

Title: Conserved non-coding regions in the chicken genome harbour functionally important variation

Authors:

Chiara Bortoluzzi^{1,*}, Christian Groß^{2,3,*}, Mirte Bosse¹, Dick de Ridder², Marcel Reinders³, Hendrik-Jan Megens¹

Affiliations:

¹ Animal Breeding and Genomics Group, Wageningen University & Research, the Netherlands

² Bioinformatics Group, Wageningen University & Research, the Netherlands

³ Bioinformatics Group, TU Delft, the Netherlands

* These authors equally contributed

Abstract:

Title:	Conserved non-coding regions in the chicken genome harbour functionally important variation
Authors:	Chiara Bortoluzzi ^{1,*} , Christian Groß ^{2,3,*} , Mirte Bosse ¹ , Dick de Ridder ² , Marcel Reinders ³ , Hendrik-Jan Megens ¹
Affiliations:	¹ Animal Breeding and Genomics Group, Wageningen University & Research, the Netherlands ² Bioinformatics Group, Wageningen University & Research, the Netherlands ³ Bioinformatics Group, TU Delft, the Netherlands
Abstract:	Following functional studies in model species, it is now known that the majority of the eukaryote genome consists of non-coding regions. However, recognizing the functional significance of non-coding variants remains difficult. Conserved non-coding sequences (CNCs) identified by comparative genomic approaches are reliable indicators of functionally constrained sequences and are likely to contain functional variation. Using this principle, we analyze CNC sequences in chicken and develop a Combined Annotation Dependent Depletion (CADD) machine learning approach for chicken (i.e. chCADD). We find that more than 6% of the chicken genome has been conserved across more than 280 million years of vertebrate evolution. We also show that, although the chicken genome is less evolutionarily conserved (~73 Mb) than the human genome, 70% of the conserved elements are non-coding and display features characteristic of genomic regions constrained by purifying selection. Variants found in CNC regions have an overall high chCADD score, which suggests that they are likely to be functional. We also use the chCADD score to identify specific subregions of higher functional importance within CNC regions and found that genes in most of these regions are associated with known diseases in human, mouse, and rat. Taken together, our findings indicate that CNC regions harbour variants of functional significance that should also be intensively surveyed and prioritized along with protein-

	<p>coding mutations in genome-wide association studies. As we show, such prioritization is nowadays possible thanks to the development of new methodological approaches, such as chCADD. Therefore, the results of this study have important implications for further functional study in chicken.</p>
<u>Submitting author (name):</u>	Chiara Bortoluzzi
<u>Submitting author (email):</u>	chiara.bortoluzzi@wur.nl