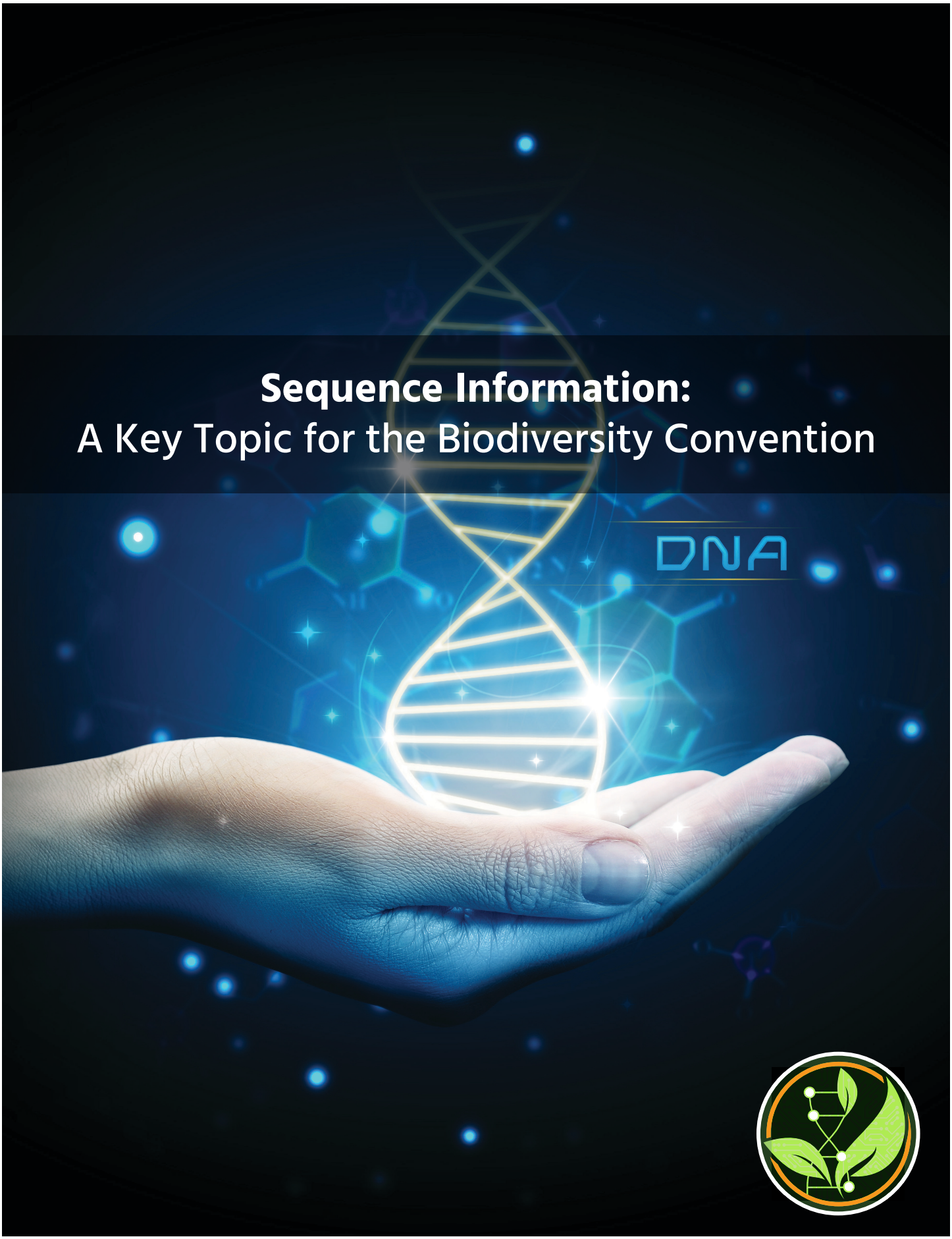
Sequence Information:
A Key Topic for the Biodiversity Convention



Sequence Information: A Key Topic for the Biodiversity Convention

2018

About African Centre for Biodiversity: The African Centre for Biodiversity (ACB) is committed to dismantling inequalities in the food and agriculture systems in Africa and our belief in people's right to healthy and culturally appropriate food, produced through ecologically sound and sustainable methods, and their right to define their own food and agricultural systems. (Please note: On 7 April 2015 the African Centre for Biosafety officially changed its name to the African Centre for Biodiversity. This name change was agreed on through consultation within the ACB to reflect the expanded scope of our work over the past few years. All ACB publications prior to this date will remain under our old name of African Centre for Biosafety and should continue to be referenced as such.)

About ETC Group: ETC Group works to address the socioeconomic and ecological issues surrounding new technologies that could have an impact on the world's marginalized people. It operates at the global political level and works closely with partner civil society organizations (CSOs) and social movements, especially in Africa, Asia and Latin America. ETC Group is headquartered in Val David, Canada, and Davao, Philippines, with offices in Mexico City and Guelph, Canada.

About Third World Network: Third World Network (www.twn.my) is an independent non-profit international network of organizations and individuals involved in issues relating to sustainable development, the South and North- South relations. TWN's goal is to deepen the understanding of the development dilemmas and challenges facing developing countries and to contribute to policy changes in pursuit of just, equitable and ecologically sustainable development. TWN's International Secretariat is in Penang (Malaysia) with offices in Kuala Lumpur (Malaysia) and Geneva (Switzerland). There are researchers based in Beijing, Delhi, Jakarta, Manila and New York. The Latin America Regional Secretariat is located in Montevideo (Uruguay) and the African Regional Secretariat is in Accra (Ghana).

This document was produced under the Building International Capacity in Synthetic Biology Assessment and Governance (BICSBAG) Project. The BICSBAG partners gratefully acknowledge the financial support from SwedBio at the Stockholm Resilience Centre, Frontier Coop Foundation and CS Fund in the production of these materials.

Visit www.synbiogovernance.org for more information and translations.

Layout: Cheri Johnson

Building International Capacity in Synthetic Biology Assessment and Governance

Sequence Information: A Key Topic for the Biodiversity Convention

Introduction

Sequence information will be on the agenda at the 14th Conference of the Parties (COP) to the Convention on Biological Diversity (CBD) in November 2018 in Egypt, and it is a topic that governments cannot afford to ignore. New applications of sequence information are transforming how genetic resources are used, and have major long term implications for the CBD, particularly for the objective of fair and equitable benefit sharing.

"Free" sequence information of a wide variety of biodiversity is increasingly becoming available, and is reducing the need for physical access to plants, microbes, animals and other living things in an growing number of research and commercial applications. As the drivers of this phenomenon, which include cheap digital sequencing, gene editing, and other biotechnological and synthetic biology approaches, continue to develop, the trend will accelerate. More and more sequences will be generated "in the field" and shared electronically, potentially without proper prior informed consent (PIC) and mutually agreed terms (MAT).

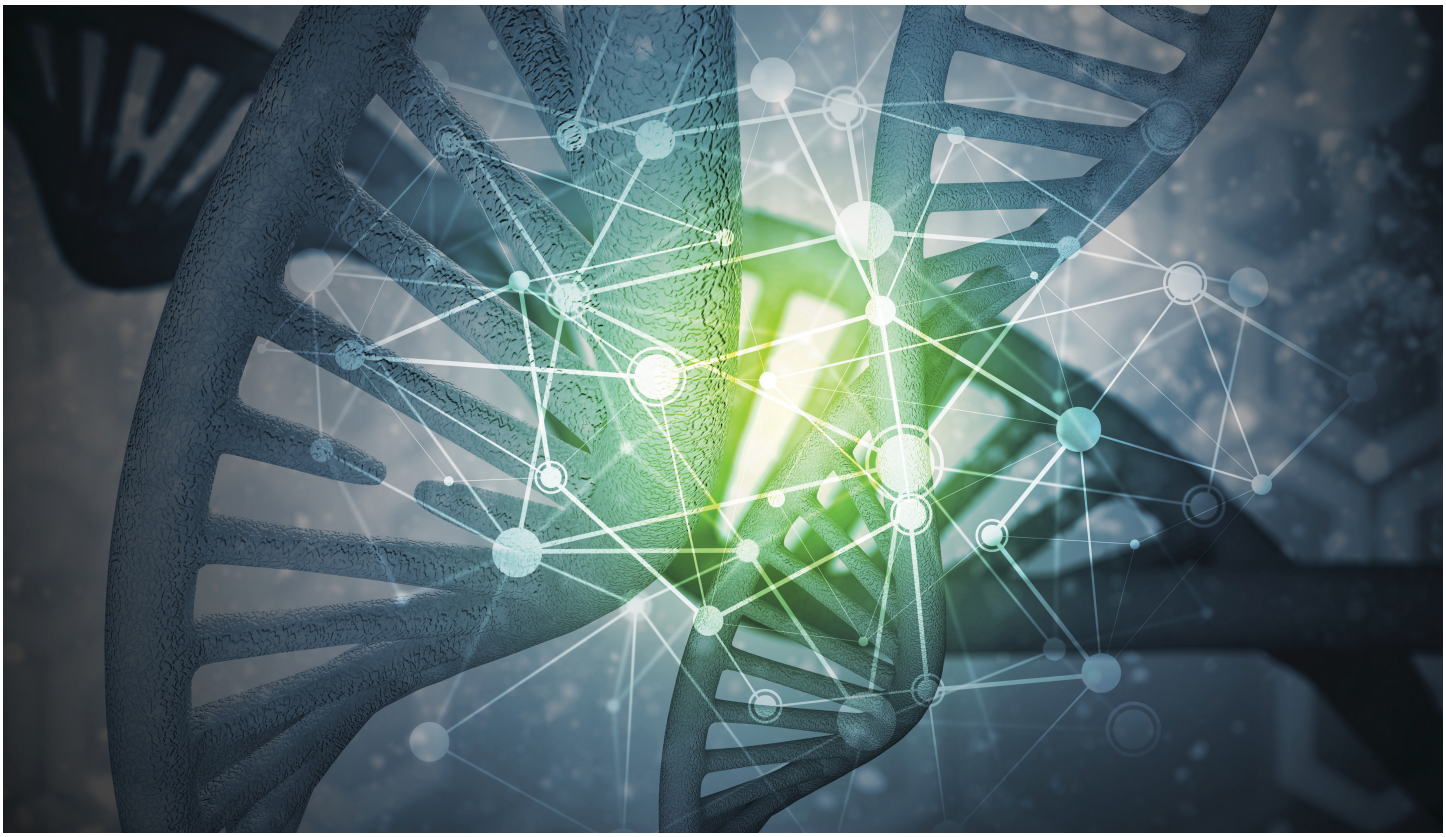
A situation is quickly arising in which biopiracy is allowed to occur because legal frameworks have not caught up with technical realities.

These rapidly advancing technologies are upending traditional approaches to access and benefit sharing under the CBD and extend across the realm of biodiversity from the smallest organisms, such as viral pathogens, to the large and complex genomes of many crops. A situation is quickly arising in which biopiracy is allowed to occur because legal frameworks have not caught up with technical realities.

Sharing of sequence information is now central to many aspects of research, but so long as that information is generated and shared without applying benefit sharing obligations, the developing country governments, farmers, and indigenous peoples that created and nurtured that diversity will lose out. National genetic resources and Indigenous Peoples' plants will be privately "mined" for profitable sequences with little or no recompense.

This impending change promises to be so stark that failing to address sequence information could undermine the entire Convention by upsetting application of its third objective, the fair and equitable sharing of the benefits arising out of the utilization of genetic resources, and the substantial efforts put forth by the CBD to date to implement it, not least the Nagoya Protocol on Access and Benefit Sharing.

Governments should recognize that it is unrealistic to expect results from the CBD's discussion on digital sequence information (DSI) if they do not make it a high priority in their preparation for the



COP. That is because the status quo is highly beneficial to the interests of user countries and the biotechnology industry, and the North will not move on its own to effectively address this threat to the CBD, as it would rather continue to benefit from free access to genetic resources in the form of a massive and growing cloud of DSI.

To stop the unfolding free-for-all on sequence information, Parties to the CBD must make the Convention current by finding a way to apply benefit sharing rules to access and use of sequences.

Otherwise, as *ex-situ* collections move to sequence their collections, researchers deploy small (even hand-held) sequencers, and online databases continue to mass-publish sequences without regard for benefit sharing and without placing any restrictions on patent claims, Parties will find that the access and benefit sharing (ABS) foundation on which the CBD was built has been washed away by a sea of big data.

Rapid advances in sequencing and synthesis

Faster synthesis and transfer: The easiest place to see the transformation currently underway is with the smallest organisms. In the health sector, cheap, deep and fast gene sequencing means that the full sequence of influenza viruses can be determined within hours of their isolation. And if that sequence is uploaded onto an Internet database or sent by email, gene synthesis technologies make it possible to recreate a living virus in less than three days, at an appropriately equipped laboratory anywhere in the world.

Thus, in some cases, entire small organisms can now be moved across the planet more quickly when transmitted on the Internet as sequence information than when physical samples are carried by a courier such as DHL or Federal Express.

Longer genomes: In addition to speed increases, the complexity of gene constructs (measured by the size of their genome) that can be synthesized from a sequence is increasing. Poliovirus, the first virus to be wholly synthesized in the lab (in 2002),

is about 7,500 nucleotides long. In November 2016, an American scientific team announced whole synthesis of adenovirus, with a genome of 34,000 nucleotides, four and a half times that of poliovirus.

The Synthetic Yeast Genome Project, an international collaboration of several labs, plans to soon have synthesized the 16 chromosomes of *Saccharomyces cerevisiae* – a total of 12 million base pairs of DNA comprising an entire eukaryotic genome.

From sequences to gene editing: Of course, it is presently not possible to synthesize more complex organisms from scratch, but that is not necessary in order for sequence information to transform ABS. By combining sequence data with synthetic biology gene-editing technologies such as CRISPR, genetic diversity from one place can be introduced into organisms in another without physical access taking place, and without a material transfer agreement (MTA).

Provider countries may thus be unwittingly allowing access to their genetic resources when sequence information of their biodiversity is placed online without adequate controls.

For example, a company might find the sequence of a therapeutic or nutritional compound in a database, and insert that sequence into a laboratory strain used in cell culture. The company might thereby produce a valuable nutrient or plant-based medicine in cell culture vats, all without ever having physically accessed the plant source or executing an MTA.

In another example, corporate crop breeders interested in making tomatoes more tolerant of dry conditions might turn their attention to the gene sequences of tomato plants from the desert coasts of Ecuador, Peru, and Chile. With enough sequence and characterization data, they might identify mutations that make tomatoes from those places more drought tolerant. Patents may claim these mutations, and gene-editing techniques can be used to introduce them into commercial cultivars for sale in North America or Europe, again potentially without any MTA or ABS arrangement.

Such medicines or climate change adapted plant varieties are merely semi-hypothetical examples; the number of other traits in other biodiversity that might similarly be accessed through data, and not physical transfer of materials, is practically infinite. Provider countries may thus be unwittingly allowing access to their genetic resources when sequence information of their biodiversity is placed online without adequate controls.

Digital sequences in the CBD and other UN fora

Managing sequence information is an issue of relevance beyond the CBD. It reaches across biology to include crops and other genetic resources addressed by the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) and the Commission on Genetic Resources for Food and Agriculture (CGRFA) of the Food and Agriculture Organization (FAO) and, in some cases, the World Health Organization (WHO), particularly (but not exclusively) its Pandemic Influenza Preparedness Framework (PIP Framework).

While the implications of sequence information have been anticipated for several years, including by the ITPGRFA, which has discussed sequences under the term "dematerialization," the CBD's deliberations on digital sequences are particularly important because of the Convention's broad scope and membership, and the importance of keeping FAO's approach to sequence information consistent with that of the Convention.

The origin of the present CBD discussion is found in the 2015 report of the Ad Hoc Technical Expert Group (AHTEG) on Synthetic Biology, which identified an important potential impact of sequence information on the Convention as "access without benefit-sharing."[1] The CBD Conference of the Parties in Mexico in December 2016 then adopted a decision on sequence information that sets in motion a plan intended to lead to a substantive decision at their next meeting in Egypt in 2018.[2]

Many developing countries at COP 13 took the position that the CBD should adopt a decision clarifying that sequence information should be treated equivalently to physical samples for the purposes of benefit-sharing.

The World Health Organization, which is considering sequence information in the context of disease research (access to and use of pathogen sequences, particularly influenza virus), has engaged in some practical – but as yet unfinished – consideration of methods to manage access to and use of sequences inside a multilateral access and benefit sharing arrangement, specifically the WHO Pandemic Influenza Preparedness Framework (PIP Framework). The question of potentially pandemic influenza sequences is expected to be discussed at the 71st World Health Assembly in May 2018.[3]

WHO has consulted with databases that host influenza sequence data and is considering "front end" approaches involving database user agreements, data hosting approaches including sequence tagging, and more difficult and complex "back end" monitoring of the appearance of multilateral system (for influenza) sequences in intellectual property claims.

Many developing countries at COP 13 took the position that the CBD should adopt a decision clarifying that sequence information should be treated equivalently to physical samples for the purposes of benefit-sharing.

Terms related to sequence information

The several international processes that are considering sequence information are using a variety of overlapping terms. Recognizing a need for consistency, the CBD has decided to review terminology before its next meeting, and it may be hoped that the terms it identifies for use will also be employed elsewhere to aid understanding and consistency. For now, various terms are used:

*"**Dematerialization**"* has been used for several years by the ITPGRFA. It refers to genetic resources and the fact that sequence information can, for some Treaty purposes, particularly access, supplant the need for physical specimens. Thus, genetic resources are being moved and used without movement of the origin material (germplasm). At recent meetings, the ITPGRFA and CGRFA have also begun to use the CBD-origin term digital sequence information (DSI, see below).

"***Genetic sequence data,"*** or GSD, is a term used by the WHO PIP Framework. Discussion of GSD under the Framework includes the RNA sequence of the influenza virus and that sequence in alternative forms, mainly modifications pertinent to research, diagnosis, and therapeutics, such as complementary DNA (cDNAs) and insertions or deletions with effects on virulence, growth in cell culture, etc.

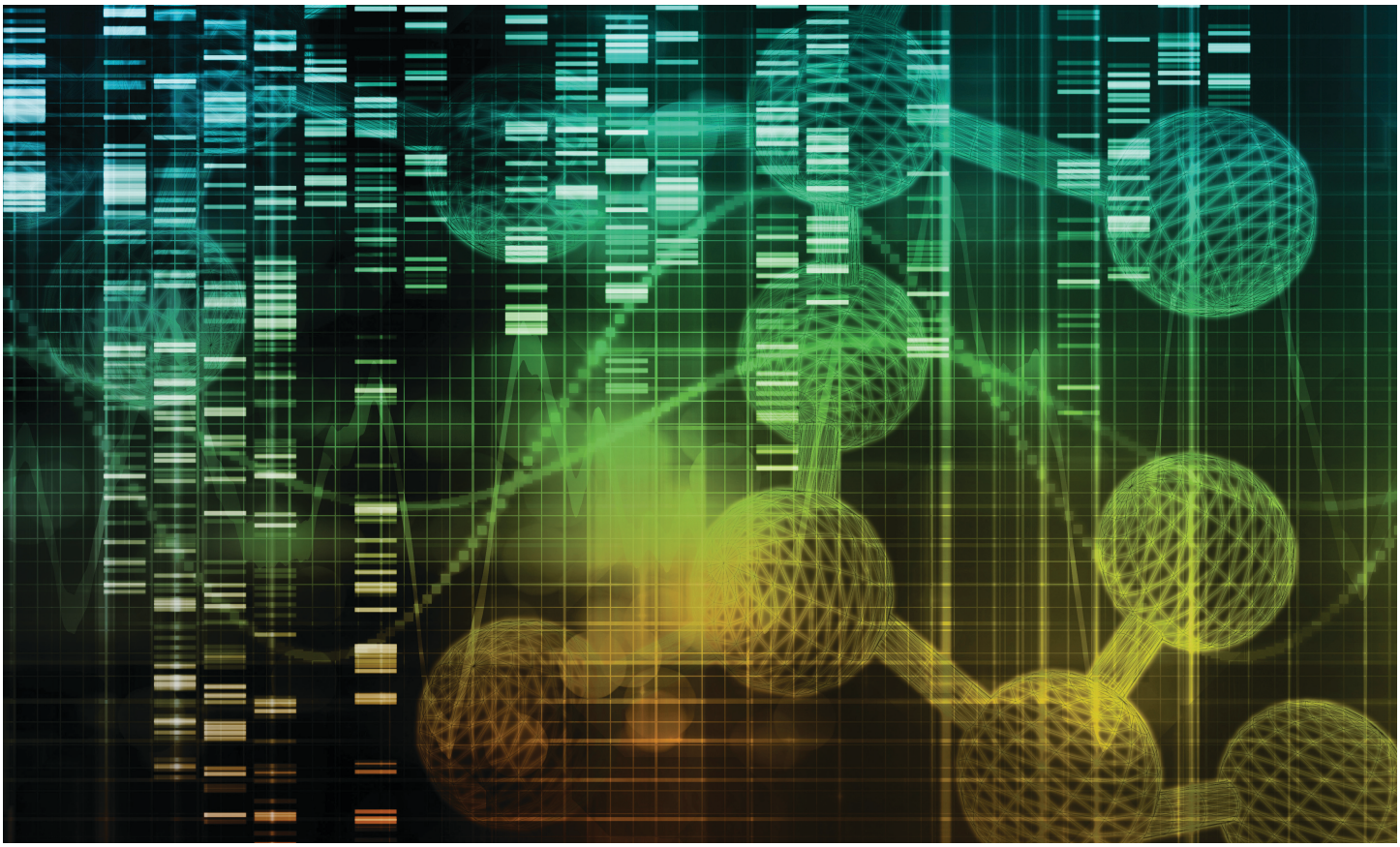
"***Digital Sequence Information,***" or DSI, is a term originating in the CBD AHTEG whose scope has not been defined but which many consider to include DNA, RNA, and amino acid / protein sequences in their various forms, as well as epigenetic and other characterization information of genetic resources. It has been pointed out that the word "digital" may be eliminated from the term since it could be limiting for reasons including that future information (computer) systems may not be "digital," and that sequence information that is not presently stored "digitally" should also be part of the discussion.

Key concepts to enforce in rules and policies for sequence information

Sequence information must be considered the equivalent of physical samples (e.g. seeds)

Accessing sequence information increasingly satisfies many of the same purposes previously served by accessing physical material, including use in the creation of new commercial products that may be placed under patent and other intellectual property rights claims. Since sequences are used this way, and will increasingly be used this way henceforth, the ABS rules that apply to physical material should also apply to sequences.

ABS agreements, including MTAs, must be updated to cover sequence data



Most ABS agreements and laws are grounded in the assumption that material physically changes hands. As presently drafted, these agreements may not be applicable to sequence information. An ABS agreement that does not address sequences may, by permitting physical access to materials, allow the recipient to generate sequence information and benefit from that information outside of the benefit sharing framework. Thus, if sequences are not treated as the equivalent of physical material in ABS agreements, biopiracy will be facilitated, and ways through which users can escape benefit sharing obligations will continue proliferating.

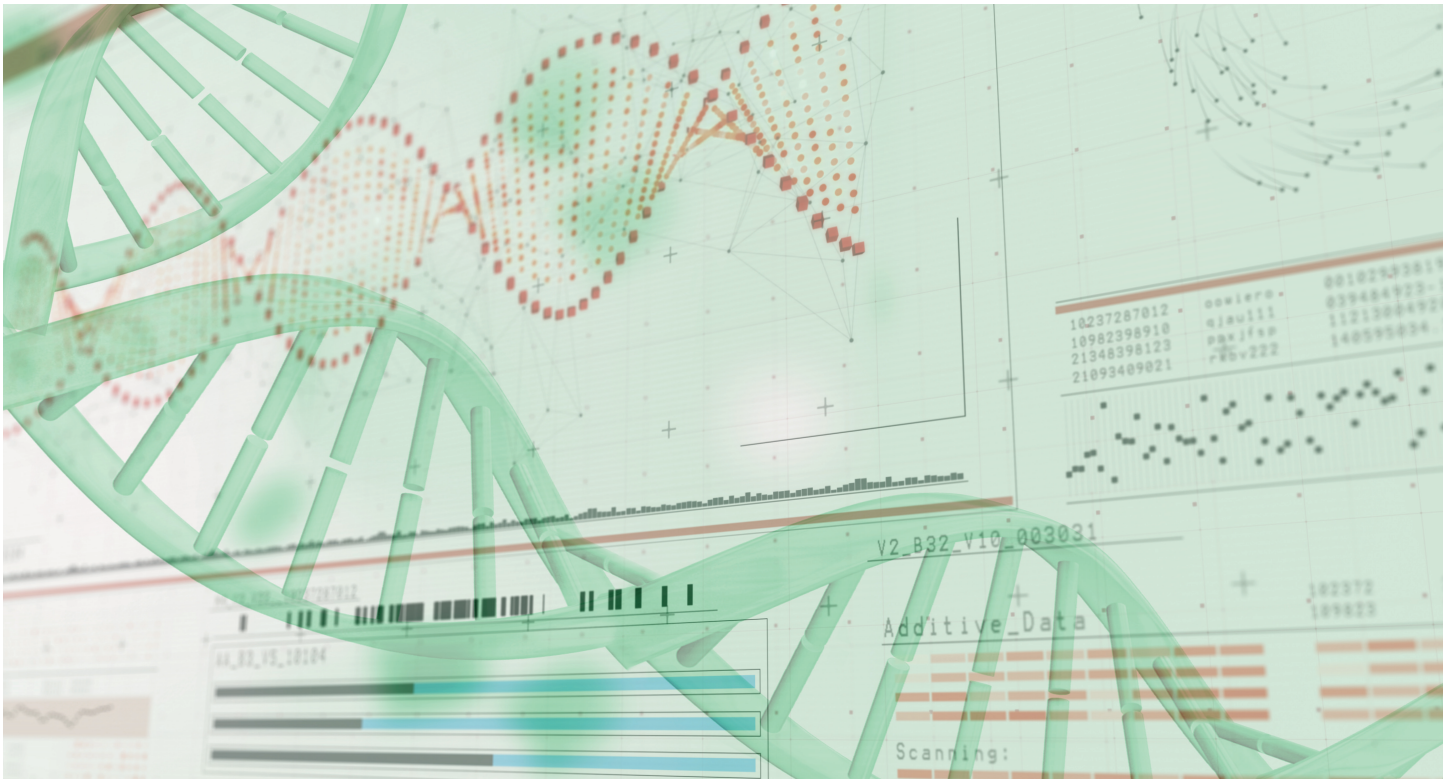
Hosts of sequence information must require users to agree to benefit sharing

Hosts of sequence information including databases like Genbank and the European Nucleotide Archive (ENA) must be required to have their users to agree to benefit sharing as a condition of access to sequence information. "Open access," or "publicly-available" databases do not, and cannot, mean "no user obligations." The CBD should develop rules for such data user agreements (e.g. "click-wrap" terms and conditions), and sequence information hosts should be required to implement them.

Sequence information includes DNA, RNA, and amino acid sequences

Genomic nucleotides are only part of the relevant sequences. The hereditary material of organisms is not just DNA but in some cases is RNA. And because of the complementarity between the molecules, and their important functions, the sequences of both must be covered. The sequences of amino acids that nucleotides encode are similarly valuable and can be used to replicate and modify natural compounds and in design of biological systems.

Beyond the "building block" sequences, there is a range of information that is also part of sequence information, and that includes characterization data on genetic resources, epigenetic, methylation and similar information, and useful and discernable data and patterns, for example, short tandem repeat (STR) counts.



Conclusion

It is, for most CBD intents and purposes, effectively impossible to recall sequence information once it has been released into the electronic realm without benefit sharing strings attached. But stopping sequence information publication *per se*, of course, should not be the objective of CBD Parties. Rather, the CBD COP needs to quickly move to clarify that sequences come with the same benefit sharing obligations as apply to physical material, and this principle may be carried over to the approaches used by FAO and WHO.

In that way, the generation and use of sequence data will not be impaired, but rather will occur among actors committed to the Convention's ABS goals. If developed countries resist or refuse to apply benefit-sharing rules to sequence information in the Convention's process, then provider countries must consider using ABS laws and agreements to place restrictions on the right of users to sequence and share sequences of genetic resources.

Once it is established by the Convention that benefit sharing fully applies to sequence information, the stage will then be set for the CBD's next great sequence information challenge: how to ensure that the capabilities of big data are put to use in a way that genuinely serves the interest of the indigenous peoples and local communities that nurture and preserve biodiversity, and not those companies that would merely use information technology to advance the presently inequitable situation.

1 See UNEP/CBD/SYNBIO/AHTEG/2015/1/3. URL: http://www.cbd.int/doc/meetings/synbio/synbioahteg-2015-01/official/synbioahteg-2015-01-03-en.pdf

2 CBD COP Decision XIII/16. Digital sequence information on genetic resources. URL: https://www.cbd.int/doc/decisions/cop-13/cop-13-dec-16-en.pdf

3 See the WHO PIP Framework website, URL: http://www.who.int/influenza/pip/en/