# Convention on Biological Diversity – Call for information:

# The use of digital sequence information on genetic resources

## Response by the Wellcome Trust and Wellcome Trust Sanger Institute

**8 September 2017**

### KEY MESSAGES

- We strongly disagree with the proposal to include digital sequence information (DSI) in the scope of the Convention on Biological Diversity (CBD) and the Nagoya Protocol. We agree fully that countries should share equitably in the benefits of research and development activities to which they contribute and which utilise sovereign genetic resources, but consider that the inclusion of DSI would fail to achieve this goal, and do far more harm than good.

- In particular, it could seriously threaten international research, development and surveillance activities relating to global health threats – which depend upon the rapid and open sharing of pathogen sequence information from countries around the world.

- Science is a global endeavour and our ability to tackle complex global challenges depends on international collaboration and the availability of research findings and data. At present, researchers around the world are sharing and utilising DSI on animals, plants and pathogens via public databases to advance research and its application for the global good (see annex I). It is absolutely vital that free and unrestricted sharing of DSI via these resources is able to continue – the imposition of any terms or conditions would seriously undermine their value and sustainability.

- In our response, we present a series of case studies to illustrate some of the many ways that DSI is being utilised in international research endeavours – which could potentially be hampered by the current proposals.

### General points

1. Wellcome is a global charitable foundation dedicated to improving health for everyone. We fund research related to health, support activities in over 70 countries and intend to spend £5 billion over the next 5 years to further our mission. The Wellcome Trust Sanger Institute uses genome sequences to advance the understanding of the biology of humans and pathogens to improve human health. We use science at scale to tackle the most challenging global health research questions.

2. Our response to this call for information focuses on how DSI from genetic resources is used in research. In this submission, we illustrate the benefits of the unrestricted use of

DSI through four case studies on projects related to epidemic response, tracking drug resistance and conservation.

3. We disagree with the proposal to include of DSI in the scope of the CBD and the Nagoya Protocol. Science is increasingly collaborative and solving complex problems like epidemics and climate change will depend on international partnerships. Many of these rely on the ability to share DSI, regardless of what country a researcher is based in. There is a risk that the Nagoya Protocol could be implemented by countries in a way that poses a barrier to these collaborations, and this would be exacerbated if DSI were included in its scope.

4. Our primary concern is on genetic resources relating to human, animal and plant pathogens, where we believe the extension of the Convention to include DSI could greatly exacerbate its potential to threaten international research, development and surveillance activities relating to existing and emerging global health threats.

5. The continuous sharing of pathogen samples and DSI and the availability of pathogen DSI in public databases is critical for ensuring timely, accurate and accessible sharing of data for epidemic risk assessment and rapid response (see case studies 1 and 2). Timely sharing is essential for generating actionable public health information about how to prevent and respond to outbreaks.

6. The availability of DSI is also essential for developing the diagnostics, vaccines and pharmaceuticals needed to detect, prevent and treat illness. For example, the development of annual seasonal flu vaccines depends on the availability of global epidemiological and virus sequence data.

7. The global influenza surveillance and response system (GISRS) is a network of public health laboratories coordinated by WHO that collect human and animal flu virus samples and share the samples and related DSI for analysis. This happens on an annual cycle and rapid sharing of both the viruses and DSI is crucial as the target strains for seasonal vaccines must be decided 5-6 months ahead of the flu season for the vaccine to be developed in time. This is a very time sensitive process which could potentially be prohibitively delayed if access and use had to be agreed on a bilateral basis.

8. The Berlin Declaration of the G20 Health Ministers recently recognised the importance of GISRS with regard to sample and data sharing[1]. It is therefore essential that this is not inadvertently threatened through including DSI in the scope of the CBD.

9. The field of genomics in particular has thrived on the basis of free and unrestricted sharing of DSI from genetic resources. For over 30 years, the European Nucleotide Archive (ENA) at EMBL-EBI has maintained and provided researchers access to genetic sequence information related to animals, humans and plants. From September 2014 to August 2017, users downloaded more than 3.5 million sequences. As well as providing

[1] http://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/G/G20-Gesundheitsministertreffen/G20_Health_Ministers_Declaration_engl.pdf

the sequence aspect of the scientific record, ENA provides a platform for the sharing and dissemination of early research data[2].

10. Any restrictions on sharing and accessing DSI from genetic resources in databases like ENA could create a major barrier to research and innovation that utilises this information. For example, this would impact efforts to collate DSI from around the world to track the emergence of drug resistance (see case study 3). The complexity of interpreting and complying with varying domestic access and benefit sharing legislation for DSI may threaten the sustainability of ENA and other databases if this dissuades or prevents researchers from sharing and using sequences.

11. Analysing DSI has a variety of applications as broad as the diversity in the natural world. The availability of DSI about all kinds of living things, from parasites and mosquitoes to bats and gorillas, in databases means they can be used by the international scientific community, sometimes in unexpected ways (see case study 4).

**Case study 1 – Global spread of dysentery – Dr Nicholas Thomson, Wellcome Trust Sanger Institute**

Dysentery kills around 525,000 children under 5 each year, around the world[3]. The most common cause of dysentery is *Shigella* bacteria, spread by contaminated food or water. The result is extreme diarrhoea which if untreated is often fatal. There are four species of *Shigella*, and most cases are caused by *S. flexneri*. Little was known about the genetic makeup of *S. flexneri* strains, which made tracking the spread of infections challenging and has significantly impeded public health efforts to control outbreaks.

351 strains of *S. flexneri*, from accredited public health laboratories in South Africa, Bangladesh, France, Vietnam and Korea were sequenced at the Sanger Institute, with the samples themselves originating from outbreaks in Bangladesh, Pakistan, Algeria, Egypt, Haiti, Senegal, Guinea, Burkina Faso, Chad, Cameroon, Ivory Coast, Korea, Madagascar, Haiti and the Dominican Republic.

By sequencing these many different strains, the researchers were able to determine that unique *S. flexneri* strains occupy distinct geographic areas[4], do not spread but are able to persist over a long time. This indicates the bacteria does not travel over wide areas and is not the underlying cause of dysenteric pandemics. These findings strongly reinforce the importance of sanitation and the provision of clean drinking water, and inform public health strategies for tackling persistent outbreaks. The potential complexity of conducting multiple bilateral agreements for access and use of DSI in this kind of research could cause delays and potentially slow down public health efforts the research seeks to inform.

**Case study 2 – Putting genomic surveillance at the heart of viral epidemic response – Professor Andrew Rambaut, University of Edinburgh**

[2] Toribio, A. L., Alako, B., Amid, C., Cerdeño-Tarrága, A., Clarke, L., Cleland, I., Cochrane, G. (2017). European Nucleotide Archive in 2016. *Nucleic Acids Research*, *45*, D32–D36. http://doi.org/10.1093/nar/gkw1106
[3] http://www.who.int/mediacentre/factsheets/fs330/en/
[4] Connor , T.R., *et al 2015* Species-wide whole genome sequencing reveals historical global spread and recent local persistence in Shigella flexneri *eLIFE* 4 e07335 doi:10.7554/eLife.07335

In recent outbreaks of disease such as Ebola and Middle East respiratory syndrome coronavirus (MERS-CoV), sequencing samples of the viral pathogen revealed critical insights into the origins of infection and the evolution and transmission of the disease.

For example, in the 2013-2016 West African Ebola outbreak, researchers collected and sequenced over 1600 Ebola virus samples to understand how the virus was evolving and the factors responsible for its transmission[5]. Analysis of the virus' DSI showed how different strains crossed borders and spread within countries. Access to a subset of this data during the outbreak led to border closures to limit its spread[6] and suggests the potential value of sequencing to control future outbreaks, if it can be shared and analysed quickly enough to inform response efforts. However, the impact of these new sequencing technologies has yet to be realised, in part due to the complexities and time taken to ship samples for sequencing and the resulting delays in the production of DSI.

Professor Andrew Rambaut is leading a Wellcome Trust funded project to develop a field-deployable virus sequencing system and accompanying information sharing platforms so that real-time viral genome sequencing can have a greater impact on the public health response for the next outbreak. Including DSI in the scope of the CBD and the Nagoya protocol could threaten the impact that this and other technologies may have on our ability to respond to the next epidemic.

**Case study 3 – Tracking resistance to artemisinin collaboration (TRAC) studies**

Each year, malaria kills about 500,000 and causes debilitating illness in over 200 million people[7]. The most effective treatment is artemisinin combination therapy, but resistance to artemisinin, which leads to treatment failure, has started to spread throughout Southeast Asia. Tracking and understanding the genetic lineage of resistance is vital for improving malaria containment and elimination, as well as patient treatment.

TRAC I and II are international collaborations, coordinated by the Mahidol-Oxford Tropical Medicine Research Unit, to gather, share and analyse 'real-time' genetic information on malaria drug resistance. The TRAC I study[8] mapped the extent and severity of artemisinin resistance in Southeast Asia and TRAC II will continue to map resistance and in addition assess the safety and efficacy of new artemisinin combination treatments. It is an international effort, as TRAC II has over 60 investigators from more than a dozen countries.

The researchers post malaria parasite DSI on TRAC and other databases such as the WorldWide Antimalarial Resistance Network (WWARN), so that other researchers can build on their findings. Rapid sharing is critical to help track and anticipate the geographic routes of drug resistance and inform national and regional patient treatment strategies to stop resistance to artemisinin spreading to other malaria endemic regions.

**Case study 4 - Conservation of endangered gorilla populations – Chris Tyler Smith, Wellcome Sanger Institute**

Mount Tschiaberimu in the Democratic Republic of Congo (DRC) is home to a highly endangered population of gorillas. With a population of 1 female, 4 males and one baby of

[5] Holmes, E.C., Dudas, G., Rambaut, A., Andersen, K.G., 2016 The evolution of the Ebola virus: Insights from the 2013-2016 epidemic *Nature 538*(193-200). doi:10.1038/nature19790
[6] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4960802/pdf/13059_2016_Article_1019.pdf
[7] http://www.who.int/malaria/publications/world-malaria-report-2016/en/
[8] Ashley, L. *et al.* 2014 Spread of Artemisinin Resistance in Plasmodium falciparum Malaria *NEJM* 371:411-423 doi: 10.1056/NEJMoa1314981

unknown sex, the colony is no longer viable and needs new gorillas to provide enough genetic variability for the colony to continue.

Living on Mount Tschiaberimu, they were thought to be mountain gorillas, but some unusual characteristics put this in question. To confirm their species, researchers at the Sanger Institute working with researchers from the US and conservationists from the DRC, took a sample from a gorilla, which was sequenced at the Sanger Institute and compared to openly available DSI from three gorilla species. The analysis showed that it was not a mountain gorilla but an eastern lowland subspecies, which was necessary information to develop a conservation strategy. The DSI of the gorilla was also deposited in a public database.

Identifying the gorilla's species was only possible because the researchers were able to access DSI of the other gorillas. Once again, our concern is that a potential knock on effect of the inclusion of DSI in the scope of the Nagoya protocol could be that complying with varying domestic access and benefit sharing legislation could cause prohibitive delays on this kind of work progressing.

**Annex I**- Examples of widely used databases of DSI from genetic resources

- European Nucleotide Archive
- GenBank
- Banana genome hub
- PATRIC
- iMicrobe
- Flybase
- WormBase
- Peanutbase
- Expression Atlas
- Rat Genome Database
- UCSC Malaria Geome Browser