

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

Yeast Sequencing: “Network” Genomics and Institutional Bridges

ABSTRACT

This paper examines the model of *network genomics* pioneered in the late 1980s and adopted in the European Commission-led Yeast Genome Sequencing Project (YGSP). It contrasted with the burgeoning large-scale center model being developed in the United States to sequence the yeast genome, chiefly as a pilot for tackling the human genome. We investigate the operation and connections of the two models by exploring a co-authorship network that captures different types of sequencing practices. In our network analysis, we focus on institutions that bridge both the European and American yeast whole-genome sequencing projects, and such concerted projects with non-concerted sequencing of yeast DNA. The institutions include two German biotechnology companies and Biozentrum, a research institute at Universität Basel that adopted yeast as a model to investigate cell biochemistry and molecular biology. Through assessing these bridging institutions, we formulate two analytical distinctions: between proximate and distal, and directed and undirected sequencing. Proximate and distal refer to the extent that intended users of DNA sequence data are connected to the generators of that data. Directed and undirected capture the extent to which sequencing was part of a specific research program. The networked European model, as mobilized in the YGSP, enabled the coexistence and cooperation of institutions exhibiting different combinations of these characteristics in contrast with the more uniformly distal and undirected large-scale

*Miguel García-Sancho, Science, Technology and Innovation Studies, University of Edinburgh, Old Surgeons' Hall, High School Yards, Edinburgh, EH1 1LZ, United Kingdom. miguel.gsancho@ed.ac.uk

The following abbreviations are used: DKFZ, Deutsches Krebsforschungszentrum (German Cancer Research Center); DNA, deoxyribonucleic acid; EMBL, European Molecular Biology Laboratory; GATC, Gesellschaft für Analyse-Technik und Consulting; LMUM, Ludwig-Maximilians-Universität München; MIPS, Martinsried Institute for Protein Sequences; SDSTC, Stanford DNA Sequencing and Technology Center; tRNA, transfer RNA; YGSP, Yeast Genome Sequencing Project.

Historical Studies in the Natural Sciences, Vol. 52, Number 3, pps. 361–400. 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.2022.52.3.361>.

centers. This contributes to broadening the historical boundaries of genomics and presenting a *thicker* historiography, one that inextricably meshes genomics with the trajectories of biotechnology and cell biology. 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, cell biology, *Saccharomyces cerevisiae*, yeast, biotechnology companies

1. INTRODUCTION

In 1996, the baker's and brewers' yeast *Saccharomyces cerevisiae* became the first eukaryotic organism to have its whole genome sequenced. The article describing this sequence appeared a year later in *Nature* and included over 600 authors based in more than 80 institutions.¹ This achievement was the outcome of various national and international projects sponsored by European, Japanese, Canadian, and US funding agencies.²

Historians and social scientists have focused on the contrasting sequencing strategies in Europe and the United States. The strategy that the European Commission (EC) promoted materialized in a concerted Yeast Genome Sequencing Project (YGSP) that started in 1989 and concluded in 1996.³ Its objective was to expand Europe's industrial and scientific capacity through the coordinated sequencing activity of a consortium of universities, research institutes, breweries, and start-up companies from the member-states of the European Union (until 1993, the European Community).⁴ The European

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

2. André Goffeau, Bart G. Barrell, Howard Bussey, Ronald W. Davis, Bernard Dujon, Horst Feldmann, Francis Galibert, et al., "Life with 6000 Genes," *Science* 274, no. 5287 (1996): 546–67; Mark Johnston, "The Yeast Genome: On the Road to the Golden Age," *Current Opinion in Genetics & Development* 10, no. 6 (2000): 617–23.

3. On the origins and planning of the YGSP, see Andre Goffeau, "Yeast Transport-ATPases and the Genome-sequencing Project," in *Comprehensive Biochemistry*, vol. 43, eds. G. Semenza and A.J. Turner (Amsterdam: Elsevier, 2004), 493–596, especially 493–94, 514–29; Philippe Goujon, *From Biotechnology to Genomes: The Meaning of the Double Helix* (Singapore: World Scientific Publishing, 2001), chaps. 6 and 7.

4. Giuditta Parolini, *Building Human and Industrial Capacity in European Biotechnology: The Yeast Genome Sequencing Project (1989–1996)* (Berlin: Technische Universität Berlin, 2018). https://depositonce.tu-berlin.de/bitstream/11303/7470/4/parolini_guiditta.pdf. The European Commission is the executive branch of the European Union, and includes the civil service organized into directorates-general responsible for different areas of policy and administration.

project distributed relatively small portions of yeast DNA (cosmid clones) among many laboratories, in line with the attempt by the EC to strengthen integration among member-states. Actors on the US-side dubbed it the “cottage industry” approach.⁵

The US National Institutes of Health (NIH) lacked this integrationist drive and sought instead to maximize efficiency in a smaller number of highly productive genome centers. Unlike the European institutions, these centers were not primarily involved in yeast research, nor were they interested in further exploiting the resulting sequence. Instead, the NIH saw yeast as a suitable pilot project in which to evaluate the technologies for sequencing the human genome. Following his appointment as associate director of the NIH Office for Human Genome Research in 1988, James Watson inaugurated two genome centers at Stanford and Washington University in St Louis, the latter of which had led the development of the yeast physical genome map in the second half of the 1980s. These institutions, along with the Sanger Institute that the Wellcome Trust founded in the United Kingdom in 1993, comprehensively sequenced full yeast chromosomes as preparation for tackling the human genome. McGill University and the Riken Institute also led projects in Canada and Japan to sequence full yeast chromosomes.⁶

The historical literature has shown how the sequencing projects emerged within a preexisting, tightly connected yeast genetics community. From the mid-twentieth century, this community established a specific yeast strain (S288C) as the standard for conducting genetic experiments. Research groups from all around the world shared and exchanged this strain, which became the basis of the genetic and physical maps that the sequencing initiatives subsequently used.⁷ One of the participants in the NIH genome

5. Pierre-Benoît Joly and Vincent Mangematin, “How Long Is Co-operation in Genomics Sustainable?” in *The Social Management of Genetic Engineering*, eds. Peter Wheale, Rene von Schomberg, and Peter Glasner (Aldershot: Ashgate, 1998), 77–90; James D. Watson, “The Human Genome Project: Past, Present, and Future,” *Science* 248, no. 4951 (1990): 44–49, 45.

6. Erika A. Szymanski, Niki Vermeulen, and Mark Wong, “Yeast: One Cell, One Reference Sequence, Many Genomes?,” *New Genetics and Society* 38 (2019): 430–50. Overall, European institutions contributed the largest volume of sequence to the published version of the yeast genome. The Sanger Institute, Stanford, and Washington University were the largest individual sequencers.

7. Erika M. Langer, “Molecular Ferment: The Rise and Proliferation of Yeast Model Organism Research” (PhD Thesis, University of California, San Francisco, 2016); Szymanski et al., “Yeast” (n.6). On similar community efforts around other model organisms, 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.

project has suggested that the sequencing enterprise was the final realization of the endeavor that pioneering yeast mapper Robert Mortimer and other colleagues started a half-century earlier on the sixteen *S. cerevisiae* chromosomes.⁸

Prior scholarship has thus placed yeast genomics within a broader research tradition and identified different sequencing strategies. Yet this literature has not fully overcome the geographically fragmented recollections of participants, as well as the limited knowledge—and documentation—of how the European and NIH initiatives coordinated their activities and interacted with other yeast sequencing efforts. The differentiated roles of the many laboratories involved in the European consortium is also a lacuna in the literature.

We address some of these shortcomings by using a new body of evidence: a network that visualizes how institutions co-authored articles that described for the first time in the published record newly determined yeast sequences submitted to the European Nucleotide Archive, the DNA Data Bank of Japan, or GenBank in the United States between 1980 and 2000. As with our analyses on *Homo sapiens* and the pig *Sus scrofa*, we generated the network and gathered the underpinning data, for which we chose a collection window to capture sequencing activity before, during and after concerted yeast genome sequencing projects.⁹

The sequencing project that the EC sponsored can be easily traced in the network, and this makes it different from other visualizations we analyze in the special issue. However, as in the human and pig networks, the yeast co-authored publications show a great deal of sequencing activity that occurred outside efforts to determine a reference genome. In the case of the yeast network, a substantial portion of the co-authors of the sequences described in the scientific literature were based in laboratories interested in biochemistry and cell biology.

We address this sequencing activity through the identification and analysis of institutions that bridge different modes of sequencing in the network: between different concerted genome projects, and between those concerted

8. Mark Johnston, “The 2002 George W. Beadle Medal Robert Mortimer and André Goffeau,” *Genetics* 164, no. 2 (2003): 422–23.

9. For a full description of the method and dataset underlying the network visualization, see 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.

projects and other areas of biological research. Due to the relatively small size of the yeast network, we were able to examine inter-cluster relationships through assessing an easily recognized set of bridging institutions.¹⁰ In social network analysis, a bridge is a connection between two groups: scholars in this field have characterized bridges as “mediators” or “brokers” that lie between different parts of the network that themselves reflect different forms of work, organizational norms, historical trajectories, geographical locations, or other factors structuring collaboration.¹¹

By seeking and providing qualitative explanations as to the origin of network bridges we offer an additional historiographical bridge to link different stories; in so doing, we present a bigger picture of yeast genomics and its role within biological research on *S. cerevisiae*. The focus of our analysis will be on two types of bridging institutions: (1) a number of German biotechnology start-up companies specializing in DNA sequencing services, and (2) Biozentrum, a Swiss contributor to yeast genome sequencing. The German companies participated in sequencing operations led by both the EC and US genome centers, as well as providing sequence data to institutions working on the biology of yeast. Biozentrum, partly because of its special funding arrangements and partly because of its institutional history, is not positioned within the tightly connected network cluster of European institutions involved in the YGSP, despite having led the sequencing of one chromosome with other members of the EC consortium.

The focus on these institutions will enable us to show something that, despite being known, is not sufficiently detailed in the literature (at least not in the English language): that German—and, more generally, central European institutions—were key players in the early years of genomics. We argue that this crucial and often overlooked role was largely based on bridging different sequencing efforts and sometimes bridging genomic science with other disciplines. The German start-up companies, particularly Gesellschaft für Analyse-Technik und Consulting (GATC) and Genotype, also present a different historical genealogy of biotechnology compared with prominent European and US companies such as Biogen or Genentech, on which scholarly

10. The connections underpinning the clusters we identify in the yeast network are far fewer than those in the pig and, especially, the human network. We therefore assess them with the qualification that the extent to which they represent communities may be weaker in some cases, compared with the other networks we analyze in this special issue.

11. Ronald S. Burt, “Structural holes and good ideas,” *American Journal of Sociology* 110 (2004): 349–99.

studies have focused.¹² The German firms provided sequencing services and occasionally technologies relating to these rather than focusing on the commercialization of new pharmaceuticals or the provision of biological materials such as enzymes. They consequently developed business models tied to their role in public initiatives such as the European YGSP.

Our examination of the bridging institutions allows us to argue that the YGSP embodied a varied ecology of sequencers. The institutions within the EC consortium, while collectively producing a whole reference genome, displayed distinct modes and motivations to sequence. In some instances, these institutions made use of the data they contributed to the YGSP for their own research. In others, they collaborated with other laboratories that used the sequences for investigating yeast biology. This diversity of sequencers and dual orientation of the sequencing—seeking to simultaneously delineate the whole genome as well as further explore the biology of the yeast cell—was a feature of the network model of genomics. The networked organization of the YGSP meshed with the more general drive toward the establishment of research networks that was characteristic of EC-funded initiatives of the 1990s.¹³

In the United States, the large-scale center model that Watson promoted for genomics was not intended to involve sequencing taking place in smaller laboratories. In Europe, however, whole-genome directed sequencing could include this kind of institution. The EC did not create specialist sequencing centers, and the new practice and organization of reference genome sequencing did not displace existing laboratories: rather, it enriched the ecology of them. This European network model of genomics did come to be occluded by the large-scale center model. Though both constituted genomics, a victor's history might only—or mainly—take the latter into account.

To characterize the more complex ecology of the networked European model of yeast genome sequencing—and in line with our investigations of human and pig genomics—we move beyond a simple distinction of sequence

12. Sally Smith Hughes, *Genentech: The Beginnings of Biotech* (Chicago: The University of Chicago Press, 2011); Nicolas Rasmussen, *Gene Jockeys: Life Science and the Rise of Biotech Enterprise* (Baltimore: Johns Hopkins University Press, 2014); Doogab Yi, *The Recombinant University: Genetic Engineering and the Emergence of Stanford Biotechnology* (Chicago: University of Chicago Press, 2015).

13. For example, see this report from the mid-1990s: Luca Guzzetti, *A Brief History of European Union Research Policy* (Brussels: European Commission Directorate-General XII, 1995).

producers and users.¹⁴ Due to the nature of the European YGSP, and the institutions identified through our network analysis, we have been able to provide examples and a fine-grained gradation of different levels of entanglement between sequence production and use: yeast sequencers that at the same time of producing data became, to a greater or lesser extent, users of that data in contrast to the more monolithic and decoupled production regime that the large-scale center model predicated.

We detail the links between actors in the YGSP and between the YGSP and external institutions using two pairs of categories: proximate and distal sequencing, and undirected and directed sequencing. By proximate and distal, we mean the extent to which the producers of DNA sequence data know in advance and are connected to the users of that data. Our distinction between undirected and directed sequencing is based on the extent to which a given actor performs sequencing to inform their research program in either an exclusive or nonexclusive fashion. Directed sequencing involves the immediate use of sequence data as part of a particular research program that shaped the production of that data, regardless of the producer also sharing the sequence with others or retaining it. Undirected sequencing, conversely, does not entail performing this activity with the aim of using the resultant data for specific and known purposes. The extent to which a given institution undertakes more undirected or directed sequencing, therefore, depends on the motivation to sequence, and the subsequent use of that sequence data.

Through examining the sequencing practices of our three bridging institutions and how they fit into a particular organizational model of producing, assembling, annotating, and distributing DNA sequence data, we can pinpoint the crucial differences between the European model of yeast genomics and the US model. There were differences in terms of motivations and aims, reflecting distinct institutional and political drivers on either side of the Atlantic. While the potential economic, social, and medical benefits of the Human Genome Project—for which the US yeast sequencing effort was a pilot—were hazy and

14. 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; James Lowe, Rhodri Leng, Gil Viry, Mark Wong, Niki Vermeulen, and Miguel García-Sancho, “The Bricolage of Pig Genomics,” this issue. The producer–user distinction arose as an actor’s category from the context of research policy, funding, and organization. From there, it has informed studies of genomics—especially human genomics—such as Stephen Hilgartner’s, even as it came into question through them: Stephen Hilgartner, *Reordering Life: Knowledge and Control in the Genomics Revolution* (Cambridge, MA: MIT Press, 2017).

conceived to lie in the future, the payoff that the EC envisaged for the YGSP was more immediate and concrete. As well as strengthening larger and more established businesses, the sequencing of *S. cerevisiae* was intended to help foster newer and smaller biotechnology companies, and to build competitiveness and collaborative connections across Europe.

GATC, Genotype, and Biozentrum all operated distinctly from the large-scale sequencing centers that the NIH founded in the United States to tackle the full yeast—and human—genomes. GATC and Genotype engaged in proximate, undirected sequencing. They positioned themselves and functioned as service providers who knew the users of those services: the European project and, often, other institutions that would apply the sequence data to cell biological and biochemical problems. Biozentrum engaged in both proximate and directed sequencing, with their sequencing being a progression of an existing program of “applied microbiology,” a term that the yeast chromosome coordinator based there, Peter Philippsen, used in a specific sense.¹⁵

In contrast, the sequencing practices of the large-scale centers that led the US yeast genome project at Stanford and Washington University were both undirected and distal from the potential users of the data.¹⁶ The general aim of yeast genome sequencing in the United States was the *eventual* translation of the sequence—to help complete the human genome as much as advance yeast biology—rather than more immediate, proximate use. This meant that the specificity of the data produced was almost never in accordance with the specific yeast-related research goals or needs of any particular recipient.

These different forms of sequencing produced data with different roles within genomics, and with distinct implications for the wider life sciences. In the European network model, sequence data constituted a means toward known—or at least, clearly conceived—research ends. The sequences were there to serve biology and the members of the EC consortium constructed them to enable the data to feed into research programs, either directly or

15. Peter Philippsen, telephone interview with Miguel García-Sancho and James Lowe, 20 September 2019. On the way Philippsen used this term, see section 4 of this paper.

16. In some instances, as Hilgartner has shown for human genomics, the distancing between the large-scale centers and the community of yeast biologists occurred gradually. This was the case for the Genome Sequencing Center at Washington University, which in its early years exhibited a relationship that was more proximal in nature with Mark Johnston, a researcher interested in yeast glucose metabolism who ended up as a leader in *S. cerevisiae* genome sequencing: Miguel García-Sancho and James W.E. Lowe, *A History of Genomics Across Species, Communities and Projects* (Palgrave Macmillan, forthcoming), chap. 2; Hilgartner, *Reordering*, (n.14), chaps. 3 and 4.

through connection to other yeast biology resources to aid the wider community. The YGSP laboratories, therefore, had to enrich and connect the sequences to other forms of biological data and knowledge at source, often in collaboration with other yeast biologists. In the US large-scale center model, on the other hand, the production and release of data was itself the main object, an end in itself, something that—supposedly—would enter the wider world and serve the necessities of a plethora of unknown users. The metaphorical term “pipelines” captured this role for data, implying it was something that just needed to travel. In this model, the task of genomics was to create the data and then provide it with the means to travel, both through the processes of its own generation and the creation of wider infrastructures through which it could move. Rather than being specifically designated to serve existing biological research, this outpouring of data and associated infrastructuring created the conditions for a data-driven biology that emerged and rapidly increased in importance with the release of the human reference genome.

In what follows, we detail our analysis of the network and examine each of the bridging institutions, from the most undirected in orientation (GATC and Genotype) to the most directed (Philippsen’s laboratory at Biozentrum). Due to the networked model they were working in, GATC and Genotype were considerably less distal than the also undirected US-based large sequencing centers, or the later-established biotechnology company Celera Genomics. These differences in user proximity, as well as Biozentrum’s directed sequencing program, epitomized the diverse ecosystem of the YGSP and cemented the EC’s networked approach to genomics. As we argue throughout the paper, this approach and its divergences from the large-scale center model contribute a distinctive genealogy to our proposed *thickened* historiography of genomics.¹⁷

2. FOLLOWING BRIDGES BETWEEN NETWORK CLUSTERS

The yeast co-authorship network presents some striking characteristics when compared to the equivalent human and pig visualizations. The number of

17. On “thick” as opposed to “thin” approaches to the history of genomics, see Leng et al., “The Sequences” (n.9). On “thickening” the historiography of genomics beyond the European network model, see James Lowe, Miguel García-Sancho, Rhodri Leng, Mark Wong, Niki Vermeulen, and Gil Viry, “Across and within Networks: Thickening the History of Genomics,” this issue.

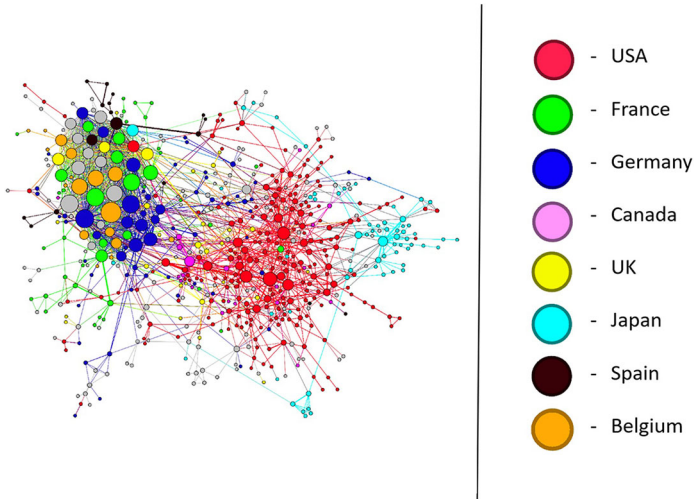


FIGURE 1. Main component of the yeast co-authorship network. We sized the nodes according to the number of publications that the represented institution co-authored with other institutions and colored them by country, as indicated in the legend on the right side of the figure; we colored the rest of the nodes gray. Figure elaborated by the authors.

institutions is considerably lower in yeast than in the other two species: 827 for yeast, against 6,014 and 1,272 for human and pig, respectively. At a visual level, the nodes that represent yeast co-authoring institutions appear more unevenly grouped than in the human and pig networks: there is a compact agglomeration of nodes on the left side of the yeast main component—the largest section of the network in which co-authorship ties connect all the nodes—that contrasts with the much more dispersed distribution of institutions across the rest of the network space (see figure 1). This appearance, to which the network metrics correspond, reflects the way the EC organized the consortium of laboratories that conducted the yeast genome sequencing work.

The European YGSP emerged in the context of a robust biotechnology agenda that the EC's Directorate-General XII, responsible for research and development, was promoting. This agenda experienced a dramatic boost in the 1980s when the political leaders of the European Community and Commission pushed for a single market and increased convergence among the member-states on many fronts. The biotechnology policy sought to improve

the competitiveness of European industry alongside a more general goal of forging greater European integration and building *capacity* at the transnational level.¹⁸

To do this, the EC fostered a model based on promoting scientific cooperation among its member-states rather than creating new research centers.¹⁹ In the case of the YGSP, this networked model materialized in an array of laboratories and centers of varying sizes that sequenced parts of chromosomes, in many cases regions they had themselves asked to cover. This meant, often, that groups interested in the biology thought to be associated with a particular region of the genome conducted the sequencing of it. Consequently, sequence data could inform the overall research aims of the group more proximately, even as its generation was part of a wider concerted effort. Coupled with the private sector involvement in the project through the Yeast Industrial Platform, this architecture sought to smooth the path for the rapid translation of results into commercial outcomes.²⁰

After an initial pilot exercise to sequence chromosome III, the overall project managers at the EC, André Goffeau and Alessio Vassarotti, distributed the remaining chromosomal assignments to coordinators who oversaw the sequencing work.²¹ These coordinators partitioned the chromosomes and sent to the consortium laboratories cosmid clones containing the fragments of DNA to sequence. Each laboratory could conduct the sequencing in any way they chose, provided they conformed to particular standards concerning

18. Parolini, "Building" (n.4); Mark F. Cantley and J. Dreux de Nettancourt, "Biotechnology Research and Policy in the European Community: The First Decade and a Half," *FEMS Microbiology Letters* 100 (1992): 25–32; Alfredo Aguilar, Etienne Magnien, and Daniel Thomas, "Thirty Years of European Biotechnology Programmes: From Biomolecular Engineering to the Bioeconomy," *New Biotechnology* 30, no. 5 (2013): 410–25; Miguel García-Sancho, "Europe and the Genome: An Overlooked Strategy for a Translational Genomics," in *Perspectives on the Human Genome Project and Genomics*, eds. Christopher Donohue and Alan C. Love (Minneapolis: University of Minnesota Press, forthcoming).

19. This also grew out of a general orientation toward networked projects deriving from EC research policy in the late 1980s; see David Dickson, "Networking: Better than Creating New Centers?," *Science* 237, no. 4819 (1987): 1106–7; Guzzetti, *A Brief History* (n.13), 168–70.

20. Alessio Vassarotti, André Goffeau, Étienne Magnien, Bronwen Loder, and Paolo Fasella, "Genome Research Activities in the EC," *Biofutur* 94 (1990): 84–90.

21. The pilot project ran 1989–1990, and the rest of the sequencing 1991–1996. Here, for the sake of simplicity, we refer to both projects together as the YGSP. Goffeau had a dual role as practicing scientist at Université Catholique de Louvain and a scientific staff member at the EC's Directorate-General XII.

quality, the use of certain materials for transferring DNA such as cosmids, and formatting and submission of data. The price paid to laboratories per base pair declined over the course of the project, and there was therefore constant pressure to increase output efficiency while maintaining and improving the quality of the data.²²

Sequencing laboratories transmitted their data to the Martinsried Institute for Protein Sequences (MIPS). MIPS was an institution based in Martinsried, a suburb of Munich, that stemmed in 1988 from the also Martinsried-based Protein Chemistry Department of the Max Planck Institute of Biochemistry.²³ Despite the initial objective of MIPS being to unify the existing protein sequence databases in Europe, Japan, and the United States, Goffeau designated it as the bioinformatics coordinator of the YGSP, thus dealing with DNA as well as amino acid data.

Crucially, however, MIPS went beyond the management of data to actually creating knowledge based on the data it received and processed.²⁴ Its first role was assessing the quality of the data by comparing it to overlapping sequences on the same chromosomal region that other participating laboratories had produced. MIPS also assembled the data using the previously produced physical maps, connected laboratories working on homologous chromosomal regions, identified particular genomic features, and annotated the sequences accordingly. Its scientists and curators then handed the results to the submitting laboratory and the designated chromosome coordinator, who would normally base future cosmid allocations on the quality of data that each sequencing group had submitted. Once MIPS handed over the sequence data, as the bioinformatics coordinator for the project they then had to publish the chromosomal sequence within six months, providing a window within which all groups involved could conduct research using their data prior to

22. Stacia R. Engel, Fred S. Dietrich, Dianna G. Fisk, Gail Binkley, Rama Balakrishnan, Maria C. Costanzo, Selina S. Dwight, et al., "The Reference Genome Sequence of *Saccharomyces cerevisiae*: Then and Now," *G3: Genes, Genomes, Genetics* 4, no. 3 (2014): 389–98.

23. *Max-Planck Gesellschaft Jahrbuch [Yearbook of the Max Planck Institutes]*, vols. 1985 (122–26), 1986 (147–49), and 1992 (138), respectively. Library of the Max Planck Institute of Biochemistry, Martinsried, consulted November 2019. Protein sequencing pioneer Pehr Victor Edman headed the Protein Chemistry Department of the Max Planck Institute of Biochemistry during the 1970s.

24. Hans-Werner Mewes, personal communication with authors, November 2021.

publication. This tended to give the sequencing groups approximately one year of exclusive use of the data they had determined.²⁵

The publications that presented the sequences explain the clustering of the YGSP institutions in our network. As the size of the reported sequences increased, the multi-institutional nature of the publications also grew due to the institutions in the consortium often sequencing more than one chromosome or chromosomal region. This intense inter-institutional collaboration translated into multiple co-authorship ties between the network nodes. The *Yeast Genome Directory*, the 1997 special issue of *Nature* that reported the complete reference sequence, contained articles describing nine of the sixteen chromosomes, with some of the papers signed by authors from more than twenty different institutions.

Looking at the structure of our co-authorship network beyond the YGSP agglomeration, the picture of yeast sequencing becomes more complicated. The institutions comprising this agglomeration account for only 17.8% of the nodes in the main component and 12.7% in the whole network.²⁶ This shows that the concerted project of the EC represented a small fraction of the overall sequencing and publishing activity around *S. cerevisiae*. Outside of the confines of this project, an array of institutions from Europe and elsewhere were sequencing yeast DNA and reporting their results in co-authored articles, some of them with the same goals as the YGSP—determining the whole yeast genome—and some as a means of achieving other biological research goals.

To further probe the structure of the network—and pave the way for subsequent qualitative work—we looked at the organization of the nodes into clusters. We identified different clusters or communities using the modularity algorithms of the network analysis software Gephi.²⁷ By doing this, we detected nineteen clusters in the yeast network and color-coded the ten largest ones in figure 2.

25. We obtained details of the organization of the YGSP from Hans-Werner Mewes, “The Bioinformatics of the Yeast Genome—A Historical Perspective,” *Yeast* 36 (2019): 161–65; Hans-Werner Mewes, interview with Miguel García-Sancho, Munich, 12 November 2019; Joly and Mangematin, “How Long” (n.5); Parolini, “Building” (n.4).

26. In contrast with the human and pig networks, where the main components include 92.7% and 80.3% of the nodes respectively, only 71.3% of the publishing institutions—590 of a total of 827—are part of the yeast’s main component. The rest are either isolates—they did not co-author with other institutions any of their articles—or form separate three- to four-node groupings.

27. Modularity algorithms divide the network into clusters—also called communities—by partitioning the network in ways to maximize the ties within partitions and minimize those between nodes in different partitions. We used the Leiden algorithm: V. A. Traag, Ludo Waltman, and Nees J. van Eck, “From Louvain to Leiden: Guaranteeing Well-Connected Communities,” *Scientific Reports* 9 (2019): article 5233.

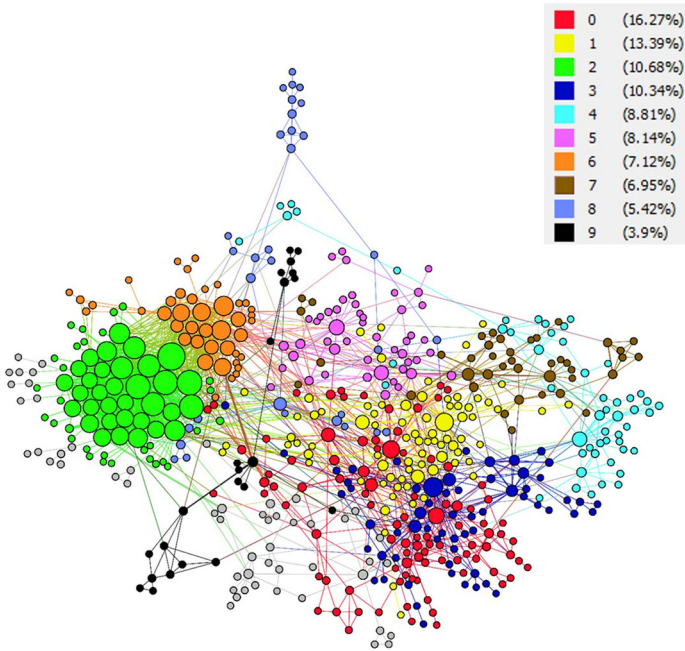


FIGURE 2. Modularity analysis of the yeast main component at resolution 1.0. We color-coded the ten largest clusters and displayed the rest in gray. Legend beside the figure indicates color code and percentage of nodes of the top ten clusters. Figure elaborated by the authors.

Almost all the institutions that participated in concerted yeast genome projects, except for the Japanese ones,²⁸ are located in three modularity groups: a mainly European community (green-colored cluster 2, on the far left); a community positioned next to it that comprises mainly Belgium, French, and German institutions (orange-colored cluster 6); and a community to the right of cluster 6 that includes Stanford University, along with the Sanger Institute and McGill University (pink-colored cluster 5).

Our modularity analysis drew two notable results. The first was the division of the agglomeration of YGSP nodes into two communities, clusters 2 and 6. All but three nodes in the core of cluster 2 represent institutions that were part

28. The Japanese sequencing projects and institutions are beyond the scope of this special issue. In all our networks, US institutions often mediate connections between Japanese and European nodes. This reflects the highly international and collaborative nature of US science, and the influence of longstanding programs of cooperation between the United States and Japan. See, for example, www.science.org/content/article/nsf-celebrates-50-years-us-japan-collaborations and www.fic.nih.gov/News/Examples/Pages/japanese-collaboration.aspx

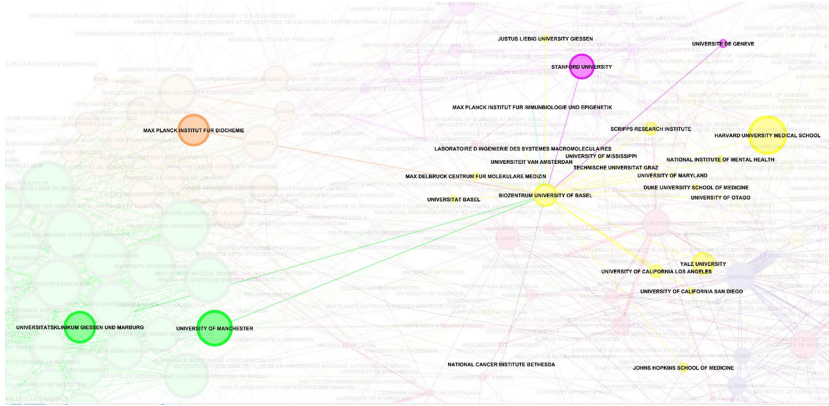


FIGURE 3. Zoomed-in network highlighting the institutions with at least one co-authorship relationship with Biozentrum. Figure elaborated by the authors. To visualize a higher resolution version, click https://www.pure.ed.ac.uk/ws/portafiles/portal/274085513/3_Yeast_Figure_3.jpg.

of the EC consortium, and only one node from the core of the neighboring cluster 6 was external to this consortium.²⁹ The only participant in the European genome project that was outside clusters 2, 5, and 6 was Biozentrum, a large biomedical research center attached to the Universität Basel and whose node is located to the right of the YGSP agglomeration, close to the geometrical center of the network (see figures 2 and 3). This distribution led to the second key finding of our modularity analysis: a relatively small number of nodes, among them those in cluster 6 and the one representing Biozentrum, mediate the connections between communities.

Biozentrum coordinated the sequencing of chromosome XIV of *S. cerevisiae* with other members of the European consortium, but because of being based in a non-member state, its funding came from the Swiss Government rather than the EC. As we will detail below, researchers at Biozentrum—including the chromosome XIV coordinator—also conducted more targeted and biologically driven sequencing work that enabled international collaborations in the field of cell biology.³⁰ In our network, this institution features in cluster 1 (yellow) and connects to institutions in five different clusters apart from its own: linking, for example, Stanford University with the European clusters 2 and 6 (see figure 3).

29. The YGSP included two non-European external partners, Kobe University in Japan and Rutgers University in the United States, both of them part of cluster 2.

30. On yeast as a unicellular eukaryotic species and the tradition of cell biological research on this organism, see Niki Vermeulen and Molly Bain, “Little Cell, Big Science: The Rise (and Fall?) of Yeast Research,” *Issues in Science and Technology* 30, no. 4 (2014): 38–46.

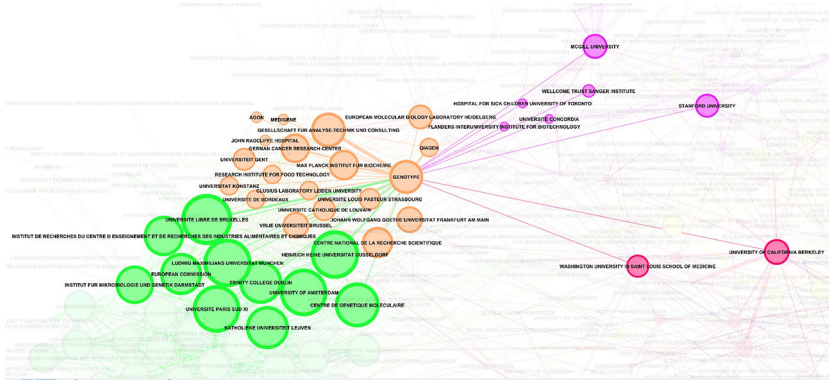


FIGURE 4. Zoomed-in network highlighting institutions with at least one co-authorship relationship with Genotype. It shows the triangle that this company forms with Gesellschaft für Analyse-Technik und Consulting and Qiagen, as well as the clusters and institutions that these companies bridge. Figure elaborated by the authors. To visualize a higher resolution version, click https://www.pure.ed.ac.uk/ws/portalfiles/portal/274085515/3_ Yeast_Figure_4.jpg.

Three other bridging institutions in our network are Qiagen, GATC, and Genotype, all of them German biotechnology companies that participated in the YGSP. They form a triangle in the orange cluster 6 (see figure 4) and have relatively strong co-authorship ties between them. Apart from bridging cluster 2 with cluster 5 (green and pink, respectively) due to their interstitial position, the three German companies also connect the YGSP institutions with the red-colored cluster 0. Clusters 5 and 0 include the three large-scale centers that led the sequencing of five yeast chromosomes: Washington University in St. Louis, Stanford University, and the Sanger Institute.³¹ The German companies also co-authored with institutions that did not participate in concerted, whole-genome projects but published yeast DNA sequences for cell biological research purposes, such as the University of California, Berkeley, in cluster 0.

Qiagen, GATC, and Genotype's pattern of co-authorships bridges the network in illuminating ways for historical research. First, the three companies connect concerted genome projects developed in Europe and America—projects that the existing literature has tended to regard as independent due to

31. If we look at the sequence submission data from which we selected our corpus of publications, Stanford, Washington University, and the Sanger Institute feature among the top four institutions in volume of yeast sequence submitted to the databases, and the latter two are also leading submitters of human sequence data.

their geographical distance and strategic differences: network organization vs. large-scale centers.³² Qiagen, GATC, and Genotype are present in six of the eight publications reporting full yeast chromosome sequences that the European consortium led, and two of the seven that either the Sanger Institute, McGill, Stanford, or Washington University first-authored.³³ Although officially, the German companies were only part of the EC consortium, they also participated in the large-scale sequencing of chromosomes XII and XVI with a small number of institutions from that consortium, whose resulting publications featured Washington and McGill University as first-authors, respectively.

Second, the three German companies connect concerted whole-genome projects with more biologically driven yeast sequencing. Beyond Washington University, cluster o mainly comprises institutions that used the single-celled eukaryote yeast as a model to conduct biomedical research. As we detail below, the German companies combined their participation in the EC consortium with collaborations involving institutions that used sequence data to investigate genes encoding biomedically relevant proteins.

The focus on those bridging institutions was crucial to moving our network analysis forward. The specific structure and smaller size of the yeast network enabled us to address interactions between clusters that Qiagen, Genotype, GATC, and Biozentrum mediated. These four institutions exhibited distinctive modes of conducting sequencing that a selective focus on concerted yeast genome projects—or a tracing of specific genealogies deriving from yeast genetics—might have obscured. In particular, GATC, Genotype, Qiagen, and Biozentrum combined the sequencing of whole yeast chromosomes within the European network model alongside collaboration with more proximate sequence users, especially biochemistry and cell biology laboratories that sought to deploy the DNA data in their research programs. In the next two sections of the paper, we qualitatively examine the work underlying the bridging co-authorships and discuss the way in which they affected—and were in turn conditioned by—the European model of networked genomics. We finally reflect on the historiographical significance of these institutions manifesting

32. Parolini, “Building” (n.4); Szymanski et al., “Yeast” (n.6); Joly and Mangematin, “How Long” (n.5).

33. The chromosomes not involving the three German companies were: VI (authored exclusively by Japanese institutions), VII and III (led by the EC consortium), V (led by Stanford), IX and XIII (led by the Sanger Institute), I (led by McGill University), and VIII (led by Washington University).

different combinations of proximate and distal, as well as directed and undirected sequencing.

3. PROXIMATE AND UNDIRECTED SEQUENCING: THE GERMAN BIOTECHNOLOGY COMPANIES

Historians have suggested that in the early years of concerted human genome projects—from the late 1980s onward—Germany did not participate to the same extent as the UK or France. Indeed, the Federal Republic of Germany was one of the last western European countries to have a national human genome project. It was launched in 1995 by the Federal Ministry of Education and Research, and the German Research Foundation, several years after the earliest European programs. The literature has attributed this late involvement to a particularly acute social suspicion of biotechnology in the 1980s and existing weakness in human genetics in the pre-unification Federal Republic, motivated by a reaction to the role of eugenics in the Nazi dystopia.³⁴

However, if we shift the historiographical focus from human to yeast, Germany was a major player in the consortium that conducted the YGSP due to its established brewing industry and research tradition in microbiology and biochemistry. This led to a large number of co-authorships of yeast sequence publications that feature prominently in our network: Germany is the joint-second most represented country in the main component with Japan, below the United States; the first in European cluster 6; and the third in cluster 2. If we take clusters 6 and 2 together, Germany is the most represented country in that YGSP-dominated agglomeration.

Germany also played an important role in the development of sequencing technologies and computational tools for sequence analysis. These lines of

34. On the connections between genetics and National Socialism, see Paul Weindling, *Health, Race and German Politics between National Unification and Nazism, 1870–1945* (Cambridge, UK: Cambridge University Press, 1989) and Benno Müller-Hill, *Murderous Science: Elimination by Scientific Selection of Jews, Gypsies, and Others, Germany 1933–1945* (Oxford: Oxford University Press, 1988). On fears of a resurgence of eugenic thought in the light of the Human Genome Project and the possibilities of biotechnology, see Daniel Kevles, “Out of Eugenics: The Historical Politics of the Human Genome,” in *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, eds. Daniel Kevles and Leroy Hood (Cambridge, MA: Harvard University Press, 1992), 3–36. On the possibilities and limitations of human genomics after the reunification of Germany, see Robert Cook-Deegan, *The Gene Wars: Science, Politics, and the Human Genome* (New York: Norton, 1994), 198–200.

work were concentrated in Heidelberg, where the Deutsches Krebsforschungszentrum (DKFZ; German Cancer Research Center) established the European Data Resource for Human Genome Research, and the European Molecular Biology Laboratory (EMBL) designed the first database to store DNA sequences, which would later become the European Nucleotide Archive. The DKFZ resource was the main repository for the Human Genome Analysis Programme, an initiative that the EC launched in 1990. During the late 1980s, the EMBL devised an automatic DNA sequencer that the Swedish company Pharmacia subsequently commercialized.³⁵ Both the EMBL and the DKFZ feature prominently in the yeast and human co-authorship networks.

Yet, it is not these institutions but our three German biotechnology companies—Qiagen, GATC, and Genotype—that play a significant role in bridging different clusters within the yeast network. These companies originated between 1984 and 1990, under the leadership of researchers who had started their careers in Universität Düsseldorf, Universität Konstanz, and the DKFZ, respectively. In what follows, we introduce qualitative evidence to account for the prominence of GATC and Genotype, whose bridging roles are the most historically relevant.

GATC was a family venture, with the brothers Fritz Jr., Peter, and Thomas building the business whose initial major selling point were the patents that their father—Universität Konstanz professor Fritz Sr.—had filed for a direct blotting electrophoresis process.³⁶ This process was the basis of a series of machines that the family developed to improve the process of transferring

35. On the DNA sequence database and the EMBL sequencer that Pharmacia commercialized, see Miguel García-Sancho, *Biology, Computing, and the History of Molecular Sequencing: From Proteins to DNA, 1945–2000* (Palgrave Macmillan, 2012), chaps. 4, 5 and 6. On the Human Genome Analysis Programme, see García-Sancho, “Europe” (n.18).

36. Fritz Pohl Sr., “Method in Which Elemental Particles Electrophoretically Migrate through a Gel onto a Collecting Surface of a Moving Belt” and “Electrophoretic Apparatus Employing a Collecting Belt Moving in Contact with a Gel,” US Patent number 4631120 and 4631122, respectively, both granted December 1986: <https://patents.justia.com/patent/4631120> and <https://patents.justia.com/patent/4631122>. The patents relate to work that Fritz Pohl Sr. and Stephan Beck, his PhD student at Konstanz, described in co-authored papers during the mid-1980s, e.g., Stephan Beck and Fritz M. Pohl, “DNA Sequencing with Direct Blotting Electrophoresis,” *The EMBO Journal* 3, no. 12 (1984): 2905–9. On Beck’s subsequent thriving career in genomics, including an appointment as Head of Human Sequencing at the Sanger Institute, see Stephan Beck, “Getting Up Close and Personal with UK Genomics and Beyond,” *Genome Medicine* 10 (2018): article 38. Also see Ulrich Falke, *Biotechnologie in Deutschland—25 Jahre Unternehmensgründungen* (Berlin: Bundesministerium für Bildung und Forschung, 2010). https://web.archive.org/web/20150711042733/https://www.bmbf.de/pub/Biotechnologie_in_Deutschland.pdf

DNA fragments from electrophoretic gels to blotting membranes, that allowed their selection and different forms of assay. The basic approach was to have the blotting membrane placed on a belt that would collect the fragments from the gel. In addition to the advantages offered in terms of preparation and handling of gels—abandoning radioactive labeling of DNA—and the ability to swap out different gels and blotting membranes for different purposes, the machines sought to ease the automation of the initial steps of the sequencing process, among them the selection of DNA fragments to be included in cosmid libraries. The sale of these machines was an early feature of the company.³⁷

Fritz Sr. remained at Konstanz until his death in 1994, while his sons ran the company external to the university, but still in Konstanz, a city in the far southwest of Germany near the border with Switzerland. While Fritz Jr. concentrated on the technical aspects of electrophoresis, and Peter from 1996 on the more business and administrative side, Thomas focused on the development of sequencing services and the growth of the company as managing director during GATC's participation in the YGSP. Thomas had worked at the EMBL for six years prior to the establishment of GATC, becoming an expert in the construction of genome libraries.³⁸

Genotype's founder was Michael Rieger, who had first worked as a microbiologist at Universität Heidelberg, then conducted genetics research at the DKFZ, and from 1983 worked in an early biotechnology company. In all these positions, Rieger's work had involved conducting a considerable amount of DNA sequencing. He was able to continually increase the amount of DNA sequence that could be read from electrophoretic gels in manual sequencing. Rieger's motivation in setting up Genotype was to lead his own institute. Despite employing only six to eight people, its sequencing output was prodigious.³⁹ Unlike GATC, Genotype did not offer a new technology for sequencing as such, but the presence of an experienced and expert sequencing scientist as well as a well-organized and motivated team. Genotype was based in

37. Stephen Oliver, "Obituary. In Memory of Fritz M. Pohl 1939–1994," *Yeast* 11 (1995): 391–92.

38. Hans Lehrach, an early developer of DNA sequencing methods and genomics advocate at the EMBL, used Pohl's libraries in his search for the Huntington's Disease gene. In this endeavor, the EMBL competed with the Massachusetts General Hospital and other members of the Huntington's Disease Consortium; see García-Sancho et al., "The Human Genome Project" (n. 14).

39. GATC grew from four or five people over the lifetime of the YGSP to c.150 by the sale of the company in 2017.

Wilhelmsfeld, a small town just northeast of Heidelberg, like GATC in the southwest of Germany.

Neither of the companies were officially part of the sequencing of chromosome III, the YGSP pilot project that involved thirty-five European laboratories.⁴⁰ Subsequently, however, GATC was able to improve their machine to produce higher output and better quality, demonstrating the new GATC 1500 to the satisfaction of the project coordinators, Goffeau and Vassarotti. The company thus became a subcontractor of Universität Konstanz, a member of the chromosome III sequencing team. GATC was involved—in this case from the beginning and along with Genotype—in the next stage of the project that began in 1992 with the sequencing of chromosomes XI and II. While GATC was a direct contractor of the EC, Genotype became a subcontractor of the Ludwig-Maximilians-Universität München (LMUM), the coordinator of chromosome II sequencing.⁴¹ From then until the mid-1990s, both GATC and Genotype received a substantial proportion of their income from involvement in European projects: the YGSP and subsequent whole-genome sequencing efforts of the thale cress *Arabidopsis thaliana* and soil bacterium *Bacillus subtilis*.⁴² This was in spite of the payment per base declining as the projects went on.

These companies did not use automated sequencers until the mid-1990s. This was not unusual across the European participants in the YGSP. Following their involvement in chromosome II, Genotype trialed some automatic sequencers but did not immediately adopt them. GATC acquired some Applied Biosystems machines shortly before the release of the *S. cerevisiae* sequence in 1996, but sought to adapt the chemistry and other parameters to refine their operation. Until then, GATC had conducted sequencing manually and limited automation to the selection of DNA fragments for libraries

40. Alessio Vassarotti and André Goffeau, “Sequencing the Yeast Genome: The European Effort,” *Trends in Biotechnology* 10 (1992): 15–18; Stephen G. Oliver, Quirina J. M. van der Aart, Maria L. Agostoni-Carbone, Michel Aigle, Lilia Alberghina, Despina Alexandraki, G. Antoine, et al., “The Complete DNA Sequence of Yeast Chromosome III,” *Nature* 357 (1992): 38–46.

41. Being a subcontractor brought in more income than directly engaging with the EC, due to German government rules stipulating that companies could receive only half of the cost from public funds (be they German or European). However, due to administrative constraints, at times it was only possible to directly contract with the EC, as GATC did during the second stage of the YGSP.

42. Thomas Pohl estimates that in 1998 and 1999, GATC received approximately 50% of their turnover from European projects; telephone interview with Miguel García-Sancho and James Lowe, 25 September 2019.

using its in-house electrophoretic blotting technology. Goffeau's calculations supported the preference for manual methods, since they indicated that manual sequencing was faster until about 1996.

During the completion of chromosomes XI and II, GATC and Genotype became reputed for the level and quality of their sequence output. They were able to share their expertise with other members of the consortium who were not specialist sequencers at the regular YGSP meetings. There, the participants discussed methods and problems, resulting in their acquiring tacit knowledge and collectively speeding up their sequencing operations. The network model, therefore, harmonized the heterogeneity—in terms of both methods and participants—of the YGSP to optimize their sequence production via the sharing of best practices. This was vital given the disparity in sequencing expertise within the European consortium, between the sequencing companies and the microbiology or cell biology laboratories, for example.⁴³

GATC and Genotype continued to participate in the sequencing of a large number of cosmids. As different coordinators managed each chromosome effort and announced the sequences in separate papers, this explains the multiple co-authorships between both companies and institutions from the European clusters 6 and 2 in our network. Yet GATC and Genotype were also part of a pool of European consortium members that co-authored the complete sequence description of chromosomes XII and XVI with Washington University, McGill University, Stanford, and the Sanger Institute (in our dataset: PMIDs 9169871 and 9169875). As well as the two German companies, other European institutions involved in those papers were the Max Planck Institute of Biochemistry, DKFZ, EMBL, AGON, MediGene, and Qiagen (the latter three also biotechnology companies based in Germany). All these institutions form part of cluster 6 whose co-authorship ties bridge the European agglomeration of the network with the large-scale sequencing centers in cluster 5.

The German companies, along with the other European institutions that co-authored chromosomes XII and XVI, therefore contributed to both the more distributed model of the YGSP and larger-scale efforts that genome centers led or participated with support from the NIH and the Wellcome

43. Michael Rieger, telephone interview with Miguel García-Sancho and James Lowe, 9 October 2019. Two bioinformatics coordinators of the YGSP confirmed to us the reliability and prowess of GATC and Genotype's scientists, and emphasized the key role of the network approach in ensuring the quality and timely delivery of the sequence data: Karl Kleine, telephone interview with Miguel García-Sancho and James Lowe, 1 November 2019; Hans-Werner Mewes, interview with Miguel García-Sancho, Munich, 12 November 2019.

Trust. The resulting publications state that the European institutions conducted physical mapping and sequencing of specific chromosomal regions. For chromosome XVI, the institutional coordinator—McGill University—asked the European consortium to map a specific area and subcontracted Washington University and the Sanger Institute to speed up the sequencing. In chromosome XII and the sequencing of other yeast chromosomes, large-scale centers consulted particular geneticists to help tackle problematic regions such as those containing repetitive sequences. For instance, LMUM's Horst Feldmann contributed to the annotation of Ty (transposable) sequence elements, while Oxford geneticist Ed Louis provided clones covering telomeric regions of many chromosomes and helped annotate the resulting sequences.⁴⁴

Involving other institutions and individuals in this way represented an organic and *ad-hoc* approach rather than a premeditated strategy. It sought to harness collaborators to deal with specialist problems, as well as ensuring that sequences were finished on time. This suggests that, in spite of their automated techniques and industrial modes of operation, Washington University and the other large-scale co-authors of the papers needed the collective, artisanal expertise from the European consortium to reach certain areas of the chromosomes. Goffeau viewed the “technical diversity” characteristic of the European project as a key advantage over “factory sequencing.”⁴⁵

Subsequent collaborations between the EC and genome centers also required an array of skills and specific technical know-how on top of large-scale sequencing: for instance, concerning various genome areas of *A. thaliana* that automatic sequencers could not properly deal with.⁴⁶ This dimension of the European model reflected the values that the architects of the sequencing networks had built into their organization and conduct. Data users, which included many of

44. For this, in the “Yeast Genome Directory” (n.1), Louis participated as a co-author on the papers for chromosomes IV, VII, XII, XV and XVI; he also provided a telomere clone for the chromosome XIV work. In addition to the chromosome II sequencing that he coordinated, Feldmann co-authored the chromosome XV paper and appeared in the acknowledgments of chromosome XIV. Mark Johnston related Feldmann and Louis's involvement in an interview with Miguel García-Sancho and James Lowe over Skype on 20 and 24 September 2020.

45. Goffeau also argued that “scientists are usually more prone to exploit scientifically the genome they have sequenced themselves (or to the mapping, sequencing, and annotation of which they have been very closely associated) than to exploit ‘anonymous’ often partial, shot-gun sequences downloaded from mercenary sequencing factories,” Goffeau, “Yeast Transport-ATPases” (n.3), 524–25.

46. The Multinational Coordinated *Arabidopsis thaliana* Genome Research Project—Progress Report: Year Six. www.nsf.gov/pubs/1997/nsf97131/nsf97131.htm

the sequencing laboratories as well as industrial partners, needed to be able to use the sequences and associated knowledge when working on their problems. This meant that the data needed to be comprehensive and of high quality. The additional, manual, and bespoke attention to particular troublesome regions was therefore worth spending time on. The European model thus stood in contrast to the commitment—both technical and political—to generate, assemble, and quickly release unrestricted sequence data at the genome centers.

As well as bridging the European and North American arms of yeast whole-genome sequencing, GATC and Genotype also connect the institutional agglomeration around the YGSP with other clusters in our network that did not participate in concerted yeast sequencing initiatives. An example of this is an article that Genotype co-authored in 1996 with the University of California, Berkeley, one of the most prolific sequence publishers within the red cluster *o*. The basis for this co-authorship was the policy of dissemination of sequence data within the European YGSP and the central role of MIPS in that process. As previously noted, producers of sequence data typically had a year of exclusive use of that data before publication. Any laboratory outside the consortium interested in a particular genomic region could, however, ask MIPS whether the sequence was available, and who had determined it. The MIPS curators, who had records of all sequence submissions, could then contact the group who possessed the data on that region and ask if they wanted to collaborate with the requesting laboratory.

This was what the group at the University of California, Berkeley, did, according to Rieger. The company transmitted the sequence directly to Berkeley and received an authorship credit in the resulting publication for it.⁴⁷ Apart from Genotype, researchers at Berkeley's Department of Molecular and Cell Biology—which is involved in virtually all of the sixty-nine yeast sequence publications of this university—co-authored the article (in our dataset: PMID 8550599). It describes the sequence of a gene (*PAN2*) that codes for a protein of the same name responsible for cleaving and binding repetitive sites of the yeast genome.⁴⁸ As well as scientific

47. Michael Rieger, telephone interview with Miguel García-Sancho and James Lowe, 9 October 2019.

48. Aside from Berkeley, other scientific exchanges in which Genotype engaged were with the EMBL, the Freie Universität Berlin, Centre national de la recherche scientifique and Institut Curie (in our dataset: PMID 8565072) and with the Universiteit van Amsterdam and ETH Zentrum, Zürich (PMID 10024662, which is not in our dataset): Michael Rieger, personal communication, November 2021.

exchanges such as this, the German companies also conducted other (often small-scale) contract sequencing services that sometimes led to co-authored publications and account for their bridging position in our network. For example, two years before their contribution to the whole sequencing of chromosome XII, GATC determined the sequence of a gene involved in the transport of substances from the yeast cell nucleus to the cytoplasm in co-authorship (in our dataset: PMID 7559750) with the EMBL and the Institute for Biochemistry of Freie Universität Berlin—the latter institution belongs to cluster 8, dark blue in figure 2 above the YGSP agglomeration.

This diversity of sequencing collaborations shows GATC and Genotype's need to expand their customer base: they had to permanently seek out new markets, especially given the continuous decline of European funding per sequenced base. Throughout the 1990s, both GATC and Genotype deployed a range of strategies within their business models, from more distal to more proximate sequencing. In distal sequencing operations, MIPS received the companies' results, and compiled, assessed, assembled, and disseminated the sequence data to chromosome coordinators before unrestricted public release. In proximate sequencing, GATC and Genotype worked with the University of California, Berkeley; the EMBL; Freie Universität Berlin; and other identifiable users that interacted with the companies more directly. In some cases, such as the co-authorship with the EMBL and Freie Universität Berlin, GATC directed the data to the research necessities of the users. In others, such as the co-authorship between Genotype and Berkeley, the sequence was undirected, that is, repurposed from the YGSP. This gradation of sequence outputs and users distinguished the German companies from the genome centers at Stanford and Washington University, which conducted only undirected and distal sequencing.

More generally, the potential for conducting proximate and directed sequencing within the YGSP created a niche for biotechnology companies in Europe. Firms such as GATC and Genotype were more viable within the networked model of genomics, in which the EC shared the sequencing assignments among a variety of institutions—some of them public and some commercial—and allowed them a period of exclusive exploitation of the data. The European sequencing companies thus embraced the nascent market of

bioinformation, which Myriad and Incyte—and later Celera Genomics—also pursued in the United States.⁴⁹

The creation of genome centers in the United States as publicly funded institutions that would tackle whole-chromosome and, at times, whole-genome sequencing made bioinformation companies adopt different business models to those in Europe. Whereas European firms could be species-agnostic service providers that needed simply to deliver the machines, clones, or sequence data that their customers had ordered from them, US companies had to compete with the genome centers and add value to any data they produced and held. This meant that the latter would concentrate primarily on human genomics, and either sell access to proprietary databases or develop therapeutic and diagnostic targets or products.⁵⁰ Another difference was that the US bioinformation companies relied on venture capital and patenting for their funding, while a main source of income for GATC, Genotype, and other European firms was their participation in publicly sponsored European projects. Both GATC and Genotype reinvested the revenues from their participation in those projects in the improvement of their sequencing capacities—including Fritz Sr.'s blotting electrophoresis machine—in order to develop proprietary technologies and expand their customer base.

Different business strategies, therefore, existed among nascent biotechnology companies in the genomics arena in Europe and the United States, conditioned by the differential models of organizing large-scale sequencing

49. Walter Gilbert, a Nobel laureate for the co-invention of the first DNA sequencing methods, created a sequencing company, the Genome Corporation, during the early years of the Human Genome Project. This firm predated the German ones but did not last long due to financial uncertainties and the stock market crash of 1987: Everett Mendelsohn, “The Social Locus of Scientific Instruments,” in *Invisible Connections: Instruments, Institutions and Science*, eds. Robert Bud, Susan E. Cozzens, and Roy F. Potter (Bellingham, WA: SPIE Optical Engineering Press, 1992), 5–22, especially 14–17. Gilbert had previously co-founded the early European start-up Biogen and later, in 1992, became Myriad’s vice-chair: Brian Dick and Mark Jones, “The Commercialization of Molecular Biology: Walter Gilbert and the Biogen Startup,” *History and Technology* 33, no. 1 (2017): 126–51; John M. Conley, Robert Cook-Deegan, and Gabriel Lázaro-Muñoz, “Myriad after Myriad: The Proprietary Data Dilemma,” *North Carolina Journal of Law & Technology* 15, no. 4 (2014): 597–637.

50. While the public and charitable funding regime of the genome centers enabled them to exclusively focus their mission on distal and undirected sequencing, US bioinformation companies needed to develop parallel or alternative strategies. As we show elsewhere in this special issue, Celera cultivated a more proximate relationship with their community of sequence users, leading to collaboration with publicly funded human and medical geneticists: García-Sancho et al., “The Human Genome Project” (n.14).

projects. As a result, the contrast between public and private actors in the United States appeared to be pronounced. Conversely, genomics in Europe was congruent with what historians have shown: that biotechnological enterprises and the commercialization of science involved profound entanglements of public and private institutions rather than dichotomous separation.⁵¹

Beyond the history of biotechnology, the investigation of the services that GATC and Genotype provided outside the European consortium enables us to broaden the historiography of yeast biology itself. The majority of the authoring institutions in our network—more than 70% of the nodes—published their articles with the aim of exploring biochemically and physiologically relevant yeast proteins rather than just compiling the corresponding DNA sequences. The traces that some of these publications left in the form of co-authorships with GATC, Genotype, and other sequence producers allow the connection of the history of the determination of the yeast genome with that of the use of that organism in cell biological research, thus complementing existing accounts that link yeast genomics to prior genetic and physical mapping of the S288C strain of *S. cerevisiae*.⁵² The next section will further uncover this genealogy with biochemical and cell biological research through qualitatively exploring another bridging institution—Biozentrum—that was itself a user of yeast sequence data.

4. DIRECTED AND PROXIMATE SEQUENCING: BIOZENTRUM

Biozentrum is a large research institute located in the Swiss city of Basel. Founded in 1971, its mission was to concentrate human and technical

51. On the entanglement of public and private actors, as well as commercial and academic practices in the history of science, see García-Sancho et al., “The Human Genome Project” (n. 14), note 19. On the emergence of biotechnology more specifically, see Jean-Paul Gaudillière, “New Wine in Old Bottles? The Biotechnology Problem in the History of Molecular Biology,” *Studies in History and Philosophy of Biological and Biomedical Sciences* 40, no. 1 (2009): 20–28; Robert Bud, “From Applied Microbiology to Biotechnology: Science, Medicine and Industrial Renewal,” *Notes and Records of the Royal Society* 64 (2010): S17–S29; Soraya de Chadarevian, “The Making of an Entrepreneurial Science: Biotechnology in Britain, 1975–1995,” *Isis* 102, no. 4 (2011): 601–33; Elizabeth Popp Berman, “Why Did Universities Start Patenting?: Institution-building and the Road to the Bayh-Dole Act,” *Social Studies of Science* 38, no. 6 (2008): 835–71.

52. Johnston, “The 2002” (n.8); Langer, “Molecular Ferment” (n.7); Szymanski et al., “Yeast” (n.6). The “Yeast Genome Directory” (n.1) reinforces this genetics-centered genealogy by including, before the contributions describing the reference sequence, an article on the linkage and physical maps of *S. cerevisiae* featuring Mortimer—a yeast chromosome mapping pioneer—as a co-author.

molecular biological capabilities to foster both basic biomedical science and research driven by the needs of the local pharmaceutical industry.⁵³ To achieve this, its founders co-located various departments of Universität Basel at the Biozentrum building and established central facilities, where researchers shared electron microscopy technologies, electrophoretic plates, or recombinant DNA methods. They also appointed thirteen founding professors by 1974, five of whom had either spent postdoctoral spells in the United States or moved from previous positions at universities there. One of them, Werner Arber, received shortly after his appointment the Nobel Prize for the co-discovery of restriction enzymes.⁵⁴

From early on, molecular microbiology was a main line of research at Biozentrum. Yet this institution promoted a broad conception of this discipline as a set of tools, techniques, and approaches rather than a narrow specialism, so molecular biological research encompassed cell biology, biochemistry, and neurobiology, among other areas. This also involved a variety of model organisms, with Arber's preference being phage viruses that infected bacteria, while another founding professor, Gottfried Schatz, introduced yeast after moving from Cornell University. Yeast squared with Biozentrum's objective of combining genetic, developmental, and metabolic approaches in the study of the cell. At Biozentrum, Schatz became the head of the Biochemistry Division and focused on the biogenesis of yeast mitochondria.⁵⁵

Another characteristic of Biozentrum was the combination of the professorial system of continental Europe with the US academic career model. The former involved teams organized around research professors with a substantial amount of scientific freedom. The latter sought long postdoctoral positions in order to favor research excellence and professional stability. At Biozentrum, this combination took the form of a professional category—project leader—

53. The Basel-based pharmaceutical industry contributed five million Swiss francs (two from Roche, one each from Ciba, Geigy, and Sandoz) to the overall construction costs, in addition to the 32.5 million that the canton where Basel is provided; Michael Bürgi, *Pharmaforschung im 20. Jahrhundert—Arbeit an der Grenze zwischen Hochschule und Industrie* (Zürich: Chronos, 2011), chapter 3.1 and on 197; Jürgen Engel, *Die Entstehung und Funktion des Biozentrums*, https://web.archive.org/web/20201102064017/https://unigeschichte.unibas.ch/fileadmin/user_upload/pdf/Engel_Biozentrum.pdf

54. Bruno Strasser, *La fabrique d'une nouvelle science, La biologie moléculaire à l'âge atomique (1945–1964)* (Florence: Leo S. Olschki, 2006), 393. See also <https://web.archive.org/web/20200626100750/https://www.biozentrum.unibas.ch/about/biozentrum-at-a-glance/history>

55. Gottfried Schatz, "Interplanetary Travels," in *Comprehensive Biochemistry*, vol. 41, ed. Giorgio Semenza and Rainer Jaenicke (Amsterdam: Elsevier, 2000), 449–530.

intended to overcome the short-term contracts of early career researchers and match American assistant professorships. To do this, research professors and fellows established work plans and found funding sources to extend the standard three-year tenures, so projects and publications could progress to a greater depth. Both the founding professors and first generation of project leaders built on their previous professional networks to attract talented postdocs to their divisions. This resulted in a significant presence of research staff from the United States, as well as visiting fellowships and other forms of academic exchange with North America.⁵⁶

Biozentrum's institutional history accounts for its peculiar position in our co-authorship network: despite coordinating the sequencing of one yeast chromosome, its node is outside the institutional agglomeration that captures the YGSP (see figures 2 and 3, above). As we discussed in our network analysis, Biozentrum's ineligibility at the time for receiving EC funds partly explains this location. Yet the co-authorship relationships that its researchers established also shaped Biozentrum's network behavior. These co-authorships included both other members of the European consortium and, crucially, many North American institutions outside the YGSP agglomeration with which its founding professors had long-lasting ties.

By qualitatively examining some of these North American co-authorships, we found connections between Biozentrum's involvement in the sequencing of the *S. cerevisiae* genome and its work outside the YGSP. For instance, researchers from the Biozentrum Division of Biochemistry co-authored two articles with the Department of Chemistry and Biochemistry at the University of California, Los Angeles, between 1998 and 2000 (in our dataset: PMIDs 9822593 and 10648604). Both publications used sequences of yeast genes to characterize two proteins—Tim9p and Tim18p—that either form part of the mitochondrial membrane or transport substances between the cell cytoplasm and this organelle. One of the articles included researchers from the Division of Biochemistry of the University of Manchester as co-authors and the other from the Department of Cell Biology at Harvard Medical School. There is a substantial degree of overlap among the Biozentrum authors in the two articles, and they both included Schatz, the head of division and leader of the project exploring mitochondrial biogenesis in yeast.

56. Kostas Tokatlidis, interview with Miguel García-Sancho, University of Glasgow, 16 August 2019, and personal communication, November 2021.

The co-authorship with the University of Manchester arose from the move of a postdoctoral fellow of Schatz, Kostas Tokatlidis, to this institution from Biozentrum. While he was based in Basel, Tokatlidis recalls the possibility of sharing techniques as a crucial factor easing the combination of yeast genetics with the investigation of the mitochondrial proteins. By that time—the mid- to late 1990s—the sharing of techniques was no longer restricted to the central facilities: it also occurred across the divisions of Biozentrum. The technicians, employed with more open-ended contracts than the postdocs, were key personnel materializing this pooling of expertise; they would move temporarily across divisions and spend time with colleagues learning the target technique, for then returning and teaching it to the academic staff of their home division. According to Tokatlidis, Peter Philippsen’s laboratory at the Division of Molecular Microbiology provided access to both techniques and databases, and this proved crucial for enabling the combination of DNA sequencing with protein biochemistry research.⁵⁷

Philippsen was the coordinator of chromosome XIV for the YGSP, an effort in which GATC also participated alongside nineteen other laboratories from the European consortium. Philippsen had completed his PhD at LMUM in the early 1970s, in a laboratory that used yeast as a model for the study of transfer RNA (tRNA), one of the molecules that mediates in the synthesis of proteins from a given DNA sequence. The head of the LMUM laboratory, Horst Feldmann, would later lead the completion of yeast chromosome II for the EC, whose publication in 1994 (in our dataset: PMID 7813418) predated that of the sequence of chromosome XIV by three years.⁵⁸

Following his PhD, Philippsen applied for a postdoctoral position at Stanford University and worked with Ronald Davis at the Biochemistry Department between 1975 and 1978. In 1973, Herbert Boyer (University of San Francisco) and Stanley Cohen (Department of Genetics, Stanford University) had devised the first recombinant DNA techniques to genetically modify bacteria. Among others, Philippsen combined these techniques with DNA

57. Kostas Tokatlidis, interview with Miguel García-Sancho, University of Glasgow, 16 August 2019. The co-authorship networks and publications we retrieved through them enabled us to conduct focused oral histories. In this one, for example, we showed a printout of the Tim9p article to Tokatlidis, and from this he was able to inform us exactly where the DNA sequence of the gene came from, based on the name of the co-authoring technician, Tina Junne.

58. On Feldmann’s contribution to the YGSP, see Horst Feldmann, “A Life with Yeast Molecular Biology,” in *Comprehensive Biochemistry*, vol. 46, eds. Vladimir P. Skulachev and Giorgio Semenza (Amsterdam: Elsevier, 2008), 275–333.

sequencing—another recent invention—to clone and determine the nucleotide structure of genes synthesizing tRNA molecules.⁵⁹ Philippsen and Davis’s collaboration was long-lasting and continued after the former returned to Europe as a Biozentrum postdoc in 1978. The oldest Biozentrum publication in our dataset (1985) is a co-authorship of Philippsen and Davis that describes the isolation of ten DNA fragments, their mapping into yeast centromeres—the structures joining the strands of the chromosomes—and the comparison of their sequences (in our dataset: PMID 2996783). Philippsen’s research during the 1980s involved sequencing from the centromeres out to the edges of the chromosomes and studying their organization and evolutionary differentiation.

In Basel, Philippsen initially joined Schatz’s Division of Biochemistry and became a project leader in 1991. Following this appointment, Philippsen moved his laboratory to Biozentrum’s Division of Molecular Microbiology and named it the Institute of Applied Microbiology. Together with Genzentrum—a research center that LMUM had established, partially following Biozentrum’s model—Philippsen’s Institute was one of the first in Europe to acquire an automatic DNA sequencer from the company Applied Biosystems.⁶⁰ This technology enabled him to lead the sequencing of chromosome XIV and also collaborate with other divisions and researchers of Biozentrum, such as Schatz and Tokatlidis.

The biannual reports of Philippsen’s Institute show that apart from producing sequences for the YGSP, he also used the data to conduct a variety of “structural and functional” analyses in yeast and other fungi. Philippsen, therefore, directed part of the sequencing operation to the research necessities of his own laboratory and other proximate users such as Schatz and Tokatlidis. The institute’s label of *applied microbiology* captured this sense of direct, proximal use rather than distal sequence translation. In referring to applied microbiology, Philippsen did not mean the basic/applied science distinction that science policy mobilizes. Neither did the term imply the translation of the data

59. Peter Philippsen, telephone interview with Miguel García-Sancho and James Lowe, 20 September 2019.

60. Apart from participating in the YGSP, Genzentrum was an active contributor to human genomics and is the seventh largest submitter of *H. sapiens* sequences between 1985 and 1995. On its history and role, see García-Sancho and Lowe, *A History* (n.16), chap. 2. See also Magnus Altschäff’s ongoing research at LMUM on the development of biotechnology in the Munich area: www.kooperation-und-konkurrenz.geschichte.uni-muenchen.de/teilprojekte1/phase1/tp-6/index.html

for eventual medical or industrial outcomes. Rather, by “applied” Philippsen referred to the immediate, direct use of sequence data for research: to act as an “information basis” that allows the exploration of a number of biological problems, ranging from cell growth to nuclear migration and mitosis.⁶¹

This direct analysis of the data was a key motivation of many research groups involved in the YGSP, most of them laboratories in the fields of biochemistry and microbiology interested in the functional properties of the sequence. Functional exploitation was one of the driving reasons for Goffeau starting the YGSP and, consequently, he incorporated this dimension from the very beginning of the sequencing operation. The identification of proteins thought to be pertinent to the cellular and biochemical processes with which individual laboratories or Goffeau himself worked was a key element not just of the agenda of the sequencing laboratories or chromosome coordinators but also of MIPS’s pre-sequence release work. As Bernard Dujon, coordinator of chromosomes XI and XV at the Institut Pasteur, put it: “most people were not interested in the genome, they only regarded it as a large collection of genes among which were those corresponding to their topic of interest.”⁶²

The separation of the whole-sequencing effort and EUROFAN, a project that ran from 1995 to 1998 and aimed to functionally explore genes discovered within the yeast genome, obscured the embeddedness of functional sequence analysis in the YGSP. The boundaries between these two projects give the appearance that the YGSP operated on a model of conducting sequencing first and then working out what to do with all that data. This was not the case, however. The continuities between the two projects, which involved many of the same institutions, reflect the outgrowth of EUROFAN from the functional goals of the YGSP, furthering them with the creation and analysis of mutant

61. For specific uses of this directed sequencing, see Biozentrum’s biennial reports over 1991–2001, which discuss sequence analysis in terms of the functional and comparative findings it enables, rather than the advancement of a domain of research in itself. Between 1991 and 1998, the reports summarized the lines of research at the Institute of Applied Microbiology as “Structural and functional analysis of genes and genomes in yeast and other fungi.” From 1998 onward, the summary changed to “Genomics as information basis for investigating dynamics of growth and nuclear migration in fungi”: *University of Basel Biozentrum: Biennial Report—Zweijahresbericht, 1991–1993, 1993–1995, 1996–1997, 1998–1999, and 2000–2001*. Personal archive of Peter Philippsen, obtained 14 October 2019.

62. Bernard Dujon, “My Route to the Intimacy of Genomes,” *FEMS Yeast Research* 19, no. 3 (2019): foz023.

yeast libraries.⁶³ These continuities also point to the success of the YGSP as a model for the exploitation of the data it generated: the proximate and directed nature of much of the sequencing work, the network organization of the laboratories involved, and the role of MIPS enabled this concurrent production and functional exploration of the sequences. A substantial part of the functional analyses appeared back to back with the sequence descriptions in the chromosome publications of the *Yeast Genome Directory*.

In contrast to GATC and Genotype, the noncommercial institutions of the European consortium retained their academic orientation, only spending part of their time on sequencing and functional analysis, with the rest concentrated on microbiological, biochemical, or cell biological research. Their sequencing was thus more directed than that of the companies we have discussed. Yet the duality of the work of all YGSP participants is worth noting: apart from supporting their own research needs—or those of other, proximate users—their sequencing was also undirected, in the sense that the data could potentially enable other unconceived uses after release in public repositories. How did the European network model, which created space for the proximate-directed applied microbiology of Philippsen's group as well as the less proximate and less directed work of GATC and Genotype, compare with the large-scale centers involved in delivering the whole yeast genome in the United States?

Take the Stanford DNA Sequencing and Technology Center (SDSTC) that Davis co-founded in 1993 with funding from the NIH. SDSTC was a large-scale sequencing institution that pursued the twin objectives of contributing toward the determination of whole yeast chromosomes and developing the technical means for increasing the scale and efficiency of sequencing. Although some co-authorships in our network suggest that SDSTC conducted some yeast genome sequencing in cooperation with European institutions and provided sequence data to other departments leading cell biological projects, it may be more appropriate to say that it was Davis himself who contributed to these initiatives.⁶⁴

63. About a quarter of the participants in the YGSP network participated in EUROFAN, and all but two of the twenty-one participating institutions in EUROFAN had authors in the *Yeast Genome Directory* (n.i): <https://cordis.europa.eu/project/id/BIO4950080>

64. For example, in articles describing collaborative cell biological research and full chromosome sequences in which he was not the coordinator, Davis listed his affiliation as Stanford's Department of Biochemistry and not SDSTC: PMID 8978028 in our dataset and chromosome IV and XVI papers in the "Yeast Genome Directory." The SDSTC affiliation appeared only when Davis's team led the sequencing effort: chromosome V paper in the "Yeast Genome Directory" (n.i).

As an institution, SDSTC differed from the Institute of Applied Microbiology in three key aspects: (1) it did not use the sequenced DNA to investigate yeast biology and rather focused on the possible scaling-up of the sequencing technologies to the human genome; (2) it released the sequence data to public databases immediately, cosmid by cosmid, rather than retaining the information for a period of time;⁶⁵ and (3) it sequenced its target chromosomes comprehensively instead of focusing on certain areas—the paper reporting the full sequence of chromosome V included co-authors exclusively based in SDSTC, while Philippsen's chromosome XIV paper featured twenty different co-authoring laboratories, each in charge of a specific region. In other words, the genome center model that the SDSTC embodied was significantly more distal and undirected than the network organization of the European sequencing effort.

These differences suggest that SDSTC, Biozentrum, and their sponsors attached distinct values to the sequences. For the NIH and SDSTC, the yeast reference genome was a basic biological resource that would constitute a future research asset. For Biozentrum and the EC, the sequences did not have to wait for future, distal users: because of the networked model, the groups that had themselves determined them, as well as industrial partners or any other collaborator they agreed to work with, could exploit the data. In the case of Biozentrum, the immediate use of the sequences materialized in co-authorships between the Division of Biochemistry and other protein chemistry and cell biology laboratories at Harvard Medical School and the University of California, Los Angeles. In the case of GATC and Genotype, the temporary restriction of access to their sequences transformed both companies into obligatory points of passage and helped them to build a reputation among potential customers, which sometimes led to co-authorships like the ones with Berkeley, the EMBL, and Freie Universität Berlin.⁶⁶

65. US genome centers including SDSTC submitted the sequences to GenBank and other repositories subsequently mirrored the data: the Stanford-based *Saccharomyces* Genome Database, the European Nucleotide Archive, and the DNA Data Bank of Japan. See Sean Walsh and Bart Barrell, "The *Saccharomyces Cerevisiae* Genome on the World Wide Web," *Trends in Genetics* 12 (1996): 276–77.

66. Other members of the European consortium kept the use of the sequences to themselves: this was the case for Feldmann's laboratory in Munich, with LMUM laying at the heart of the European cluster 2 in our network and not developing a significant bridging role with other modularity groups.

Conversely, SDSTC's undirected releasing of data to public databases was based on the expectation that others would be able to exploit the sequences for scientific, medical, and industrial purposes. Yet, in this model, there was not much further concern with the sequence users, nor with the prospective uses of the sequence data. For SDSTC, the exploitation of the sequence data would occur in a more distal way: this institution and the other yeast genome centers had less direct contact with the users of the sequence.⁶⁷ Furthermore, the genome centers identified the translation of the sequence more with the use of the data and the technologies for human genomics, and less with the research necessities of the yeast community. The success story of the Human Genome Project, especially in the form it took from the late 1990s, led scientists and commentators to regard the undirected and distal sequencing that the large-scale center model predicated as definitional of genomics and a means of demarcating it from non-genomic research. Our history of the YGSP and the European network model shows that this was not always the case and that there is room for a genomics conducted in more proximate and directed manners.

5. CONCLUSIONS

This paper has shown the historiographical potential of investigating the network model of conducting genomics that the EC advanced, chiefly in sequencing the genome of the yeast *S. cerevisiae* between 1989 and 1996. We have done this by analyzing the co-authored publications that this project fostered between institutions involved in the genome effort and others exploring yeast for other biological research purposes. A relatively small number of bridging institutions channeled the co-authorship ties between bespoke and concerted yeast genome sequencing. By quantitatively, visually, and qualitatively exploring their bridging role in a co-authorship network, we have attempted to bridge different strands both within and across the history of genomics and the life sciences.

67. The distance from the users increased in genome centers that were not based in universities and could not easily collaborate with departments investigating yeast biology. This was the case for the Sanger Institute, whose node is less connected with others than Stanford and Washington University in our network.

Our bridging institutions showed deep interactions between the genome sequencing effort and other biochemical and cell biological research on yeast. By looking at the detail of these interactions, we complemented existing narratives that present the full sequence of *S. cerevisiae* as the culmination of yeast genetics research. The focus of the literature on US institutions that contributed to linkage and physical mapping of yeast chromosomes and later developed genome centers—such as Washington University and Stanford—suggests a lineage between the genetic characterization of this organism and its later whole-genome sequencing. Yet, without denying this, the laboratories of the European network and their connections show the importance of broader biological investigations of the yeast cell as both a motivation to sequence its genome and a set of research questions that the resulting data could help address in an innovative way. These investigations used the DNA sequences as tools to target other research objects, such as proteins or organelles. Delineating this pathway between cell biological research and whole-genome yeast sequencing contributes an additional genealogy to our proposed thicker history of genomics, in addition to the continuities and connections between medical genetics and human genomics that we revealed earlier in this special issue.⁶⁸

Furthermore, the role of two German companies—GATC and Genotype—as bridging institutions in our network adds previously unexplored lineages to the history of biotechnology. GATC and Genotype’s business models came to focus on the provision of DNA sequencing services. As such, GATC and Genotype’s potential sources of funding differed from those available to Genentech, Biogen, and other historical exemplars that had proliferated earlier in the United States and, to a lesser extent, Europe. The German companies were more dependent on ongoing genome projects and less on venture capital investments. Given the funding structure of those projects, investigating the activity of GATC and Genotype reveals new entanglements of the public and the private in their provision of a service to publicly sponsored genomic initiatives. The companies initially tailored this service provision to the networked, cottage industry approach of the genome projects that the EC sponsored. However, as the co-authorship patterns in the sequencing of yeast

68. On thickening or widening the historiographical scope of genomics, see Leng et al., “The Sequences” (n.9). On the genealogies between medical genetics and human genomics, see García-Sancho et al., “The Human Genome Project” (n.14).

chromosomes XII and XVI show, larger-scale sequencing endeavors in North America also needed the smaller-scale technologies of GATC and Genotype.

The continued importance of the distributed, networked organization in Europe during the early to mid-1990s confirms the contingency of the model of the genome center, something we show more explicitly in the context of human genomics.⁶⁹ The comprehensive and dedicated sequencing of genomes at institutions that would not themselves participate in the use of the data only became the dominant model during the late stages of the completion of the human reference genome (1996–2003). Before that, genome centers co-existed and competed with more distributed strategies, as the simultaneous US and European approaches to yeast sequencing reveal.

This means that up to the late 1990s, genome sequencing did not necessarily have to be undirected and distal. Biozentrum, the German companies, and all the sequence producers we have analyzed within the European project contributed to a resource—the yeast reference genome—aimed at a large community of biologists. Yet their sequencing work was also directed, to a greater or lesser extent, to the necessities of more specific, proximate users that ranged from their own laboratories to other divisions of their own research centers or potential customers. Stanford and the other yeast genome centers, on the contrary, conceptualized the translation of the sequence in a much more abstract way since their users were more distal. They sought the large-scale production of a sequence with a range of potential uses—not least informing the determination of the reference human genome—but without specific direct and identified recipients.

The European model thus fostered a wide array of actors with different relationships to the sequence data they produced, and to the other members of the network. The funding arrangements—per base pair and with sequencing assignments distributed upon quality assessment of previously produced data—enabled companies like GATC and Genotype to contribute large percentages of the total yeast reference genome while growing their own businesses. They could do this alongside other members of the network like Biozentrum, who were producing sequence data for their own particular research purposes as well as contributing it to the total product of the network. The networked model therefore offered the opportunity for a wide array of methods, motivations, and outputs of sequencing activity, all of

69. García-Sancho et al., “The Human Genome Project” (n.14).

them relying on common standards, quality control mechanisms, and an embargoed data release policy. In particular, in allowing small laboratories to participate, this model also allowed small-scale sequencing specialists to thrive.

There was some disagreement over the future of this networked organization of sequencing at the final conference of the European YGSP, in 1996. While two scientists discussed the merits of the network-based approach and outlined the conditions for it to continue to work, two administrators believed that the scale required for future efficiency in sequence production meant that networks of smaller-scale laboratories would cease to play a role.⁷⁰

Indeed, the EC changed their funding policy in 2000. They would no longer contract with companies like GATC and Genotype and pay them per base as before, but would fund only labor and materials.⁷¹ Further, they shifted focus from whole-genome sequencing to projects pursuing “post-genomics” and “translation.”⁷² The change in the funding model of the EC forced a bifurcation. For the specialist sequencing providers, it was up (concentration) or out (diversification). Both GATC and Genotype tried the latter. Through participating in a consortium of five German companies called the Gene Alliance—that originated in 1998—GATC also attempted to pursue the former, while still maintaining the advantages of the network model. In the Gene Alliance, constituent companies remained independent and able to pursue their own approaches to sequencing, but in combining their forces they would achieve the capacity to compete with more concentrated centers.⁷³ The Gene Alliance was successful in getting some contracts for genome sequencing and establishing collaborations with companies working on pharmacogenomics and agricultural biotechnology but was in abeyance by the mid-2000s.⁷⁴

70. Programme of the Final European Conference of the Yeast Genome Sequencing Network, Trieste, 25–28 September 1996, 24, 31, 80–84. Personal archive of Karl Kleine, obtained 22 November 2019.

71. Thomas Pohl, telephone interview with Miguel García-Sancho and James Lowe, 25 September 2019.

72. See <https://cordis.europa.eu/programme/id/FP6-LIFESCIHEALTH>

73. Robert Koenig, “German Biotech Form Gene Venture,” *Science* 280, no. 5366 (1998): 999–1000. The other four companies were AGOWA, Biomax Informatics, MediGenomix, and Qiagen.

74. García-Sancho and Lowe, *A History* (n.16), chap. 2.

The reasons for the demise of the Gene Alliance are a subject for further inquiry. Its fate meant that the heterogeneity of different combinations of proximate-distal and directed-undirected sequencing (sometimes in the same laboratory) could not persist, especially in the light of funding becoming concentrated in a handful of large-scale genome centers in the mid- to late 1990s. A bifurcation toward institutions conducting either proximate-directed or distal-undirected sequencing was the result, with scientists and commentators often (and narrowly) understanding genomics as just involving the latter.

The sharpening of distinctions throughout the 1990s between a small number of large-scale sequencing centers conducting distal-undirected sequencing and a plethora of existing small-scale laboratories conducting proximate-directed sequencing for specific research purposes encourages the view that these analytical terms are in practice always paired as opposed categories. When addressing human genomics in this special issue, we detailed connections between these two modes of sequencing in the context of the full determination and annotation of the sequence of human chromosome 7.⁷⁵ In this paper, we have added to this the manifestation of different combinations enabled by a particular organizational model of genomics research. This networked model never failed in practice but ceased to exist primarily because of the projected timescales for organisms with larger genomes.

The genome center model led to the rapid production of a human reference genome sequence, published ahead of schedule in 2004. However, the medical and industrial exploitation of that sequence has proved slower than expected. Is this because of the large-scale center model's more abstract and future-oriented concept of translation? The context of the production of data shapes how fungible it is, and the context of particular scientific problems (including those in industry and medicine) conditions the requirements of data and their form. Taken together, these considerations create difficulties when the domains and context of production and use of sequence data are separated and distal. This then establishes the problematic of the *translational gap* between the production of data and its use and further exploitation. In light of this, we may reflect on the EC's genome projects of the 1990s as a model, uniting as it did the production of sequence with many different and unrestricted kinds of use in a heterogeneous ecology of institutions: proximate as well as distal, with direct as well as undirected potential uses in mind.

75. García-Sancho et al., "The Human Genome Project" (n.14).

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/>