



HAL
open science

Tracking and Mapping the Use of CRISPR-Cas Systems for Research in China

Guillaume Levrier

► **To cite this version:**

Guillaume Levrier. Tracking and Mapping the Use of CRISPR-Cas Systems for Research in China. The Crispr Journal, 2021, 4 (3), pp.321-338. 10.1089/crispr.2021.0015 . hal-04309075

HAL Id: hal-04309075

<https://hal.science/hal-04309075>

Submitted on 27 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tracking and Mapping the Use of CRISPR-Cas Systems for Research in China

Guillaume Levrier

ORIGINAL ARTICLE

10.1089/crispr.2021.0015

Abstract

Scientists have floated the idea of a “Sputnik 2.0” technological race between the People’s Republic of China (PRC) and the United States of America in CRISPR-based genome editing research. This quantitative analysis of articles published between 2010 and 2020 shows that research centers based in the PRC have succeeded in making CRISPR-based genome editing a standard tool. A corpus of 18,863 academic documents containing the acronym CRISPR in their abstract shows that although PRC-based research institutions were slower to start publishing on CRISPR, they have now outpaced the publication rate of institutions located in the European Union (EU). While U.S.-based institutions have kept their leading position in basic research, PRC-based research has become momentous in agriculture-related fields. This corpus hence illustrates how deeply the international landscape of life sciences research has shifted since the Human Genome Project, mostly to the PRC’s advantage.

Introduction

Genome editing, a set of techniques derived from a capacity to trigger an organism’s DNA repair mechanisms following an endonuclease-induced disruption at a targeted site, has become mainstream¹ in research contexts between 2010 and 2020. Much of this success as a tool for life sciences academic research has been attributed to the technological developments following the discovery of Cas proteins’ potential as tracking and cleaving mechanisms.

Some commentators have argued over the existence of a new “Sputnik 2.0” race² between the US and China. According to this view, different regions of the world would be competing to achieve what could be called a “nuclease supremacy”. While this analogy might appear to be far-fetched, the People’s Republic of China (PRC) policy interest in life sciences has indeed skyrocketed over the past twenty years. A first stone was set with the ninth five-year plan (1996-2001), which mentioned the importance of biotechnology. The thirteenth five-year plan (2016-2020) explicitly lists genome editing techniques (基因编辑技术) as part of the “advanced biotechnologies” (发展先进高效生物技术). The same document argues that developing such technologies are to make the PRC an international leader in this field (提升我国生物技术前沿领域原创水平, 抢占国际生物技术竞争制高点) (see Appendix 1). Recent publications³ show that PRC-based institutions have made considerable progress in crop improvement, notably using base-editing.

This study seeks to determine to what extent research institutions in the PRC have taken advantage of CRISPR-based genome editing as a platform for research. Using the open-source software PANDORÆ (see Methods section), all the publications available on Elsevier’s Scopus service containing the acronym CRISPR in their abstract for the years 2010 to 2020 were retrieved and analyzed. These requests yielded 18 863 results, which make up the core corpus used for this research.

Material and Methods

The corpus was retrieved and visualized using PANDORÆ’s automated request feature (FLUX) to Elsevier’s Scopus service. This software is free and open source, which makes data interpretation and reproducibility easier compared to commercial solutions. Its features ad-hoc data cleaning (through the CHÆROS processes) and visualization (through TYPES) algorithms [all these processes and PANDORÆ’s complete code are available online on GitHub], making the use of other solutions such as Derwent Data Analyzer, Clarivate analytics, Bibexcel, InCites or CiteSpace redundant. However, PANDORÆ normalizes the data it retrieves to the CSL-JSON (Citation-Style Library – JavaScript Object Notation) format, and can hence yield explorable visualizations based on these different data sources or sorting algorithms.

An important caveat, directly linked to the nature of the database used, must be acknowledged: this dataset is relevant in terms of orders of magnitude only.

The metadata made available by Scopus is rich but necessarily imperfect. Some affiliations might be improperly listed, others might be missing. It does not provide users with a reliable way to distinguish reviews from research articles. Identical documents might be listed several times under different names, or as different versions of the same work. Chinese authors' names are latinized and abbreviated by the service, which makes authorship attribution difficult to manage. Whenever possible, PANDORÆ's processes use DOIs to sort data, and tries to use the longest available name form. However, one should not overestimate the precision of this data set. Although it is relevant to explore developments of thousands of articles on the scope of decade, as is done in this article, more data cleaning work would be needed to support more precise assertions. It should also be noted that PRC and CN (CN is the ISO 3166-1 alpha-2 code for China) are used interchangeably in this article, both designating China. CN being a two-letter abbreviation, it is used in the code alongside the other two area abbreviations (EU and US), which can prove practical for substring selection purposes.

To respect application programming interface (API) rate limits, a total of 11 requests were made, following the syntax hereunder. The last two characters were changed according to the year requested. In this URL syntax, the %20 corresponds to a blank space, which can be misleading in this formula. Once parsed, the end of the sequence "%2020XX" becomes "_20XX," where _ represents a blank space. The requests were made on January 22, 2021. Requests made at a different date will yield different results, as articles are corrected, retracted, or retrospectively added.

```
ABS(CRISPR)%20AND%20PUBYEAR%20%3D%2020XX
```

Each request was geocoded using PANDORÆ's matching algorithm. Each encoded affiliation city was attributed a longitude and latitude when available. Each resulting data set was then converted to the CSL-JSON format and sent to a Zotero library instance for hand curation.

To provide more insights on the use of CRISPR among disciplines and subfields, the ISSN of the publication in which each retrieved document was published was then stored in a dedicated data map. Each discrete ISSN was then requested through Scopus' ISSN API service, which yielded its profile. These journals profiles contain (for this specific corpus) 0 to 8 different keywords qualifying their editorial line. For this dataset, this request yielded a total of 212 subfields sorted in 27 categories (see Appendix 2 for the full list).

The eleven resulting collections were then exported from this library and rebuilt using a Go language script

written for that purpose. The `affilRebuild.go` (see supplements) script iterates over all the documents contained in all the CSL-JSON files available in the same directory. Each affiliation listed in a document is then added to a map whose keys are four potential targets: EU, US, CN and Other. The Go map is hence declared as a typed variable (such as `map[string][]map[string]interface{}`). The first string is the area key, and the value is a slice of Go-parsed JSON objects, in our case articles retrieved from Scopus and encoded in CSL-JSON by PANDORÆ. If the country belongs to pre-Brexit EU28, it will be added to the EU category. The script then outputs a series of CSL-JSON documents by country (or, in the case of EU country, an "EU" file), in which are listed all the documents which have at least one affiliation located to that country.

Following this script's logic, a document listing several affiliations will be duplicated as many times in the sorting map. The datasets produced by this script hence do not weigh documents, but instead weigh affiliations. A document with no affiliation listed will be dismissed. To provide more insights on this weighting, the same script also produces a document profile (Fig. 1). This profile is a map of the listed affiliation structure of each article. Two main features appear: most of the articles retrieved list only one affiliation; and most of the articles which list two or more affiliations list two affiliations in the same block (EU, US, CN/PRC or Other).

This dataset provides a global overview of how CRISPR has been written in article abstracts in different countries over the past decade. By coupling this data to Scopus' "cited by" data (as of January 2021) for each article and its the journal keyword qualifications, a much deeper perspective can be produced. The `fieldRebuild.go` script hence merges the US, EU and CN datasets with each article's citation count and their journal's keyword to generate a series of graphs per field (see Appendix 2).

The datasets were then re-imported into PANDORÆ for visualization through its TYPES processes. The geotype (Figs 4-5) represents articles which list at least one affiliation in PRC and maps them by city. On these figures, a city's radius is a logarithmic function of the number of articles which list an affiliation within its walls. This log function is necessary to ensure a very wide range of cities, some hosting numerous large institutions listed as affiliations by thousands of articles in this field, others hosting a single laboratory relevant to that same field, remain visible on a single visualization. Red edges between cities represent at least one instance in which institutions from these cities have collaborated on an article. The chronotype (Fig 6,7) styles a circular timeline based on the weighted dataset: each dot listed is one instance of an article per affiliation. Red edges connect instances with same authors. To make up for the weak quality of the date encoding in the original dataset,

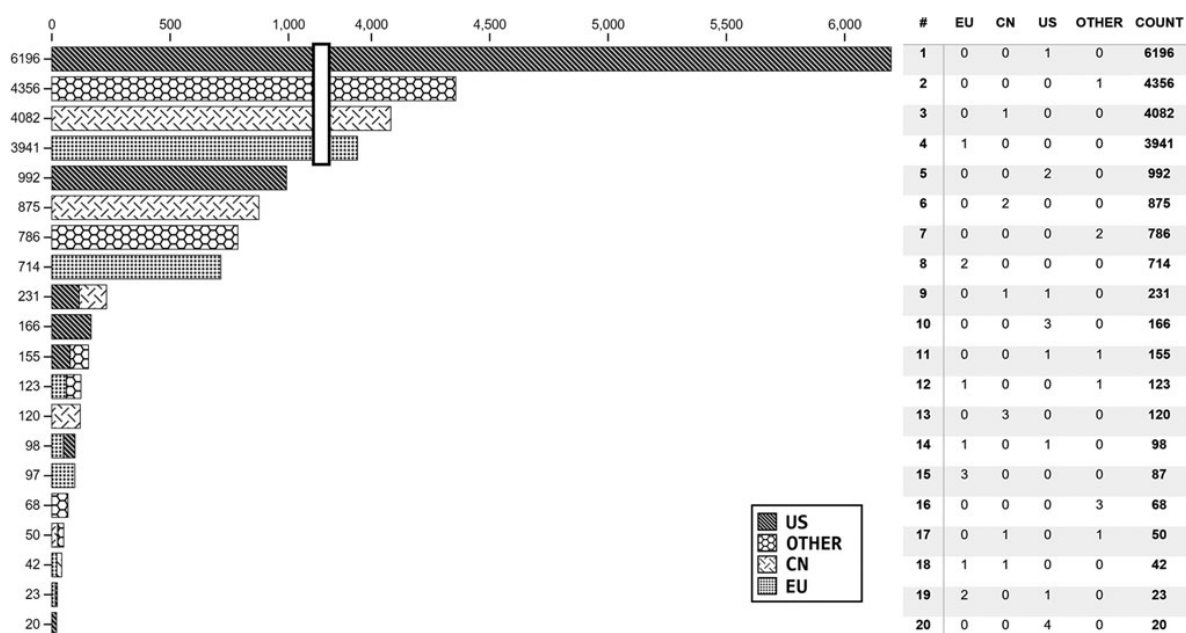


FIG. 1. Document affiliations profiles, showing the 20 most prevalent profiles. The rectangle on the bar chart indicate a broken scale. A document profile is a specific mix of listed affiliations, using four categories of affiliations (CN, EU, US, and Other). The most prevalent profile in the data set is 1 U.S. institution listed. The ninth most prevalent is 1 CN institution listed and 1 U.S. institution listed. Here, CN is equivalent to PRC. CN; PRC, People’s Republic of China.

PANDORÆ substitutes any missing month and day by the 15th of January. The beginning-of-year cluster is hence a data quality artefact, though such a cluster might exist (in much smaller proportions). The fieldtype (Fig 10) is a series of stacked area charts displaying the number of papers published in the field (dark color) per area each year and the cumulative citation count of articles published in that area that year (light colors) as of early 2021. An important methodological point must be highlighted here: if an article was published in a journal qualified with X keywords (here, X being comprised between 0 and 8), it will figure in each of the X relevant graphs (one for each subfield identified by that keyword). This can be particularly impactful for highly cited papers published in high factor journals with multiple keywords (see Fig 8). However, the keyword distribution (Fig 4) shows that its profile is similar for the three areas. Three features must however be acknowledged:

- the percentage of EU publications in journals with no keywords is higher than the other two areas, which means that the EU publications will be underrepresented in Fig 10 and Appendix 2.
- the percentage of US publications with only one keyword is higher than the other two areas, which means that it is less represented in cumulative terms (on the totality of Appendix 2).
- The “Global” category is an average of US, EU and CN, and does not take into account the “Other” area.

Results

The retrieved corpus illustrates the rise of CRISPR in academic abstracts between 2010 and 2020 (Fig. 2). The rise in publications validates the idea that a “CRISPR decade” did occur in life sciences research between these dates. Through the affiliation listings, and the profile weighting detailed in the “Methods” section, one can also observe that US-based research institutions have collectively published more articles than any other region in the world (Fig. 3). They also make up most the hundred most cited articles in the corpus (Fig 5). A second look shows that the PRC has caught-up with, and arguably surpassed, the EU in terms of number of publications. While intra-regional collaboration weighting artifacts logically deepen the gap between regions, it appears unlikely that they would do so to such an extent that it would narrow a gap of 362 items (which would still amount to more than 300 items using average weightings).

The affiliations’ geocoding highlights the importance of international and national cooperation. Cooperation between research institutions in the US and in the PRC seem to be much deeper than cooperation between the ones based in the EU with their Chinese counterparts (Fig. 6). Another observation is that intra-PRC collaboration appears to be mostly to the benefit of Beijing-based institutions (Fig 7). A more quantitative approach gives additional insights (Fig. 8). Intra-PRC collaborations are strongly carried by the collaborations listing the Chinese Academy of

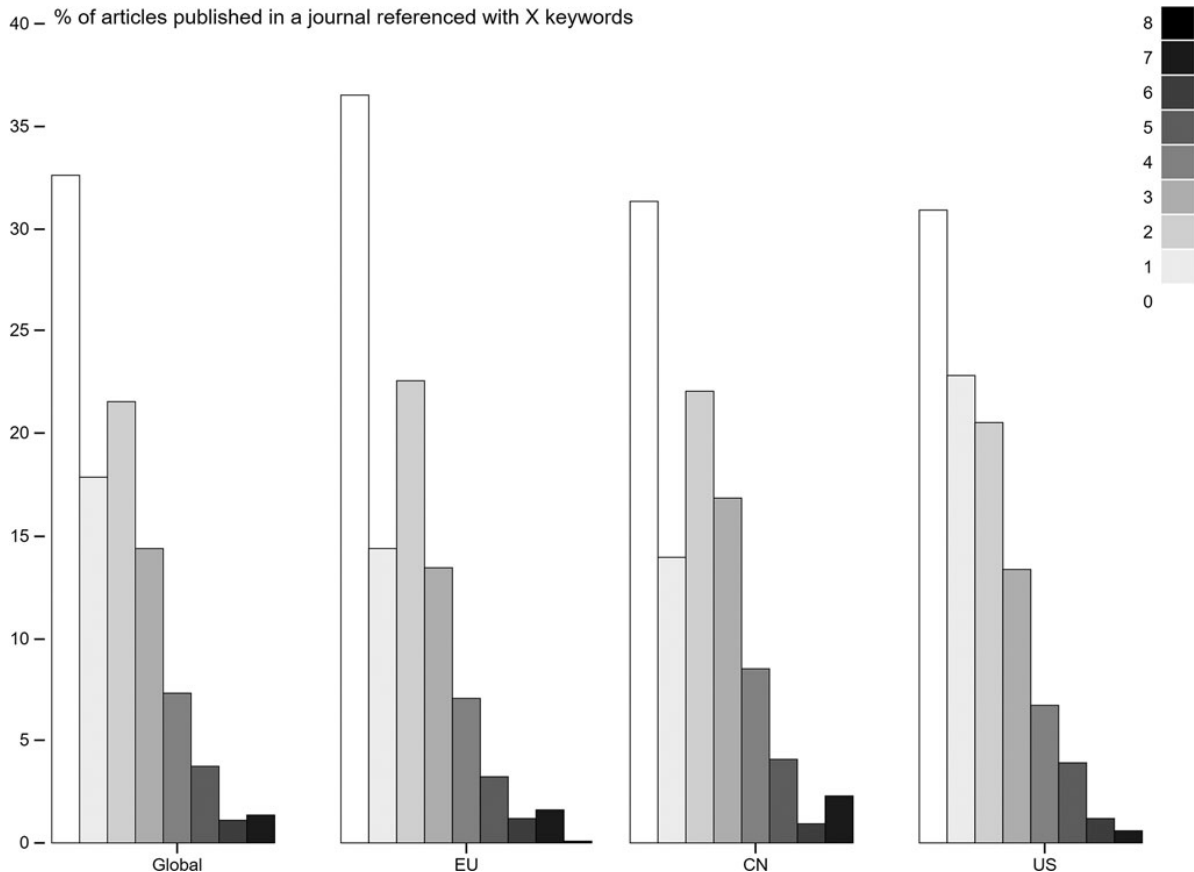


FIG. 2. Number of registered keywords distribution per area. This graph shows for example that >35% of the articles written by authors with an EU affiliation are published in journals with no Scopus-registered keyword. Global is EU, CN, US average, excluding other countries.

Science as co-affiliation). This is in the same order of magnitude as the US average (12 articles). The EU average is significantly higher (24 articles), which can be explained by the listing of several large national research institutions (such as France's Centre National de la Recherche Scientifique). International collaborations account for a very limited number of total collaboration (see blue curve in Fig. 8 and right part of the table). Out of the 30 international institutions that collaborate most with PRC-based research centers, 20 are based in the US.

A chronological perspective (Fig.9) on the dataset shows that teams that have started using CRISPR in all three areas have kept using it over the years. Zooming in on two periods (Fig. 10) confirms that collaborations between PRC and US teams through time seem to be more dynamic than with EU institutions. A more in-depth analysis (Fig 11) of this chronological data shows that affiliations in the PRC have not only caught up in "mainstream" CRISPR disciplines, such as Molecular Biology, Biochemistry, Cell Biology or Biotechnology, but have also become prominent in more specific fields, with a strong focus

on Agricultural sciences. Areas such as Plant science, Agronomy and Crop Science, Animal science and Zoology, Insect science and others feature leading PRC academic institutions in terms of both publications and citations. This is consistent with previous geographical studies of genome editing intellectual property⁴.

The main difference between authors in the PRC, the US and the UE seems to be on high-impact articles. While their publication profiles are similar for the average article (Fig. 12 a.,b.,c.), the higher-impact articles published by US-affiliated authors (Fig. 12 d.) are singularly more impactful than the higher-impact articles of either EU-based or PRC-based authors, though EU-based authors appear to still have a slight edge on that measure. This is confirmed by ranking the most 'productive' affiliations of each area (Fig. 13). The US would be in a similar position as the EU and the PRC if not for the affiliations based in the Boston area and the San Francisco Bay area, which on their own account for five times the 20 first most productive institutions of either the PRC or EU area.

DOI	Cited by	CN	US	EU
10.1126/science.1231143	7504		■	
10.1126/science.1225829	6280		■	
10.1126/science.1232033	5053		■	
10.1038/nprot.2013.143	4299		■	
10.1016/j.cell.2014.05.010	2696		■	
10.1126/science.1258096	2430		■	
10.1126/science.1247005	2258		■	
10.1016/j.cell.2013.04.025	2172		■	
10.1016/j.cell.2013.02.022	2156		■	
10.1016/j.tibtech.2013.04.004	1957		■	
10.1016/j.cell.2013.08.021	1883		■	
10.1038/nbt.2623	1747		■	
10.1038/nbt.2501	1721		■	
10.1038/nbt.2842	1712		■	
10.1016/j.cell.2013.06.044	1640		■	
10.1016/j.cell.2015.09.038	1537		■	■
10.1126/science.1246981	1482		■	
10.1038/nature15514	1306	■		
10.1038/nrmicro2577	1295		■	
10.1038/nbt.2508	1295		■	
10.1038/nature17946	1266		■	
10.1126/science.1179555	1255			■
10.1038/nature09886	1247			■
10.7554/eLife.00471	1201		■	
10.1073/pnas.1208507109	1170			■
10.1038/nature14136	1168		■	
10.1038/nbt.2507	1143			
10.1038/nature09523	1134			
10.1038/nature16526	1132		■	
10.1038/nbt.2675	1086		■	
10.1038/nature10886	1078		■	
10.1038/nbt.2808	1077		■	
10.1038/nrmicro3569	1001		■	
10.1038/nbt.3437	993		■	
10.1016/j.cell.2013.08.022	987		■	
10.1016/j.cell.2013.12.001	943		■	
10.1038/nbt.3117	903		■	
10.1093/nar/gkt135	860		■	
10.1038/nbt.3199	853		■	
10.1038/nature24644	847		■	
10.1038/nature13011	845		■	
10.1016/j.cell.2014.02.001	813		■	
10.1016/j.cell.2014.09.014	799		■	
10.1038/nbt.2969	799	■		
10.1016/j.cell.2015.04.004	794			■
10.1101/gr.162339.113	787		■	
10.1038/nature14592	775		■	
10.1101/gr.171322.113	772		■	
10.1038/nmeth.2649	750		■	
10.1016/j.stem.2013.11.002	736			■
10.1038/nbt.3026	695		■	
10.1073/pnas.1308335110	679		■	
10.1038/nrg3686	664		■	
10.1126/science.aac7041	663		■	
10.1016/j.molp.2015.04.007	651	■		
10.1016/j.cell.2014.01.027	648	■		
10.1038/nmeth.2598	629		■	
10.1093/nar/gkt780	622		■	
10.1126/science.aad5143	622		■	
10.1016/j.cell.2014.09.039	618		■	
10.1093/bioinformatics/btu048	617		■	
10.1038/nbt.2908	610		■	
10.1007/s13238-015-0153-5	602	■		
10.1038/nbt.2884	601		■	
10.1038/nbt.3198	601			■
10.1038/nrg2749	594		■	
10.1038/nature14580	594		■	
10.1056/NEJMoa1502214	590		■	■
10.1126/science.aaf5573	588		■	
10.1038/nmeth.2532	577		■	
10.1038/nbt.2800	572		■	
10.1126/science.aad5177	566		■	
10.1534/genetics.113.152710	562		■	
10.1126/science.aam9321	560		■	
10.1016/j.cell.2015.11.015	556		■	
10.1038/nrg3899	556		■	
10.1038/nature13579	555		■	
10.7554/eLife.04766	539		■	
10.1038/nbt.2889	535		■	
10.1016/j.biotechadv.2014.12.006	530			■
10.1126/science.1247997	522		■	
10.1126/science.aad5725	519		■	
10.1093/nar/gkt520	517		■	
10.1186/s13059-016-1012-2	514		■	
10.1038/cr.2013.45	510	■		
10.1016/j.cell.2015.12.035	510		■	
10.1038/nature16490	507		■	
10.1146/annurev-genet-110410-132430	502		■	
10.1038/nmeth.2681	501		■	
10.1038/nature21405	500		■	
10.1038/nature26155	492		■	
10.1016/j.cell.2014.11.052	491		■	
10.1038/nmeth.3284	491		■	
10.1101/gr.122705.111	489		■	
10.1093/nar/gku410	481		■	
10.1016/j.celrep.2013.06.020	479			■
10.1038/nbt.3439	475		■	
10.1038/nm.3802	465		■	
10.1038/cr.2013.122	460		■	
10.1038/nmeth.2641	459		■	

FIG. 3. Top 100 articles of the corpus by citations. Affiliations are attributed as one or more per area.

These observations can also be confirmed through a different perspective, that of ranking journals in which these articles have been published by number of papers published and number of citations (Fig. 14 a & b). Unsurprisingly, high impact-factor journals have mostly published highly cited articles written by authors linked by a US-based affiliation (Fig 14 b). However, PRC-based authors do publish comparatively more in journals such as *Frontiers in Plant Science* and *Plant Biotechnology Journal*, which is consistent with the keyword-based analysis presented in Fig 11.

These quantitative perspectives on the PRC CRISPR research landscape can hence be summarized as follows:

- PRC research institutions were slower to start publishing articles using CRISPR than US or EU based institutions.
- The publication rate of PRC institutions has kept accelerating between 2015 and 2020. It remains below the rate of US institutions but is now higher than the rate of EU institutions.
- The average article appears to have the same impact in the three compared areas. However, US institutions based in either the Boston or San Francisco areas have a major edge on very high impact articles.
- Publication in the PRC have a level of atomicity comparable to the US's. The average PRC institution published 15 articles relevant to this corpus over the past decade. When collaborating with other institutions, they mostly collaborated with other PRC-based institutions, and appear to have remained loyal to these institutions.
- PRC-based institutions do have an edge in specific fields. This is especially salient for Agriculture-related sciences.



FIG. 4. PANDORÆ geotype showing documents of the corpus listing at least one PRC-based affiliation. Edges link cities hosting collaborating institutions.



FIG. 5. PANDORÆ geotype, focused on Beijing for the period 2019–2020.

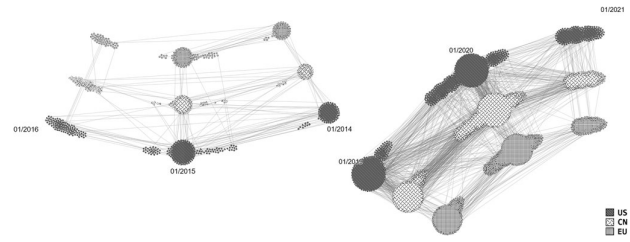
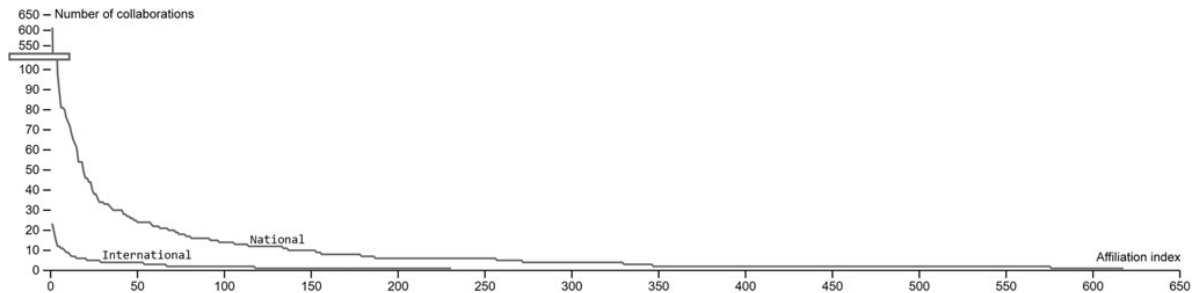


FIG. 6. PANDORÆ chronotype clamped focus on 2014–2015 and 2019–2020.



CN

Num.	Affiliation Name
607	University of Chinese Academy of Sciences
246	Chinese Academy of Sciences
126	Institute of Plant Physiology and Ecology, Shanghai Institute f...
98	Chinese Academy of Agricultural Sciences
89	ShanghaiTech University
81	Fudan University
81	Zhejiang University
80	Institute of Zoology Chinese Academy of Sciences
76	Ministry of Agriculture of the People's Republic of China
74	Institute of Biochemistry and Cell Biology, Shanghai Institute ...
72	Guangzhou Institute of Biomedicine and Health
68	Sun Yat-Sen University
65	Institute of Biophysics Chinese Academy of Sciences
63	South China Agricultural University
61	Institute of Genetics and Developmental Biology Chinese Academy...
54	Huazhong Agricultural University
54	Ministry of Health of People's Republic of China
54	Central South University
49	Fujian Agriculture and Forestry University
46	Ministry of Education China
46	Tianjin Medical University
44	Institute of Microbiology Chinese Academy of Sciences
44	Tsinghua University
40	Wuhan University
38	School of Life Sciences Fudan University
38	Chinese Academy of Medical Sciences & Peking Union Medical Coll...
36	Nankai University
34	Xuanwu Hospital, Capital Medical University
34	Zhejiang University School of Medicine
34	Northwest A&F University

International

Num.	Affiliation Name	Country
23	Harvard Medical School	United States
19	Emory University School of Medicine	United States
15	Dana-Farber Cancer Institute	United States
12	Harvard T.H. Chan School of Public Health	United States
12	University of Rwanda	Rwanda
11	Brigham and Women's Hospital	United States
11	Massachusetts General Hospital	United States
10	Aarhus Universitet	Denmark
9	Center for Discovery and Innovation	United States
9	Kyoto University	Japan
8	Charles Sturt University	Australia
7	University of North Dakota	United States
7	David Geffen School of Medicine at UCLA	United States
7	University of California, Davis	United States
6	University of Illinois at Urbana-Champaign	United States
6	UT Southwestern Medical Center	United States
6	University of California, San Francisco	United States
6	University of California, Los Angeles	United States
6	Johns Hopkins School of Medicine	United States
6	University of Los Angeles	United States
5	Chinese University of Hong Kong	Hong Kong
5	The Pirbright Institute	United Kingdom
5	City University of Hong Kong	Hong Kong
5	Duke-NUS Medical School Singapore	Singapore
5	Center for Medical Genetics and Primary Health Care	Armenia
5	Children's Hospital Boston	United States
5	University of Copenhagen, Faculty of Health Sciences	Denmark
5	Stanford University School of Medicine	United States
4	Wayne State University	United States
4	Loma Linda University School of Medicine	United States

FIG. 7. Line graph: number of collaborations per institution, ranked to most collaborative to least. The rectangle indicates a broken scale. Table: Top 30 national (left) and international (right) collaborations with PRC-based research institutions.

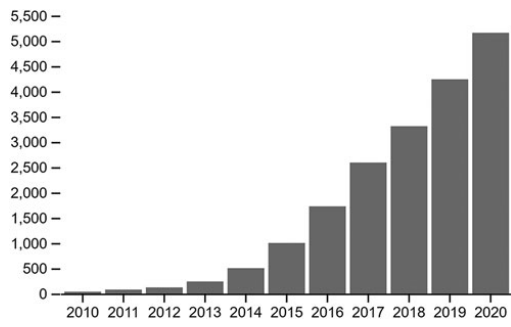


FIG. 8. Number of Scopus-listed publications containing the acronym "CRISPR" in their abstract per year from 2010 to 2020, included.

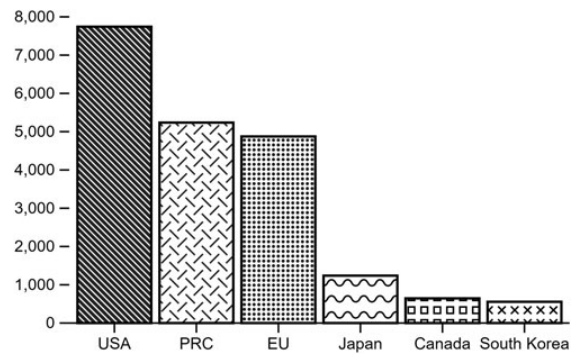


FIG. 9. Number of affiliation-weighted documents per category of countries. If a document lists one affiliation in one category, one point will be added to this category. If a document lists two affiliations in one category and a third one in another category, a total of three points will be distributed accordingly.

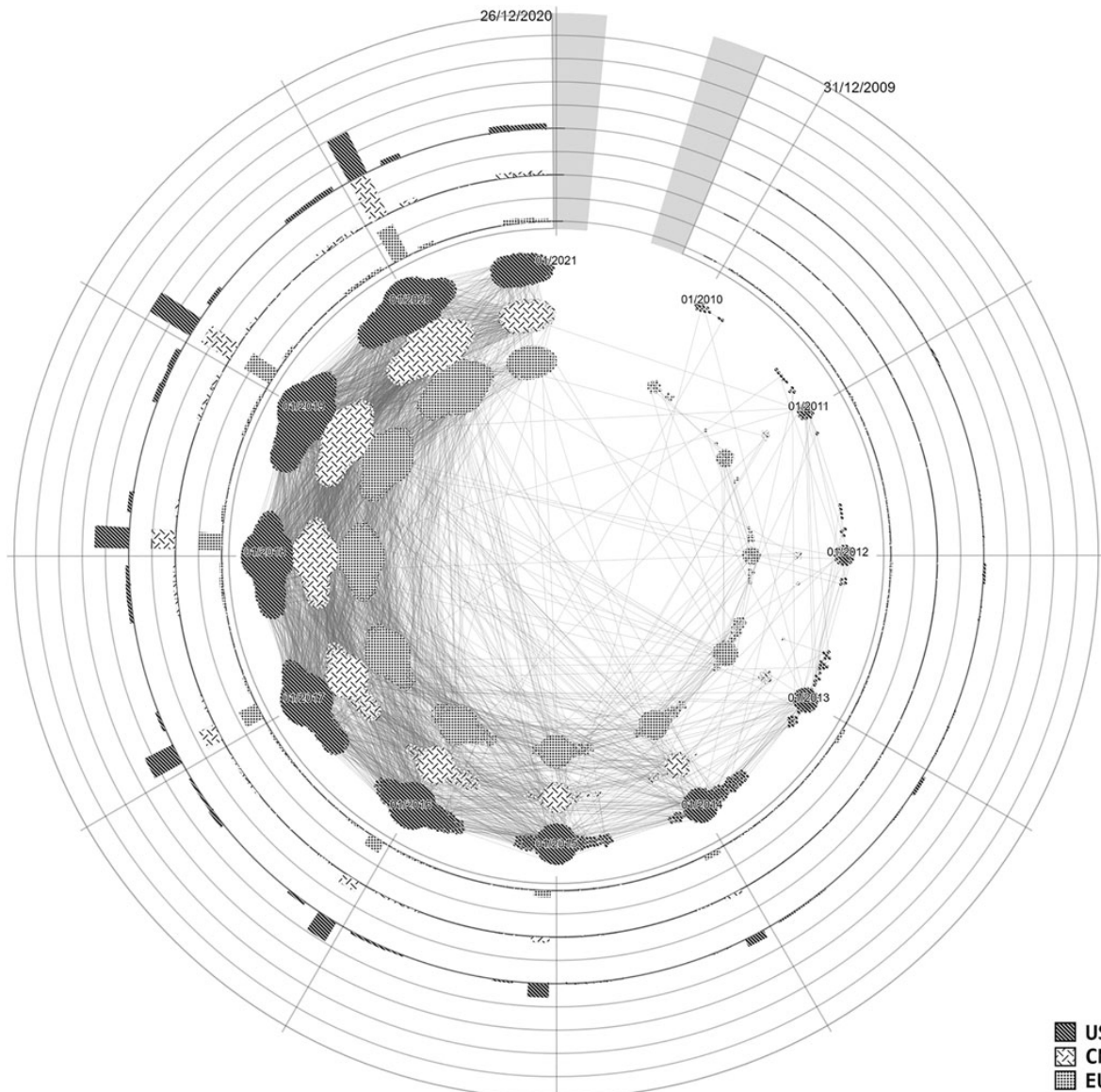


FIG. 10. PANDORÆ chronotype showing the affiliation-weighted data sets for US, PRC and EU28 over the duration of the chosen timeframe.

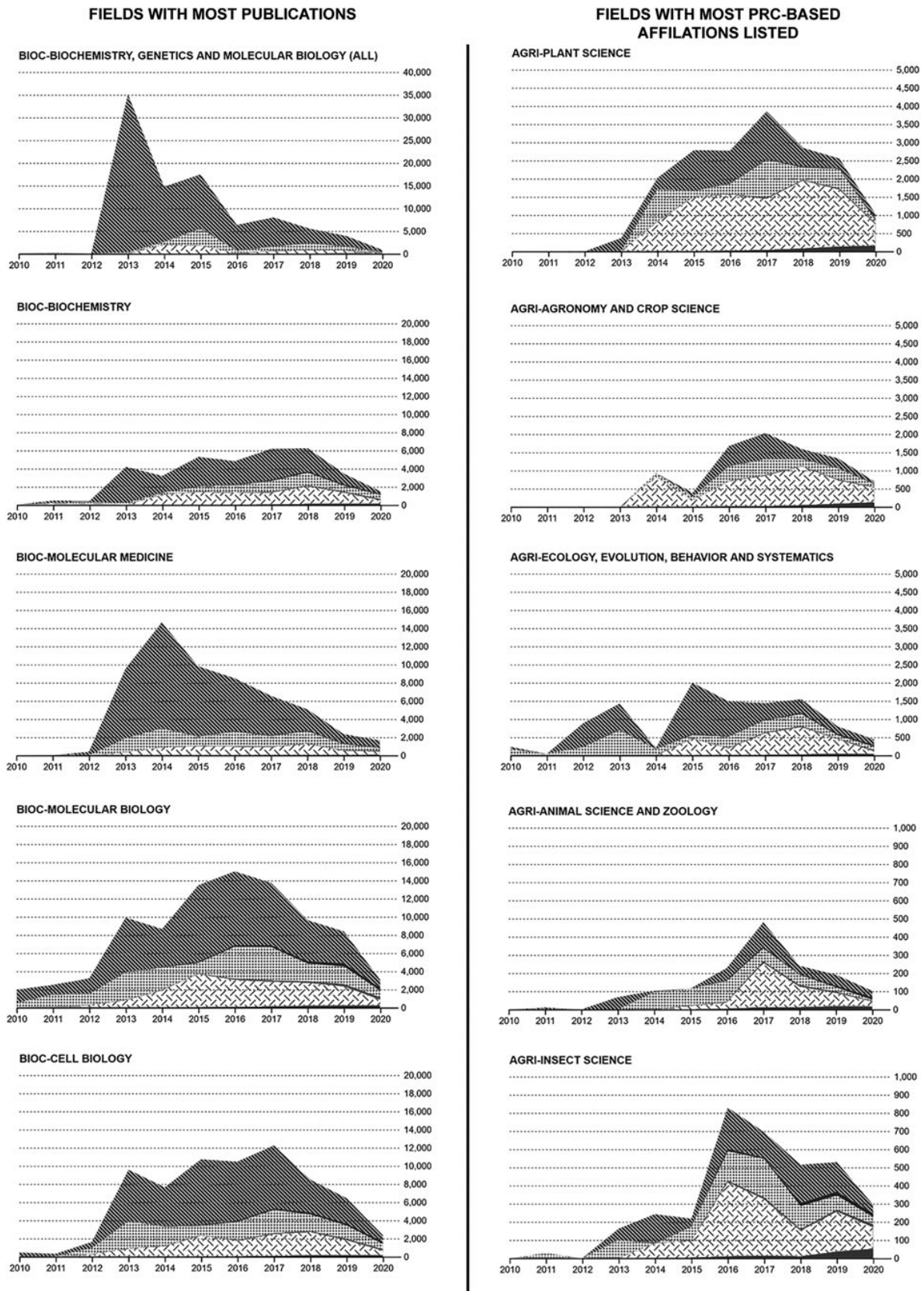


FIG. 11. Stacked area chart of publications (shades) and citations (textures) per area between 2010 and 2020. Strokes represent the US, the woven pattern represents the PRC, and circles represent the EU. Left figures fields with most publications, right figures fields dominated by PRC-based affiliations. Citation data are as of January 2021. All data are computed per year.

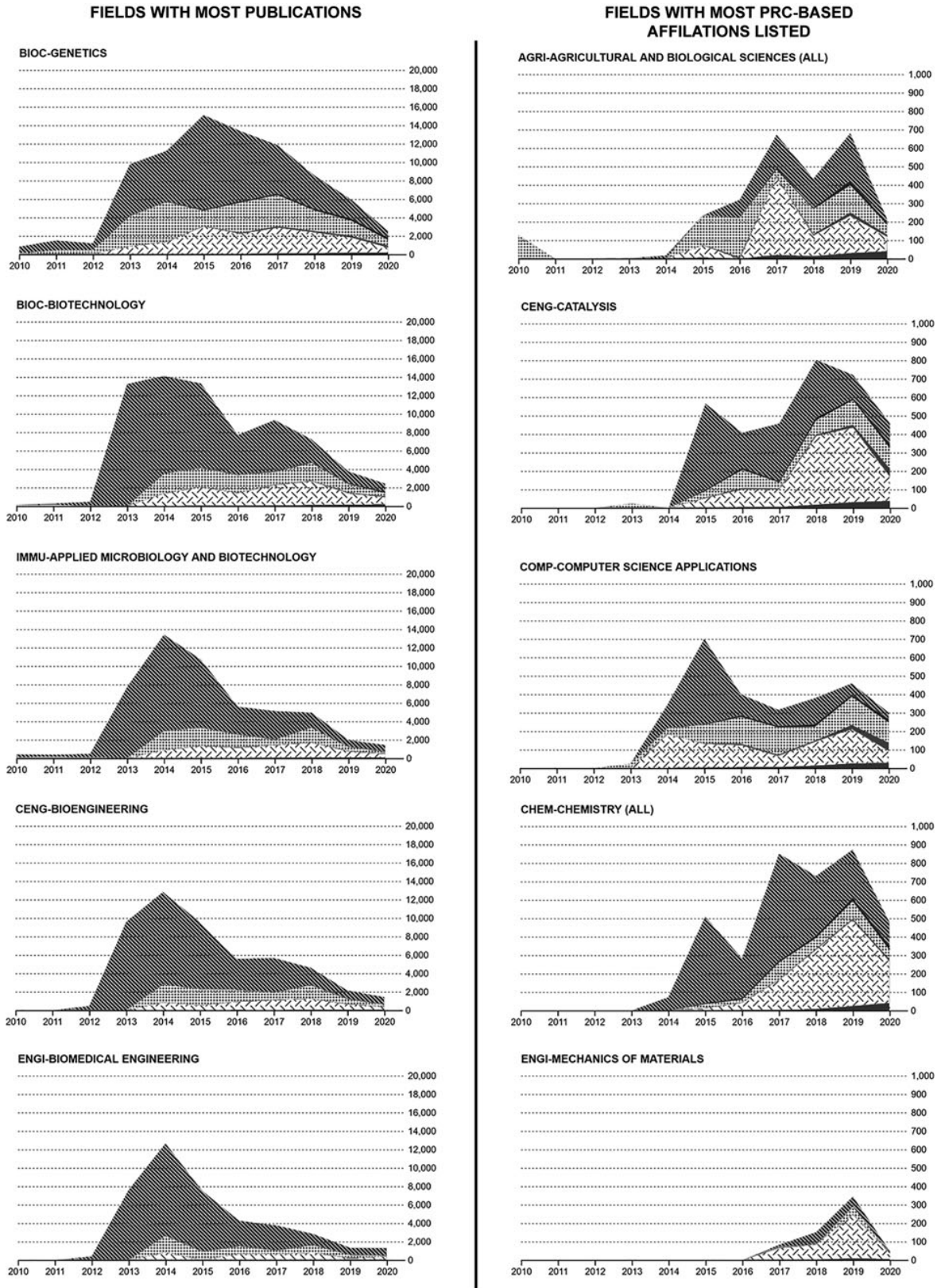


FIG. 11.(Continued).

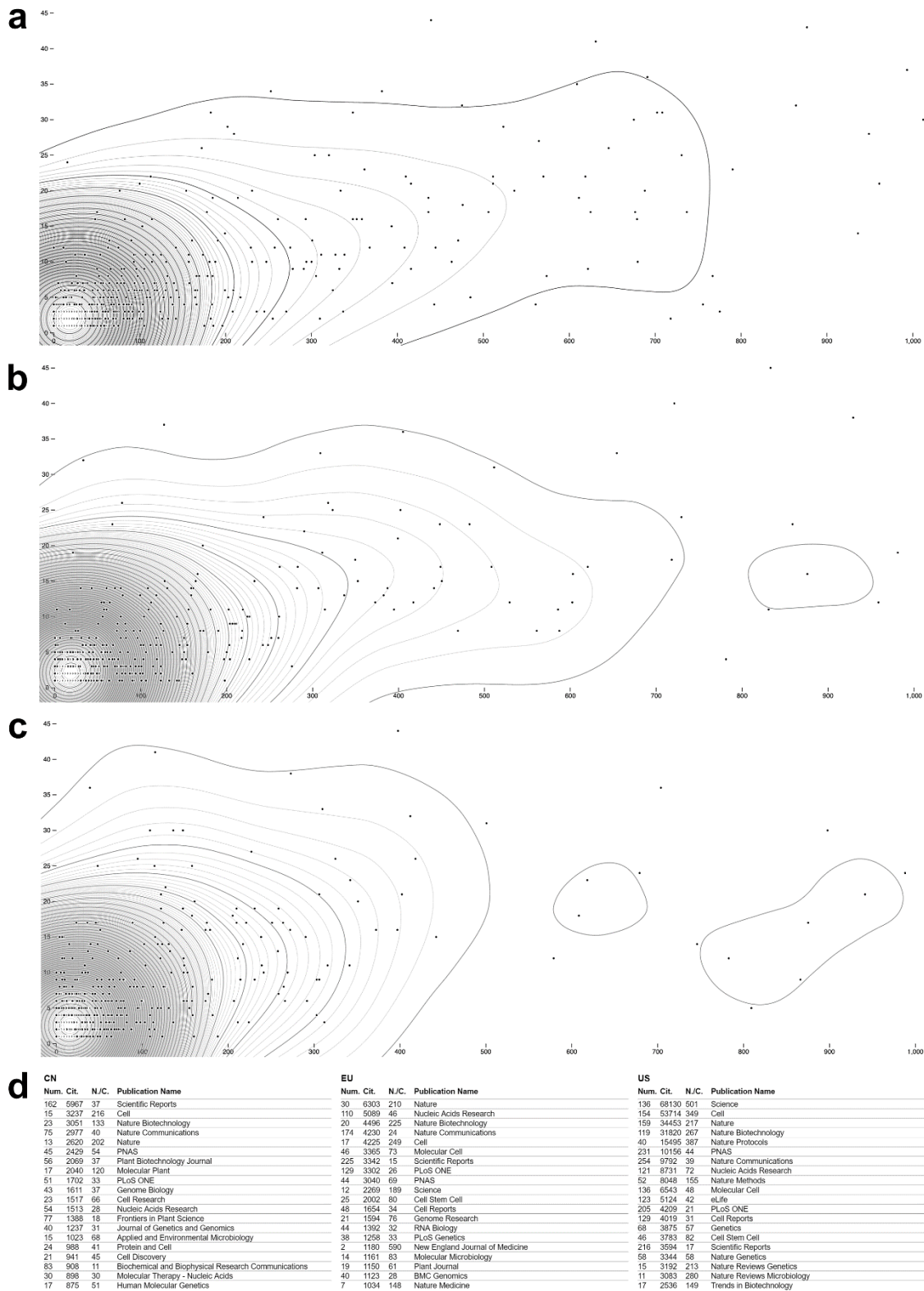


FIG. 12. Contour density by marching squares of journals for articles published by US-based (a), EU-based (b), and PRC-based (c) authors. X-axis figures citations, Y-axis the number of articles. The chosen extents aim to display the repartition of average articles for each area. (d) Shows the performance of top articles for each area, listing the number of articles published in this journal (Num.), the cumulative number of citations of these articles (Cit.), the number of articles divided by the cumulative citations (N./C.), and the publication's name.

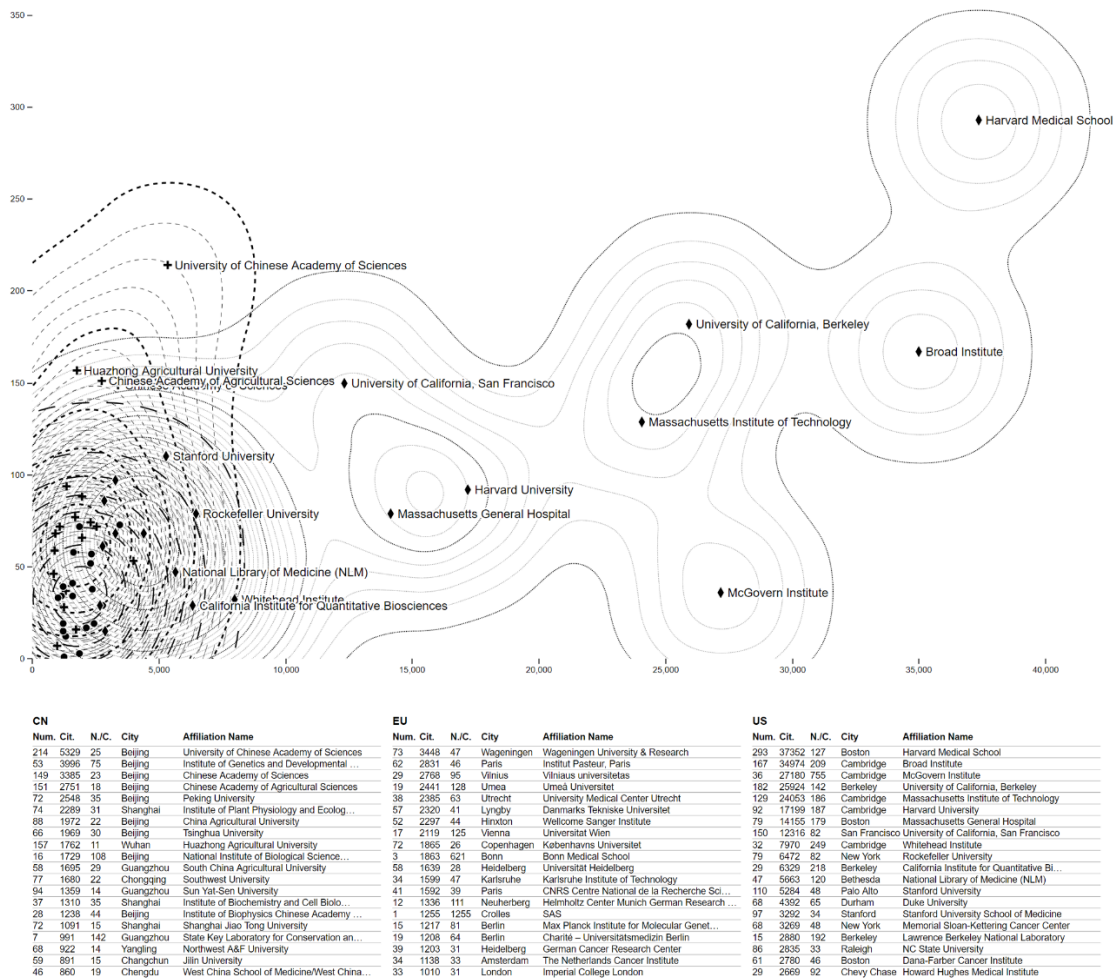


FIG. 13. Contour density by marching squares of US (continuous - diamonds), PRC (short dashes - crosses), and EU (long dashes - circles) most productive affiliations in terms of cumulative citations. X-axis measures the number of citations, Y-axis the number of published articles. Table: tabular representation of the same data, showing the number of articles published by authors listing this affiliation (Num.), the cumulative number of citations of these articles (Cit.), the number of articles divided by the cumulative citations (N./C.), the city in which the affiliation is based in according to Scopus and the affiliation's name.

Discussion

The PRC was the only “developing country” to participate to the International Human Genome Project (HGP). Its contribution amounted to 30 centimorgans, or 1% of the total length. This was considered an impressive feat⁵ for the country at the time. The data presented in this article shows that the scale of PRC’s contribution to life sciences in the decade leading to 2020 has radically changed. It has now replaced the EU as the main challenger to the US, and possibly its main partner, in this area. Does this meteoric rise give substance to the idea that a race for “nuclease supremacy” would currently be under way? Is genome editing considered to be a technological platform so transformative that it could rank as a strategic asset? Do policymakers in the US and in China believe CRISPR to be relevant to the future of sovereign power?

The dynamics of CRISPR’s adoption as a lever for more epistemic progress do show that different countries have invested discretely in its early promises. For some part, these differences can be explained by structural factors. The US hosts world-leading universities in the life sciences and has managed to keep them competing for the top position in genome editing research. At the other end of the spectrum, EU countries have largely failed to challenge the US’s leadership and to ramp up its funding for research programs. Brexit seems poised to weaken scientific Europe’s position further. This sharply contrasts with the PRC’s position, as PRC-based institutions are becoming ever more present in CRISPR-powered research.

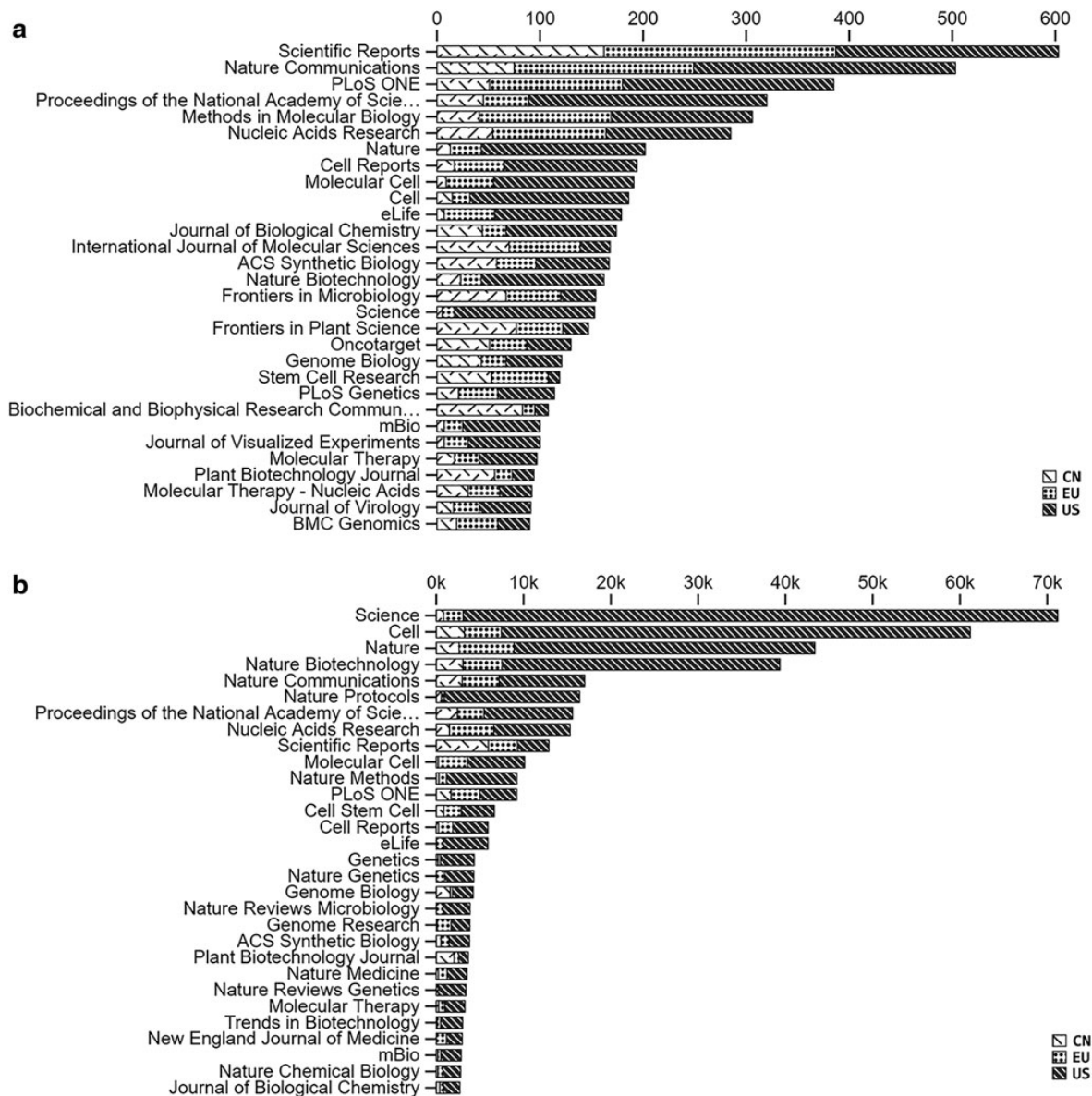


FIG. 14. Stacked bar chart ranking publications that published (a) the most articles from the corpus (b) articles with most cumulative citations in the corpus. Bars are stacked by area of author affiliation.

The PRC's rise in these research fields is not uniform, which shows it was able to capitalize on its preexisting structures to build a comparative advantage in some fields. US-based institutions have built up on their early discoveries to take the lead in terms of number of publications and citations in the life sciences, making it hard for PRC-based labs to challenge their positions. But this advance relies a lot on the past glory of setting the first landmark articles for a new method: publications since 2018 show a much more levelled landscape. In the meantime, some PRC institutions have focused their efforts on Agriculture-related fields, with great success in Plant sciences and Agronomy and Crop sciences. In doing so, they have kept their edge in these fields and made it sharper with CRISPR.

This progress did not happen deus ex machina. The 2000-2010 decade did host the rise of leading PRC-based institutions specialized in DNA sequencing⁶. The results of this scientific and industrial investment continue to bear fruit today⁷. But if the PRC-based laboratories were so prompt in taking full advantage of what genome editing newly allowed, it is also due to the very nature of CRISPR-Cas techniques. The "slower, cheaper, more efficient" slogan is slowly fading into irrelevance, as other engineerable nucleases are less and less resorted to outside of therapeutic contexts. Laboratories now use CRISPR in the same way they would have used PCR in the previous decades⁸. Would not that be the relevant time to invest in searching the next biotechnological revolution?

Table 1. Anonymous written contributions to the « Etats Généraux de la bioéthique » online platform set up by the French government (first half of 2018)

FR	EN
<p>il faut améliorer nos embryons de sorte que les élèves soient tous brillants et deviennent ingénieurs et autres bac+8 _ _ _ La Chine s'y attaque!</p> <p>Non médecin et banal citoyen, j'ai fait séquencer mon génome par une compagnie -hélas non-française- puisque ce n'est pas encore possible en France.</p> <p>Donc, ces données sont stockées quelque part aux États-Unis où en Chine ou ailleurs.</p> <p>Les États-Unis et la Chine ont pris une avance considérable dans ce domaine très important et il est temps pour l'Europe de se mettre au diapason.</p> <p>Les enjeux actuels:sociaux; économiques et politiques..</p> <p>mondiaux aussi: que fera la loi Française face àç la question du clonage (voir Chine et USA)</p> <p>La frontière entre réparer et augmenter est floue. [...] De plus, en admettant que ces technologies soient fortement encadrées en France, elles ne le seront pas ailleurs, aux USA ou en Chine. Qui peut empêcher une personne s'y rendre et modifier son génome ?</p> <p>Le problème c'est que personne ne sait ce que veut dire capacités "normales". C'est subjectif. Ce qui est normal pour quelqu'un ne l'est pas pour un autre. D'autant plus que si on encadre ces technologies en France, elles seront autorisées ailleurs, aux USA ou en Chine. N'importe qui pourra s'y rendre et s'augmenter, puis revenir en France. Aucune loi ne pourra les empêcher.</p> <p>Mais, toute loi française pour interdire les augmentations ne s'appliqueront qu'en France. Quid des autres pays ? Qui pourra empêcher le citoyen lambda français d'aller aux USA ou en Chine, modifier son génome, et revenir en France ?</p>	<p>We must enhance our embryos so that students can all be brilliant, become engineers and highly skilled people _ _ _ China is starting to do it!</p> <p>As a non-MD and average citizen, I had my genome sequenced by a company – Alas a non-French one- since it isn't possible yet in France. So, these data are stashed somewhere in the US or in China or elsewhere.</p> <p>The US and China have taken a considerable advance in this very important domain, and it is time for Europe to catch up.</p> <p>The current stakes: social, economic, political...</p> <p>Global too: what will the French law do with regards to cloning (see China and USA)</p> <p>The frontier between fixing and augmenting is blurry. [...] In addition, even if these technologies are strongly framed in France, they will never be elsewhere, in the US or in China. Who can prevent someone to go there and have their genome modified?</p> <p>The problem is that no one wants "normal" capacities. It's subjective. What is normal for someone isn't for someone else. Plus, if we limit these technologies in France, they will be authorized elsewhere, in the US or in China. Anyone will be able to go there and augment themselves, and then come back to France. No law will be able to prevent that.</p> <p>But, all French laws to prevent augmentations will only be applied in France. What about other countries? Who will prevent the average French citizen to go to the US or to China, modify their genome, and come back to France?</p>

This question, as illustrated by the diversity of reactions⁹ to a recent international report, is proving to be a political minefield. Controversies around governmental responses to the covid-19 pandemic might also have a durable impact on the future of molecular biology. This very specific context might explain why the recently published fourteenth 5-year plan only mentions “genetics” once, as part of a longer list containing other cutting-edge technology such as AI, quantum technologies or hydrogen energy [在类脑智能、量子信息、基因技术、未来网络、深海空天开发、氢能与储能等前沿科技和产业变革领域，组织实施未来产业孵化与加速计划，谋划布局一批未来产业].

Outside of China, the idea of a new Sputnik race lives on. The fact that the DARPA has funded genome-editing related projects¹⁰ can be interpreted as indicating that some policymakers in the US consider the country to be in a scientific competition deserving State intervention. But a race with whom? Science Magazine reported in 2020¹¹ that out of the 189 scientists investigated by the NIH for failure to “disclose financial ties to foreign governments”, 93% were targeted for alleged ties with the PRC. Fifty-four scientists have resigned or been fired. Such uncontrolled technological transfers to the PRC are apparently considered to be a threat by US authorities.

The PRC's new status has also already been, to some extent, as acknowledged by citizens interested in these issues. The prospective uses of CRISPR-Cas techniques were listed as one of the topics for debate during the French government's public consultation over the revamping of its bioethics framework, which happened in the first half of 2018. The presented array

of comments (Table 1) highlights how the US and the PRC have become, in the view of the French citizens who chose to take part in this exercise, equivalent in terms of both perceived technical capacity and perceived lack of appropriate regulatory frameworks. The underlying idea, that the two areas are developing these techniques at breakneck speed to reap its (hypothetical) benefits, indicates that this race also exists in the minds of those who consider themselves to be spectators to it.

Conclusion

This quantitative study finds that the PRC-based institutions appear to be both the main rivals and best partners of US-based institutions in genome editing research, and that PRC research is particularly salient in Agriculture-related fields. The swiftness of the PRC's research focused progress and the reaction of US authorities both substantiate the idea of a “technological race” going on over genome editing between the two countries.

These findings call for a qualitative comparison of the policies implemented in the UE, the US and China to promote genome editing research. Such research could elucidate what “genome editing” is to policymakers, and how government policies have been discussed, designed and used to fuel this rapid progress.

References

1. Collins FS, Doudna JA, Lander ES, et al. Human molecular genetics and genomics—Important advances and exciting possibilities. *N Engl J Med* 2021;384:1–4.
2. Cyranoski D. CRISPR gene-editing tested in a person for the first time. *Nature News* 2016;539:479.
3. Gao C. Genome engineering for crop improvement and future agriculture. *Cell* 2021;184:1621–1635. DOI: 10.1016/j.cell.2021.01.005.
4. Martin-Laffon J, Kuntz M, Ricroch AE. Worldwide CRISPR patent landscape shows strong geographical biases. *Nat Biotechnol* 2019;37:613–620.
5. Wang X, Xia Z, Chen C, et al. The international Human Genome Project (HGP) and China's contribution. *Protein Cell* 2018;9:317–321.
6. Wu J, Xiao J, Zhang R, et al. DNA sequencing leads to genomics progress in China. *Sci China Life Sci* 2011;54:290–292.
7. Dun Y. Chinese DNA sequencing company raises a record \$1 billion. *Nikkei Asia*. Available online at: <https://asia.nikkei.com/Business/Startups/Chinese-DNA-sequencing-company-raises-a-record-1-billion> (last accessed February 22, 2021).
8. Rabinow P. *Making PCR: A Story of Biotechnology*. Chicago: University of Chicago Press. 2011.
9. Angrist M, Barrangou R, Baylis F, et al. Reactions to the National Academies/Royal Society Report on Heritable Human Genome Editing. *CRISPR J* 2020;3:332–349.
10. Safe Genes. Available online at: <https://www.darpa.mil/program/safe-genes> ((last accessed March 22, 2021).
11. Mervis Jun. 12 J, 2020, Pm 6:00. Fifty-four scientists have lost their jobs as a result of NIH probe into foreign ties. *Science AAAS*. 2020. Available online at: <https://www.sciencemag.org/news/2020/06/fifty-four-scientists-have-lost-their-jobs-result-nih-probe-foreign-ties> (last accessed March 22, 2021).

Acknowledgments

Many thanks to the reviewers for their perspectives which helped me build a much deeper article than I had initially envisioned. Many thanks also to Sciences Po's METAT team for their advice and technical help in the early days of PANDORÆ and to Lou Safra, Thierry Vedel, and Jean Chiche for their insights on early versions of this work.

Funding Information

This research is funded by the Agence Nationale de la Recherche, grant ComIngGen - 18-CE38-0007-01.

PANDORÆ and the scripts used for this paper are open, free and available on GitHub.