Roots of misunderstanding

Aylwyn Scally

Zusammenfassung

Wurzeln des Missverständnisses

Ich erörtere den interdisziplinären Dialog zwischen Genetik und Geschichte, der sich auf das Gebiet der Paläogenomik konzentriert. Hierbei betrachte ich einige der Faktoren, die zur Fehlkommunikation zwischen diesen Bereichen beitragen, und wie Forscher auf die damit verbundenen Herausforderungen reagieren können, einschließlich der Unterschiede in der Veröffentlichungspraxis, der DNA-Genealogie-Industrie und dem allgemeinen Verständnis von Abstammung.

Summary

I discuss the interdisciplinary dialogue between genetics and history, centred on the field of paleogenomics. I consider some of the factors contributing to miscommunication between these fields and how researchers can respond to the challenges they pose, including differences in publication practices, the DNA genealogy industry, and the public understanding of ancestry.

Introduction[1]

Genomes carry information about the ancestry and demography of past individuals, and are thus a source of historical evidence. In principle, this evidence is now more accessible than ever before, thanks to the development of whole-genome sequencing. However, due to a combination of demographic and methodological factors, most analyses of present-day genetic data are informative about events on a timescale of thousands rather than hundreds of years, and so until recently there has been a relative absence of genetic information on historical timescales (as opposed to prehistoric or earlier periods). There are exceptions, such as methods focusing on very rare genetic variants, which are more likely to have arisen recently (Schiffels et al. 2016), but such studies by their nature require large numbers of samples. And in many cases, there have been considerable demographic and environmental changes, including large-scale migration within and from outside the region studied, particularly during and since the medieval period. Thus, while ancestral geographic correlations do remain in present-day genetic data (Novembre et al. 2008; Leslie et al. 2015), various factors make it harder to use such data to discern subtle past effects or relationships.

The ability to sequence ancient and historical DNA has altered this picture dramatically, producing a growing number of samples which record the genetic ancestry and relatedness of past individuals from different locations and walks of life. Importantly (and barring one or two exceptions, such as King et al. 2014), it is not the ancestries themselves that are of interest so much as their association with contextual factors. Combined with additional sources of archaeological and written evidence, the potential for multidisciplinary studies is enormous, particularly on regional and local scales rather than the continental scale of many initial ancient DNA studies (Hui et al. 2023). Indeed, such interdisciplinarity is essential, for as with any evidence, genetic data alone provide a partial view, affected by biases and gaps that depend on context and time. Relating them to questions of historical or archaeological interest requires conjecture and interpretation, which change as further evidence, genetic or otherwise, becomes available.

As was the case with other new technologies in the past, the power and potential of ancient DNA has seen it rapidly established as an important archaeological tool. And as archaeologists become proficient in its methods, so too historians of the early medieval period, who make frequent use of archaeological evidence, are engaging with genetic data. Equally, as geneticists develop methods focused on historical timescales and data, research collaborations spanning genetics, archaeology and history are growing in number.

However, such interactions face several obstacles. Prior to the advent of ancient DNA, some of the frustrations experienced by archaeologists were captured, only half jokingly, in the remark that when geneticists presented their findings, they reported things that either were already known or were obviously wrong. Few researchers are as familiar as they should be with the relevant literature of other fields, but sometimes even the self-awareness of this is lacking. There remains a tendency for geneticists, particularly when communicating research to the public, to reach for outdated historical narratives and cherry-picked examples. On the other

1 Acknowledgements: John Robb, Lara Cassidy and Pontus Skoglund for useful discussions and comments.

hand, frustrations can also get entrenched and become an obstacle in themselves. Were we to formulate some guidelines for interdisciplinary research, one of them might be that if something seems ›obviously wrong‹, the misunderstanding might well be yours!

A frequent misunderstanding arises from the differences in how models are perceived and constructed. In genetics, particularly population genetics, models are deliberately abstracted and simplified as a means to interpret data quantitatively. The truth, whatever it may be, is understood to be more complex than could be represented in a useful model. Moreover, sometimes a paper may use simple models not because the authors are oblivious to the complexities of the problem but because the size of initial datasets, or novel methods, do not yet support inference under more complex models. Simple models are useful, often because insight comes from comparing how they respond to data, rather than from inferring the specific values of their parameters. Genetic population models, for example, are typically parameterised by a single ›population size‹ parameter, which ostensibly represents the number of individuals in the population, but when inferred from real data is also influenced by, and thus potentially informative about, other factors such as kinship, population structure and natural selection. Similarly, the common ancestry of multiple populations may be modeled by a simple tree, as if their separation originated in a series of clean and instantaneous splits; but inference under such a model can be a way to reveal the underlying complexities of gradual separation and cross-population gene flow. Some archaeologists are familiar with such approaches, but to others they can come across as naïve or even deliberate reductionism, and it is not uncommon to find exasperated criticism of genetic analyses along these lines.

These and other issues, differences in technical language and jargon, or the ways in which research questions are posed and addressed, are typical of the challenges in any interdisciplinary exchange. But there are also particular factors that exacerbate the problems of dialogue between these fields. Here I will touch on two of them: differences in scholarly publication practices, and the impact of commercial DNA ancestry testing.

Publication practices

The first ancient human whole genome sequence, from a Neanderthal individual who lived more than 38 000 years ago, was published in 2010 (Green et al. 2010) and was a landmark scientific achievement. Since then, the rapidly increasing number of ancient sequences and their transformative potential for understanding human prehistory have continued to attract public and media interest. Studies of ancient DNA and other developments in the new field of paleogenomics have appeared frequently in journals such as *Nature* and *Science*. Such attention may be receding as the field matures, but high-profile publication remains a realistic goal for many practitioners, particularly when presenting data from previously unstudied regions or time periods, or applying it to a long-standing question for the first time.

This experience has shaped the expectations and publishing practices of the field, in ways which contribute to miscommunication with other researchers, particularly non-scientists. Many features of high-profile journal publication have long been recognised as problematic: compressed papers, which serve primarily as an abstract for a lengthy and technical supplement; limits on the numbers of references and figures; and above all, the pressure to present research and its conclusions as groundbreaking, which is further heightened when communicating results to the press and media. Sometimes the fanfare associated with publication gives the impression, particularly outside the field, that the authors intend this to be a definitive and even final statement on the subject, when in fact it is merely a first examination. And, unfortunately, it is often what is said in the media rather than in the paper itself (the tone of which may be far more circumspect) that colleagues react to. Concerns about public misunderstanding make this fraught, particularly in the context of misrepresentation by racists, ethno-nationalists and pseudoscientists.

It is important to note that many paleogenomic studies are co-authored by both geneticists and archaeologists, and increasingly also by historians. One might imagine this would mitigate many of the problems of cross-disciplinary communication. Unfortunately, in practice it can lead to further problems, particularly where data suggest a new hypothesis or seem to favour one over another. Sometimes such collaborations are lopsided, with archaeologists regarded merely as a source of samples, or geneticists as quantitative support for an existing position. Even where, within a collaboration, there is mutual understanding of cross-disciplinary methods and aims, communicating its findings to a diverse audience is a different challenge. Papers often end up combining genetic content, which may be technical and indecipherable by non-experts, with a fully-formed archaeological or historical synthesis. The intention may be to help readers understand the implications of genetic analyses in historical or archaeological terms; but the result is often to present an argument whose basis and assumptions are impenetrable to its intended audience, yet which nevertheless purports to favour, with the authority of science, one side of a contentious debate over another.

Differences in publishing timescales, and the fact that paleogenomics is a new and rapidly changing field, also foster miscommunication. For historians, it is not uncommon to engage with arguments published decades previously if they remain current and still shape viewpoints or teaching on a topic. By contrast, the genetic approaches of two decades ago are very often obsolete. Genomic data from mitochondrial DNA and chromosome Y, for example, are still informative and useful in some contexts, but the simplistic interpretations they once supported are often not. It can come as a surprise to geneticists when a contemporary critique of approaches in genetics cites and debates such publications. But while they may be ignored by geneticists now, outside the field they appear as part of its relatively recent literature. Furthermore, some older publications are cited frequently by pseudo-scientists and pseudo-historians, thereby sustaining their influence on popular understanding and forcing historians to deal with this legacy of earlier contributions. The long half-life this gives to misleading ideas is not sufficiently appreciated within genetics.

The DNA ancestry industry

Ancestry occupies a complex and often ambiguous place in society and culture, but also a very prominent one. Popular enthusiasm for genetic genealogy has fuelled the enormous growth of the consumer DNA-testing industry over the last decade, with tens of millions of customers worldwide (Regalado 2019). The reports these companies provide to their customers, identifying distant cousins or assigning percentages of ancestry in various geographical and cultural categories, are now a significant part of how many people think about ancestry.

There are important differences between genetic ancestry, which describes the inheritance of DNA itself, and genealogical ancestry, represented by an individual's family tree or pedigree (Mathieson/Scally 2020), but methods based on genetic data necessarily rely on genetic ancestry as a marker of genealogical ancestry. Thus, when consumer genetics companies identify cousin relationships between participants, this is based on long stretches of identical genome sequence representing recent shared ancestry, and when they report percentages of ancestry, these are based on genetic similarity to groups of participants in their databases.

One can certainly question how well such reports are understood by customers. Arguably, the larger companies are relatively open about the methodologies used and interpretations given. But the marketing of these services is another matter. Advertisements vary in tone, from the frivolous, with ancestry as something relatively inconsequential and fun – the way many people view horoscopes, say – to the more serious, implying a deeper and more significant relationship between genomes and human traits. Customers are invited to plan holidays and diets, choose sports teams and construct aspects of their identity based on ancestry. Above all, they are exhorted to discover their roots.

The reception of these messages varies for different audiences, and they are targeted accordingly. For example, in societies or communities impacted by the legacy of colonialism or transatlantic slavery, the marketing reflects the fact that connections between ancestry, genetics and identity can be a highly consequential matter (Abel/Schroeder 2020). But in each case the messages play into and reinforce enduring misconceptions about ancestry. In the 21st century we may no longer aspire to trace our genealogies, like medieval princes, to mythological figures such as Noah and Adam, but the appetite for stories of personal descent from kings and emperors, or from Vikings, Celts and other historical peoples, remains undiminished. In this way, the popular concept of ancestry combines two notions which, in reality, are somewhat in opposition: the notion of origins and roots, and the uniqueness of an individual's identity.

There is of course a sense, or more accurately a time interval, in which everyone's ancestry is indeed unique. No two people have exactly the same pedigree, unless they are full siblings[2]. Differences in recent ancestry are thus literally familiar; one person may have an Irish grandparent, say, while another may have Italian ancestors a few generations further back. Intuitively, this ancestry is projected to older timescales, the first person claiming descent from the Celts perhaps, and the other from the Romans. But the intuition is misleading. The arithmetic of genealogical ancestry, in which the number of ancestors roughly doubles every generation back in time, means that inevitably there is a point in time when the number of ancestors, for any individual today, exceeds the total population alive at that time. This is the *genetic isopoint* (although as defined it is a feature of genealogical rather than genetic ancestry), and prior to it, all our pedigrees coincide.

Models of demography and global gene flow place the isopoint for present-day humans surprisingly recently, perhaps only a few thousand years ago (Rohde et al. 2004). In other words, for the vast majority of the human past, genealogical ancestry is a shared story rather than something which differentiates people. Moreover, on regional rather than global scales, pedigrees merge even more recently; for example, within Europe this occurs only seven or eight centuries ago, meaning that everyone with European ancestry is descended genealogically from Charlemagne, the Vikings, the Romans and so on. More importantly perhaps, simplistic claims about the present-day coherence of ancient or ›indigenous‹ European ancestries, which beset the popular reception of historical scholarship on the early medieval period, are undermined. For while indigeneity is applicable in regions and societies impacted by colonisation in the last few centuries, it is far less so in relation to events much further in the past. These observations are further reinforced by evidence from ancient and historical DNA, suggesting that gene flow and migration have been ubiquitous and, in some cases, very substantial in many parts of the world (Olalde et al. 2018; Skoglund/Mathieson 2018). Thus, genetic evidence is not primarily about the genetic origins of present-day nations, nor typically about the lineages of royal families and warrior clans.

Among geneticists there is concern that the marketing and presentation of consumer genetics services contributes to public misconceptions. However most take it as obvious that these services have little relevance to current research and thinking in human genetics. They are unaware of the extent to which it has come to represent their field, not only in the public mind, but also for many researchers in other disciplines, including history, anthropology and the social sciences. Critiques of human genetics from these fields frequently reference problems or misrepresentations associated with DNA ancestry tests and their reception (Hakenbeck 2019; Abel 2022).

One consequence is that an obsession with ancestry, categorisation and the simplistic representation of human diversity is projected back onto genetics as a field. And of course, geneticists do make use of categories and group labels, for example as descriptors of data (many of which come in turn from governmental or public health collection practices), or to facilitate statistical inference and modelling (Birney et al. 2021). Very rarely, or never, do such categories map straightforwardly onto the complex structure of human

2 And even siblings, if they are not twins, differ in genetic ancestry due to variation in the inheritance of chromosomal segments from their parents.

genetic diversity, and the simplistic ideas of human evolution and diversity in which they do are long outdated. But the terminology of a method often outlasts its use, and once disseminated into other contexts, can rejuvenate the old ideas from which it derives. Terms like admixture, population and ancestry, whose meanings carry different implications and broader associations outside genetics, are now part of the public conversation about human genetic diversity, thanks in part to their use by the DNA ancestry industry.

Conclusion

It can be expected that many of the difficulties discussed here will recede as the field of paleogenomics matures, as researchers are trained in its methods and become more familiar with the literature of multiple fields, and as studies move from high-profile journals to focus on more detailed, regional or local cohorts. Indeed, for the dialogue between genetics and archaeology, where quantitative and scientific methods are well established, much of this is already well underway, with research moving from population-level concerns to those of kinship and social relations. For historians it will require a greater appreciation and understanding of quantitative modelling and inference-based approaches to evidence, and the ways in which they are discussed and presented.

But, as the issues around DNA ancestry testing demonstrate, a substantial part of the dialogue between these fields is mediated through their public engagement (or lack thereof) and the impact it has on the reception of scholarship on historical and archaeological questions. For geneticists in particular there are several implications. It requires a better appreciation of the fact that, while ideas of biological race and arguments based on them may no longer be explicit in the language and theory of population genetics, they remain very much present in public discourse worldwide. This in turn requires them to think more carefully about – and change if necessary – the terminology used, not only in outreach but also in research itself, from which figures and results are frequently drawn and misrepresented (Carlson/Harris 2020). Above all, it necessitates a more concerted effort to anticipate and counter misconceptions about human genetics and ancestry, and to challenge the misrepresentations from which they stem.

Bibliography

Abel 2022
S. Abel, Permanent Markers. Race, Ancestry, and the Body after the Genome (Chapel Hill, NC 2022).

Abel/Schroeder 2020
S. Abel/H. Schroeder, From Country Marks to DNA Markers: The Genomic Turn in the Reconstruction of African Identities. Current Anthr. 61,22, 2020, 198–209, doi:10.1086/709550.

Birney et al. 2021
E. Birney/M. Inouye/J. Raff/A. Rutherford/A. Scally, The language of race, ethnicity, and ancestry in human genetic research. arXiv preprint 2106,10041, 2021, doi:10.48550/arXiv.2106.10041.

Carlson/Harris 2020
J. Carlson/K. Harris, Quantifying and contextualizing the impact of bioRxiv preprints through automated social media audience segmentation. PLOS Biol. 18,9, 2020, e3000860, doi:10.1371/journal.pbio.3000860.

Green et al. 2010
R. E. Green/J. Krause/A. W. Briggs/T. Maricic/U. Stenzel et al., A draft sequence of the Neandertal genome. Science 328,5979, 2010, 710–722, doi:10.1126/science.1188021.

Hakenbeck 2019
S. E. Hakenbeck, Genetics, archaeology and the far right: an unholy Trinity. World Arch. 51,4, 2019, 517–527, doi:10.1080/00438243.2019.1617189.

Hui et al. 2023
R. Hui/C. L. Scheib/E. D'Atanasio/S. A. Inskip/C. Cessford et al., Medieval social landscape through the genetic history of Cambridgeshire before and after the Black Death. bioRxiv 531048, 2023, doi:10.1101/2023.03.03.531048.

King et al. 2014
T. King/G. Fortes/P. Balaresque/M. G. Thomas/D. Balding et al., Identification of the remains of King Richard III. Nature Commun. 5,5631, 2014, doi:10.1038/ncomms6631.

Leslie et al. 2015
S. Leslie/B. Winney/G. Hellenthal et al., The fine-scale genetic structure of the British population. Nature 519, 2015, 309–314, doi:10.1038/nature14230.

Mathieson/Scally 2020
I. Mathieson/A. Scally, What is ancestry? PLOS Genetics 16,3, 2020, e1008624, doi: 10.1371/journal.pgen.1008624.

Novembre et al. 2008
J. Novembre/T. Johnson/K. Bryc/Z. Kutalik/A. R. Boyko et al., Genes mirror geography within Europe. Nature 456, 2008, 98–101, doi:10.1038/nature07331.

Olalde et al. 2018
I. Olalde/S. Brace/M. E. Allentoft/I. Armit/K. Kristianson et al., The Beaker phenomenon and the genomic transformation of northwest Europe. Nature 555, 2018, 190–196, doi:10.1038/nature25738.

Regalado 2019
A. Regalado, More than 26 million people have taken an at-home ancestry test. MIT Technology Rev., 2019, <https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/> (01.06.2023).

Rohde et al. 2004
D. Rohde/S. Olson/J. Chang, Modelling the recent common ancestry of all living humans. Nature 431, 2004, 562–566, doi:10.1038/nature02842.

Schiffels et al. 2016
S. Schiffels/W. Haak/P. Paajanen/B. Llamas/E. Popescu et al., Iron age and Anglo-Saxon genomes from East England reveal British migration history. Nature Commun. 7,10408, 2016, doi:10.1038/ncomms10408.

Skoglund/Mathieson 2018
P. Skoglund/I. Mathieson, Ancient genomics of modern humans: the first decade. Annu. Rev. Genomics and Human Genetics 19, 2018, 381–404, 2018, doi:10.1146/annurev-genom-083117-021749.

Address

Dr. Aylwyn Scally
University of Cambridge
Department of Genetics
Downing Street
Cambridge CB2 3EH
United Kingdom
aos21@cam.ac.uk
ORCID: https://orcid.org/0000-0002-0807-1167