SCIENTIFIC REPORTS

OPEN

Avian Toll-like receptor allelic diversity far exceeds human polymorphism: an insight from domestic chicken breeds

Zuzana Świderská [1,2], Adéla Šmídová[1], Lucie Buchtová[1], Anna Bryjová[1,3], Anežka Fabiánová[1], Pavel Munclinger[1] & Michal Vinkler [1]

Immune genes show remarkable levels of adaptive variation shaped by pathogen-mediated selection. Compared to humans, however, population polymorphism in animals has been understudied. To provide an insight into immunogenetic diversity in birds, we sequenced complete protein-coding regions of all Toll-like receptor (*TLR*) genes with direct orthology between mammals and birds (*TLR3*, *TLR4*, *TLR5* and *TLR7*) in 110 domestic chickens from 25 breeds and compared their variability with a corresponding human dataset. Chicken *TLRs* (*chTLRs*) exhibit on average nine-times higher nucleotide diversity than human *TLRs* (*hTLRs*). Increased potentially functional non-synonymous variability is found in *chTLR* ligand-binding ectodomains. While we identified seven sites in *chTLRs* under positive selection and found evidence for convergence between alleles, no selection or convergence was detected in *hTLRs*. Up to six-times more alleles were identified in fowl (70 *chTLR4* alleles vs. 11 *hTLR4* alleles). In *chTLRs*, high numbers of alleles are shared between the breeds and the allelic frequencies are more equal than in *hTLRs*. These differences may have an important impact on infectious disease resistance and host-parasite co-evolution. Though adaptation through high genetic variation is typical for acquired immunity (e.g. *MHC*), our results show striking levels of intraspecific polymorphism also in poultry innate immune receptors.

Domestic chickens and humans have a lot in common. Given their joint history, the domestic chicken is now an abundant and widespread species around the world. Both humans and chickens now serve as basic research model species representing genome references for birds and mammals. Yet domestic chicken breeds remain rarely studied from the perspective of evolutionary immunology. This is despite Darwin himself pointed out that variation in domestic fowl populations provides an excellent system for investigating evolution through natural and artificial selection[1]. Selection is particularly strong at shaping host-pathogen interactions, where different genes can be selected for genetic diversity. In the light of these facts, it is surprising how little attention has been paid to immunologically-relevant genetic variation in the highly diversified yet phenotypically standardised breeds of domestic fowl[2–5].

Unlike the many laboratory and commercial lines more typically studied[6,7], many domestic chicken breeds exhibit high genetic and phenotypic variation[8–10]. Several hundreds of breeds are now recognised worldwide, many of which are only locally distributed, having been maintained as stable phenotypic forms for centuries[11]. These traditional fowl breeds were originally domesticated from free-living red junglefowl subspecies, with possible admixture of other *Gallus* species, on multiple occasions in different regions of Asia[12]. This, together with distinct trade-driven migration routes and selection for different human needs in different environments, may have diversified the breeds phenotypically as regards pathogen resistance.

Previous studies suggest that the chicken genome is approximately two-times richer in exon polymorphism than the human genome[13,14]. From an evolutionary perspective, variation in innate immune receptor genes, which form a direct molecular interface between pathogens and their hosts, is particularly appealing since major

[1]Charles University, Faculty of Science, Department of Zoology, Viničná 7, Prague, 12843, Czech Republic. [2]Charles University, Faculty of Science, Department of Cell Biology, Viničná 7, Prague, 12843, Czech Republic. [3]The Czech Academy of Sciences, Institute of Vertebrate Biology, v.v.i., Květná 8, Brno, 60365, Czech Republic. Correspondence and requests for materials should be addressed to M.V. (email: michal.vinkler@natur.cuni.cz)

evolutionary adaptations among polymorphic variants can be predicted[15]. Toll-like receptors (TLRs) act as innate immunity sensors responsible for detection of invading pathogen ligands during early phases of an infection[16]. TLRs are type I transmembrane proteins present either on the cell surface or in the intracellular compartments. They typically consist of a pathogen-recognition horseshoe-shaped ectodomain, a short segment spanning the membrane and an intracellular toll/interleukin-1 receptor (TIR) signalling domain[17].

TLRs are encoded by a multigene family which is only partially conserved across vertebrates, e.g. people and chickens have similar numbers of TLR genes[18], but only four functionally distinct TLRs show direct orthology between both species[19]: endosomal viral-dsRNA-sensing TLR3[20,21]; TLR4 detecting bacterial lipopolysaccharide (LPS) and various other pathogen-derived and host-derived compounds on cell surfaces[22,23]; cell-surface-based bacterial-flagellin-sensing TLR5[24,25]; and endosomal viral-ssRNA-sensing TLR7[26]. The other TLRs may be duplicated (e.g. chicken TLR1 and TLR2[27]), pseudogenised (chicken TLR8[26]) or unique in either of the species (e.g. human TLR9[28]; or chicken TLR15[29] and TLR21[30]). Although human TLR7 and TLR8 are closely related, they slightly differ in their natural ligand preferences[31–34].

Although usually unable to avoid expression of TLR ligands, pathogens in many cases have succeeded in evolving structural modifications that impair recognition by TLRs[35]. Co-evolution with pathogens can then select for diversification in TLR alleles through specific adaptations to ligand variants[15]. Accordingly, most parts of the TLR molecule remain highly conservative due to purifying selection, while other parts, such as the ligand-binding regions, exhibit striking variability, both at the interspecific and intraspecific levels[36–40]. This variation could affect disease resistance[41,42].

In this study, we compare genetic variability and evolutionary patterns in *TLR3*, *TLR4*, *TLR5* and *TLR7* in humans (the only other species with large-sample intraspecific TLR diversity data publicly available), represented by 25 world-wide populations, and domestic chickens, represented by 25 traditional breeds. Information on sequence variation in these receptors is used to show differences in levels of potentially functional variation and the number of sites under positive selection between humans and domestic chickens. Furthermore, we also compare data on allele frequencies and allele sharing. Besides this, we examined the patterns of *TLR* variation with respect to a neutral mitochondrial marker and linked population structure in chicken *TLRs* (*chTLR*) to neutral population structure based on 19 microsatellites (allowing us comparison with previously published evidence for chicken breeds). In doing so, this study provides a pioneering insight into understanding the remarkable levels of variation in non-human immunogenetics when investigated outside traditional commercial and inbred models.

Results

Comparison of genetic variation in human and chicken *TLRs*. We sequenced complete protein-coding DNA sequences (CDSs) for all *TLRs* with direct orthologues between mammals and birds (*TLR3*, *TLR4*, *TLR5* and *TLR7*) in 110 chickens (25 breeds) to compare variability with a corresponding set of 110 randomly-selected humans sampled from 25 populations around the world as part of the 1000 Genomes Project[43]. For a complete description of SNVs detected in chicken *TLRs* (*chTLR*), including their frequencies, see Supplementary Table S1. In most *chTLRs* (*TLR3*, *TLR4* and *TLR7*) we observed between 2.2× and 4× more single nucleotide variants (SNVs) than in human *TLRs* (*hTLRs*), with 17 SNVs in *hTLR3* vs 38 SNVs in *chTLR3*; 9 SNVs in *hTLR4* vs 27 SNVs in *chTLR4*; and 5 SNVs in *hTLR7* vs 20 SNVs in *chTLR7* (Fig. 1a, Table 1). Only in *TLR5* did we detect slightly more SNVs in *hTLR5* (22 SNVs) than in *chTLR5* (19 SNVs). Interestingly, 86% of the *hTLR5* SNVs were very rare variants with frequencies below 5%, while only 58% of the SNVs were below 5% frequency in *chTLR5* (Fig. 2). This is reflected in the nucleotide diversity (π), which was 2.5× higher in *chTLR5* than *hTLR5* (Fig. 1a, Table 1). While the frequency of SNVs in all *hTLR* genes was highly skewed, those found in *chTLRs* had a more equal variant representation and more SNVs of medium frequencies. This was also true for non-synonymous single nucleotide variants (nsSNVs) potentially affecting TLR structure and function (Fig. 2). The higher number of SNVs and more equal variant frequency meant that *chTLRs* exhibited nucleotide diversity up to 21.5× higher than *hTLRs* (Table 1). Unlike *hTLRs*, we found no sequences with internal STOP codons in *chTLRs*. In contrast to *TLRs*, part of the mitochondrial control region previously used as a neutral marker of variability in humans (hypervariable segment I; *HVS-I*) harboured 3.7× more SNVs in humans than chickens (Fig. 1a). Our results do not appear to be affected by the particular selection of the 110 individuals in the chicken or human populations. In chickens, we tested the effect of variation in the numbers of individuals across the breeds on those breeds where we have at least 6 animals per breed represented in the data set (six breeds: Araucana AR, Brahma BH, Czech Golden Pencilled CZ, La Flèche LF, Rosecomb Bantam RO, and Sebright SE). The results show that the average chicken nucleotide diversity does not importantly change with the number of individuals used per breed (2, 4 or 6): *chTLR3* $\pi_{110} = 0.00302$, $\pi_{(2\,ind\,\times\,6\,br)} = 0.00300$, $\pi_{(4\,ind\,\times\,6\,br)} = 0.00296$, $\pi_{(6\,ind\,\times\,6\,br)} = 0.00301$; *chTLR4* $\pi_{110} = 0.00258$, $\pi_{(2\,ind\,\times\,6\,br)} = 0.00241$, $\pi_{(4\,ind\,\times\,6\,br)} = 0.00250$, $\pi_{(6\,ind\,\times\,6\,br)} = 0.00244$; *chTLR5* $\pi_{110} = 0.00111$, $\pi_{(2\,ind\,\times\,6\,br)} = 0.00122$, $\pi_{(4\,ind\,\times\,6\,br)} = 0.00114$, $\pi_{(6\,ind\,\times\,6\,br)} = 0.00117$; *chTLR7* $\pi_{110} = 0.00106$, $\pi_{(2\,ind\,\times\,6\,br)} = 0.00099$, $\pi_{(4\,ind\,\times\,6\,br)} = 0.00100$, $\pi_{(6\,ind\,\times\,6\,br)} = 0.00096$. Also increasing the human dataset to the full sample of 2504 people represented in the 1000 Genomes Project did not importantly alter our estimates of human *TLR* nucleotide diversity: *hTLR3* $\pi_{110} = 0.00044$ vs. $\pi_{2504} = 0.00042$, *hTLR4* $\pi_{110} = 0.00012$ vs. $\pi_{2504} = 0.00016$, *hTLR5* $\pi_{110} = 0.00044$ vs. $\pi_{2504} = 0.00045$, and *hTLR7* $\pi_{110} = 0.00018$ vs. $\pi_{2504} = 0.00019$.

Location and physicochemical properties of the coding variants. While nsSNVs were more evenly distributed across genes in *hTLRs*, almost all nsSNVs (with three exceptions, two in *chTLR3* and one in *chTLR7*) in *chTLRs* were located in the ligand-binding ectodomains (see Supplementary Fig. S1 and Supplementary Table S1). Based on 3D model measurements, thirteen chicken amino acid substitutions neighboured the predicted functional sites of TLRs with topological proximities of less than 10 Å (Fig. 3, Supplementary Table S1). Twelve nsSNVs found in *chTLR3*, six in *chTLR4*, six in *chTLR5* and two in *chTLR7* resulted in important differences in residue physicochemical properties, and hence may substantially influence the resultant ligand-binding features
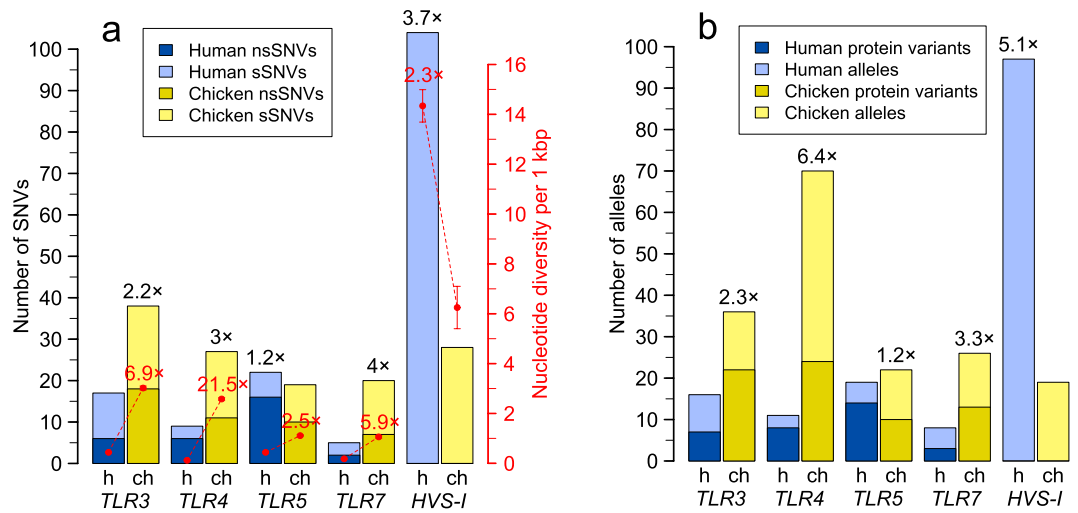
2

Figure 1. Genetic variability in human and chicken Toll-like receptors (*TLRs*) and hypervariable segment I (*HVS-I*) of the mitochondrial DNA. (**a**) Comparison of numbers of single nucleotide variants (SNVs; left axis) and nucleotide diversity (right axis; red) found in human (h; blue) and chicken (ch; yellow) *TLRs* and *HVS-I*. Synonymous variants are shown in light colours and non-synonymous variants in dark colours. (**b**) Comparison of numbers of alleles and protein variants of *TLRs* found in human (h; blue) and chicken (ch; yellow) populations and number of alleles found in *HVS-I*. Total numbers of TLR protein variants are highlighted in dark colours, while numbers of alleles are represented by the total height of each bar (combination of light and dark colours). Numbers above the bars and points indicate the fold differences between values for humans and chickens. Error bars denote standard deviations (in *TLR* genes the bars are too short to be visible).

Gene	Sp	L (bp)	SNVs	nsSNVs	k	$\pi \pm$ SD	A	PV	Hd $\pm$ SD
TLR3	h	2715	17	6	1.187	0.00044 ± 0.00002	16	7	0.759 ± 0.016
	ch	2691	38	18	8.120	0.00302 ± 0.00008	36	22	0.903 ± 0.011
TLR4	h	2520	9	6	0.310	0.00012 ± 0.00002	11	8	0.244 ± 0.038
	ch	2532	27	11	6.540	0.00258 ± 0.00005	70	24	0.947 ± 0.010
TLR5	h	2577	22	16	1.143	0.00044 ± 0.00003	19	14	0.709 ± 0.025
	ch	2586	19	10	2.866	0.00111 ± 0.00004	22	10	0.854 ± 0.013
TLR7	h	3150	5	2	0.566	0.00018 ± 0.00002	8	3	0.481 ± 0.037
	ch	3180	20	7	3.363	0.00106 ± 0.00005	26	13	0.849 ± 0.017

Table 1. Diversity statistics of human and chicken *TLRs*. Sp: species (h: human, ch: chicken); L (bp): sequence length in base pairs; SNVs: number of single nucleotide variants; nsSNVs: number of non-synonymous single nucleotide variants; k: average number of nucleotide differences between two sequences; π: nucleotide diversity (average number of nucleotide differences per site between two sequences); SD: standard deviation; A: number of alleles; PV: number of protein variants; Hd: haplotype diversity.

of the receptors (Fig. 3, Supplementary Table S1). Four amino acid substitutions in chTLR3, one in chTLR4, two in chTLR5 and one in chTLR7 were predicted by the PROVEAN analysis (score lower than the cut-off of −2.5) to affect the biological functions of the receptors (Fig. 3, Supplementary Tables S1 and S2). All of them were located in the ligand-binding ectodomain, but only three lied close to the predicted functional sites (<10 Å). Five of these non-conservative substitutions were consistent with the above-mentioned radical substitutions changing the physicochemical properties of the molecules. This was different than in hTLRs, where out of the six non-conservative substitutions detected by PROVEAN (one in hTLR3, one in hTLR4 and five in hTLR5), only three were located in the ectodomain, while the other three lied in the signalling TIR domain. Two of these substitutions were STOP codons (one in hTLR4 and one in hTLR5), three radically changed the residuum properties and only one was conservative (Supplementary Table S2).

Detection of positive selection acting on *TLRs*. In total, we identified five sites in chTLR4 (A26V, R261K, K343R, Y383H and Q611R) and two in chTLR7 (I121V and S447G) that were under significant positive selection (IFEL: p < 0.05; FUBAR: posterior probability > 0.95; Supplementary Table S3). Two sites in chTLR4 (Y383H and Q611R) and the two sites in chTLR7 were recognised based on the two independent statistical approaches. In addition, we found three sites in chTLR3 (S17F, S399R, K513R), one site in chTLR4 (S23C) and two sites in chTLR5 (S409T and A619E) where positive selection was detected with only marginal non-significance (IFEL: 0.05 < p < 0.1; FUBAR: posterior probability 0.90 < p < 0.95). In contrast, we identified only two putatively
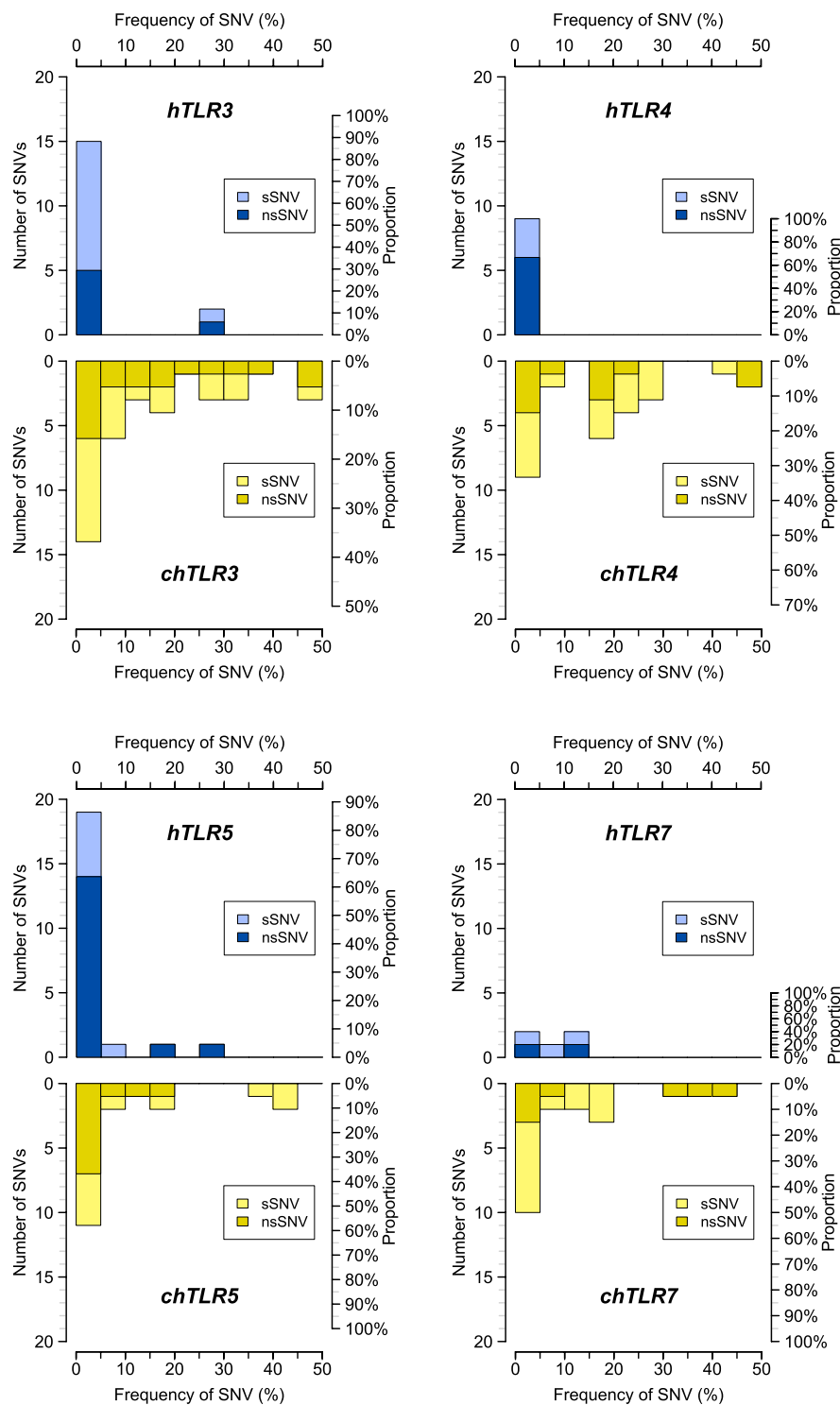
Figure 2. Comparison of minor variant frequencies in human and chicken *TLR* single nucleotide variants (SNVs). Human *TLRs* (*hTLRs*) are highlighted in blue and chicken *TLRs* (*chTLRs*) in yellow, while allele frequencies of equivalent TLRs are shown specularly. Synonymous variants are shown in light colours and non-synonymous variants in dark colours. The left axis shows number of SNVs, while the right axis indicates the proportion of SNVs on a relative scale.

positively selected sites in hTLRs (hTLR4: T399I; hTLR5: I644F), both being marginally non-significant in the tests. All positively selected sites in both species were located in the ligand-binding ectodomains, with the exception of I644F in hTLR5 which lay in the transmembrane domain (Fig. 3). None of the positively selected sites was
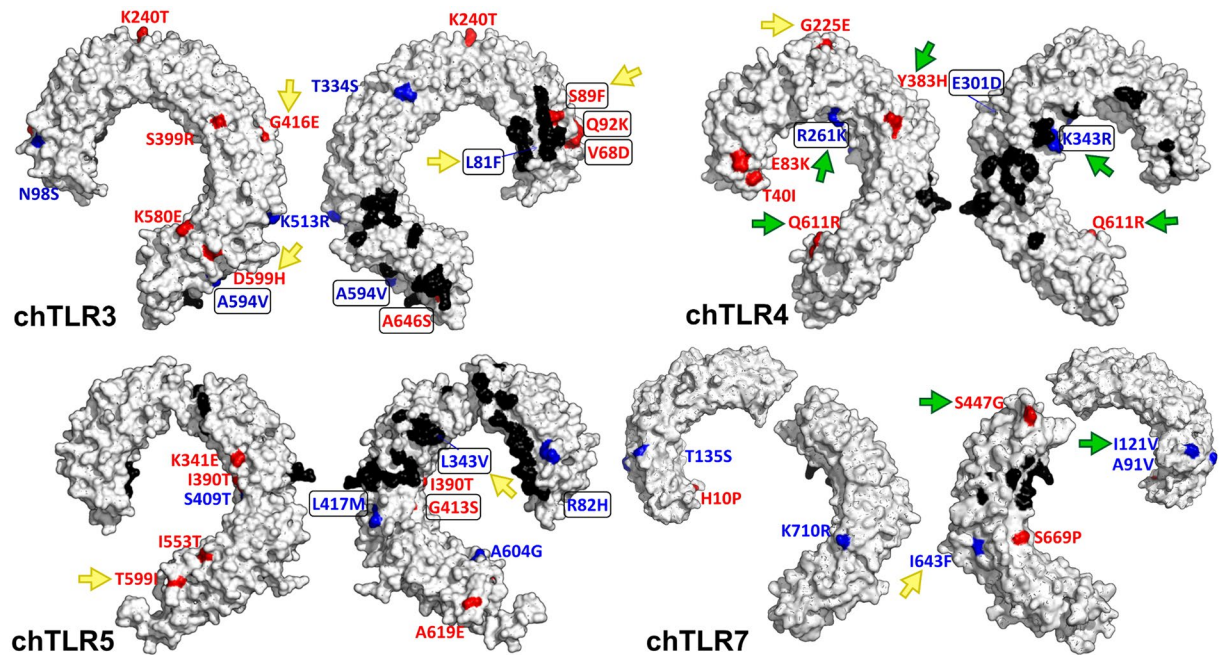
Figure 3. Projection of amino acid substitutions onto 3D models of chicken TLR ectodomains. Radical alternations are shown in red, conservative in blue and functionally important sites known from mammalian[68,90,91] or fish[92] studies are highlighted in black. Substitutions closer than 10 Å to functional sites are highlighted by a rectangle, non-conservative sites (PROVEAN) are indicated with yellow and positively selected sites (IFEL, FUBAR) with green arrows.

consistent with the non-conservative sites with putatively dramatic functional effects detected by the PROVEAN analysis (Supplementary Table S2 and S3).

Allelic variability in *TLRs* and its distribution across chicken breeds.   We identified $1.2\times$ to $6.4\times$ higher levels of *TLR* haplotype diversity (Hd) in chickens than in humans (Fig. 1b, Table 1; for a complete list of *chTLR* alleles and corresponding protein variants see Supplementary Table S4). Only in *hTLR7* (site 665), *chTLR3* (sites 495 and 1781), *chTLR4* (site 903) and *chTLR5* (site 2001) was there significant evidence for recombination between alleles ($p < 0.01$; confirmed by both SBP and GARD methods). Species-specific difference in allele number had a clear impact on the shape of the haplotype networks, with networks constructed from *hTLR* sequences exhibiting star-like patterns and networks constructed from *chTLR* sequences showing far more complex patterns (Supplementary Fig. S2). Haplotype diversity ranged from 0.244 to 0.759 in *hTLRs*, being in all cases lower than in *chTLRs* (Hd range 0.849–0.947; Table 1). The variation of individual human populations is generally low over all the receptors (Fig. S2). Similar to the nucleotide diversity, also in the case of haplotype diversity our results do not appear to be affected by the particular selection of the 110 individuals in the chicken or human population. In chickens, the effect of variation in the numbers of individuals across the breeds (the same six breeds as above) on the haplotype diversity was minor: *chTLR3* $\mathrm{Hd}_{110} = 0.903$, $\mathrm{Hd}_{(2\,\mathrm{ind}\times6\,\mathrm{br})} = 0.877$, $\mathrm{Hd}_{(4\,\mathrm{ind}\times6\,\mathrm{br})} = 0.892$, $\mathrm{Hd}_{(6\,\mathrm{ind}\times6\,\mathrm{br})} = 0.885$; *chTLR4* $\mathrm{Hd}_{110} = 0.947$, $\mathrm{Hd}_{(2\,\mathrm{ind}\times6\,\mathrm{br})} = 0.913$, $\mathrm{Hd}_{(4\,\mathrm{ind}\times6\,\mathrm{br})} = 0.941$, $\mathrm{Hd}_{(6\,\mathrm{ind}\times6\,\mathrm{br})} = 0.942$; *chTLR5* $\mathrm{Hd}_{110} = 0.854$, $\mathrm{Hd}_{(2\,\mathrm{ind}\times6\,\mathrm{br})} = 0.855$, $\mathrm{Hd}_{(4\,\mathrm{ind}\times6\,\mathrm{br})} = 0.864$, $\mathrm{Hd}_{(6\,\mathrm{ind}\times6\,\mathrm{br})} = 0.863$; *chTLR7* $\mathrm{Hd}_{110} = 0.849$, $\mathrm{Hd}_{(2\,\mathrm{ind}\times6\,\mathrm{br})} = 0.855$, $\mathrm{Hd}_{(4\,\mathrm{ind}\times6\,\mathrm{br})} = 0.825$, $\mathrm{Hd}_{(6\,\mathrm{ind}\times6\,\mathrm{br})} = 0.824$, and increasing the human dataset to the full sample of 2504 people represented in the 1000 Genomes Project did not importantly alter our estimate of the human *TLR* haplotype diversity: *hTLR3* $\mathrm{Hd}_{110} = 0.759$ vs. $\mathrm{Hd}_{2504} = 0.747$, *hTLR4* $\mathrm{Hd}_{110} = 0.244$ vs. $\mathrm{Hd}_{2504} = 0.289$, *hTLR5* $\mathrm{Hd}_{110} = 0.709$ vs. $\mathrm{Hd}_{2504} = 0.711$, and *hTLR7* $\mathrm{Hd}_{110} = 0.481$ vs. $\mathrm{Hd}_{2504} = 0.481$.

In *chTLRs*, many alleles were shared between chicken breeds (Supplementary Fig. S2 and S3 and Supplementary Table S4). Despite this allele sharing, distinguishable population structure was revealed based on both neutral loci (19 microsatellites) and *TLRs* alleles (Supplementary Fig. S4). Unlike microsatellites, however, variation in the allelic frequency of *chTLRs* did not allow detailed resolution of breed identity with increasing number of dividing groups (K).

Convergent evolution in *chTLRs*.   We detected signals of convergent evolution in *chTLR* alleles, giving origin to six protein variants in chTLR4 and two in chTLR7 (Fig. 4, Supplementary Table S5). In contrast, no signs of convergence were identified in human *TLRs* or any other chicken *TLRs*.

Discussion

Immune-related genes involved in host defence are remarkable for their high levels of genetic polymorphism. Although typically investigated in genes of the vertebrate Major Histocompatibility Complex *(MHC)*, innate immune receptors may also harbour striking levels of interspecific as well as intraspecific evolutionarily adaptive variation. Our results indicate that while this pattern remains elusive in human innate immune genes it can be
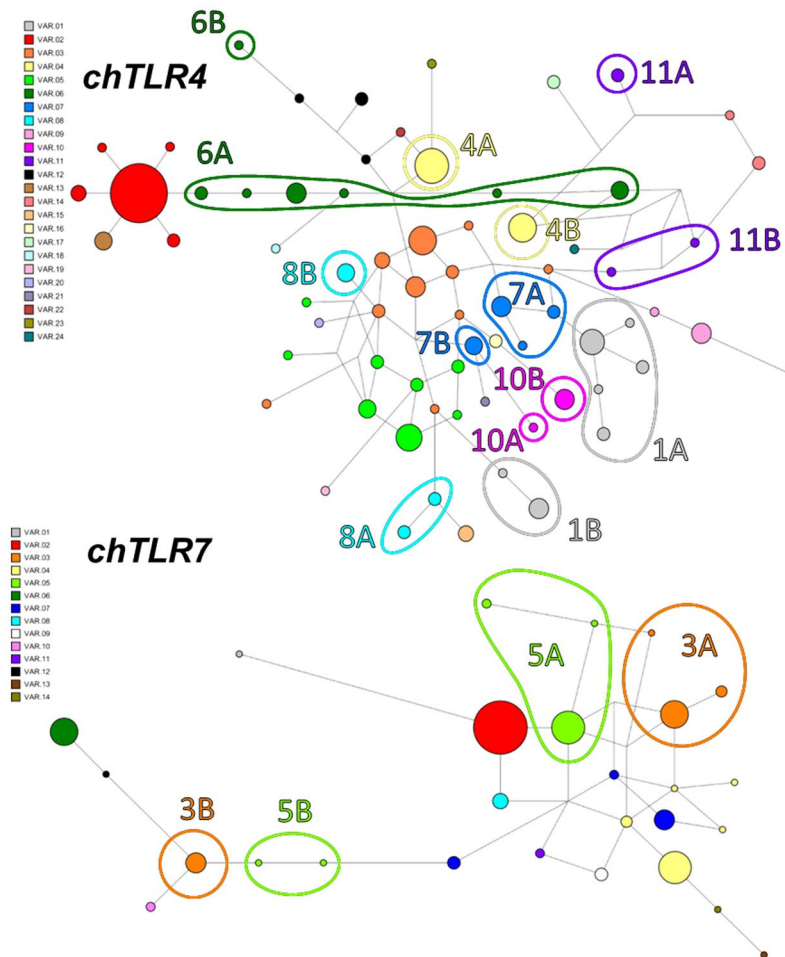
Figure 4. Convergent evolution in chicken *TLR4* and *TLR7*, as indicated by identical protein variants originating from unrelated alleles. Alleles encoding the same protein variant are marked in the haplotype network with the same colour. Distinct clusters of convergent alleles are numbered, highlighted with circles and marked with A or B. (For distances between clusters see Supplementary Table S5).

clearly documented in domestic fowl. Since their domestication from wild red junglefowl *(Gallus gallus)* at least 4000 years ago[44,45], chickens have evolved in a shared environment with humans[46], coming into contact with joint microbial challenges including viral (e.g. avian influenza viruses[47]) or bacterial infections (e.g. *Salmonella*[48], *Campylobacter*[49]; *Helicobacter*[50,51] and others[52]). However, probably due to differences in their ancestral populations and distinct modes of recent selection, the two species are differentially equipped to face these threats. Therefore, chickens and humans may respond very differently to the same set of pathogens that infect both species (e.g. when infected with gastrointestinal pathogens[53]). Here we show that chicken *TLR (chTLR)* genes from population sets of comparable sizes exhibit up to 20× higher nucleotide diversity in up to 6× more alleles than human *TLRs (hTLRs)*. As an example, Georgel *et al.* detected only four *TLR4* alleles in a French human population[54], compared with the 70 *TLR4* alleles now detected in European chickens.

Previous genomic research has shown that, in general, the chicken genome is about 2× richer in exon polymorphism than the human genome[13,14]. Interestingly, we found even higher levels of genetic variation in chicken *TLRs*, with up to 4× more single nucleotide variants (SNVs) in *chTLRs* than *hTLRs*. The levels of human *TLR* nucleotide diversity indicated in our study (0.00012–0.00044) are fully comparable to those previously reported by Barreiro *et al.* (0.00027–0.00071)[55]. While the nucleotide diversity we report for *chTLRs* (0.00106–0.00302) is about 30× lower than that in the variable region of the human (0.024–0.071)[56], chicken (0.0371–0.0400)[57] and red junglefowl *MHC* (0.068–0.101)[58], it is still far higher than either the gene average[59] or levels typical for non-coding regions in humans[60]. Thus, though the distinct patterns of *TLR* variation observed in humans and chickens partially reflect the general features of their genomes, the contrast is higher than expected. This is further supported by the contrasting trend in mitochondrial *HVS-I* variation, which is almost 4× richer in human SNVs. Interestingly, it was in this same *HVS-I* region that the lower genetic variation of humans compared to other primates was first documented[61]; yet here we observe higher variation in humans than chickens. Our results show that while inter-breed differences (not tested in this study) change with increasing sample size within the breeds, the above-mentioned difference between species described in our dataset holds true even with changing sample sizes used for the analysis. This importantly supports the plausibility of our conclusions. However, we do

not claim that the difference we found between human and chicken *TLRs* represents a general difference between birds and mammals. In wild house mouse *TLR4* the levels of nucleotide diversity were previously found intermediate to the values detected in this study for humans and chickens[62].

TLR5, with 22 human SNVs and 19 chicken SNVs, appears to break the general trend outlined above. In order to fully comprehend *TLR5* variability, however, it is important to consider SNV frequency. Since 86% of *hTLR5* variability comprises very rare variants, final nucleotide diversity (π) in *TLR5* is actually more than $2\times$ higher in chickens than humans. This pattern suggests weaker negative selection in *hTLR5* compared to other *hTLRs*. In agreement with this hypothesis, STOP-codon-encoding substitutions have been identified in *hTLR5*[63] and some avian species[64]. We found no *chTLR5* pseudogenes, suggesting a conservative essential function in chickens.

Only around 60% of chicken protein-coding genes are known to possess a single human orthologue[65]. There are several *TLRs* with orthologues missing or with several paralogues[66]. Therefore, to provide a biologically meaningful human-chicken comparison, the *TLR* genes investigated in this study were selected to represent direct orthologues with confirmed and conserved ligand-specificity. In chickens, most nsSNVs were located in regions encoding the ectodomains, suggesting a potential significance for TLR ligand-binding specificity. We recorded 26 non-synonymous (ns) SNVs that have an important impact on residue physicochemical properties, and 13 amino acid substitutions topologically neighbouring receptor functional sites. Of these, five amino acid substitutions (V68D, S89F, Q92K and A646S in chTLR3 and G413S in chTLR5) belong to both these groups and represent candidate sites putatively altering ligand specificity. Furthermore, the S89F variation in TLR3 was identified as a non-conservative substitution by the PROVEAN analysis.

The functional significance of some nsSNVs in chickens is further supported by the results of selection analysis. In contrast to hTLRs, we identified significant positive selection acting on the ligand-binding ectodomains in chTLR4 and chTLR7. While the functional significance of the putative positively selected TLR7 positions remains unclear, there are several candidate sites in TLR4 with high support for a functional effect. In particular, sites 261 and 343 are topological neighbours to the predicted MD-2-dimmerisation and LPS-binding residues[67,68]. Site 343 has previously been identified as under positive selection at the interspecific level in Galloanserae[40], with positive selection at neighbouring sites also being reported in other avian and mammalian taxa[39,69,70]. Positions 383 and 611 in chTLR4, which harbour non-conservative substitutions that have been recognised as positively selected using several independent approaches (despite lying out of the predicted functional sites), also appear to be relevant. Sites adjacent to position 611 have also been reported as under selection in other vertebrates[39,69]. Finally, residues 383 and 611 have been identified as responsible for differences in salmonellosis resistance in chickens[42].

Ongoing positive selection, therefore, appears stronger in chickens than humans, as supported by the distinct SNV frequency and allelic variability patterns in *chTLRs* and *hTLRs*. Since both SNVs and alleles show highly skewed frequency distributions in *hTLRs*, while those in *chTLRs* exhibit more equal SNV and allele representation, we assume that negative selection on *TLRs* is stronger in humans, while *chTLRs* appear to be under stronger balancing selection. From this perspective, *chTLRs* are closer in their co-evolution with pathogens to *MHC* than *hTLRs*[71]. Additional support for this view can be gained from the star-like *hTLR* haplotype network patterns (typical for post-selective-sweep populations). In contrast, the complex *chTLR* net-like patterns are indicative of either recombination or convergence. Correspondingly, haplotype diversity is low in *hTLRs*, while the levels in *chTLRs* resemble those of *MHC*[57]. Despite the generally higher recombination rates reported for the chicken genome[65], our evidence of only four recombination breakpoints in 154 *TLR* alleles does not suggest that high levels of *chTLR* allelic variability were gained through recombination. Hence, convergent evolution may have been more important in *chTLR* evolution, especially in *TLR4*.

Of the total number of 154 *chTLR* alleles in four genes, 100 did not occur in more than one breed. Most of these 100 were low frequency alleles. In contrast, alleles that are more frequent are largely shared between breeds, which may have resulted from ancestral polymorphism maintained across selective breeding events or from inter-breed gene flow. Nevertheless, population structure is distinguishable, with both neutral microsatellites and *TLR*s suggesting differentiation in the breeds. This confirms previous reports of differentiation in chicken breeds using microsatellites[8], though the shared polymorphism in *chTLRs* seems to be higher than in neutral markers. Apparently, some of the investigated chicken breeds possess high genetic diversity that could not be reliably fully captured with the limited sample used in this study. Therefore, this study has no ambition to compare the breeds one to another. Yet, recognising this phenomenon of inter-breed variation in immunogenetic variability, our study opens the way to further targeted research in larger population datasets.

The emergence of the domestic chicken was associated with inter(sub)specific hybridisation[44,72,73] increasing their genetic diversity, similarly to humans[74]. Through their recent evolutionary history, both humans and chickens have been selected for survival under similar novel pathogenic environments. Unlike humans, that have been shown to exhibit relatively low levels of genetic variation[61], chickens may be more diversified given their heterogeneous origin and artificial selection for diversity based on human cultural and economic needs. Taken together, our study documents that the broad spectrum of geographically distinct, locally-adapted chicken populations maintains tremendous immunogenetic variation in *TLRs*. Shaped by natural selection, the diversity shown by chicken *TLRs* far exceeds that in human *TLRs*, providing also an insight into the genetic basis of breed-specific variation in resistance to various infectious diseases[75]. While the data from chickens and humans cannot be directly generalised to birds and mammals, our results highlight that there are important interspecific differences in the levels of putatively functional genetic polymorphism in the innate immune receptors across vertebrates.

Methods

Genetic samples.   Genetic samples of 110 chickens representing 25 randomly chosen traditional domestic breeds were provided by hobby breeders from the Czech Republic, Slovakia, Germany, France and Italy. The breeds selected were reported to have their origin in the Indian subcontinent, southeast Asia, Europe and South America[76,77]. Although the relatedness of these breeds is unknown and the pedigree information is missing, the

breed-flock variation secured sampling of unrelated individuals. For each breed, 1–14 tissue (blood, feather or muscle) samples (average n = 4.4) were collected and stored in 96% ethanol at −20 °C. After analysis, all samples were deposited in the Genetic Bank of the Department of Zoology, Charles University, Prague (GRbio Institution Code: ZCU). For detailed information on the samples used see Supplementary Dataset 1. This research was carried out in accordance with Czech legislation (Act No 246/1992 Coll., on the protection of animals against cruelty) and approved by the Ethical Committee of the Faculty of Science, Charles University (Reference no. 34712/2010–30).

Molecular analysis of sequence variation in selected chicken genes. For sequencing of complete protein-coding DNA sequences (CDSs) of four chicken *TLRs* (*chTLRs*; 2691 bp of *chTLR3*, 2532 bp of *chTLR4*, 2586 bp of *chTLR5* and 3180 bp of *chTLR7*) and 1–521 bp of hypervariable segment I (*HVS-I*; reference GenBank ID NC_001323.1) of the mitochondrial DNA control region describing the main chicken mitochondrial haplogroups[44] we used Sanger sequencing that provides longer reads than most NGS techniques. Alleles were identified using PHASE[78] and verified by cloning. Unique *chTLR* allele sequences were deposited in GenBank (IDs listed in Supplementary Table S4). For detailed protocols see Supplementary Methods online.

Sequence variation in selected human genes. Variability of human *TLRs* (*hTLRs*) and *HVS-I* was assessed based on sequence data available through the 1000 Genomes Project Phase 3[43], downloaded using the Data Slicer Tool. We extracted all *hTLR* variable sites in CDSs for all individuals. Due to the necessity of including two alleles of *hTLR7*, which is localised on the X chromosome in humans, only females were pre-filtered. A subset of 110 individuals from 25 populations was then randomly selected in R v. 3.2.1[79]. As with chickens, the population sample size varied around an average of n = 4.4 individuals per population. For detailed information on the human sequences used see Supplementary Methods online.

Genetic diversity measurements in sequence data. Nucleotide diversity (π), average number of nucleotide differences (k) and haplotype diversity (Hd) for all selected *TLRs* and *HVS-I* were calculated in DnaSP v. 5.10.01[78]. Haplotype networks for all *TLRs* were constructed in Network v. 4.6.1.2. (Fluxus Technology) using a median-joining algorithm[80]. Additional properties, such as breed/population and *TLR* protein variant, were visualised in Network publisher v. 2.0.0.1 (Fluxus Technology). Alignments were screened for recombination breakpoints using the SBP (Single Breakpoint) and GARD (Genetic Algorithm for Recombination Detection) tools, available on the Datamonkey server[81].

Prediction of the effects of amino acid substitutions on protein function. PROVEAN (Protein Variation Effect Analyzer) v 1.1[82] was used to predict the functional effect of the amino acid substitutions detected in the TLRs in our data set. The algorithm blasted and used between 181 and 265 homologous sequences from the NCBI NR protein database to predict the level of conservation of the individual sites in the particular TLRs. As a threshold for the effect on biological function we took the default value of delta alignment score of −2.5.

Selection analysis. Positive selection was tested using tools available on the Datamonkey server[83]. Individual sites under diversifying and purifying selection were detected using the codon-based maximum likelihood methods, FUBAR[84] and IFEL. Results with P-values < 0.05 (IFEL) or posterior probabilities > 95% (FUBAR) were considered significant.

Protein structures. In order to ascertain the precise location of the variation in proteins, we adopted the approach previously described by Vinkler *et al*.[40]. Briefly, amino acid sequences coding the ligand-binding ectodomains of *chTLRs* were used to generate 3D models in the I-TASSER on-line tool[85]. Since proteolytic cleavage between leucine rich repeat 14 and 15 in the TLR7 endosomal domain has been shown to be essential for murine activation[86] and human[87] TLR7, we modelled the two parts of TLR7 separately. Models with highest C-scores were selected for analysis. Functional sites reported in the literature were visualised together with the variable sites, the distances between them being measured using the PyMOL Molecular Graphics System v. 1.5 (Schrödinger, LLC). The accuracy of the 3D models was checked using the ModFOLD Model Quality Assessment Server v. 4.0[88]. All structures achieved high levels of confidence at P < 0.05 and Global model quality scores > 0.384. Details on amino acid classification adopted are provided in Supplementary Methods online.

Analysis of chicken population structure. The sample set of 110 chickens were genotyped for 19 unlinked microsatellite markers. Genetic population structure for both *chTLR* alleles and microsatellites was determined based on the Bayesian clustering approach implemented in STRUCTURE v. 2.3.4[89]. For detailed description of our approach see Supplementary Methods online.

Data Availability
Unique *chTLR* allele sequences were deposited in GenBank (GenBank IDs KU235138- KU235484).

References
1. Darwin, C. *The Variation of Animals and Plants Under Domestication*. (Murray J., 1868).
2. Huang, Y. Q. *et al*. Haplotypic variation and characteristics across the toll-like receptor 3 locus in chickens. *Anim. Genet.* **43**, 343–347 (2012).
3. Ruan, W., Wu, Y. & Zheng, S. J. Different genetic patterns in avian Toll-like receptor (TLR)5 genes. *Mol. Biol. Rep.* **39**, 3419–3426 (2012).
4. Ruan, W., An, J. & Wu, Y. Polymorphisms of Chicken TLR3 and 7 in Different Breeds. *PLoS ONE* **10**, e0119967 (2015).
5. Ruan, W. K., Wu, Y. H., An, J. & Zheng, S. J. Polymorphisms of chicken Toll-like receptors 4, 15, and 21 in different breeds. *Poult. Sci.* **91**, 2512–2516 (2012).

6. Hillel, J. *et al*. Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. *Genet. Sel. Evol. GSE* **35**, 533–557 (2003).
7. Muir, W. M. *et al*. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *Proc. Natl. Acad. Sci.* **105**, 17312–17317 (2008).
8. Lyimo, C. M. *et al*. Global diversity and genetic contributions of chicken populations from African, Asian and European regions. *Anim. Genet.* **45**, 836–848 (2014).
9. Lyimo, C. M. *et al*. Maternal genealogical patterns of chicken breeds sampled in Europe. *Anim. Genet.* **46**, 447–451 (2015).
10. Weigend, S. & Romanov, M. n. Current strategies for the assessment and evaluation of genetic diversity in chicken resources. *Worlds Poult. Sci. J.* **57**, 275–288 (2001).
11. Delany, M. E. Genetic diversity and conservation of poultry. In *Poultry genetics, breeding, and biotechnology* (eds. Muir, W. M. & Aggrey, S. E.) 257–281 (CABI Pub, 2003).
12. Groeneveld, L. F. *et al*. Genetic diversity in farm animals – a review. *Anim. Genet.* **41**, 6–31 (2010).
13. Sachidanandam, R. *et al*. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
14. Wong, G. K.-S. *et al*. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432**, 717–722 (2004).
15. Vinkler, M. & Albrecht, T. The question waiting to be asked: Innate immunity receptors in the perspective of zoological research. *Folia Zool.* **58**, 15–28 (2009).
16. Medzhitov, R. Toll-like receptors and innate immunity. *Nat. Rev. Immunol.* **1**, 135–145 (2001).
17. Kang, J. Y. & Lee, J.-O. Structural Biology of the Toll-Like Receptor Family. In *Annual Review of Biochemistry*, Vol. 80 (eds. Kornberg, R. D., Raetz, C. R. H., Rothman, J. E. & Thorner, J. W.) **80**, 917–941 (Annual Reviews, 2011).
18. Velová, H., Gutowska-Ding, M. W., Burt, D. W. & Vinkler, M. Toll-like receptor evolution in birds: gene duplication, pseudogenisation and diversifying selection. *Mol. Biol. Evol.* **39**, 2170–2184 (2018).
19. Cormican, P. *et al*. The avian Toll-Like receptor pathway–subtle differences amidst general conformity. *Dev. Comp. Immunol.* **33**, 967–973 (2009).
20. Karpala, A. J., Lowenthal, J. W. & Bean, A. G. Activation of the TLR3 pathway regulates IFNβ production in chickens. *Dev. Comp. Immunol.* **32**, 435–444 (2008).
21. Schwarz, H. *et al*. Chicken toll-like receptor 3 recognizes its cognate ligand when ectopically expressed in human cells. *J. Interferon Cytokine Res.* **27**, 97–101 (2007).
22. Keestra, A. M. & Putten, J. P. M. van. Unique Properties of the Chicken TLR4/MD-2 Complex: Selective Lipopolysaccharide Activation of the MyD88-Dependent Pathway. *J. Immunol.* **181**, 4354–4362 (2008).
23. Kogut, M. H. *et al*. Expression and function of Toll-like receptors in chicken heterophils. *Dev. Comp. Immunol.* **29**, 791–807 (2005).
24. Iqbal, M. *et al*. Identification and functional characterization of chicken toll-like receptor 5 reveals a fundamental role in the biology of infection with Salmonella enterica serovar typhimurium. *Infect. Immun.* **73**, 2344–2350 (2005).
25. Keestra, A. M., de Zoete, M. R., van Aubel, R. A. M. H. & van Putten, J. P. M. Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin. *Mol. Immunol.* **45**, 1298–1307 (2008).
26. Philbin, V. J. *et al*. Identification and characterization of a functional, alternatively spliced Toll-like receptor 7 (TLR7) and genomic disruption of TLR8 in chickens. *Immunology* **114**, 507–521 (2005).
27. Temperley, N. D., Berlin, S., Paton, I. R., Griffin, D. K. & Burt, D. W. Evolution of the chicken Toll-like receptor gene family: A story of gene gain and gene loss. *BMC Genomics* **9**, 62 (2008).
28. Yilmaz, A., Shen, S., Adelson, D. L., Xavier, S. & Zhu, J. J. Identification and sequence analysis of chicken Toll-like receptors. *Immunogenetics* **56**, 743–753 (2005).
29. Higgs, R. *et al*. Induction of a novel chicken Toll-like receptor following Salmonella enterica serovar Typhimurium infection. *Infect. Immun.* **74**, 1692–1698 (2006).
30. Roach, J. C. *et al*. The evolution of vertebrate Toll-like receptors. *Proc. Natl. Acad. Sci. USA* **102**, 9577–9582 (2005).
31. Tanji, H., Ohto, U., Shibata, T., Miyake, K. & Shimizu, T. Structural reorganization of the Toll-like receptor 8 dimer induced by agonistic ligands. *Science* **339**, 1426–1429 (2013).
32. Tanji, H. *et al*. Toll-like receptor 8 senses degradation products of single-stranded RNA. *Nat. Struct. Mol. Biol.* **22**, 109–115 (2015).
33. Shibata, T. *et al*. Guanosine and its modified derivatives are endogenous ligands for TLR7. *Int. Immunol.* **28**, 211–222 (2016).
34. Zhang, Z. *et al*. Structural Analysis Reveals that Toll-like Receptor 7 Is a Dual Receptor for Guanosine and Single-Stranded RNA. *Immunity* **45**, 737–748 (2016).
35. Reddick, L. E. & Alto, N. M. Bacteria Fighting Back: How Pathogens Target and Subvert the Host Innate Immune System. *Mol. Cell* **54**, 321–328 (2014).
36. Alcaide, M. & Edwards, S. V. Molecular Evolution of the Toll-Like Receptor Multigene Family in Birds. *Mol. Biol. Evol.* **28**, 1703–1715 (2011).
37. Downing, T., Lloyd, A. T., O'Farrelly, C. & Bradley, D. G. The Differential Evolutionary Dynamics of Avian Cytokine and TLR Gene Classes. *J. Immunol.* **184**, 6993–7000 (2010).
38. Smirnova, I., Poltorak, A., Chan, E. K., McBride, C. & Beutler, B. Phylogenetic variation and polymorphism at the Toll-like receptor 4 locus (TLR4). *Genome Biol.* **1**, research002 (2000).
39. Vinkler, M., Bryjová, A., Albrecht, T. & Bryja, J. Identification of the first Toll-like receptor gene in passerine birds: TLR4 orthologue in zebra finch (*Taeniopygia guttata*). *Tissue Antigens* **74**, 32–41 (2009).
40. Vinkler, M., Bainová, H. & Bryja, J. Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. *Genet. Sel. Evol.* **46**, 72 (2014).
41. Beaumont, C. *et al*. Effect of two candidate genes on the Salmonella carrier state in fowl. *Poult. Sci.* **82**, 721–726 (2003).
42. Leveque, G. *et al*. Allelic Variation in TLR4 Is Linked to Susceptibility to Salmonella enterica Serovar Typhimurium Infection in Chickens. *Infect. Immun.* **71**, 1116–1124 (2003).
43. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
44. Miao, Y.-W. *et al*. Chicken domestication: an updated perspective based on mitochondrial genomes. *Heredity* **110**, 277–282 (2013).
45. Tixier-Boichard, M., Bed'hom, B. & Rognon, X. Chicken domestication: from archeology to genomics. *C. R. Biol.* **334**, 197–204 (2011).
46. Larson, G. & Fuller, D. Q. The Evolution of Animal Domestication. *Annu. Rev. Ecol. Evol. Syst.* **45**, 115–136 (2014).
47. Van Reeth, K. Avian and swine influenza viruses: our current understanding of the zoonotic risk. *Vet. Res.* **38**, 243–260 (2007).
48. Guard-Petter, J. The chicken, the egg and Salmonella enteritidis. *Environ. Microbiol.* **3**, 421–430 (2001).
49. Hermans, D. *et al*. Poultry as a Host for the Zoonotic Pathogen Campylobacter jejuni. *Vector-Borne Zoonotic Dis.* **12**, 89–98 (2012).
50. Stanley, J. *et al*. Helicobacter pullorum sp. nov.-genotype and phenotype of a new species isolated from poultry and from human patients with gastroenteritis. *Microbiol. Read. Engl.* **140**(Pt 12), 3441–3449 (1994).
51. Javed, S., Gul, F., Javed, K. & Bokhari, H. Helicobacter pullorum: An Emerging Zoonotic Pathogen. *Front. Microbiol.* **8**, 604 (2017).
52. Hampson, D. J. The Spirochete Brachyspira pilosicoli, Enteric Pathogen of Animals and Humans. *Clin. Microbiol. Rev.* **31**(UNSP), e00087–17 (2018).
53. Barrow, P. A. The paratyphoid salmonellae. *Rev. Sci. Tech. Int. Off. Epizoot.* **19**, 351–375 (2000).

54. Georgel, P., Macquin, C. & Bahram, S. The Heterogeneous Allelic Repertoire of Human Toll-Like Receptor (TLR) Genes. *PLoS ONE* **4**, (2009).
55. Barreiro, L. B. *et al.* Evolutionary Dynamics of Human Toll-Like Receptors and Their Different Contributions to Host Defense. *PLoS Genet.* **5**, (2009).
56. Buhler, S. & Sanchez-Mazas, A. HLA DNA Sequence Variation among Human Populations: Molecular Signatures of Demographic and Selective Events. *PLOS ONE* **6**, e14643 (2011).
57. Wang, Y. *et al.* Sequence variations of the MHC class I gene exon 2 and exon 3 between infected and uninfected chickens challenged with Marek's disease virus. *Infect. Genet. Evol.* **21**, 103–109 (2014).
58. Worley, K. *et al.* Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. *Immunogenetics* **60**, 233–247 (2008).
59. Li, W. H. & Sadler, L. A. Low nucleotide diversity in man. *Genetics* **129**, 513–523 (1991).
60. Zhao, Z., Yu, N., Fu, Y.-X. & Li, W.-H. Nucleotide Variation and Haplotype Diversity in a 10-kb Noncoding Region in Three Continental Human Populations. *Genetics* **174**, 399–409 (2006).
61. Gagneux, P. *et al.* Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc. Natl. Acad. Sci.* **96**, 5077–5082 (1999).
62. Fornuskova, A., Bryja, J., Vinkler, M., Macholán, M. & Piálek, J. Contrasting patterns of polymorphism and selection in bacterial-sensing toll-like receptor 4 in two house mouse subspecies. *Ecol. Evol.* **4**, 2931–2944 (2014).
63. Wlasiuk, G., Khan, S., Switzer, W. M. & Nachman, M. W. A History of Recurrent Positive Selection at the Toll-Like Receptor 5 in Primates. *Mol. Biol. Evol.* **26**, 937–949 (2009).
64. Bainová, H. *et al.* First evidence of independent pseudogenization of Toll-like receptor 5 in passerine birds. *Dev. Comp. Immunol.* **45**, 151–155 (2014).
65. Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
66. Boyd, A., Philbin, V. J. & Smith, A. L. Conserved and distinct aspects of the avian Toll-like receptor (TLR) system: implications for transmission and control of bird-borne zoonoses. *Biochem. Soc. Trans.* **35**, 1504–1507 (2007).
67. Ohto, U., Fukase, K., Miyake, K. & Shimizu, T. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proc. Natl. Acad. Sci. USA* **109**, 7421–7426 (2012).
68. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* **458**, 1191–1195 (2009).
69. Areal, H., Abrantes, J. & Esteves, P. J. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol. Biol.* **11**, 368 (2011).
70. Grueber, C. E., Wallis, G. P. & Jamieson, I. G. Episodic Positive Selection in the Evolution of Avian Toll-Like Receptor Innate Immunity Genes. *PLoS ONE* **9**, (2014).
71. Key, F. M., Teixeira, J. C., de Filippo, C. & Andrés, A. M. Advantageous diversity maintained by balancing selection in humans. *Curr. Opin. Genet. Dev.* **29**, 45–51 (2014).
72. Eriksson, J. *et al.* Identification of the Yellow Skin Gene Reveals a Hybrid Origin of the Domestic Chicken. *PLOS Genet.* **4**, e1000010 (2008).
73. Liu, Y.-P. *et al.* Multiple maternal origins of chickens: out of the Asian jungles. *Mol. Phylogenet. Evol.* **38**, 12–19 (2006).
74. Dannemann, M., Andrés, A. M. & Kelso, J. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am. J. Hum. Genet.* **98**, 22–33 (2016).
75. Anderson, K. L., Brewer, M. T., Rasmussen, M. A. & Carlson, S. A. Identification of heritage chicken breeds with diminished susceptibility to intestinal colonization by multiple antibiotic-resistant Salmonella spp. *Livest. Sci.* **182**, 34–37 (2015).
76. Pavel, I. & Tuláček, F. *Vzorník plemen drůbeže (Catalogue of poultry breeds)*. (Český svaz chovatelů 2006).
77. Scrivener, D. *Poultry Breeds and Management: An Introductory Guide*. (Crowood Press, Limited, 2008).
78. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinforma. Oxf. Engl.* **25**, 1451–2 (2009).
79. R Core Team. A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna). Available at: https://www.r-project.org/. (2015).
80. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
81. Pond, S. L. K., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. W. Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. *Mol. Biol. Evol.* **23**, 1891–1901 (2006).
82. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLOS ONE* **7**, e46688 (2012).
83. Pond, S. L. K. & Frost, S. D. W. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinforma. Oxf. Engl.* **21**, 2531–3 (2005).
84. Murrell, B. *et al.* FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol. Biol. Evol.* **30**, 1196–1205 (2013).
85. Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–181 (2015).
86. Kanno, A. *et al.* Essential role for Toll-like receptor 7 (TLR7)-unique cysteines in an intramolecular disulfide bond, proteolytic cleavage and RNA sensing. *Int. Immunol.* **25**, 413–422 (2013).
87. Hipp, M. M. *et al.* The Processed Amino-Terminal Fragment of Human TLR7 Acts as a Chaperone To Direct Human TLR7 into Endosomes. *J. Immunol.* **194**, 5417–5425 (2015).
88. McGuffin, L. J., Buenavista, M. T. & Roche, D. B. The ModFOLD4 server for the quality assessment of 3D protein models. *Nucleic Acids Res.* **41**, W368–W372 (2013).
89. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945–959 (2000).
90. Liu, L. *et al.* Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* **320**, 379–381 (2008).
91. Wei, T. *et al.* Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. *Protein Sci.* **18**, 1684–1691 (2009).
92. Yoon, S. *et al.* Structural Basis of TLR5-Flagellin Recognition and Signaling. *Science* **335**, 859–864 (2012).

Acknowledgements

Author Contributions

M.V. designed the research, Z.Ś., A.Š., L.B., A.B. and A.F. conducted the study and Z.Ś. and P.M. analysed the data. Z.Ś. and M.V. wrote the main manuscript text, Z.Ś. prepared all figures and all authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-36226-1.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.