWR) and low weight relaxed (LWR) (Dunnington et al. 2013). All generations were hatched in the same incubators and reared in the same pens on the same diet. Pooled semen was used to produce each generation of relaxed lines.

Sequencing and genome alignments

The sequencing data used in this dataset was originally reported in (Lillie et al. 2017a). In short, DNA for the genomic analyses was prepared from blood samples collected from 9-30 individuals from each line and pooled in equimolar ratios prior to library construction. Genome sequencing library construction and sequencing was carried out by SciLifeLab (Uppsala, Sweden) using two lanes on an Illumina Hiseq 2500. Reads were aligned to the *Gallus gallus* genome (Galgal5; INSDC Assembly GCA_000002315.3, Dec 2015) using BWA (Li and Durbin 2009). Genomes were sorted and duplicates were marked and removed with Picard (v1.92; http://picard.sourceforge.net). GATK (v3.3.0; McKenna et al. 2010) was used for realignment around indels. GATK UnifiedGenotyper was used to generate allele calls at all sites (option: emit all sites) and with ploidy = 30 (18 for LWS generation 50 as only 9 individuals went into this pool) to account for the pooled genome sample. Sites were

188 filtered to only include those with >10 and <100 reads, wherefrom allele frequency,
189 heterozygosity, and pairwise F_{ST} between all populations were calculated. Samtools (Li et al.
190 2009; v1.1; Li 2011) was used to generate mpileup files for PoPoolation2 (v1.201; Kofler et
191 al. 2011), which was used to calculate F_{ST} over 1000 bp sliding windows with 50% overlaps
192 between the population samples using the Karlsson et al. (2007) method, with minimum count
193 3, minimum coverage 10, maximum coverage 100, and minimum coverage fraction 1.
194 Genome alignments were visualized in IGV (v2.3.52; Robinson et al. 2011; Thorvaldsdottir et
195 al. 2013).
196
197 **Differentiated regions**
198 As reported in (Lillie et al. 2017a), differentiated regions were identified by employing an
199 empirical F_{ST} threshold of 0.953, representing the top 5% F_{ST} values in generation 55. Windows
200 with F_{ST} values above this threshold were clustered into differentiated regions when they were
201 less than 100 kb from one another. Clusters with less than 2 SNPs or less than 100 kb were
202 removed from the dataset to retain only the stronger candidate regions. Mean and median
203 heterozygosity were calculated for each line within each differentiated region. We used the
204 Variant Effect Predictor (VEP) (McLaren et al. 2016) available from Ensembl (Aken et al.
205 2017) to investigate potential functionality of candidate alleles. Haplotype structure within
206 regions of interest were visualized using adjusted allele frequencies (Lillie et al. 2017b) (similar
207 to the allele polarization step in the haplotype-block reconstruction approach used by Franssen
208 et al. 2017). This approach adjusts allele frequencies within the sequenced lines to the
209 generation of lowest haplotypic complexity, such that allele frequencies across all lines would
210 be adjusted to 1-AF, for sites where allele frequency > 0.5 in the generation of lowest haplotypic
211 complexity. In most cases, the generation of lowest haplotypic complexity contains one fixed
212 extended haplotype within the region of interest, with raw allele frequencies equal to ~0 or ~1,

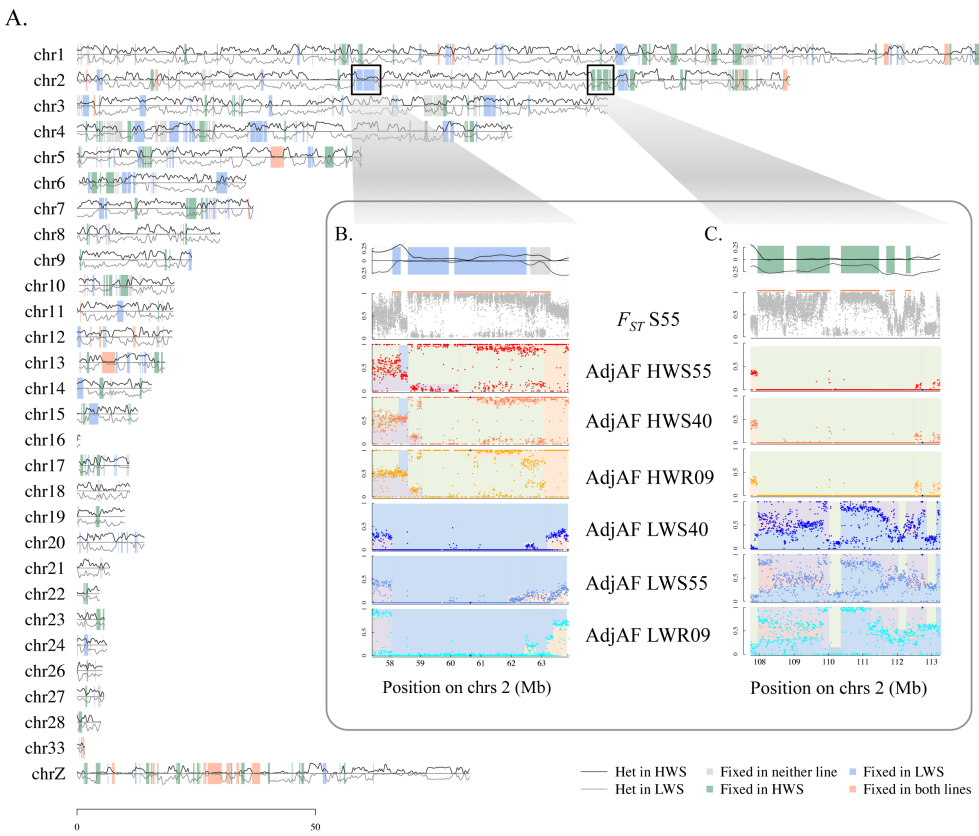
generating adjusted allele frequencies equal to ~0. These were then plotted using custom R scripts.

Data availability

Pooled genome data generated for this study are available via Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) under bioProject: PRJNA516366; bioSample: SAMN10787895; and accessions: SRR8480632-SRR8480641.

RESULTS

Alignment coverage of the reference genome after read alignment was between 91.22 % and 91.44 % across the sequenced pools. As reported earlier (Lillie et al. 2017a), 395 differentiated regions between the HWS and LWS lines in selected generation 55 were identified from the clusters of high differentiation ($F_{ST} > 0.953$; Figure 2). These regions covered a total of 174.5 Mb, or 14.2% of the genome, which is an increase from the 244 differentiated regions identified in selected generation 40 (99.6 Mb or 8.1% of the genome; Figure S1).



230

231 Figure 2. Heterozygosity across chromosomes of divergently selected Virginia body weight
232 chicken lines.

233 A. Heterozygosity in generation 55 for high weight selected (HWS; black line above x-axis)
234 and low weight selected (LWS; grey line below x-axis) presented across the chicken
235 chromosomes within the differentiated regions shaded using color code: grey where neither
236 line is fixed (); green where only HWS is fixed (); blue where only LWS is fixed within the
237 differentiated region; red where both lines are fixed.

238 B. F_{ST} and heterozygosity patterns across chromosome 2:58-63 Mb for selected generations
239 55 and 40 and relaxed generation 9.

240 C. F_{ST} and heterozygosity patterns across chromosome 2:108-113 Mb for selected generations
241 55 and 40 and relaxed generation 9.

242 Panels within B and C insets:

243 Panel 1: Detail of heterozygosity trace from chromosome map. Panel 2: Mean F_{ST} within 1 kb

windows between HWS and LWS at generation 55 indicated with grey points; region of differentiation indicated with the orange line above F_{ST} plot; allele frequencies of SNP markers with association to body weight indicated with blue diamonds. Panel 3-8: Mean adjusted allele frequency (adjAF) of 5kb windows in HWS generation 55 (HWS55) / HWS generation 40 (HWS40) / high weight relaxed generation 9 (HWR9) / LWS generation 55 (HLWS55) / LWS generation 40 (LWS40) / low weight relaxed generation 9 (LWR9). Shaded colors within these plots have been used to highlight the runs of adjusted allele frequencies that contribute to different haploblocks.

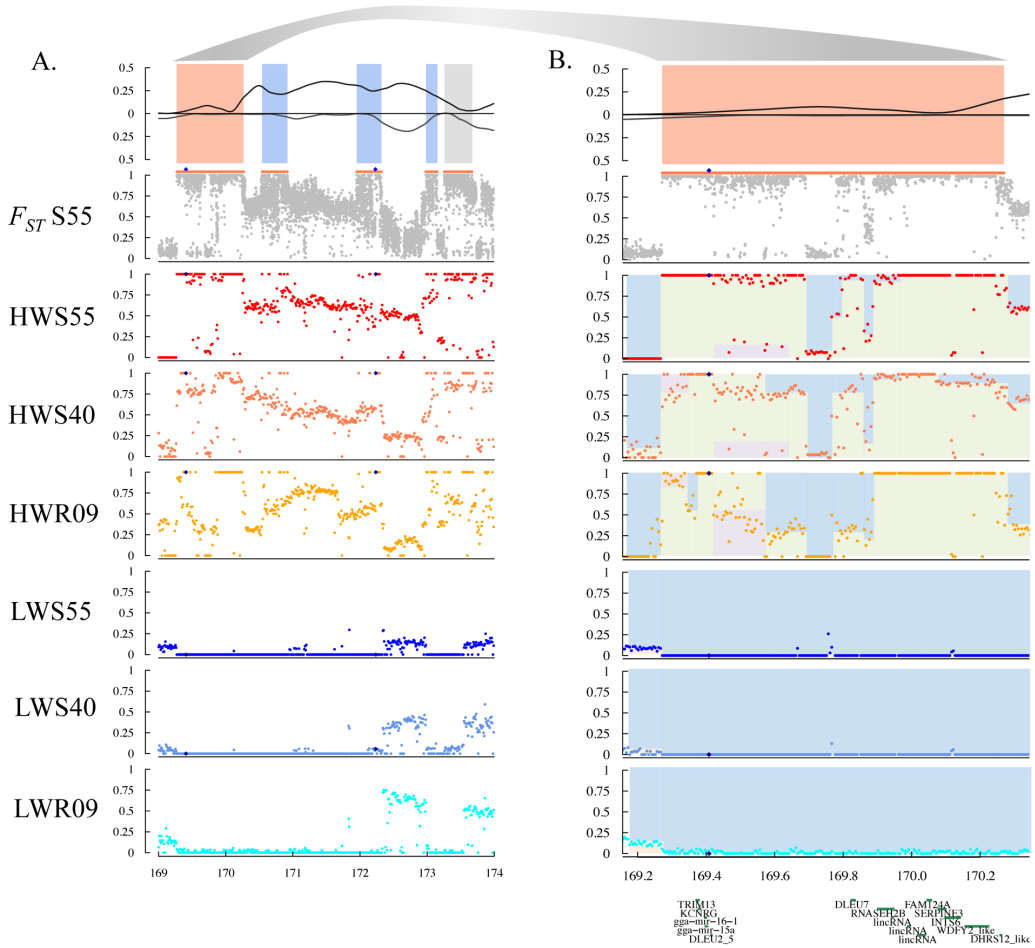
Candidate selective sweeps are often polymorphic in one of the selected lines

Compared to the genome-wide trend, there was a decline in heterozygosity at the extremes of high FST (S2 Figure). Few regions of differentiation showed fixation in both lines; rather, it was often the case that there was fixation across an extended region in one line, while many nucleotide positions in the region still segregated in the other (Figure 3; S3 Figure). Of the differentiated regions greater than 0.5 Mb in length, 34 regions (33 %) were close to fixation in HWS (mean heterozygosity <= 0.1) while LWS continued to segregate (mean heterozygosity > 0.1), 33 regions (32 %) were close to fixation in LWS (mean heterozygosity <= 0.1) while HWS continued to segregate (mean heterozygosity > 0.1), and 37 regions (35 %) were close to fixation for alternative haplotypes in both lines (mean heterozygosity <= 0.1 in both). This demonstrated that, while one extended haplotype was fixed in one line, multiple haplotypes continue to segregate in the other. Fixation was as common in HWS as LWS, a trend that extended to regions with confirmed associations for the selected trait, 8-week body weight (Figure 3; S3 Figure) (Zan et al. 2017).

Figure 3. Mean heterozygosity in HWS (above the x-axis; grey) and LWS (below the x-axis; black) at generation 55 within differentiated regions greater than 0.5 Mb in length. Differentiated regions overlapping known associations with 8-week body weight (Zan et al. 2017) are indicated in orange. Regions presented are sorted in order of increasing cumulative heterozygosity across both lines with increased heterozygosity in HWS.

Mosaic haplotypes in the *growth1* QTL

To illustrate the genomic mosaicism observed in many differentiated regions, we investigated the *growth1* QTL in detail (Jacobsson et al. 2005; Zan et al. 2017). A long region of divergence was observed between 169.3 Mb and 173.7 Mb on chromosome 1 (in total 4.4 Mb), where a single extended haplotype was close to fixation in LWS by generation 40, whereas the pattern of polymorphism in the HWS suggest that multiple haplotypes segregate in this line (Figure 4). Approximately 28% of the nucleotide positions within this region were highly divergent between HWS and LWS at generation 55. In total, 10,148 from 36,934 sites had differences between the lines in allele frequency that were greater than 0.9. This level of divergence implies that the long haplotype fixed in LWS is not present in the HWS at generation 55 because it has been selected out.



285

286 Figure 4. A. F_{ST} and heterozygosity patterns across chromosome 1:169-174 Mb for selected
287 generations 55 and 40 and relaxed generation 9 of the Virginia body weight chicken lines. B.
288 F_{ST} and heterozygosity patterns across chromosome 1:169.2-170.3 Mb for selected
289 generations 55 and 40 and relaxed generation 9 with gene positions. Panel 1: Heterozygosity.
290 Panel 2: Mean F_{ST} within 1 kb windows (grey points) in selected generation 55; defined
291 region of differentiation indicated with orange line above F_{ST} plot; blue diamonds indicate
292 allele frequencies of SNP markers with significant association to body weight from Zan et al.
293 (2017) . Panel 3-8: Mean adjusted allele frequency of 5kb windows in HWS55 (panel 3),
294 HWS40 (panel 4), HWR09 (panel 5), LWS55 (panel 6), LWS40 (panel 7), LWR09 (panel 8).
295 Colors within the plots are used to highlight the runs of adjusted allele frequencies that

296 contribute to different haploblocks.

297

298

299 We checked for the presence of a known insertion that has been previously associated with
300 increased body weight in other chicken populations (Jia et al. 2016). By examining the soft-
301 clipped reads (S4 Figure), we observed that this variant also segregated in our population
302 within the short region fixed for divergent haplotypes in both LWS and HWS (approximately
303 50 kb in length; GGA1: 169,370,000-169,420,000). Contrary to the expectation that this
304 insertion would dominate in the HWS line, it was instead present on the long haplotype fixed
305 by generation 40 in LWS, and entirely absent in HWS. Furthermore, this insertion is linked to
306 the LWS allele of SNP marker rs14916997, which was associated with low body weight (Zan
307 et al. 2017).
308 To identify candidate linkages and thus other potential functional variants, the *growth1* QTL
309 was evaluated for further sub-haplotype fixations. In the nearby region (GGA1: 169,950,000-
310 170,220,000), the HWS and LWS were also fixed by generation 55. Bounded by these two
311 highly differentiated regions are 15 annotated genes (S1 Table), with 8 missense mutations
312 (S2 Table). One missense mutation with predicted (SIFT) deleterious effect was present in
313 HWS, located within the sixth exon of Ribonuclease H2 Subunit B gene (RNASEH2B), as
314 well as the small (18 bp) deletion in LWS at GGA1: 169,407,811, which was absent from
315 HWS.

316

317 **DISCUSSION**

318 There was an overall increase (244 to 395) in the number of differentiated regions between
319 the HWS and LWS from generation 40 to generation 55. This increase is consistent with the
320 continuing phenotypic divergence between the lines resulting from the ongoing response to
321 selection in the HWS (average 56-day body-weights 1264 g and 1506 g, respectively)

contrasted to the plateau in the LWS at 142 g (Lillie et al. 2017a).

Candidate selective sweeps were often polymorphic in one of the selected lines

Within highly differentiated candidate selective sweep regions, we observed that an extended haplotype was often fixed in one line while multiple haplotypes continued to segregate in the other, thus resembling a duality in these sweep signatures. These patterns are likely to be shaped by the intensity of selection on the locus, as well as the haplotype frequencies that were present in the base population. For example, haplotypes with equal but opposite effects present at the same frequencies at the onset of selection would be expected to result in relatively equal lengths of fixation (homozygosity) in both lines, and thus a high F_{ST}. This pattern was observed only a few times (Figure 3), and thus appears to have been a relatively rare event. More likely is that the haplotypes in the founder population had different effect sizes and were present at different frequencies when selection was initiated to develop the high and low body weight chicken lines. The long runs on homozygosity found only in one line within regions of differentiation therefore likely represent the haplotypes with the largest effect, with the signature possibly being amplified by being at a low frequency at the onset of selection. These patterns suggest that the effect size of functional variants, as well as founding haplotype frequencies, are likely to have influenced the selective signature in the genome of the Virginia body weight chicken lines. The lack of line-bias in these signatures suggest that the selection plateau reached in the LWS at about generation 35, that physiologically is due to a disrupted food-consumption in some birds and an inability to enter egg production at less than 1000 g (Siegel and Dunnington 1985; Siegel and Dunnington 1987; Jambui et al. 2017a), has not made any major impact on these patterns.

A challenge in selective sweep studies is to discriminate which differentiated regions are likely due to drift or selection. The influence of drift on these lines has been investigated in a previous study, which showed that the population size was sufficiently large to prevent

genetic drift from overriding the effect of selection for the loci with the larger effects and that the probability of fixation for alternative haplotypes by drift was very low (Johansson et al 2010). Many regions of differentiation are present both within and outside of those known to be associated with body-weight (Sheng et al. 2015; Lillie et al. 2017a; Zan et al. 2017). Although it is known that the already associated regions only explain part of the variation in body-weight (Zan et al. 2017), the observation made here that fixation for alternative haplotypes is seldom the case even within regions of differentiation makes it difficult to quantify the expected influence of drift in shaping genomic signatures. By implementing a high F_{ST} cutoff, integrating association mapping results from previous studies in these lines, and comparing gene ontology and associations from other chicken populations, we have endeavored to build more confidence in differentiated regions than can be obtained separately using the approaches.

Mosaic haplotypes

Domestication of the chicken began roughly 7,000 years ago, predominately from red junglefowl, *Gallus gallus* (West and Zhou 1988; Sawai et al. 2010), but potentially also with contributions from *Gallus sonneratii* (grey) and *Gallus layafetii* (Sri Lankan) (Eriksson et al. 2008; Groeneveld et al. 2010; Tixier-Boichard et al. 2011). With subsequent distribution of chickens via human dispersal and trade, local breed formation would subject established chicken populations to genetic drift and selection to suit the environment and human-imposed selection for breed standards and production traits (Lyimo et al. 2014). Regardless of expectations that founder effects, bottlenecks, and selection would impact genetic diversity, nucleotide diversity and substitution rates were comparable for red jungle fowl and domestic chickens. This implies that domestication did not result overall in a substantial genome-wide loss of diversity in the species and had only a minor effect in an evolutionary context (Tixier-Boichard et al. 2011).

Thus, recombination of divergent haplotypes from European and Asian populations occurred while producing the WPR breed, forming the genetic substrate available for artificial selection for high and low body weight (Figure 1). During the several decades of WPR breeding that preceded the initiation of the Virginia body weight chicken lines, recombinant haplotypes are likely to have arisen, ultimately producing finely grained mosaics of the chromosomes entering the founders of the Virginia body weight lines. It is on this standing variation which the bi-directional selection experiment has acted (Figure 1). Short haplotype mosaics within the longer selective sweep regions were revealed when comparing the HWS and LWS genomes. The *growth1* QTL region on chromosome 1 is a clear example, illustrating the haplotype mosaicism of sweep signatures observed across the genome.

Mosaic haplotypes in the *growth1* QTL

The *growth1* QTL was first defined in a microsatellite-based analysis of an F_2 cross between the HWS and LWS, with a length of almost 110 cM (Jacobsson et al. 2005). This QTL was confirmed in SNP-based analyses of the F_2 (Wahlberg et al. 2009) and fine mapped in the F_2-F_8 (Besnier et al. 2011; Brandt et al. 2016) and the F_{15} generation of the Advanced Intercross Line (Zan et al. 2017). In this latest study, two associations were found within *growth1*, with the SNP rs14916997 (GGA1: 169,408,309) having the strongest association (Zan et al. 2017). From the pooled genome data, it is evident that a major haplotype, between 169.3 Mb and 173.7 Mb (in total 4.4 Mb), is close to fixation in LWS (Figure 4). This haplotype was already close to fixation by generation 40, possibly indicating that selection has been relatively strong, and the resulting pattern resembles a hard sweep in LWS. Contrastingly, multiple haplotypes segregate in HWS, with a mixture of nucleotide positions that are divergently fixed when compared to LWS, intermixed with positions that segregate for both LWS and unique HWS alleles.

399 We observed a high level of divergence between HWS and LWS within this region, implying
400 that the long haplotype fixed in LWS is not present in the HWS at generation 55 because it
401 was selected out of this line. Highly dissimilar regions are intermixed with short, but
402 continuous, stretches where the HWS and LWS haplotypes are nearly identical. Additionally,
403 the boundaries between these shared and differentiated regions are distinct. They likely
404 represent historical recombination events shared between one or more selected haplotypes,
405 rather than being multiple classic hard sweeps, as these are expected to result in a gradual
406 breakdown of population-wide linkage disequilibrium.
407 This interpretation is supported by other observations. First, for sharing of such interspersed
408 haplotypes to be possible between the lines, formative recombination events must have
409 occurred prior to the onset of bidirectional selection. Second, as several shared haplotype
410 segments were sometimes observed in the divergent regions, such multiple events must have
411 happened on the same haplotypes. Third, because the shared regions are often short (10s to a
412 few 100 kb), the events are unlikely in a population with as small a population-size as the
413 Virginia body weight chicken lines (effective population size, $N_e \sim 35$) (Marquez et al. 2010).
414 Finally, if the recombination events occurred during the selection experiment, selection on the
415 recombinant haplotypes must have been strong in order to only retain them in the lines. Such
416 is unlikely given the highly polygenic genetic architecture of 8-week body weight and the
417 dilution of selection pressure across the many loci. Therefore, these haplotype mosaics most
418 likely represent recombinant founder haplotypes resulting from their population history,
419 including the formation of the WPR as a breed. Although recombination cannot be
420 completely ruled out as a contributor to haplotype mosaicism, without individual-based
421 sequencing, we cannot confirm or eliminate a role for recombination events in HWS and
422 LWS.
423 Positive selection of this long haplotype to fixation in the LWS, coupled with negative
424 selection for its removal from HWS, may serve as an explanation for why previous studies

425 have seen a transgressive effect on 8-week body weight for SNP markers within this region.
426 Zan et al. (2017) reported that while the HWS allele at the rs14916997 SNP marker
427 (GGA1:169,408,309 bp) was associated with an increased 8-week body weight (additive
428 genetic effect approximately 26 grams), the HWS allele at the nearby SNP marker
429 rs316102705 (GGA1: 172,235,208 bp) was associated with a decrease in 8-week body weight
430 (additive genetic effect of approximately -7 grams).
431 Notably, this region on chicken chromosome GGA1 appears often in association studies
432 carried out in other populations, including comb traits (Shen et al. 2016), egg weight (Yi et al.
433 2015), feed intake (Yuan et al. 2015), abdominal fat percentage (Abasht and Lamont 2007;
434 Sheng et al. 2013), shank metrics (Sheng et al. 2013), and growth and body weight at
435 numerous life stages (Xie et al. 2012; Sheng et al. 2013; Abdalhag et al. 2015; Zhang et al.
436 2015). This may reflect a shared ancestral variant that has spread in domestic populations due
437 to its beneficial effect, or that this is a gene rich region associated with many functionally
438 important genes. Recently, an insertion (GGA1: 169,399,420) located upstream of the miR-
439 15a-16 precursor was strongly associated with growth traits in an Xinghua & White Recessive
440 Rock F_2 cross, where presence of this variant results in an altered hairpin formation, reduction
441 of miR-16 expression, and increased body weight, bone size, and muscle mass (Jia et al.
442 2016). Although this insertion was also present in multiple chicken breeds, and at high
443 frequencies in broiler breeds (Jia et al. 2016), in our population, the insertion was present in
444 LWS on the long, fixed haplotype, and was absence from the HWS, contrary to expectations
445 that the insertion would substantially increase body weight.
446 This observation may be explained by alternative hypotheses; i.e. that this insertion: i) has an
447 effect also in our lines but is linked to another polymorphism with a stronger opposite effect,
448 ii) does not have an effect in our lines due to genetic background, iii) does not have an effect
449 at all, suggesting that earlier reports revealed association between this polymorphism and the
450 studied traits, rather than causation. Alternatively, as is often the case, the functional variant

may lie outside coding regions. Nevertheless, further work within this QTL region will be required to fully characterized the functional variant responsible for the large difference in 8-week body weight in the Virginia body weight lines.

CONCLUSIONS

Over the course of long-term bidirectional selection, the Virginia body weight lines have experienced significant changes impacting behavioral, neurological, metabolic, and developmental processes (Dunnington and Siegel 1996). With a concerted effort to understand the fundamental genomics underlying these changes, we have employed linkage and association mapping, selective-sweep analyses using high-density SNP data and pooled genome sequencing across generations of the selected lines, relaxed lines, and a derived advanced intercross line (Wahlberg et al. 2009; Johansson et al. 2010; Pettersson and Carlborg 2010; Besnier et al. 2011; Pettersson et al. 2011; Ahsan et al. 2013; Pettersson et al. 2013; Sheng et al. 2015; Brandt et al. 2016). It has become clear that the difference in body weight between the selected lines relies on small to moderate effects from many loci, with the most recent analysis revealing 20 contributing loci (Zan et al. 2017).

Pooled genome resequencing revealed distinct hallmarks of selection in regions that were highly differentiated in the lines after 55 generations of long-term experimental selection. More often than not, the regions show the persistence of haplotypic diversity in one line, contrasted by fixation in the other, as demonstrated in the *growth1* QTL region. Despite selection acting on the same pool of standing genetic variants, the genomic signatures of selection thus resemble classic hard sweeps in one line, contrasted to a mosaic pattern of divergence in the other. This haplotype mosaic emerges from recombination of multiple divergent founder haplotypes, probably shaped by historical bottlenecks, crossbreeding, and inbreeding, which forms the standing genetic variation available in this selection experiment.

ACKNOWLEDGEMENTS

For help with genome sequencing, we thank Uppsala University SNP&SEQ Technology Platform (Science for Life Laboratory), a national infrastructure supported by the Swedish Research Council (VR-RFI) and the Knut and Alice Wallenberg Foundation. For computational resources, we thank the Uppsala Multidisciplinary Center for Advanced Computational Science and the associated Next generation sequencing Cluster and Storage (UPPNEX) project, funded by the Knut and Alice Wallenberg Foundation and the Swedish National Infrastructure for Computing (SNIC; project id: b2015010). This work was supported by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (DNR 221-2013-450 to ÖC) and various sources at Virginia Tech. Lastly, individuals at Virginia Tech for a range of roles in producing and maintaining the chickens across more than half a century.

AUTHORS' CONTRIBUTIONS

ÖC and PBS designed the study. ML performed analyses. CFH and PBS designed the pedigrees and managed the Virginia body weight chickens, collected samples, and provided DNA. ML and ÖC wrote the manuscript. All authors read and approved the final manuscript.

REFERENCES

Abasht, B., and S. J. Lamont, 2007 Genome-wide association analysis reveals cryptic alleles as an important factor in heterosis for fatness in chicken F-2 population. Anim. Genet. 38: 491-498. https://doi.org/10.1111/j.1365-2052.2007.01642.x

Abdalhag, M. A., T. Zhang, Q. C. Fan, X. Q. Zhang, G. X. Zhang *et al.*, 2015 Single nucleotide polymorphisms associated with growth traits in Jinghai yellow chickens. Genet. Mol. Res. 14: 16169-16177. https://doi.org/10.4238/2015.December.8.6

Ahsan, M., X. Li, A. E. Lundberg, M. Kierczak, P. B. Siegel *et al.*, 2013 Identification of candidate genes and mutations in QTL regions for chicken growth using bioinformatic analysis of NGS and SNP-chip data. Front. Genet. 4: 226. https://doi.org/10.3389/fgene.2013.00226

Aken, B. L., P. Achuthan, W. Akanni, M. R. Amode, F. Bernsdorff *et al.*, 2017 Ensembl 2017. Nucleic Acids Res. 45: D635-D642. https://doi.org/10.1093/nar/gkw1104

Berg, J. J., and G. Coop, 2014 A population genetic signal of polygenic adaptation. PLoS Genet. 10: 25. https://doi.org/10.1371/journal.pgen.1004412

Besnier, F., P. Wahlberg, L. Rönnegård, W. Ek, L. Andersson *et al.*, 2011 Fine mapping and replication of QTL in outbred chicken advanced intercross lines. Genet. Select. Evol. 43: 10. https://doi.org/10.1186/1297-9686-43-3

Brandt, M., M. Ahsan, C. F. Honaker, P. B. Siegel, and Ö. Carlborg, 2016 Imputation-based fine-mapping suggests that most QTL in an outbred chicken advanced intercross body weight line are due to multiple, linked loci. G3. https://doi.org/10.1534/g3.116.036012

Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 467: 587-U111. https://doi.org/10.1038/nature09352

Chan, Y. F., F. C. Jones, E. McConnell, J. Bryk, L. Bunger *et al.*, 2012 Parallel selection mapping using artificially selected mice reveals body weight control loci. Curr. Biol. 22: 794-800. https://doi.org/10.1016/j.cub.2012.03.011

Coop, G., J. K. Pickrell, J. Novembre, S. Kudaravalli, J. Li *et al.*, 2009 The role of geography in human adaptation. PLoS Genet. 5. https://doi.org/10.1371/journal.pgen.1000500

Dohner, J. V., 2001 *The encyclopedia of historic and endangered livestock and poultry breeds*. New Haven: Yale Univ. Press.

Dunnington, E. A., C. F. Honaker, M. L. McGilliard, and P. B. Siegel, 2013 Phenotypic responses of chickens to long-term, bidirectional selection for juvenile body weight-

528 Historical perspective. Poult. Sci. 92: 1724-1734. https://doi.org/10.3382/ps.2013-
529 03069
530 Dunnington, E. A., and P. B. Siegel, 1996 Long-term divergent selection for eight-week body
531 weight in White Plymouth Rock chickens. Poult. Sci. 75: 1168-1179.
532 Eriksson, J., G. Larson, U. Gunnarsson, B. Bed'hom, M. Tixier-Boichard *et al.*, 2008
533 Identification of the Yellow skin gene reveals a hybrid origin of the domestic chicken.
534 PLoS Genet. 4: 8. https://doi.org/10.1371/journal.pgen.1000010
535 Franssen, S. U., N. H. Barton, and C. Schlötterer, 2017 Reconstruction of haplotype-blocks
536 selected during experimental evolution. Mol. Biol. Evol. 34: 174-184.
537 https://doi.org/10.1093/molbev/msw210
538 Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov, 2015 Recent selective sweeps in
539 north american *Drosophila melanogaster* show signatures of soft sweeps. PLoS
540 Genet. 11: 32. https://doi.org/10.1371/journal.pgen.1005004
541 Groeneveld, L. F., J. A. Lenstra, H. Eding, M. A. Toro, B. Scherf *et al.*, 2010 Genetic
542 diversity in farm animals - a review. Anim. Genet. 41: 6-31.
543 https://doi.org/10.1111/j.1365-2052.2010.02038.x
544 Hermisson, J., and P. S. Pennings, 2005 Soft sweeps: Molecular population genetics of
545 adaptation from standing genetic variation. Genetics 169: 2335-2352.
546 https://doi.org/10.1534/genetics.104.036947
547 Hernandez, R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton *et al.*, 2011 Classic
548 selective sweeps were rare in recent human evolution. Science 331: 920-924.
549 https://doi.org/10.1126/science.1198878
550 Jacobsson, L., H. B. Park, P. Wahlberg, R. Fredriksson, M. Perez-Enciso *et al.*, 2005 Many
551 QTLs with minor additive effects are associated with a large difference in growth
552 between two selection lines in chickens. Genet. Res. 86: 115-125.
553 https://doi.org/10.1017/s0016672305007767

554 Jambui, M., C. F. Honaker, and P. B. Siegel, 2017a Correlated responses to long-term
555 divergent selection for 8-week body weight in female White Plymouth Rock chickens:
556 Sexual maturity. Poult. Sci. 96: 3844-3851. https://doi.org/10.3382/ps/pex224
557 Jambui, M., C. F. Honaker, and P. B. Siegel, 2017b Selection for juvenile body weight in
558 chickens: Standardizing for scaling. Poult. Sci. 96: 2562-2568.
559 https://doi.org/10.3382/ps/pex080
560 Jia, X. Z., H. R. Lin, Q. H. Nie, X. Q. Zhang, and S. J. Lamont, 2016 A short insertion
561 mutation disrupts genesis of miR-16 and causes increased body weight in
562 domesticated chicken. Sci. Rep. 6: 11. https://doi.org/10.1038/srep36433
563 Johansson, A. M., M. E. Pettersson, P. B. Siegel, and Ö. Carlborg, 2010 Genome-wide effects
564 of long-term divergent selection. PLoS Genet. 6.
565 https://doi.org/10.1371/journal.pgen.1001188
566 Kaplan, N. L., R. R. Hudson, and C. H. Langley, 1989 The hitchhiking effect revisited.
567 Genetics 123: 887-899.
568 Karlsson, E. K., I. Baranowska, C. M. Wade, N. H. C. Salmon Hillbertz, M. C. Zody *et al.*,
569 2007 Efficient mapping of mendelian traits in dogs through genome-wide association.
570 Nat. Genet. 39: 1321-1328. https://doi.org/10.1038/ng.2007.10
571 Kofler, R., P. Orozco-terWengel, N. De Maio, R. V. Pandey, V. Nolte *et al.*, 2011
572 PoPoolation: a toolbox for population genetic analysis of next generation sequencing
573 data from pooled individuals. PLoS One 6.
574 https://doi.org/10.1371/journal.pone.0015925
575 Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association
576 mapping and population genetical parameter estimation from sequencing data.
577 Bioinformatics 27: 2987-2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Lillie, M., Z. Y. Sheng, C. F. Honaker, L. Andersson, P. B. Siegel *et al.*, 2017a Genomic signatures of sixty years of bidirectional selection for 8-week body weight in chickens. Poult. Sci. pex383. https://doi.org/10.3382/ps/pex383

Lillie, M., Z. Y. Sheng, C. F. Honaker, B. J. Dorshorst, C. M. Ashwell *et al.*, 2017b Genome-wide standing variation facilitates long-term response to bidirectional selection for antibody response in chickens. BMC Genomics 18: 13. https://doi.org/10.1186/s12864-016-3414-7

Lyimo, C. M., A. Weigend, P. L. Msoffe, H. Eding, H. Simianer *et al.*, 2014 Global diversity and genetic contributions of chicken populations from African, Asian and European regions. Anim. Genet. 45: 836-848. https://doi.org/10.1111/age.12230

Marquez, G. C., P. B. Siegel, and R. M. Lewis, 2010 Genetic diversity and population structure in lines of chickens divergently selected for high and low 8-week body weight. Poult. Sci. 89: 2580-2588. https://doi.org/10.3382/ps.2010-01034

Maynard-Smith, J., and J. Haigh, 1974 Hitch-hiking effect of a favorable gene. Genet. Res. 23: 23-35. https://doi.org/10.1017/s0016672300014634

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297-1303. https://doi.org/10.1101/gr.107524.110

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie *et al.*, 2016 The Ensembl Variant Effect Predictor. Genome Biol. 17: 14. https://doi.org/10.1186/s13059-016-0974-4

Pettersson, M., F. Besnier, P. B. Siegel, and Ö. Carlborg, 2011 Replication and explorations of high-order epistasis using a large advanced intercross line pedigree. PLoS Genet. 7. https://doi.org/10.1371/journal.pgen.1002180

Pettersson, M. E., and O. Carlborg, 2010 Dissecting the genetic architecture of complex traits and its impact on genetic improvement programs: lessons learnt from the Virginia chicken lines. R. Bras. Zootec. 39: 256-260. https://doi.org/10.1590/s1516-35982010001300028

Pettersson, M. E., A. M. Johansson, P. B. Siegel, and Ö. Carlborg, 2013 Dynamics of adaptive alleles in divergently selected body weight lines of chickens. G3 3: 2305-2312. https://doi.org/10.1534/g3.113.008375

Pritchard, J. K., and A. Di Rienzo, 2010 Adaptation - not by sweeps alone. Nat. Rev. Genet. 11: 665-667. https://doi.org/10.1038/nrg2880

Pritchard, J. K., J. K. Pickrell, and G. Coop, 2010 The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. Curr. Biol. 20: R208-R215. https://doi.org/10.1016/j.cub.2009.11.055

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander *et al.*, 2011 Integrative genomics viewer. Nat. Biotech. 29: 24-26. https://doi.org/10.1038/nbt.1754

Sawai, H., H. L. Kim, K. Kuno, S. Suzuki, H. Gotoh *et al.*, 2010 The origin and genetic variation of domestic chickens with special reference to junglefowls *Gallus g. gallus* and *G. varius*. PLoS One 5: 11. https://doi.org/10.1371/journal.pone.0010639

Schrider, D. R., and A. D. Kern, 2017 Soft sweeps are the dominant mode of adaptation in the human genome. Mol. Biol. Evol. 34: 1863-1877. https://doi.org/10.1093/molbev/msx154

628 Shen, M. M., L. Qu, M. Ma, T. C. Dou, J. Lu *et al.*, 2016 Genome-wide association studies
629 for comb traits in chickens. PLoS One 11: 18.
630 https://doi.org/10.1371/journal.pone.0159081
631 Sheng, Z. Y., M. E. Pettersson, C. F. Honaker, P. B. Siegel, and Ö. Carlborg, 2015 Standing
632 genetic variation as a major contributor to adaptation in the Virginia chicken lines
633 selection experiment. Genome Biol. 16. https://doi.org/10.1186/s13059-015-0785-z
634 Sheng, Z. Y., M. E. Pettersson, X. X. Hu, C. L. Luo, H. Qu *et al.*, 2013 Genetic dissection of
635 growth traits in a Chinese indigenous x commercial broiler chicken cross. BMC
636 Genomics 14: 12. https://doi.org/10.1186/1471-2164-14-151
637 Siegel, P. B., 1962 Selection for body weight at 8 weeks of age. 1. Short term response and
638 heritabilities. Poult. Sci. 41: 954-962.
639 Siegel, P. B., and E. A. Dunnington, 1985 Reproductive complications associated with
640 selection for broiler growth in *Poultry Genetics and Breeding.* , edited by W.G. Hill,
641 J.M. Mason and D. Hewitt. British Poultry Sci. Ltd. Longman Group, Harlow.
642 Siegel, P. B., and E. A. Dunnington, 1987 Selection for growth in chickens. Critical Reviews
643 in Poultry Biology 1: 1-24.
644 Stephan, W., T. H. E. Wiehe, and M. W. Lenz, 1992 The effect of strongly selected
645 substitutions on neutral polymorphism - analytical results based on diffusion-theory.
646 Theor. Popul. Biol. 41: 237-254. https://doi.org/10.1016/0040-5809(92)90045-u
647 Teshima, K. M., and M. Przeworski, 2006 Directional positive selection on an allele of
648 arbitrary dominance. Genetics 172: 713-718.
649 https://doi.org/10.1534/genetics.105.044065
650 The American Standard of Perfection, 1957 American Poultry Assoc., Davenport, IA USA
651 Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov, 2013 Integrative Genomics Viewer
652 (IGV): high-performance genomics data visualization and exploration. Brief.
653 Bioinform 14: 178-192. https://doi.org/10.1093/bib/bbs017

Tixier-Boichard, M., B. Bed'hom, and X. Rognon, 2011 Chicken domestication: From archeology to genomics. C. R. Biol. 334: 197-204. https://doi.org/10.1016/j.crvi.2010.12.012

Wahlberg, P., Ö. Carlborg, M. Foglio, X. Tordoir, A. C. Syvanen *et al.*, 2009 Genetic analysis of an F-2 intercross between two chicken lines divergently selected for body-weight. BMC Genomics 10: 13. https://doi.org/10.1186/1471-2164-10-248

West, B., and B. X. Zhou, 1988 Did chickens go north - New evidence for domestication. Journal of Archaeological Science 15: 515-533. https://doi.org/10.1016/0305-4403(88)90080-5

Xie, L., C. L. Luo, C. G. Zhang, R. Zhang, J. Tang *et al.*, 2012 Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. PLoS One 7: 9. https://doi.org/10.1371/journal.pone.0030910

Yi, G. Q., M. M. Shen, J. W. Yuan, C. J. Sun, Z. Y. Duan *et al.*, 2015 Genome-wide association study dissects genetic architecture underlying longitudinal egg weights in chickens. BMC Genomics 16: 14. https://doi.org/10.1186/s12864-015-1945-y

Yuan, J. W., K. H. Wang, G. Q. Yi, M. Ma, T. C. Dou *et al.*, 2015 Genome-wide association studies for feed intake and efficiency in two laying periods of chickens. Genet. Select. Evol. 47: 13. https://doi.org/10.1186/s12711-015-0161-1

Zan, Y., Z. Sheng, M. Lillie, L. Rönnegård, C. F. Honaker *et al.*, 2017 Artificial selection response due to polygenic adaptation from a multilocus, multiallelic genetic architecture. Mol. Biol. Evol. msx194. https://doi.org/10.1093/molbev/msx194

Zhang, G. X., Q. C. Fan, J. Y. Wang, T. Zhang, Q. Xue *et al.*, 2015 Genome-wide association study on reproductive traits in Jinghai Yellow Chicken. Anim. Reprod. Sci. 163: 30-34. https://doi.org/10.1016/j.anireprosci.2015.09.011