

Understanding Digital Sequence Information

From Genetic Resources to DSI

Moisés Cortés-Cruz

DNA & Genomics Laboratory

National Genetic Resources Center

National Forestry, Crops and Livestock Research Institute

SEMARNAT
SECRETARÍA DE
MEDIO AMBIENTE
Y RECURSOS NATURALES



Al servicio
de las personas
y las naciones



Strengthening of
National Capacities for
the implementation of the
Nagoya Protocol (ABS)

THE ABS
CAPACITY
DEVELOPMENT
INITIATIVE



3rd International ABS Dialogue

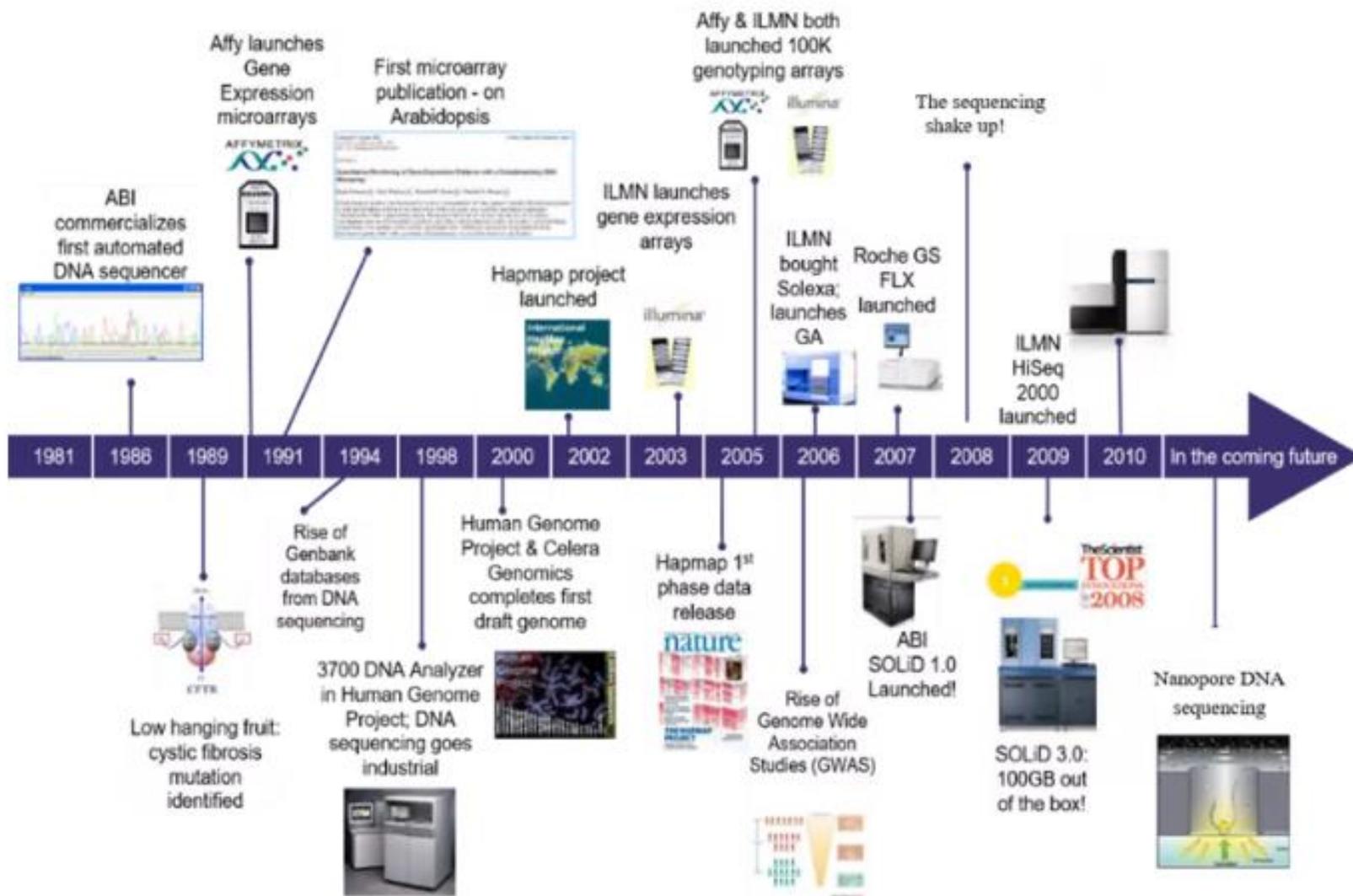
Mexico City, May 25th 2018

- **How DSI is obtained and managed?**
- **DSI for characterization and development of new plant varieties**

Can DSI be understood?

- There was a consensus that the term “DSI” is not the appropriate term to refer to these types of information (AHTEG on DSI, February 2018).

How DSI is obtained?



Examples of DSI

Whole genome sequencing

- ✓ Whole genome (re)sequence is a method to analyze a specific polymorphism from a species with a known reference genome. It provides the most comprehensive map of an organism's genetic make up.

Whole exome sequencing

- ✓ Whole exome sequencing is a method that selectively analyzes coding regions (exons). It is a cheaper and effective way of sequencing compared to WGS, that analyzes the entire genome.

Whole transcriptome sequencing

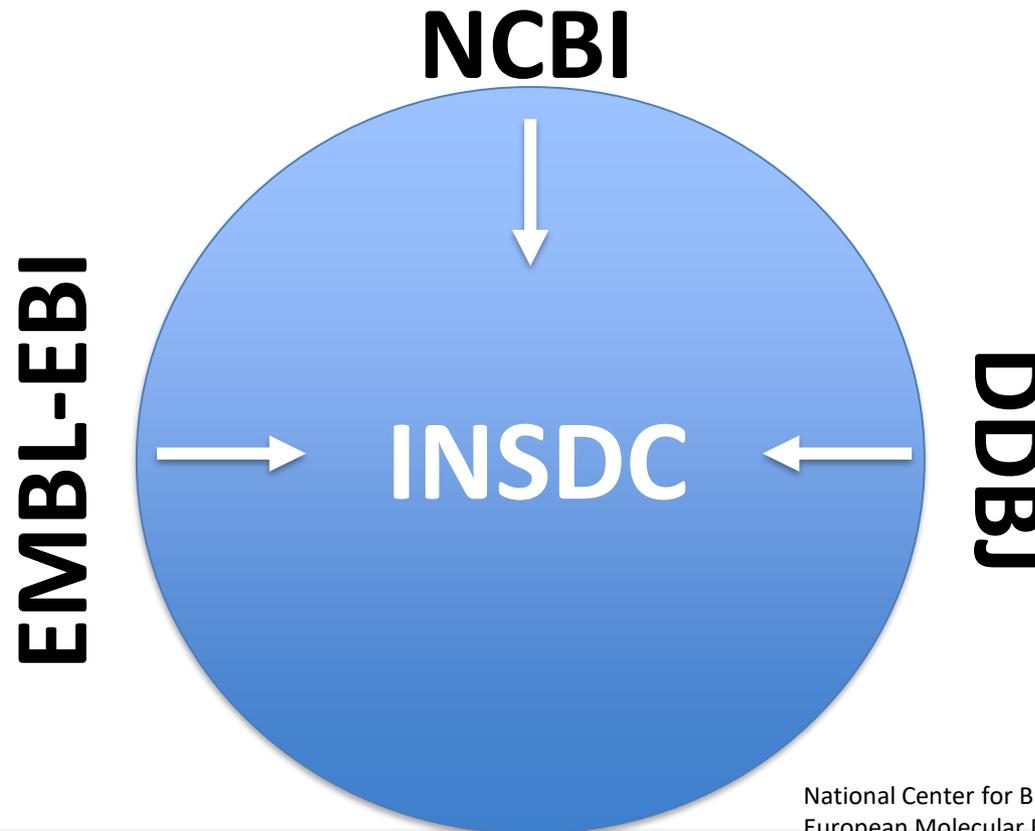
- ✓ Whole transcriptome sequencing is a major advance in the study of gene expression compared to traditional microarray-based approaches in that it provides a comprehensive view of a cellular transcriptional profile at a given biological moment enabling not only qualitative but also quantitative analysis of the transcriptome. It also provides information about diverse variations occurring at RNA level of post-transcriptional modifications such as splice variants and isoforms.

Expression profiling sequencing

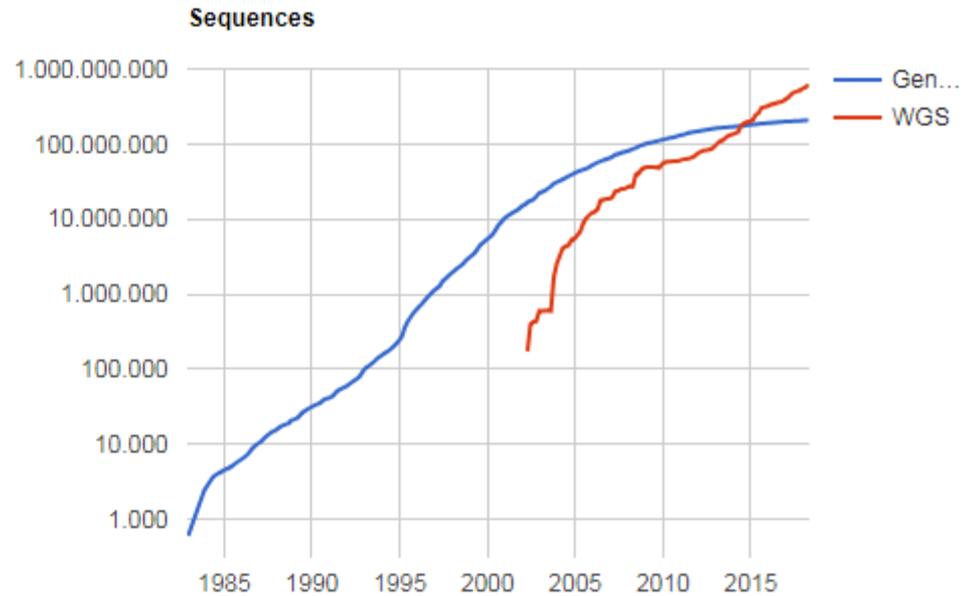
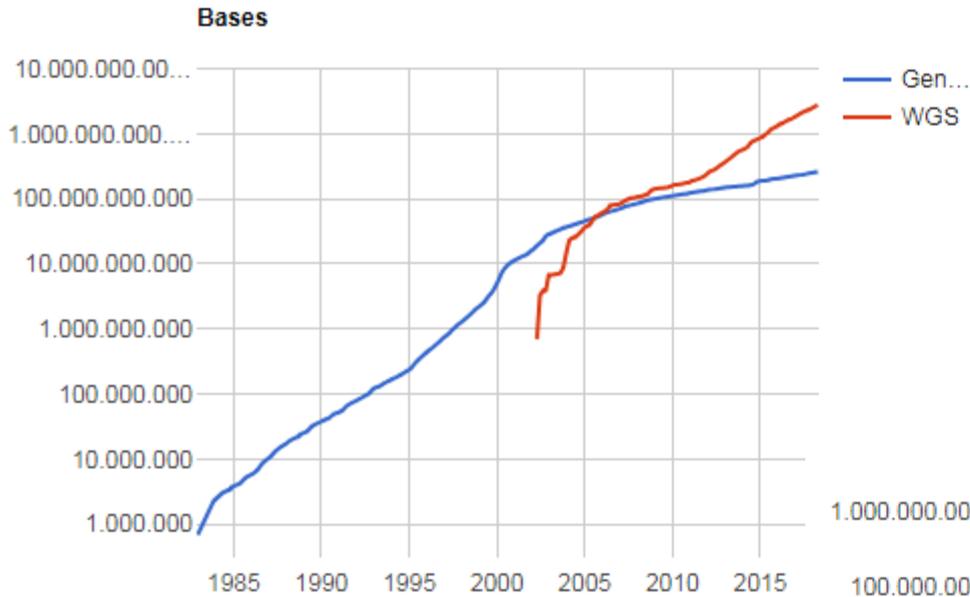
- ✓ Expression profiling sequencing gives most comprehensive insight about the transcription level of individual mRNAs which is then further related to the expression level of proteins.

Management of DSI

Basically, there are three main databases where DSI is stored (freely available over the internet)



GeneBank and WGS Statistics



GeneBank and WGS Statistics

Number of sequences available through the INSDC as of April 2018 is:

GeneBank		WGS	
Bases	Sequences	Bases	Sequences
260 189 141 631	208 452 303	2 784 740 996 536	621 379 029

- A Google scholar search on DNA sequencing [AND] plant characterization, shows over 15,300 scientific papers published between Jan 1, 2013 and May 24, 2018.
- SNP markers are one of the most used sequencing technologies for plant characterization (genetic diversity studies, varietal characterization, GWAS, etc).
- EDV and the TPP (UPOV 1978 vs 1991).

Genebanks (GB)

- GB harbor a large store of genetic diversity, with the potential to accelerate crop breeding
- Sequencing of model (and non-model) organisms (reference genomes)
- Systematic exploration and use of crops diversity for genetic research and breeding
- Identification of useful accessions for breeding
- Extraction of accessions from GB for targeted phenotyping

Core Collections

- Molecular characterization (new methodologies, GBS)
- Phenotypic evaluation
- CC development
- Availability in GB for breeders
- Efficient breeding schemes

- DSI holds a great importance for conservation and sustainable use of genetic resources biodiversity.
- DSI can be used for several objectives (diversity, breeding, taxonomy, etc.).
- Currently, the majority of DSI is allocated in open access databases.
- However, not all countries have the same capability to generate, analyze and utilize DSI.
- Although there is a high number of crop genomes sequenced, there is still a lot of work to be able to fully understand the function of all identified coding regions (putative, homologous and functional genes).

Thank you!

SAGARPA
SECRETARÍA DE AGRICULTURA,
GANADERÍA, DESARROLLO RURAL,
PESCA Y ALIMENTACIÓN



inifap
Instituto Nacional de Investigaciones
Forestales, Agrícolas y Pecuarias



National Genetic Resources Center
National Forestry, Crops and Livestock Research Institute
Blvd. de la Biodiversidad 400
Tepatlán de Morelos, Jalisco, MEXICO
www.inifap.gob.mx

