Parasitology

cambridge.org/par

A roll of the dice: pathogen–host interaction and the evolution of disease susceptibility

Paul Capewell[1], Philipp Olias[2,3] and Brian Shiels[4]

[1]School of Molecular Biosciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK; [2]Institute of Animal Pathology, Vetsuisse Faculty, University of Bern, Bern, Switzerland; [3]Institute of Veterinary Pathology, Justus Liebig University, Giessen, Germany and [4]School of Biodiversity, One Health and Veterinary Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

Review Article

Cite this article: Capewell P, Olias P, Shiels B (2026) A roll of the dice: pathogen–host interaction and the evolution of disease susceptibility. *Parasitology* **153**, 325–332. https://doi.org/10.1017/S0031182025101492

Received: 7 August 2025
Revised: 19 November 2025
Accepted: 13 December 2025
First published online: 9 January 2026

Keywords:
DNA-binding proteins; epigenesis; evolution; gene expression regulation; host–pathogen interactions; *Theileria annulata*

Corresponding author: Brian Shiels; Email: brian.shiels@glasgow.ac.uk

Abstract

The probability that a disease will manifest is highly variable. Susceptibility to disease is influenced by genetic background, environment and lifestyle choices. In this review, we put forward the premise that evolution of disease susceptibility may be partially influenced by the interaction of divergent pathogen DNA-binding proteins with variable binding sites in the host genome. The hypothesis put forward is derived from recent data obtained from work on the protozoan parasite, *Theileria annulata*, together with related evidence from viral and bacterial pathogens that have been postulated to modulate host epigenome architecture. The pathogen proteins highlighted have the potential to mimic functions of mammalian epigenome organisers linked to a range of disease syndromes. It is feasible, therefore, that the evolutionary relationship between pathogen and host impacts susceptibility to a range of conditions, such as autoimmune disorders and cancer, which are not directly linked to pathogen infection.

Introduction

Disease manifestation varies widely between individuals, even when infected with the same pathogen. Pathologies can be linked to specific genetic mutations where the occurrence of disease is virtually guaranteed. The outcome for many conditions, however, is more unpredictable. Moreover, whether a disease occurs or not can be influenced by external factors, such as environment and lifestyle, which appear to operate independent of genetic information inherited by the affected individual. Understanding how these factors promote disease in some individuals but not others is key to the prediction of disease risk and promotion of health in a population.

The adaptation of hosts and pathogens to display mild disease or asymptomatic infection is theorised to occur via a co-evolutionary arms race that promotes an equilibrium between host survivability and pathogen transmission, reducing pathology and increasing tolerance to infection (Glass et al., 2012; Sironi et al., 2015). In contrast, novel infection of hosts that have not co-evolved with a pathogen would be expected to be more susceptible to severe disease (Glass et al., 2012; Seal et al., 2021). This has been shown for numerous zoonotic infections and by the transfer of high-yield, exotic livestock breeds into endemic regions of disease (Glass, 2001; Seal et al., 2021). Despite co-evolution of pathogen–host interaction towards milder disease outcomes, there is still the unpredictable potential for members of a tolerant population to suffer acute disease and for those of susceptible breeds to show limited clinical signs (Glass et al., 2005).

To address how variability in disease susceptibility can be generated, our work has aimed to identify factors that promote the markedly different outcomes that occur when different cattle breeds are infected with the tick-borne parasite *Theileria annulata*, the causative agent of tropical theileriosis. Data indicate that susceptibility to this disease is, in part, mediated by direct interactions between pathogen-derived factors and the host genome that influence chromatin structure and gene expression. The importance of the epigenetic landscape is often overlooked when seeking to identify susceptibility-linked heritable traits but could make a significant contribution given the known role of the epigenome in numerous infectious and non-infectious diseases (Davis et al., 2021).

Pathogenesis and cellular transformation in tropical theileriosis

Tropical theileriosis is a disease of economic importance to livestock production and is endemic over much of Asia and North Africa (Minjauw and McLeod, 2003). Acute disease results in dysregulation of the immune response and lung pathology that mirrors numerous viral diseases. Both large and small production systems are susceptible, and the impact on poorer subsistence farmers is catastrophic. European cattle breeds (*Bos taurus*), such as Holstein, show a greater

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

susceptibility to acute disease than breeds native to endemic regions, like Sahiwal (*Bos indicus*) (Glass et al., 2012). This should be expected, as native breeds that have co-evolved with *Theileria* would be closer to reaching an equilibrium in the trade-off between host survivability and pathogen transmissibility. Efforts have been made, including the national programme in India (Singh, 2015), to improve productivity by crossbreeding tolerant native cattle breeds with European breeds. While this has mitigated losses, infected crossbred animals still show clinical signs and lower production than European breeds (Larcombe et al., 2019).

Previous studies have confirmed that Sahiwal cattle are tolerant to *T. annulata* infection when compared to Holstein cattle (Glass et al., 2005). A pro-inflammatory response (cytokine storm) has been observed in infected Holstein calves, allowing postulation that acute disease pathology is linked to parasite-mediated dysregulation of the host immune response (Glass et al., 2012). In support of that premise, transcriptome analysis on *T. annulata*-infected leukocytes, *B. taurus* (Holstein) vs *B. indicus* (Sahiwal), has found that tolerance is associated with altered expression of hundreds of infection-associated bovine genes that operate in innate immunity and oncogenesis, demonstrating a multifaceted phenotype (Jensen et al., 2008; Larcombe et al., 2022). However, despite showing a statistically significant breed associated trend across both sets of samples, many genes showed a degree of variable transcript level within each breed. Given the complicated nature of the data, it is not clear how differential susceptibility to tropical theileriosis arises. One model for this is that divergent interaction between pathogen factors and host chromatin promote different epigenetic landscapes that lead to a wide range of phenotypes, both between and within breed: from tolerance to acute disease.

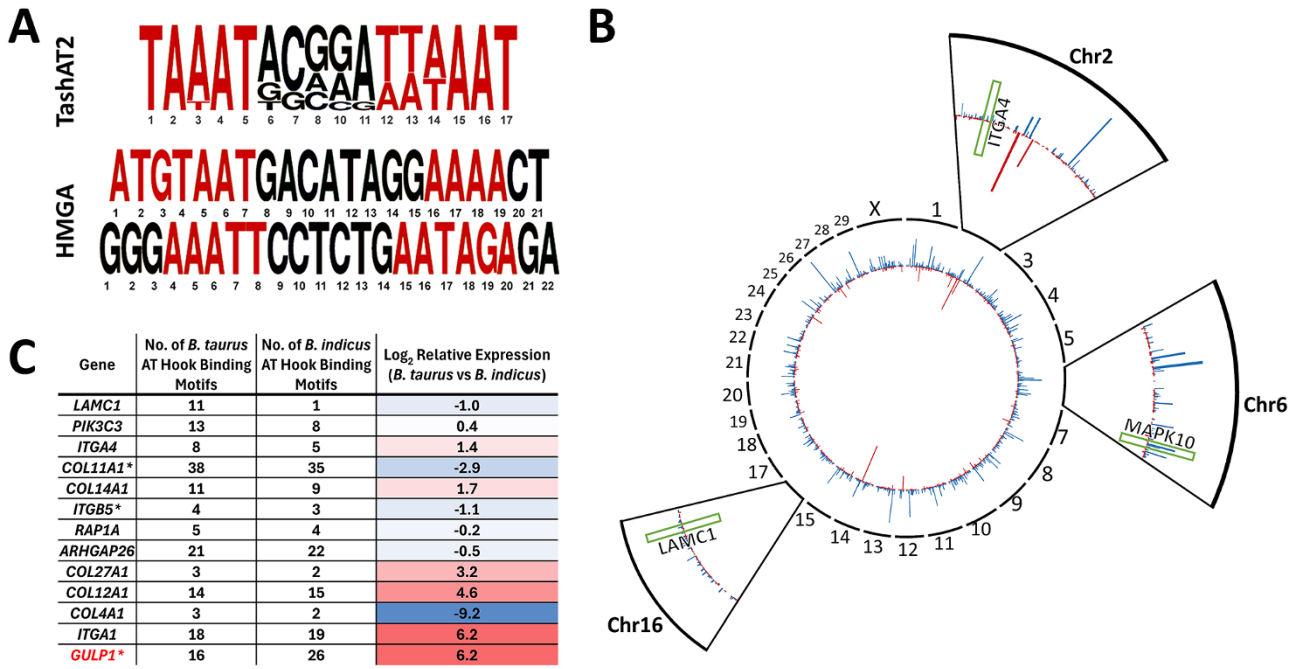
The major point of host–pathogen interaction associated with pathology in tropical theileriosis is the macroschizont-infected leukocyte. Establishment of the infected cell is accompanied by transformation into a cancer-like cell that metastasises to several organs, including the lungs. Destruction of the lymphoid system and massive pulmonary oedema then occur (Irvin and Morrison, 1987), and differences in breed susceptibility have been linked to these events (Chaussepied et al., 2010). Transformation of leukocytes occurs via constitutive activation of bovine transcription factors, such as NF-κB and AP1, to promote infected cell survival and oncogenesis (Shiels et al., 2006), but these factors also act as pro-inflammatory mediators. Activation of these host cell 'Janus-like' transcription factors is modulated by parasite-dependent reorganisation of the infected cell gene expression network (Durrani et al., 2012). Thus, while some NF-κB target genes show elevated expression, others are negatively regulated, and expression of interferon-stimulated genes is moderated (Oura et al., 2006; Durrani et al., 2012). More recent work extends these findings and indicates that differential modulation of the infection-associated transcription network is linked to breed susceptibility (Larcombe et al., 2022). Thus, significant differences (<0.1 FDR) in the expression of hundreds of infection-associated genes were identified between macroschizont-infected cell lines derived from 6 infected Sahiwal calves compared to 5 Holsteins. Pathway analysis highlighted an enrichment of genes and pathways associated with innate immunity and cholesterol biosynthesis, and a modulated type I interferon response was prominent. These results demonstrate that the expression of genes with potential to impact the outcome of infection is actively manipulated by the parasite in a breed-dependent manner. Because most intracellular pathogens hijack the host cell machinery to construct a niche that promotes their survival, the information gained from the *Theileria* model may apply to other pathogen–host interactions.

A common pathogen hijacking strategy deployed to establish an intracellular niche is the export of molecules that co-opt regulation of host cell gene expression. Pathogens subvert regulation of gene expression by corrupting host transcription factor (TF) activity and by altering host chromatin structure to modulate TF accessibility (Cheeseman and Weitzman, 2015; Davis et al., 2021). *Theileria annulata* likely deploys both strategies. Direct manipulation of transcription factor state is demonstrated by constitutive activation of NF-κB via hijacking of the IκB kinase signalosome (Heussler et al., 2002), while parasite factors encoded in *TashAT* gene cluster are indicated as modulators of host chromatin structure (Swan et al., 2001; Shiels et al., 2006; Larcombe et al., 2022). TashAT proteins travel to infected leukocyte nuclei, bind AT-rich DNA and contain AT-hook domains that are organised in a manner similar to mammalian high mobility group A (HMGA) proteins. HMGAs operate by binding and affecting various AT-rich genetic (DNA) motifs to modify chromatin structures and methylation, regulating transcription factor accessibility and gene expression. HMGAs therefore act as hubs of nuclear function (Reeves, 2001) and have a major role in development, stem cell maintenance, cholesterol biosynthesis and the inflammatory response (Treff et al., 2004; Henriksen et al., 2010; Vignali and Marracci, 2020). These pathways are also differentially modified by *Theileria* infection, depending on whether the infected cell line was derived from a tolerant or susceptible host (Larcombe et al., 2022). Due to the similarity between mammalian HMGAs and parasite TashAT proteins, we proposed that they act in a related manner and that the pathogen modifies host chromatin and TF accessibility by binding AT-rich genetic motifs to influence the epigenetic landscape and disease susceptibility. In support of this model, many loci ascribed as causal variants for disease susceptibility have been linked to regions of chromatin accessibility (Maurano et al., 2012).

Breed-specific DNA motifs and selection of epigenetic architecture

Expression of a TashAT encoding gene (*TashAT2*) in a transfected bovine macrophage cell line altered transcript levels of more than 800 bovine genes (Durrani et al., 2023). Pathway analysis indicated involvement in focal adhesion, transforming growth factor (TGF)-β, axon guidance and integrin signalling with similarity to pathways regulated by mammalian HMGA2 (focal adherence, TGF-β, integrin signalling) (Singh et al., 2015), supporting postulation that the parasite factor acts as an HMGA mimic. Moreover, 184 genes ($P < 0.1$) modulated by TashAT2 also display differential expression between Sahiwal and Holstein infected cells, with enrichment of genes operating in Integrin signalling ($P = 0.04$) and Axon guidance ($P = 0.02$). A further link was established by comparison of the published *B. taurus* and *B. indicus* genomes for AT-rich nucleotide motifs bound by TashAT2 (Larcombe et al., 2022). The results showed that the pattern (number and positions) of the TashAT2-bound consensus motif differs substantially between the 2 genomes (Figure 1) and that more than 45% of genes differentially expressed between Sahiwal and Holstein infected cells feature at least one putative TashAT2 binding motif ($P < 0.001$). Within this cohort, there is enrichment in pathways modulated by infection and differences in the number of TashAT2 binding motifs between genes of *B. indicus* vs *B. taurus*. An example is

https://doi.org/10.1017/S0031182025101492 Published online by Cambridge University Press



C

Gene	No. of *B. taurus* AT Hook Binding Motifs	No. of *B. indicus* AT Hook Binding Motifs	Log_2 Relative Expression (*B. taurus* vs *B. indicus*)
LAMC1	11	1	-1.0
PIK3C3	13	8	0.4
ITGA4	8	5	1.4
*COL11A1**	38	35	-2.9
COL14A1	11	9	1.7
*ITGB5**	4	3	-1.1
RAP1A	5	4	-0.2
ARHGAP26	21	22	-0.5
COL27A1	3	2	3.2
COL12A1	14	15	4.6
COL4A1	3	2	-9.2
ITGA1	18	19	6.2
*GULP1**	16	26	6.2

Figure 1. Differences in the number and location of AT-hook binding site motifs between *Bos taurus* and *Bos indicus* genomes (A) Consensus DNA sequence motif bound by parasite TashAT2 compared to mammalian HMGA. AT rich regions are in red. (B) CIRCOS summary plot of genes displaying a different number of TashAT2 binding sites in *Bos indicus* (red) and *Bos taurus* (blue). Zoomed views for chromosomes 2, 8 and 16: genes in the EGF and integrin signalling pathways with different numbers of TashAT2 binding site motifs and differing in expression between Holstein and Sahiwal infected cells are indicated by green boxes (ITGA4, MAPK10 and LAMC1). (C) Integrin pathway genes differentially expressed between *T. annulata* infected *Bos taurus* vs *Bos indicus* infected cell lines displaying different numbers of predicted TashAT2 binding site motifs (data derived from Larcombe et al., 2022). Relative expression is shaded using a red-to-blue gradient, with red indicating higher values and blue indicating lower values. GULP1 (highlighted red) has been experimentally validated as a modulated target gene in a TashAT2 transfected-BoMac cell line (Durrani et al., 2023).

the EGF/PI3K pathway, linked to transformation of the infected leukocyte and modulated by mammalian AT-hook proteins (Wang et al., 2021). Furthermore, altered expression of integrin signalling genes is associated with virulence of the *Theileria* infected leukocyte (Haidar et al., 2015) and can occur via TashAT2 (Singh et al., 2015) and HMGA1 (Reeves et al., 2001). Based on these findings, it can be postulated that variability in the arrangement and, potentially sequence, of motifs bound by AT-hook proteins vary between breeds and possibly individual animals. Such variability could impact on epigenome architecture and the resulting expression level of multiple genes linked to disease susceptibility. How the variability in motif arrangement within the host genome is generated is not known. The nucleotide motif bound by HMGA AT-hook proteins and TashAT2 is only semi-conserved and consists of a pair of AT-rich sequences interspersed with a region higher in GC content (Swan et al., 2001; Su et al., 2020; Larcombe et al., 2022). It is probable that these motif patterns have a level of variability between breeds and possibly within breed populations. It is also possible that evolution of the TashAT2-bound motif pattern in the host genome could be influenced by selective pressure, as interaction with a lethal pathogen infection could remove, over time, individuals with patterns that give rise to fatal disease (Tseng et al., 2021). Mammalian HMGA proteins are known to operate in important developmental pathways and are linked to haematopoiesis, epithelial–mesenchymal transition and adipogenesis (Su et al., 2020; Vignali and Marracci, 2020). Thus, selective breeding for particular traits may influence the pattern of motifs (and epigenome architecture) encoded by genomes of cattle breeds that are differentially susceptible to pathogen infection.

Genetic diversity and the evolution of pathogen–host interaction

An additional potential contributor to altered infected cell gene expression is diversity in the pattern of AT-hook domains across TashAT2/3 alleles. In the analysis of (Weir et al., 2010), Ten major arrangements were identified across 8 sequenced isolates. These findings have been validated, by analysis of parasite genomes derived from multiple clonal cell lines. The results identified TashAT2/3 alleles displaying variable numbers and spacing of AT-hooks. Moreover, dN/dS analysis provides evidence for positive selection of amino acid substitution primarily within the first 2 AT-hook domains of TashAT2 and towards the C-terminus of the protein (Figure 2). The negatively charged C-terminal tail of HMGA proteins show variable phosphorylation patterns that are thought to influence DNA binding efficiency or protein partner binding, which may contribute to cellular transformation (Sgarra et al., 2009; Su et al., 2020). In analogy, the substitutions highlighted for the C-terminal region of TashAT2 overlap with predicted PEST motifs (Shiels et al., 2006) and phosphorylation sites for protein kinases. Post-translational modification of PEST regions in nuclear proteins are considered important for partner binding and modulation of function, including oncogenic activity (Sarfraz et al., 2021).

The data for selection of variant pathogen AT-hook domains, together with evidence that TashAT alleles display geographical substructuring (Weir et al., 2010), implies that *T. annulata* has co-evolved with local host genetic backgrounds to recognise divergent DNA target motifs. In this scenario, the random probability of disease would be influenced by both the genetic background of the infecting parasite populations and the genome of the infected host,

https://doi.org/10.1017/S0031182025101492 Published online by Cambridge University Press

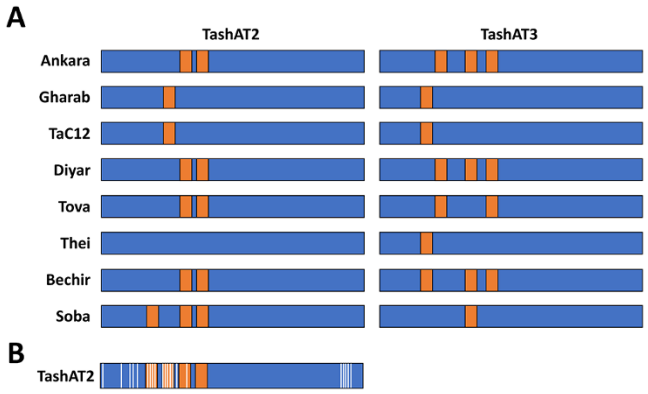


Figure 2. Amino acid diversity and rearrangement of AT-hook motifs in TashAT2/3 alleles (A) Pattern of AT-hook motifs predicted from allelic DNA sequences of TashAT2 and TashAT3. The regions containing the TashAT locus were extracted from 8 publicly available genomes (ENA project ID: PRJEB65114) previously aligned to the C9 Ankara reference strain using BEDtools. The number and positions of AT-hook motifs were identified using SMART and the core (GRP) motif for each putative site validated manually. Relative positions of each AT-hook domain are denoted Orange. (B) The TashAT2 sequences were compared using JCoDA to identify sites of positive selection of amino acid substitution. The relative positions of sites with a probability of ($dN/dS > 1$) greater than 0.9 (white) were overlaid on a composite cartoon of the TashAT2 alleles.

leading to a wide range of infection outcomes, between and within animal breeds. Over evolutionary time the pathogen–host relationship has produced *B. indicus* genomes that possess patterns of (AT-rich) DNA binding motifs that, in general, promote tolerance; while European, Bos taurine breeds, selected for productivity traits in the absence of pathogen pressure, display greater susceptibility. The model predicts that (1) parasite TashAT proteins contribute to the construction of chromatin architectures that promote gene expression changes linked to a spectrum of disease outcomes; (2) differences in the pattern of motifs that interact with AT-hook proteins are linked to breed morphological traits and disease susceptibility and (3) the binding of variant TashAT factors to variant host genome contributes to disease susceptibility.

Analogous viral and bacterial epigenetic modulators

A relevant question in relation to the above model is whether it is applicable for other intracellular pathogen–host interactions. Evidence for multiple pathogen factors manipulating the host epigenome has been generated (Silmon De Monerri and Kim, 2014; Cheeseman and Weitzman, 2015; Davis et al., 2021). The 2 examples of pathogen nucleomodulins most related to *T. annulata* are the Epstein–Barr nuclear antigen 1 (EBNA1) of Epstein Barr Virus (EBV) and the ankaryin A (AnkA) protein of *Anaplasma phagocytophilum* (Table 1).

Like *Theileria*, EBV infection is linked to a transformed host cell phenotype and associated pathology (Burkitt's lymphoma, Hodgkin lymphoma). Latent virus infection requires attachment of the EBV episome to host chromosomes. This is achieved, in part, via the DNA binding functions of EBNA1 (Frappier, 2012). DNA binding has been shown to involve at least 3 domains. Two domains (CBD1 and CBD2) possess AT-hook motifs that show similarity to mammalian HMGA1a (Kanda et al., 2013) and *Theileria* TashAT2/3. Furthermore, the episomal tethering function of EBNA1 can be replicated by HMGA1. Genome-wide mapping of EBV episome attachment sites identified enrichment of AT-rich DNA, flanking C-terminal EBNA1 DNA binding domain (DBD) sites and the epigenetic H3K9me3 mark (Kim et al., 2020). These sites were associated with transcriptional repression of genes with neural functions or genes operating in the protein kinase A pathway. Thus, EBNA1 binds to host DNA and, like TashAT2, has been postulated to act as an analogue of HMGAs (Coppotelli et al., 2013) to modulate the host cell epigenetic landscape resulting in altered gene expression. In further analogy to TashAT2, study of EBNA1 gene sequences from different viral strains identified a high level of sequence divergence (Bell et al., 2008). While evidence for amino acid substitution in the AT-hook domains is limited, the majority (68%) of isolates predict alterations in the C-terminal DBD. It was proposed that these substitutions may be linked to the evolution and geographical origin of the viral isolate.

Anaplasma phagocytophilum is a tick-borne, Gram-negative bacteria placed in the order Rickettsiales. It is the cause of anaplasmosis in domestic animals and humans. Human granulocytic anaplasmosis is the third most common tick-borne disease in Europe and the USA. The outcome on infection is variable; most patients have mild clinical signs but for 36%, infection results in hospitalisation, with 7% of clinical cases requiring intensive care (Dumler, 2012). *Anaplasma phagocytophilum* infects neutrophils and, like *Theileria*, pathology is linked to a proinflammatory cytokine storm with similarity to COVID-19 (Ramanujam et al., 2021). In further analogy to *T. annulata*, infection of the host cell results in activation of PI3K/AKT and NF-κB pathways to protect against apoptosis (Sarkar et al., 2012). An extensive reprogramming of gene expression (Borjesson et al., 2005) that moderates production of proteins involved in innate immunity operates (Garcia-Garcia et al., 2009b).

Modulation of gene expression is generated through alteration of host cell chromatin architecture. Wide-scale manipulation of the epigenome is mediated by the AnkA protein that is secreted and translocated to the neutrophil nucleus via an importin-b-, RanGTP-dependent mechanism (Kim et al., 2022). AnkA manipulates the epigenome by binding to runs of AT-rich DNA and recruiting histone deacetylase 1 (HDAC1) to the *CYBB* promoter region. Recruitment of HDAC1 causes silencing of *CYBB* expression, promoting pathogen survival (Rennoll-Bankert et al., 2015). DNA binding is achieved by 4 central ankyrin repeats, together with a C-terminal AT-hook like domain with a core RGR motif. Chromatin immunoprecipitation (ChIP) studies indicate that AnkA acts genome wide to generate large scale changes to the epigenome of the infected cell (Garcia-Garcia et al., 2009a; Dumler et al., 2016). Moreover, AnkA binds to regions of the genome similar to those of matrix attachment regions and lamina associated domains. Therefore, AnkA could act as an analogue of host transcription factors, such as the special AT-rich binding-protein-1 (SATB1), also a target of HIV-1, trans-activator of transcription modulatory activity (Kumar et al., 2005), that organise chromatin architecture into topological domains within the nucleus, that in turn, influence the transcriptional programme of the cell (Sinclair et al., 2014).

Similar to TashAT2/3 and EBNA1, AnkA displays a high degree of sequence divergence across bacterial strains (Scharf et al., 2011). Most nucleotide changes are nonsynonymous, predicting amino acid substitution. While no geographical subclustering has been found, different AnkA gene clusters correlate with strain tropism for different host species, suggesting a level of co-evolution and adaptation. There is no evidence, to date, that diversity in AnkA impacts on disease outcome. However, functional diversity is

https://doi.org/10.1017/S0031182025101492 Published online by Cambridge University Press

Table 1. Comparison of 3 protozoal, viral and bacterial nucleomodulins that act as (putative) analogues of host architectural transcription factors

Pathogen	Effector protein	Host target sequence	Putative analogue	Proposed mode of action	Amino acid diversity
Theileria annulata	TashAT2	AT-rich DNA motifrelated to HMGA binding site.	HMGA1/2	Binds to host DNA. Modulates chromatin structure; Integrin signalling, TGFβ/SMAD3, EGF/PIK3 pathways.	Yes: AT hooks, C-terminal PEST.
Epstein-Barr virus	EBNA1	AT-rich DNA flanking DBD binding sites.	HMGAs	Binds to host DNA. Modulates epigenetic protein kinase A pathways.	Yes: DBD domain.
Anaplasma phagocy-tophilum	AnkA	AT-rich DNA of matrix attachment regions.	SATB1	Binds to host DNA. Modulates chromatin organisation & epigenetic landscape. Repression of CYBB expression and other host cell defence genes.	Yes: correlation with strain tropism for host species.

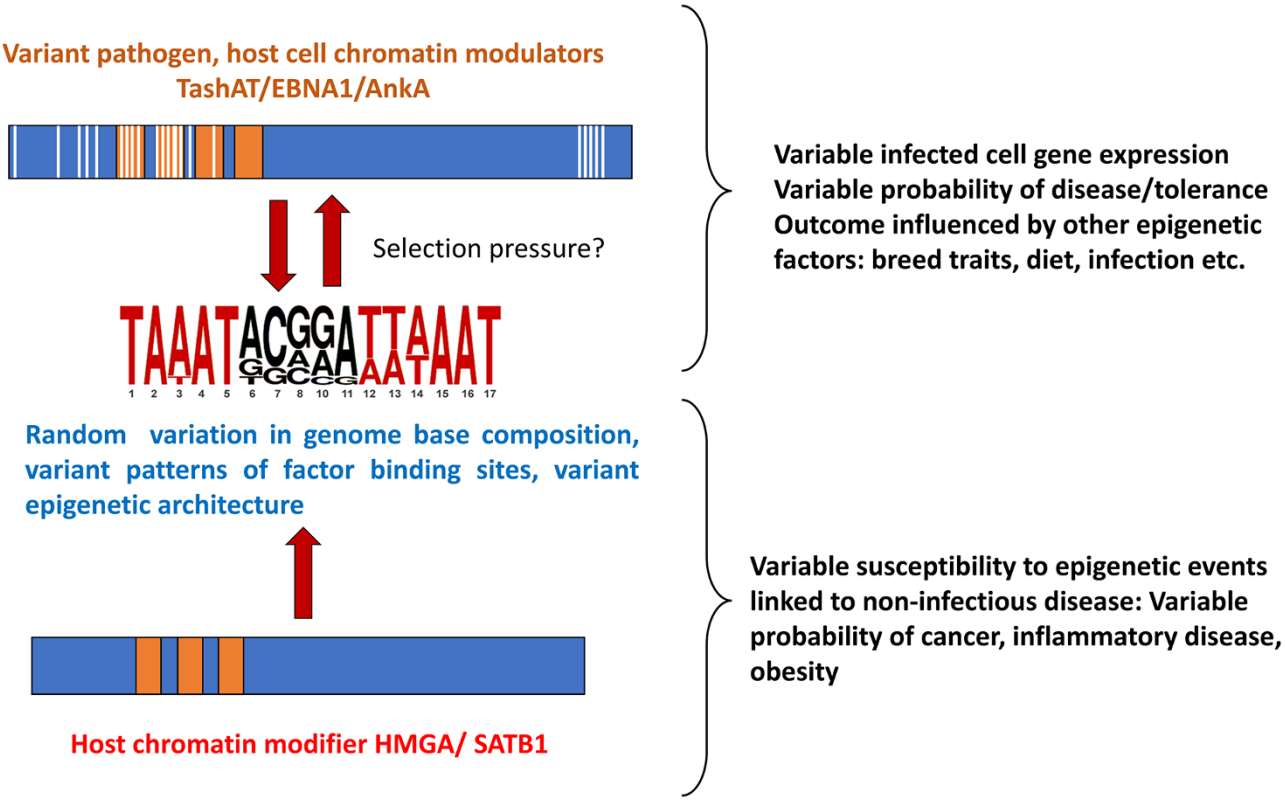


Figure 3. Overview of the proposed model for pathogen-host interaction and evolution of disease susceptibility.

thought possible (Scharf et al., 2011): and it may be noteworthy, that enriched binding of AnkA to DNA upstream of target gene transcriptional start sites displayed significant variability between the 3 different human genomes investigated (Dumler et al., 2016).

Conclusions

How genome diversity and environmental factors interact to influence susceptibility to disease is not fully understood. The model postulated (outlined in Figure 3) whereby variant DNA motifs interact with divergent regulatory proteins to construct variable epigenetic landscapes, suggests a link between genome, environment and lifestyle which could be important for understanding how the response to infection can vary so widely. For example, the infected cell epigenome and susceptibility to theileriosis may be influenced by host genome, parasite genome, livestock environment/nutrition and infection of the bovine host by other pathogens. Moreover, pathogen factors may mimic host molecules (e.g. HMGA, SATB1) (Coppotelli et al., 2013; Sinclair et al., 2014; Durrani et al., 2023) that function as base composition readers (Quante and Bird, 2016) and shape chromatin architecture during embryonic development (Goolam and Zernicka-Goetz, 2017; Panchal et al., 2020; Su et al., 2020; Vignali and Marracci, 2020). These host factors have been implicated in a range of conditions, including neoplasia and disease linked to immune dysfunction (Reeves, 2001; Treff et al., 2004; Henriksen et al., 2010; Frömberg et al., 2018; Panchal et al., 2020; Vignali and Marracci, 2020). Based on the premise that pathogen infection could influence selection of genomes with particular patterns of motifs to which the host factors bind: further study of pathogen-host interaction has potential to provide information on how variable susceptibility to diseases unrelated to infection has evolved, with the caveat that individual outcomes can still depend on just a roll of the dice.

https://doi.org/10.1017/S0031182025101492 Published online by Cambridge University Press

Limitations of the model and future perspectives

The model outlined above has limitations and requires additional data. To date, it is based on 2 reference genomes representing *B. taurus* and *B. indicus*. Analysis of genome sequence representing multiple breeds of both subspecies is required; with the proviso that a degree of diversity in the TashAT2 binding site motif pattern (number, position and potentially sequence) is predicted between different breeds and between individuals within a breed. It may also be of interest to identify the gene expression profile of infected leukocytes, and the binding site pattern in the genome, of water buffalo (*Bubalus bubalis*). This could provide information on whether evolution of tolerance to *T. annulata* displayed by native cattle breeds and *Bubalus bubalis* reflects related or divergent mechanisms.

While pathogen modulation of the host epigenome may be a relatively common strategy (Bierne and Pourpre, 2020; Silmon De Monerri and Kim, 2014; Cheeseman and Weitzman, 2015), it is unknown how applicable the above model is to intracellular pathogens, in general. Consensus AT-hook domains are prevalent in eukaryotes but thought to be rare in prokaryotes (Aravind, 1998). AT-hook proteins are encoded by herpes viruses and AT-hook domains have been reported as shared between viral, host and prokaryotic DNA binding proteins (Dreyfus et al., 2018). Multiple extended nucleic acid binding AT-hook motifs, and other protein domains with potential to act as base composition readers, have been described (Filarsky et al., 2015; Quante and Bird, 2016). The model, therefore, may be applicable to other intracellular pathogens. Even if limited to the 3 examples given above; over evolutionary time, a considerable proportion of the human and animal population will have been exposed to these pathogen–host interactions.

Additional data could be obtained by: the assay for transposase-accessible chromatin with high-throughput sequencing of *B. indicus* and *B. taurus* infected leukocytes (Buenrostro et al., 2015), to compare regions of chromatin accessibility; ChIP-sequencing (Johnson et al., 2007), to functionally identify genomic locations that TashAT2 protein binds; comparison of the gene expression profile modulated by distinct TashAT2/3 alleles (Durrani et al., 2023); investigation of whether TashAT2 binding site diversity is driven by selection (Sironi et al., 2015) or has evolved passively (McShea, 1994).

Author contributions. B.R.S. conceived and wrote the initial draft of the manuscript. B.R.S., P.C., and P.R. all contributed to further drafting and editing. Additional analysis was performed by P.C.

Financial support. B.R.S. was awarded a grant from the Biotechnology and Biological Sciences Research Council (BBSRC) https://www.ukri.org/councils/bbsrc/ Grant BB/L004739/1. The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests. The authors declare there are no conflicts of interest.

Ethical standards. Not applicable.

References

Aravind L (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Research* **26**, 4413–4421. https://doi.org/10.1093/nar/26.19.4413

Bell MJ, Brennan R, Miles JJ, Moss DJ, Burrows JM and Burrows SR (2008) Widespread sequence variation in Epstein-Barr virus nuclear antigen 1 influences the antiviral T cell response. *The Journal of Infectious Diseases* **197**, 1594–1597. https://doi.org/10.1086/587848

Bierne H and Pourpre R 2020 Bacterial Factors Targeting the Nucleus: The Growing Family of Nucleomodulins.Toxins (Basel) **12**, 220. https://doi.org/10.3390/toxins12040220

Borjesson DL, Kobayashi SD, Whitney AR, Voyich JM, Argue CM and DeLeo FR (2005) Insights into pathogen immune evasion mechanisms: *Anaplasma phagocytophilum* fails to induce an apoptosis differentiation program in human neutrophils. *The Journal of Immunology* **174**, 6364–6372. https://doi.org/10.4049/jimmunol.174.10.6364

Buenrostro JD, Wu B, Chang HY and Greenleaf WJ (2015) ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology* **109**, 21.29.1–21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Chaussepied M, Janski N, Baumgartner M, Lizundia R, Jensen K, Weir W, Shiels BR, Weitzman JB, Glass EJ, Werling D and Langsley G (2010) TGF-b2 induction regulates invasiveness of *Theileria*-transformed leukocytes and disease susceptibility. *PLOS Pathogens* **6**, 1001197. https://doi.org/10.1371/journal.ppat.1001197

Cheeseman K and Weitzman JB (2015) Host–parasite interactions: An intimate epigenetic relationship. *Cellular Microbiology* **17**, 1121–1132. https://doi.org/10.1111/cmi.12471

Coppotelli G, Mughal N, Callegari S, Sompallae R, Caja L, Luijsterburg MS, Dantuma NP, Moustakas A and Masucci MG (2013) The Epstein–Barr virus nuclear antigen-1 reprograms transcription by mimicry of high mobility group A proteins. *Nucleic Acids Research* **41**, 2950–2962. https://doi.org/10.1093/nar/gkt032

Davis SZ, Hollin T, Lenz T and Le Roch KG (2021) Three-dimensional chromatin in infectious disease—A role for gene regulation and pathogenicity? *PLOS Pathogens* **17**, e1009207. https://doi.org/10.1371/journal.ppat.1009207

Dreyfus DH, Farina A and Farina GA (2018) Molecular mimicry, genetic homology, and gene sharing proteomic "molecular fingerprints" using an EBV (Epstein-Barr virus)-derived microarray as a potential diagnostic method in autoimmune disease. *Immunologic Research* **66**, 686–695. https://doi.org/10.1007/s12026-018-9045-0

Dumler JS (2012) The biological basis of severe outcomes in *Anaplasma phagocytophilum* infection. *FEMS Immunology & Medical Microbiology* **64**, 13–20. https://doi.org/10.1111/j.1574-695X.2011.00909.x

Dumler JS, Sinclair SH, Pappas-Brown V and Shetty AC (2016) Genome-wide *Anaplasma phagocytophilum* AnkA-DNA interactions are enriched in intergenic regions and gene promoters and correlate with infection-induced differential gene expression. *Frontiers in Cellular and Infection Microbiology* **6**, 97. https://doi.org/10.3389/fcimb.2016.00097

Durrani Z, Kinnaird J, Cheng CW, Brühlmann F, Capewell P, Jackson A, Larcombe S, Olias P, Weir W and Shiels B (2023) A parasite DNA binding protein with potential to influence disease susceptibility acts as an analogue of mammalian HMGA transcription factors. *PLOS ONE* **18**, e0286526. https://doi.org/10.1371/journal.pone.0286526

Durrani Z, Weir W, Pillai S, Kinnaird J and Shiels B (2012) Modulation of activation-associated host cell gene expression by the apicomplexan parasite *Theileria annulata*. *Cellular Microbiology* **14**, 1434–1454. https://doi.org/10.1111/j.1462-5822.2012.01809.x

Filarsky M, Zillner K, Araya I, Villar-Garea A, Merkl R, Längst G and Németh A (2015) The extended AT-hook is a novel RNA binding motif. *RNA Biology* **12**, 864–876. https://doi.org/10.1080/15476286.2015.1060394

Frappier L (2012) The Epstein-Barr virus EBNA1 protein. *Scientifica* **2012**, 1–15. https://doi.org/10.6064/2012/438204

Frömberg A, Engeland K and Aigner A (2018) The special AT-rich sequence binding protein 1 (SATB1) and its role in solid tumors. *Cancer Letters* **417**, 96–111. https://doi.org/10.1016/j.canlet.2017.12.031

Garcia-Garcia JC, Barat NC, Trembley SJ and Dumler JS (2009b) Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. *PLoS Pathogens* **5**, e1000488. https://doi.org/10.1371/journal.ppat.1000488

Garcia-Garcia JC, Rennoll-Bankert KE, Pelly S, Milstone AM and Dumler JS (2009a) Silencing of host cell *CYBB* gene expression by the nuclear effector AnkA of the intracellular pathogen *Anaplasma phagocytophilum*.

Infection and Immunity **77**, 2385–2391. https://doi.org/10.1128/IAI.00023-09

Glass EJ (2001) The balance between protective immunity and pathogenesis in tropical theileriosis: What we need to know to design effective vaccines for the future. *Research in Veterinary Science* **70**, 71–75. https://doi.org/10.1053/rvsc.2000.0428

Glass EJ, Crutchley S and Jensen K (2012) Living with the enemy or uninvited guests: functional genomics approaches to investigating host resistance or tolerance traits to a protozoan parasite, *Theileria Annulata, in Cattle. Veterinary Immunology and Immunopathology* **148**, 178–189. https://doi.org/10.1016/j.vetimm.2012.03.006

Glass EJ, Preston PM, Springbett A, Craigmile S, Kirvar E, Wilkie G and Brown CGD (2005) *Bos taurus* and *Bos indicus* (Sahiwal) calves respond differently to infection with *Theileria annulata* and produce markedly different levels of acute phase proteins. *International Journal for Parasitology* **35**, 337–347. https://doi.org/10.1016/j.ijpara.2004.12.006

Goolam M and Zernicka-Goetz M (2017) The chromatin modifier Satb1 regulates cell fate through Fgf signalling in the early mouse embryo. *Development* **144**, 1450–1461. https://doi.org/10.1242/dev.144139

Haidar M, Whitworth J, Noé G, Liu WQ, Vidal M and Langsley G (2015) TGF-β2 induces Grb2 to recruit PI3-K to TGF-RII that activates JNK/AP-1-signaling and augments invasiveness of *Theileria*-transformed macrophages. *Scientific Reports* **5**, 15688. https://doi.org/10.1038/srep15688

Henriksen J, Stabell M, Meza-Zepeda LA, Lauvrak SA, Kassem M and Myklebost O (2010) Identification of target genes for wild type and truncated HMGA2 in mesenchymal stem-like cells. *BMC Cancer* **10**, 329. https://doi.org/10.1186/1471-2407-10-329

Heussler VT, Rottenberg S, Schwab R, Küenzi P, Fernandez PC, McKellar S, Shiels B, Chen ZJ, Orth K, Wallach D and Dobbelaere DAE (2002) Hijacking of host cell IKK signalosomes by the transforming parasite *Theileria*. *Science* **298**, 1033–1036. https://doi.org/10.1126/science.1075462

Irvin AD and Morrison WI (1987) Immunopathology, immunology and immuno- prophylaxis of *Theileria* infections. In Soulsby EJL (ed.), *Immune Responses in Parasitic Infections: immunology, Immunopathology and Immunoprophylaxis Vol III Protozoa Baton Rouge*. Florida: CRC Press Inc, 223–274.

Jensen K, Paxton E, Waddington D, Talbot R, Darghouth MA and Glass EJ (2008) Differences in the transcriptional responses induced by *Theileria annulata* infection in bovine monocytes derived from resistant and susceptible cattle breeds. *International Journal for Parasitology* **38**, 313–325. https://doi.org/10.1016/j.ijpara.2007.08.007

Johnson DS, Mortazavi A, Myers RM and Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502. https://doi.org/10.1126/science.1141319

Kanda T, Horikoshi N, Murata T, Kawashima D, Sugimoto A, Narita Y, Kurumizaka H and Tsurumi T (2013) Interaction between basic residues of Epstein-Barr Virus EBNA1 protein and cellular chromatin mediates viral plasmid maintenance. *Journal of Biological Chemistry* **288**, 24189–24199. https://doi.org/10.1074/jbc.M113.491167

Kim K-D, Tanizawa H, De Leo A, Vladimirova O, Kossenkov A, Lu F, Showe LC, Noma K and Lieberman PM (2020) Epigenetic specifications of host chromosome docking sites for latent Epstein-Barr virus. *Nature Communications* **11**, 877. https://doi.org/10.1038/s41467-019-14152-8

Kim Y, Wang J, Clemens EG, Grab DJ and Dumler JS (2022) *Anaplasma phagocytophilum* Ankyrin A Protein (AnkA) enters the nucleus using an Importin-β-, RanGTP-dependent mechanism. *Frontiers in Cellular and Infection Microbiology* **12**, 828605. https://doi.org/10.3389/fcimb.2022.828605

Kumar PP, Purbey PK, Ravi DS, Mitra D and Galande S (2005) Displacement of SATB1-bound histone deacetylase 1 corepressor by the human immunodeficiency virus type 1 transactivator induces expression of interleukin-2 and its receptor in T cells. *Molecular and Cellular Biology* **25**, 1620–1633. https://doi.org/10.1128/MCB.25.5.1620-1633.2005

Larcombe SD, Capewell P, Jensen K, Weir W, Kinnaird J and Glass EJ (2022) Susceptibility to disease (tropical theileriosis) is associated with differential expression of host genes that possess motifs recognised by a pathogen DNA binding protein. *PLoS ONE* **17**, 0262051. https://doi.org/10.1371/journal.pone.0262051

Larcombe SD, Kolte SW, Ponnudurai G, Kurkure N, Magar S, Velusamy R, Rani N, Rubinibala B, Rekha B, Alagesan A, Weir W and Shiels BR (2019) The impact of tick-borne pathogen infection in Indian bovines is determined by host type but not the genotype of *Theileria annulata*. *Infection, Genetics and Evolution* **75**, 103972. https://doi.org/10.1016/j.meegid.2019.103972

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R and Stamatoyannopoulos JA (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195. https://doi.org/10.1126/science.1222794

McShea DW (1994) Mechanisms of large-scale evolutionary trends. *Evolution* **48**, 1747–1763. https://doi.org/10.1111/j.1558-5646.1994.tb02211.x

Minjauw B and McLeod A (2003) Tick-borne diseases and poverty: The impact of ticks and tick-borne diseases on the livelihoods of small-scale and marginal livestock owners in India and eastern and Southern Africa. UK: Research Report. *DFID Animal Health Programme*, Centre for Tropical Veterinary Medicine, University of Edinburgh. http://r4d.dfid.gov.uk/PDF/Outputs/RLAHTickBorn_Book.pdf

Oura CAL, McKellar S, Swan DG, Okan E and Shiels BR (2006) Infection of bovine cells by the protozoan parasite *Theileria annulata* modulates expression of the ISGylation system. *Cellular Microbiology* **8**, 276–288. https://doi.org/10.1111/j.1462-5822.2005.00620.x

Panchal O, Wichmann G, Grenman R, Eckhardt L, Kunz-Schughart LA, Franke H, Dietz A and Aigner A (2020) SATB1 as oncogenic driver and potential therapeutic target in head & neck squamous cell carcinoma (HNSCC). *Scientific Reports* **10**, 8615. https://doi.org/10.1038/s41598-020-65077-y

Quante T and Bird A (2016) Do short, frequent DNA sequence motifs mould the epigenome? *Nature Reviews Molecular Cell Biology* **17**, 257–262. https://doi.org/10.1038/nrm.2015.31

Ramanujam D, Nasrullah A, Ashraf O, Bahr M and Malik K (2021) Human Granulocytic Anaplasmosis as a COVID-19 Mimicker. *European Journal of Case Reports in Internal Medicine* **8**, 13. https://doi.org/10.12890/2021_003047

Reeves R (2001) Molecular biology of HMGA proteins: Hubs of nuclear function. *Gene* **277**, 63–81. https://doi.org/10.1016/S0378-1119(01)00689-8

Reeves R, Edberg DD and Li Y (2001) Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells. *Molecular and Cellular Biology* **21**, 575–594. https://doi.org/10.1128/MCB.21.2.575-594.2001

Rennoll-Bankert KE, Garcia-Garcia JC, Sinclair SH and Dumler JS (2015) Chromatin-bound bacterial effector ankyrin A recruits histone deacetylase 1 and modifies host gene expression: AnkA recruits HDAC1 to modify *CYBB* expression. *Cellular Microbiology* **17**, 1640–1652. https://doi.org/10.1111/cmi.12461

Sarfraz M, Afzal A, Khattak S, Saddozai UAK, Li H, Zhang Q, Madni A, Haleem KS, Duan S, Wu D, Ji S and Ji X (2021) Multifaceted behavior of PEST sequence enriched nuclear proteins in cancer biology and role in gene therapy. *Journal of Cellular Physiology* **236**, 1658–1676. https://doi.org/10.1002/jcp.30011

Sarkar A, Hellberg L, Bhattacharyya A, Behnen M, Wang K, Lord JM, Möller S, Kohler M, Solbach W and Laskay T (2012) Infection with *Anaplasma phagocytophilum* Activates the Phosphatidylinositol 3-Kinase/Akt and NF-κB Survival Pathways in Neutrophil Granulocytes. *Infection and Immunity* **80**, 1615–1623. https://doi.org/10.1128/IAI.05219-11

Scharf W, Schauer S, Freyburger F, Petrovec M, Schaarschmidt-Kiener D, Liebisch G, Runge M, Ganter M, Kehl A, Dumler JS, Garcia-Perez AL, Jensen J, Fingerle V, Meli ML, Ensser A, Stuen S and Von Loewenich FD (2011) Distinct host species correlate with *Anaplasma phagocytophilum ankA* gene clusters. *Journal of Clinical Microbiology* **49**, 790–796. https://doi.org/10.1128/JCM.02051-10

Seal S, Dharmarajan G and Khan I (2021) Evolution of pathogen tolerance and emerging infections: A missing experimental paradigm. *eLife* **10**, e68874. https://doi.org/10.7554/eLife.68874

https://doi.org/10.1017/S0031182025101492 Published online by Cambridge University Press

Sgarra R, Maurizio E, Zammitti S, Lo Sardo A, Giancotti V and Manfioletti G (2009) Macroscopic differences in HMGA oncoproteins post-translational modifications: C-terminal phosphorylation of HMGA2 affects its DNA binding properties. *Journal of Proteome Research* **8**, 2978–2989. https://doi.org/10.1021/pr900087r

Shiels B, Langsley G, Weir W, Pain A, McKellar S and Dobbelaere D (2006) Alteration of host cell phenotype by *Theileria annulata* and *Theileria parva*: Mining for manipulators in the parasite genomes. *International Journal for Parasitology* **36**, 9–21. https://doi.org/10.1016/j.ijpara.2005.09.002

Silmon De Monerri NC and Kim K (2014) Pathogens hijack the epigenome. *The American Journal of Pathology* **184**, 897–911. https://doi.org/10.1016/j.ajpath.2013.12.022

Sinclair SH, Rennoll-Bankert KE and Dumler JS (2014) Effector bottleneck: Microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure. *Frontiers in Genetics* **5**, 274. https://doi.org/10.3389/fgene.2014.00274

Singh CV (2015) Cross-breeding in cattle for milk production: achievements, challenges and opportunities in India-A Review. *Advances in Dairy Research* **4**, 3. https://doi.org/10.4172/2329-888X.1000158

Singh I, Ozturk N, Cordero J, Mehta A, Hasan D, Cosentino C, Sebastian C, Krüger M, Looso M, Carraro G, Bellusci S, Seeger W, Braun T, Mostoslavsky R and Barreto G (2015) High mobility group protein-mediated transcription requires DNA damage marker γ-H2AX. *Cell Research* **25**, 837–850. https://doi.org/10.1038/cr.2015.67

Sironi M, Cagliani R, Forni D and Clerici M (2015) Evolutionary insights into host–pathogen interactions from mammalian sequence data. *Nature Reviews Genetics* **16**, 224–236. https://doi.org/10.1038/nrg3905

Su L, Deng Z and Leng F (2020) The mammalian high mobility group protein AT-Hook 2 (HMGA2): biochemical and biophysical properties, and its association with adipogenesis. *International Journal of Molecular Sciences* **21**, 3710. https://doi.org/10.3390/ijms21103710

Swan DG, Stern R, McKellar S, Phillips K, Oura CAL, Karagenc TI, Stadler L and Shiels BR (2001) Characterisation of a cluster of genes encoding *Theileria annulata* AT hook DNA-binding proteins and evidence for localisation to the host cell nucleus. *Journal of Cell Science* **114**, 2747–2754. https://doi.org/10.1242/jcs.114.15.2747

Treff NR, Pouchnik D, Dement GA, Britt RL and Reeves R (2004) High-mobility group A1a protein regulates Ras/ERK signaling in MCF-7 human breast cancer cells. *Oncogene* **23**, 777–785. https://doi.org/10.1038/sj.onc.1207167

Tseng CC, Wong MC, Liao WT, Chen CJ, Lee SC, Yen JH and Chang SJ (2021) Genetic Variants in Transcription Factor Binding Sites in Humans: Triggered by Natural Selection and Triggers of Diseases. *International Journal Molecular Sciences* **22**, 4187. https://doi.org/10.3390/ijms22084187

Vignali R and Marracci S (2020) *HMGA* genes and proteins in development and evolution. *International Journal of Molecular Sciences* **21**, 654. https://doi.org/10.3390/ijms21020654

Wang X, Wang J and Wu J (2021) Emerging roles for HMGA2 in colorectal cancer. *Translational Oncology* **14**, 100894. https://doi.org/10.1016/j.tranon.2020.100894

Weir W, Karagenç T, Baird M, Tait A and Shiels BR (2010) Evolution and diversity of secretome genes in the apicomplexan parasite *Theileria annulata*. *BMC Genomics* **11**, 42. https://doi.org/10.1186/1471-2164-11-42

https://doi.org/10.1017/S0031182025101492 Published online by Cambridge University Press