JASs Journal of Anthropological Sciences
Open Access
forum Vol. 103 (2025), pp. 213 - 216
doi. 10.4436/jass10315

Reimagining human population genetics in the age of Artificial Intelligence

Alex Mas-Sandoval[1] & Matteo Fumagalli[2,3]

1) Laboratorio de Evolución Humana, Universidad de Burgos, España
e-mail: amas@ubu.es

2) School of Biological and Behavioural Sciences, Queen Mary University of London, UK

3) The Alan Turing Institute, London, UK

Population genetics at a methodological crossroads

Population genetics is the discipline that studies genetic variation within and between populations and changes in genetic diversity over time, providing a fundamental framework for understanding and modelling evolutionary processes. Based on mathematically tractable models, it offers valuable insights into how mutation, natural selection, genetic drift, and migration shape genetic variation. However, these models are based on simplifying assumptions that often overlook the complexities of real populations. Although they enable powerful analytical tools, population genetic models also restrict the types of data that can be used and the scope of questions that can be answered.

In recent years, deep learning, a branch of artificial intelligence (AI), has reshaped how population geneticists analyse data, generate hypotheses, and model evolutionary processes (Korfmann et al. 2023). Unlike traditional model-based approaches, deep learning enables data-driven methods capable of capturing complex patterns in large genomic datasets with flexible assumptions. At the same time, the shift to data-driven methods facilitates the integration of diverse data types, such as environmental and phenotypic information, with genomic data. However, it also raises important ethical and political questions about how these tools are developed, shared, interpreted, and applied.

In this commentary, we highlight how AI is transforming human population genetics, focussing on methodological innovations and ethical challenges. We argue that AI should drive a broader rethinking of the field beyond the limits of classical models and should enable a shift toward a more comprehensive and critical approach to population genetics and evolutionary biology.

Data-driven inference to break the Wright-Fisher mould

The Wright-Fisher model, named after Sewall G. Wright and Ronald A. Fisher, has shaped population genetics for nearly a century. Among others, Wright and Fisher established the foundations of modern evolutionary synthesis and provided a mathematical framework for population genetics. However, they were also involved with eugenicist ideas, with Fisher being particularly explicit in his writings and correspondence (Bodmer et al. 2021). Wright-Fisher models, along with their extensions, allow predictions of allele frequency dynamics and genetic variation over time under core assumptions: constant population size, random mating and absence of population structure, and neutrality. While analytically convenient, these assumptions rarely hold in real populations and particularly in non-European populations, as they reflect an implicit Eurocentric framing of population structure.

The simplifications of these models enable the study of large-scale empirical data by monitoring a limited set of summary statistics. However, reducing the multidimensionality of genomic

data to a few parameters also entails substantial information loss. Deep learning, on the contrary, allows neural networks to be trained on raw genomic data and abstract features. By directly learning the mapping between high-dimensional data and the model parameters, it can model far more complex demographic and evolutionary scenarios. At the same time, a limitation of deep learning is the need for sufficient training data, usually generated through computational simulations. Flexible and efficient simulation tools such as SLiM (Haller and Messer, 2019) now make it feasible to produce synthetic training data under scenarios that include natural selection, non-constant population sizes, admixture, population structure, non-random mating, and cultural dynamics. Following this approach, deep learning has already been successfully applied to identify regions under selection, infer demographic history, and classify complex evolutionary scenarios (Sanchez et al. 2021).

Most of the algorithms successfully deployed to study population genetic data are based on convolutional neural networks. These implementations treat an alignment of genotypes as an image and slide small pattern detectors across that image to learn its local haplotype structure, blocks of shared ancestry, recombination breakpoints, introgressed tracts, or signals of natural selection. Because convolutional neural networks use the full alignment rather than predefined summary statistics, they can learn useful features directly from the data and often exceed the accuracy of established statistical methods (Flagel et al. 2019). Other AI technologies recently adopted in the field include generative models, graph neural networks, and large language models. Generative models are of particular interest due to their joint ability to synthesise new data and infer evolutionary parameters (Yelmen and Jay 2023).

Another promising approach is multimodal AI, which allows for the integration of genetic data with diverse data types, including both other omics data (e.g. epigenetic markers, 3D genome architecture, transcriptomic profiles) and multiple sources of information (e.g. environmental variables, historical records, linguistic data, archaeological findings). Advances in multimodal AI enable the simultaneous analysis of these heterogeneous datasets, moving beyond state-of-the-art approaches by contextualising genetic variation within broad biological and social frameworks.

These methodological shifts are especially relevant for human populations, where genetic variation reflects not only migration and natural selection, but also the imprint of social boundaries and stratification. Such processes can generate assortative mating and sex-biased demographic contributions, leaving signatures often overlooked by classical models but now detectable through deep-learning approaches (Mas-Sandoval et al. 2023). Therefore, machine learning can be used to embed social patterns in the inferential pipeline, making it possible to test whether the observed genetic variation reflects structured rules of mating and migration grounded in historical and social processes.

Creating space for conceptual innovation

The integration of machine learning into population genetics not only is transforming the types of analysis that can be performed, it is also reshaping how researchers spend their time. Tasks that once demanded extensive manual effort, such as data formatting, feature engineering, and model parameterisation, can now be partially automated. Then, by alleviating routine computational demands, AI enables scientists to dedicate more time and creative energy to developing innovative ideas and synthesising them into new theoretical models.

The reduction of procedural burden creates space for deeper conceptual and exploratory work. This expanded cognitive space fosters greater creativity in evolutionary biology, providing an opportunity to rethink foundational concepts and generate novel hypotheses. Foundational contributions —from natural selection and modern synthesis to neutral theory, endosymbiotic theory or the discovery of archaic

admixture through ancient DNA, among others— have profoundly reshaped our understanding of evolutionary processes and human history. However, the field has lately focused on pursuing methodological advances to achieve more precise inference within established frameworks, often at the expense of theoretical innovation. Creating space for conceptual creativity may generate insights that better capture the dynamic interplay between genetics, culture, and environment, advancing a more integrated understanding of human evolution within its complex contexts.

Artificial intelligence in an inegalitarian world

The integration of machine learning into population genetics presents not only methodological opportunities but also significant ethical and political challenges. Machine learning models are shaped by the data with which they are trained. In genomics, this raises specific concerns. When datasets are unbalanced or historically biased, such as those dominated by individuals of European descent, models may yield inaccurate or misleading inferences for under-represented populations. In applications such as medical genetics, forensic identification, or genetic ancestry inference, such misclassifications can have material consequences and may reinforce existing social and structural inequalities (Chen et al. 2023).

These concerns are particularly relevant in human population genetics, a field that addresses questions of identity, ancestry, and diversity. The selection of research questions, the choice of populations to sample, and the interpretation of results are never neutral decisions. They are shaped by institutional priorities, funding structures, and cultural biases of the researcher, all of which influence which genomes are studied and the way variation is explained. As machine learning becomes increasingly central to the field, it is essential to consider how computational tools may amplify or obscure specific narratives about human history.

A further dimension of inequality arises from the computational and infrastructural demands of deep learning. Although AI may seem to reduce certain barriers that scientists face, such as programming expertise or English proficiency, training large-scale models still requires high-performance computing, substantial energy demands, and specialized technical skills, resources that remain unevenly distributed across global research institutions. This asymmetry creates barriers for researchers working in regions with limited infrastructure and can restrict the range of perspectives and research agendas that inform the field. Promoting equity in the application of AI to population genetics will require investment in open-source tools, scalable, and low-resource algorithms, and data governance frameworks that support fair and inclusive scientific collaboration (Armenteras 2021).

At the same time, advances in AI allow the integration of multiple forms of data to address multilayered questions about the interaction between genetic structure and social, ecological, or cultural variables. However, the interpretation of such models requires conceptual frameworks that extend beyond statistical or computational inference. Although machine learning can identify associations across heterogeneous datasets, it does not assess their causal significance, historical origins, or ethical implications. These tasks remain the responsibility of researchers and must be informed by interdisciplinary knowledge and critical analysis.

Through the metaphor of the cyborg—a figure that blurs the boundaries between human and machine, nature, and culture—Donna Haraway anticipated that technologies are not neutral tools, but hybrid constructions in which social assumptions are embedded (Haraway, 1985). Since knowledge is always partial and situated within the cultural and political contexts in which it is produced, algorithms cannot be regarded as autonomous actors. Instead, they must be interpreted, directed, and held accountable by the communities that design and deploy them. Situating hypotheses within epistemological diversity and acknowledging the frameworks through which genetic data are analysed can clarify the assumptions embedded in

computational models. Moreover, because human population genetics inevitably engages with questions of identity, belonging, and ancestry, adopting an explicit decolonial approach is essential to confront the biases inherited from the troubled past of the field.

Toward cutting-edge, creative and committed research

AI is transforming population genetics, expanding the boundaries of the questions we ask ourselves about evolutionary processes. It enables integration across biological, cultural, and environmental dimensions and creates space for creativity and new theoretical horizons.

These opportunities carry clear responsibilities. The biases in the datasets, the environmental costs of the computation, and the default cultural assumptions embedded in research questions must be explicitly addressed. In response, we advocate for human population genetics that is technically innovative, intellectually expansive, and ethically grounded. Only by meeting these standards can AI contribute to cutting-edge, creative, and committed research.

References

Armenteras D (2021) Guidelines for healthy global scientific collaborations. Nat Ecol Evol 5: 1193–1194. https://doi.org/10.1038/s41559-021-01496-y

Bodmer W, Bailey RA, Charlesworth B, et al (2021) The outstanding scientist, R.A. Fisher: his views on eugenics and race. Heredity 126:565–576. https://doi.org/10.1038/s41437-020-00394-6

Chen RJ, Wang JJ, Williamson DFK, et al (2023) Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng 7:719–742. https://doi.org/10.1038/s41551-023-01056-8

Flagel L, Brandvain Y, Schrider DR (2019) The unreasonable effectiveness of convolutional neural networks in population genetic inference. Mol Biol Evol 36:220–238. https://doi.org/10.1093/molbev/msy224

Haller BC, Messer PW (2019) SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. Mol Biol Evol 36:632–637. https://doi.org/10.1093/molbev/msy228

Haraway DJ (1985) A manifesto for cyborgs: science, technology, and socialist-feminism in the 1980s. Social Rev 80:65-107.

Korfmann K, Gaggiotti OE, Fumagalli M (2023) Deep learning in population genetics. Genome Biol Evol 15:evad008. https://doi.org/10.1093/gbe/evad008

Mas-Sandoval A, Mathieson S, Fumagalli M (2023) The genomic footprint of social stratification in admixing American populations. eLife:e84429. https://doi.org/10.7554/eLife.84429

Sanchez T, Cury J, Charpiat G, et al (2021) Deep learning for population size history inference: Design, comparison and combination with approximate Bayesian computation. Mol Ecol Resour 21: 2645–2660. https://doi.org/10.1111/1755-0998.13224

Yelmen B, Jay F (2023) An overview of deep generative models in functional and evolutionary genomics. Annu Rev Biomed Data Sci 6:173–189. https://doi.org/10.1146/annurev-biodatasci-020722-115651