

**JAMES LOWE, MIGUEL GARCÍA-SANCHO, RHODRI LENG,  
MARK WONG, NIKI VERMEULEN AND GIL VIRY\***

## **Across and within Networks: Thickening the History of Genomics**

---

### **ABSTRACT**

In this essay, we reflect on how the findings of the preceding papers enabled us to *thicken* the history of genomics. We have expanded the number of dimensions across which our historical work operated beyond extending the dimension of time. Building on this, we argue that the history of genomics became synchronically entangled with a range of communities, target species, and research agendas—among them yeast biochemistry, pig and human immunology, systematics, medical genetics, and agricultural genetics. We make sense of these entanglements with analytic categories to characterize modes of organizing and conducting sequencing, and the relationships between the practices of sequencing and the objectives of those collaborating around it: horizontal and vertical, proximate and distal, directed and undirected, as well as intensive and extensive sequencing. Our categories emerged as we analyzed and qualitatively interpreted datasets and co-authorship networks. Throughout this special issue, we have characterized genomics as a set of tools that open up connections between actors, institutions, experimental organisms, and historically contingent forms of research. We contend that presenting genomics in this way emphasizes the agency of the communities that mobilized the sequence data and offers a fresh perspective for addressing the medical and agricultural translation of that data. We close by proposing how we can develop our mixed-methods approach through the establishment of a domain ontology that would allow information on sequence submissions and publications to be connected to other forms of data, thus expanding

\*James Lowe, Science, Technology and Innovation Studies, University of Edinburgh, Old Surgeons' Hall, High School Yards, Edinburgh, EH1 1LZ, United Kingdom. [jwelowe1985@gmail.com](mailto:jwelowe1985@gmail.com)

The following abbreviations are used: DNA, deoxyribonucleic acid; HGP, Human Genome Project; INRA, Institut National de la Recherche Agronomique; IMPRH, INRA-Minnesota porcine Radiation Hybrid panel; PiGMap, Pig Gene Mapping Project; SNP, Single-Nucleotide Polymorphism.

---

*Historical Studies in the Natural Sciences*, Vol. 52, Number 3, pps. 443–475. ISSN 1939-1811, electronic ISSN 1939-182X. © 2022 by the Regents of the University of California. All rights reserved. Please direct all requests for permission to photocopy or reproduce article content through the University of California Press's Reprints and Permissions web page, <https://www.ucpress.edu/journals/reprints-permissions>. DOI: <https://doi.org/10.1525/hsns.2022.52.3.443>.

the range of evidence available for historical analysis. This essay is part of a special issue entitled *The Sequences and the Sequencers: A New Approach to Investigating the Emergence of Yeast, Human, and Pig Genomics*, edited by Michael García-Sancho and James Lowe.

KEY WORDS: genomics, sequencing, mixed-methods, social network analysis, collaboration, quantitative methods, qualitative methods

## 1. INTRODUCTION: THE TASK OF HISTORICIZING GENOMICS

In 2020, Michel Morange published an extensively revamped version of his acclaimed *History of Molecular Biology*. He explained that his reason for undertaking a major update—rather than just issuing a new edition of the 1994 version—was to account for the achievements that had occurred in the intervening twenty-six years. Among these were the culmination of the Human Genome Project (HGP) and the emergence of post-genomic fields such as systems and synthetic biology. Morange wondered whether these achievements represented a rupture with the molecular biological paradigm or rather continuations—more or less pronounced—of its fundamental tenets. He concluded that, although biology had radically changed in the first two decades of the twenty-first century—notably due to the availability of large amounts of data—there was no field of research “in which molecular descriptions and explanations” had become “obsolete.”<sup>1</sup>

In Morange’s view, molecular biology had operated through its history by introducing its tools and models into the problems of other disciplines. This tendency became clearer from the mid-1970s onward when molecular biology dissolved into fields such as developmental biology. By addressing these transitions, and the theories, methods, and explanations that molecular biology had offered to other disciplines since its original formulation in the mid-twentieth century, Morange believed that historians could shed new light on claims of novelty in today’s life sciences.

Morange’s argument resonates with a thesis that historians of molecular biology formulated during the 1990s: that molecular biology was never, strictly speaking, a discipline, and any disciplinarity it possessed was partial and evanescent. Molecular biology was, rather, a set of tools that researchers from other disciplines—such as genetics or microbiology—adopted and adapted to

1. Michel Morange, *The Black Box of Biology: A History of the Molecular Revolution* (Cambridge, MA: Harvard University Press), on 386.

problems that the life sciences had been addressing for decades.<sup>2</sup> The disciplinary status of molecular biology was the result of a rhetorical game during the 1950s and 1960s in which its advocate scientists successfully promoted the necessity and novelty of a field fostering their proposed molecular explanation of life, thus securing support and funding for new laboratories, journals, and research programs.<sup>3</sup> Subsequently, this emerging infrastructure developed its own conceptual apparatus with which to interpret and investigate life, and new analytical techniques specific to the task of grasping the fine details of cellular mechanisms and processes. It is these analytical and conceptual tools that multifarious other disciplines have adopted, their molecularization accompanying the diminishment of the distinct disciplinarity of molecular biology.<sup>4</sup>

The genealogies we have uncovered in this special issue suggest that the history of genomics experienced a similar pattern. While, in the 1990s, historians probed the origin myth of the molecular revolution, showing that molecular approaches to life existed well before the emergence of molecular biology, we have both diversified the history of genomics and queried its disciplinary status. The establishment of genomics as a differentiated field with its own institutions and programs such as the genome centers and the HGP was also the result of rhetoric and persuasion by, among others, one of the self-appointed founders of molecular biology, James Watson.<sup>5</sup> As he did with

2. See, for example, Soraya de Chadarevian and Jean-Paul Gaudillière, eds., "The Tools of the Discipline: Biochemists and Molecular Biologists," special issue of *Journal of the History of Biology* 29 (1996).

3. Pnina G. Abir-Am, "The Politics of Macromolecules: Molecular Biologists, Biochemists, and Rhetoric," *Osiris* 7 (1992): 164–91; Jean Gayon, "Is Molecular Biology a Discipline?," in *History and Epistemology of Molecular Biology and Beyond: Problems and Perspectives*, preprint (#310) (Berlin: Max Planck Institute for The History of Science, 2006), 249–52.

4. Michel Morange, "The Historiography of Molecular Biology," in *Handbook of the Historiography of Biology*, eds. Michael Dietrich, Mark Borrello, and Oren Harman (Cham, Switzerland: Springer International Publishing, 2018); Hans-Jörg Rheinberger, "Recent Science and Its Exploration: The Case of Molecular Biology," *Studies in History and Philosophy of Biological and Biomedical Sciences* 40, no. 1 (2009): 6–12.

5. Along with other scientists, Watson positioned himself as member of a self-conscious "genomics vanguard." On this vanguard and the genesis of genomics as a discipline, see Stephen Hilgartner, *Reordering Life: Knowledge and Control in the Genomics Revolution* (Cambridge, MA: MIT Press, 2017), 27–30, 38–41, 47–61, 91–110. On the formation of disciplines more generally, and their role in structuring and shaping knowledge: Jan Golinski, *Making Natural Knowledge: Constructivism and the History of Science* (Chicago: University of Chicago Press, 2008 [1998]), 66–70; Ilana Löwy, "On Hybridizations, Networks and New Disciplines: The Pasteur Institute and the Development of Microbiology in France," *Studies in the History and Philosophy of Science* 25 (1994): 655–88; Robert E. Kohler, *From Medical Chemistry to Biochemistry: The Making of a Biomedical Discipline* (Cambridge, UK: Cambridge University Press, 1982).

molecular biology in the 1960s and biotechnology in the 1970s, Watson succeeded in persuading funders and policymakers of the suitability of pursuing the determination of the reference human genome sequence through a new form of scientific organization that we have called the large-scale center model. Yet before Watson and other supporters of this model established the new regime—and parallel to it—a variety of life scientists adopted and mobilized genome mapping and DNA sequencing practices that differed in their organization from the large-scale centers. We have addressed through the preceding papers the mapping and sequencing models of cell biochemists, immunologists, systematics researchers, and geneticists with medical and agricultural orientations. This diversity shows that, as much as constituting a discipline in its own right, genomics was a set of tools that penetrated and transformed other life sciences disciplines. The histories of the communities and disciplines that adopted genomic tools are thus co-constitutive and inseparably intertwined with the history of genomics itself.<sup>6</sup>

Genomic tools push researchers to treat the genome as a distinct biological object and foster the development of infrastructures and new scientific roles. These infrastructures include databases, the figure of the curator, and a new discipline that emerged and developed hand in hand with genomics: bioinformatics.<sup>7</sup> The tools include DNA sequencing, assembly, and annotation techniques—from more manual and artisanal to high-throughput and automated—as well as browsers and software, and new means to query databases to find pertinent sequence data.

Researchers can assemble these tools, roles, and infrastructures differently, and the resulting assemblages do not need to promote a highly concentrated and large-scale approach to genomics. Just as life scientists from disparate fields used the tools and techniques of molecular biology to advance research agendas beyond the interests of self-declared pioneers such as Watson, the resources generated under the umbrella of genomics have also transcended their

6. Elsewhere we have formulated a new term—“genomicists”—to emphasize the diversity of these communities of practitioners and their agency—sometimes collaborative and sometimes competitive—in making the history of genomics: Miguel García-Sancho and James W. E. Lowe, *A History of Genomics Across Species, Communities and Projects* (London: Palgrave Macmillan, forthcoming), esp. chap. 1.

7. On the tight coupling of the history of genomics, the development of particular kinds of databases, and the discipline of bioinformatics, see Hallam Stevens, *Life Out of Sequence: A Data-Driven History of Bioinformatics* (Chicago: University of Chicago Press, 2013). On the rise and significance of curators in genomics infrastructures, see Sabina Leonelli, *Data-Centric Biology: A Philosophical Study* (Chicago: University of Chicago Press, 2016).

mobilization at large-scale centers. As well as uncovering the ways in which genomics unfolded outside concerted projects to determine whole reference genomes, our perspective has enabled us to find tools, techniques, and infrastructures that distinct communities produced, assembled, and repurposed. This includes various kinds of genomic maps and mapping methods; catalogs of clinically relevant variants drawn from the vertical sequencing strategies of medical genetics; the bricoleurs of pig genomics generating and readapting radiation hybrid panels and DNA libraries; and the bioinformatic processes and coordination that the Martinsried Institute for Protein Sequences led within the European Yeast Genome Sequencing Project. These approaches to genomics also included distinct organizational models of coordinating sequencing, such as the network genomics of the European Commission, the chromosome workshops and disease consortia of medical geneticists, the jamboree meetings of Celera Genomics, and the special relationship between a well-established grouping of institutions engaged in pig genetics with the Sanger Institute, a genome center.

The diversity of assemblages in which genomics materialized paves the way for adding new dimensions to Morange's portrayal. By looking at the ways that different research communities adopted DNA sequencing tools, we have presented lineages of genomics beyond the theories, models, and experimental approaches of molecular biology. Prior literature had already documented the use of protein and later DNA sequences in evolutionary biology, a field that during the 1970s and early 1980s witnessed a debate about the status of molecular data compared to more traditional morphological evidence.<sup>8</sup> Our contribution, apart from adding new genealogies, has been deploying a mixed-methods approach that helps find them more systematically. Crucially, we have also related these genealogies to each other, and connected genomic practice to medical, agricultural, and other life science research goals.

The existing historiography has tended to extend the temporal horizon diachronically and position genomics in a longer-term history punctuated by continuities and change. We offer an alternative lens that *thickens* as well as stretches time: it renders visible the multiple, simultaneous, and rapidly changing assemblages of capabilities, resources, and practices underlying genomics,

8. Edna Suárez-Díaz, "Making Room for New Faces: Evolution, Genomics and the Growth of Bioinformatics," *History and Philosophy of the Life Sciences* 32 (2010): 65–89; Bruno Strasser, "Collecting, Comparing, and Computing Sequences: The Making of Margaret O. Dayhoff's Atlas of Protein Sequence and Structure, 1954–1965," *Journal of the History of Biology* 43 (2010): 623–60.

and captures synchronic as well as diachronic processes. Examples of thick synchronicity are the collaboration of medical geneticists and Celera Genomics around human chromosome 7, the contribution of European institutions to the large-scale sequencing of yeast chromosome XII, the entanglements of sequence production and use that both chromosome endeavors entailed—the former for clinical purposes and the latter for cell biological research—and the work of Washington University and the Sanger Institute across yeast, human, and—in the case of the latter institution—whole-genome pig sequencing. Our approach and methodology enabled us to detect and analyze these synchronic connections in the preceding papers, which a more case study–based history may have overlooked.

It was thus our continuous shuttling between quantitative data and qualitative inquiry—mediated by our visual and metric analysis of co-authorship networks—that thickened the historiographical boundaries of genomics. A key stimulus of this process was our early realization that the main publishers of DNA sequences differed from the main submitters and that publications describing sequences were not just proxies of the submission of those sequences to central data repositories. Further examination of those publications revealed an ecosystem of institutions that collaborated in both the determination of sequences and their use in a variety of research programs. The publications, therefore, conveyed a complex story that went beyond the mere production and submission of DNA sequences; they represented a point of entry to investigate the entanglement between the production of sequences and their use in scientific, medical, and agricultural practices.<sup>9</sup>

Most of the publishers of human sequences were based in medical genetics institutions that the large-scale center model of the HGP had recast as mere users of the reference human genome. Yet their publications revealed that these institutions were not only users but also producers of what we categorized as *vertical* sequences: sequence data concerning variation at specific chromosomal positions linked to genetic diseases. These vertical sequences and knowledge of the clinical consequences of their variability were crucial to reconnect the *horizontal* reference genome data to medical problems.<sup>10</sup>

9. Rhodri Leng, Gil Viry, Miguel García-Sancho, James Lowe, Mark Wong, and Niki Vermeulen, “The Sequences and the Sequencers: What Can a Mixed-Methods Approach Reveal about the History of Genomics?,” this issue.

10. Miguel García-Sancho, Rhodri Leng, Gil Viry, Mark Wong, Niki Vermeulen, and James Lowe, “The Human Genome Project as a Singular Episode in the History of Genomics,” this issue.

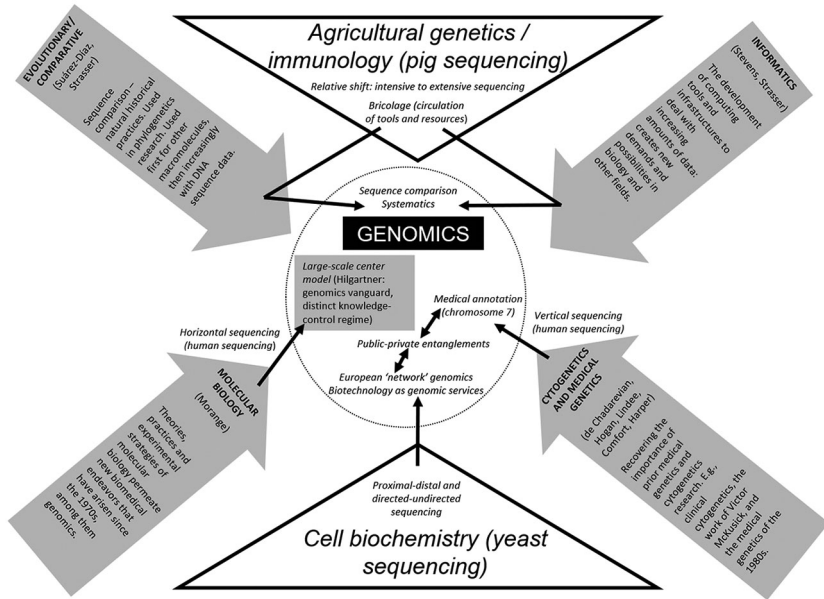
The co-authorship ties of the yeast publications enabled us to fractionate sequencing practices according to the degree of proximity of the production of data to specific goals and users. We did this by comparing the large-scale center model with the network organization that characterized the European Yeast Genome Sequencing Project. Some of the large-scale centers in the United States had little interest in yeast biology. We argued that their sequencing work was undirected to specific research objectives and distal from final users of the yeast sequences. By contrast, the network genomics of the European Commission involved a variety of institutions ranging from sequencing companies that did not participate in the use of the sequences, to biochemistry and cell biology laboratories that were themselves the primary users of the data and exploited the sequences they produced to further their research objectives.<sup>11</sup>

Finally, our analysis of the pig publications provided a temporal dimension that further complicated producer-user dynamics. By examining *bricolaging* processes embedded in our co-authorship network, we argued that the production of a reference genome was just one manifestation among many of the repurposing of materials and tools that pig genomics entailed. Other repurposing practices included the characterization of breeds, populations, and families of pigs different from the ones embodied in the reference genome. The release of the reference genome, nevertheless, enabled a shift in the nature of bricolaging and collaborations from an intensive emphasis on a specific type of pig—the object of the reference data—toward more extensive sequencing being conducted across breeds, populations, and families.<sup>12</sup>

A key aspect of our methodology is that we derived our findings from the quantitative, visual, and qualitative analysis of data rather than any preselected case study. We drew out our analytical categories and proposed genealogies during our ongoing mixed-methods interrogation of datasets and networks. Horizontal and vertical, proximate and distal, directed and undirected, intensive and extensive sequencing were interpretative concepts to account for institutions and ties that do not fit with the well-characterized division of labor associated with the large-scale center model. These concepts enabled us to both detail and operationalize the entanglement between sequence production and

11. Miguel García-Sancho, James Lowe, Gil Viry, Rhodri Leng, Mark Wong, and Niki Vermeulen, “Yeast Sequencing: ‘Network’ Genomics and Institutional Bridges,” this issue.

12. James Lowe, Rhodri Leng, Gil Viry, Mark Wong, Niki Vermeulen, and Miguel García-Sancho, “The Bricolage of Pig Genomics,” this issue.



**FIGURE 1.** Diagram depicting the existing genealogies of genomics as established in the scholarly literature (in the grayed-out arrows and box) and some of the contributions we have made to thicken the history of genomics, represented in italics. Figure 2, below, depicts the further consequences of our thickening of the history of genomics. Figure elaborated by the authors.

use that prior scholarship had observed.<sup>13</sup> This operationalization took the form of identifying genealogies between genomics and medical genetics, immunology, cell biochemistry, systematics, and livestock breeding, thus *thickening* the continuities that other historians had proposed with molecular and evolutionary biology (see figure 1).

Our approach, however, still relies on case studies that we selected from our networks. In the next section, we reflect on the relationship between the networks and the case studies, drawing on existing debates on *macro* and *micro* perspectives in both the historiography of science and network analysis.<sup>14</sup> Our

13. See, for instance, Hilgartner, *Reordering* (n.s), 140; and Bruno J. Strasser, *Collecting Experiments: Making Big Data Biology* (Chicago: University of Chicago Press, 2019), chaps. 3 and 5.

14. See, for example, H. Floris Cohen, ed., “The ‘History Manifesto’ and the History of Science,” *Isis* 107, no. 2 (2016): 309–10; Bonnie Erickson, “Social Networks and History: A Review Essay,” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 30 (1997): 149–57.



case studies, as we contend, are not more representative than others that previous scholarship has addressed. What makes them novel is the way that we selected them and, crucially, their embeddedness in the co-authorship network as an entity that points to global dynamics in the organization of genomics, without, of course, ever fully representing them.

We then argue that, for our thick historiography to develop further, it is essential to fashion mixed-methods approaches that capture not only sequencing and sequences but also other forms of genomic practices and data. We outline some strategies to move beyond the results of this special issue, for example in the development of a domain ontology that would connect—in a formal and stable way—our submissions and publications to other forms of data of interest for the historical study of genomics. Finally, we conclude that our perspective of genomics as a set of tools entangled with the history of various life science disciplines and practices—rather than a field in itself that needs to be *applied*—may offer a fresh perspective to the problem of medical and agricultural translation of sequence data.

## 2. THE NETWORK AND THE CASE STUDIES

Questions concerning the interpretation and meaning of our co-authorship networks became apparent from the moment we began analyzing them. This analysis was the collective endeavor of a multidisciplinary team that included historians of science, as well as quantitative and qualitative social scientists. From day one, it was clear that the team members differed in their understanding of what the analysis of the network involved and, especially, achieved. While some of us prioritized examining the network's global structural properties, others believed the focus should be on exploring co-authoring institutions or co-authorship relationships that were particularly striking against the background of our qualitative historical knowledge.

These productive differences in our team reflected a longstanding debate on how to use network visualizations in historical research. Some scholars favor “looking hard at tie structure” and “only then” asking “how structural position relates to attributes” of a qualitative nature. Others would rather use the network as a map to identify actors (nodes) and connections (ties) that they can subsequently investigate through other historical sources, such as archives,

oral histories, or primary and secondary literature.<sup>15</sup> The first approach predominantly draws on metrics such as centrality and density scores that reflect properties of the whole network. The second seeks to identify potential case studies in the network through the selection of clusters, temporal partitions, or filters that ease visualization.

Within their still-recent use of network visualizations, historians of science have mainly favored the network-as-a-map approach.<sup>16</sup> This is due to the configuration of history of science as a discipline and its traditional preference for carefully delimited case studies over wider and more loosely characterized accounts, either geographically or chronologically. Since the so-called anthropological turn of the 1970s, history of science has borrowed considerably from the theoretical and methodological tools of the social sciences, thus incorporating thick and empirically grounded descriptions as their main source of evidence.<sup>17</sup> This has led the literature to refrain, with notable exceptions, from big-picture narratives as potentially leading to overgeneralizations and redolent of a history of ideas approach that historians of science have struggled to distance themselves from.<sup>18</sup>

15. Erickson, "Social Networks" (n.14), 150. These different approaches embody a deeper ontological debate on the nature and properties of networks, as well as individual or collective agency. For a Science and Technology Studies perspective on these debates, see Sally Wyatt, Staša Milojević, Han Woo Park and Loet Leydesdorff, "Intellectual and practical contributions of scientometrics to STS," in *The Handbook of Science and Technology Studies*, ed. Ulrike Felt, Rayvon Fouché, Clark A. Miller and Laurel Smith-Doerr (Cambridge, MA: MIT Press, 2017), 87–112.

16. See, for instance, Yves Gingras, "Revisiting the 'Quiet Debut' of the Double Helix: A Bibliometric and Methodological Note on the 'Impact' of Scientific Publications," *Journal of the History of Biology* 43 (2010): 159–81; Michael Pettit, Darya Serykh, and Christopher D. Green, "Multispecies Networks: Visualizing the Psychological Research of the Committee for Research in Problems of Sex," *Isis* 106 (2015): 121–49. In both cases, the authors use network visualizations to identify lines of research in the scientific fields they are exploring, and generate questions that require qualitative historical tools to address.

17. See the following debate: Sheila Jasanoff, "Reconstructing the Past, Constructing the Present: Can Science Studies and the History of Science Live Happily Ever After?," *Social Studies of Science* 30, no. 4 (2000): 621–31; Lorraine Daston, "Science Studies and the History of Science," *Critical Inquiry* 35, no. 4 (2009): 798–813; and a response: Peter Dear and Sheila Jasanoff, "Dismantling Boundaries in Science and Technology Studies," *Isis* 101, no. 4 (2010): 759–74.

18. Some scholars consider that this reluctance to pursue big-picture approaches has triggered unintended consequences, such as fragmentation of the historiography of science and inhibition of intra- and interdisciplinary communication: see papers by Robert E. Kohler, Paula Findlen, Steven Shapin, and David Kaiser in "Focus: The Generalist Vision in the History of Science," special issue of *Isis* 96, no. 2 (2005): 224–51.

History of science has thus tended to draw its general claims from the interpretation of case studies. To avoid the shortcomings of a strict case study approach, historians of science have developed theoretical tools to transcend the specificity of investigating bounded and concrete events. Rather than presenting case studies in isolation, their investigations have emphasized the “heuristic value” of these events and their capacity to “reveal structures and connections that remain invisible in more wide-sweeping accounts.”<sup>19</sup> A complementary technique to look beyond one or various case studies is the *longue-durée* framework that Morange advocates for the history of the life sciences. Both within and outside the life sciences—and beyond the contemporary period that Morange explores—*longue-durée* perspectives stress continuities and patterns not only across case studies but also throughout the historical periods in which they unfolded.<sup>20</sup>

Building on this, a number of broader narratives have emerged within the historiography of science. They generally adopt categories that provide an interpretative thread through case studies over a long timeframe. A paradigmatic example is John Pickstone’s ways of knowing and working, the two categories through which he addressed the production of scientific, technological, and medical knowledge from the Renaissance to the twentieth century. More recently, Bruno Strasser has tackled the history of the life sciences throughout the twentieth century by exploring instances of hybridization of the practices of comparing and experimenting. This has enabled him to show continuities between the global DNA sequence databases of the 1980s and earlier repositories in natural history museums.<sup>21</sup>

19. Soraya de Chadarevian, “Microstudies versus Big Picture Accounts?,” *Studies in History and Philosophy of Biological and Biomedical Sciences* 40 (2009): 13–19, on 16; see also Robert E. Kohler, “A Generalist’s Vision,” *Isis* 96, no. 2 (2005): 224–29. Other historians have critiqued the extent to which drawing general conclusions from case studies is possible: Peter Galison, “Ten Problems in History and Philosophy of Science,” *Isis* 99, no. 1 (2008): III–24.

20. Morange’s proposed *longue-durée* concerns only contemporary biomedicine and contrasts with earlier uses of this framework: Morange, “The Historiography” (n.4). On *longue-durée* perspectives in the historiography of science, the modern sciences and the life sciences more generally, see Frederick L. Holmes, “The *Longue Durée* in the ‘History of Science’,” *History and Philosophy of the Life Sciences* 25, no. 4 (2003): 463–70; Mathias Grote, “What Could the ‘Longue Durée’ Mean for the History of Modern Sciences?,” *Archives Ouvertes Working Paper Series* (2015), <https://halshs.archives-ouvertes.fr/halshs-01171257>; Mathias Grote, “Petri Dish versus Winogradsky Column: A *Longue Durée* Perspective on Purity and Diversity in Microbiology, 1880s–1980s,” *History and Philosophy of the Life Sciences* 40, no. 1 (2018): II.

21. John V. Pickstone, *Ways of Knowing: A New History of Science, Technology, and Medicine* (Chicago: University of Chicago Press, 2001); Strasser, *Collecting Experiments* (n.13).

An additional opportunity to widen narrative lenses are the digital data-banks that proliferate in the humanities, and the social and natural sciences: rapidly growing online collections that offer historians all sorts of data, from bibliographic references to demographic information or DNA sequences.<sup>22</sup> Our special issue and the mixed-methods approach underlying it represents a step in this direction. The data underpinning our networks encompassed a thirty-five-year timeframe and required the triangulation of millions of sequence submissions and thousands of publications that institutions all over the world co-authored.<sup>23</sup> This multiplied the range and breadth of case studies we could work with and enabled us to trace our lineages between genomics and other disciplinary or real-world problems. However, as we have emphasized throughout the special issue, the data and networks in themselves fell short in capturing the entire history of genomics: because of the way we collected and constructed them, they included artifactual records and omitted crucial genealogies of genomics research. Rather than being comprehensive or more objective representations of genomics, they constituted a platform from which we could address the methodological challenges of historicizing data.

One of these challenges is to steer between what Stephen Gaukroger has called the twin pitfalls of undercontextualization and overcontextualization. Through our awareness of the way in which the networks were decontextualized abstractions of the practices that generated the sequence submissions and publications in the first place, and in qualitatively recontextualizing the networks and the parts and patterns therein, we have been able to evaluate what they showed and concealed of the history of genomics. The networks led us to identify and examine parts of them—case studies on particular clusters, for example—that we used, in turn, to illuminate our understanding and appreciation of the patterns and structure of the wider network: the whole.

In conjunction with the qualitative research that the identification of these key parts or patterns of the network prompted, this process of zooming in and

22. Jane Maienschein, "Time, Impact, and the Need for Digital History and Philosophy of Science," *Isis* 107, no. 2 (2016): 344–45; Ivan Flis, Evina Steinová, and Paul Wouters, "Digital Humanities Are a Two-Way Street," *Isis* 107, no. 2 (2016): 346–48; Eugene Garfield, Alexander I. Pudovkin, and V. S. Istomin, "Why Do We Need Algorithmic Historiography?," *Journal of the American Society for Information Science and Technology* 54, no. 5 (2003): 400–12.

23. On widening geographical, as well as temporal, historiographical boundaries, see Ana Barahona, "Local, Global, and Transnational Perspectives on the History of Biology," in *Handbook of the Historiography of Biology*, eds. Michael Dietrich, Mark Borrello, and Oren Harman (Cham, Switzerland: Springer International Publishing, 2018).

out allowed us to embed the parts—the clusters or case studies—in a wider context that broadened their historical significance. According to Gaukroger, this wider context does not need to be a comprehensive, fully detailed picture of the narrative historians seek to convey; it just needs to be sufficiently developed to place the case studies within a broader web of signifiers with which they relate—in terms of both similarity and contrast.<sup>24</sup> Our networks provided this web within which our case studies acquired heuristic value and revealed broader historiographical trends.

It was through their embeddedness in the networks that we saw our case studies as representative of different ways of organizing the practice of sequencing: horizontal and vertical in human; intensive and extensive in pig; and distal and proximate, as well as directed and undirected, in yeast. These different ways of sequencing became the analytical categories with which we historically probed the generality of the large-scale center model that shaped the last stages of the HGP.<sup>25</sup> Yet a difference between our categories and those established in other broad narratives of the historiography of science was that, rather than bringing them in at the outset of the investigations, we developed our ways of sequencing along with the analysis of patterns and properties in the co-authorship networks, as well as qualitative historical research.

In the human network, the exploration of the strong ties between medical schools and hospitals led us to propose vertical sequencing as an alternative approach to the horizontal strategy that had characterized the production of the reference human genome sequence. In the yeast network, the institutions involved in the European Yeast Genome Sequencing Project showed, through their co-authorship relationships, more heterogeneity and flexibility in their organization than the genome centers: proximate and directed, as well as distal and undirected sequencing. In the pig network, by examining the

24. Stephen Gaukroger, "Undercontextualization and Overcontextualization in the History of Science," *Isis* 107, no. 2 (2016): 340–42. On the complex dialectic between qualitative historical analysis, large datasets, and network visualizations, see also Kenneth D. Aiello and Michael Simeone, "Triangulation of History Using Textual Data," *Isis* 110 (2019): 522–37; Deryc T. Painter, Bryan C. Daniels, and Jürgen Jost, "Network Analysis for the Digital Humanities: Principles, Problems, Extensions," *Isis* 110 (2019): 538–54.

25. On the importance of choosing analytical categories that do not reify dichotomies that dominant narratives have established—such as production and use as the main organizational principles of genomics research—see Carla Nappi, "The Global and Beyond: Adventures in the Local Historiographies of Science," *Isis* 104 (2013): 102–10.

co-authorship ties and publishing patterns, we observed a shift in balance from intensive sequencing before the production of the reference genome to extensive sequencing afterward, accompanied by a change in models of collaboration.

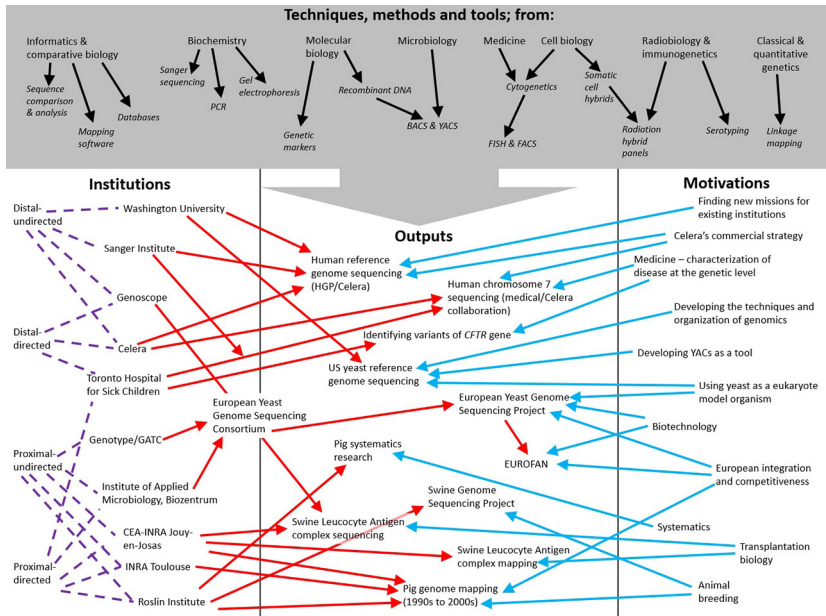
Overall, it was the whole set of co-authorship ties with all the other nodes in the networks that shaped the position of these institutions and configured the spaces they interlinked. This spatial configuration, along with the evolution of the co-authorship ties over time, led us to realize that the large-scale genome center model represented only a small part of the sequencing practices that our networks captured: it operated within a limited timeframe and in a small number of institutions across the three networks.

Our history of genomics has thus decentered a time period—covering the HGP and the implementation of the large-scale center model—that had been the focus of most of the existing accounts.<sup>26</sup> We have done so by both historicizing the HGP and bringing into the equation other forms of conducting genomics that occurred before, during, and after the sequencing of the human genome. Revealing these other forms of genomics has enabled us to thicken existing narratives and present the history of genomics as inextricably entangled with the trajectories and goals of cell biochemistry, systematics, immunology, and medical and agricultural genetics, among other fields (see figure 2). We have also expanded the range of species considered: yeast and pig, as well as human.<sup>27</sup>

These other forms of genomics, however, are still largely other forms of conducting DNA sequencing. In what follows, we argue that in order to further thicken the history of genomics we need to move from sequencing to other genomic practices and data. We discuss sources and methodological strategies to do so, and propose the creation of a domain ontology that integrates and connects various forms of data with potential to shed light on the history of genomics.

26. For a de-centering exercise in the historiography of early modern science, see Andrew Cunningham and Perry Williams, "De-Centring the 'Big Picture': The Origins of Modern Science and the Modern Origins of Science," *The British Journal for the History of Science* 26 (1993): 407–32.

27. On how considering non-human species may illuminate the analysis of biomedical practice: Sabina Leonelli, "When humans are the exception: Cross-species databases at the interface of biological and clinical research," *Social Studies of Science* 42, no. 2 (2012): 214–236.



**FIGURE 2.** Diagrammatic depiction of the thickened history of genomics that we have introduced in this special issue. It portrays the diversity of techniques (and sources thereof), as well as the multiple outputs, organizational modes, and motivations to sequence. The italicized labels indicate techniques, methods, and tools. Figure elaborated by the authors.

### 3. CONSIDERATIONS FOR FURTHER THICKENING THE HISTORY OF GENOMICS

Our thickening of the history of genomics so far has relied on two main anchors: sequence submissions and publications, with the latter rooted in the former. Sequences and sequencing, however, are not the only elements of genomics and, as one of us argued elsewhere, historians need to transition from a “thin” viewpoint to considering the “thick” array of practices involved in genomics research, among them the creation of DNA libraries and maps.<sup>28</sup> Through this special issue, we have extended this thickening approach from the sequences themselves to the process of sequencing, including the subjects, objects, and boundaries of those sequencing processes as well. Our categories—of horizontal and vertical, proximate and distal, directed and

28. James W. E. Lowe, “Sequencing through thick and thin: Historiographical and philosophical implications,” *Studies in History and Philosophy of Biological and Biomedical Sciences* 72 (2018): 10–27.

undirected, intensive and extensive sequencing—encompass subjects, objects, boundaries, and processes, and helped us to augment our narrative lens when addressing sequencing practices. We will build on this to outline how other forms of data—and capturing the relationships between different forms of data—can move beyond what we have done, which has implications for other areas of the history of science without ready access to the kinds of data we have exploited in this special issue.

### 3.1. Moving beyond DNA Sequences

We have argued that DNA sequence submissions to public databases and their associated first publications constitute a useful platform to historicize genomics. Indeed, it is in part through our analyses of these data that forms of genomic work and genealogies beyond the HGP first came to our attention—for instance, the lineage of pig extensive sequencing and systematic research. Systematic surveys of biodiversity in pig and other species often involve new sequencing and the downloading of—and commensurating with—previously produced sequence data. Yet other relevant genealogies of genomics such as continuities with quantitative genetics and linkage mapping became clear only through qualitative research in archives and oral histories with scientists, some of whom we identified thanks to our networks.

We have been able to relate these other forms of genomic work to the co-authorship ties, for example by linking the story of medical genetics research to the sequencing of progressively larger regions of the genome, or by relating the pig genome mappers to the pig genome sequencers. In articulating a distinction between thick and thin sequencing, we emphasized the impossibility of demarcating DNA sequencing from other forms of genomic work without implicitly or explicitly recapitulating artificial divisions of the processes, organizational forms, temporalities, and spatialities involved. A clear example of this is genome mapping: undertaking this practice often implies sequencing, and sequencing has historically depended on mapping.<sup>29</sup>

29. Lowe, “Sequencing” (n.28); Adam Bostanci, “Sequencing Human Genomes,” in *From Molecular Genetics to Genomics: The mapping cultures of twentieth-century genetics*, ed. Jean-Paul Gaudillière and Hans-Jörg Rheinberger (Abingdon: Routledge, 2004), 158–179; Soraya de Chadarevian, “Mapping the worm’s genome. Tools, networks, patronage,” in *From Molecular Genetics to Genomics: The mapping cultures of twentieth-century genetics*, ed. Jean-Paul Gaudillière and Hans-Jörg Rheinberger (Abingdon: Routledge, 2004), 95–110.



Our accounts have not ignored or sidelined these forms of genomic work beyond sequencing. Yet so far, we have discussed all of them as they relate to DNA sequence submissions, which act as a center of gravity of this special issue. Other centers of gravity may occasion different investigative paths and historical accounts. Producing other datasets—or other types of analytical tools and resources—to capture the diversity of forms of work around genomics would be a way of creating alternative centers of gravity. This would expand the narrative frame of existing accounts and would likely result in new genealogies of genomics, and the forging of connections between them.

This is not, however, a simple task. DNA sequence submissions have the advantages of being discrete, unitized, and held in relatively well-supported databases with dedicated institutional support. The databases are not fully coherent or comprehensive, but they are expansive. We can be confident they include most sequences generated after certain dates, at least in institutions whose funding bodies or governing policies mandated timely submission. Other forms of genomic work may not be amenable to extraction in the interests of producing equivalent kinds of datasets to the ones we have for DNA sequence submissions. Here we discuss three of these: genome mapping and the production of other *omic* databases; the generation of genomic resources and tools; and genomic prediction and commercially sensitive research.

### Genome Mapping and Other Omic Databases

Many different kinds of genomic maps aided sequencing practices and also required the production of sequence data for their construction. These maps include genetic or linkage maps, cytogenetic maps, and radiation hybrid maps. While the maps may still be extant in published form, many of the databases that stored their underlying data are now extinct or off-line. Apart from the mapping information, such databases included submitter details and other relevant metadata.

The organizational and technical nature of the mapping enterprises conditioned the extraction and recording of the resulting data. For example, direct submission of genes to the UK Human Genome Mapping Project or mapped markers based on the INRA-Minnesota porcine Radiation Hybrid panel (IMpRH) included information such as the identity of the laboratory submitting it.<sup>30</sup> Curated resources focused on linking mapping assignments to putative functions such as Online Mendelian Inheritance in Man and the Pig

30. Lowe et al., “The Bricolage” (n.12), section 4.2.

Quantitative Trait Locus Database still exist, and include mapping data and links to publications that scholars can exploit in a similar way to our sequence submissions.<sup>31</sup> Yet in other projects (such as the Pig Gene Mapping Project—PiGMap—and the genetic and physical mapping of yeast) the tracing of individual submissions is complicated. This is because the initiatives made mapping assignments based on the collation and analysis of data from multiple contributing laboratories (in the case of PiGMap) or a small group from which the eventual maps remain but not data on individual assignments. Resultant publications thus tend to be at the level of the overall collaboration, like the *Yeast Genome Directory*,<sup>32</sup> and the PiGMap cytogenetic and genetic map papers.<sup>33</sup> The data submitted is not therefore in a form as to easily facilitate tracing collaborations beyond what qualitative research can already demonstrate.

A rich potential source of data that would capture more functional aspects of genomics are databases specific to particular species, such as ClinVar for humans, the *Saccharomyces* Genome Database for yeast, and the Pig Quantitative Trait Locus Database and Porcine Translational Research Database for the pig.<sup>34</sup> There is immense historical promise in the use of such databases, even considering the practicalities and necessary selectivity of abstracting some of the rich, interconnected data. There is, however, a danger of being seduced by the depth of the data that such repositories include. These databases are partitioned according to individual species, on an even more radical level than the European Nucleotide Archive, which organizes its entries by species while encompassing all of them. This infrastructural aspect reflects the coalescence of communities around particular organisms for purposes ranging from their status as model organisms (yeast), animal models (pigs), or translational resources (humans, yeast, and pigs).<sup>35</sup>

31. [www.ncbi.nlm.nih.gov/omim](http://www.ncbi.nlm.nih.gov/omim) and [www.animalgenome.org/cgi-bin/QTLDdb/SS/index](http://www.animalgenome.org/cgi-bin/QTLDdb/SS/index)

32. Yeast Genome Directory, *Nature* 387, no. 6632 S (1997).

33. Alan L. Archibald, Chris S. Haley, Judy F. Brown, Sandra Couperwhite, Heather A. McQueen, David Nicholson, Wouter Coppieters, et al., “The PiGMap Consortium Linkage Map of the Pig (*Sus scrofa*),” *Mammalian Genome* 6 (1995): 157–75; Martine Yerle, Yvette Lahbib-Mansais, Clemens Mellink, André Goureau, Philippe Pinton, Geneviève Echard, Joël Gellin, et al., “The PiGMap Consortium Cytogenetic Map of the Domestic Pig (*Sus scrofa domestica*),” *Mammalian Genome* 6 (1995): 176–86.

34. [www.ncbi.nlm.nih.gov/clinvar](http://www.ncbi.nlm.nih.gov/clinvar); [www.yeastgenome.org](http://www.yeastgenome.org); <http://199.133.11.115/fmi/webdl/Porcine%20Translational%20Research%20Database>

35. On community databases, see Sabina Leonelli and Rachel A. Ankeny, “Re-thinking Organisms: The Impact of Databases on Model Organism Biology,” *Studies in History and Philosophy of Biological and Biomedical Sciences* 43, no. 1 (2012): 29–36.

Pursuing further study separately on individual species due to the availability of rich resources on them risks suppressing lines of research that highlight the porosity of the boundaries of yeast, human, and pig.<sup>36</sup> There were significant overlaps between yeast and human genomics on the US side, for instance, and the Sanger Institute was a presence in human, yeast, and pig genomics. Researchers and funders used sequencing results in one organism to create resources (e.g., Yeast Artificial Chromosomes) and ways of organizing sequencing (e.g., the use of yeast as a pilot for human genome sequencing in the United States) that were deployed for different species. Sequence and sequence-aligned data pertaining to any one species may serve as a resource for the further production, elaboration, and sense-making of genomic data for another.<sup>37</sup>

As the authors of this special issue all worked across the human, yeast, and pig networks, we were better able to apprehend the kinds of differences in practice and organization that led us to the formulation of the analytical categories that have helped us to thicken the history of genomics. For any one species, therefore, we could not adequately characterize genomics except with reference to practices that make use of, and engage with, the genomics and genomes of other species. The use of the species as an organizing principle may make research tractable, just as it does the organization of data into databases. Finding ways in which species divisions do not unduly account for or channel historical analysis is, however, crucial to properly documenting the trans-specific and cross-specific aspects of a thicker history of genomics.

### Genomic Resources and Tools

The dissemination of materials such as DNA samples, radiation hybrid panels, Bacterial Artificial Chromosome libraries, DNA probes and primers, and software for mapping and sequence assembly has been a crucial part of the enterprise of genomics. It constitutes a material form of collaboration that the

36. Historians of science working on animal experimentation have argued for the necessity of focusing on work across species: Rachel Mason Dentinger and Abigail Woods, "Introduction to 'Working Across Species,'" *History and Philosophy of the Life Sciences* 40 (2018): 30; Carrie Friese and Adele E. Clarke, "Transposing Bodies of Knowledge and Technique: Animal Models at Work in Reproductive Sciences," *Social Studies of Science* 42, no. 1 (2012): 31–52; Miguel García-Sancho and Dmitriy Myelnikov, "Between Mice and Sheep: Biotechnology, Agricultural Science and Animal Models in Late-Twentieth Century Edinburgh," *Studies in History and Philosophy of Biological and Biomedical Sciences* 75 (2019): 24–33.

37. James W. E. Lowe, "Humanising and Dehumanising Pigs in Genomic and Transplantation Research," *History and Philosophy of the Life Sciences* (Forthcoming).

co-authorship relationships we have identified through our datasets and networks reflect only obliquely. The qualitative work that followed from our analysis of the pig network enabled us to find a spreadsheet of recipients of a Bacterial Artificial Chromosome library distributed by the Resource Center of the French Institut National de la Recherche Agronomique (INRA) station in Jouy-en-Josas. We extracted data concerning markers that the recipients mapped using the IMpRH panel. These records document the circulation of materials in a directional fashion, from the center of production to the users, and in the case of IMpRH, submission of data back to the center. They do not provide a basis for network analyses in and of themselves. However, scholars can relate or incorporate subjects, objects, and processes identified through this type of analysis to networks they have created from alternative sources.

A further area of interest is the source of DNA for incorporation in these resources and tools, or used in mapping and sequencing. Typically, co-authors disclose the strain of yeast or breed of pigs on which they conducted a sequencing study in the text of the resulting publications. Scholars investigating these practices may thus extract the DNA sources either manually from a corpus of publications, or via text-mining algorithms.<sup>38</sup> They may further incorporate the data into a network that reflects the distribution of breeds and strains, as well as different forms of work and organization of the groups involved.

### **Genomic Prediction and Commercially Sensitive Research**

The advent of tools such as microarrays or SNP chips have enabled large-scale genotyping of individuals: testing for the presence or absence of particular genomic variants. This has led to the development of methods of genomic prediction (of aspects of the phenotype from the genotype) including Genome Wide Association Studies and genomic selection. These both rely on ascertaining statistical relationships between known genetic variants (especially single base changes to DNA called single-nucleotide polymorphisms, or SNPs) and phenotypic variation. This work builds on already established genomic data, resources, and tools. In turn, it produces considerable amounts of new data, concerning variation among individuals in particular.

38. We have piloted some textual analyses of the titles and abstracts of our corpus of publications. These shed some light on the source of DNA and, more generally, the motivations behind the sequencing practices. A collaboration with bibliometric scholars using natural language processing methods could lead to more precise classifications and analyses in the future.

While archival resources may reveal the origins of the SNPs identified and included in such chips,<sup>39</sup> the actors that produce much of the data deriving from their use either have a commitment to privacy (for example, of patients), commercial confidentiality, or only making data available that is necessary for publication. These forms of data may therefore fall outside of the realms that require deposition into a public database accompanied with relevant metadata. In those cases, disclosure of funding arrangements, for instance between publicly funded research institutions and private companies, may be a revealing source of data. An additional possibility is triangulating those insights with other forms of data such as co-listing on patent applications, co-attendance at conferences, and prior working or training relationships.

As noted in the first paper of this special issue, not all of the submissions in our datasets have institutional attributions, due to some information being missing in the sequence repositories from which we extracted them.<sup>40</sup> Interestingly, a substantial number of these records list a patent instead of a publication or a submitter, especially in the yeast and pig datasets. Because of their fundamental differences with publications, we had to exclude the patents from our analysis. Patents do represent an important link between production and use of sequences, however, so exploring them may prove a fruitful future research avenue.

Mapping co-patenting relationships is possible using online resources such as “Lens.org.”<sup>41</sup> We generated an experimental patent co-application network for the pig, using the search term “scrofa.” The resulting visualization depicted two separate large clusters: one of mainly European-based researchers, the other primarily of US-based researchers. The link between the two clusters was one person, Graham Plastow, of the Pig Improvement Company, a breeding firm that had barely figured in the publication co-authorship network. This helped to make the absences in our datasets and networks starker, and prompted us to conduct an oral history interview with him.

However, patent data has its own partialities. For example, while using data on co-inventorship may help to establish networks of collaboration engaged in translating research, and links between academia and industry, prior scholarship suggests that inclusion as a co-inventor in a patent is often more exclusive

39. For example, in papers of the Pig SNP Working Group that Lawrence Schook provided to us.

40. Leng et al, “The Sequences” (n.9).

41. www.lens.org

than co-authorship. Female and junior researchers may consequently become less visible in such analyses.<sup>42</sup>

Additional sources to identify broader professional networks are *festschrifts* or other published tributes,<sup>43</sup> doctoral theses, and some grant applications. In the application to the United States Department of Agriculture (USDA) to obtain funds for pig genome sequencing, for example, each of the named collaborators had to fill out a “Conflict of Interest List.” In this document, they indicated which of the other applicants they had co-authored or collaborated with over the previous four years, which were doctoral or postdoctoral supervisors or supervisees of other applicants, and any other (especially financial) relationships they may have had.<sup>44</sup>

Conference proceedings often list attendees, as well as providing abstracts—the authors of which may not encompass all attendees, or even be attendees, though. Conferences and workshops were key to the development of genomics (e.g., the workshops on human chromosome mapping), so compiling data from these may alleviate some of the issues we have identified in the discussion above. Yet such lists of conference attendees are only partially available: informal meetings or workshops may be less well represented. Therefore, exclusive use of this kind of source may obscure certain forms of communication and collaboration in favor of others.

### 3.2. A Proposal

The challenges that the three examples above present indicate that a simple transference of the approach we have taken for DNA sequence data to other artifacts of genomic work may not be possible in most cases. This is likely to also be the case for other forms of scientific activity of scholarly interest. The iterativity of our mixed-methods approach depended in part on a relatively

42. Philippe Ducor, “Coauthorship and Coinventorship,” *Science* 289, no. 5481 (2000): 873–75; Francesco Lissoni, Fabio Montobbio, and Lorenzo Zirulia, “Inventorship and Authorship as Attribution Rights: An Enquiry into the Economics of Scientific Credit,” *Journal of Economic Behavior & Organization* 95 (2013): 49–69; Carolin Haeussler and Henry Sauermann, “Credit Where Credit Is Due? The Impact of Project Contributions and Social Factors on Authorship and Inventorship,” *Research Policy* 42, no. 3 (2013): 688–703.

43. See, for example, Jack C. M. Dekkers, Susan J. Lamont, and Max Rothschild, eds., *From Jay Lush to Genomics: Visions for Animal Breeding and Genetics*, proceedings of conference held at Iowa State University, Ames, Iowa, 16–18 May 1999.

44. Alan Archibald’s personal papers; Partition – “Pig Genome Sequencing USDA Application – August 2005,” obtained 17 May 2017.

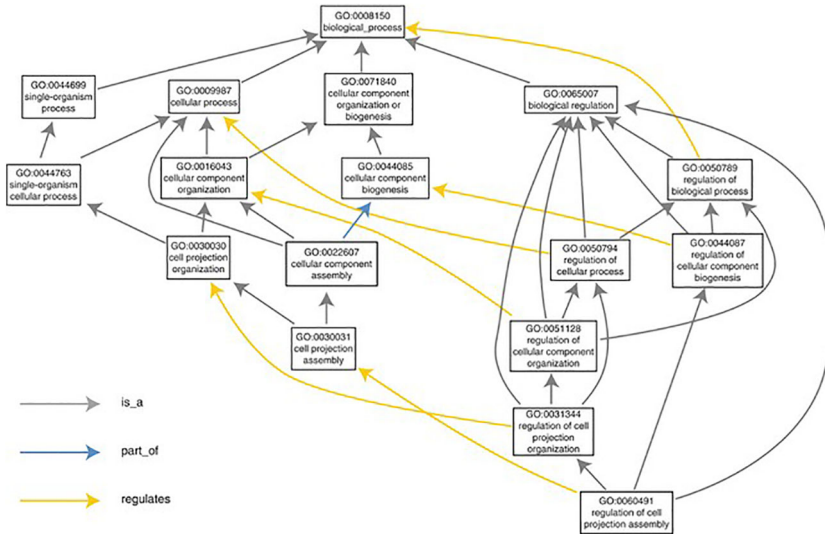
sharp demarcation between the quantitative datasets and our qualitative research. Attempting to access and use other potential forms of data may blur these distinctions rather than enabling working across them.

The generation of data always involves some processes of socio-technical construction. In our case, upstream construction was in the hands of the sequence submitters and curators of the European Nucleotide Archive. Our work, which involved making choices at multiple stages, was one of data extraction, formulation, cleaning, and analysis. Studying the forms of work we indicate above with other types of data would likely involve producing the data in the first place. This would often require deriving quantitative data from qualitative research on the transfer of materials, formation of exchange networks, results of genomic testing, and so forth. We could gather some data, for example on citation networks, in a less qualitative way.

There are two potential problems with these types of newly constructed datasets. One is that we would be overly dependent on pre-specifying a particular object of inquiry and therefore foreclosing the possibilities open to an investigation. Using sequence submissions also falls into that trap to some extent, but as sequencing is a very basic form of work that a variety of actors conduct and use for a myriad of purposes, it does not channel further inquiry that restrictively. The second problem is that these alternative datasets would be far less extensive than the DNA sequence submissions and publications, at least concerning the period after funding and journal policies made submission to public databases mandatory.

A way of addressing these problems is to use new and existing datasets in combination with each other. The creation and curation of a domain ontology, much like the ones already available in the natural sciences, would enable such a use. Domain ontologies are ways of organizing and representing different forms of data concerning a particular kind of object or phenomenon. Gene Ontology (see figure 3) and the Mammalian Phenotype Ontology are examples of them in genetics research.<sup>45</sup> Rather than the nodes of our network visualizations, different types of actors and entities would feature as objects in the ontology, with attributes assigned to them, and different kinds of relations assigned among them.

45. <http://geneontology.org> and <https://www.ebi.ac.uk/ols/ontologies/mp>



**FIGURE 3.** Extract of a set of objects and different relations depicted between them in the Gene Ontology. Figure reproduced from Pascale Gaudet, Nives Škunca, James C. Hu, and Christophe Dessimoz, “Primer on the Gene Ontology,” in *The Gene Ontology Handbook*, eds. Christophe Dessimoz and Nives Škunca (New York: Springer, 2017), 25–37, on 26. This work is licensed under CC-BY-4.0 (<https://creativecommons.org/licenses/by/4.0>). We reproduce the image without modification.

There would be several advantages to this approach. As with the natural science ontologies, our proposed ontology could incorporate different forms of data, since the objects do not need to be of the same kind, nor the relations equivalent. An individual person may be an object; so may an institution. An individual person could have the relations of “was a PhD student of” another person, “worked at” an institution, or “authored/co-authored” a given publication. Provided there was an agreed curatorial basis for categories and assignments, and a process that allowed the incorporation of new data, the ontology could gradually grow and incorporate one small qualitatively derived dataset (or, indeed, quantitatively derived dataset) at a time.

Scholars and other users could then derive quantitative datasets and networks by extracting only those sets of objects, attributes, and relations pertinent for their purposes. Such an ontology would enable the addition and commensuration of data from sources that would be problematic to integrate outside the ontology, such as some of the rich databases with different curatorial standards we listed in section 3.1. It would also ensure that species are not an organizing principle of inquiry; the identity of species (and subspecific



variants) that individuals and institutions investigate and compare with others in publications and projects would be included, but only as one class sharing the ontology space with many others.

Though never likely to be exhaustive or complete, such an ontology would provide the extent and coverage missing in small individual datasets, and provide multiple independently derived sources to correct for the partiality of any one dataset. Such an ontology may originate in an individual project but end up becoming a shared resource that other groups may use and enrich, in much the same way we did with our datasets.<sup>46</sup> Much like our datasets, networks, and qualitative evidence, such an ontology would not constitute an end-point but rather an opening to the generation of further questions, observations, and lines of investigation.<sup>47</sup> With every addition of new data and relations, the patterns evident in the ontology would change. Further, scholars could connect this ontology to others and create an ecosystem of interlinked ontologies, again as in the natural sciences.

Table 1 presents an initial list of ontology classes (categories of objects) and relations based on the work we have conducted so far.

**TABLE 1.** Indicative List of Classes and Relations with Which a History of Genomics Ontology Could Begin to Be Populated\*

<b><i>Classes (indicative objects that are members of the class)</i></b>	<b><i>Relations</i></b>
Person	
Researcher (Lap-Chee Tsui)	Was a PhD student of
Administrator/official (André Goffeau)	
Institution	
Sequencing center (Sanger Institute)	Worked at
University (Ludwig-Maximilians-Universität München)	Did PhD at
Research institute (Roslin Institute)	
Medical school (Harvard Medical School)	
Small-scale sequencing company (Genotype)	
Large-scale sequencing company (Celera)	
Funding body (Wellcome Trust)	Received funding from

(continued)

46. The publication and submission data on which we based our co-authorship networks are available to the community without restrictions at <https://datashare.is.ed.ac.uk/handle/10283/3517>.

47. And, potentially, the integration and cross-pollination of research silos: Sabina Leonelli, "Bio-ontologies as Tools for Integration in Biology," *Biological Theory* 3, no. 1 (2008): 7–11.

TABLE 1. (continued)

Classes (indicative objects that are members of the class)	Relations
Methods and techniques (Radiation hybrid mapping)	Used the method/technique of
Project (Yeast Genome Sequencing Project)	Participated in Was named member of Was on steering committee of
Map (The PiGMaP consortium linkage map of the pig, <i>Sus scrofa</i> )	Mapped
Reference sequence ( <i>Saccharomyces cerevisiae</i> S288c RefSeq Genome: <a href="http://www.ncbi.nlm.nih.gov/genome/15?genome_assembly_id=22535">www.ncbi.nlm.nih.gov/genome/15?genome_assembly_id=22535</a> )	Sequenced Annotated Outcome of
Gene (human <i>CFTR</i> —cystic fibrosis gene)	Sequenced Isolated/characterized
Gene polymorphism(s) ( <i>IGF2</i> Q mutation in <i>Sus scrofa</i> )	Sequenced Cataloged
Publication (PubMed ID: 12690205)	Authored/co-authored Cited by Outcome of
Meeting/conference (Human Gene Mapping Workshop 9)	Organized/ co-organized Attended
Species ( <i>S. cerevisiae</i> )	Worked on
Subspecies population/strain (S288C strain of <i>S. cerevisiae</i> )	Worked on

\*We have indented subordinate classes below the class of which they are a part. Examples of ontology objects are provided in parentheses for classes and subordinate classes.

It is crucial for us to approach such an ontology—or federation of ontologies—critically and with a view to its limitations, in the same way as we have done with our dataset and networks. The classes and relations are discrete categories that must be sufficiently inclusive or indicative of a large enough number of individuals to be meaningful but also specific enough to be useful and not connote an excessive and unwieldy population. This means that only some of the classes and relations of potential historical interest will be chosen for inclusion in the ontology. Such selectivity poses the risk of reifying classes and relations as being particularly significant to the exclusion of others that may be salient but difficult to include within the logic of an ontology. As

Hallam Stevens observes with respect to biological ontologies, such architectures require the creation of “standard, computable objects” and consequently flatten some aspects of biology into data. This does not always comport with the ways biologists conceive the objects and concepts they explore, so we could conjecture that this may be as problematic—if not more—for historians and social scientists.<sup>48</sup>

An ontology requires continuous updating of its architecture, the classes and relations it includes, new additions of individual objects, and the introduction of fresh relations.<sup>49</sup> However, the possibility remains of it presenting a static rather than dynamic picture. We have described and evaluated our efforts to historically animate datasets and networks that represent the totality of a particular time period. Our closeness to the processes underlying the construction of the datasets and networks, and our ability to play with them, to partition them, and to make sense of them in conversation with our ongoing qualitative research, considerably aided this dynamization. For those who make small contributions to an ontology without an appreciation of its construction and evolution, or who access the data to gain an insight into genomics without previously immersing themselves in it, such an animation of the contents may be more difficult to achieve.

Furthermore, the kinds of classes and relations that ontologies portray do not allow for ambiguous or nuanced interpretation, specificity, or items that may exist on a continuum. Those items included must be recognized matters of fact, and this may lead to particular kinds of sources being favored over others, affecting what is represented in the ontology, and therefore the direction of research taken by users of it. As Stevens contends for biological ontologies—an argument that is, again, applicable to history and the social sciences—“the standardization and data-ization” that these tools foster and the ways they portray both objects and relations affect how researchers “think” and “what they do” with their evidence.<sup>50</sup>

48. Stevens, *Life* (n.7), chap. 5, quote on 108.

49. The revisability of ontologies does not merely entail an accumulation of data, however, but can reflect a shift in understanding or theory: Stevens, *Life* (n.7), 126. On the notion that the ongoing development of life science ontologies constitutes a form of theorizing, see Sabina Leonelli, “Classificatory Theory in Data-Intensive Science: The Case of Open Biomedical Ontologies,” *International Studies in the Philosophy of Science* 26, no. 1 (2012): 47–65.

50. Stevens, *Life* (n.7), 127. For a similar concern with the source material for seeding ontologies, see William Bechtel, “Using the Hierarchy of Biological Ontologies to Identify Mechanisms in Flat Networks,” *Biology & Philosophy* 32 (2017): 627–49.

There are also practical issues concerning the proposal, curation, and verification of classes and relations, the ongoing entering of data, evaluation of its quality, general maintenance, and promotion of ontologies as tools for historical research. While new ontologies in the life sciences can draw upon well-established sets of classes and relations—as well as norms of management and design—these may not be applicable or appropriate for the humanities and social sciences. Here, different attitudes to the realism of categories and objects pertain, for which capturing change, contingency, and contextuality are just as important as, if not more important than, establishing a canonical body of consensus knowledge.<sup>51</sup> Cultivating the appropriate set of norms, practices, and content for ontologies in the history of science must therefore be a creative, collaborative, and reflexive endeavor.

For the construction of an ontology such as this, as well as for the deployment of more ambitious mixed-methods approaches, historical research will need to be more interdisciplinary, networked, and team based. Jamborees analogous to those that genomics researchers promoted could serve to populate the ontology *in situ*, or else train and network a cadre of contributors and curators.<sup>52</sup> While this would be a novel departure, ontology-related responsibilities are not radically distinct from the roles historians already take on as peer reviewers or journal editors, if less established. Pioneers should create a process to identify absences and flag them, and build scope for the overall ontology to evolve, analogous to the way life science initiatives such as the Gene Ontology have.<sup>53</sup> This mode of working could foster more fluid communication across the humanities and social and natural sciences. Access to computational methods of analysis may help historians to use the large datasets of natural scientists as evidence and

51. Though for a discussion of ways of reaching practical consensus in spite of theoretical and/or empirical disagreements among scientific practitioners about the classification of objects and their relations, see Beckett Sterner, Joeri Witteveen, and Nico Franz, “Coordinating Dissent as an Alternative to Consensus Classification: Insights from Systematics for Bio-ontologies,” *History and Philosophy of the Life Sciences* 42 (2020): article 8. On implementing an ontology for evolutionary biology, a historical science, see Francisco Prosdocimi, Brandon Chisham, Enrico Pontelli, Arlin Stoltzfus, and Julie D. Thompson, “Knowledge Standardization in Evolutionary Biology: The Comparative Data Analysis Ontology,” in *Evolutionary Biology: Concept, Modeling, and Application*, ed. Pierre Pontarotti (Berlin and Heidelberg: Springer-Verlag, 2009), 195–214.

52. On these two models of annotation jamborees, see García-Sancho and Lowe, *A History* (n. 6), chap. 6.

53. Sabina Leonelli, Alexander D. Diehl, Karen R. Christie, Midori A. Harris, and Jane Lomax, “How the Gene Ontology Evolves,” *BMC Bioinformatics* 12 (2011): article 325.

mobilize different narratives around them, thus potentially influencing science policy problems such as those concerning the translation of research results.

#### 4. CONCLUSIONS

This concluding essay has offered a synoptic reflection on the datasets and networks we analyzed separately through the special issue. We have done this by looking at the wider implications of our mixed-methods approach—within and beyond the historiography of genomics—and identifying other possible data sources and visualizations to both continue our investigations of genomics and extend them to other cognate scientific fields. Both the co-authorship networks we have created and the ontology we propose for the future offer an image of genomics as a complex ecosystem of institutions of different kinds, sequencing DNA for different purposes, and displaying different organizational models depending on the characteristics of the communities involved, among them medical geneticists, cell biochemists, immunologists, and agriculturally inclined geneticists. This image contrasts with the usual portrayal of genomics in both academic literature and participants' accounts as a unified field organized according to the HGP framework.

Our more diverse portrayal has consequences for the ongoing historiographical attempts at placing genomics within longer-term temporal frameworks in the history of the life sciences. The HGP-centric vision derives from a *thin* understanding of genomics as a field that both gathered and extended the progress that molecular biology had achieved throughout the second half of the twentieth century. In this vision, the historical endeavor of genomics built on the accumulated knowledge and methods of molecular biology, mobilizing existing techniques to produce DNA sequence data, especially about the human genome and to a lesser extent about model organisms that would help in understanding *Homo sapiens*. This narrative implied that other disciplines would use the data to answer their ongoing research questions in a novel fashion.

Our networks offer a less linear and more web-like image of genomics. Historical interpretation of these visualizations and their underlying data shows a variety of connections irreducible to molecular biology or the mere transition from large-scale sequence production to downstream use. In our web-like model (see figure 2, above), a plurality of actors generated different outcomes of which the HGP and the genome centers were just one example. They had different motivations to sequence and, accordingly, conducted their

work in distinct ways—horizontally or vertically, intensively or extensively—and showed different degrees of involvement with the resulting data: more proximate or distal, more directed or undirected. More importantly, these actors both produced DNA sequences and used them to tackle a variety of agricultural, medical, and cell biological problems. Rather than addressing a translational gap that the prior production of a human or other species' reference genome had created, they were translating sequence data as they produced it either in isolation or as they collectively contributed to a fuller description of their target genome.

From this thicker historiographical standpoint, genomics resembles more a set of tools entangled with existing life sciences disciplines and less a field of its own. Operating outside of the prominent knowledge-control regime of the genome centers and accompanying centralized infrastructures, these existing disciplines offer the possibility for new appreciations of the impact of genomics on the life sciences and the worlds they touch.

Considering genomics as an imperial enterprise in which the core—tellingly located in the United States and the United Kingdom, and increasingly in new centers of power such as China—dominates the periphery and transforms it in its own image, has been a fruitful path for scholarship over the course of three decades. But in some respects, it accepts too readily the maps of the new world of the life sciences that imperial genomics advocates have charted. An alternative is to consider genomics as a new element added to existing and continuing practices and research programs in the wider life sciences, in which the history and sociology of these areas, and their ongoing orientation toward particular problems, conditions any transformative potential.<sup>54</sup> In this way, the historiography of genomics can take a turn long predated by colonial and imperial histories, which have sought to interpret how the overwhelming forces of colonization and imperialism nevertheless confronted societies and cultures that dealt with the new demands and influences in distinct ways, and additionally affected the core and its imperial projects in unexpected fashions.<sup>55</sup>

54. In this respect, the extent to which genomics functions as a *supplementation* of these existing research systems is an open question; we use the sense of supplementation as mobilized in Hans-Jörg Rheinberger, *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube* (Stanford, CA: Stanford University Press, 1997), 4.

55. For general surveys that discuss some of these historiographical developments in terms of forms of connectivity and spatial configurations, see Alan Lester, "Imperial Circuits and Networks: Geographies of the British Empire," *History Compass* 4, no. 1 (2006): 124–41; Sanjay Subrahmanyam, *Explorations in Connected History: From the Tagus to the Ganges* (Oxford: Oxford

Policies that seek to improve the medical translation of sequence data have identified the “infusion of genomic methods and approaches across the life sciences” as a key accomplishment.<sup>56</sup> In this special issue, we have shown that this infusion was not just a new strategy that followed the completion of the human and other species’ reference sequences, marking the transition from genomics to post-genomics research. Our perspective of genomics as a set of tools beyond the HGP and the concerted production of a reference sequence at large-scale centers implies that the infusion of genomic technologies and the use of genomic data in the life sciences were, to a large extent, ongoing processes that did not need specific post-genomic policies. Rather than a new problem that arose after the field of genomics completed its historical endeavor—to produce and freely disseminate reference sequences—the translation of this information was and still is an ongoing process in genetics, cell biological, and systematics research laboratories that adopted genomic technologies and both produced and used sequence data.

Translation is a nonlinear, recursive, and dynamic process that requires careful attention to the ways in which actors generate, package, mobilize, deploy, and integrate data to produce fresh knowledge claims of practical import.<sup>57</sup> Considerable recent work has established how data enter into infrastructures where other researchers retrieve and re-use them in a variety of

---

University Press, 2005). An example of this kind of history is a wide-ranging account that tries to identify different forms of encounter between indigenous peoples and imperial power: Philip D. Morgan, “Encounters between British and ‘Indigenous’ Peoples, c. 1500–c. 1800,” in *Empire and Others: British Encounters with Indigenous Peoples, 1600–1850*, eds. Martin Daunton and Rick Halpern (London: UCL Press, 1999), 42–78. In numerous cases, whole societies and cultures have been destroyed, and were therefore unable to exercise even the constrained agency and resistance that such works foreground. In deploying this imperial metaphor to characterize genomics, we do not want to imply that the dominance of a particular organization or mode of scientific practice is in any way analogous to colonial violence in its effects on individuals and communities. The absence of exploitation and the limitations of conceiving of genomics as the core and the rest of the life sciences as a periphery adds to the disanalogy.

56. Eric D. Green, Chris Gunter, Leslie G. Biesecker, Valentina Di Francesco, Carla L. Easter, Elise A. Feingold, Adam L. Felsenfeld, et al., “Strategic Vision for Improving Human Health at The Forefront of Genomics,” *Nature* 586, no. 7831 (2020): 683–92, on 690.

57. Jamie Lewis, Jacki Hughes, and Paul Atkinson, “Relocation, Realignment and Standardisation: Circuits of Translation in Huntington’s Disease,” *Social Theory & Health* 12, no. 4 (2014): 396–415; James W.E. Lowe, Sabina Leonelli, and Gail Davies, “Training to Translate: Understanding and Informing Translational Animal Research in Pre-Clinical Pharmacology,” *TECNOSCIENZA: Italian Journal of Science and Technology Studies* 10, no. 2 (2019): 5–30.

different, perhaps unconceived, contexts.<sup>58</sup> This scholarship has addressed so-called data-centric biology as a counterpoint to more traditional forms of biological research in which the quantity of data generated is relatively small and targeted toward the particular aims of a given observational or experimental study. James Griesemer, however, identifies an intermediate range of practice between these: that of the “datapoint-centric” (his emphasis).<sup>59</sup>

To paraphrase his argument in the conceptual terminology we have used in this special issue, datapoint-centric biology involves more proximal and/or directed sequencing, with data practices oriented toward concrete research purposes and questions, rather than primarily satisfying the data-centric requirement to make data mobilizable and interoperable, standardized, and in bulk. We have touched on some areas of research that fall under this datapoint-centric umbrella: the cataloging of variants by the Cystic Fibrosis Genetic Analysis Consortium that we discussed, for example.<sup>60</sup>

Much of what we have talked about appears to lie at the intersection of data-centric and datapoint-centric science. At this intersection, we see researchers working to create and adapt genomic tools and data to contribute to the tackling of their own research problems or to open up new ones. The network of European yeast sequencers, the medical geneticists teaming up with Celera, and the consortium of pig geneticists working with the Sanger Institute all sought to exploit and interact with the institutions and infrastructures of data-centric biology and its fruits. To do so required new organizational configurations and ways of working to establish connections and traffic at these nexuses. In forging these connections, the yeast, pig, and human life scientists we discussed diverged from the dominant large-scale center model of genomics research. Their forms of organization related to, took from, and contributed toward a data-centric world they did not create or control but that nevertheless provided opportunities as well as shaped the scientific environment around them in ways they had to adapt to. We suggest that in focusing on this intersection and how institutions and communities creatively constructed, managed, and navigated it, we have contributed insights into processes of translation and, in some cases, how they can stutter.

58. Leonelli, *Data-Centric* (n.7). Mobility and interoperability are key elements of this: Sabina Leonelli, “Learning from Data Journeys,” in *Data Journeys in the Sciences*, eds. Sabina Leonelli and Niccolò Tempini (SpringerOpen, 2020), 1–24.

59. James Griesemer, “A Data Journey through Dataset-Centric Population Genomics,” in *Data Journeys in the Sciences*, eds. Sabina Leonelli and Niccolò Tempini (SpringerOpen, 2020), 45–167.

60. García-Sancho et al., “The Human Genome Project” (n.10).



All this suggests that, as much as proposing future strategies, translational research policies also need to look at the past and reappraise the narratives and disciplinary status of genomics. The genealogies we have uncovered in this special issue, along with the analytical categories and datasets accompanying them, contribute to this much needed historical and critical reassessment of policymaking. At a scholarly level, they also provide the basis for thickening the history of genomics through augmenting the increasingly well-established genealogies underlying it with an approach that emphasizes the ramification of its practices and outputs: the mobilization, re-articulation, and embedding of genomics across the life sciences and beyond.

### Acknowledgments

See a full list of people and institutions whose support has been essential at the end of the introductory article of this special issue "The Sequences and the Sequencers: What Can a Mixed-Methods Approach Reveal about the History of Genomics?" The research and writing of this paper were sponsored by the "TRANSGENE: Medical Translation in the History of Modern Genomics" Starting Grant, funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, grant agreement No. 678757. For more details on the project, see <https://transgene.sps.ed.ac.uk/>