Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Diverse genetic conflicts mediated by molecular mimicry and computational approaches to detect them

Shelbi L. Russell [a,b], Gabriel Penunuri [a,b], Christopher Condon [a,b,*]

[a] Genomics Institute, University of California Santa Cruz, Santa Cruz, CA, United States
[b] Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA, United States

A B S T R A C T

In genetic conflicts between intergenomic and selfish elements, driver and killer elements achieve biased survival, replication, or transmission over sensitive and targeted elements through a wide range of molecular mechanisms, including mimicry. Driving mechanisms manifest at all organismal levels, from the biased propagation of individual genes, as demonstrated by transposable elements, to the biased transmission of genomes, as illustrated by viruses, to the biased transmission of cell lineages, as in cancer. Targeted genomes are vulnerable to molecular mimicry through the conserved motifs they use for their own signaling and regulation. Mimicking these motifs enables an intergenomic or selfish element to control core target processes, and can occur at the sequence, structure, or functional level. Molecular mimicry was first appreciated as an important phenomenon more than twenty years ago. Modern genomics technologies, databases, and machine learning approaches offer tremendous potential to study the distribution of molecular mimicry across genetic conflicts in nature. Here, we explore the theoretical expectations for molecular mimicry between conflicting genomes, the trends in molecular mimicry mechanisms across known genetic conflicts, and outline how new examples can be gleaned from population genomic datasets. We discuss how mimics involving short sequence-based motifs or gene duplications can evolve convergently from new mutations. Whereas, processes that involve divergent domains or fully-folded structures occur among genomes by horizontal gene transfer. These trends are largely based on a small number of organisms and should be reevaluated in a general, phylogenetically independent framework. Currently, publicly available databases can be mined for genotypes driving non-Mendelian inheritance patterns, epistatic interactions, and convergent protein structures. A subset of these conflicting elements may be molecular mimics. We propose approaches for detecting genetic conflict and molecular mimicry from these datasets.

1. Introduction to genetic conflict and molecular mimicry

Genetic conflicts between genes, genomes, and individuals drove the evolution of many emergent biological phenomena, from mobile element-repressing heterochromatin, to endosymbiont-derived organelles, to protein mimicry [1–4]. Extant selfish genetic elements include selfish chromosomes, plasmids, organelles, cell lineages (cancer), and transposable elements. Intergenomic genetic conflicts occur between viruses, symbiotic bacteria, and symbiotic fungi and their hosts (Fig. 1A). In all of these examples, conflicts are waged at the molecular level to promote the survival, abundance, and transmission of an intergenomic or selfish driving element to the detriment of sensitive target elements. When a driver is lethal to sensitive elements, it is termed a killer. Although essential proteins involved in replication and

transmission can theoretically be targeted through novel mechanisms, mimicry of existing molecular mechanisms provides a fast-track for driving elements to bias cellular functions [2,3]. For example, diverse bacterial pathogens and viruses gain access to host cells by co-opting small receptor-binding motifs and full domains to induce cellular uptake or niche formation.

Many forms of molecular mimicry exist for mediating intergenomic conflicts between pathogens and hosts, as well as intragenomic conflicts between selfish elements in a single genome. Mimicry may be based on shared short motifs or complex structures. The genes encoding molecular mimics can arise in divergent organisms *de novo* by mutation and convergence on target functions and/or structures. Or they can arise through horizontal transmission from another lineage [2,5]. *We define molecular mimics to be DNA, RNA, or proteins that have converged on a*

function, structure or sequence, to co-opt a process imparted by a target element, resulting in the redirection of the products, rewards, or implications of that process to the genetic element encoding the mimic. Molecular mimics are distinguished as "perfect" or "imperfect" based on how closely they reproduce the full mechanism of the element or phenotype being imitated. Studying mechanistic trends among these elements provides insight into how likely molecular mimics are to evolve, which functional strategies they employ, and how we might identify them in genetic datasets.

Molecular mimic detection has historically been challenging because validation involves carefully controlled experiments. Genetic conflict distorts allele frequencies from Mendelian expectations, permitting signatures of conflict to be detected from genome-wide population

sequencing data [7,8]. However, only a subset of these elements will use molecular mimicry. Driving and sensitive haplotypes can be large and in strong linkage due to structural variation, such as chromosomal inversions, that prevents recombination. This linkage makes ascertaining the causative distortion loci in these multigenic haplotypes difficult. Elucidating the functional molecular basis of the distortion, which is required for molecular mimic identification, is exceedingly challenging because few organisms are genetically and developmentally tractable. Thus, knowledge of molecular mimics in genetic conflicts is limited and biased towards genetically tractable systems.

Until genome-wide, culture-independent screens are leveraged to find mimics broadly, the occurrence and distribution of molecular mimicry across life is informed through careful experiments in a subset
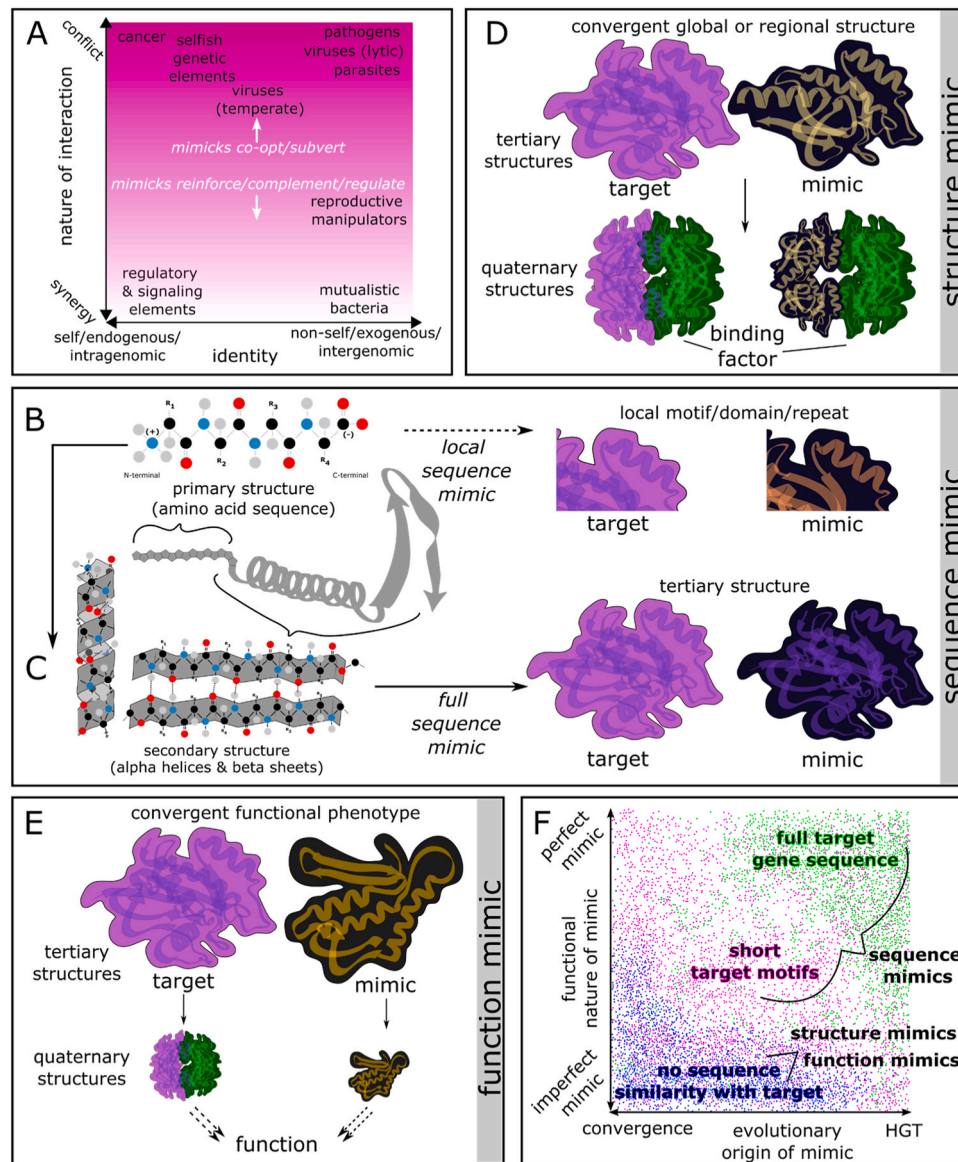

Fig. 1. Models of molecular conflict evolution and selection for molecular mimics. A) Ecological molecular mimic categories are distinguished by the nature of the interaction, *i.e.*, whether they conflict or synergize with organismal fitness, and the genetic relationship with the selfish element, *i.e.*, whether they are from the same genome or different, potentially highly divergent, genomes. B-E) Idealized illustrations of the three types of molecular mimics: B,C) sequence-based, D) structure-based, and E) functional mimics. F) A hypothetical distribution of molecular mimic genes arranged by their evolutionary origin and their functional nature. Mimics arise by complete convergence on target features from a dissimilar starting point or HGT from an exogenous source. Functionally, mimics may be perfect, if they fully and completely recapitulate the activity of the target they mimic, or imperfect if they have invented their own ways of inducing functional outcomes. Full genes have a low probability of evolving from convergence alone, whereas short motifs of a few residues evolve more readily from scratch. Functional mimics employ imperfect mimicry by definition. Perfect mimics are often sequence mimics, but may be very good structural mimics as well. Structural and functional mimics that initially evolve by convergence may be subsequently horizontally transferred to new genomes for new conflicts.

of model systems. Model systems for genetic conflict, ranging from selfish *Drosophila* chromosomes to SARS-CoV-2 viruses, have been studied well enough at the mechanistic level to inform on the possible forms of molecular mimicry. These investigations have revealed diverse manifestations of mimicry, prompting its description since the early 2000s [5,9–11]. By the 2010s, the power of whole genome sequencing and accumulating genomic data in public databases made *in silico* screens for molecular mimicry possible [12–15]. The power of these approaches strengthens every year, as more genomes across the tree of life are sequenced and as new tools, such as AlphaFold [16], are developed for computationally predicting and comparing molecular function [17]. Full knowledge about the distribution of molecular mimicry in nature awaits genome mining studies across the tree of life. Soon the necessary genomic data and protein-interaction tools will be available to address this question organism-agnostically.

In the sections below, we explore 1) what types of molecular mimics are expected to occur based on evolutionary theories, 2) empirical examples of mimicry in nature, and 3) approaches for identifying genetic conflicts involving molecular mimics from genome-wide sequencing data (full list of literature in Table S1). Through this theoretical and empirical survey, we conclude that convergently evolved mimics are generally limited to short motifs in intergenomic conflicts. Full-length sequence and structure molecular mimics arise in intragenomic conflicts via gene duplication events and in intergenomic conflicts via horizontal gene transfer (HGT). Functional mimics tend to involve entirely novel genes and pathways to reconstruct the mimicked molecular function. The few well-characterized examples of functional mimicry occur in intergenomic conflicts. Through leveraging the bioinformatic approaches for mimicry detection we present in the final section, we hope that more molecular mimics will be identified and these hypotheses about their distributions can be empirically tested in the future.

2. Distribution of molecular mimicry across life: expectations from molecular evolutionary theory

Genetic conflicts can be mediated by molecular mimicry at functional, structural, and genetic (*i.e.,* sequence) levels [5] (Fig. 1B-E). Mimicry of cellular and molecular functions at its highest level does not require similar protein structures or genetic sequences. If a conflicting element has evolved an entirely new pathway to accomplish the same functional goals, this is termed functional mimicry [18]. Often, molecular mimics evolve similar protein structures as their target due to constraints on binding and interaction with other factors in the functional molecular network. Proteins engaging in structural mimicry share similar 3-dimensional conformations, despite not sharing extensive sequence similarity at the amino acid level. They arise convergently, from divergent elements that fold into conformations similar to the structure and activity of mimicked elements [7]. Compared to coevolving convergent functions and structures, sequence-based molecular mimicry provides a quick path to coopting target functions. Sequence-level mimicry can occur via single domains, multi-domain repeats, or full-length homology that translates to identical structures upon expression of those elements [5,18–20].

The expected distribution of molecular mimics across life depends on phylogenetic constraints, the complexity of the mimicked element, and the relationship between conflicting elements (*i.e.,* intragenomic vs intergenomic conflict). The gene content of every genome is structured by its phylogenetic history, with potential exceptions due to HGT. The infinite sites model of molecular evolution [21] predicts that each mutation will occur only once across evolutionary history because mutation rates are generally low enough to make a mutation at any given position exceedingly unlikely [21,22]. Although there are key examples of recurrent mutations at specific residues due to strong selection and convergent evolution [23–25]. Thus, stretches of high sequence identity between divergent genomes are more likely to be due to the reshuffling

of genetic variation among genomes by recombination and HGT than by random mutations. *Overall, these factors inform our expectations that the rate of convergent sequence-based mimic evolution to be inversely proportional to gene length in intergenomic conflicts.* These trends do not apply to intragenomic conflicts, as selfish mimics can be created by sequence duplication events in one genome (discussed in Section 3). The frequency of structural and functional mimics may also be related to the length of the underlying genetic elements in intergenomic conflicts, as the evolution of long elements may require more time and coevolutionary pressures than short elements to acquire the necessary substitutions.

Driver elements experience evolutionary pressures to mimic target functions, structures, or sequences, which stem from the conserved, endogenous motifs used by the target genome itself [26]. Endogenous molecular networks rely on conserved binding sites and signaling motifs to make efficient use of gene products across diverse molecular pathways. This conservation of interaction parameters enabled the evolution of diverse tissue types, developmental programs, and environmentally-driven phenotypic plasticity across multicellular eukaryotes [27–29]. However, it also selected for conflicting elements that happened across conserved motifs, structures, or functions and could use them for their own advantage. Viruses and host-associated bacteria contain a high abundance of eukaryotic-like domains due to both convergence and HGT [18,20,30]. The HGT domains often function to mimic homologous elements in the target genome. For example, the eukaryotic-like sphingosine-phosphate lyase from *Legionella pneumophila* intracellular bacteria was horizontally transferred from one of their protozoan host ancestors and functions to inhibit macrophage autophagy [18].

In the next section, we detail examples of molecular mimicry in genetic conflicts at sequence, structural, and functional levels to assess whether the current empirical data support their expected distributions. These examples are organized by mechanistic impact on their target's cell biology. Overall, we observe distinct molecular mimic distributions between intergenomic and intragenomic conflicts that are consistent with theoretical expectations (summarized in Fig. 1F).

3. Empirical examples of molecular mimicry in nature

3.1. Sequence-based mimics

The evolution of sequence-based molecular mimicry is largely structured by selection for convergent short motifs, full sequence duplications, and elements that have undergone HGT. Mimicry of short motifs is a common trend among viruses, pathogenic bacteria, and cancerous cell lineage proteins [19,31,32]. Whereas, convergent full-length sequence mimics are more common in intragenomic, selfish conflicts [33–35]. The number of substitutions required to reach high sequence identity bound these distributions. Full-length homology between a mimic and a divergent target's nucleic or amino acid sequence suggests that the driving element arose via HGT, and subsequently coevolved for intergenomic conflict [5].

3.1.1. Recurrently evolved SLiM ligand mimics hijack eukaryotic cell biology

Eukaryotic proteomes leverage short linear motifs (SLiMs) that are three to 15 residues long for protein-protein interactions including signaling, post-translational modifications (PTMs), and localization. These motifs are highly conserved, but are surrounded by lower-conservation sequences that modify the specificity of SLiM-protein interactions. SLiMs often occur in intrinsically disordered regions of proteins, allowing the binding surfaces to be accessed [26,31]. These short elements, and their DNA and RNA precursors, enable genomes to leverage existing molecular networks to evolve new functions and phenotypic plasticity [39]. Indeed, the human proteome is estimated to encode more than 100,000 SLiMs [26]. These endogenous mimics

potentially drive mimicry in genetic conflict indirectly through increased selection pressures for intergenomic and selfish elements to evolve highly-utilized target/host motifs.

A common theme among bacterial pathogens seeking to modify eukaryotic cells is mimicry of the ligand RGD (arginine-glycine-aspartate) SLiM domain that binds to integrin, an essential host cell-adhesion receptor [32]. Integrin signaling induces actin polymerization to strengthen cellular adhesion to bound substrates (Fig. 2A). This is critical to normal tissue differentiation and is co-opted in many cancers for metastasis and tumor formation [40]. By recognizing the small RGD domain in various ligands, integrin is able to bind many factors, including laminin, fibronectin, and tenascin glycoproteins. Intergenomic elements mimicking integrin ligands include the gastrointestinal pathogen *Helicobacter pylori*, which manipulates host integrin signaling through a RGD motif in its CagL protein [41] (Fig. 2B). CagL is an extracellular component of the *H. pylori* type IV secretion system (T4SS) that induces CagA effector protein injection into host cells upon integrin binding to induce cellular differentiation [32,42]. The mammalian cell entry protein (Mce) Mce3C from *Mycobacterium tuberculosis* contains a RGD motif that binds β2 integrin to induce local actin rearrangements that permit bacterial invasion [37] (Fig. 2C). In addition to the RGD motif, many of these proteins contain other motifs mimicking other extracellular matrix (ECM) interactions that modulate the outcome of integrin binding [41]. For example, the SARS-CoV-2 spike protein S1 contains an RGD motif that binds to host integrin and ACE2 receptors to induce cell spreading, adhesion, and proliferation [43,44].

Histone tail SLiM mimicry is a common trend among viruses and cancer cell lineages. These short, degenerate, fast-evolving sequences facilitate histone-protein binding to modulate chromatin compaction and transcriptional regulation [19,45]. Numerous viral proteins mimic histone motifs to hijack host gene expression, thereby repressing anti-viral gene expression and upregulating their own replication [45]. The histone H3 lysine 4 (H3K4)-like motif of the Influenza non-structural protein 1 (NS1) binds to the human PAF1 transcription elongation complex (hPAF1C) to suppress transcriptional elongation [46]. The NS1 histone mimic is imperfect, as its methylation can not be reversed by histone H3K4 demethylase [6]. SARS-CoV-2 encodes at least two proteins that mimic host histone SLiMs, the ORF8 and envelope (E) proteins. ORF8 mimics host histone H3 ARKS (alanine, arginine, lysine, serine) motifs to disrupt epigenetic regulation, promoting chromatin compaction and increasing virus copy number [38] (Fig. 3A,B). Mimicry of healthy histone tail binding is also leveraged by intragenomic conflicts, including cancer. Basal-like breast cancer cells overexpress the transcription activator Twist, which contains a histone H4 motif. Inappropriate twist expression mimics the epithelial-mesenchymal transition (EMT) pathway, which is required for altered cell-to-cell contacts in tumors and metastatic cells [47]. Mimicking an EMT transition is strongly selected in cancer because migration to new somatic niches enables cancer cells to increase their frequency in the population.

3.1.2. De novo full-length sequence mimics arise from gene duplications

Intragenomic conflicts provide examples of full-length sequence mimicry that arose via gene duplication. Many of these duplication-based drivers operate via imperfect molecular mimicry of the original, functional element to obstruct its function. Gene duplications can result in inhibitory interactions between the original gene and its duplicate through dominant negative interactions [34] or RNA-interference (RNAi) from the negative-sense strand. If these duplicates are expressed in the appropriate tissues at the right times, these interactions can promote the transmission of the driving element and create endogenous genomic conflict *de novo*.

In *Drosophila*, many meiotic driver, responder, and suppressor elements arose from gene duplication events on autosomes or sex chromosomes [33] and some of these duplicated genes function as molecular mimics. Meiotic drive is an intragenomic conflict strategy by which a selfish chromosome biases its transmission into the gametes that will contribute genetic material to the next generation, either through disrupting segregation or poisoning target chromosomes. *Drosophila melanogaster*'s autosomal *Segregation Distorter* system is one of the best-studied examples of male meiotic drive that employs a driver-responder system. In normal gametogenesis, sperm chromatin must be unwound from large nuclear histones and repackaged with small protamines to fit into the small confines of the sperm head nucleus. Nuclear import of protamines and other cofactors requires high amounts of nuclear RAs-related Nuclear protein (Ran) bound to GTP (RanGTP), which are maintained by the *RanGTPase Activating Protein* (RanGAP). The selfish SD haplotype encodes a driver locus derived from a partial duplication of RanGAP (Sd-RanGAP), lacking the nuclear export signal and SUMO domain. This truncation traps Sd-RanGAP in the nucleus where it mimics RanGAP function in the wrong compartment: Sd-RanGAP catalyzes the hydrolysis of nuclear RanGTP, disrupting RanGTP/GDP gradients, and blocking transport through nuclear pore channels (NPCs) (Fig. 3C,D). Excess RanGDP blocks nuclear envelope and heterochromatin formation, which explains how this single gene can disrupt both protamine import and chromatin repackaging during spermatogenesis, respectively [33,48]. This protein localizes to both spermatids during cellularization, but it is ineffective on the SD haplotype, which also encodes an insensitive responder locus. Although the mechanism is still being elucidated, it is clear the tandem repeat length of the satellite responder locus underlies the severity of the target haplotype's response to drive [49].

The *D. melanogaster* genome contains examples of gene duplication-derived mimics in intragenomic conflicts that are older than the SD system, and have reached fixation across populations. These include the *Stellate* (*Ste*) system, which is derepressed in hybrid matings between males lacking the Y-linked repressor element and females encoding the X-linked driving element. In normal spermatogenesis, the X-linked testis-specific regulatory beta subunit of casein kinase II (CkIIβ) controls
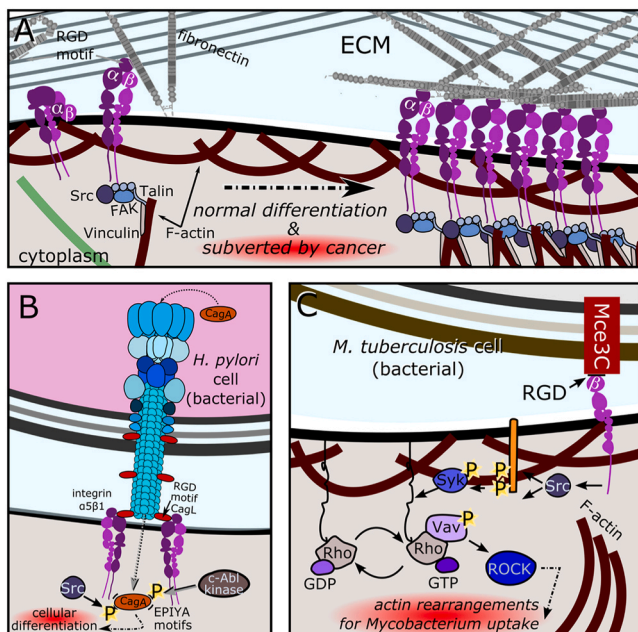

Fig. 2. Molecular mimicry of SLiM motifs mediate intragenomic and intergenomic genetic conflicts at the cell surface. A) Fibronectin-integrin binding in the extracellular matrix (ECM) mediates cellular adhesion, migration, and differentiation in normal cells, and can be co-opted by metastatic cancerous cell lineages for metastasis. B) The CagL protein in the *H. pylori* T4SS mimics integrin-binding RGD motifs. Once injected into the host cytoplasm, the CagA protein's EPIYA motifs trigger phosphorylation, leading to functional mimicry of the Ras signaling pathway [36]. C) The *M. tuberculosis* Mce3C membrane protein mimics RGD-integrin binding, inducing actin polymerization and bacterial engulfment [37].
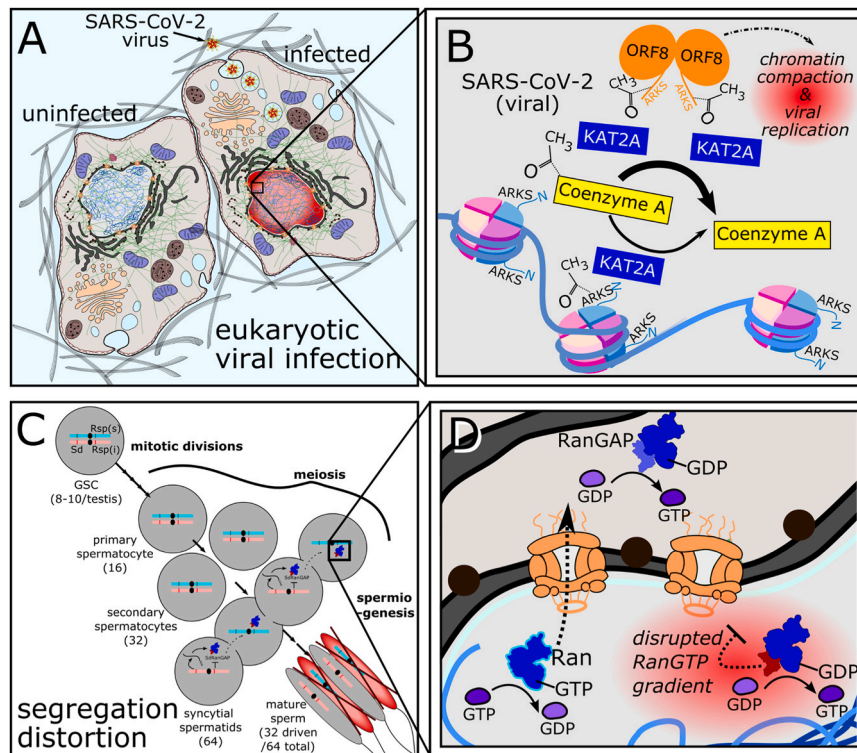
Fig. 3. Molecular mimicry of SLiM motifs and large domains mediate genetic conflicts in the somatic and germline nucleus. A) Eukaryotic cell infected with SARS-CoV2 virions, which express B) ORF8 proteins containing ARKS motifs. This motif mimics histone tails and alters histone acetylation rates by co-opting the KAT2A lysine acetyltransferase for viral acetylation [38] (note wider rate arrow for viral acetylation compared to host acetylation). C) Male germline autosomal segregation distortion in *Drosophila* depends on two loci: the Sd driver (red chromosome) and the sensitive Rsp locus (cyan chromosome). Sd-RanGAP is produced by Sd chromosomes, and diffuses into non-Sd-containing sperm through cytoplasmic bridges prior to cellularization. D) In the presence of the Rsp(sensitive) sequence, truncated Sd-RanGAP localizes to and is trapped in the nucleus, mimicking GTP hydrolysis, but in the wrong compartment. Disrupted RanGTP gradients block protamine import and stop spermiogenesis in sensitive haplotypes (blue).

casein kinase II (CkII) phosphorylation of Wnt pathway components. Extant *D. melanogaster* X chromosomes carry up to 400 partial copies of the N-terminus of CkIIβ (termed the Ste locus) that bind to the alpha subunit to produce crystalline structures in male premeiotic germ cells that prevent their development. These inhibitory mimic proteins are blocked by the expression of a Y-linked repressor locus, termed suppressor of stellate (*Su(Ste)*). This repressor is a duplication of *Ste* that contains a transposon insertion positioned downstream. This insertion event created a promoter that induces the transcription of the opposite strand of *Su(Ste)*, leading to double-stranded RNA (dsRNAs) accumulation. These dsRNAs are cleaved into repeat-associated small RNAs (rasi-RNAs), which function in the piwi-interacting RNA pathway to inhibit *Ste* expression [50].

3.1.3. Horizontally transmitted full-gene mimics are repurposed for conflict across life

Despite the low likelihood of identical genes evolving multiple times by chance, many elements are well-distributed across taxa, indicating that HGT is a powerful force in shuffling genetic variation that can lead to molecular mimicry. Sequence-based molecular mimics of eukaryotic repeats and domains are found in many bacterial and viral genomes, including ankyrin repeats, leucine rich repeats, tetratricopeptide repeats, and pentatricopeptide repeats, SRC Homology 3 (SH3) domains, and RING domains [18,30]. Similarly, toxin-antitoxin (TA) systems likely evolved among mobile genetic elements, and have subsequently transferred to bacterial, archaeal, and eukaryotic chromosomes [51].

Receptor proteins enable cellular life to sense and respond to stimuli, but leave cells vulnerable to viral entry and manipulation. The *Vibrio cholerae* genome encodes the cytoplasmic receptor-transcription factor Vibrio quorum modulator A (VqmA$_{VC}$) that binds the quorum sensing autoinducer ligand 3,5-dimethylpyrazin-2-ol (DPO). DPO binding to VqmA$_{vc}$, enables *V. cholerae* to sense and respond to population size changes. At high *V. cholerae* densities, excess DPO binding suppresses the expression of virulence effectors and biofilm factors. The lysogenic temperate vibriophage of *V. cholerae* VP882 also encodes a copy of VqmA, termed VqmA$_{phage}$, that it acquired through a HGT event. Phage-encoded VqmA$_{phage}$ binds DPO, induces the viral lytic program at peak bacterial densities [52]. Thus, through acquisition of a bacterial quorum sensing receptor by HGT, vibriophage are able to mimic host density-dependent growth and transmission responses.

TA systems are excellent examples of highly effective, compact, and versatile elements for controlling cellular outcomes that are present at high frequencies in mobile elements such as transposons and plasmids. These two-element systems consist of a protein-encoded toxin and an antitoxin that is an RNA or protein that directly or indirectly neutralizes the toxin [53]. TA systems induce plasmid addiction and viral infection-induced abortions in bacteria, as well as post-segregational killing of cells that lack the antidote in both bacteria and eukaryotes [51]. While TA systems are common players in genomic conflicts, molecular mimics are more common in suppression mechanisms than in the TA systems themselves. For example, the bacterium *Pectobacterium atrosepticum* hosts bacteriophage in the family Myoviridae and encodes a TA system to induce altruistic suicide of virus-infected clonal lineages. However, the bacteriophage acquired the RNA-based antidote gene from the TA system by HGT, and uses it to mimic the host's antitoxin activity to repress suicide, permitting its replication and lysis from host cells [53, 54]. Similarly, in *Schizosaccharomyces pombe* fungi, a TA system is employed by selfish chromosomes to poison spore haplotypes lacking the antidote gene. Through non-allelic gene conversion, a non-selfish chromosome acquired the protein-based antidote, endowing it with

toxin resistance through antidote mimicry [55].

3.2. Structural mimics

Similar protein structures can be encoded by highly divergent amino acid sequences, enabling convergent evolutionary routes to shared molecular functions. Viruses and pathogenic bacteria have explored this structural space in their evolution, producing a plethora of novel strategies for mimicking host molecular interactions [8]. However, selfish and intergenomic conflicts with linked fitness may be more restricted in the structures they can evolve, as protein mimics can be antigenic and induce autoimmunity or coevolutionary responses [5,56]. Thus, some structural mimicry may result from constrained selection for elements that reappropriate functions without triggering cellular or humoral immunity.

Structural mimicry also enables genetic elements in conflict to diverge from their own endogenous interaction networks, while converging on the function of their target. This strategy simultaneously enables mimicked function while preventing the disruption of existing essential networks in the conflicting genome [57]. Examples of structural mimics have been known for decades, despite the difficulty in detecting them prior to accurate *ab initio* protein structure prediction tools, through protein crystal structures and binding data. Recent advances in protein structure and interaction predictions are enabling the discovery of many more structural mimics, both across genomes and taxa [17,58,59].

3.2.1. Repurposed ancient domains for molecular mimicry

Ancient, conserved molecular machinery essential to housekeeping functions provides a slow-moving target for genetic conflict to co-opt function. For example, both bacteria and eukaryotes use PTMs, such as phosphorylation, to modulate the activity of their enzymes and signaling proteins. Through employing their own PTMs on host proteins, conflicting intergenomic elements can induce large downstream impacts on host signaling pathways. In contrast to the small motifs that trigger PTMs on bacterial and viral effector proteins by eukaryotic enzymes, larger repeats and domains are required for exogenous elements to conjugate PTMs onto eukaryotic factors [32].

Protein tyrosine phosphatase (PTP) domains catalyze the removal of phosphate from activated residues, are present in all domains of life, and have been leveraged in genetic conflicts. For example, PTP domains are encoded on *Salmonella's* SptP and *Yersinia's* YopH bacterial effector proteins. These effectors are injected into the host cytoplasm through protein secretion systems, where they dephosphorylate host proteins to restore cellular homeostasis and prevent phagocytosis, respectively [60,61]. Some viruses encode type III dual function PTP domains, which can dephosphorylate activated serine and threonine residues, in addition to tyrosine. Overall the structures of these enzymes strongly resemble their eukaryotic targets [60], which likely reflects both deep conservation of PTP function as well as convergence on the host structural conformation.

3.2.2. Convergent structural mimics among membrane-bound and secreted effector proteins

Ligand structural mimicry enables genetic elements in conflict to bind target receptors in order to block immune activation, induce cellular signaling, and gain access to targeted cells. As discussed above, eukaryotic extracellular matrix (ECM) proteins bind to integrin and other cell surface receptors to induce normal development and cellular maintenance. Structural mimicry of ECM binding surfaces is a common strategy leveraged by intergenomic elements to promote internalization, and provides an alternative mechanism to sequence-based RGD-motif mimicry. For example, the *Yersinia pseudotuberculosis* bacterial Invasin protein is a membrane-embedded trimeric autotransporter adhesin that mimics the structure of eukaryotic integrin αv to bind to integrin β1 for cell surface attachment. Similarly, West Nile Virus glycoprotein E

mimics the fibronectin FN10 domain for integrin αVβ3 binding [15]. Through searching proteomic databases for potential binding partners of the coronavirus spike protein receptor binding motifs, a new epidermal growth factor (EGF) receptor binding motif was discovered that mimics host receptors and may determine cellular tropism [62].

Cell surface glycan modifications are commonly mimicked among cancer cell lineages, viruses, and bacterial toxins. In non-selfish *in vivo* cell biology, glycan modifications play significant roles in cellular adhesion and signaling cascades that impact differentiation, metabolism, immunity, and other processes [63]. Loss of cell surface glycans enables cancer cells to evade immune detection. Glycan co-option bestows different metastatic lineages with the ability to colonize new tissues and form tumors [64]. Bacterial pathogens such as *Campylobacter jejuni, Neisseria* species*, H. pylori,* and some *E. coli* strains possess cell walls composed of lipooligosaccharides (LOS), lipopolysaccharides (LPS), or capsular polysaccharides (CPS) decorated with carbohydrate modifications that mimic eukaryotic lectin-binding glycans. For example, an altered glycan linkage in *Neisseria meningitidis* serogroup B capsules enables recruitment of host factor H, which is required to deactivate the immune complement pathway induced by bacterial infection. In *Neisseria gonorrhoeae,* LOS with broken sialic acid linkages bind to the lactosamine residues of eukaryotic asialoglycoprotein receptors, triggering clathrin-mediated cell entry. The O-antigen of *H. pylori's* LPS membrane is a mimic of Lewis blood group antigens, which enables colonization and invasiveness [65]. In viruses, glycan mimicry often takes the form of a glycan shield that enables immune evasion. For example, HIV envelope proteins and Influenza A hemagglutinin glycoproteins add host-derived N-linked glycans near critical residues to sterically block neutralizing antibody binding [66] (Fig. 4A). Enveloped viruses are able to mimic apoptosis cues by collecting phosphatidylserine from organelle membranes and displaying it to receptors such as T cell immunoglobulin and mucin receptor 1 (TIM1) to trigger uptake by phagocytic cells [67].

Vesicle traffic is vital for cellular homeostasis and is commonly manipulated in intergenomic conflicts. Through the regulation of membrane budding, cytoskeletal transport, and fusion with cellular organelles, vesicle traffic plays vital roles in many cellular processes. SNARE (soluble N-ethylmaleimide-sensitive factor (NSF) attachment protein receptor) proteins mediate and regulate vesicle fusion with target membranes. Emerging evidence indicates that diverse pathogens, including *Chlamydia* and *Legionella,* use SNARE protein mimicry to prevent fusion with autophagosomes, expand the replicative niche, and control membrane traffic (Fig. 4B). SNARE proteins consist of transmembrane domains connected with linker sequences to coiled-coil motifs that enable membrane fusion. Based on the sequence and structural diversity among *Legionella* LegC SNARE mimics, and their ability to complement the loss of endogenous SNAREs, the fusion mechanism appears to be highly permissible to structural variation [68]. LegC fusion with liposome vesicles containing the endosomal arginine (R)-SNARE vesicle-associated membrane protein 4 (VAMP4) provides extra membrane and cytosolic resources for bacterial niche expansion. However, LegC is an imperfect SNARE mimic. Once LegC fuses with the VAMP4 protein, it cannot be disassembled by the disassembly machinery and rendered inactive.

Many different actin and microtubule-associated proteins are mimicked by pathogenic bacteria and viruses for movement within a eukaryotic cell [69]. Actin nucleation proteins, such as the Wiskott-Aldrich syndrome (WASP) family of proteins, are essential for building and recycling the cytoskeleton and vesicles. Microtubule-associated proteins permit long-range directed transport of cellular cargos in normal cell biology. Intergenomic conflicts have evolved mimics for both of these processes. The membrane-associated ActA protein from *Listeria monocytogenes* is a structural mimic of WASP actin nucleation proteins, which enables the bacteria to polymerize actin, propelling themselves from cell to cell. Following phosphorylation by host casein kinase II (CK2), ActA interacts with the
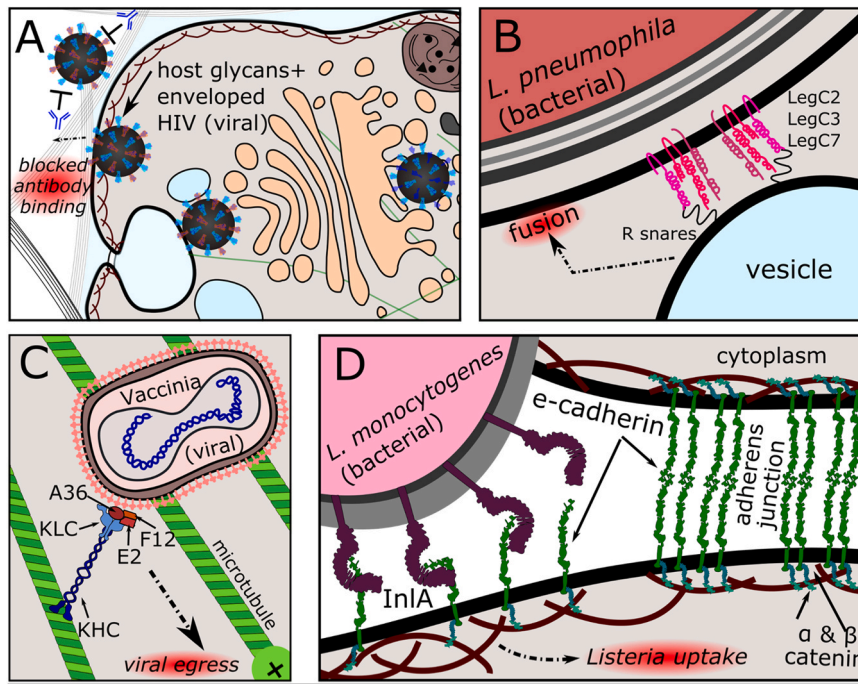
Fig. 4. Intergenomic conflicts leverage numerous examples of structural and functional molecular mimicry for entry to and egress from host cells. A) HIV lipid envelopes are decorated with proteins that undergo N-linked glycosylation as they mature, enabling them to mimic host-self signals and evade immune detection. B) Intracellular *L. pneumophila* encodes SNARE protein mimics that control host vesicle fusion. C) Mature vaccinia virus encodes motor protein adapter mimic proteins that co-opt KHC-mediated plus-end microtubule transport for egress. D) To promote intracellular uptake, *L. monocytogenes* functionally mimics adherens junction membrane tension to restructure the actin cytoskeleton.

actin-related protein (ARP) 2/3 complex through a region that displays structural similarity to the C region of the VCA domain of WASP proteins to induce polymerization. While these phosphorylation sites are shared among divergent WASP proteins, the rest of the protein is unique to *L. monocytogenes*, indicating that this mimic evolved by convergence [70]. The microtubule cytoskeleton also offers opportunities for intergenomic conflicts to co-opt cellular transport: Vaccinia virus encodes a linker protein mimic that enables it to commandeer host Kinesin heavy chain motor proteins for transport to the plasma membrane from the cytoplasm for cell exit (Fig. 4C). Similarly, HIV's capsid protein contains a region that is a structural mimic of the microtubule-associated End-binding1 (Eb1) protein, enabling it to bind to Cytoplasmic linker protein-170 (CLIP-170) for dynein-mediated minus-ended microtubule transport to the nucleus [69].

3.3. Functional mimics

The molecular armaments of genetic elements in conflict do not need to share sequence or structural similarity with the factors they are mimicking to hijack cellular functions. This functional mimicry is likely the only available strategy for some element types. For example, a sequence-based or structural mimic may not be possible for endogenous proteins that leverage intrinsically disordered regions for dynamic interactions. Importantly, mimic categories are not mutually exclusive. The bacterial CagA protein from *H. pylori* mimics a tyrosine phosphorylation SLiM motif at the sequence level for proximal molecular interactions, but mimics a Grb2-associated binder (Gab) family adapter protein at the functional level for signaling consequences [36]. Thus, more examples of functional mimicry will likely be reported as the phenotypic consequences of genetic conflict are fully characterized in different systems.

3.3.1. Convergent charge distributions for DNA mimicry

More than a dozen independently evolved bacterial and viral proteins have been reported to mimic host DNA to subvert protein-DNA binding. DNA-binding proteins are required by cellular genomes to regulate genome replication, gene expression, and epigenetic modifications. These binding proteins have evolved to recognize DNA phosphate backbones generally or a sequence of nucleotide residues specifically [71]. Proteins that functionally mimic DNA binding do so by evolving helical strands of negatively charged phosphate residues, with one or two rows of either aspartic acid or glutamic acid residues. The majority of viral DNA mimics inhibit host proteins from targeting invading genomes. For example, bacteriophage T7 uses its Overcome classical restriction (Ocr) protein to block the host type I restriction modification (R/M) system from cleavage. Ocr proteins dimerize to form a curved structure that mimics the angle of bent B-form DNA and its negatively charged surface to bind and block the R/M enzyme [72].

3.3.2. Horizontally transferred extracellular matrix functional mimics

Selection for extracellular ligand mimics is strong in intergenomic conflicts, resulting in the evolution of functional ligand mimics, in addition to the sequence and structural ligand mimics discussed previously. The adherens junctions is an emergent structural complex created through protein-protein binding between adjacent cells. Binding proteins include the transmembrane glycoprotein E-cadherin and intracellular components including p120-catenin, β-catenin and α-catenin. In normal adherens junctions, repeating E-cadherin subunits bind themselves across the junction (Fig. 2A). These cellular adhesion structures not only provide scaffolding and structure for cells, they communicate environmental signals and mediate cooperative cellular behaviors, such as epithelial barrier formation and maintenance [73,74]. The food-borne pathogen *L. monocytogenes* uses its internalin protein InlA as a functional mimic of E-cadherin to gain entry to the host cell. LnlA binding to E-cadherin triggers junction recycling through clathrin or caveolin-mediated endocytosis, engulfing the bacteria in the process [75] (Fig. 4D). The InlA protein evolved as a novel structure consisting of fifteen leucine-rich repeats (LRRs) surrounding an inter-repeat region

that binds E-cadherin, which arose by ancient horizontal gene transfer, followed by divergence and high rates of recombination among *Listeria* strains [76].

3.3.3. Bacteria convergently evolved a functional mimic system for modifying ubiquitination

Some host-associated bacteria have evolved a structurally and mechanistically novel set of proteins to leverage eukaryotic ubiquitination pathways for their own use, illustrating deep convergent functional mimicry.

Post-translational protein ubiquitination for signaling and protein-turnover is a eukaryotic innovation that is not employed by bacteria [77]. Despite this, two novel families of bacterial effector ligases, the HECT-like and NEL families of E3 Lys48-specific ligases from *Salmonella's* SopA and *Shigella's* IpaH9.8 proteins, can polyubiquitinate host inflammatory proteins. Novelty often makes these ubiquitin linkages irreversible because host proteins cannot cleave them. Ubiquitin cleavage is also leveraged by bacterial functional mimics to block immune signaling. For example, *Legionella* uses its effector transglutaminase MvcA to catalyze the non-canonical deubiquitination of the E2 ubiquitin-conjugating enzyme UBE2N, blocking its activity, preventing it from ubiquitinating bacterial proteins [78]. Given that these bacterial-evolved strategies of modifying ubiquitin signaling are entirely

novel and interact with elements of eukaryotic ubiquitination pathways that are still poorly understood, such as through Lys6 polyubiquitin linkages, study of pathogenic mechanisms will reveal endogenous mechanisms [61].

4. Detecting molecular mimicry in genetic conflict computationally

Detecting evidence of genetic conflict and molecular mimicry computationally requires robust coevolutionary theory, ample empirical data, and well-developed interaction models (Fig. 5, Supplemental Figures S1 and S2). While molecular mimics have been sought through sequence-based data mining for at least 30 years [11,61], advancements in protein structural prediction models recently enabled screens for mimics at the structure level [8]. *In silico* approaches for screening for genetic conflict and molecular mimicry are important because they expand our knowledge of the distribution and frequency of elements among taxa and the likelihood for mimicry to evolve. Indeed, expanded knowledge of protein domain distributions among bacteria corrected the assumption that bacteria lack protein tyrosine phosphatase domains; now it is well understood that they arose in bacteria, and not by horizontal gene transfer [61].
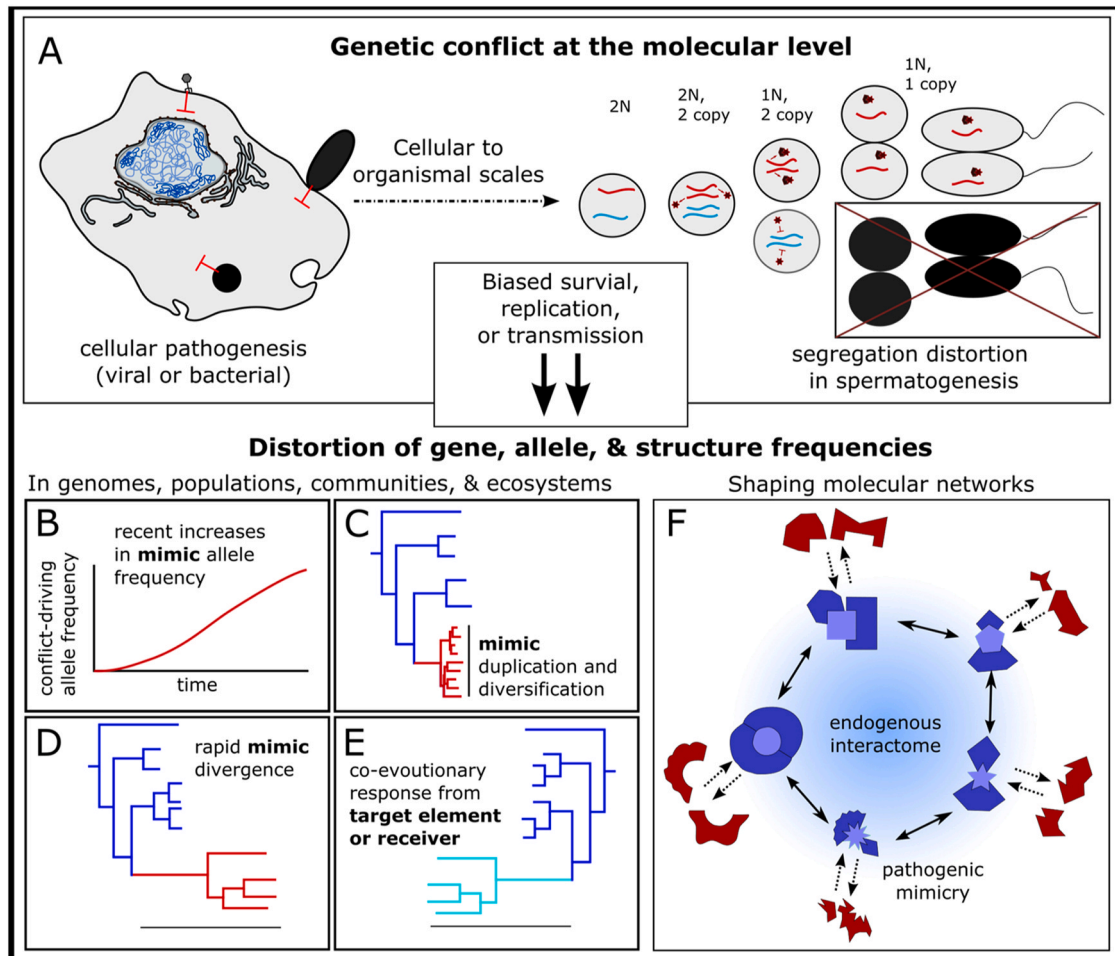

Fig. 5. Computational approaches for detecting molecular mimicry. A) Genetic conflicts at the cellular and organismal levels can be detected through B-E) population genomic data and evolutionary inference because of the influence biased survival, replication, and transmission have on allele frequencies. Mimic elements in conflict are depicted in red and target elements of mimicry are depicted in blue. B) Non-Mendelian inheritance can be detected through allele frequency measurements. Mimic phylogenetic patterns can reveal C) duplication and diversification and D) divergence from their non-mimic relatives. Through co-evolutionary analysis, arms races can be detected between mimic divergence (in D) and E) target divergence from untargeted relatives. F) Protein structure and interaction predictions from genomic data can be leveraged to detect structural arms races between targets (blue) and mimics (red). See Supplemental Figs. S1 and S2 for summaries of molecular mimic evolutionary sources and screening guidelines, respectively.

4.1. Detecting sequence-level mimics

Molecular mimic identification at the sequence level is a powerful approach for identifying conflicting genetic elements that may encode the ability to co-opt the cell biology of their target [31,32]. Indeed, these approaches have identified tapeworm candidate mimics of stickleback proteins [79], as well as malaria, filarial nematode, and *Legionella* mimic candidates of human proteins [12,80]. Sequence-level mimics of target elements can be detected at the gene, domain, or motif level, either among protein, RNA, or DNA sequences, depending on the mimic type, degree of relatedness, and functional sequence requirement. Therefore, motif, domain, and BLAST searches of candidate sequence-level mimics should be performed as a first-pass search in conflict-driven molecular mimic screens. However, there is a lower limit on the length of sequence mimicry that can be detected bioinformatically. For example, PTM-triggering sequences are likely too small and heterogeneous outside of the modified residue to detect bioinformatically [26].

Databases for short motifs, repeats, and domains exist, which can be leveraged to look for shared sequences between conflicting elements and their targets. The wealth of genomic resources made available by next-generation genome sequencing and decades of functional assays enabled researchers to inventory the molecular interaction modules that manifest organismal phenotypes in the Eukaryotic Linear Motif (ELM) resource [31]. Using the ELM prediction tool, intergenomic mimics of host protein SLiMs can be predicted in conflicting genomes [81]. Mimic proteins that travel through nuclear pore complexes require nuclear localization signals (NLSs) [82], which exhibit a variety of forms. NLSs can be predicted from amino acid sequences by hidden markov models or machine learning approaches [83]. In contrast to NLSs, nuclear export signals (NESs) only share leucine-rich regions, and may be better predicted structurally.

Genomic surveys for large structural changes, such as duplicated genes, tandem repeats, mobile element insertions, and novel chromosomes or plasmids, can be employed to detect selfish elements that may use molecular mimicry to achieve a transmission advantage. For example, gene duplications are one of the hallmarks of genetic conflict in meiotic drive systems [33], and can be easily detected by deviations in reference re-sequencing depth [34]. Toxin-antitoxin genes often evolve on plasmids and spread via transposable elements and viruses to chromosomes, making their sequence-based detection relatively simple through alignment to toxin and mobile element databases and *ab initio* repeat finders [84]. As antidote-mimicking resistance loci spread by horizontal gene transfer and gene conversion, genomes should also be screened for partial TA elements. Accessory chromosomal sequences, such as B chromosomes, that may encode mimic elements can be identified by their variable presence among tissues and their divergence relative to the endogenous chromosomes [4].

4.2. Detecting structure-level mimics

The advent of accurate protein structural prediction models such as AlphaFold [85,86] enabled whole proteome screens for structural mimic candidates, *ab initio*. Protein structural predictions and structural alignment algorithms have been in development for decades, and have been instrumental in identifying similar protein structures between pathogens and hosts [15]. Structural mimic candidates are recognizable through 1) alignment between mimic and target protein structures [87], 2) interactions between the mimic and its interacting partners [16], and 3) competitive interactions between the mimic and targets for the interacting partners (*e.g.,* [17,59]). Through combining these well-developed conceptual frameworks with recent innovations in structural prediction, more structural mimics can be identified.

Regional structural mimic prediction was useful for understanding the evolution of host tissue infection affinities, *i.e.,* tropisms, in the rapidly evolving viral SARS-CoV2 pandemic. In addition to the integrin-binding fibronectin-like RGD motif, the coronavirus spike glycoprotein also contains a receptor binding motif (RBM) that modulates viral cellular tropism. Structural prediction and alignment allowed SARS-CoV2 researchers to identify host protein structures that are similar to viral spike protein RBMs, resulting in highly specific epidermal growth factor (EGF)-like hits that bind similarly to the factors they mimic in docking simulations [58]. Many other regional structures could be mined for in viral and bacterial genomic data through receptor-informed reverse searches for mimicked ligand binding surfaces, and other specific interaction predictions.

While structural protein predictions are now quite reliable for many secondary and tertiary structures, intrinsically disordered domains [88] and genome-wide protein-protein docking interactions are still difficult to predict [16]. Disordered proteins may mediate imperfect mimicry mechanisms through their conformational flexibility and SLiMs enrichment [26]. So, this subset of the proteome should be further mined for motif-based mimics, such as histone mimics [45]. Signatures of molecular mimicry have been detected in aggregate across conflicting genomes through interaction predictions [8], suggesting that at least some mimicked interactions can be detected *ab initio*.

4.3. Detecting function-level mimics

Functional mimicry is far more difficult to detect from individual genomic datasets than sequence or structural mimics because predictions about the nature of mimicry need to be identified experimentally before they can be linked to their genetic determinants. Fortunately, genetic conflict leaves marks on populations that can be leveraged to detect unknown elements and mechanisms. The short and long-term consequences of selfish element proliferation and transmission can be seen through phenomena such as repetitive element expansions, skews in allele frequencies (Fig. 5B), changes in chromosome number, and alterations in sex ratio [3]. Correlated coevolutionary patterns, such as accelerated evolution or constraint, between genes encoded in interacting genomes can also be leveraged to detect signals of genetic conflict and mimicry (Fig. 5C-E).

Through leveraging genome-wide signals of biased genetic transmission, recent studies have detected evidence of genetic conflicts that were unresolvable with smaller datasets. Through a scan across thousands of publically available eukaryotic genomes, transposable element expansions were shown to drive intron gain in diverse eukaryotes, from basal protists to copepods [89]. By leveraging bulk gamete genome sequencing and comparing the parental allele frequencies to those of somatic tissue, researchers have been able to detect skews in parental allele frequency transmission in *Arabidopsis* hybrid plants that are due to unknown segregation distortion mechanisms [90]. Using a similar approach, *Zea mays* ssp. *mexicana* was recently shown to encode a male-specific segregation distorter that functions through RNAi-mediated targeting of an essential pollen lipase [91]. While these results do not resolve mechanistic details about the nature of the genetic elements in conflict or mimicry, they do provide candidate loci for further experimental investigation.

Advances in multi-omics approaches enable the identification of specific interacting factors between conflicting genomes in medium to high-throughput. If conflicting genomes are known *a priori*, then patterns of correlated coevolution can be investigated through phylogenetic branch length tests (Fig. 5D,E). Tests such as this have been used to detect molecular mimicry in sexual conflicts between abalone egg coat proteins and their sperm lysin binding partners, which trigger vitelline envelope breakdown and fertilization [92]. Targeted assays for post-translational modification mimics, such as lectin microarrays to assay glycan binding, can identify structural mimics of non-protein factors [93]. If epitopes and binding surfaces are known, targeted structural searches and *in silico* docking assays can be run on the conflicting genome to identify candidate binding partners among the expressed ligands, receptors, and other proteins (Fig. 5F).

5. Conclusion

Broad knowledge about the distribution and occurrence of molecular mimics across known intragenomic and intergenomic genetic conflicts informs on the likelihood of these strategies evolving in newly discovered interactions. Recent advances in database mining, evolutionary inference, and protein structure prediction offer new strategies for identifying molecular mimics in genetic conflict computationally. Here, we linked empirical data from these limited, yet growing, datasets with theoretical predictions about the likelihood of convergent evolution to show: In intergenomic conflicts deploying molecular mimicry, convergence drives the evolution of short genetic elements for sequence-based mimicry. Longer intergenomic sequence mimics are generally obtained through HGT. Structural and functional mimics exhibit more variable trends that are likely attributed to the duration and strength of selection pressures on the coevolved function (*e.g.*, short ligand-binding sequence mimics versus long ligand structural mimics). Importantly, functional and structural convergence in intergenomic conflicts may be ancient. For example, deeply rooted bacterial protein tyrosine phosphatases and ubiquitin ligases coevolved to mimic eukaryotic processes [60,61,78]. These trends do not hold for intragenomic conflicts that leverage molecular mimicry because the close relationship between driver and target enables sequence duplication events to drive the rapid evolution of any length mimic.

While an impressive number of molecular mimics are known to play a role in genetic conflicts, our knowledge about their taxonomic distribution is still limited by the availability of genomic data and tractable systems for functional testing. With machine-learning approaches improving our ability to predict patterns from observed data, our capacity to detect mimicry at the DNA, RNA, protein, and post-translational levels is likely to improve rapidly. Once we have a better sense of the global distribution of molecular mimics, these phenomena can be explored for emergent properties suggestive of common selective regimes and adaptive potential. For example, mimicry rings, in which divergent lineages converge on a similar mimicry pattern due to similar selective pressures [5], may exist for molecules and shape taxon-specific pressures on host sequence and structural variation.

Declaration of Competing Interest

None

Author contributions

SLR: study conception and design, literature analysis, figure illustration, and writing. GP: literature analysis and writing. CC: literature analysis and writing

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.semcdb.2024.07.001.

References

[1] J.H. Werren, Selfish genetic elements, genetic conflict, and evolutionary innovation, Proc. Natl. Acad. Sci. 108 (2011) 10863–10870, https://doi.org/10.1073/pnas.1102343108.

[2] S.E. Massey, B. Mishra, Origin of biomolecular games: deception and molecular evolution, J. R. Soc. Interface 15 (2018) 20180429, https://doi.org/10.1098/rsif.2018.0429.

[3] R.N. McLaughlin, H.S. Malik, Genetic conflicts: the usual suspects and beyond, J. Exp. Biol. 220 (2017) 6–17, https://doi.org/10.1242/jeb.148148.

[4] J.A. Ågren, A.G. Clark, Selfish genetic elements, PLOS Genet. 14 (2018) e1007700, https://doi.org/10.1371/journal.pgen.1007700.

[5] N.C. Elde, H.S. Malik, The evolutionary conundrum of pathogen mimicry, Nat. Rev. Microbiol. 7 (2009) 787–797, https://doi.org/10.1038/nrmicro2222.

[6] S. Qin, Y. Liu, W. Tempel, M.S. Eram, C. Bian, K. Liu, G. Senisterra, L. Crombet, M. Vedadi, J. Min, Structural basis for histone mimicry and hijacking of host proteins by influenza virus protein NS1, Nat. Commun. 5 (2014) 3952, https://doi.org/10.1038/ncomms4952.

[7] L. Aravind, L.M. Iyer, A.M. Burroughs, Discovering biological conflict systems through genome analysis: evolutionary principles and biochemical novelty, Annu. Rev. Biomed. Data Sci. 5 (2022) 367–391, https://doi.org/10.1146/annurev-biodatasci-122220-101119.

[8] N.S. de Groot, M. Torrent Burgas, Bacteria use structural imperfect mimicry to hijack the host interactome, PLoS Comput. Biol. 16 (2020) e1008395, https://doi.org/10.1371/journal.pcbi.1008395.

[9] Ø.H. Holen, R.A. Johnstone, The evolution of mimicry under constraints, Am. Nat. 164 (2004) 598–613.

[10] A. Alcami, Viral mimicry of cytokines, chemokines and their receptors, Nat. Rev. Immunol. 3 (2003) 36–50, https://doi.org/10.1038/nri980.

[11] C.E. Stebbins, J.E. Galán, Structural mimicry in bacterial virulence, Nature 412 (2001) 701–705, https://doi.org/10.1038/35089000.

[12] D. Burstein, T. Zusman, E. Degtyar, R. Viner, G. Segal, T. Pupko, Genome-scale identification of legionella pneumophila effectors using a machine learning approach, PLOS Pathog. 5 (2009) e1000508, https://doi.org/10.1371/journal.ppat.1000508.

[13] S. Xu, C. Zhang, Y. Miao, J. Gao, D. Xu, Effector prediction in host-pathogen interaction based on a Markov model of a ubiquitous EPIYA motif, BMC Genom. 11 (2010) 1–15, https://doi.org/10.1186/1471-2164-11-S3-S1.

[14] A.C. Doxey, B.J. McConkey, Prediction of molecular mimicry candidates in human pathogenic bacteria, Virulence 4 (2013) 453–466, https://doi.org/10.4161/viru.25180.

[15] N. Drayman, Y. Glick, O. Ben-nun-Shaul, H. Zer, A. Zlotnick, D. Gerber, O. Schueler-Furman, A. Oppenheim, Pathogens use structural mimicry of native host ligands as a mechanism for host receptor engagement, Cell Host Microbe 14 (2013) 63–73, https://doi.org/10.1016/j.chom.2013.05.005.

[16] P. Bryant, G. Pozzati, A. Elofsson, Improved prediction of protein-protein interactions using AlphaFold2, Nat. Commun. 13 (2022) 1265, https://doi.org/10.1038/s41467-022-28865-w.

[17] V. Muthye, J.D. Wasmuth, Proteome-wide comparison of tertiary protein structures reveals molecular mimicry in Plasmodium-human interactions, Front. Parasitol. 2 (2023) https://www.frontiersin.org/articles/10.3389/fpara.2023.1162697 (accessed October 21, 2023).

[18] S. Mondino, S. Schmidt, C. Buchrieser, Molecular mimicry: a paradigm of host-microbe coevolution illustrated by legionella, e01201-20, mBio 11 (2020), https://doi.org/10.1128/mBio.01201-20.

[19] N.E. Davey, G. Travé, T.J. Gibson, How viruses hijack cell regulation, Trends Biochem. Sci. 36 (2011) 159–169, https://doi.org/10.1016/j.tibs.2010.10.002.

[20] A.C. Frank, Molecular host mimicry and manipulation in bacterial symbionts, FEMS Microbiol. Lett. 366 (2019) fnz038, https://doi.org/10.1093/femsle/fnz038.

[21] J. Ma, A. Ratan, B.J. Raney, B.B. Suh, W. Miller, D. Haussler, The infinite sites model of genome evolution, Proc. Natl. Acad. Sci. 105 (2008) 14254–14261.

[22] A. Harpak, A. Bhaskar, J.K. Pritchard, Mutation rate variation is a primary determinant of the distribution of allele frequencies in humans, PLOS Genet 12 (2016) e1006489, https://doi.org/10.1371/journal.pgen.1006489.

[23] D.L. Stern, The genetic causes of convergent evolution, Nat. Rev. Genet. 14 (2013) 751–764, https://doi.org/10.1038/nrg3483.

[24] P.V. Markov, M. Ghafari, M. Beer, K. Lythgoe, P. Simmonds, N.I. Stilianakis, A. Katzourakis, The evolution of SARS-CoV-2, Nat. Rev. Microbiol. 21 (2023) 361–379, https://doi.org/10.1038/s41579-023-00878-2.

[25] S.G. Giulieri, R. Guérillot, S. Duchene, A. Hachani, D. Daniel, T. Seemann, J.S. Davis, S.Y. Tong, B.C. Young, D.J. Wilson, T.P. Stinear, B.P. Howden, Niche-specific genome degradation and convergent evolution shaping Staphylococcus aureus adaptation during severe infections, eLife 11 (2022) e77195, https://doi.org/10.7554/eLife.77195.

[26] P. Tompa, N.E. Davey, T.J. Gibson, M.M. Babu, A million peptide motifs for the molecular biologist, Mol. Cell 55 (2014) 161–169, https://doi.org/10.1016/j.molcel.2014.05.032.

[27] K. Kajala, M. Gouran, L. Shaar-Moshe, G.A. Mason, J. Rodriguez-Medina, D. Kawa, G. Pauluzzi, M. Reynoso, A. Canto-Pastor, C. Manzano, V. Lau, M.A.S. Artur, D.A. West, S.B. Gray, A.T. Borowsky, B.P. Moore, A.I. Yao, K.W. Morimoto, M. Bajic, E. Formentin, N.A. Nirmal, A. Rodriguez, A. Pasha, R.B. Deal, D.J. Kliebenstein, T.R. Hvidsten, N.J. Provart, N.R. Sinha, D.E. Runcie, J. Bailey-Serres, S.M. Brady, Innovation, conservation, and repurposing of gene function in root cell type development, 3348.e19, Cell 184 (2021) 3333, https://doi.org/10.1016/j.cell.2021.04.024.

[28] K.R. Nitta, A. Jolma, Y. Yin, E. Morgunova, T. Kivioja, J. Akhtar, K. Hens, J. Toivonen, B. Deplancke, E.E.M. Furlong, J. Taipale, Conservation of transcription factor binding specificities across 600 million years of bilateria evolution, eLife 4 (2015) e04837, https://doi.org/10.7554/eLife.04837.

[29] S. Li, H.G. Dohlman, Evolutionary conservation of sequence motifs at sites of protein modification, J. Biol. Chem. 299 (2023), https://doi.org/10.1016/j.jbc.2023.104617.

[30] J.E. Martyn, L. Gomez-Valero, C. Buchrieser, The evolution and role of eukaryotic-like domains in environmental intracellular bacteria: the battle with a eukaryotic

cell, FEMS Microbiol. Rev. 46 (2022) fuac012, https://doi.org/10.1093/femsre/fuac012.

[31] M. Kumar, S. Michael, J. Alvarado-Valverde, B. Mészáros, H. Sámano-Sánchez, A. Zeke, L. Dobson, T. Lazar, M. Örd, A. Nagpal, N. Farahi, M. Käser, R. Kraleti, N. E. Davey, R. Pancsa, L.B. Chemes, T.J. Gibson, The eukaryotic linear motif resource: 2022 release, Nucleic Acids Res. 50 (2022) D497–D508, https://doi.org/10.1093/nar/gkab975.

[32] H. Sámano-Sánchez, T.J. Gibson, Mimicry of short linear motifs by bacterial pathogens: a drugging opportunity, Trends Biochem. Sci. 45 (2020) 526–544, https://doi.org/10.1016/j.tibs.2020.03.003.

[33] C. Courret, C.-H. Chang, K.H.-C. Wei, C. Montchamp-Moreau, A.M. Larracuente, Meiotic drive mechanisms: lessons from *Drosophila*, Proc. R. Soc. B Biol. Sci. 286 (2019) 20191430, https://doi.org/10.1098/rspb.2019.1430.

[34] A. Levasseur, P. Pontarotti, The role of duplications in the evolution of genomes highlights the need for evolutionary-based approaches in comparative genomics, Biol. Direct 6 (1) (2011) 12, https://doi.org/10.1186/1745-6150-6-11.

[35] A.A. Vogan, S.L. Ament-Velásquez, A. Granger-Farbos, J. Svedberg, E. Bastiaans, A. J. Debets, V. Coustou, H. Yvanne, C. Clavé, S.J. Saupe, H. Johannesson, Combinations of Spok genes create multiple meiotic drivers in Podospora, eLife 8 (2019) e46454, https://doi.org/10.7554/eLife.46454.

[36] C.M. Botham, A.M. Wandler, K. Guillemin, A transgenic drosophila model demonstrates that the helicobacter pylori caga protein functions as a eukaryotic gab adaptor, PLoS Pathog. 4 (2008) e1000064, https://doi.org/10.1371/journal.ppat.1000064.

[37] Y. Zhang, J. Li, B. Li, J. Wang, C.H. Liu, Mycobacterium tuberculosis Mce3C promotes mycobacteria entry into macrophages through activation of β2 integrin-mediated signalling pathway, Cell. Microbiol. 20 (2018) e12800, https://doi.org/10.1111/cmi.12800.

[38] J. Kee, S. Thudium, D.M. Renner, K. Glastad, K. Palozola, Z. Zhang, Y. Li, Y. Lan, J. Cesare, A. Poleshko, A.A. Kiseleva, R. Truitt, F.L. Cardenas-Diaz, X. Zhang, X. Xie, D.N. Kotton, K.D. Alysandratos, J.A. Epstein, P.-Y. Shi, W. Yang, E. Morrisey, B.A. Garcia, S.L. Berger, S.R. Weiss, E. Korb, SARS-CoV-2 disrupts host epigenetic regulation via histone mimicry, Nature 610 (2022) 381–388, https://doi.org/10.1038/s41586-022-05282-z.

[39] N.E. Davey, M.S. Cyert, A.M. Moses, Short linear motifs – ex nihilo evolution of protein regulation, Cell Commun. Signal. 13 (2015) 43, https://doi.org/10.1186/s12964-015-0120-z.

[40] H. Hamidi, J. Ivaska, Every step of the way: integrins in cancer progression and metastasis, Nat. Rev. Cancer 18 (2018) 533–548, https://doi.org/10.1038/s41568-018-0038-z.

[41] J. Conradi, N. Tegtmeyer, M. Woźna, M. Wissbrock, C. Michalek, C. Gagell, T. Cover, R. Frank, N. Sewald, S. Backert, An RGD helper sequence in CagL of helicobacter pylori assists in interactions with integrins and injection of CagA, Front. Cell. Infect. Microbiol. 2 (2012) https://www.frontiersin.org/articles/10.3389/fcimb.2012.00070 (accessed October 18, 2023).

[42] T. Hayashi, H. Morohashi, M. Hatakeyama, Bacterial EPIYA effectors - Where do they come from? What are they? Where are they going?: bacterial EPIYA effector family, Cell. Microbiol. 15 (2013) 377–385, https://doi.org/10.1111/cmi.12040.

[43] E.G. Norris, X.S. Pan, D.C. Hocking, Receptor-binding domain of SARS-CoV-2 is a functional αv-integrin agonist, J. Biol. Chem. 299 (2023), https://doi.org/10.1016/j.jbc.2023.102922.

[44] P. Simons, D.A. Rinaldi, V. Bondu, A.M. Kell, S. Bradfute, D.S. Lidke, T. Buranda, Integrin activation is an essential component of SARS-CoV-2 infection, Sci. Rep. 11 (2021) 20398, https://doi.org/10.1038/s41598-021-99893-7.

[45] U. Schaefer, J.S.Y. Ho, R.K. Prinjha, A. Tarakhovsky, The "Histone Mimicry" by pathogens, Cold Spring Harb. Symp. Quant. Biol. 78 (2013) 81–90, https://doi.org/10.1101/sqb.2013.78.020339.

[46] I. Marazzi, J.S.Y. Ho, J. Kim, B. Manicassamy, S. Dewell, R.A. Albrecht, C. W. Seibert, U. Schaefer, K.L. Jeffrey, R.K. Prinjha, K. Lee, A. García-Sastre, R. G. Roeder, A. Tarakhovsky, Suppression of the antiviral response by an influenza histone mimic, Nature 483 (2012) 428–433, https://doi.org/10.1038/nature10892.

[47] J. Shi, Y. Wang, L. Zeng, Y. Wu, J. Deng, Q. Zhang, Y. Lin, J. Li, T. Kang, M. Tao, E. Rusinova, G. Zhang, C. Wang, H. Zhu, J. Yao, Y.-X. Zeng, B.M. Evers, M.-M. Zhou, B.P. Zhou, Disrupting the interaction of BRD4 with diacetylated twist suppresses tumorigenesis in basal-like breast cancer, Cancer Cell 25 (2014) 210–225, https://doi.org/10.1016/j.ccr.2014.01.028.

[48] L.F. Gingell, J.R. McLean, A protamine knockdown mimics the function of *Sd* in *Drosophila melanogaster*, G3 Genes. 10 (2020) 2111–2115, https://doi.org/10.1534/g3.120.401307.

[49] M. Herbette, X. Wei, C.-H. Chang, A.M. Larracuente, B. Loppin, R. Dubruille, Distinct spermiogenic phenotypes underlie sperm elimination in the Segregation Distorter meiotic drive system, PLOS Genet. 17 (2021) e1009662, https://doi.org/10.1371/journal.pgen.1009662.

[50] C.D. Malone, R. Lehmann, F.K. Teixeira, The cellular basis of hybrid dysgenesis and Stellate regulation in Drosophila, Curr. Opin. Genet. Dev. 34 (2015) 88–94, https://doi.org/10.1016/j.gde.2015.09.003.

[51] A. Burga, E. Ben-David, L. Kruglyak, Toxin-antidote elements across the tree of life, Annu. Rev. Genet. 54 (2020) 387–415, https://doi.org/10.1146/annurev-genet-112618-043659.

[52] J.E. Silpe, B.L. Bassler, A host-produced quorum-sensing autoinducer controls a phage lysis-lysogeny decision, 280.e13, Cell 176 (2019) 268, https://doi.org/10.1016/j.cell.2018.10.059.

[53] D. Jurènas, N. Fraikin, F. Goormaghtigh, L. Van Melderen, Biology and evolution of bacterial toxin–antitoxin systems, Nat. Rev. Microbiol. 20 (2022) 335–350, https://doi.org/10.1038/s41579-021-00661-1.

[54] T.R. Blower, T.J. Evans, R. Przybilski, P.C. Fineran, G.P.C. Salmond, Viral evasion of a bacterial suicide system by RNA–based molecular mimicry enables infectious altruism, PLoS Genet. 8 (2012) e1003023, https://doi.org/10.1371/journal.pgen.1003023.

[55] M.A. Bravo Núñez, J.J. Lange, S.E. Zanders, A suppressor of a wtf poison-antidote meiotic driver acts via mimicry of the driver's antidote, PLOS Genet. 14 (2018) e1007836, https://doi.org/10.1371/journal.pgen.1007836.

[56] J. English, S. Patrick, L.D. Stewart, The potential role of molecular mimicry by the anaerobic microbiota in the aetiology of autoimmune disease, Anaerobe 80 (2023) 102721, https://doi.org/10.1016/j.anaerobe.2023.102721.

[57] N.S. de Groot, M.T. Burgas, Bacteria use structural imperfect mimicry to hijack the host interactome, PLOS Comput. Biol. 16 (2020) e1008395, https://doi.org/10.1371/journal.pcbi.1008395.

[58] C.A. Beaudoin, A.R. Jamasb, A.F. Alsulami, L. Copoiu, A.J. van Tonder, S. Hala, B. P. Bannerman, S.E. Thomas, S.C. Vedithi, P.H.M. Torres, T.L. Blundell, Predicted structural mimicry of spike receptor-binding motifs from highly pathogenic human coronaviruses, Comput. Struct. Biotechnol. J. 19 (2021) 3938–3953, https://doi.org/10.1016/j.csbj.2021.06.041.

[59] G. Lasso, B. Honig, S.D. Shapira, A sweep of earth's virome reveals host-guided viral protein structural mimicry and points to determinants of human disease, Cell Syst. 12 (2021) 82–91.e3, https://doi.org/10.1016/j.cels.2020.09.006.

[60] F. Böhmer, S. Szedlacsek, L. Tabernero, A. Östman, J. den Hertog, Protein tyrosine phosphatase structure–function relationships in regulation and pathogenesis, FEBS J. 280 (2013) 413–431, https://doi.org/10.1111/j.1742-4658.2012.08655.x.

[61] D. Salomon, K. Orth, What pathogens have taught us about posttranslational modifications, Cell Host Microbe 14 (2013) 269–279, https://doi.org/10.1016/j.chom.2013.07.008.

[62] C.A. Beaudoin, A.R. Jamasb, A.F. Alsulami, L. Copoiu, A.J. van Tonder, S. Hala, B. P. Bannerman, S.E. Thomas, S.C. Vedithi, P.H.M. Torres, T.L. Blundell, Predicted structural mimicry of spike receptor-binding motifs from highly pathogenic human coronaviruses, Comput. Struct. Biotechnol. J. 19 (2021) 3938–3953, https://doi.org/10.1016/j.csbj.2021.06.041.

[63] C.A. Custódio, J.F. Mano, Cell surface engineering to control cellular interactions, ChemNanoMat 2 (2016) 376–384, https://doi.org/10.1002/cnma.201600047.

[64] A. Varki, R. Kannagi, B. Toole, P. Stanley, Glycosylation Changes in Cancer, in: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger (Eds.), Essent. Glycobiol., 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015 http://www.ncbi.nlm.nih.gov/books/NBK453023/ (accessed October 31, 2023).

[65] H. de Jong, M.M.S.M. Wösten, T. Wennekes, Sweet impersonators: molecular mimicry of host glycans by bacteria, Glycobiology 32 (2022) 11–22, https://doi.org/10.1093/glycob/cwab104.

[66] N.L. Miller, T. Clark, R. Raman, R. Sasisekharan, Glycans in virus-host interactions: a structural perspective, Front. Mol. Biosci. 8 (2021). ⟨https://www.frontiersin.org/articles/10.3389/fmolb.2021.666756⟩ (accessed October 31, 2023).

[67] A. Amara, J. Mercer, Viral apoptotic mimicry, Nat. Rev. Microbiol. 13 (2015) 461–469, https://doi.org/10.1038/nrmicro3469.

[68] X. Shi, P. Halder, H. Yavuz, R. Jahn, H.A. Shuman, Direct targeting of membrane fusion by SNARE mimicry: convergent evolution of Legionella effectors, Proc. Natl. Acad. Sci. 113 (2016) 8807–8812, https://doi.org/10.1073/pnas.1608755113.

[69] D. Seo, D.B. Gammon, Manipulation oF Host Microtubule Networks by Viral Microtubule-associated Proteins, Viruses 14 (2022) 979, https://doi.org/10.3390/v14050979.

[70] R.L. Lamason, M.D. Welch, Actin-based motility and cell-to-cell spread of bacterial pathogens, Curr. Opin. Microbiol. 35 (2017) 48–57, https://doi.org/10.1016/j.mib.2016.11.007.

[71] D. Yesudhas, M. Batool, M.A. Anwar, S. Panneerselvam, S. Choi, Proteins recognizing DNA: structural uniqueness and versatility of DNA-binding domains in stem cell transcription factors, Genes 8 (2017) 192, https://doi.org/10.3390/genes8080192.

[72] H.-C. Wang, C.-H. Ho, K.-C. Hsu, J.-M. Yang, A.H.-J. Wang, DNA Mimic proteins: functions, structures, and bioinformatic analysis, Biochemistry 53 (2014) 2865–2874, https://doi.org/10.1021/bi5002689.

[73] A. Hartsock, W.J. Nelson, Adherens and tight junctions: Structure, function and connections to the actin cytoskeleton, Biochim. Biophys. Acta BBA - Biomembr. 1778 (2008) 660–669, https://doi.org/10.1016/j.bbamem.2007.07.012.

[74] T. Shibata-Seki, M. Nagaoka, M. Goto, E. Kobatake, T. Akaike, Direct visualization of the extracellular binding structure of E-cadherins in liquid, Sci. Rep. 10 (2020) 17044, https://doi.org/10.1038/s41598-020-72517-2.

[75] M. Bonazzi, M. Lecuit, P. Cossart, Listeria monocytogenes internalin and E-cadherin: from structure to pathogenesis, Cell. Microbiol. 11 (2009) 693–702, https://doi.org/10.1111/j.1462-5822.2009.01293.x.

[76] M. Ragon, T. Wirth, F. Hollandt, R. Lavenir, M. Lecuit, A.L. Monnier, S. Brisse, A new perspective on listeria monocytogenes evolution, PLOS Pathog. 4 (2008) e1000146, https://doi.org/10.1371/journal.ppat.1000146.

[77] X. Grau-Bové, A. Sebé-Pedrós, I. Ruiz-Trillo, The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin, Mol. Biol. Evol. 32 (2015) 726–739, https://doi.org/10.1093/molbev/msu334.

[78] T.G. Franklin, J.N. Pruneda, Bacteria make surgical strikes on host ubiquitin signaling, PLOS Pathog. 17 (2021) e1009341, https://doi.org/10.1371/journal.ppat.1009341.

[79] F.O. Hebert, L. Phelps, I. Samonte, M. Panchal, S. Grambauer, I. Barber, M. Kalbe, C.R. Landry, N. Aubin-Horth, Identification of candidate mimicry proteins involved in parasite-driven phenotypic changes, Parasit. Vectors 8 (2015) 225, https://doi.org/10.1186/s13071-015-0834-1.

[80] P. Ludin, D. Nilsson, P. Mäser, Genome-wide identification of molecular mimicry candidates in parasites, PLOS ONE 6 (2011) e17546, https://doi.org/10.1371/journal.pone.0017546.

[81] P. Hraber, P.E. O'Maille, A. Silberfarb, K. Davis-Anderson, N. Generous, B. H. McMahon, J.M. Fair, Resources to discover and use short linear motifs in viral proteins, Trends Biotechnol. 38 (2020) 113–127, https://doi.org/10.1016/j.tibtech.2019.07.004.

[82] J. Lu, T. Wu, B. Zhang, S. Liu, W. Song, J. Qiao, H. Ruan, Types of nuclear localization signals and mechanisms of protein import into the nucleus, Cell Commun. Signal. 19 (2021) 60, https://doi.org/10.1186/s12964-021-00741-y.

[83] M. Bernhofer, T. Goldberg, S. Wolf, M. Ahmed, J. Zaugg, M. Boden, B. Rost, NLSdb-major update for database of nuclear localization signals and nuclear export signals, Nucleic Acids Res. 46 (2018) D503–D508, https://doi.org/10.1093/nar/gkx1021.

[84] C. Goubert, R.J. Craig, A.F. Bilat, V. Peona, A.A. Vogan, A.V. Protasio, A beginner's guide to manual curation of transposable elements, Mob. DNA 13 (2022) 7, https://doi.org/10.1186/s13100-021-00259-7.

[85] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, Nucleic Acids Res. 50 (2022) D439–D444, https://doi.org/10.1093/nar/gkab1061.

[86] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.

[87] M. van Kempen, S.S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C.L.M. Gilchrist, J. Söding, M. Steinegger, Fast and accurate protein structure search with Foldseek, Nat. Biotechnol. (2023) 1–4, https://doi.org/10.1038/s41587-023-01773-0.

[88] C.J. Wilson, W.-Y. Choy, M. Karttunen, AlphaFold2: a role for disordered protein/region prediction? Int. J. Mol. Sci. 23 (2022) 4591, https://doi.org/10.3390/ijms23094591.

[89] L. Gozashti, S.W. Roy, B. Thornlow, A. Kramer, M. Ares, R. Corbett-Detig, Transposable elements drive intron gain in diverse eukaryotes, Proc. Natl. Acad. Sci. 119 (2022) e2209766119, https://doi.org/10.1073/pnas.2209766119.

[90] R. Corbett-Detig, P. Medina, H. Frérot, C. Blassiau, V. Castric, Bulk pollen sequencing reveals rapid evolution of segregation distortion in the male germline of Arabidopsis hybrids, Evol. Lett. 3 (2019) 93–103, https://doi.org/10.1002/evl3.96.

[91] B. Berube, E. Ernst, J. Cahn, B. Roche, C. de, S. Alves, J. Lynn, A. Scheben, A. Siepel, J. Ross-Ibarra, J. Kermicle, R. Martienssen, Teosinte Pollen Drive Guides maize Domest. Evol. Rna (2023), https://doi.org/10.1101/2023.07.12.548689, 2023.07.12.548689.

[92] J.E. Aagaard, S.A. Springer, S.D. Soelberg, W.J. Swanson, Duplicate abalone egg coat proteins bind sperm lysin similarly, but evolve oppositely, consistent with molecular mimicry at fertilization, PLOS Genet 9 (2013) e1003287, https://doi.org/10.1371/journal.pgen.1003287.

[93] S. Chen, R. Qin, L.K. Mahal, Sweet systems: technologies for glycomic analysis and their integration into systems biology, Crit. Rev. Biochem. Mol. Biol. 56 (2021) 301–320, https://doi.org/10.1080/10409238.2021.1908953.