Genome-wide Linkage Analysis and Association Study Identifies Loci for Polydactyly in Chickens

Yanfa Sun*†#1, Ranran Liu*§1, Guiping Zhao*§, Maiqing Zheng*§, Yan Sun*, Xiaoqiong Yu*, Peng Li*, Jie Wen*§2

*Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, P.R. China; †College of Life Science, Longyan University, Longyan, Fujian 364012, P.R. China; §State Key Laboratory of Animal Nutrition, Beijing 100193, P.R. China; #Fujian Provincial Key Laboratory of Preventive Veterinary Medicine and Biotechnology, Longyan, Fujian 364012, P.R. China

1Yanfa Sun and Ranran Liu contributed equally to this work.

2Corresponding author

*§Yuanmingyuan West Road No. 2, Haidian District, Beijing 100193, P.R. China

†#Dongxiao North Road No. 1, Xinluo District, Longyan, Fujian 364012, P.R. China
LOCi FOR CHICKEN POLYDACTYLY

Key words: chicken, polydactyly, linkage analysis, GWAS, candidate genes

Corresponding author: Jie Wen, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Yuanmingyuan West Road No. 2, Haidian District, Beijing 100193, P.R. China. Tel: +86-010-62815856; Fax: +86-010-62895351; Email: jiewen@iascaas.net.cn.
ABSTRACT Polydactyly occurs in some chicken breeds, but the molecular mechanism remains incompletely understood. Combined genome-wide linkage analysis and association study (GWAS) for chicken polydactyly helps identify loci or candidate genes for the trait, and potentially provides further mechanistic understanding of this phenotype in chickens, and perhaps other species. The linkage analysis and GWAS for polydactyly was conducted using an F2 population derived from Beijing-You chickens and commercial broilers. The results identified two QTLs through linkage analysis and seven SNPs through GWAS, associated with the polydactyly trait. One QTL located at 35 cM on the GGA2 was significant at the 1% genome-wise level and another QTL at the 1% chromosome-wide significance level was detected at 39 cM on GGA19. A total of seven SNPs, four of 5% genome-wide significance (P < 2.98×10^{-6}) and three of suggestive significance (5.96×10^{-5}) were identified, including two SNPs (GGaluGA132178 and Gga_rs14135036) in the QTL on GGA2. Of the identified SNPs, the eight nearest genes were sonic hedgehog (SHH), limb region 1 homolog (mouse) (LMBR1), dipeptidyl-peptidase 6, transcript variant 3 (DPP6), thyroid stimulating hormone, beta (TSHB), sal-like 4 (Drosophila) (SALL4), par-6 partitioning defective 6 homolog beta (C. elegans) (PARD6B), coenzyme Q5 (COQ5), and tyrosine 3-monoxygenase/tryptophan 5-monoxygenase activation protein, etapolyptide (YWHAH). The GWAS supports earlier reports of the importance of SHH and LMBR1 as regulating genes for polydactyly in chickens and other species, and identified others, most of which have not previously been associated with limb development. The genes and associated SNPs revealed here provide detailed information for further exploring the molecular and developmental mechanisms underlying polydactyly.
Polydactyly is a common limb malformation in human, mouse, chicken and other vertebrates characterized by supernumerary digits. In chickens, where the usual number is four, polydactyly is the phenotype with a fifth digit on one or both feet (Warren, 1944). Five digits is a breed characteristic in some chickens, including Beijing-You, Silkies, Dorking, Houdan and Sultan. The chicken polydactyly locus was mapped to chromosome 2 (GGA2) through linkage analysis, in a region homologous to human Chr7q36 and mouse Chr5 (Pitel et al., 2000). This region is known to be important for human and mouse polydactyly traits and other limb malformation (Dundar et al., 2001; Dobbs et al., 2005; Lettice et al., 2005). Huang et al. (2006) found that the T1254C polymorphism in exon 13 of LMBR1 was strongly associated with polydactyly in chickens. Another study has shown complete association of this same SNP (ss161109890) with the polydactyly phenotype in a manner that was concordant with the dominant and incompletely penetrant nature of this trait (Dorshorst et al., 2010). This study also searched the highly conserved zone of polarizing activity (ZPA) regulatory sequence (ZRS) for other SNPs that might explain this trait and found none. Several polydactylous breeds were compared, with the Silkie and Sultan breeds both having this SNP while three other breeds with polydactyly did not (Faverolle, Houdan and Dorking). Due to the history of these breeds, they suggested there may be more than one allele or locus controlling polydactyly in the chicken, although in all of these breeds polydactyly still behaves as a single gene trait (Dorshorst et al., 2010).

The causality of the ss161109890 SNP located within the ZRS was subsequently demonstrated using SHH allelic imbalance and electroporation of mutant ZRS experiments. The ectopic expression of SHH in the anterior ZPA is consistent with known causes of polydactyly in other species (Dunn et
The related genes for polydactyly were identified to better understand the molecular and developmental mechanisms underlying polydactyly. Genes related to hedgehog signaling, sonic hedgehog (SHH), and limb region 1 homolog (mouse) (LMBR1), are important regulatory genes for limb development (Lettice et al., 2012) and polydactyly traits in human, mouse and chicken (Clark et al., 2001; Maas and Fallon, 2004; Sagai et al., 2004; Sagai et al., 2005; Huang et al., 2006; Albuisson et al., 2010; Dunn et al., 2011). Other candidate genes (e.g. EN2, HOXA10-13, GLI3, WNT2, and WNT16) for the polydactyly trait were also found in different species (Gorbach et al., 2010). Quinn et al., (2012) identified a previously unrecognized role for Zic3 in regulating limb digit number via its modifying effect on GLI3 and SHH expression levels in mice. Kalsoom et al. 2013 have recently identified a novel zinc-finger gene (ZNF141) associated with autosomal recessive postaxial polydactyly type A through whole genome sequencing combined with homozygosity mapping and array comparative genomic hybridization (aCGH) analysis. These studies found the gene for one polydactyly allele/locus, but the molecular mechanism and other polydactyly alleles/loci remain incompletely understood (Robb and Delany, 2012).

Both linkage analysis and genome-wide association study (GWAS) can be used to map the QTLs underlying complex traits. The GWAS, based on high throughput SNP genotyping technologies, has revealed loci underlying a variety of interesting traits in domestic animals (Pearson et al., 2008) and it is more powerful than traditional linkage analysis in detecting genetic variants. In chickens, GWAS have been performed for growth and meat quality (Gu et al., 2011; Xie et al., 2012; Sun et al., 2013; Liu et al., 2013), egg production and quality (Liu et al., 2011), and other traits.

In the present study, both linkage analysis and GWAS have been performed for the polydactyly
trait using the CAAS chicken F2 resource population derived from the cross between a local Chinese chicken breed (Beijing-You) and a commercial broiler line (Cobb-Vantress) (Sun et al., 2013) to identify and provide information for allele/locus for polydactyly in chickens.

MATERIALS AND METHODS

Animals and phenotypes

This research complied with the Guidelines for Experimental Animals established by the Ministry of Science and Technology (Beijing, China). The Chinese Academy of Agricultural Science (CAAS) chicken F2 resource population was derived from a cross between Beijing-You (BJY) chickens and Cobb broilers (CB). The BJY is a Chinese indigenous breed with a 25% to 31% incidence of five digits on one or both feet. The CB (Cobb-Vantress) is a commercial broiler breed with four digits on each foot. Six BJY males were each mated to 12 CB females. Six males and 20 females of their offspring, the F1 generation, then produced the F2 chickens using artificial insemination and only non-related females to avoid inbreeding. The chickens from the three generations were raised in stair-step cages under the same routine conditions of nutrition and management at the conservation farm of the Institute of Animal Sciences (Sun et al., 2013).

A total of 400 chickens from three generation (including 367 F2 chickens from 20 full-sib families in five batches) were used. Blood was collected into acid citrate dextrose anticoagulant tubes from the brachial vein on d 56. The number of digits, four or five on one or both feet, was recorded. Two out of six BJY males in the F0 generation had polydactyly. The pedigrees with polydactyly traits are shown in Table S1.

Genotyping and SNP quality
Genomic DNA (gDNA) was extracted from blood samples using the phenol-chloroform method. Genotyping was performed with 50 ng/µl gDNA using the Illumina 60K Chicken SNP Beadchip by DNA LandMarks Inc. (Saint-Jean-sur-Richelieu, PQ, Canada). A total of 39 were excluded due to sample call rate < 90%. A total of 13,985 SNPs were removed for failing to meet one or more of the following conditions: SNP call rate < 90%, minor allele frequency (MAF) < 3%, Hardy-Weinberg equilibrium (HWE) test p of < 10^{-6} and SNPs with no assigned chromosome or linkage group. After quality control measures, 42,585 SNPs distributed among 29 chromosomes (including the Z chromosome) and one linkage group (LGE22) remained. The average physical distance between two neighbouring SNPs was approximately 20.4 Kb (Sun et al., 2013).

**Linkage analysis**

The genetic linkage map was constructed using 19 full-sib families (Six males and 12 females from F0, five males and 20 females from F1, and 148 males and 152 females from F2) and 42,585 SNPs with the improved version of CRI-MAP (2.503a, run in a 64-bit Unix system). The method was described in previous studies (Groenen et al., 2009; Elferink et al., 2010). In genetic map construction, one full-family was excluded due to a computational problem.

In order to reduce the effect of linkage disequilibrium (LD) on the results, 6518 independent SNPs were acquired in all autosomal chromosomes using the indep-pairwise option, with a window size of 25 SNPs, a step of 5 SNPs, and r^2 threshold of 0.2 (Sun et al., 2013). QTL analysis for the polydactyly trait used these same SNPs and web-based GridQTL software (Allen et al., 2012). A least-squares regression model was used for single-QTL analysis, including the fixed effects of sex, hatch, and family, along with additive and dominance coefficients for the putative QTL. Significance thresholds were calculated using a permutation test (Churchill & Doerge 1994). A total of 10000
permutations were computed to determine 5% and 1% chromosome-wide significance levels. 5% and 1% genome-wide significance levels were calculated following the Bonferroni correction:

\[ P_{\text{genome-wide}} = 1 - (1 - P_{\text{chromosome-wide}})^{\frac{Ga}{Gc}} \]

Where Ga is the length genetic map of each chromosome, and Gc is the length of the genetic map of all autosomal chromosomes. Confidence intervals for QTL positions were estimated by bootstrapping, as presented by Visscher et al. (1996).

**Statistical analysis for GWAS**

The population structure was assessed by multidimensional scaling (MDS) analysis using PLINK software 1.07 (Purcell et al., 2007), as described previously in detail (Gu et al., 2011 and Sun et al., 2013). Pairwise identity-by-state (IBS) distances were calculated between all individuals using the 6518 independent SNP markers, and MDS components were obtained using the mds-plot option based on the IBS matrix.

The GWAS for polydactyly with 328 chickens from the F2 generation and 42,585 SNPs was performed using the compressed mixed linear model (MLM) (Zhang et al., 2010) and carried out using Tassel 3.0 software (Bradbury et al., 2007). The statistical model was:

\[ Y_{ijklmn} = \mu + C_{1j} + S_k + B_{l} + G_m + K_n + e_{ijklmn}, \]

where, \( Y_{ijklmn} \) are phenotypic values, \( \mu \) is the common mean, \( C_{1j} \) is the effect of the first principal component, \( S_k \) is the effect of sex, \( B_{l} \) is the effect of batch of hatching \((l=1 - 5)\), \( G_m \) is the effect of the SNP, \( K_n \) is the random effect of the relative kinship matrix, which was constructed by matrix simple matching coefficients based on the independent SNPs, and this step was followed by compression (Zhang et al., 2010), and \( e_{ijklmn} \) is the random residual effect.

The genome-wide significance p-value threshold was determined by the “LD adjusted”
The Bonferroni method (Duggal et al., 2008), which was calculated on the basis of the estimated number of independent markers and LD blocks in all chromosomes. The F2 population was estimated to have 16,760 “independent” tests, based on the “solid spine of LD” algorithm with a minimum D′ value of 0.8 calculated by Haploviev 4.1 (Barrett et al., 2005). The two threshold p-values were established as $5.96 \times 10^{-5}$ (1/16,760) for suggestive significance, and $2.98 \times 10^{-6}$ (0.05/16,760) for 5% genome-wide significance. The Manhattan plot, showing genome-wide p-values of GWAS, was produced using R 2.13.2 software with the “gap” package (Zhao, 2007).

RESULTS

Linkage analysis

Before linkage analysis, the genetic linkage map was constructed using the CAAS chicken F2 population. A total of 42,585 SNP markers (each marker had 4 to 603 informative meiosis, average of 254), about 74% of all markers on the SNP beadchip, and 337 individuals from three generation and 19 full-sib families were used to construct the linkage map. As shown in Table 1, the total length of the sex-average map was 3040.8 cM and there was no significant difference in total length between the male and female sex-specific maps (2929.2 and 2930.3 cM). The recombination rate of the map was 2.9 cM/Mb. The detailed information for the map is shown in Table S2.

A total of three hundred chickens from the F2 generation and 6518 independent SNPs (Table S3) in autosomal chromosomes were used in the linkage analysis for polydactyly. Quantitative trait loci locations and effects for the polydactyly trait are shown in Table 2. One QTL located at 35 cM (confidence interval from 23 to 63 cM) on GGA2 was significant at the 1% genome-wise level (Figure S1); it explained 9.34% of the phenotypic variation. Another QTL at the 1%
chromosome-wide significance level was detected at 39 cM on GGA19 and it explained 5.35% of the phenotypic variation.

**GWAS results**

After 39 samples were excluded because of low sample call rate and one lost phenotypic record, 327 chickens (286 with four digits, and 41 with five digits) from the F2 generation were used for further analysis.

The MDS analysis of the 16,760 SNPs using the first two MDS components showed that the chickens within each full-sib family were clustered together (Sun *et al.*, 2013). To correct for population stratification, the first MDS component was used as a covariate in the model as suggested (Price *et al.*, 2006; Gu *et al.*, 2011; Sun *et al.*, 2013).

As shown in Figure 1 and Table 3, four SNPs associated with polydactyly were of genome-wide significance and three SNPs had suggestive significance. These SNPs explained 6 to 9% of the phenotypic variation.

The most significant SNP, GGaluGA132178, was located on GGA2 at 6641.2 Kb (P = 7.94 × 10^{-7}), 9.5 Kb downstream of dipeptidyl-peptidase 6, transcript variant 3 (*DPP6*). Another SNP, Gga_rs14135036, located on GGA2 at 7628.2 Kb was suggestively associated with polydactyly (P = 5.96 × 10^{-5}). This SNP was in close proximity to other candidate genes for polydactyly, 113.9 Kb downstream of sonic hedgehog (*SSH*) and 240.7 Kb upstream of limb region 1 homolog (mouse) (*LMBRI*).

The SNP Gga_rs13606164, located at 2579.1 Kb on GGA26, is 35.6 Kb downstream of the thyroid stimulating hormone, beta (*TSHB*) gene, had highly significant association with the polydactyly (P < 2.98×10^{-6}). Two other SNPs (GGaluGA181512 and Gga_rs16175635), located
within a 13.3 Kb segment on GGA20, were of genome-wide significance \( (P < 2.98 \times 10^{-6}) \); one was 176.3 Kb upstream of the Sal-like 4 \((Drosophila)\) \((SALL4)\) gene and the other was 176.4 Kb from the Par-6 partitioning defective 6 homolog beta \((C. elegans)\), \((PARD6B)\) gene.

Two SNPs \((\text{Gga}_{rs}15023766\) and \(\text{Gga}_{rs}14093104)\) with suggestive association (both \(p\)-values were \(2.74 \times 10^{-5}\)) with polydactyly were located within a 484.99 Kb segment on GGA15; one was 22.9 Kb downstream of coenzyme Q5 \((COQ5)\), and the other was 2.8 Kb upstream of tyrosine 3-monoxygenase/tryptophan 5-monoxygenase activation protein, etapolypeptide, \((YWHAH)\).

**DISCUSSION**

In the present study, the QTLs and candidate genes for polydactyly were identified in chickens using both linkage analysis and GWAS. A high-density genetic linkage map is the basis for linkage analysis. The current map is comprised of 29 chromosomes and one linkage group, with a total size of 3040.8 cM. The map size was similar with the high resolution chicken genetic map using 13340 SNPs and 1619 chickens \((\text{Elferink et al.}, 2010)\). The present map ensures accuracy and precision of linkage analysis. In the present study, two QTLs (one located at 35 cM on GGA2 and another located at 39 cM on GGA19) were identified through linkage analysis. The QTL on GGA2 was also found in previous studies in chicken \((\text{Dunn et al.}, 2011; \text{Yang et al.}, 2013)\) and other species.

For GWAS analysis, the compressed MLM was one of the most effective methods for controlling false positives in GWAS \((\text{Zhang et al.}, 2010; \text{Huang et al.}, 2010)\) and ensured reliability of the results. The population-based GWAS has more power than traditional family-based linkage analysis \((\text{Klein}, 2007)\) and it can detect more SNPs of interest that are in or near potential candidate genes. The GWAS results obtained here were in accord with the linkage analysis on GGA2 and
obtained more detail of relevant SNPs. The GWAS also found additional potential loci on other chromosomes. Seven significant SNPs were identified near the potential candidate genes or the important regulatory genes for polydactyly. The genes, SHH and LMBR1 on GGA2, already known to be important for polydactyly in chickens and other species, were identified again here. Variations of LMBR1 were found to influence digits numbers in chickens (Huang et al., 2006; Hang et al., 2007). Both LMBR1 and SHH, were shown to be essential for normal development of the chicken limb (Lettice et al., 2003; Lettice et al., 2005; Dunn et al., 2011). The SHH and LMBR1 were also in the QTL region identified here by linkage analysis. Another gene DPP6, near the QTL peak in GGA2, was also identified here. Mutations of DPP6 produce human disease and cause movement disorders (Tanaka et al., 2011).

The gene SALL4 encodes a putative zinc finger transcription factor and plays a role in modulating embryonic stem cell pluripotency and early embryonic development (Zhang et al., 2006). Variations of SALL4 are associated with upper limb abnormality in humans (Kohlhase et al., 2002; Paradisi et al., 2007). Given what is known of the roles of this gene in limb development, the present GWAS findings suggest that it is a potential candidate gene for polydactyly in chickens.

Four other genes, TSHB, PAR6B, COQ5, and YWHAH, are involved in growth and development, cell differentiation or signal transduction (Phillips, 2004; Nolan et al., 2008; Alarcon et al., 2010; Dibrov et al., 1997; Grover et al., 2009), but there has been no suggestion that they are involved in limb development. The present results indicate that they might be potential candidate genes for polydactyly and so they are deserving of further study in this regard.

In conclusion, two QTLs and eight genes (SHH, LMBR1, DPP6, TSHB, SALL4, PAR6B, COQ5 and YWHAH) were identified through combined linkage analysis and GWAS for the
polydactyly phenotype in chickens. In addition to confirming earlier studies in chickens and other
species that \textit{SHH} and \textit{LMBRI} are important regulating genes for polydactyly, the other genes
identified here provide additional information for potentially exposing the molecular mechanisms
and developmental biology underlying polydactyly in chickens. As new candidate genes, especially
the \textit{SALL} gene, they will be targets for future work.

\section*{ACKNOWLEDGEMENTS}

The authors would like to acknowledge W. Bruce Currie (Emeritus Professor, Cornell
University) for his suggestions on presentation. The research was supported by grants: National Key
Technology R&D Program (2011BAD28B03), National High-tech R&D Program (2011AA100301),
the Agricultural Science and Technology Innovation Program (ASTIP-IAS04) and Longyan
University Scientific Research Fund for the Doctoral Young Scholars.

\section*{REFERENCES}

Alarcon, V. B., 2010 Cell polarity regulator PARD6B is essential for trophectoderm formation in the

Albuisson, J., B. Isidor, M. Giraud, O. Pichon, T. Marsaud \textit{et al.}, 2010 Identification of two novel
mutations in Shh long-range regulator associated with familial pre-axial polydactyly. Clinical

bioinformatics application to the cloud. Digital Research, September 10-12, Oxford, UK.

Barrett, J., B. Fry, J. Maller and M. Daly, 2005 Haploview: analysis and visualization of LD and


syndrome maps to chromosome 7q36. Journal of medical genetics 38: 304-309.


Elferink, M. G., P. van As, T. Veenendaal, R. P. Crooijmans and M. A. Groenen, 2010 Regional differences in recombination hotspots between two chicken populations. BMC genetics 11: 11.


Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng et al., 2010 Genome-wide association studies of 14 agronomic traits in rice landraces. Nature genetics 42: 961-967.


Pitel, F., R. Bergé, G. Coquerelle, R. Crooijmans, M. A. M. Groenen et al., 2000 Mapping the Naked Neck (NA) and Polydactyly (PO) mutants of the chicken with microsatellite molecular markers. Genetics Selection Evolution 32: 73-86.


Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira et al., 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81: 559-575.

Quinn, M. E., A. Haaning and S. M. Ware, 2012 Preaxial polydactyly caused by Gli3
haploinsufficiency is rescued by Zic3 loss of function in mice. Human molecular genetics 21: 1888-1896.

Robb, E., and M. Delany. 2012. The expression of preaxial polydactyly is influenced by modifying genetic elements and is not maintained by chromosomal inversion in an avian biomedical model. Cytogenetic Genome Res. 136:50-68.


Sun, Y., G. Zhao, R. Liu, M. Zheng, Y. Hu et al., 2013 The identification of 14 new genes for meat quality traits in chicken using a genome-wide association study. BMC Genomics 14: 458.


Table 1 Linkage map lengths and recombination rates

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Length</th>
<th>Sex-average</th>
<th>Sex-specific</th>
<th>Recombination rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGA1</td>
<td>200.9</td>
<td>450.2</td>
<td>460.5</td>
<td>440.8</td>
</tr>
<tr>
<td>GGA2</td>
<td>154.8</td>
<td>308.4</td>
<td>328.7</td>
<td>288.5</td>
</tr>
<tr>
<td>GGA3</td>
<td>113.6</td>
<td>253.4</td>
<td>269.9</td>
<td>237.3</td>
</tr>
<tr>
<td>GGA4</td>
<td>94.2</td>
<td>199.9</td>
<td>211.2</td>
<td>188.7</td>
</tr>
<tr>
<td>GGA5</td>
<td>62.2</td>
<td>155.6</td>
<td>163.3</td>
<td>148.6</td>
</tr>
<tr>
<td>GGA6</td>
<td>37.3</td>
<td>102.7</td>
<td>110.2</td>
<td>95.3</td>
</tr>
<tr>
<td>GGA7</td>
<td>38.3</td>
<td>103.9</td>
<td>102.8</td>
<td>105.2</td>
</tr>
<tr>
<td>GGA8</td>
<td>30.6</td>
<td>90.9</td>
<td>92.5</td>
<td>89.4</td>
</tr>
<tr>
<td>GGA9</td>
<td>25.5</td>
<td>98.8</td>
<td>107.1</td>
<td>90.1</td>
</tr>
<tr>
<td>GGA10</td>
<td>22.5</td>
<td>82.2</td>
<td>86.2</td>
<td>78.3</td>
</tr>
<tr>
<td>GGA11</td>
<td>21.9</td>
<td>63.9</td>
<td>68.2</td>
<td>59.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>GGA12</td>
<td>20.5</td>
<td>75.1</td>
<td>79.5</td>
<td>70.7</td>
</tr>
<tr>
<td>GGA13</td>
<td>18.9</td>
<td>60.5</td>
<td>66.3</td>
<td>55.6</td>
</tr>
<tr>
<td>GGA14</td>
<td>15.8</td>
<td>67.8</td>
<td>71.3</td>
<td>64.4</td>
</tr>
<tr>
<td>GGA15</td>
<td>13.0</td>
<td>53.7</td>
<td>58.1</td>
<td>49.4</td>
</tr>
<tr>
<td>GGA16</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>GGA17</td>
<td>11.2</td>
<td>49.8</td>
<td>51.5</td>
<td>48.2</td>
</tr>
<tr>
<td>GGA18</td>
<td>10.9</td>
<td>48.0</td>
<td>48.7</td>
<td>47.4</td>
</tr>
<tr>
<td>GGA19</td>
<td>9.9</td>
<td>51.4</td>
<td>55.9</td>
<td>46.8</td>
</tr>
<tr>
<td>GGA20</td>
<td>13.9</td>
<td>54.3</td>
<td>56.1</td>
<td>52.5</td>
</tr>
<tr>
<td>GGA21</td>
<td>6.9</td>
<td>48.5</td>
<td>48.4</td>
<td>48.7</td>
</tr>
<tr>
<td>GGA22</td>
<td>3.9</td>
<td>51.6</td>
<td>52.0</td>
<td>51.3</td>
</tr>
<tr>
<td>GGA23</td>
<td>6.0</td>
<td>51.5</td>
<td>48.9</td>
<td>54.1</td>
</tr>
<tr>
<td>GGA24</td>
<td>6.4</td>
<td>47.9</td>
<td>44.3</td>
<td>51.5</td>
</tr>
<tr>
<td>GGA25</td>
<td>2.0</td>
<td>58.4</td>
<td>63.6</td>
<td>54.1</td>
</tr>
<tr>
<td>GGA26</td>
<td>5.1</td>
<td>48.3</td>
<td>47.5</td>
<td>49.2</td>
</tr>
<tr>
<td>GGA27</td>
<td>4.7</td>
<td>49.6</td>
<td>47.6</td>
<td>51.6</td>
</tr>
<tr>
<td>GGA28</td>
<td>4.5</td>
<td>48.7</td>
<td>53.1</td>
<td>44.4</td>
</tr>
<tr>
<td>LGE22</td>
<td>0.9</td>
<td>37.6</td>
<td>35.5</td>
<td>40.5</td>
</tr>
<tr>
<td>GGZ</td>
<td>74.6</td>
<td>228.3</td>
<td>-</td>
<td>228.3</td>
</tr>
<tr>
<td>Total autosomal</td>
<td>956.7</td>
<td>2812.5</td>
<td>2929.2</td>
<td>2702.0</td>
</tr>
<tr>
<td>Total Length</td>
<td>1031.3</td>
<td>3040.8</td>
<td>2929.2</td>
<td>2930.3</td>
</tr>
</tbody>
</table>

1physical length of the chromosome was based on the position of the last marker in the
WASHUC2 build (2006).

<table>
<thead>
<tr>
<th>Location (cM/Mb)</th>
<th>Confidence interval (cM/Mb)</th>
<th>F-ratios(^1)</th>
<th>A ± SE(^2)</th>
<th>D ± SE</th>
<th>PV (%)(^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGA2 (35/0.8)</td>
<td>23.0 - 63.0/ 0.5 - 2.2</td>
<td>15.88**</td>
<td>0.17 ± 0.03</td>
<td>0.05 ± 0.04</td>
<td>9.34</td>
</tr>
<tr>
<td>GGA19 (39/0.5)</td>
<td>22.0 - 45.0/0.3- 0.7</td>
<td>9.06*</td>
<td>0.03 ± 0.03</td>
<td>0.18 ± 0.06</td>
<td>5.35</td>
</tr>
</tbody>
</table>

\(^1\) 1% Chromosome-wide significance, and ** 1% Genome-wide significance.

\(^2\) A, Additive; D, Dominance.

\(^3\) PV = Percent of phenotypic variance explained by the QTL. PV% = \([\text{MSR} - \text{MSF}/\text{MSR}] \times 100\) where MSR-residual mean square in the reduced model, MSF-residual mean square in the full model.
<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNPs</th>
<th>Position (bp)</th>
<th>P_value&lt;sup&gt;1&lt;/sup&gt;</th>
<th>P_adjust&lt;sup&gt;2&lt;/sup&gt;</th>
<th>MAF&lt;sup&gt;3&lt;/sup&gt;</th>
<th>R&lt;sup&gt;2&lt;/sup&gt;&lt;sup&gt;4&lt;/sup&gt;</th>
<th>Nearest gene&lt;sup&gt;5&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>GGaluGA132178</td>
<td>6641213</td>
<td>7.94E-07</td>
<td>0.01</td>
<td>C/A-0.10</td>
<td>0.09</td>
<td>DPP6 D 9.5 Kb</td>
</tr>
<tr>
<td>2</td>
<td>Gga_rs14135036</td>
<td>7628179</td>
<td>5.96E-05</td>
<td>1.00</td>
<td>A/G-0.31</td>
<td>0.06</td>
<td>SHH D 113.9 Kb; LMBR1 U 240.7 Kb</td>
</tr>
<tr>
<td>15</td>
<td>Gga_rs15023766</td>
<td>9240112</td>
<td>2.74E-05</td>
<td>0.46</td>
<td>C/A-0.17</td>
<td>0.07</td>
<td>COQ5 U 22.9 Kb</td>
</tr>
<tr>
<td>15</td>
<td>Gga_rs14093104</td>
<td>8755174</td>
<td>2.74E-05</td>
<td>0.46</td>
<td>G/A-0.17</td>
<td>0.07</td>
<td>YWHAH U 2.8 Kb</td>
</tr>
<tr>
<td>20</td>
<td>GGaluGA181512</td>
<td>12887215</td>
<td>2.08E-06</td>
<td>0.03</td>
<td>C/A-0.19</td>
<td>0.09</td>
<td>SALL4 U 176.3 Kb</td>
</tr>
<tr>
<td>20</td>
<td>Gga_rs16175635</td>
<td>12900524</td>
<td>2.08E-06</td>
<td>0.03</td>
<td>A/G-0.19</td>
<td>0.09</td>
<td>PARD6B D 176.4 Kb</td>
</tr>
<tr>
<td>26</td>
<td>Gga_rs13606164</td>
<td>2579119</td>
<td>1.47E-06</td>
<td>0.02</td>
<td>A/G-0.14</td>
<td>0.09</td>
<td>TSHB U 35.6 Kb</td>
</tr>
</tbody>
</table>

<sup>1</sup>P_Value from MLM models; <sup>2</sup>LD-adjusted Bonferroni P-value; <sup>3</sup>MAF, minor allele frequency of the first allele; <sup>4</sup>R<sup>2</sup>, the proportion of phenotypic variation explained by the SNP; <sup>5</sup>U indicates the SNP is upstream of the gene and D is downstream.
Figure legends

Figure 1 Manhattan plot showing the association of all SNPs with the polydactyly trait. SNPs are plotted on the x-axis according to their position on each chromosome against association with these traits on the y-axis (shown as -log10 P-value). The black line indicates genome-wise significance of suggestive association (P = 5.96 ×10^{-5}), and the red line indicates genome-wise 5% significance with a P-value threshold of 2.98×10^{-6}.

Supporting information

Table S1 The pedigrees with polydactyly traits in the CAAS chicken F2 population.

Table S2 Genetic map of the CAAS chicken F2 population.

Table S3 Independent SNPs used in linkage analysis and multidimensional scaling analysis.

Figure S1 Polydactyly QTL analysis on GGA2 in the CAAS chicken F2 population. The bold line shows the 1% genome-wide significance threshold.