

Review

Not peer-reviewed version

Potential Applications of Computerised Algorithms and Implementation of AI Models into Microbiological, Evolutionary and Pharmaceutical Research

[Theodor-Nicolae Carp](#)*

Posted Date: 23 December 2024

doi: 10.20944/preprints202412.1882.v1

Keywords: AI; bioinformatics; microbiology; evolutionary biology; immunology; pharmacology; viral mutation; immune adaptation; research catalysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Review

Potential Applications of Computerised Algorithms and Implementation of AI Models into Microbiological, Evolutionary and Pharmaceutical Research

Theodor-Nicolae Carp

Independent Academic Researcher, Address: Watford, Hertfordshire, England, United Kingdom of Great Britain; theodore.nicholas100@gmail.com

Abstract: Artificial intelligence has sharply and broadly evolved since the invention of the computer and its eventual widespread distribution throughout the world. It has led to the replacement of traditional, long-term clinical practices to computerised procedures bringing rapid results. Ultimately, this led to the development of bioinformatics, helping geneticists sequence whole genomes and proteomes within a short time frame. Subsequently, biostatisticians receive novel tools to rapidly build statistical projections in a record amount of time as well. The overall key word behind computerised advancement of biological, genetic and medical research is “time”, as bioinformatics and biotechnology represent two key “enzymes” for the “chemical reaction” of clinical research. As a result, the field of epidemiology has received a novel, all-time opportunity to rapidly build statistical projections regarding the probability of future epidemic or even pandemic outbreaks, and the scope of the current paper is to academically raise awareness of such an opportunity, and to briefly describe examples through which such process could be performed from a general, biomedical science-based perspective. The central scope of the such research efforts could be the investigation of past single-nucleotide polymorphism events that have occurred in certain polymorphic viral agents responsible for the onset of clinically powerful infectious diseases, both indirectly, through the suppression of first-line and second-line host immunity, as well as directly, through the causation of clinical disease and even decrease in the quality of the induced third-line immune activity, to bring an overall system of scientific projection that will bring above-threshold levels of accuracy in the calculation of statistical probabilities regarding the onset of future epidemics and pandemics. It is important to observe the manner microbial agents substantially contribute to the onset of significant and severe clinical infectious diseases, not only by directly inducing the disease and targeting the central, adaptive elements of the host immune system, but also by causing a transient delay in the activation of the peripheral, innate elements of the host immunity, which play a critical role in the creation of an optimal bridge of signalling to the adaptive immunity, meaning that an optimal innate immune activation process plays a crucial role in a timely and qualitative activation of the adaptive immune system.

Keywords: AI; bioinformatics; microbiology; evolutionary biology; immunology; pharmacology; viral mutation; immune adaptation; research catalysis

Introduction

Microbial evolution has played a major role in the onset of public health-related crises that led to a significant number of disease cases and induced casualties among the world population, as numerous polymorphic infectious agents have evolved by adapting their genes to environmental changes, which are represented in this case by the adaptation of human and animal immunity to such infectious events. Currently, it seems that several polymorphic microbes tend to take an evolutionary

advantage over the innate immune system, more than the adaptive immune system, since the latter is known as an “effector” immune department as it directly counteracts the pathophysiological effects of the microbe, whilst the first is known as the “inducer” or “bridging” immune department, helping the adaptive immune system undergo a timely and proportionate activation process. Given the current context of an unbalanced between advanced microbial evolution and an under competing human first-line and second-line immunity, it may now be more important to develop softwares and algorithms capable of projecting the emergence of polymorphic microbial variants with novel, advanced capabilities of hijacking such human innate immunity, which plays a critical role in the creation of a competent bridge to the core, adaptive immune elements and pathways. There is a highly complex context in which viral genomes express specific proteins that are responsible for the prevention of their detection by Pattern-Recognition Receptors (PRRs), such as Toll-Like Receptors (TLRs) 3, 7 and 8, as well as Melanoma Differentiation-Associated protein 5 (MDA5) and Retinoic acid-Inducible Gene I (RIG-I), leading to a lack of detection of both Pathogen-Associated Molecular Patterns (PAMPs), which constitute part of the viral genome, and Damage-Associated Molecular Patterns (DAMPs), which represent toxins produced by the virus. As a result, the overall activation of genes encoding the interferon system may be delayed for a significant amount of time, leading to an unrestricted increase and distribution of the viral load during the initial stages of infection, favouring the onset of severe immunopathogenesis, particularly in people with one or more underlying health condition that already impact the quality of their immune activation. Often, such viral genomes synthesise proteins that directly antagonise the activation of the interferon system and of its byproduct cytokine pathways, and such initial interferon deficiency will lead to an overproduction of interferon glycoproteins once cells begin detecting the viral load, as this often occurs only by the time the viral load is present in multiple tissues, organs and even organ systems. A highly relevant example may be observed in the case of the recent SARS-CoV-2-induced COVID-19 pandemic, in which the virus was seen to produce non-structural proteins (NSPs), like NSP1, NSP10, NSP14 and NSP16. Whilst NSP10, NSP14 and NSP16 were observed to interact with each other and indirectly inhibit the interferon system by preventing the activation of various important PRRs, NSP1 was observed to directly inhibit the host interferon-based signalling process independently. NSP1 may be a major viral protein, not only in the case of the COVID-19 pandemic, but also in various Influenza A disease variants, as it was discovered that NSP1 binds to the NSP1-binding protein (NSP1-BP) and causes a decrease in the quality of NSP1-BP’s molecular activity that brings benefit to the viral non-structural protein in cause. Furthermore, with each round of viral mutation, it was observed that the virus gained fresh capabilities of inhibiting the interferon response, despite previously-built immune memory, which often resulted in repeated contraction of clinical disease. Such events of repeated contraction of infectious clinical disease were broadly observed in many other infectious diseases caused by polymorphic viruses, such as the Influenza A Virus (IAV). There may now be an existing statistical probability that there is a proportional relationship between the duration of a pre-symptomatic stage of an infectious disease, such as SARS-CoV-2-induced COVID-19, and an increased incidence of severe pathogenesis and pathophysiology, given the fact that direct and indirect microbial self-camouflaging look to play a major role in transiently preventing the bridging between innate and adaptive immune activation, whose effects can sometimes be catastrophic for the quality of the eventually induced overall immune response - particularly for patients with underlying immune and perhaps also non-immune health conditions. It is also possible that the incidence of post COVID-19 syndrome, which is a chronic inflammatory disease that occurs as a result of autoimmune activation, is also proportional with the lengthy pre-symptomatic phase of the viral disease. Likewise, alongside the buildup for AI-derived statistical research regarding probabilities that microbial SNPs may directly induce host pathophysiology and impact the integrity and quality of the adaptive immune response, it may also be crucially important for bioinformaticians to develop softwares and statistical projection algorithms for the purpose of predicting future mutations that stimulate the virus to freshly inhibit host innate immune activation, as this often can lead to a repeated onset of clinical infectious disease. Such programs and algorithms may also be developed to predict zoonotic events that may be statistically probable to lead to epidemic or even

pandemic outbreaks. Just as the progressive development of technology and AI has helped catalyse the process of molecular diagnostic procedures, such as Polymerase Chain Reaction, by turning it from traditional (PCR) to real-time (RT-qPCR), so it may be possible, in a similar fashion, for bioinformaticians and biostatisticians to catalyse the process of projecting statistical probabilities concerning the onset of future epidemics or even pandemics caused by specific microbial agents. Even such a process of Real-Time qPCR has been undergoing considerable optimisations to make it even faster and more accurate than during the initial stages of the RT-qPCR molecular testing implementation (Gadkar, V.y, and Filion, M., 2014).

Discussion

Paradoxically, Single-Nucleotide Polymorphisms (SNPs) constitute both critical ways of biological adaptation seen throughout evolutionary history, as well as a concerning type of genetic mutations that have been occurring in diverse life forms, with consequences that often severely impact the lifespan of both the life form itself and the lifespan of other surrounding organisms and microorganisms. SNPs are known to involve three types of genetic mutations with regards to the structure of the gene: insertions, deletions and substitutions, as well as three types of mutations with regards to gene functionality: silent, missense and nonsense, with the latter two often bringing a severe molecular impact upon the production of the involved proteins. Nonetheless, SNPs are also known to bring effective results for specific life forms, helping them better adapt to novel environmental changes and remain in the circle of natural selection. In the areas of virology and epidemiology, SNPs are widely known to involve the creation and adaptation of novel variants of microorganisms, becoming or remaining at least as capable of causing severe clinical illness in their target organisms as before. For example, it has recently been discovered that the emergence of novel variants of microbes through recurring SNP events bring an impact of a continuing pathogenesis of oncogenic diseases so major that the phenomenon has been regarded as a hallmark of cancer induction (Lythgoe et al., 2022). The topic of frequent occurrence of SNP events in microbial agents could not have been more relevant for the recent SARS-CoV-2-induced COVID-19 pandemic, as it has been recorded that SNP events in various important parts of the viral genome occurred during each round of variant emergence and natural selection, which often involved a weakening of the average induced pathophysiology, but an increase of the molecular abilities of the virus to evade the host interferon system (Suh S. et al., 2022). A major sign of such an aspect is that important early variants of the SARS-CoV-2, highly polymorphic viral agent, were discovered through an SNP genotyping molecular testing procedure (Harper H. et al., 2021). A similar trend may be observed with regards to the IAV agent, which is known to cause the flu respiratory disease, as researchers discovered viral variants using SNP genotyping as well (Duan S. et al., 2011). Furthermore, it was discovered that the IFN-gamma-encoding human gene underwent SNP during adaptation to the H5N1 strain of the IAV virus, which helped the immune system become resistant to such viral strain (Bin Ji et al., 2015). It may likewise be that SNP changes are also at the core of human adaptation to microbial evolution against host immunity. Likewise, virologists, epidemiologists, clinical bioinformaticians and biostatisticians could share their clinical field for the purpose of helping scientific research project the occurrence of certain moments in which an SNP event causative of significant microbiological adaptation occurs in the future. Given the fact that numerous polymorphic viruses have been prioritising a targeting of the first- and second-line immune weak points in their adaptation process against the central, third-line host immunity, it could be important to analyse the history of SNP events in viral genes encoding proteins that bring such molecular effects, to ultimately project the occurrence of future SNP events of potential concern for the evolutionary state of the host innate immune system. The innate and adaptive departments of the host immune system both play their unique role in shaping the overall defence system of the host organisms, and the overall clinical objective of improving the human evolutionary trajectory against pathogenic agents with tricky ways of evading the immune system is to help create a better and wider bridge between the two immune departments, potentially describing a "United Immune System" concept. Such an aspect may involve a direct address of the weak evolutionary points that are likely

existent in important first-line and second-line, innate immune systems, such as the interferon system and locally even the complement system, whilst continuing efforts to develop therapeutic and immunising agents that continue to aid the third-line, adaptive immune system build more effective signals, as well as a wider, more durative and fortified memory. Firsthand, it had been theorised that the innate immune system brings speedy and broad-spectrum effects of microbial clearance and that it lacks specificity and memory, unlike the adaptive immune system, which is deemed as the immune department having specificity and displaying long-term memory against certain pathogenic agents. Nonetheless, it was only recently discovered that the innate immunity actually displays considerable extents of both “specificity” and “memory”, and that the adaptive immunity also displays “unspecificity” after all. Likewise, there may now be a possibility that vaccine-based research efforts will bring a wider inclusion of major innate immune elements, such as Type I and Type III Interferon glycoproteins, as they would not only represent vaccine adjuvants, but immunising agents themselves. It may be through such a wider inclusion of innate immunity into vaccinology that the concept of “United Immune System” will be successfully implemented over time (Carp T., 2024).

Underlying health conditions, which are both immune and non-immune in nature, have altogether contributed to the phenomenon of delayed and exaggerated Type I and Type III Interferon-based autocrine and paracrine signalling, leading to the development of a disrupted balance between produced pro-inflammatory and anti-inflammatory cytokines as a result of a more chaotic signalling rate to genes that synthesise cytokines as such. Type I and Type III Interferons may be deemed as pre-cytokine immune glycoproteins, whilst Type II Interferons would more likely be deemed as part of the cytokine family, given the fact that the production of Type II Interferons is dependent upon the activation extent of Interferon-Stimulated Genes (ISGs), which themselves synthesise various cytokines and whose activation is resulted from the autocrine and paracrine signalling of Type I and Type III IFN glycoproteins to the IFNAR1/IFNAR2 receptor complex and the IFNL1R/IL10R2 receptor complex respectively, which is followed by the signalling transduction process that ends with the phosphorylation of the transcription factors for the ISG family. Henceforth, it may be important to differentiate the difference between Type II Interferons, which are produced by natural lymphocytes as a result of their ISG cytokine protein-based recruitment, and Type I and III Interferons, which are produced and signalled by several types of immune cells, such as plasmacytoid dendritic cells (pDCs). As a result, it may be that delays in the timing of IFN I and III synthesis, exocytosis and signalling may bring the most problematic effects of delay and exaggeration of the overall immune activation, as the quality of IFN II-based expression seems to depend upon the quality of IFN I and III-based production and successful signalling. Whilst it is important to mention that a normalised synthesis of Type II Interferons is more directly responsible for a normalised recruitment of adaptive lymphocytes that include major types of B- and T- Lymphocytes, it is as important to mention that Type I and Type III IFNs being at the highest point of the foundation of immune activation ultimately plays the most important role in the adequate synthesis of IFN II, meaning that a normalised recruitment of adaptive lymphocytes is ultimately mostly dependent upon the adequate activation of Interferon-encoding genes (INGs) that express classes of Interferon glycoprotein as such. In short, it is statistically probable that Type I and Type III Interferons are included in the group of immune proteins that are most foundational for the adequate development of the bridge between the innate and the adaptive immune departments as soon as the initial infective moment occurs, and it is possible that the best approach paving the way to a successful buildup of a foundational immunological tree implicates the detection of the immune elements that play the most and the least foundational roles respectively. The classification of immune elements as pre-cytokine, cytokine and post-cytokine could further support such scientific and intellectual efforts, as they can only increase the resolution of an immunological tree as such. It is thereby critical for such AI-derived algorithms to include the potential scientific fact above in order for the best research efforts for the development of pharmaceutical, therapeutic and vaccine-based approaches to ultimately be performed. It seems that an evolutionary push for the adaptation of human immunity to the SNP changes of polymorphic viruses in their genes suppressive of human Type I and Type III IFN-encoding genes (INGs) involves two parts: a more direct, immune system-based approach that

frontally increases the sensitisation of such INGs, foundationally by using recombinant Type I and Type III IFNs, and a less direct, pathogen-based approach that helps the process of the sensitisation of the ING activity by means of the utilisation of wholly inactivated or lysed microbial genomes that only contain slightly-activated genes involved in the suppression of such IFN production. A harmonic cooperation of the two approaches may bring the highest rate of efficacy amongst all candidates.

Furthermore, a substance named protollin was tested through a first-phase, double-blind clinical trial, for the delay in the onset of the Alzheimer's Disease neurodegenerative disease, which is known to be caused by aggregates of beta-amyloid misfolded protein in the encephalon. Namely, the substance was administered in a small dosage nasally, and the results were promising, as helper CD4+ and cytotoxic CD8+ T-Lymphocytes were recruited to the encephalon, where they activated microglial cells and prevented the lysis and deposition of astrocytes, helping in the process of destroying both soluble and insoluble beta-amyloid aggregates before the clinical onset of the neurodegenerative disease, which often occurs in relatively elderly patients (Frenkel D. et al., 2005). Furthermore, collected clinical evidence even suggested low-dose protollin plays a role as an immunising agent against Alzheimer's Disease, as it brought prophylactic immune effects against beta-amyloid aggregate-induced neuronal pathogenesis (Frenkel D. et al., 2008). Likewise, protollin may be used for a delay in the onset of other protein-related degenerative diseases, including Retinitis Pigmentosa, which is caused by the pathogenic action of Rhodopsin aggregates against the rod photoreceptors upon the retinal cells, essentially damaging and destroying an increasing count of cells with rods and eventually causing a complete loss of vision, oftentimes by the time the patients reach an age of 40-45. Such an application may be feasible due to the proximity of the optical system to the central nervous system, though simultaneously, medical and clinical caution is specifically required in such a case due to the fact that the retina is sensitive to any slight extent of immune activation. Moreover, a low dose of protollin may even be used alongside a low dose of Type I and Type III Interferon glycoproteins to treat both natural and adaptive lymphocytes to optimise their immune function and integrity. In the case of HIV-induced AIDS, a treatment of adaptive lymphocytes with protollin as well may further sharpen their defence mechanisms against the retrovirus, as well as improve their immune efficacy, potentially bringing an utmost "evolutionary punch of self-defence" against microbial agents like HIV. Likewise, protollin, Type I IFNs and Type III IFNs, if used in low concentrations, may both individually and together constitute whole immunising agents against various diseases that involve various extents of immune activation (Carp T., 2024). Moreover, there are several plant-based proteins that play various roles in sensitising the activation of Type I and Type III Interferon-encoding genes (INGs) and of Interferon-Stimulated Genes (ISGs), and likewise, principles of Translational Medicine and Pharmacology may be used to develop small drug-like molecules with the aim of sensitising such elements of the interferon system to the best of the research abilities, whilst preventing the onset of any adverse event to the same best of the scientific abilities. The plants that synthesise such interferon-stimulating proteins include: *Silybum marianum*, *Astragalus membranaceus*, *Schisandra chinensis*, *Agaricus blazei*, *Ganoderma lucidum*, *Morinda citrifolia*, *Aloe vera* and *Foeniculum vulgare* (Dacia Plant, 2021). Overall, clinical researchers may bring an effort of utilising a combination of various Type I and Type III Interferon-stimulating proteins to create an overall, proportional "evolutionary punch of self-defence" that will support the human immune system in dominating the microbial mechanisms of evading first- and second-line immunity once again.

Initially, the UniProt program was used to collect the amino acid sequences of the concerned viral protein, before the Clustal Omega program was used to perform multiple sequence alignment involving the listed proteins above. The above programs were used because the research involved whole protein sequences, without requiring a BLAST- and FASTA-based sequencing to detect the protein source of specific amino acid sequences (Pearson et al., 2014). Derivatives of MEGABLAST and G-BLASTN were also used to increase the speed and accuracy of such AI-mediated procedures (Zhao, K. and Chu, X., 2014). Alongside the BLAST and FASTA, basic-level processes of nucleotide and protein sequencing, similarities may also researched and predicted using the Structured Query

Language (SQL) programming method, which constitutes a back-end computerised approach of building operations to determine the existence or absence of statistical significance that specific protein sequences are homologous in nature, and such an approach may be required in the context of analysing and predicting the occurrence of specific SNPs in pathogenic microbes of public health concern, due to the need of further research catalysis as a result of the highly complex spectrum of factors that lead to such SNP occurrences (Pearson et al., 2017). A large degree of homology has been observed in all protein variants listed above, as the percentage of identity exceeded 30% and the e-value was smaller than 0.001 respectively, indicating the substantial nature of the hit. When the e-value is smaller than 0.001, the p-value implying the null hypothesis that there is no association between the protein is smaller than 0.05, meaning that the null hypothesis ought to be rejected. In multiple sequence alignment, it is important to search for three key amino acids: Serine (Ser/S), Histidine (His/H) and Aspartic Acid (Asp/D). Serine is known to have a non-polar, aliphatic R group, whilst Histidine is known to have a non-polar, aromatic R group, whilst Aspartic Acid contains a polar, negatively-charged R group. These three key amino acids are found in sequences of 3-10 amino acids that constitute the identical regions of the amino acid sequences, and they make up the major active site residues that facilitate the hydrolytic chemical reaction of the peptide bond (conserved motifs). In other words, genetic changes play a foundational role in the determination of the biochemical configuration of translated and folded proteins.

With regards to statistical projections, computer softwares like SPSS could be used for the collection of all necessary statistical calculated data in order to build finalised patterns of mathematical prediction with an above-threshold level of reliability. Through the utilisation of softwares like Microsoft Excel and SPSS, statistical tests that include Chi-Squared tests, as well as paired and unpaired T-tests, are performed, with scientists often managing to determine the relationship of correlation and causality without difficulty or significant temporal expenditure likewise. By the statistical analysis of past evolutionary viral processes and their single nucleotide polymorphisms (SNPs), scientists could build predictive algorithms for future viral mutations and determine statistical probability about whether there will be certain mutations causative of above-threshold levels of gain-of-function in viral proteins not directly causative of immunopathology, but of transient cellular immune silencing. There are various environmental, microbiological and proteomic factors that influence the morbidity and case-fatality ratios of epidemic and pandemic diseases, and the core underlying influence lies within the boundaries of evolutionary biology. One example in which computerised systems may be updated to involve artificial intelligence (AI) models would implicate the update the integrated engine of the Laboratory Information Management System (LIMS) - which stores genetic data from hospital patients - comprising a data pool for processed Real-Time Quantitative Polymerase Chain Reaction (RT-qPCR) test results - through the inclusion of a franklin.ai model, which was created by the Sonic Healthcare UK organisation and designed to automatise the entry and processing of genetic data from patients suspected of cancer. Likewise, franklin.ai increases the storage of the collected data through the increase of the frequency of the data processing per day (Smith S. C. et al., 2024). Likewise, such an AI model or similar ones may be included in other software systems that contain data collected from genetic sequencing and the assembly of numerous phylogenetic trees, in order for the computerised process of biological and statistical research to be further catalysed. Such AI models could also be used for the pools of statistical data in order to help the speed of statistical projections to continue increasing. Namely, AI models like franklin.ai could be installed into the integrated engine of biological, genomic and statistical systems, which normally contains multiple kinds of data formats, where exchange regularly takes place, and the overall objective would be to catalyse such process of data exchange and processing in order to increase the storage of such systems, though it is probable for the model to be more suitable for a catalysis of genetic data entry. Nonetheless, such an AI-related update may represent a significant example of how manual processes of data entry can be automatized, helping computerised research processes become faster and more accurate. By such processes, it may be more feasible for bioinformaticians and biostatisticians to determine statistical probabilities that mutations

in specific regions of microbial genomes will occur and that novel infectious disease outbreaks would occur as a result.

The scope of the present study is to investigate the extent to which the transient capabilities of microbes to hijack main innate immune pathways influence the rates of morbidity and lethality caused by the induced infectious disease afterward. To effectively perform such a study, we will investigate the relationship between single-nucleotide polymorphisms in genes encoding various viral proteins and their consequent adaptation to novel immune memory. Specifically, we will observe the relationship between such viral mutations and changes in the biochemical profile of secondary and tertiary structures of such proteins, given the existence of various biochemical versions of amino acid-based side chains, which play a critical role in the determination of the protein's stability, functionality and overall state of activity. It may currently be deemed as statistically probable that the more functional and active such viral proteins are, the more capable is the overall viral genome to evade critical elements of the host immune system. Afterward, it would be necessary to build phylogenetic trees of such viral proteins to observe their evolutionary history over a large, inter-era extent of time. It would be possible to perform such a function using the available databases for pairwise sequence and multiple sequence alignment. After the output results are published, the implicated researchers may then observe the amino acid sequence of each aligned protein, as well as the degree of homology between each protein variant. Then, the secondary structure of each aligned protein can be analysed in detail using various available softwares. There are multiple features within such proteins that require research attention, including the existence of either Serine, Histidine or Aspartate in the conserved motifs, as well as the existence or absence of polarity, the existence or absence of an electric charge, the extent of repetition of amino acids whose R groups have the same electric charge or extent of polarity. Likewise, the biochemical profile of the aligned proteins is directly influenced by their genetic makeup, which in turn is directly influenced by SNPs. Throughout such analysis, researchers may develop statistical programmes to estimate probabilities that future mutations in specific microbial agents will lead to unique zoonotic processes or even cause the selection of a previous variant into one causative of major epidemic or pandemic illnesses. The overall process can be repeated for each species of the viral proteins implicated, and for each individual virus, there may be separate statistical projections made regarding any occurrence of future mutations leading to the development of new public health-related disease outbreaks. One important example of a factor leading to such consequences would be the frequency of SNPs in genetically active and influential areas of each viral genome. Furthermore, there may be a wide spectrum of environmental factors that may need to be taken into consideration during such processes of projecting the occurrence of specific SNP events in microbial agents. Whilst it is important to involve a broad methodology that is based on Artificial Intelligence, it is equally important to continue basing scientific and statistical research on efforts of human intellectual and wisdom, given that it is crucial to continue thinking outside of the box, and implicitly, outside of the algorithms that are based on an already-enclosed box of past analyses, in order for better scientific discovery to continue arising. Overall, it is important to apprehend the high level of complexity that such biostatistical research may be displaying, meaning that intellectual efforts may require to be extensive in nature and that research teams may need to be broad in both numbers and diversity of scientific background.

Conclusions

The topic of single-nucleotide polymorphism in both pathogenic agents causative of concerning human and animal diseases, and host organisms adapting to "updates" in microbial adaptation could not have been more relevant than currently, given the context of evolving illnesses that present diverse mechanisms severely impacting the health state of the host organisms, both short- and long-term. With regards to the development of projective softwares and algorithms, it is firstly important to search for the protein sequences on their matching databases. It may be important to also utilise back-end programming languages, such as SQL, to further catalyse the process of predicting protein homologues where necessary. Then, it is critical to detect past occurrences of single-nucleotide

polymorphisms, and such a process does require a degree of genome sequencing. Analysing the genetic profile of each viral protein of concern is crucial in the determination of a thorough explanation involving the biochemical profiles and changes of each protein. Various softwares involving Artificial Intelligence may now already be available to perform such functions. Furthermore, such two steps can be taken bi-directionally and such an analytical nature would be required in the case of complex evolutionary patterns behind advanced levels of adaptation to human and animal immunity. It is important to apprehend the fact that even a few mutations may be enough to significantly change the biochemical profile of secondary and tertiary structure proteins, making them either pronouncedly more stable and active in their new conformations, or in the contrary, much less. Often, it may also be important to research quaternary structure proteins, as they consist of multiple tertiary structure proteins that have interacted with each other and joined together, just as in immunology, multiple distinct cells that constitute part of the immune system often join together to form a megacell during exposure to various infectious agents. Such factors will likewise markedly influence the evolutionary growth of such proteins, and implicitly, of their viral producers. The scope of the research is to review the evolutionary context behind the pendulum of competition between microbial pathogenesis and mechanisms of immune counteraction, and analyse the relationship between microbial SNPs and the enhanced/suppressed stability and activity of such viral non-structural proteins, whose change seem to play a crucial role in determining the severity of future epidemic outbreaks, due to the strong relationship between the extent and duration of transient viral immune evasion, and the average case-fatality ratio of the caused infectious diseases. Such potential applications of bioinformatics and biostatistics into microbiology and epidemiology could very well include microbial mutation events that involve genes directly responsible for the onset of clinical infectious disease, as well as the decrease in the quality of adaptive immune responses. To offer an example regarding the importance of a computerised catalysis of clinical research, an update of the Polymerase Chain Reaction molecular testing from a traditional version to a technological, real-time variant, has helped researchers detect specific microbial agents or major genes in hours, rather than days. Likewise, a computerised catalysis of epidemiological projective research from a traditional version to a technological one may help scientists build more detailed and accurate statistical probabilities regarding the onset of future public health crises, and most importantly, potentially in an exponentially more timely manner. Such intellectual, scientific and digital efforts could be merged for an overall development of a projective system that may help humanity better prepare for future public health crises, by bringing relevant statistical estimations and probabilities that specific viruses will undergo specific SNP events that would bring a major change in their polymorphic evolutionary trajectory.

References

1. Wolff, T., O'Neill, R. E., & Palese, P. (1998). NS1-Binding protein (NS1-BP): a novel human protein that interacts with the influenza A virus nonstructural NS1 protein is relocalized in the nuclei of infected cells. *Journal of virology*, 72(9), 7170–7180. <https://doi.org/10.1128/JVI.72.9.7170-7180.1998>
2. Lythgoe, M. P., Mullish, B. H., Frampton, A. E., & Krell, J. (2022). Polymorphic microbes: a new emerging hallmark of cancer. *Trends in microbiology*, 30(12), 1131–1134. <https://doi.org/10.1016/j.tim.2022.08.004>
3. Rollin, J., Bester, R., Brostaux, Y., Caglayan, K., De Jonghe, K., Eichmeier, A., Foucart, Y., Haegeman, A., Koloniuk, I., Kominek, P., Maree, H., Onder, S., Posada Céspedes, S., Roumi, V., Šafářová, D., Schumpp, O., Ulubas Serce, C., Sömera, M., Tamisier, L., Vainio, E., ... Massart, S. (2023). Detection of single nucleotide polymorphisms in virus genomes assembled from high-throughput sequencing data: large-scale performance testing of sequence analysis strategies. *PeerJ*, 11, e15816. <https://doi.org/10.7717/peerj.15816>
4. Harper, H., BurrIDGE, A., Winfield, M., Finn, A., Davidson, A., Matthews, D., Hutchings, S., Vipond, B., Jain, N., COVID-19 Genomics UK (COG-UK) Consortium, Edwards, K., & Barker, G. (2021). Detecting SARS-CoV-2 variants with SNP genotyping. *PloS one*, 16(2), e0243185. <https://doi.org/10.1371/journal.pone.0243185>

5. Henriques-Santos, B. M., Farjun, B., Corrêa, I. A., Figueiredo, J. B., Fidalgo-Neto, A. A., & Kuriyama, S. N. (2022). SARS-CoV-2 Variant Determination Through SNP Assays in Samples From Industry Workers From Rio de Janeiro, Brazil. *Frontiers in microbiology*, *12*, 757783. <https://doi.org/10.3389/fmicb.2021.757783>
6. Ghosh, N., Saha, I., Sharma, N., Nandi, S., & Plewczynski, D. (2021). Genome-wide analysis of 10664 SARS-CoV-2 genomes to identify virus strains in 73 countries based on single nucleotide polymorphism. *Virus research*, *298*, 198401. <https://doi.org/10.1016/j.virusres.2021.198401>
7. Suh, S., Lee, S., Gym, H. et al. A systematic review on papers that study on Single Nucleotide Polymorphism that affects coronavirus 2019 severity. *BMC Infect Dis* *22*, 47 (2022). <https://doi.org/10.1186/s12879-022-07034-w>
8. Roy-Vallejo E, Fernández De Córdoba-Oñate S, Delgado-Wicke P, Triguero-Martínez A, Montes N, Carracedo-Rodríguez R, Zurita-Cruz N, Marcos-Jiménez A, Lamana A, Galván-Román JM, Villapalos García G, Zubiaur P, Ciudad M, Rabes L, Sanz M, Rodríguez C, Villa A, Rodríguez JÁ, Marcos C, Hernando J, Díaz-Fernández P, Abad F, de los Santos I, Rodríguez Serrano DA, García-Vicuña R, Suárez Fernández C, P. Gomariz RP, Muñoz-Calleja C, Fernández-Ruiz E, González-Álvaro I, Cardeñoso L, and the PREDINMUN-COVID Group (2023) Occurrence of SARS-CoV-2 viremia is associated with genetic variants of genes related to COVID-19 pathogenesis. *Front. Med.* 10:1215246. doi: <https://doi.org/10.3389/fmed.2023.1215246>
9. Kim, Y. C., Won, S. Y., & Jeong, B. H. (2021). The first association study of single-nucleotide polymorphisms (SNPs) of the *IFITM1* gene with influenza H1N1 2009 pandemic virus infection. *Molecular & cellular toxicology*, *17*(2), 179–186. <https://doi.org/10.1007/s13273-021-00123-y>
10. Mirzo, S. M., Kumar, A., Sharma, N. K., Li, L., Balshaw, R., Plummer, F. A., Luo, M., & Liang, B. (2022). NFκB1 Polymorphisms Are Associated with Severe Influenza A (H1N1) Virus Infection in a Canadian Population. *Microorganisms*, *10*(10), 1886. <https://doi.org/10.3390/microorganisms10101886>
11. Duan, S., Boltz, D. A., Li, J., Oshansky, C. M., Marjuki, H., Barman, S., Webby, R. J., Webster, R. G., & Govorkova, E. A. (2011). Novel genotyping and quantitative analysis of neuraminidase inhibitor resistance-associated mutations in influenza A viruses by single-nucleotide polymorphism analysis. *Antimicrobial agents and chemotherapy*, *55*(10), 4718–4727. <https://doi.org/10.1128/AAC.00316-11>
12. Ji, B., Sun, T. T., Ma, Z. L., Lu, Q. F., Hu, W. L., Jian, Z. H., Jiang, Y., & Chen, P. F. (2015). Possible Association of IFN-γ Gene -316A/G SNP with Humoral Immune Response to Killed H5N1 HPAI Vaccine in a Red Junglefowl Population. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research*, *35*(11), 844–849. <https://doi.org/10.1089/jir.2015.0018>
13. Wang, J., Luo, Y., Katiyar, H., Liang, C., & Liu, Q. (2024). The Antiviral Activity of Interferon-Induced Transmembrane Proteins and Virus Evasion Strategies. *Viruses*, *16*(5), 734. <https://doi.org/10.3390/v16050734>
14. Liao, Y., Goraya, M. U., Yuan, X., Zhang, B., Chiu, S. H., & Chen, J. L. (2019). Functional Involvement of Interferon-Inducible Transmembrane Proteins in Antiviral Immunity. *Frontiers in microbiology*, *10*, 1097. <https://doi.org/10.3389/fmicb.2019.01097>
15. Taylor, J. P., Blum, S. I., Graffeo, H. C., Shang, Q., Qiu, S., Green, T. J., Botta, D., Lund, F. E., & Tse, H. M. (2024). The Type 1 Diabetes-Associated Single Nucleotide Polymorphism rs1990760 in IFIH1 Is Associated with Increased Basal Type I IFNs and IFN-stimulated Gene Expression. *Journal of immunology (Baltimore, Md. : 1950)*, *213*(10), 1415–1428. <https://doi.org/10.4049/jimmunol.2400344>
16. Looney, B. M., Xia, C. Q., Concannon, P., Ostrov, D. A., & Clare-Salzler, M. J. (2015). Effects of type 1 diabetes-associated IFIH1 polymorphisms on MDA5 function and expression. *Current diabetes reports*, *15*(11), 96. <https://doi.org/10.1007/s11892-015-0656-8>
17. Lincez, P. J., Shanina, I., & Horwitz, M. S. (2015). Reduced expression of the MDA5 Gene IFIH1 prevents autoimmune diabetes. *Diabetes*, *64*(6), 2184–2193. <https://doi.org/10.2337/db14-1223>
18. Chistiakov, D. A., Voronova, N. V., Savost'Anov, K. V., & Turakulov, R. I. (2010). Loss-of-function mutations E6 27X and I923V of IFIH1 are associated with lower poly(I:C)-induced interferon-β production in peripheral blood mononuclear cells of type 1 diabetes patients. *Human immunology*, *71*(11), 1128–1134. <https://doi.org/10.1016/j.humimm.2010.08.005>

19. Frenkel, D., Puckett, L., Petrovic, S., Xia, W., Chen, G., Vega, J., Dembinsky-Vaknin, A., Shen, J., Plante, M., Burt, D. S., & Weiner, H. L. (2008). A nasal proteosome adjuvant activates microglia and prevents amyloid deposition. *Annals of neurology*, 63(5), 591–601. <https://doi.org/10.1002/ana.21340>
20. Frenkel, D., Maron, R., Burt, D. S., & Weiner, H. L. (2005). Nasal vaccination with a proteosome-based adjuvant and glatiramer acetate clears beta-amyloid in a mouse model of Alzheimer disease. *The Journal of clinical investigation*, 115(9), 2423–2433. <https://doi.org/10.1172/JCI23241>
21. Vello, F., Filippini, F., & Righetto, I. (2024). Bioinformatics Goes Viral: I. Databases, Phylogenetics and Phylodynamics Tools for Boosting Virus Research. *Viruses*, 16(9), 1425. <https://doi.org/10.3390/v16091425>
22. Hölzer, M., & Marz, M. (2017). Software Dedicated to Virus Sequence Analysis “Bioinformatics Goes Viral”. *Advances in virus research*, 99, 233–257. <https://doi.org/10.1016/bs.aivir.2017.08.004>
23. Ibrahim, B., McMahon, D. P., Hufsky, F., Beer, M., Deng, L., Mercier, P. L., Palmarini, M., Thiel, V., & Marz, M. (2018). A new era of virus bioinformatics. *Virus research*, 251, 86–90. <https://doi.org/10.1016/j.virusres.2018.05.009>
24. Hu, G., & Kurgan, L. (2019). Sequence Similarity Searching. *Current protocols in protein science*, 95(1), e71. <https://doi.org/10.1002/cpp.71>
25. Pearson W. R. (2014). BLAST and FASTA similarity searching for multiple sequence alignment. *Methods in molecular biology (Clifton, N.J.)*, 1079, 75–101. https://doi.org/10.1007/978-1-62703-646-7_5
26. Chen, Y., Ye, W., Zhang, Y., & Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic acids research*, 43(16), 7762–7768. <https://doi.org/10.1093/nar/gkv784>
27. Zhao, K., & Chu, X. (2014). G-BLASTN: accelerating nucleotide alignment by graphics processors. *Bioinformatics (Oxford, England)*, 30(10), 1384–1391. <https://doi.org/10.1093/bioinformatics/btu047>
28. Warnow T. (2021). Revisiting Evaluation of Multiple Sequence Alignment Methods. *Methods in molecular biology (Clifton, N.J.)*, 2231, 299–317. https://doi.org/10.1007/978-1-0716-1036-7_17
29. Edgar R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
30. Edgar R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, 5, 113. <https://doi.org/10.1186/1471-2105-5-113>
31. Duricki, D. A., Soleman, S., & Moon, L. D. (2016). Analysis of longitudinal data from animals with missing values using SPSS. *Nature protocols*, 11(6), 1112–1129. <https://doi.org/10.1038/nprot.2016.048>
32. Shek, D. T., & Ma, C. M. (2011). Longitudinal data analyses using linear mixed models in SPSS: concepts, procedures and illustrations. *TheScientificWorldJournal*, 11, 42–76. <https://doi.org/10.1100/tsw.2011.2>
33. Dudley, W. N., Benuzillo, J. G., & Carrico, M. S. (2004). SPSS and SAS programming for the testing of mediation models. *Nursing research*, 53(1), 59–62. <https://doi.org/10.1097/00006199-200401000-00009>
34. Pearson, W. R., & Mackey, A. J. (2017). Using SQL Databases for Sequence Similarity Searching and Analysis. *Current protocols in bioinformatics*, 59, 9.4.1–9.4.22. <https://doi.org/10.1002/cpbi.32>
35. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic acids research*, 51(D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>
36. Goetsch, W., Beerenwinkel, N., Deng, L., Dölken, L., Dutilh, B. E., Erhard, F., Kaderali, L., von Kleist, M., Marquet, R., Matthijnsens, J., McCallin, S., McMahon, D., Rattei, T., Van Rij, R. P., Robertson, D. L., Schwemmler, M., Stern-Ginossar, N., & Marz, M. (2021). ITN-VIROINF: Understanding (Harmful) Virus-Host Interactions by Linking Virology and Bioinformatics. *Viruses*, 13(5), 766. <https://doi.org/10.3390/v13050766>
37. Sharma, D., Priyadarshini, P., & Vrati, S. (2015). Unraveling the web of viroinformatics: computational tools and databases in virus research. *Journal of virology*, 89(3), 1489–1501. <https://doi.org/10.1128/JVI.02027-14>
38. Smith, S., & Dennis, B. (1991). HL7: facilitating multiple approaches to systems integration. *Computers in healthcare*, 12(2), 59–61.
39. Achard, F., Vaysseix, G., & Barillot, E. (2001). XML, bioinformatics and data integration. *Bioinformatics (Oxford, England)*, 17(2), 115–125. <https://doi.org/10.1093/bioinformatics/17.2.115>
40. Bruhn, R. E., & Burton, P. J. (2003). Designing XML schemas for bioinformatics. *BioTechniques*, 34(6), . <https://doi.org/10.2144/03346st03>

41. Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular diversity*, 25(3), 1315–1360. <https://doi.org/10.1007/s11030-021-10217-3>
42. Zhu H. (2020). Big Data and Artificial Intelligence Modeling for Drug Discovery. *Annual review of pharmacology and toxicology*, 60, 573–589. <https://doi.org/10.1146/annurev-pharmtox-010919-023324>
43. Chopra, H., Baig, A. A., Gautam, R. K., & Kamal, M. A. (2022). Application of Artificial Intelligence in Drug Discovery. *Current pharmaceutical design*, 28(33), 2690–2703. <https://doi.org/10.2174/1381612828666220608141049>
44. Chopra, H., Annu, Shin, D. K., Munjal, K., Priyanka, Dhama, K., & Emran, T. B. (2023). Revolutionizing clinical trials: the role of AI in accelerating medical breakthroughs. *International journal of surgery (London, England)*, 109(12), 4211–4220. <https://doi.org/10.1097/JS9.0000000000000705>
45. Koski, E., & Murphy, J. (2021). AI in Healthcare. *Studies in health technology and informatics*, 284, 295–299. <https://doi.org/10.3233/SHTI210726>
46. Singh, H., Nim, D. K., Randhawa, A. S., & Ahluwalia, S. (2024). Integrating clinical pharmacology and artificial intelligence: potential benefits, challenges, and role of clinical pharmacologists. *Expert review of clinical pharmacology*, 17(4), 381–391. <https://doi.org/10.1080/17512433.2024.2317963>
47. Zhavoronkov, A., Vanhaelen, Q., & Oprea, T. I. (2020). Will Artificial Intelligence for Drug Discovery Impact Clinical Pharmacology?. *Clinical pharmacology and therapeutics*, 107(4), 780–785. <https://doi.org/10.1002/cpt.1795>
48. Thomas, S., Abraham, A., Baldwin, J., Piplani, S., & Petrovsky, N. (2022). Artificial Intelligence in Vaccine and Drug Design. *Methods in molecular biology (Clifton, N.J.)*, 2410, 131–146. https://doi.org/10.1007/978-1-0716-1884-4_6
49. Ananya, Panchariya, D. C., Karthic, A., Singh, S. P., Mani, A., Chawade, A., & Kushwaha, S. (2024). Vaccine design and development: Exploring the interface with computational biology and AI. *International reviews of immunology*, 43(6), 361–380. <https://doi.org/10.1080/08830185.2024.2374546>
50. Hassoun, S., Jefferson, F., Shi, X., Stucky, B., Wang, J., & Rosa, E. (2022). Artificial Intelligence for Biology. *Integrative and comparative biology*, 61(6), 2267–2275. <https://doi.org/10.1093/icb/icab188>
51. Bahrami, A. A., Payandeh, Z., Khalili, S., Zakeri, A., & Bandehpour, M. (2019). Immunoinformatics: In Silico Approaches and Computational Design of a Multi-epitope, Immunogenic Protein. *International reviews of immunology*, 38(6), 307–322. <https://doi.org/10.1080/08830185.2019.1657426>
52. Smith, S. C., Diaz-Perez, J. A., Mochel, M. C., Billings, S. D., Fernandez, L., & Poklepovic, A. S. (2024). A High-grade PML::JAK1 Fusion Sarcoma. *The American journal of surgical pathology*, 10.1097/PAS.0000000000002326. Advance online publication. <https://doi.org/10.1097/PAS.0000000000002326>
53. Watanabe, T., Suzuki, N., Tomonaga, K., Sawa, H., Matsuura, Y., Kawaguchi, Y., Takahashi, H., Nagasaki, K., & Kawaoka, Y. (2019). Neo-virology: The raison d'être of viruses. *Virus research*, 274, 197751. <https://doi.org/10.1016/j.virusres.2019.197751>
54. Dennehy J. J. (2014). What ecologists can tell virologists. *Annual review of microbiology*, 68, 117–135. <https://doi.org/10.1146/annurev-micro-091313-103436>
55. Komarova N. L. (2007). Viral reproductive strategies: How can lytic viruses be evolutionarily competitive?. *Journal of theoretical biology*, 249(4), 766–784. <https://doi.org/10.1016/j.jtbi.2007.09.013>
56. Ruth, B., Peter, S., Ibrahim, B., & Dittrich, P. (2024). Revealing the hierarchical structure of microbial communities. *Scientific reports*, 14(1), 11202. <https://doi.org/10.1038/s41598-024-61836-3>
57. Coudray-Meunier, C., Fraisse, A., Martin-Latil, S., Delannoy, S., Fach, P., & Perelle, S. (2016). A Novel High-Throughput Method for Molecular Detection of Human Pathogenic Viruses Using a Nanofluidic Real-Time PCR System. *PloS one*, 11(1), e0147832. <https://doi.org/10.1371/journal.pone.0147832>
58. Gadkar, V.y, & Filion, M. (2014). New Developments in Quantitative Real-time Polymerase Chain Reaction Technology. *Current issues in molecular biology*, 16, 1–6.
59. Carp, T. N. (2024). Potential Innovations in Modern-Day Human and Animal Vaccine Development. Preprints. <https://doi.org/10.20944/preprints202407.2158.v5>

60. Carp, T. N. (2024). Calibrating Human Immunity in the Context of Advanced Microbial Evolution and Self-Camouflaging. Preprints. <https://doi.org/10.20944/preprints202411.2046.v4>
61. Dacia Plant (2021), *Interferonate tablets*, doi: <https://www.daciaplant.ro/interferonat-comprimat.html>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.