#### **Special Event on Genomics Information**

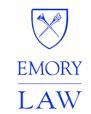
#### **Genomic information: technological & institutional dimensions**

#### **Scoping study:**

#### **Technological Landscape**

Todd Kuiken, North Carolina State University Eric Welch, Arizona State University Margo Bagley, Emory University Selim Louafi, CIRAD



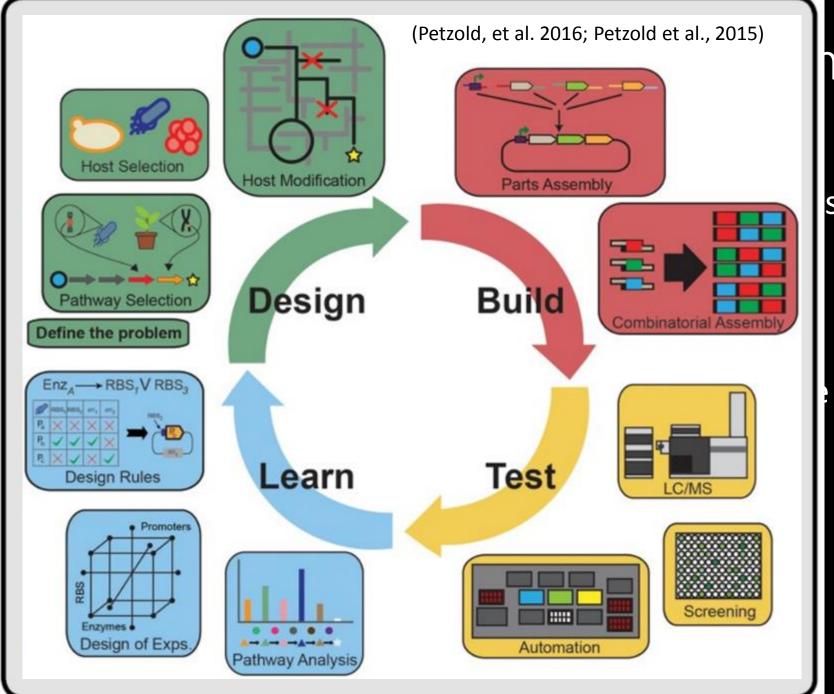






Sy

- Synthe incorpo
- There a
  - Mos
- Unless changing



hic

s that

rapidly



Farmers in Engand begin cross-breeding plants to create hybrids. Today, virtually all commercially available fruits, vegetables and grains have been altered by humans.

1940s

Breeders begin to seek new methods of introducing genetic variation to plants and the term 'genetic engineering' is coined.



1970: The discovery of type II restriction enzymes, generates wheat varieties that increase yields by 70%; they become a fundamental tool in genetic engineering.

**1971**: First synthesis of a gene, by US scientists.

**1972**: First recombinant DNA molecules - the birth of contemporary biotechnology.

1973: World's first transgenic animal

1977: Transgenic *E. coli* used to produce insulin.

1990: The enzyme used to make cheese is one of the first food products to be manufactured using recombinant DNA technology.

1994: The first commercialised genetically engineered crop: the Flavr SavrTM tomato.

1997: Dolly the sheep – the world's first cloned mammal - is born in Scotland.



vears ago

Humans begin the process of genetic modification by selecting plants and animals with desirable

800s traits.

Darwin's Origin of Species is published and Mendel describes how traits are passed between generations, providing the scientific basis for selective breeding.



In England, Watson and Crick describe the structure of DNA accelerating efforts to understand and manipulate genetic material.



Researchers develop more precise methods of genetic engineering.

**1982:** First genetically engineered plant is reported, in the US.

**1986:** These genetically engineered plants are grown in the field for the first time.



2000s



New plant breeding techniques, such as cisgenesis and intragenesis, are developed, originating in the Netherlands.

2009: Draft genome sequence of the potato published – the world's third most important crop.

**2010:** First fully synthetic, self-replicating bacterial cell is produced by American scientist Craig Venter.

2012: Draft genome of wheat is published.

2013: Scientists in the US pioneer genome editing using CRISPR/Cas9

- Many elements are entirely novel
- Involves designing and constructing new biological "parts", "devices" and "systems"
- Can involve much larger-scale interventions
- Pace of change is increasing

European Commission Future Brief, Synthetic Biology and biodiversity: http://ec.europa.eu/environment/integration/research/newsalert/pdf/synthetic biology biodiversity FB15 en.pdf

## Moving from physical to digital

- Prior to DSI technologies a researcher interested in an organisms genetic information would need to obtain the organism and then have it sequenced
  - Today its possible to download an organisms sequence information from a database or academic journal and have it synthesized by a DNA synthesis company (or increasingly synthesize it themselves)
  - Breeders and researchers have characterized numerous alleles (or mutations) associated with plant domestication and improvement, and have identified the genes and phenotypic differences between crops and their wild relatives.
  - According to some researchers, these mutations are the "raw material on which selection can operate making species adaptation and long-term evolution possible" (Nogue et al., 2016).
  - This 'raw material' is equivalent to computer codes that can be analyzed, reprogrammed and theoretically used either within the plant it was obtained from, or within a different species of plant.
  - This type of description or abstraction of the plant's genetic make-up is part of the scientific/engineering philosophy surrounding synthetic biology
- However many factors determine whether its possible to build living material from DSI and whether it will function in a living system:
  - Cost
  - Time available
  - Ease of access to and quality of DNA synthesis and construction
  - Quality of the DNA libraries and digital sequence information in them
  - Quality of the "parts" (if in physical format)

# Proliferation and decentralization of data (material)

- In the past physical material was accessed through collections held in gene banks, botanical gardens or private/public collections
- Fueled by decreasing costs and improvements in computer software (AI) we have seen a shift towards digital collections that include complete genomes and "parts" registries
  - Some are public databases (iGEM registry), some are public/private partnerships and others are individual researchers collections
  - Some of the databases also include the physical "parts" along with the digital file (iGEM)
- There has been a push (in part) towards open data and open exchange of data across the globe
  - This is driven by the need to access different data sets in order to understand and utilize the "parts" or functionalities that are encoded in the genomes of organisms
  - It's not clear whether these databases will remain "open" as our understanding and utilization of these parts increases
- The field seems to be moving towards a highly distributed service oriented model in which foundries and DNA synthesis companies build their own collections of sequence data and then produce the physical constructs based on those digital files

### Monitoring DSI?

- As discussed vast amounts of DSI are being collated in different databases and repositories (Genbank, Adgene, NCBI, iGEM and others)
- While some interviewees mentioned data sharing and use agreements there was no indication that such agreements are widespread or imposed by all database operators
- In theory database access could be tracked
  - some interviewees suggested that some database operators have been and will continue to be resistant to implementing such tracking
- It's not clear what you should track
  - Many ways that partial sequence information (parts) can be combined
  - The same sequence (or part) could be present in multiple organisms

#### "New" technologies and plant breeding

- Advances in genetic technologies, machine learning (AI), and growing databases of DSI could impact plant breeding
  - The rise of DNA construction and genome editing technologies (i.e. CRISPR/Cas9) could facilitate the construction of multiple variants that involve alterations to multiple genes across an organism (National Academies of Sciences, 2017)
  - Many of these technologies are still young and will depend on our understanding of the plant species, complexity of the traits of interest and how they interrelate
- Researchers are increasingly able to scan DSI and identify traits of interest across multiple species
  - Need access to digital libraries of DNA sequence data
  - Do not (necessarily) need access to the physical material
    - Once the traits are identified they can be synthesized
  - Many researchers we spoke with suggested they still need/prefer to go back to the physical material from which the trait was identified in
    - This could present a problem for both research and ABS if that origin information is not tracked with the DSI
      - We found that this origin information varies by DSI database/collections

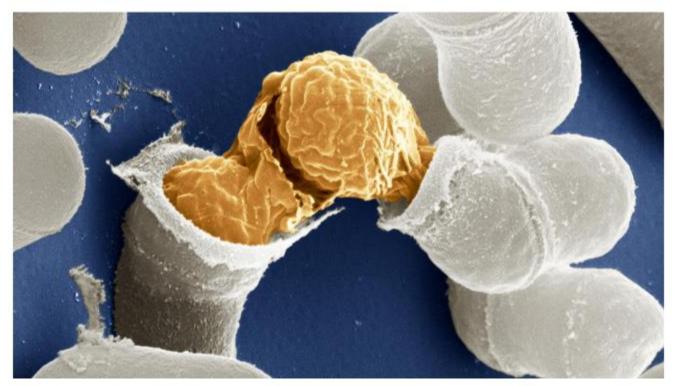
## Deskilled/Democratized/Decentralized

Synthetic Biology based on standard parts Iceland Greenland Kazakhstan Mongolia Turkey Algeria Libya Egypt Western Salura Saudi Arabia Mauritania Mali Nigeria DR Congo Zambia Mozambique Madagascar Google My Maps

IN THE LAB

#### College students try to hack a gene drive - and set a science fair abuzz

By IKE SWETLITZ @keswetlitz / DECEMBER 14, 2016

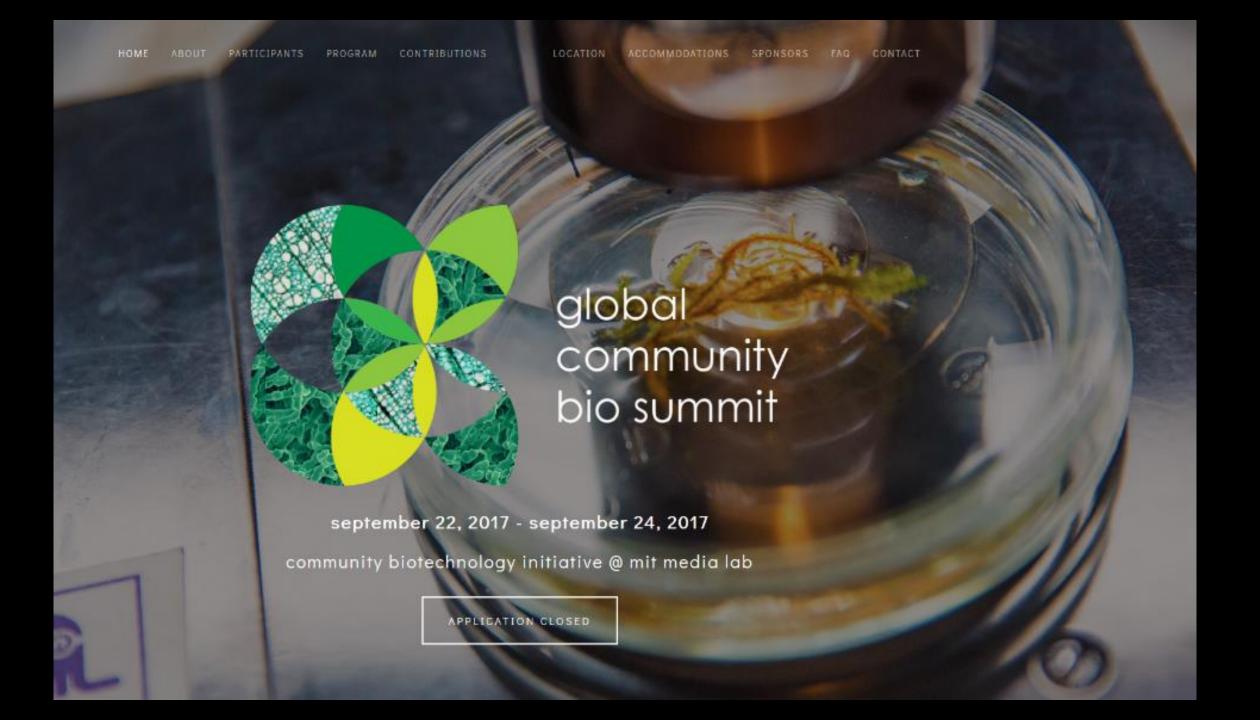


College students worked with yeast cells like these, attempting to insert a "gene drive." JUERGEN BERGER/MARIA LANGEGGER VIA NIH

#### May 2008 – 1<sup>st</sup> DIYbio meet-up







## Final Thoughts....

- 1. The digitization era is producing large amounts of sequence data that is widely available and easily exchanged
  - 1. High throughput and automation of screening has enabled researchers to "screen thousands to billions of variants of an organism for function or phenotype" (National Academies of Sciences, 2017)
  - 2. Mining plant genomic information (data) for gene editing purposes in agriculture
  - 3. Mining for plant genomic information (data) for use outside of agriculture
- 2. While many researchers we spoke with still require or prefer the physical material for their work, there was a suggestion that the separation between material and data (DSI) is increasing
- 3. Demand for screening technologies is increasing and moving towards "omics approaches that are agnostic to the type of organism being tested
- 4. There is a community building dimension to "synbio"
  - 1. Most people know each other
  - 2. They also need each other (both in terms of data and techniques)
  - 3. De-skilling/Democratization
    - 1. iGEM
    - 2. DIYbio

## Thank You....

Todd Kuiken, Ph.D. <a href="mailto:tkuiken@ncsu.edu">tkuiken@ncsu.edu</a>
@drtoddoliver
+1-919-515-2593



