Sequencing BGI

Citation for published version:
Wang, Kai, Shen, X & Williams, R 2020, 'Sequencing BGI: The evolution of expertise and research organisation in the world's leading gene sequencing facility', *New Genetics and Society*.
https://doi.org/10.1080/14636778.2020.1843148

Digital Object Identifier (DOI):
10.1080/14636778.2020.1843148

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
New Genetics and Society

Sequencing BGI: the evolution of expertise and research organisation in the world's leading gene sequencing facility

By Kai Wang, Xiaobai Shen* and Robin Williams
*Corresponding author. Email: x.shen@ed.ac.uk

Abstract

The increasing importance of computational techniques in post-genomic life science research calls for new forms and combinations of expertise that cut across established disciplinary boundaries between computing and biology. These are most marked in large scale gene sequencing facilities. Here new ways of organising knowledge production, drawing on industrial models, have been perceived as pursuing efficiency and control to the potential detriment of academic autonomy and scientific quality.

We explore how these issues are played out in the case of BGI (Beijing Genomics Institute prior to 2008. In Mandarin 华大基因 - *Hua Da Jiyin* - Big China Genome), which is today the world's largest centre for gene sequencing research. Semi-detached from traditional academic institutions, BGI has developed distinctive models for organising research and for developing expertise, informed by information technology sector practices, that differ from existing models of interdisciplinary research in academic institutions.

1 Introduction

The application of statistical and computational techniques is seen as crucial to the decoding of the human genome (Garcia-Sancho 2012). The subsequent resurgence of life science has involved the increasing automation of gene sequencing and made available escalating volumes of data (Leonelli 2012, Vermeulen 2012). These changes have been accompanied by the emergence of large-scale life science facilities, enabling new models of scientific research organisation. Though there have been few detailed studies of these developments (Vermeulen 2016), some have pointed to the influence of industrial models, involving increasing specialisation with new forms and combinations of expertise, which are seen as undermining traditional academic models of disciplinary knowledge production (Kleinman and Vallas 2006, Stevens 2011, Highfield 2016). Bartlett, Lewis and Williams (2016) flag the possibility that "values from industry are infiltrating academic ones … which might extend to different ideas of reward and success." Collins et al. (2003)

highlighted the concerns amongst scientists that the large-scale industrial model of genomic research would not be attractive to the strongest scholars.

This paper examines the case of BGI (formerly Beijing Genomics Institute), today the world's largest gene sequencing research centre. Currently BGI Research Institute has over 1000 staff as part of a larger Group with over 6000 employees.[1] This paper explores how in BGI, relatively detached from traditional disciplinary academic institutions, distinctive arrangements have emerged in terms of the formation and deployment of expertise - how expert researchers are recruited, how they develop their expertise and careers - and how research is organised. These seem to diverge from models of research observed in large-scale Bioinformatics facilities in the West (Vermeulen 2016). BGI recruits its researchers from undergraduate and Masters programmes, breaking from the widely established pattern based on research training through a University PhD. An authoritative editorial in *Nature* (Anon. 2010) strongly criticised BGI's training model and argued that "the burden of proof" lay with BGI to demonstrate that its organisation prepared its "student-workers to meet the wide range of skills needed by industry and academia" (Anon. 2010:7). We consider this question and its implications for the future of expertise in life science research.

1.1 Goals and Organisation of the Study

This paper presents the most detailed empirical study of BGI to date. It explores the distinctive models for the formation of expertise and the organisation of research work which underpin BGI's strategy for delivering high quality, low cost gene sequencing. We draw insights from a tradition of scholarship that explores how expertise is generated, mobilised, validated and traded in the knowledge economy (Fleck 1988). This perspective views expertise as acquired significantly through experience-based learning (Fuller and Unwin 2010), rather than just formal training, through the development of techniques for the utilisation of advanced equipment (epistemic practices) (Nerland and Jensen 2012; Eyal 2013).

Opportunities to conduct fieldwork in BGI have been limited. A number of academic studies of BGI have recently appeared (Wong 2016; Fischer 2017; Stevens 2018), which contribute, to a greater or lesser degree, to the primary evidence base.[2] The apparently most empirically-detailed of these (Stevens

[1] The total staff number of BGI Group is recorded in its official website as over 6000 (https://www.bgi.com/us/company/about-bgi/, sampled 22 February 2020). According to the latest Annual Report of BGI Genomics (BGI Genomics, 2018), BGI Genomics Co. Ltd has 3150 staff mainly based in Shenzhen, Hong Kong, and other cities in China, with 628 R&D personnel included.

[2] As these three papers do not include formal methodological accounts, we cannot directly assess their empirical base. However we can examine the primary empirical material cited in their papers.
Stevens (2018:87) reports "interviewing scientists, visiting the … labs … and attending conferences, workshops, talks, and other activities at the lab."

2018), though drawing upon visits to BGI and interviews with staff, notes that access was "necessarily restricted and incomplete in some respects" requiring ethnography to be "supplemented by media sources". (Stevens 2018: 87-88). This paper significantly extends this existing evidence base.

We conducted three rounds of fieldwork in BGI in Shenzhen over the period 2010 – 2017. The first author visited BGI's headquarters and gene sequencing facility in its original industrial premises in Shenzhen's Yantian district. He undertook two rounds of interviews with a wide array of research and development staff in March 2010 and with BGI Managers and Directors in February 2011. The co-authors visited BGI's new purpose-built showcase premises in Shenzhen's Dapeng Peninsular in April 2017, for extended discussions with a group of BGI managers and researchers, to track further developments in BGI's strategy and research organisation.

Interviews were recorded and tapes transcribed. Chinese language interviews were translated. A list of the 16 respondents, detailing their role, specialty, work unit and date of interview is included as an appendix (*Table 2: BGI Shenzhen Directors, Managers and Research Staff interviewed).* Interviewee identities are pseudonymised. This primary research was supported by continued desk research including the extensive body of online grey literature in both Chinese and English language. The research team analysed this body of material to identify and code key themes emerging from the fieldwork. All members of the research team contributed to writing up and analysis of the results.

1.2 Post-Genomic Transformations in Life Science Research

Advances in computational techniques were central to the development of techniques for sequencing the human genome (Garcia-Sancho, 2012), and the subsequent emergence of modern post-genomic biology. The alliance between biology and computing was initially restricted to a specialised community, willing and able to invest in developing shared understandings of the scientific and technical issues at stake. However, as the scale and pace of sequencing have increased, information technology has become key to collecting, storing and analysing the accelerating flows of gene sequence and other information. Genomics (coupled with related efforts in proteomics, systems biology etc.), revolving around large-scale data analysis and computer models, is becoming key across many areas of biological research (Calvert 2010). Radical new conceptions have been articulated of the future of life science and its exploitation centred around the reuse of experimental data. The UK Biotechnology and Biological Sciences Research Council even proposed that increasing adoption of technologies and tools to store, analyse and re-use experimental data would lead

---

Fischer (2017) cites interviews with two of BGI's charismatic leaders (Wan Jiang and Wang Jun), offered as a way of generating comparative insights in relation to his more detailed study of Singapore's Genome Institute of Science.
Wong (2016) cites media sources and trade press together with as posts on online fora. Including translated "online forums discussing BGI recruitment and Chinese media interviews". (Idem. S111)

to growth of ‘dry’ science alongside the traditional ‘wet’ science of test tubes and reagents (BBSRC 2003).

In this post-genomic world, new generations of increasingly-automated gene sequencing equipment have offered dramatic improvements in the speed and efficiency of sequencing. ‘3rd-Generation’ gene sequencers, with their highly parallelised analysis methods, offer very accurate sequencing at fraction of the cost ($1000 compared to over $1 Million) than the previous ‘Sanger’ sequencing method (Kulski 2016). Alongside these developments we note the increasing salience of large-scale gene-sequencing facilities able to afford and share the large capital costs of equipment and information infrastructure and also better placed to exploit these machines efficiently. Alongside these innovations in equipment and techniques we thus find significant changes in the organisation of research. Larger facilities have a particular opportunity and incentive to establish new working methods and procedures to improve the speed/efficiency /accuracy of gene screening and information processes. This includes the scope to recruit new forms of expertise (reflecting the increasing importance of computational specialists) and establish new combinations of knowledge to achieve more elaborate divisions of labour and responsibility.

The formation of new forms of expertise was seen to cut across entrenched discipline-based arrangements for training and developing the careers of researchers and ways of organising research. The Human Genome Project (HGP) had been a key driver behind the emergence of Bioinformatics through what Bartlett et al. (2016:189) characterise as a “shotgun marriage” between biology and computer science: two strongly entrenched disciplinary fields with different methods and presumptions about the world and different reward and value systems (Lewis, Bartlett and Atkinson 2016). Bartlett et al. (2016), examining struggles for epistemic authority over the emerging field of bioinformatics, suggest that biologists have retained cultural power as legitimate interpreters of biological world. Conversely the important contributions of diverse computational specialists in creating protocols, developing algorithms and data curation in processing data from large scale biological experiments - might not be recognised and thus rewarded/ valorised for example when the work was published in a biological science journal. Lewis et al. (2016) note that these hybrid or fractional scientists (idem: 471) may be overlooked or “’hidden’ in the middle” (Lewis et al. 2016: 487).

These issues might however come to be seen as temporary phenomena that may be expected to be resolved as these new bioinformatic roles become established, supported by new structures for acquiring and valorising expert status and developing careers for these new forms of expertise. Bartlett et al. (2016) explore how these developments unfold over time, distinguishing several generations of bio-informaticians – (passing from its earliest stages, through adolescence [2002-6] to adulthood [2007-11]). These discussions suggest the need for processual, evolutionary analyses of these developments through historical and longitudinal studies that can capture the temporal dimension and changing dynamics and stakes. This paper seeks to contribute to this account of a rapidly evolving set of developments.

1.3 New models of research work in large scale bioinformatics facilities

Large scale bioinformatics centres are strategic sites where these new expert roles are being developed and elaborated. Studying them provides an opportunity to examine these processes in detail.

Gene sequencing involved different methods of research organisation characterised not just by their larger scale but also by a higher division of labour. Collins, Morgan and Patrinos (2003:286) observed how in the UK Sanger Institute "at first everyone did everything," following the tradition of manual sequencing groups. … However, it soon became apparent… that, for the sake of efficiency and accuracy, it was best to recruit staff of varying skills—from sequencing technology to computer analysis—and to allocate the work accordingly."

Concerns were expressed about these changes in the organisation of biological research. Stevens (2011, 2013) offers a critical account of the industrial model of organisation of science in a commercial facility (The Broad Institute in Cambridge MA). Stevens suggests (2011:31) that the "importation of computers into the life sciences has brought with it not only changes in practice, but also a transformation of the values and knowledge regimes of biology. The descriptions of work at The Broad show how the 'business' modalities of the computer changed the kind of work that is performed and the kinds of questions that are asked and answered". Stevens (2011) points to "a fundamental transformation" in biological knowledge production, which has become commodified: applying business principles and technology to increase productive output, decrease variability and increase scrutiny in order to produce "high quality, high quantity" certified and valuable knowledge (Stevens 2011:30).

Though commercial gene sequencing facilities had achieved high levels of efficiency in sequencing, questions have been raised about the quality and scientific value of outputs of large-scale and in particular commercial laboratories. In an influential account in the leading journal *Science,* Collins, Morgan and Patrinos (2003), who were key players in the HGP, stress the need for this work to be 'science driven', reflecting continuing concerns about the scientific contribution of the industrial model of scientific research and of commercial players.

These views have raised questions about the scientific contribution of BGI: is BGI just a 'genomics factory' as suggested by Cyranoski (2010) or does it represent a breakthrough in the way we do science? As Fischer (2018:276) notes: "There is an enormous amount of mythology about BGI", with polarised narratives based upon simplistic projections.

This study accordingly offers a determinedly empirical account. It focuses upon the ways in which BGI has elaborated and refined in practice a distinctive model for recruiting and training a highly specialised workforce and deploying it effectively in its research activities. We will explore how BGI has creatively experimented with Western organisational models and practices (Wong 2016, Stevens 2018) to create new models of scientific research organisation geared towards BGI's particular strategy and context.

2 Case study of BGI

2.1 Origins and History of BGI; Growth Strategy

Beijing Genomics Institute was set up in Beijing in August 1998, as an independent, non-governmental research institute - to prepare China to become a part of the HGP, the global "big Science" public-private partnership launched in the aftermath of the whole-genome shotgun sequencing efforts and patenting claims of Craig Venter/Celera Genomics (Chen 2013). A research proposal presented by the Chinese scientist, Huangming Yang, at the Human Genome Conference at Cambridge University in September 1999, successfully paved the way for China to enter the international HGP scheme. China's involvement in a 'one percent share' of the draft mapping was officially announced by the International Human Genome Organisation three months later. This success placed BGI in a leading position to work with China's other national Human Genome Centres with three million RMB of government funding in November 1999.

In June 2000, the international HGP completed its first 'rough draft' of the entire human genome. 30 million out of the entire 3 billion base pairs of the human genetic structure were mapped by BGI. Following that, BGI also successfully sequenced the coronavirus responsible for the 2003 worldwide SARS epidemic.

BGI has become China leading genomics institute (described in Mandarin as Hua Da Jiyin [华大基因 - Big China Genome Technology Co.]). State support placed BGI in a unique and powerful position to gain further resources. BGI's co-founder, Prof Huanming Yang, has been selected to be an academician of China Academy of Sciences (CAS) since 2007, which affords "the right to put forward suggestions on the decision-making of the State on major science and technology issues" [3]

In 2010, China Development Bank provided a $1.5 billion, 10-year loan in 'collaborative funds', which enable BGI to acquire all of the 128 latest and fastest (*HiSeq 2000*) next generation sequencers produced by *Illumina* in the beginning of 2010 (Fox and King 2010). In 2013, it acquired the Californian company, Complete Genomics, a life sciences company which had been developing and supplying DNA sequencing platforms since 2006 (Larsson 2013). This move brought BGI from 40% to 50% of the global gene sequencing market (Mahajan 2014). BGI, and its sequencing subsidiary MGI, in addition developed its own third generation sequencers (BGISEQ-500 in 2015, and MGISEQ-2000 and MGISEQ-200 in 2017) which offer strong competition to Illumina's products (Senabouth et al. 2019).

[3] http://english.casad.cas.cn/Me/OR/200905/t20090515_3152.html sampled 10 February 2020.

Though acquiring the prestigious status as a part of the Chinese Academy of Science (CAS) in 2004 in Beijing, this institutional framework proved too restrictive.[4] In 2007, BGI relocated its headquarters to Shenzhen, the most economically vibrant city in southern China, to benefit from investment from Shenzhen Municipal government as well as central government. It subsequently expanded at an unprecedented pace. The move helped BGI secure investment, space and managerial freedom for its further development.[5] Initially housed in unprepossessing industrial buildings (a former shoe factory) in Shenzhen's Yantian district, BGI recently moved in to ultramodern showcase premises, purpose-built, in landscaped gardens in the Dapeng Peninsular to host the China National Gene Bank (initiated in 2011, opened in 22 September 2016). It is now by far the world's largest genetics research centre. BGI focuses on high scale, high quality, affordable gene sequencing services and associated data infrastructures, primarily to pursue research, often in collaboration with other universities. but also supporting various other income generating activities including genetic screening services for clinical diagnostics and pharmaceutical drug development and other commercial operations.

BGI Group has today proliferated into ten interlinked organisations encompassing a wide range of activities including production of its sequencing technology platform, data infrastructures and commercial service as well as research, and a college which offers a growing number of joint degree programmes through tie-ins with established Chinese universities (including Tsinghua, Peking, and Wuhan).[6] Today it has sequencing centres in Europe and North America as well as China and a global network of partners/centres in over 100 countries and regions. Thus only around half (3,150) of BGI group's total current 6000 staff are employed in BGI Genomics in Shenzhen.[7] The 2017 flotation of BGI Genomics in the Shenzhen Stock Exchange valued the company at RMB 54.57 billion (US$8.29 billion).[8]

2.2 Recruitment and career development of BGI Research Institute staff

2.2.1 Recruitment

[4] For example CAS institutes had a maximum of 150 staff (Cyranoski 2010)

[5] Though officially awarded the status of a 'state agency' by the Shenzhen Municipal Government they are a state agency in name only. As GX, prenatal diagnosis developer and research project head in Healthcare Platform (Interviewed 17th March 2010) told us "*We are under the management model of enterprise rather than institutional organization*". This ambiguous status gave BGI considerable flexibility as "a nonprofit [that] operates several businesses (Larsen 2013:7)

[6] BGI's structure is complex and has evolved rapidly with its diverse range of activities. As well as BGI Research Institute (now termed Shenzhen Hua Da Gene Research Institute), BGI Group contains three additional non-profit subsidiaries (BGI College, China National GeneBank, GigaScience – its publishing arm, BGI Medicine). BGI has four commercial subsidiaries (BGI Genomics (undertaking testing and research services), MGI (producing sequencing equipment) FGI (Forensic services), and Health. (https://en.genomics.cn/en-about.html sampled 24 FEB 2020)

[7] https://www.bgi.com/us/company/about-bgi/, last sampled 22 February 2020

[8] https://www.weekinchina.com/2017/11/under-the-microscope/ Sampled 4 July 2019

This is a new institution that has grown quickly.[9] The rapid recruitment of younger researchers means that their average age is low - only 26 in 2011 (Frank 2011). This had only risen to 29 six years later by the time of our final 2017 visit by when the head of BGI Research was only 32.[10]

Most BGI Research Institute staff have a background in biology. Table 1 provides a breakdown of the disciplinary backgrounds of BGI research staff in the BGI-Shenzen site studied.

INSERT TABLE 1 NEAR HERE

BGI has recruited "an army of young bioinformaticians" (Cyranoski 2010). BGI recruits bioinformaticians without PhDs. They recruit people doing first degrees or MScs in bioinformatics. Potential recruits are often attracted through placements that give them a chance to collaborate with BGI on projects (and perhaps even contribute to a paper) before graduation and then move to BGI. Thus BGI recruited 17 of the 30 students who enrolled in its 2009 summer school , including CH, already a Project Investigator. CH (interviewed 16th March 2010) recalled that "*in 2009. … when I was still studying in University I applied for this summer school and attended the interview and then I received my acceptance and came here during my summer holidays. The summer school was designed to last a few months. I deemed it to be a good chance to observe and learn from the researches on site and truly I had an unforgettable and productive summer holiday here in BGI-Shenzhen. After this summer school I graduated from university and decided to join the institute.*"

At the time of our final interviews (April 2017), BGI was recruiting people with a computing background to provide the specific skills needed to handle big data. Experienced BGI researchers observed that [Computer scientists] "*don't need to be experts in biology. They don't need a huge knowledge. They only need a simple understanding of the processes".* (NY statistician in BioBig Data Group interviewed 11th April 2017) OC (Director of Bio-BigData Group, interviewed 11th April 2017) added the further observation at this point that "*if you tell them too much they may get confused by the terminology*".

Leaders of a new project might spend a day explaining the basic issues to the data scientists before starting to work on a model. More experienced staff, who understood both sides of the problem, could then advise on which models were better. NY (statistician in BioBig Data Group interviewed 11th April 2017) further observed that is it harder for biologists to understand computer modelling - as they lack the mathematical and computing skills - than for AI/computer specialists to apply their models to biological problems.

[9] BGI's 2018 Annual Report (BGI Genomics Co 2019:49) shows the number of staff conducting research as growing rapidly 2016 448 2017 523 2018 628;
[10] Interview with Director of BioBig Data Group, Institute of Research, 11/04/2017.

BGI then train up these recruits internally.[11] Training is provided in the methods and tools applied by BGI and is achieved in the course of solving problems and undertaking research within research teams.

One project head (GX, Prenatal Diagnosis Developer, interviewed 17th March 2010) told us: "*in our training model. The trainees are able to work in the changing situations and pass on the knowledge and skills they received to newcomers… they are able to handle most of the R&D issues raised in our research project team. Moreover they are good enough to train the newcomers after them."*

Staff development revolves around undertaking research projects, in the course of which, candidates develop and demonstrate their capabilities (which are evaluated in staff assessment). LL, Director, System of Science and Technology (interviewed 17th Feb. 2011) noted that "*the training we provide is not isolated from practice. Actually we evaluate based on practice. Specifically after assigning the practical tasks to the researchers we judge our researchers by the approaches they adopt in the R&D practices*"

BGI had established a training centre (called 'Bronzeman'), which, rather distinctively, was organised around the modular groups addressing specific research processes (discussed in section 2.3). Training was focused upon difficulties that were encountered in carrying out research projects. Researchers from other units were also encouraged to team up to solve problems.

This model involves elements of occupational specialisation, differing from established professional models of knowledge work, by being centred around the acquisition by specialised occupations of knowledge about the specific procedures and tools within the organisation. This is typical of contexts where there is limited labour mobility, occupations have not become standardised and expert status is leveraged through within the enterprise through an Internal labour market (Procter and Williams 1988; Lee et al. 2012). BGI found it more effective to appoint graduates and MSc students from Bioinformatics than PhDs from biology and the train them up. On-the-job training (rather than recruiting Biology PhDs) was seen as a way of getting not just the right skills but also the right orientation – a focus on practice in which new recruits were enrolled in BGI's ways of working.

LL, Director, System of Science and Technology (interviewed 17th Feb. 2011), referred us to difficulties they had encountered when recruiting "*well-established researchers, professors or experts*" from outside large-scale genomic research from conventional, discipline-based, academic institutions. He went on to suggest that: "*if they do not try to change themselves,* [conventional academic] *biologists might find it is hard to work here in the BGI*". GX (Prenatal Diagnosis Developer, interviewed 17th March 2010) also highlighted the difficulties they had encountered when recruiting established disciplinary specialists, noting that "*it is hard to get them enrolled*".

[11] In 2011 BGI Group established BGI College "to improve disciplines, industry and talents via real world projects" and act as an advocate of innovative "practical project training to enable industry development" https://en.genomics.cn/en-about.html last sampled 24 Feb 2020

BGI staff thus develop their knowledge and expertise through their involvement in projects – improving their understanding of particular tools and processes and how they can be applied in particular investigations in practice (rather than, for example, through formal training/certification). Staff develop their capabilities and reputation, and are able to navigate BGI's informal internal career structures, through their involvement in a succession of projects. Here BGI managers recognized competing trajectories for expertise development between encouraging specialisation and generalisation. Specialisation was beneficial for achieving higher levels of expertise and was encouraged, but had to be balanced against the risk it might reduce role flexibility and willingness to embrace change. Thus MX, Marketing Director for China, (interviewed 18th February 2011) noted that "*on some occasions we need to get the disciplinary specialisation extremely specialised. One benefit of doing so is to enable the researcher to be highly focused on his or her disciplinary expertise in exploring cutting edge technologies… After a long period of time however we notice that researchers following such a specialised direction progress well in furthering understanding and application of expertise on their own specialised disciplines but at the cost of narrowing the horizontal field of view. Consequently we need to maintain the balance in modifying the structure of systems … so as to facilitate both the vertical exploration and the horizontal interactions.*

LL, Director, System of Science and Technology, (interviewed 17th Feb. 2011) also highlighted the tension between specialisation and broader roles: *"fine specialization .. prioritises stability rather than embracing changes. Following this directions, researchers build and enhance their R&D capability vertically based on their disciplinary background. The researchers of this kind thus rise, albeit slowly… The other direction is to develop broadly. The priority of organizational management following this direction is to provide chances and opportunities for such a tendency rather than to decide the actual assignment for the researchers in the R&D practices*"

The development of expert careers and capabilities within BGI is reflected in a strong internal labour market. Staff are retained not so much through high salaries (BGI's salaries were described in our 2017 visit as "only average" for the life science sector) but by the provision of job security and an attractive environment for employees. Thus BGI runs a kindergarten (though not in 2017 a primary school as staff were still so young then that they have mainly pre-school children). BGI offered low price apartments nearby (a short journey away by foot or by the firm's shuttle bus service).[12]

Though most employees pursue a career in BGI, some may move into industry or may set up companies of their own. Those wanting to go into academia could develop a PhD in collaboration with BGI's partner universities.

---

[12] The 12 RMB/m^2 monthly rent for BGI apartments in 2017 was less than a quarter of city centre rents. They are a short walk from the old BGI's premises or a short ride to the new premises on BGI's shuttle bus.

As we will see in section 2.3, BGI's strategies for developing and deploying expertise were closely related to the ways in which research projects were organized.

2.2.2 Paradoxical disciplinary contributions and identities in BGI

BGI managers' strategies for staff development needed to manage a number of tensions and paradoxes (e.g. as discussed above, regarding specialization). Another paradox was evident in relation to the contributions of different disciplines to BGI's work.

The contribution of biology was clearly foregrounded in directing the research and in publication. However, *de facto*, interdisciplinary bioinformatician and computational skills were seen as key to the practical challenges BGI confronted in undertaking high-volume sequencing.

Overall within BGI, the role of bioinformatics was described in a positive way – that stressed their mission-critical contribution in terms of their ability to find useful information. At the same time, computational specialists, though emphasizing their crucial supporting role in processing data, were keen to foreground the leading role of biologists. Thus DL (an IT specialist from the Special Force, working on sequencing software engineering, interviewed 16th March 2010) noted: "*The research project we are working on now is biological in nature. Biologists play the managing role in the research project at a macro level. These biologists know what they want and which kinds of data. But it is highly possible that they are not able to achieve the goals if they do all jobs by themselves – the data, information and results could be out of their reach. Under such circumstances, computer scientists, mathematicians and others are needed to collect and process the data and information that the biologists want. I accept that the biologists are the predominant members in the research project team and researchers from other related disciplines are more supporting than deciding*"

Though computation and bioinformatics contributions were internally foregrounded, biological knowledge was particularly emphasized where outputs are valorised in publication. Here, as we see below, BGI targeted top biology journals. DL (IT specialist from the Special Force, interviewed 16th March 2010) explained that *"in seeking* [publication] *opportunities basically we have a unified direction – journals in the biological science field. … The BGI Shenzhen itself is a biological research institute which is the decisive feature… Besides the research projects we are working on are all biological in nature. If we try to contribute articles to journals in engineering research fields for example I do not think the peer reviewers are able to understand our articles properly."*

Though there is an emphasis on interdisciplinarity – this had not led to the emergence of completely homogenised transdisciplinary experts. Instead the various different traditions are recognised as making different contributions.[13]

LL, Director, System of Science and Technology (interviewed 17th Feb. 2011) observed: "*Genomics and bioinformatics are extremely interdisciplinary in nature. Researchers' disciplinary backgrounds, other than computer science or mathematics, do not really work with these new disciplines. Computer science provides working skills and mathematics a way of thinking. But biology is not particularly useful here".*

However as Wang (2012:116) noted, in his study of interdisciplinarity in BGI, "*disciplinary boundaries do not lose their distinctiveness at the same rate… researchers with non-biology backgrounds tend to rate the role of biologists higher than those they themselves played.*"

Players mobilise identities internally within BGI around their specific skill and knowledge contribution to BGI's work. However when researchers present their role externally this is not in terms of their disciplinary background but is based on their contribution to BGI's research projects. Authors, including those with disciplinary backgrounds in computer science etc., are indeed working on biological research projects. Specific disciplinary identities and capabilities are not foregrounded when presenting outcomes to external audiences. The identity of a research project team member at BGI is seen as strong enough to establish the authors' affiliation. For BGI the affiliation of researchers to a research project is seen as more important than the disciplinary background and standing of the researcher. Thus FL, a Physicist, working as BioBig Data Project Coordinator (interviewed 16th March 2010) notes "*The identity I am to use in attending a conference is decided by the particular research project I am working on in BGI-Shenzhen. I see no direct relation with my disciplinary backgrounds. External communications are truly directed by the research project*".

2.3 Organisation of BGI Research Institute

2.3.1 Flexible evolving models based on project organisation

BGI is distinctive not just because of its scale but also in the ways it organises research activities. BGI's approach to research organisation is closely linked to the ways it deploys and develops expertise and also to the scientific and commercial goals of BGI.

In a context in which there were no established models of how best to organise this kind of facility, BGI managers have had to work out their own strategies and adapt them in the light of experience. In this, BGI has drawn heavily on templates in research labs in the West. BGI's co-founder saw the origins of the BGI project in the time he spent at the University of Washington Seattle Department of

[13] For example, GX (interviewed 17th March 2010), who was responsible for the prenatal diagnosis kit development, had a team including researchers with backgrounds in medicine, biology, computer science, and other fields.

Genomic Science (Fischer 2018). However these models have been subjected to "Chinese gaze and Chinese appropriation" (Wong 2016:S109)[14]. BGI has creatively experimented with Western organisational models and practices (Wong 2016, Stevens 2018) and has elaborated and refined in practice a distinctive model that matches the very specific strategy of BGI (Fischer 2018).[15] This kind of creative recombination is, of course, a key component of Schumpeter's view of innovation. However we note the sustained experimentation at large scale that has been a distinctive feature of China's successful recent modernisation (Shen et al. 2019).

BGI, in common with other organisations in new China, illustrates this experimental and emphatically pragmatic approach. BGI has been organised on a determinedly pragmatic manner, described by (LL, Director, System of Science and Technology, interviewed 17th Feb. 2011) as on the basis of "*consensus and common sense*" rather than the formal rules and regulations that apply to China's research and higher education system.[16] BGI managers stress that they are learning as they go along and have a constantly changing and evolving organisation structure.

> "*The BGI makes continuous efforts in modifying its* [management] *structure. We have no alternative but to do so, for what we are doing is unprecedented and no-one knows which way is the best one. Even if there is some way which might suit other institutes quite well it cannot be the best way for BGI… we are exploring our own way*" (MX, Director of Marketing in China Area interviewed 18th February 2011).

BGI has taken inspiration from organisational templates from cutting edge research and development facilities in IT and life sciences in the West. This was evident in BGI's ultra-modern new building in Dapeng which features breakout spaces and canteens to encourage informal exchanges. New practices accompany the novel architecture of the building. We were told that BGI Director and co-founder, Prof Wang Jian, when in Shenzhen, will frequently leave his office to discuss particular challenges with research staff. In our 2017 field trip we were shown a 5 metre long whiteboard in the communal space in the BGI building covered with notes of recent brainstorming sessions with BGI's leaders. Other initiatives to promote internal discussion include seminars and also informal 'tea meetings'. These were described to us as "*similar to the organizational form of 'tea meeting' widely adopted by R&D sectors within some well-established IT corporations."* (LL, Director, System of Science and Technology, interviewed 17th Feb. 2011). It is intriguing to observe that the

[14] (Wong 2016:S109) suggests that "rather than to continue to imagine an exceptional Chinese space opaque to Western knowledge, we ought to recognize that China has already created an exceptional space where Western capital and Western knowledge production have both been subjected to the Chinese gaze and Chinese appropriation".

[15] Thus Fischer (2018: 276) observes: "BGI is as much a creation of Danish and US scientific networks as of internal Chinese ones".

[16] For example China's education and research funding system was organised on rigid disciplinary lines following the Category of Academic Degree and Educational Disciplines by the Ministry of Education and Disciplinary Codes of the National Natural Science Foundation of China.

commonplace accounts of the benefits of coffee queues in promoting unplanned interactions in Western organisations become translated in the Chinese setting to revolve around tea![17]

BGI has developed flexible ways to organise research activities around project-based work. A central role is played by project managers who "*will coordinate all steps from initial experiment design to the information analysis at the completion stage"*. (KY, Manager, Marketing in the UK, interviewed 16th Feb 2011)

BGI managers emphasise that they are not very hierarchical.[18] They contrast themselves to the university system in which senior staff drive the research agenda. Instead within BGI there is a *de facto* internal market for proposals for research. "*If you have a good idea you can present it to the leading group – if it matches their plan about how BGI will develop it may get taken up*" (PZ, Director of S&T Office, interviewed 11th April 2017). In this way early career staff can develop their reputation within BGI through success in projects and go on to secure positions as project leader or international coordinator.

Manager (KY, Manager, Marketing in the UK, interviewed 16th Feb 2011) notes that promotion is driven by their performance across a series of projects: "*performance on the previous project serves as a primary indicator in qualifying one to be leader of the next project. Projects are varying in size in the BGI. Directing a big project in BGI normally works on the premise that accumulating experience in managing smaller projects is sufficient. "*

BGI had sought to establish an ecology through which people could progress their status and careers by demonstrating success in practice in "*achieving organisational goals…."* LL (Director, System of Science and Technology, interviewed 17th Feb. 2011) explained that "*Researchers are encouraged to try different posts within the ecology of the R&D system in the BGI. … All managers in our SST [*System of Science and Technology] *are qualified by their excellent performance in previous R&D practice. They all have qualified themselves in practice…. Doing this is risky for you do not know the outcomes in the beginning as you use the real projects to provide researchers with training opportunities. … We dare to take this risk for we attempt to create an open platform and a free atmosphere for the R&D as well as systematic opportunities for all researchers to work and exchange ideas together*".

Low hierarchy did not mean that work was not carefully managed. *Au contraire*, BGI operates a *dual management system.* R&D directors are in charge of the scientific direction of multiple concurrent projects. Projects are also subject to organisation-wide scrutiny by finance specialists who carry out project evaluation. In addition, the HR department plans recruitment and staff deployment to projects (LL, Director, System of Science and Technology,

---

[17] Stevens (2018) notes that BGI deliberately cultivates elements from Silicon Valley including the availability of cheap healthy snacks on every floor. Teamwork and health (through exercise, diet and gene screening services) are encouraged in tandem.

[18] Stevens (2018:100) notes that "a sort of antihierarchy is enforced: business suits are forbidden at headquarters".

interviewed 17th Feb. 2011). Project organization allows greater scrutiny over progress. BGI undertakes assessments of the probable success rate of individual project teams. Major programmes would be organized into multiple project teams working in parallel to guarantee higher success rate.

Specialisation could be productive in heightening capabilities but this could narrow the ability of a researcher to contribute to related fields. To redress this, BGI had developed a strategy they conceptualised as modular organisation of research (see below). Rather than developing expertise in all tasks, groups would develop capabilities in a sub-set of interconnected activities thereby developing bundles of expertise that might be applied in adjacent settings. "*we do not name a collective based on similarity of disciplinary background… rather we call those focusing on the same technology a technology family. Such a family creates its own organizational form to make joint efforts in handling technology*". (LL, Director, System of Science and Technology, interviewed 17th Feb. 2011)

Our fieldwork revealed a range of such problem-oriented groups. Some were formally established groups, notably:

- *System of Science and Technology*, a specialised department involved in designing, organising and coordinating research activities around particular research or service projects or groups of projects with a targeted market.
- *Special Force*[19] brought together interdisciplinary teams with diverse backgrounds - biology, computer science, mathematics, physics. It was formed to address particular research challenges in sequencing projects of bio-information analysis of sequencing data.

Various ad-hoc groups were established with more specific purposes. Examples include Group of Bioinformatics for Rice Sequencing Project, Group of Data Publication for Rice Sequencing Project, Group of Database Design for Cotton Sequencing. Participation and perceived contribution to the work of these groups brought a sense of '*honour*'. Such status and reputational benefits were widely adopted as a systematic incentive in BGI.[20]

MX Director of Marketing (interviewed 18th February 2011) noted that a new conception for modular work had been proposed the previous year, whereby "*modules can keep on improving their R&D capacity by improving their R&D capability by splitting and reorganising.*" Modular organisation enabled abilities developed in one project to be applied and extended in similar projects. *"For instance"*, as the Manager of marketing in the UK (KY interviewed 16th Feb. 2011) observed: "*the leader of a project [on the genomics of the falcon] was previously in*

[19] The title *Special Force* was chosen to be attractive to a cohort whose average age was surprisingly young, around 25.

[20] The range of avenues through which positive contributions received recognition included bulletin board, posters and reports in BGI's internal magazine (A Drop of BGI, 'Huada Diandi ' in Chinese). The names and photos of model staff were displayed on a board entitled the 'Hall of Fame'. Praise from the top, which made able researchers famous among all staff members, was particularly important for neophytes.

charge of a project researching genomes of ducks. It becomes easier for this leader given the previous experience in managing research activities around a bird."

Though LL (Director, System of Science and Technology, interviewed 17th Feb. 2011) did not wish to counterpose BGI's methods to the traditional ways of organizing research - emphasising that: "*Our perspective is by no means a negation of the traditional way of scientific research*" - his account goes on to suggest a very different approach to traditional conceptions of research as driven by theory and hypothesis testing. Instead BGI's work is focussed on the accumulation of data. "*On most occasions we merely regard what to do as a general direction. We accumulate data towards this direction and then we have our findings based on the data. However we rarely form hypotheses at the very beginning stage without data*"...

BGI is involved in a hugely complex pattern of research activities. They have a strategy committee that meets with research groups to discuss the kinds of research and services that they might be able to undertake.

2.3.2 Balancing research and commercialisation/exploration/exploitation

In contrast to Stevens's (2011) account of the commodification of a commercial genomics facility in the USA, BGI simultaneously pursued BOTH cutting edge scientific contributions AND an efficient sequencing service that could sustain itself in the market. This involved attending simultaneously to both research outputs and costs of service (Cyranoski 2010).

JL Experimental Investigator (interviewed 18 Mar 2010) observed "*as a research institute, the BGI-Shenzhen needs to keep its research at a cutting edge level. Meanwhile its research outputs, including the state-of-the-art technologies it develops, need to be commercialised and further industrialised. To achieve its goals in these 2 sides, the BGI is really in need of a very good meeting point which will connect the 2 sides ideally. Consideration of costs means much more to industrialisation than it means to pure research... efforts are more directly made in the optimisation of research outputs so as to make them more appealing in the market.*"

One consequence of their strategy to develop and sustain very large-scale sequencing is that they simultaneously pursued a mix of projects in which there were trade-offs between commercial and scientific motives. These trade-offs were reflected in differential charging – with lower charges being agreed for scientifically promising studies if their collaborators lacked funds. Yang Huanming, BGI Chairman and co-founder, (cited in Cyranoski 2010), "*hopes that collaborators will pay half of the estimated costs of the genomes they want sequenced and then publish jointly, but for interesting projects he will cover 70% or even all of the cost if the collaborators lack funding.* Wong (2016) sees this as representing the commodification of research.[21] In contrast, Leonelli (2012) flags

[21] Wong (2016) argues that "BGI's blatant openness with its offer of lower-cost service in exchange for shared authorship" represents "the conversion of a credit economy into a commodified, money economy" (Wong 2016:S108).

that the increasing scale and data intensity of research is "changing what counts as good science" (Leonelli, 2012:3)

As we see in the next section, the key driver for BGI was the scientific impact (reflected by papers in top journals from major investigations, often in partnership with leading international research groups) that could be leveraged from the large data sets it could relatively quickly and cheaply generate with its efficient sequencing capabilities. These capabilities also underpinned BGI's competitiveness in health (BGI runs gene screening services for hospitals) and commercial services. BGI is in the process of building up its commercial strategy. It seeks to develop commercial industrial applications through industrially-oriented activities in core target sectors: "*the main directions of the BGI-Shenzhen's industrialisation efforts are: agriculture, bio-energy; healthcare and some other. However these industrial applications in fact are now at the stage of incubation. Hopefully they will become the core competitiveness of the BGI in the near future, otherwise our current R&D model will not be able to support the whole organisation in the long run*" (FL, BioBig Data Project Coordinator interviewed 16th March 2010)

Despite these efforts, by 2017, BGI Group was still getting half of its funding from government (with more coming from Shenzhen than national government) rather than its various commercial services.

2.4 Research Outputs

BGI is very aware of the problems that might confront a large-scale life science facility and especially one in China far away from traditional centres of excellence in biology in terms of attracting leading scientists and getting work published in leading international journals. In a context in which many Chinese academics experienced difficulties getting their work published in leading international (i.e. Western) journals, BGI has even developed its own journal: *gigascience* - an open access, open data, open peer-review journal focusing on 'big data' research from the life and biomedical sciences.[22]

BGI staff are expected to publish. It has an annual target of 40 papers in top journals. However publication targets are for groups not individuals. They have in-house rules about how they will record the contribution of different authors to papers according to their inputs.

BGI has an international strategy to develop research networks and partnerships in a number of areas. As BGI vice president, Xiuqing Zhang, noted (cited in Fox & Kling 2010:190): *"Our goal is to build partnerships and collaborations around the world that contribute to our global society. Creating solutions that enhance agriculture and food production, for example, are a key focus for us."*

BGI's large-scale sequencing efforts started off with a few high profile projects :

[22] https://academic.oup.com/gigascience

the rice genome in 2001-2002, the Severe Acute Respiratory Syndrome (SARS) virus in 2003, the Asian Human Genome and the Giant Panda Genome and subsequently broadened out to include economically-important targets: potatoes, chickpeas, silkworms, soft shell turtles, asparagus, chickens (Normile 2012, Mahajan 2014). By 2019, BGI could point to 2097 publications[23] including many papers in top-tier journals, such as *Nature, Science* and *Cell*. [24] Figure 1 shows the number of papers with BGI authors published each year for 2009-2018 in Nature (40), Science (29) and Cell (12).

INSERT FIGURE 1 HERE

Access to BGI's unparalleled volumes of sequencing data has become a key attractant in establishing international partnerships with leading researchers and institutions worldwide. BGI has been successful in leveraging their huge genomic databases and facilities to secure highly influential publications. In 2018, the BGI chairman, CEO and Board chair were listed amongst the most highly cited 1% of researchers in their field over the last decade.[25]

3 Analysis and conclusions

Large scale sequencing laboratories and other centralised repositories are in the vanguard of visions of transformation of life science.

In Western academic contexts, the introduction of new tools to generate and analyse data at large scale is beginning to reshape scientific practice. However the adjustment has been gradual and relatively diffuse. Though we find an increasing array of information related roles in the laboratory, we do not (yet) see the kinds of wholesale transformations of disciplinary boundaries and ways of organising research as was suggested by those who predicted a separation between wet and dry biology (BBSRC 2003). Large–scale facilities, and in particular commercial laboratories, provide a setting in which more distinctive

[23] https://www.bgi.com/resources/bgi-publications sampled 21 Mar 2019
[24] Early publications included some papers with very high impact :
Yu, J (2002), 'A draft sequence of the rice genome', *Science*, 296 (5565) 79-92. 5 Apr 2002
Wong, GKS et al. (2004) 'A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms' *Nature* 432 (7018) 717-722 9 Dec.
Wang, J et al. (2008),' The diploid genome sequence of an Asian individual' *Nature*, 456 (7218) 60-65. 6 Nov.
Zhang, GJ (2014) 'Comparative genomics reveals insights into avian genome evolution and adaptation', *Science*, 346 (6215) 1311-1320. 12 Dec
A striking feature of these publications is the very large numbers of authors (around 100). While the papers on the chicken and avian evolution included a significant minority of non-Chinese authors, the gene sequence of rice and of an Asian individual seems (perhaps uncoincidentally) to have been an exclusively Chinese project!
[25] Dr. Jian Wang, Chairman of BGI; Dr. Huanming Yang, Chairman of Board of Directors, BGI; and Dr. Xun Xu, CEO of BGI were listed as highly cited researchers (HCR) in category of molecular biology and genetics in recently published list by Clarivate Analytics. The HCR 2018 list is compiled of more than 6,000 scientific elites from around the world whose citation records were among the top 1% of most-cited publications in their field over a decade.
http://en.genomics.cn/news/show_news?nid=105743
www.twitter.com BGI International @bgigroup_intl posted Dec 3, 2018

kinds of epistemic culture might be established. Some large-scale facilities have adopted distinctive models of research organisation (Collins et al. 2003) Commercial operations in particular have applied strategies from industrial research and innovation to standardise, routinize and cheapen processes (Stevens 2011). Parallels have been drawn between these developments and the increasing salience in physics of large-scale facilities. Studies have addressed how these large collaborations have been able to deploy large numbers of staff with diverse disciplinary backgrounds (Galison 1997). Is Biology going to be built around (and indeed might it carry forward) this model of Big Science?

Vermeulen (2016) conversely draws attention to constraints in the life sciences against this kind of 'supersizing' of biology (the dispersed nature of the research materials; dispersed scientific communities; fractured funding mechanisms) which mean that large-scale collaboration in life science continues to have a multi-site networked form. These constraints may however be mitigated in the case of gene sequencing facilities that are able to exploit economies of scale rooted in the capital intensity of automated equipment and data infrastructures, process efficiencies, economies of scope for data re-use (and in BGI's case in its involvement in building new analytic equipment).

BGI represents an exceptionally large-scale and radical experiment in the organisation of life-science research. The outcomes from this experiment are still emerging. Our findings do not, however, support accounts of a simple trajectory towards the commodification of laboratory processes (Stevens 2011, Wong 2016). Instead BGI sought to work out how to apply these powerful automated techniques to new scientific challenges and to create useful and scientifically valuable knowledge from analysing and interpreting data, as part of a complex and evolving hybrid development strategy that also included exploring commercial opportunities.

What was most striking about BGI's evolving efforts to recruit, train and effectively deploy computational and life science expertise was the need to balance a number of factors which some have portrayed as being in tension (Collins et al. 2003, Kleinman and Vallas 2006, Stevens 2011). BGI managers have elaborated and continually adapt strategies to manage complex and paradoxical requirements. BGI pursued efficient sequencing services while at the same time leveraging large-scale data sets to make significant scientific contributions, reflected in publications in high status journals. Goals of advancing knowledge and commercial exploitation were pursued simultaneously through a complex mix of projects.[26]

New models of work organisation are emerging in BGI, loosely adapted from forms emerging in innovative research and development settings in the USA

---

[26] These complex goals are conveyed in BGI's mission statement:"
Mission - Trans-omics for a better life.
Vision - To be a world leader in the age of life sciences.
Core Values – Curiosity, Application of Knowledge, Working for the Betterment of Mankind.
https://en.genomics.cn/en-about.html last sampled 24 Feb 2020

(Chen 2013, Wang 2016, Fischer 2018).[27] They reflect an established Chinese strategy to draw on successful experiences of others in the world for achieving our own goals (借他山之石,逐已身之玉). These models – revolving around team-based project work - are in many ways less hierarchical than traditional academic research organisations (Steven 2018). BGI's modular ecosystem of projects was subject to a dual management system, addressing both research goals and resource costs. Researchers were mainly recruited without a PhD from life science, bioinformatics and computing backgrounds and then trained up "on the job" in BGI's processes. Such staff proved more willing/able to work in BGI's distinctive research culture than conventional academic specialists e.g. with a PhD in Biology. An internal market for research suggestions offered some scope for autonomy for researchers and enables early career staff to develop research trajectory and reputation.

These developments represent a distinctive model for the formation and deployment of expertise in life science, closely linked to the goals and organisation of research and sequencing activities. BGI's evolving strategies differ sharply from traditional discipline-based small-scale academic research. They moreover diverge significantly from patterns that have been identified in bioinformatic research in the West (Lewis and Bartlett 2013; Bartlett, Lewis and Williams 2016) and in large-scale academic and commercial genomic facilities (Stevens 2011, 2013, Hilgartner 2017).

We have highlighted some distinctive features of the epistemic system that has evolved to meet the stabilised instrumental goals of BGI. The models for the creation, deployment and development of expertise emerging in BGI differ from the models prevailing in contexts dominated by traditional disciplinary structures. Scholars and commentators have emphasised the potential costs and risks of cross-disciplinary careers, with the implication that excellent researchers need to return to academe and to disciplines to validate their contribution. However BGI has established itself and continues to sustain its position as a highly successful large-scale facility. Staff are recruited at an earlier stage in their professional formation and induced into the particular organisational culture and working methods of BGI. They are able to develop careers through an internal labour market – by demonstrating their success in contributing to BGI's collective goals across a series of projects. Elements from various disciplines are enrolled and activated to carry out and improve the various processes involved in sequencing research. The epistemic endeavours and contributions of researchers are oriented towards BGI's research projects and broader mission. They do not necessarily make contributions to the conceptual core of individual disciplines.

We do not, however, find wholesale homogenization of expertise in this process. There is still a distinct division of labour and knowledge across BGI, but this is elaborated in relation to the exigencies of the knowledge production process and the techniques and technologies deployed. Knowledge is problem-focused and

[27] BGI founder Wang Jian saw this development as emerging from his time at the University of Washington in Seattle which kicked off his efforts to build a "bridge between Beijing and the U.S." (Fischer 2018:276).

linked to the evolving modular organization of research tasks. Capabilities are validated through contributions to BGI's overall outputs rather than through conventional academic disciplinary structures. Wang (2012) describes this as a 'paradiscipline', in the sense of 'being beside and beyond discipline'.

BGI's model of work organization shares some features with the Broad Institute – the commercial sequencing facility described by Stevens (2011) - in terms of team based work organization and the development of internal careers by developing capabilities on-the-job. Stevens suggests commodification pressures in Broad have resulted in an industrial model, through the application of scientific management techniques of lean production. In BGI, speed and scale are pursued as a means of developing their commercial services and also securing top-level scientific contributions. BGI is not just a genomics factory (c.f. Cyranoski 2010).

BGI has been concerned to amass data through both research and sequencing services. Much of their work is not driven primarily by research questions at this stage. At the same time, the access BGI offers to its ever-increasing data repositories creates a powerful attractant for research communities worldwide. Leading international academic science groups have formed partnerships with BGI and have been very successful in leveraging BGI's data resources and facilities to generate publications in leading journals. [28]

This study offers opening snapshots of a development that is still at a relatively early stage and that continues to adapt and evolve. Our analysis has identified a number of tensions. These are likely to persist but may be redistributed by further reorganisations. It will be important to track further these developments in this exceptional site and explore how they fit within the evolving world of high-volume life-science research.

The study supports the case made by Bartlett et al. (2016) for longitudinal investigation of these emerging developments. At the same time, it calls into question the many studies of interdisciplinary research which, conducted from the perspective of a discipline-based institutional context, have been quick to see departures from this model as risky for the individuals involved in terms of loss of autonomy and career prospects. Such perspectives, focusing on the orderly (discipline-based) delivery of arrays of knowledge in higher education, have tended to underplay the complex and unruly assemblages of knowledge and capability in real-world research processes. In contrast in this large-scale research setting, stabilised instrumental goals – balancing scientific and

[28] See for example papers in *Nature* and *Science:*
Qin, J et al 2010 "A human gut microbial gene catalogue established by metagenomic sequencing' *Nature* (464) 59-65
Rasmussen, M et al. 2010 'Ancient human genome sequence of an extinct Palaeo-Eskimo' *Nature* (463) 757-762
Li, R. (2010) 'The sequence and de novo assembly of the giant panda genome' *Nature (*463), 311-7
Xi, Q. et al. 2009 'Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (Bombys) *Science* 326 (5951) 433-6.

commercial goals - have allowed the emergence of a distinctly different mode of organising research and of forming and deploying expertise.

3.1 Limitations and implications for future research

The findings presented here capture some of the opening stages in a fascinating story that is still unfolding with within BGI and the wider field of life science research facilities. Additional research is needed to track these developments over time and establish their significance. For example we have pointed to the recruitment of diverse information specialists (computer science, bioinformatics) and their acquisition through internal training of the specific skill combinations needed to fulfil carry out BGIs processes and how they develop careers largely through an internal labour market. But will this be sustained?[29] Alternatively will its research staff seek employment in industry or in more traditional academic settings requiring some harmonisation and interoperability of roles? Will BGI staff wanting careers in mainstream research and educational institutes need to seek more conventional qualification for example by obtaining a PhD? These developments merit continued attention over an extended duration.

This paper revolves around a detailed study of a single facility. Though we can derive insights by contrasting these findings with accounts of facilities elsewhere, comparative studies would have particular value.

****

Acknowledgements

We wish to thank BGI for providing access for our fieldwork in 2010-11 and again in 2017. KaiWang would also like to thank Professor Huanming Yang from BGI and his PhD supervisors, Professor John Dupre, Professor Barry Barnes, Professor Sabina Leonelli for their support during his study in ESRC Centre for Genomics in Society (Egenis) at Exeter University. We gratefully acknowledge helpful feedback from Dr Hannah Farrimond at Exeter Univer and Dr Jane Calvert, Dr Nik Vermeulen and colleagues at Science, Technology and Innovation Studies at the University of Edinburgh.

4 References

Anon. (2010) Do scientists really need a PhD?. *Nature* **464,** 7 https://www.nature.com/articles/464007a doi:10.1038/464007a

Barry, A., Born, G., and Weszkalnys, G. (2008) 'Logics of interdisciplinarity', *Economy and Society*, 37, pp. 20–49

[29] Experts relying solely upon Internal Labour Markets run the risk of becoming dependent workers (Procter and Williams 1998).

Bartlett, Andrew, Jamie Lewis and Matthew L Williams (2016) 'Generations of Interdisciplinarity in Bioinformatics', *New Genetics and Society* 35(2): 186–209.

BBSRC (2003) *"Towards predictive biology" Bioscience for Society: A Ten-Year Vision for bioscience,* The Biotechnology and Biological Sciences Research Council: Swindon.

BGI Genomics Co. Ltd (2019) 'Annual Report of BGI Genomics Co. Ltd 2018' (in Chinese). Retrieved from: http://www.cninfo.com.cn/new/disclosure/detail?orgId=9900031781&announcementId=1206094024&announcementTime=2019-04-25 last sampled 23 February 2020.

Bruce, Ann, Lyall, Catherine , Tait, Joyce and Williams, Robin (2004), Interdisciplinary integration in Europe : the case of the Fifth Framework programme. *Futures*, 36, 2004 , p. 457-70

Calvert, J. (2010) 'Systems Biology, Interdisciplinarity and Disciplinary Identity', in J. N. Parker, N. Vermeulen, and B. Penders (eds), *Collaboration in the New Life Sciences*, Farnham: Ashgate, pp. 201–218.

Chen, Shu-Ching Jean (2013) 'Genomic Dreams Coming True in China'. *Forbes Asia*. Posted 28 Aug. 2013. Sampled 21 Mar 2019 https://www.forbes.com/sites/forbesasia/2013/08/28/genomic-dreams-coming-true-in-china/#24008fd22760

Collins, Francis S., Morgan, Michael and Patrinos, Aristides (2003), 'The Human Genome Project: Lessons from Large-Scale Biology' *Science,* 11 April 2003 Vol. 300 pp. 286-91

Davies, Kevin (2010) 'BGI Americas Offers Sequencing the Chinese Way', *Bio-IT World*, September 28. http://www.bio-itworld.com/2010/issues/sept-oct/bgi-us.html last sampled 10 February 2020

Eyal, G.(2013), For a sociology of expertise: The social origins of the autism epidemic, *American Journal of Sociology*, 118 (4): 863–907

Fischer, M.M.J. (2018), 'A Tale of Two Genome Institutes: Qualitative Networks, Charismatic Voice, and R&D Strategies—Juxtaposing GIS Biopolis and BGI', *Science, Technology and Society*, 23, (2) pp. 271-288.

Fleck, J. (1988), 'Expertise: knowledge, power and tradeability', chap 7, pp. 143-172, in Williams, R, Faulkner, W., and Fleck, J. (eds) *Exploring Expertise*, , Basingstoke: Macmillan.

Fox, J. and Kling, J. (2010). 'Chinese institute makes bold sequencing play', *Nature Biotechnology*. 28 (3): 189–191.

Frank, Lone (2011), China is rewriting the book on genome research,

News Week, 24th April
https://www.newsweek.com/china-rewriting-book-genome-research-66513
last sampled 10 February 2020

Fuller, A. & Unwin, L. (2010), 'Knowledge Workers As The New Apprentices: The Influence Of Organisational Autonomy, Goals And Values On The Nurturing Of Expertise', *Vocations And Learning* 3: 203–222.

Galison, P. (1997). *Image & logic: A material culture of microphysics.* Chicago: The University of Chicago Press.

Garcia-Sancho, Miguel (2012), *Biology, Computing and the History of Molecular Sequencing: From Proteins to DNA, 1945–2000,* Basingstoke: Palgrave Macmillan.

Harvey, M., and A. McMeekin (2002), *UK Bioinformatics: Current Landscapes and Future Horizons.* London: Department of Trade and Industry

Highfield, Roger (2016) 'What's Wrong With Craig Venter?' *New Republic,* 2 February 2016 available online at
https://newrepublic.com/article/128977/whats-wrong-craig-venter

Hilgartner, S., (2017), *Reordering Life: Knowledge and Control in the Genomics Revolution,* Cambridge, MA: MIT Press

Kleinman, Daniel Lee and Steven Vallas (2006) 'Contradiction in Convergence: Universities and Industry in the Biotechnology Field', pp. 35-62 in Scott Frickel and Kelly Moore (eds.), *The New Political Sociology of Science: Institutions, Networks, and Power*. Madison: The University of Wisconsin Press.

Kulski, Jerzy K. (2016). Next-Generation Sequencing — An Overview of the History, Tools, and "Omic" Chap.1, p 4-59 in Jerzy K Kulski (ed.), *Applications, Next Generation Sequencing - Advances, Applications and Challenges*, IntechOpen, DOI: 10.5772/61964. Available from: https://www.intechopen.com/books/next-generation-sequencing-advances-applications-and-challenges/next-generation-sequencing-an-overview-of-the-history-tools-and-omic-applications

Larson , Christina (2013) 'Inside China's Genome Factory' *MIT Technology Review* 11 February 2013.
https://www.technologyreview.com/s/511051/inside-chinas-genome-factory/

Lee, Hsing-fen, Miozzo, Marcela and Larédo, Philippe (2012) 'Job mobility of science and engineering PhDs: 'movers' and 'stayers' and implications for knowledge flows to industry' Paper presented at *DRUID 2012 Conference* (19-21 June 21, CBS, Copenhagen, Denmark) *available online at*
https://conference.druid.dk/acc_papers/fai70r8o69h3eoylev1x93ko6kd0.pdf
Last sampled 10 February 2020

Leonelli, S (2012) 'Introduction: Making sense of data-intensive research in the biological and biomedical sciences'. *Studies in History and Philosophy of Science*

Part C: Studies in the History and the Philosophy of the Biological and Biomedical Sciences 43(1): 1–3.

Lewis, J., and A. Bartlett. 2013. 'Inscribing a Discipline: Tensions in the Field of Bioinformatics.' *New Genetics and Society* 32 (3): 243–263.

Lewis, J., Bartlett, J. and Atkinson, P. (2016), 'Hidden in the Middle: Culture, Value and Reward in Bioinformatics', *Minerva*, 54:471–490.

Mahajan, Neelima (2014) 'Genetics Research: Decoding BGI's Genes' CKGSB Knowledge, Posted 20 Jan 2014. Sampled 21 Mar 2019 http://knowledge.ckgsb.edu.cn/2014/01/20/technology/genetics-research-decoding-bgis-genes/

Nerland, Monika and Jensen, Karen (2012), 'Epistemic practices and object relations in professional work' *Journal of Education and Work*, 25(1), 101-120.

Normile, D., 2012. China's sequencing powerhouse comes of age. *Science*, 335 (6068), 516–519.

Senabouth, A., Andersen, S., Shi, Q., Shi, L., Jiang, J., Zhang, Z., Wing, K., Daniszewski, M., Lukowski, S., Hung, S., Nguyen, Q., Fink, L., Beckhouse, A., Pébay, A., Hewitt, A., and Powell, J. (2019) Comparative performance of the BGI and Illumina sequencing technology for single-cell RNA sequencing, *bioRxiv* preprint first posted online Feb. 17, 2019. https://www.biorxiv.org/content/biorxiv/early/2019/02/24/552588.full.pdf , (accessed on 11/03/2019)

Shen, Xiaobai, Williams, Robin, Zheng, S., Liu, Y., Li, Y. & Gerst, M., (2019) 'Digital online music in China – a "laboratory" for business experiment ' *Technological Forecasting & Social Change*, 139, 235-249.

Stevens, H. (2011) 'On the means of bio-production: Bioinformatics and how to make knowledge in a high-throughput genomics laboratory', *BioSocieties* 6(2): 217–242.

Stevens, H. (2013), *Life Out of Sequence: A Data-Driven History of Bioinformatics*, University of Chicago Press: Chicago, Illinois.

Stevens, H. (2018), "Starting up Biology in China: Performances of Life at BGI," *Osiris* 33, 85-106. https://doi.org/10.1086/699235

Vermeulen, N (2016), 'Big Biology: Supersizing science during the emergence of the 21st century', *Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin*, vol. 24, no. 2, pp. 195-223. https://doi.org/10.1007/s00048-016-0141-8

Wang, Kai (2012) *Interdisciplinary Research as Collective Interaction: An Investigation of Interdisciplinarity in the R&D Sector of China's Biotechnology Industry*; Unpublished PhD Thesis, University of Exeter.

Wong, Winnie (Yin Wong) (2016) "Speculative Authorship in the City of Fakes" *Current Anthropology* 58, Supplement 15, S103-S112

Williams, Robin and Procter, Rob (1998) 'Trading places : a case study of the formation and deployment of computing expertise' In : Williams, Robin et al (eds) *Exploring expertise*. Basingstoke, Macmillan, Chap. 13, p.197-222