



An integrated eco-evolutionary framework to predict population-level responses of climate-sensitive pathogens

Amy M Campbell^{1,2}, Chris Hutton¹, Craig Baker-Austin²,
Ronny van Aerle² and Jaime Martinez-Urtaza^{2,3}

It is critical to gain insight into how climate change impacts evolutionary responses within climate-sensitive pathogen populations, such as increased resilience, opportunistic responses and the emergence of dominant variants from highly variable genomic backgrounds and subsequent global dispersal. This review proposes a framework to support such analysis, by combining genomic evolutionary analysis with climate time-series data in a novel spatiotemporal dataframe for use within machine learning applications, to understand past and future evolutionary pathogen responses to climate change. Recommendations are presented to increase the feasibility of interdisciplinary applications, including the importance of robust spatiotemporal metadata accompanying genome submission to databases. Such workflows will inform accessible public health tools and early-warning systems, to aid decision-making and mitigate future human health threats.

Addresses

¹ School of Ocean and Earth Science, University of Southampton, National Oceanography Centre, Southampton, UK

² Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Weymouth, UK

³ Department of Genetics and Microbiology, Autonomous University of Barcelona, Barcelona, Spain

Corresponding author: Martinez-Urtaza, Jaime
(jaime.martinez.urtaza@uab.cat)

Current Opinion in Biotechnology 2023, 80:102898

This review comes from a themed issue on **Environmental Biotechnology**

Edited by **Jaime Martinez Urtaza** and **Luis F. De León**

For complete overview of the section, please refer to the article collection, "**Environmental Biotechnology (2023)**"

Available online 3 February 2023

<https://doi.org/10.1016/j.copbio.2023.102898>

0958-1669/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Climate change has implications for human health, in terms of heat vulnerability, food and water insecurity and

exposure to extreme weather events [1,2]. Additionally, these changing environmental conditions are altering the epidemiology of diseases caused by climate-sensitive pathogens that exist in natural ecosystems [1,3], wherein pathogen life cycles are intrinsically linked to environmental conditions (e.g. *Vibrio parahaemolyticus*, a water-borne bacteria causing gastroenteritis [4]), or their host vector is governed itself by the environment (e.g. *Borrelia burgdorferi*, the tick-borne bacteria causing Lyme disease [5,6]). Known mechanisms of how diseases are driven by climate change include seasonal shifts and expansions of transmission periods [1,3], and spatial range expansions of both vectors and pathogens themselves (transcontinental, latitudinal and polewards) [4–10]. However, recent years have seen the transcontinental expansion of certain pathogen variants with pandemic potential [9,10]. The mechanisms of such emergence are unknown, as little research has moved beyond well-established ecological relationships to quantify potential evolutionary responses of the pathogen to changes in their native environment.

Specifically, there is a need to characterise how evolutionary drivers of individual strains or clonal groups (such as mutation and horizontal gene transfer) and demography can be modulated by climate change. Environmental variability, such as temperature changes, leads to pathogen variants either thriving in the new conditions, due to phenotypic plasticity, or adapting [11], indicated by the presence of candidate genes under positive selection [12] that correlate with particular bioclimatic conditions [13], which can lead to clonal replacement by well-adapted variants. Such adaptation could offer a range of benefits, including enhanced survivability, fitness or transport potential.

There is currently no climate-sensitive disease modelling tool that considers a genetic component [14]. The next generation of tools will need to evaluate environmental drivers against observed demographic events (e.g. emergence, bottlenecks, population expansions and migrations) or the uptake of particular genetic material, to retrospectively identify selection pressures introduced by climate change [15], and identify subsequent real-world responses of pathogen populations — which have previously only been hypothesised [6,16] or explored in lab-based studies of adaption and tolerances [17,18].

Box 1 Limitations of existing, individual ecological and evolutionary approaches

Existing ecological models that have been used to explore how species are responding to climate change, have certain limitations when applied to capture the full dynamics of complex evolutionary responses of climate-sensitive pathogens. Often, they assume the species to be an unchanging entity, through optimum niche habitat shifts and dispersal based on physiological limitations of the species, without considering any evolutionary component or opportunities for adaption. Models are based on explicit occurrence/abundance population data [11] and do not allow the introduction of genomic data. Additionally, recently studies exploring evolutionary responses to climate change in terms of population survivability [13] or heterogeneous intra-species responses [38,43], are developed for larger, slower-evolving taxa, such as plants or animals [38,44,45]. Pathogens, however, exhibit rapid evolutionary change on faster timescales with significant horizontal gene transfer, requiring adapted workflows as those that exist are not directly applicable. Finally, certain ecological models act at a species level — but pathogens may not respond homogeneously at a species level, as particularly well-adapted clonal groups emerge from highly variable genomic backgrounds, with multiple variants coexisting in a population, requiring the exploration of individual variant responses and population dynamics. We explore some examples of ecological models used for similar applications below:

- **Species Distribution Models** — also known as ecological niche modelling, the niche of a species is estimated and mapped based on statistical correlations between species occurrence and environmental variables, mostly using Maximum Entropy or Generalised Linear Models [46]. These are known to underestimate [47] or overestimate [48] range boundaries based on simplistic assumptions that ignore adaption or geographic barriers. Evolutionary aspects are omitted — species are treated as having uniform requirements, which is inconsistent with our knowledge of heterogeneous groups of individuals existing in a constant state of change within a single pathogenic species [9].
- **Hybrid Genomic Species Distribution Models** — which instead of species occurrence, can focus on a particular genomic process, for example, identifying specific SNP niches, through geographical correlation between specific SNP frequencies and environmental variables, for example, [13,43–45,49]. Recent studies have been able to add an adaptive evolutionary component in this way, but this still requires combination with further genomic analysis to provide further evolutionary or demographic information, that is, migration rates, evolutionary rates and effective population size [11] to understand population-level responses, with limited methodologies available to do so. Studies are geared towards genetic vulnerability studies of animal/plant species based on discrepancies in allele presence (assuming current populations are well-adapted to local environments) — whereas we need to understand opportunistic pathogen evolution that can rapidly respond to such changes.
- **Metapopulation Models** — these aim to project demographic consequences of populations to changing environmental variables, allowing disconnected patches of species to have independent dynamics and dispersal connections [50]. However, the demography of pathogen populations is largely more complex and dynamic than those of larger taxa the models are developed for, with highly variable genomic backgrounds of many more distinct genotypes, and high rates of horizontal gene transfer between these, upon which certain well-adapted ones emerge — making it difficult to redeploy such models for pathogen evolution. Understanding connections between populations would require the explicit integration of phylogenetic analysis within the model.

Similarly, genomic approaches that aim to quantify how species respond to climate change are explored by Waldvogel et al. [11] who only elucidate a subset of evolutionary processes rather than fully integrating eco-evolutionary dynamics [11]. New approaches, specific to pathogens, have emerged recently, for example, detecting bacterial DNA transfer providing adaption to new environments [51], but require a spatiotemporal element to facilitate integration with environmental data to explicitly pinpoint such selective pressures and responses.

The gap between ecological models and such genomic approaches needs to be bridged with an infrastructure that facilitates seamless integration of eco-evolutionary processes.

This review proposes an option to overcome these limitations through a framework able to integrate genomic data, satellite-derived climate data, epidemiological data and pre-existing tools within machine learning applications to provide novel, interdisciplinary insights into current and future evolutionary pathogen responses to climate change. This framework will allow us to develop real-time public health tools and early-warning systems that can anticipate the future climate-driven emergence of well-adapted pathogenic strains and mitigate future pandemics, ultimately increasing the resilience of health systems to climate-related risks [19].

Beyond ecology

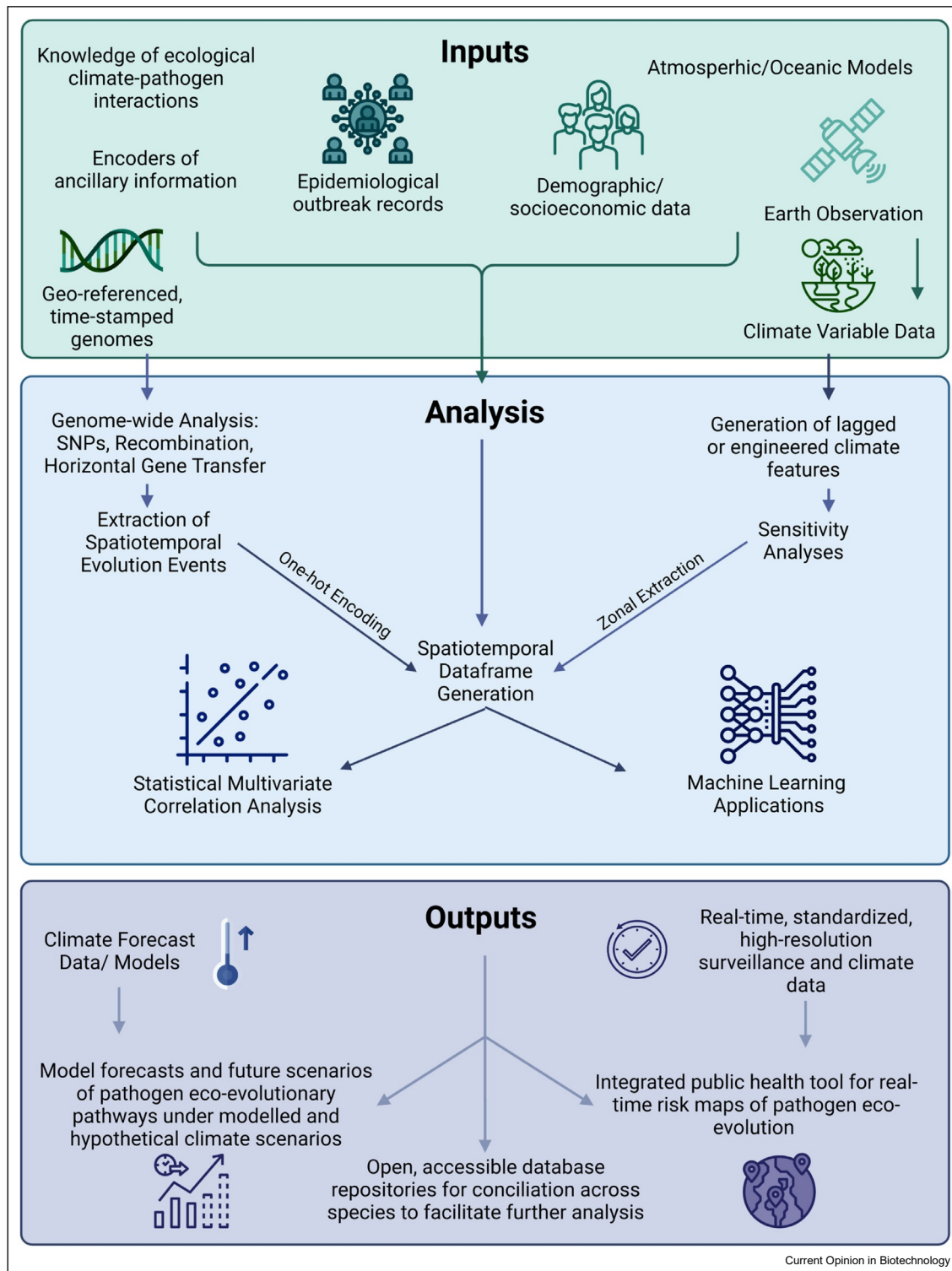
Existing ecological models, that aim to quantify how species will respond to climate change, are not sufficient to integrate genomic data and capture the full dynamics of pathogen-specific evolutionary responses to climate change (Box 1). The advent of eco-evolutionary models is necessary — while ecological data can give insights

into optimum range shifts and geographic abundance, evolutionary genomics, in the form of genomic and phenotypic data, are necessary to provide population-level insights and predictions into genetic adaption, phenotypic plasticity and demography shifts in response to climate change [11].

An eco-evolutionary framework

The framework outlined here (Figure 1) proposes that various pre-existing software are redeployed to integrate ecological dynamics and evolutionary processes, in the context of climate-sensitive pathogens. The seamless integration of phylogenetic analysis, climate timeseries, and additional data or encoders into a novel dataframe will facilitate explicit statistical investigation of complex eco-evolutionary relationships between climate change and pathogens. The dataframe will provide training data for machine learning applications to facilitate forecasting of future pathogen epidemiology, for example, optimum evolution zones and population-specific forecasts. Near-

Figure 1



An eco-evolutionary framework to explore changing epidemiology of climate-sensitive pathogens. A range of interdisciplinary inputs are explored, which are then integrated in a spatiotemporal dataframe, to be used as inputs in machine learning applications. Outputs will be implemented in integrated tools for climate-sensitive pathogen surveillance and forecasts. Created with BioRender.com.

Box 2 Spatiotemporal data requirements and analysis limitations

The toolkit requires genomes (tree tips) to be georeferenced with latitude and longitude values, and time-stamped in the metadata to allow inference of the spatiotemporal context of phylogenetic tree branches and internal tree nodes [15]. Genomic metadata on public repositories is currently insufficient. Here, we strongly recommend the 'Latitude/Longitude', 'Location' and time-stamped 'Collection Date' (in a 'dd-mm-yyyy' format) fields are completed during genomic data submission to databases.

A range of software currently exists that can visualise spatiotemporal relationships by projecting phylogenetic trees in geographical space [23,24,52–54], however, they cannot explicitly facilitate spatiotemporal correlation analysis of driver variables in-house. These inferences would be strengthened with integrated workflows that allow hypotheses to be tested [15] — currently this is only possible for static, univariate correlation analysis between environmental variables and allele frequencies [55] or phylogeny branches [24], omitting complex multivariate relationships between the constantly evolving pathogen and its dynamic environment.

real-time surveillance and climate data will be continuously re-fed into models for reanalysis and refinement. Such a template will be invaluable in the field of epidemiology and human health, but is flexible enough to be applied to a range of climate-sensitive species to increase our understanding of contemporary biodiversity and extinction processes amidst climate change [15].

Inputs**Climate data**

Spatiotemporal data of a range of environmental variables affecting environmental pathogens are required for retrospective analysis, to identify conditions that drive evolutionary shifts. This requires high-resolution data (as pathogen niches are specific and short-lived due to rapid replication) and extensive coverage (to elucidate introduction and transport mechanisms [20]). While field data collection at this scale is unfeasible, solutions can be found in the field of remote sensing that utilises continuous satellite observations to provide consistent, spatial coverage of climate variables for analysis, at up to a 1 km and hourly resolution [21]. Finally, climate forecast data and projections will become increasingly important towards the end of the workflow to predict future eco-evolutionary pathways for pathogens once these relationships have been established.

Genomic data

Tools need to integrate analysis of evolutionary and demographic timescales against timeseries of climate data to interrogate possible interactions and anticipate future evolutionary pathways. To facilitate this analysis, robust genomic surveillance is necessary — ideally providing a wide range of sequenced pathogen genomes from which to infer evolutionary processes. The challenging spatiotemporal requirements of genomic surveillance for this application are discussed in Box 2.

Additional input datasets

Additional datasets to obtain, dependent on user requirements, include human demographic data (shipping and transport routes, exposure), for novel analysis into the global dynamics of pathogens amidst human mobility [22], and epidemiological data, which, if available

and supported by robust surveillance systems, are also useful for cross-validation.

Analysis**Genome-wide analysis to extract evolutionary events**

Bayesian phylogenetic analysis of molecular sequences is utilised to reconstruct phylogenies to pinpoint the time and location of evolutionary events and demographic shifts, to explore their associations with climate anomalies, requiring dated tips with discrete (country or region) or continuous location data (co-ordinates). Phylogenetic outputs (maximum clade credibility trees) are converted into more functional outputs (e.g. dataframes, KML) for further spatiotemporal analysis using various tools [23,24]. Demographic shifts in dominant sequence types and population structure are identified using multilocus sequence typing along a timeseries or clusters of single-nucleotide polymorphisms (SNPs). Genome-wide association studies (GWAS) identify the presence of genes with evolutionary or adaptive significance, to elucidate population-specific responses used by pathogens to thrive and establish themselves in geographically distinct, dynamic environments. Annotating the trend and distribution of accessory gene presence reveals selective signatures and horizontally transferred genetic material — with genes under positive selection [12], suggesting a benefit that outweighs the pace of genetic drift [25]. Recombination events and rates are identified using various software [26–29]. These 'evolutionary metrics' are all converted into useable formats in the next stage for downstream analyses.

Spatiotemporal dataframe generation

Existing spatiotemporal software to explore climate-driven evolution is unable to identify explicit, complex associations (Box 2). This framework advocates for the generation of a fit-for-purpose spatiotemporal dataframe, similar to that used in Campbell et al. [30] but featuring evolutionary-dependent variables, to deploy in various statistical applications and provide appropriate training data for machine learning applications.

Each data point in the nested spatiotemporal dataframe is aggregated by a time unit, such as month, and spatial unit, such as district. Evolutionary metrics and gene presence are converted into a binary format (attributed a 0 or 1), known as ‘one hot encoding’, to indicate if events were observed in this particular spatiotemporal unit. Each step is then appended with associated relevant climate data and appropriate lagged or engineered features, using a zonal extraction function (as used in Ref. [30]), alongside any further discrete or continuous ancillary data, for example, socioeconomic data. This seamless integration of environmental and genomic data offers analysis potential for a range of evolutionary components against climate data in combination, which is lacking in pre-existing models (Box 1).

Statistical multivariate correlation analysis

Multivariate regression (using models such as ordinary least squares, principal component analysis and generalised additive models) can expose statistically significant relationships between the environmental variables and evolutionary metrics. Nonlinear associations can be revealed using gradient boosting trees (which can reveal the rank and relative importance of each variable in driving the evolutionary metric) and partial dependence plots to show the marginal effect of each variable on the evolutionary metric of interest. Significant relationships are those within a 95% confidence interval ($p = < 0.05$). The dataframe can be split into spatially or temporally distinct subgroups to explore spatial or seasonal variations in correlation coefficients. However, the majority of these relationships will likely be complex, and so statistical analysis alone may struggle to explore thresholds and limiting factors of variables interacting in a multivariate landscape.

Machine learning applications for forecasting

Evolutionary biology is extending beyond only reconstructing the past, towards harnessing this to predict future evolutionary processes and dynamics for individual populations [31]. Implementing this requires an effective method of integrating a complex range of interdisciplinary data, spanning all aspects of climate eco-evolutionary analysis, to make forecasts — a complex task that has proved difficult in previous studies [32,33].

Machine learning (ML), in which the system can ‘learn’ relationships between components of the model, is a computationally feasible way to effectively synthesise climate and genomic datasets and draw meaningful predictions from heterogeneous, population-specific nonlinear interactions, functioning on a complicated spatiotemporal, multivariate landscape. ML offers a range of modelling options itself, or has the ability to enhance the potential of existing ecological models by overcoming the limitations mentioned in Box 1,

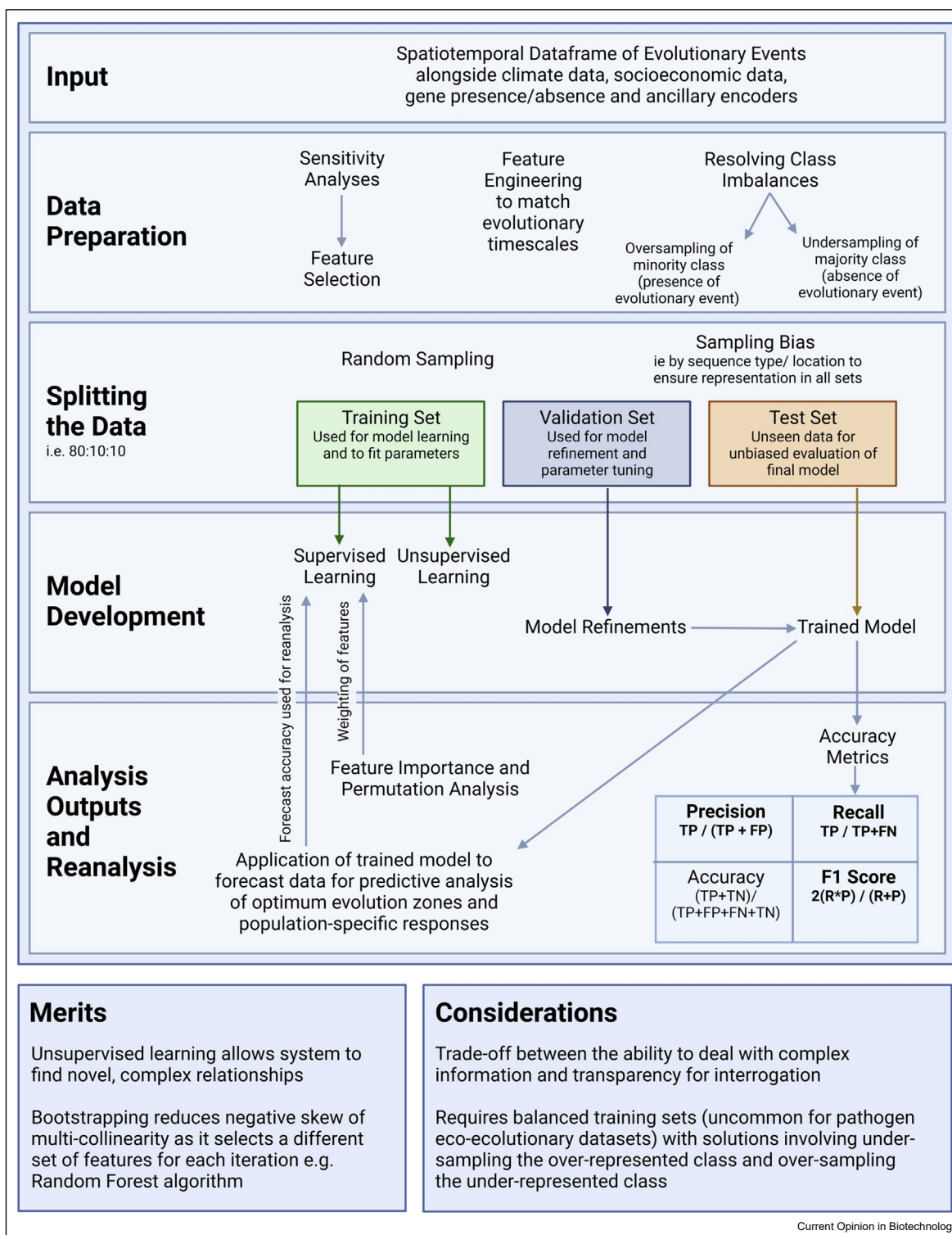
accommodating for heterogeneous and complex pathogen dynamics. ML has been utilised to forecast population-level heat tolerances [34–37], environmental drivers of adaption [38] and diffusion [33] in slower-evolving taxa, but it has not been fully exploited to elucidate the evolutionary responses of pathogens to climate change, likely due to the absence of a toolkit to combine genomic and climate datasets into a cohesive format for predictive applications [34], which our framework aims to solve.

Specifically, ML tools will be developed on the proposed novel spatiotemporal dataframe (Figure 2). The spatiotemporal dataframe will be split 80%-10%-10% (or similar) into a training dataset, validation dataset (for model refinement) and unseen test dataset (to draw accuracy metrics from), respectively. The ML model results can be interrogated to extract feature importance and infer the relative contribution of each climate variable to the final predictions. Trained and refined ML models can then be applied to long-term or real-time forecast data (or hypothetical future climate scenarios) to anticipate a range of evolutionary events — including the likelihood of mutations, gene transfer or presence of particular adaptive traits giving an indication on the future global ecology and evolution of a range of pathogens. Epidemiological and surveillance data will be key to validating the trained models, by using real-world outbreak data to give indications of forecast performance [30], and for reanalysis to provide continuous model refinement.

Looking forward: output opportunities and limitations

The outputs of the framework can then be used for various resulting applications, within academia or policy and decision-making — particularly the scientific results should be converted into user-friendly public health tools. The accuracy of these outputs is dependent on the input datasets, which will therefore benefit from increased data quality, availability and standardisation. It is critical to fund, develop and maintain active surveillance systems in the future, to provide an abundance of genomes from which to identify new epidemic clones and pinpoint when and how they emerge successfully from a population comprising many variants, with the recent Covid-19 pandemic giving an insight into what can be achieved with high-quality, near-real-time genomic sequencing to identify the location of emerging strains and consequent outbreaks [39,40]. Data are often scattered across a range of community-specific repositories or paper supplements [41,42] making interdisciplinary research difficult and requiring fit-for-purpose interdisciplinary databases for open data sharing and conciliation of results of climate-sensitive pathogens or other species [41].

Figure 2



Implementation of Machine Learning within the eco-evolutionary framework, highlighting merits and considerations of the analysis. (TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, R = recall, P = precision). Created with BioRender.com.

Concluding remarks

This novel eco-evolutionary framework targets knowledge gaps in the analysis of how climate change is driving climate-sensitive pathogens that exist in the environment, providing an option to overcome previous limitations. The integration of genomic analysis in the next generation of tools will enhance currently well-characterised ecological relationships with variant-specific evolutionary responses to climate change, to understand how strains emerge from genetically diverse populations, spread to new locations and become dominant. This review advocates for genomes to be accompanied by better spatiotemporal metadata to facilitate such analysis. ML applications have been explored as facilitators to provide forecasting capabilities of future pathogen evolution under climate change scenarios. These results require follow-through to public health tools and databases to increase our understanding and predictive capabilities of how pathogens and other species involved in disease transmission pathways (e.g. vectors and hosts), to ultimately reduce the human health burden posed by climate-sensitive pathogens.

Editorial disclosure statement

Given his role as Guest Editor, Jaime. Martinez Urtaza had no involvement in the peer review of the article and has no access to information regarding its peer review. Full responsibility for the editorial process of this article was delegated to Luis F. De León.

Data Availability

No data were used for the research described in the article.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the Natural Environmental Research Council [Grant number NE/S007210/1] and Centre for Environment, Fisheries and Aquaculture Science (Cefas) internal Seedcorn funding. J. Martinez-Urtaza is funded by grant PID2021-127107NB-I00 from Ministerio de Ciencia e Innovación (Spain).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Romanello M, et al.: **The 2021 report of the lancet countdown on health and climate change: code red for a healthy future.** *Lancet* 2021, **398**:1619-1662.
2. Barros VR, et al.: **Climate Change 2014 Impacts, Adaptation, and Vulnerability Part B: Regional Aspects: Working Group II Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.** Cambridge University Press; 2014:1-1820.
3. Chao DL, Roose A, Roh M, Kotloff KL, Proctor JL: **The seasonality of diarrheal pathogens: a retrospective study of seven sites over three years.** *PLoS Negl Trop Dis* 2019, **13**:e0007211, <https://doi.org/10.1371/journal.pntd.0007211>
4. Martinez-Urtaza J, et al.: **Epidemiological investigation of a foodborne outbreak in Spain associated with U.S. West coast genotypes of *Vibrio parahaemolyticus*.** *SpringerPlus* 2016, **5**:1-8, <https://doi.org/10.1186/s40064-016-1728-1>
5. Roy-Dufresne E, Logan T, Simon JA, Chmura GL, Millien V: **Poleward expansion of the white-footed mouse (*Peromyscus leucopus*) under climate change: implications for the spread of Lyme disease.** *PLoS One* 2013, **8**:e80724, <https://doi.org/10.1371/journal.pone.0080724>
6. Ogden NH, Ben Beard C, Ginsberg HS, Tsao JI: **Possible effects of climate change on ixodid ticks and the pathogens they transmit: predictions and observations.** *J Med Entomol* 2020, **58**:1536-1545, <https://doi.org/10.1093/jme/tjaa220>
7. Baker RE, et al.: **Infectious disease in an era of global change.** *Nat Rev Microbiol* (4) 2021, **20**:193-205, <https://doi.org/10.1038/s41579-021-00639-z>
8. Carlson CJ, et al.: **Climate change increases cross-species viral transmission risk.** *Nature* 2022, **607**:555-562, <https://doi.org/10.1038/s41586-022-04788-w>
9. Martinez-Urtaza J, et al.: **Genomic variation and evolution of *Vibrio parahaemolyticus* st36 over the course of a transcontinental epidemic expansion.** *Mbio* 2017, **8**:e01425-17, <https://doi.org/10.1128/mBio.01425-17>
10. Baker-Austin C, et al.: ***Vibrio* spp. infections.** *Nat Rev Dis Prim* 2018, **4**:8, <https://doi.org/10.1038/s41572-018-0005-8>
11. Waldvogel A-M, et al.: **Evolutionary genomics can improve prediction of species' responses to climate change.** *Evol Lett* 2020, **4**:4-18, <https://doi.org/10.1002/evl3.154>.
The authors discuss a range of genomic approaches that can and have been used to investigate how species will respond to climate change and improve prediction models. Includes a discussion on the key challenges for future eco-evolutionary models.
12. Storfer A, Patton A, Fraik AK: **Navigating the interface between landscape genetics and landscape genomics.** *Front Genet* 2018, **9**:68.
13. Aguirre-Liguori JA, Ramírez-Barahona S, Gaut BS: **The evolutionary genomics of species' responses to climate change.** *Nat Ecol Evol* (10) 2021, **5**:1350-1360, <https://doi.org/10.1038/s41559-021-01526-9>.
The authors provide a methodology (accompanied by source code) to incorporate local adaption into species distribution models for a wild plant species.
14. Stewart Ibarra A, Diaz A, Ryan S, Lippi C, Dunbar W: **Landscape mapping of software tools for climate-sensitive infectious disease modelling.** *Tech. Rep.*, Inter-American Institute for Global Change Research and Wellcome; 2022.
15. Fenderson LE, Kovach AI, Llamas B: **Spatiotemporal landscape genetics: investigating ecology and evolution through space and time.** *Mol Ecol* 2020, **29**:218-246.
The authors provide an overview and recommendations for analytical approaches that integrate genomic and environmental data for population structure, demography, phylogeography, metapopulation connectivity and adaptation.
16. Casadevall A, Kontoyiannis DP, Robert V: **On the emergence of *Candida auris*: climate change, azoles, swamps, and birds.** *MBio* 2019, **10**:e01397-19.
17. Deatherage DE, Kepner JL, Bennett AF, Lenski RE, Barrick JE: **Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures.** *Proc Natl Acad Sci USA* 2017, **114**:E1904-e1912, <https://doi.org/10.1073/pnas.1616132114>

18. Montanez I, et al.: **Analysis of *Vibrio harveyi* adaptation in sea water microcosms at elevated temperature provides insights into the putative mechanisms of its persistence and spread in the time of global warming.** *Sci Rep* 2019, **9**, <https://doi.org/10.1038/s41598-018-36483-0>
 19. Pörtner, H. O. et al.: **Climate Change 2022: Impacts, Adaptation and Vulnerability**, 3056, 2022.
 20. Kopprio GA, et al.: **Vibrio and bacterial communities across a pollution gradient in the Bay of Bengal: unraveling their biogeochemical drivers.** *Front Microbiol* 2020, **11**:594, <https://doi.org/10.3389/fmicb.2020.00594>
 21. Plummer S, Lecomte P, Doherty M: **The ESA Climate Change Initiative (CCI): a European contribution to the generation of the global climate observing system.** *Remote Sens Environ* 2017, **203**:2-8.
 22. Lemey P, et al.: **Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza h3n2.** *PLoS Pathog* 2014, **10**:e1003932, <https://doi.org/10.1371/journal.ppat.1003932>
 23. Bielejec F, et al.: **Spread3: interactive visualization of spatiotemporal history and trait evolutionary processes.** *Mol Biol Evol* 2016, **33**:2167-2169.
 24. Dellicour S, Rose R, Faria NR, Lemey P, Pybus OG: **Seraphim: studying environmental rasters and phylogenetically informed movements.** *Bioinformatics* 2016, **32**:3204-3206.
 25. Brockhurst MA, et al.: **The ecology and evolution of pangenomes.** *Curr Biol* 2019, **29**:20.
 26. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B: **Rdp4: detection and analysis of recombination patterns in virus genomes.** *Virus Evol* (1) 2015, **1**, <https://doi.org/10.1093/ve/vev003>
 27. Vaughan TG, et al.: **Inferring ancestral recombination graphs from bacterial genomic data.** *Genetics* 2017, **205**:857-870.
 28. Didelot X, Lawson D, Darling A, Falush D: **Inference of homologous recombination in bacteria using whole-genome sequences.** *Genetics* 2010, **186**:1435-1449.
 29. Didelot X, Wilson DJ: **Clonalframe: efficient inference of recombination in whole bacterial genomes.** *PLoS Comput Biol* 2015, **11**:e1004041, <https://doi.org/10.1371/journal.pcbi.1004041>
 30. Campbell AM, Racault M-F, Goult S, Laurenson A: **Cholera risk: a machine learning approach applied to essential climate variables.** *Int J Environ Res Public Health* 2020, **17**:9378.
 31. Lässig M, Mustonen V, Walczak AM: **Predicting evolution.** *Nat Ecol Evol* 2017, **1**:1-9.
 32. Tabachnick WJ: **Climate change and the arboviruses: lessons from the evolution of the dengue and yellow fever viruses.** *Annu Rev Virol* 2016, **3**:125-145, <https://doi.org/10.1146/annurev-virology-110615-035630>
 33. Jacquot M, Nomikou K, Palmarini M, Mertens P, Biek R: **Bluetongue virus spread in Europe is a consequence of climatic, landscape and vertebrate host factors as revealed by phylogeographic inference.** *Proc R Soc B: Biol Sci* 2017, **284**:20170919.
- The authors fitted phylogeographic models based on environmental variables to identify spatiotemporal drivers of diffusion bluetongue virus by hosts.
34. Cortés AJ, López-Hernández F, Osorio-Rodríguez D: **Predicting thermal adaptation by looking into populations' genomic past.** *Front Genet* 2020, **11**:564515, <https://doi.org/10.3389/fgene.2020.564515>.
- The authors discuss how recent thermal-driven selective responses can be inferred by coalescent methods and reconstructed using GWAS. Machine learning is then discussed for its potential to boost our predictive capabilities in this context, informing long-term predictions of thermal adaptive responses.
35. Garner J, et al.: **Genomic selection improves heat tolerance in dairy cattle.** *Sci Rep* 2016, **6**:1-9.
 36. Sukumaran S, Crossa J, Jarquin D, Lopes M, Reynolds MP: **Genomic prediction with pedigree and genotypic environment interaction in spring wheat grown in South and West Asia, North Africa, and Mexico.** *G3: Genes, Genomes, Genet* 2017, **7**:481-495.
 37. Juliana P, et al.: **Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat.** *Theor Appl Genet* 2019, **132**:177-194.
 38. Layton KKS, et al.: **Genomic evidence of past and future climate-linked loss in a migratory arctic fish.** *Nat Clim Chang* 2021, **11**:158-165, <https://doi.org/10.1038/s41558-020-00959-7>.
- The authors retrospectively identify past climate-associated declines in a migratory fish population using effective population size estimates and associate these with decreases in temperature. Genomic diversity and machine learning are used to identify vulnerable populations at the edges of optimum temperature ranges.
39. Akther S, et al.: **CoV genome tracker: tracing genomic footprints of Covid-19 pandemic.** *bioRxiv* 2020, <https://doi.org/10.1101/2020.04.10.036343> 2020-04.
 40. Forster P, Forster L, Renfrew C, Forster M: **Phylogenetic network analysis of SARS-CoV-2 genomes.** *Proc Natl Acad Sci* 2020, **117**:9241-9243.
 41. Waldvogel A-M, Schreiber D, Pfenninger M, Feldmeyer B: **Climate change genomics calls for standardized data reporting.** *Front Ecol Evol* 2020, **8**:242, <https://doi.org/10.3389/fevo.2020.00242>
 42. Culina A, et al.: **Navigating the unfolding open data landscape in ecology and evolution.** *Nat Ecol Evol* 2018, **2**:420-426, <https://doi.org/10.1038/s41559-017-0458-2>
 43. Exposito-Alonso M, et al.: **Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*.** *Nat Ecol Evol* 2018, **2**:352-358, <https://doi.org/10.1038/s41559-017-0423-0>
 44. Bay RA, et al.: **Genomic signals of selection predict climate-driven population declines in a migratory bird.** *Science* 2018, **359**:83-86, <https://doi.org/10.1126/science.aan4380>.
- The authors identified genes that were indicative of successful climate adaption in the yellow warbler bird, based on their associations with migratory behaviour. They also found populations with the greatest shifts in allele frequencies were declining, suggesting this was due to a failure to adapt effectively.
45. Fitzpatrick MC, Keller SR: **Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation.** *Ecol Lett* 2015, **18**:1-16, <https://doi.org/10.1111/ele.12376>.
- An example of a hybrid genomic species distribution study, using a tree species as an exemplar. The authors find a particular gene that is associated with local adaption to temperature.
46. Elith J, Leathwick JR: **Species distribution models: ecological explanation and prediction across space and time.** *Annu Rev Ecol Syst* 2009, **40**:677-697.
 47. Kearney M, Porter WP, Williams C, Ritchie S, Hoffmann AA: **Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia.** *Funct Ecol* 2009, **23**:528-538.
 48. Reside AE, Critchell K, Crayn DM, Goosem M, Goosem S, Hoskin CJ, Sydes T, Vanderduys EP, Pressey RL: **Beyond the model: expert knowledge improves predictions of species' fates under climate change.** *Ecol Appl* 2019, **29**:e01824, <https://doi.org/10.1002/eap.1824>
 49. Lowry DB, Lovell JT, Zhang L, Bonnette J, Fay PA, Mitchell RB, et al.: **QTL x environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient.** *Proc Natl Acad Sci* 2019, **116**:12933-12941.
 50. Jenouvrier S, Che-Castaldo J, Wolf S, Holland M, Labrousse S, LaRue M, Wienecke B, Fretwell P, Barbraud C, Greenwald N, Stroeve J, Trathan PN: **The call of the emperor penguin: legal**

- responses to species threatened by climate change. *Glob Change Biol* 2021, **27**:5008-5029, <https://doi.org/10.1111/gcb.15806>
51. Arnold BJ, Huang IT, Hanage WP: **Horizontal gene transfer and adaptive evolution in bacteria.** *Nat Rev Microbiol* 2022, **20**:206-218, <https://doi.org/10.1038/s41579-021-00650-4>
 52. Hadfield J, et al.: **Nextstrain: real-time tracking of pathogen evolution.** *Bioinformatics* 2018, **34**:4121-4123.
 53. Argimón S, et al.: **Microreact: visualizing and sharing data for genomic epidemiology and phylogeography.** *Microb Genom* 2016, **2**:e000093.
 54. Parks DH, et al.: **Gengis 2: geospatial analysis of traditional and genetic biodiversity, with new gradient algorithms and an extensible plugin framework.** *PLoS One* 2013, **8**:e69885.
 55. Günther T, Coop G: **Robust identification of local adaptation from allele frequencies.** *Genetics* 2013, **195**:205-220, <https://doi.org/10.1534/genetics.113.152462>