Executive summary of the report on the use of digital sequence information on genetic resources for food and agriculture

French Foundation for Biodiversity Research

View the full report (in French): https://www.fondationbiodiversite.fr/wp-content/uploads/2019/11/FRB-Rapport-DSI-2019.pdf

ABSTRACT

The field of life sciences has benefited from significant technological breakthroughs in matter of sequencing since the second half of the 20th century. As genetics makes increasing use of digital techniques, Digital Sequence Information (DSI) is becoming an important element for molecular biology in the era of genomics. Information technology (IT) is used for both processing information and its transfer to international databanks (see Figure 1).

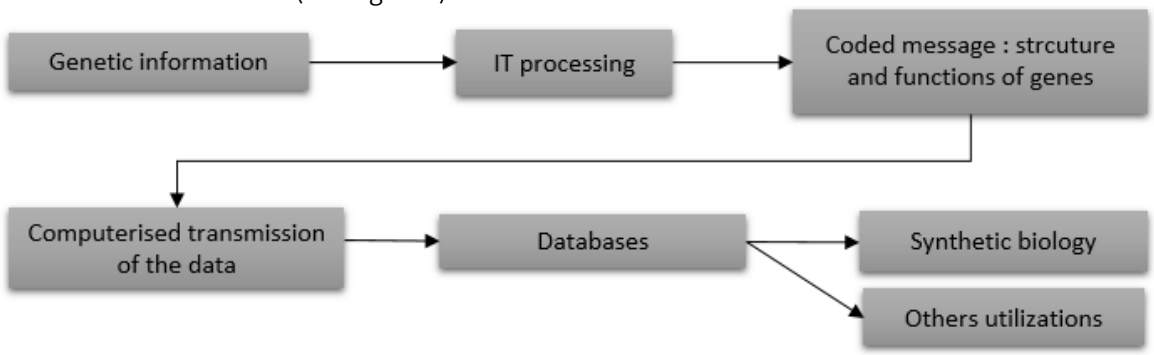

Figure 1: A close relationship between genetics and the digital[1].

The Convention on Biological Diversity (CBD), signed in Rio de Janeiro on June 5, 1992, established a contractual access and benefit-sharing (ABS) regime between suppliers and users of genetic resources based on the reaffirmation of the sovereignty of the countries (Article 15 of the CBD). The Nagoya Protocol related to the CBD aims to ensure access to genetic resources and the fair and equitable sharing of benefits arising from their use. The increasing reliance on international databanks for storing and sharing DSI raises the question on their legal status. At present, there is no official regulation as towards DSI databases. This issue is therefore at the heart of the debate between signatory parties of the Convention on Biological Diversity (CBD), the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) and the United Nations Convention on the Law of the Sea.

Within the framework of the 14th Meeting of the Conference of the Parties to the Convention on Biological Diversity (CBD COP 14) that took place between 10 and 22 November, in Egypt, the French Ministry of Agriculture and Food commissioned the Foundation for Research on Biodiversity to prepare an analytic report on the use of DSI on genetic resources for food and agriculture (plant, animal, aquatic, forest, microorganisms, invertebrates)[2].

The importance of DSI and their particular relevance to food security as well as to climate change adaptation and mitigation was stressed by the Commission on Genetic Resources for Food and Agriculture (CGRFA). This report has the aim to inform the public on the subject of DSI on genetic resources for food and agriculture (GRFA), their users and their usage.

Main results

The main findings of the report were therefore:

- to propose an appellation to replace the term "digital sequence information" with "digital data on genetic resource sequences" or "digital sequence data";
- to define a typology following the chronology of the bioinformatics protocol for processing sequencer data outputs: raw data, cleaned data, analysed data;
- to list the main applications of this digital data.

[1] Rey A. (2017). Le traitement de l'information génétique par le droit. L'exemple de l'information liée à la biodiversité, Thèse, Université de Montpellier.

[2] The report was launched during a seminar that took place on Monday, 8 October, at the House of Oceans in Paris. The seminar was attended by almost 50 people with different backgrounds (representatives of ministries, diplomats, researchers, industrialists, journalists). The seminar's meeting minutes and the speakers' presentations are available on the FRB website.

This is a recent concept, the implications of which have recently been discussed, particularly in the context of the development of new genome manipulation technologies, and for which there is no officially recognized definition[3].

The DSI on GRFAs cover several interests: from the study of genetic diversity to genetic characterization. Those interests vary according to the type of genetic resource considered.

Microorganism sequencing projects are the most popular given the small scale of microorganism genomes, inferior to that of plants or animals and therefore requiring less time to sequence and analyses. Issues arising from the use of DSI on GRFAs also depend on the commercial applications that may result but genetic selection programs are gradually integrating the entire spectrum of GRFAs. The discovery of markers of interest (sex, resistance to a parasite, etc.) allows for an early selection of offsprings.

Sequencing data are for example the basis for the work:

- characterization for the conservation of local poultry breeds (e. g. BioDivA project);
- on disease control in the shellfish sector (e. g. VIVALDI project);
- epidemiological monitoring of bees or control of colony collapse syndrome (e. g. BEEHOPE project)
- varietal improvement in cultivated species (e. g. SUNRISE and Genius projects);
- study of microbial diversity within the dairy chain (e. g. CNIEL)
- to facilitate the domestication of yeast for the agri-food industry (e. g. Bakery project)
- the rapid development of new strategies for the selection and production of forest varieties.

[3] Moreover, the definition of this notion is itself at the heart of the discussions and issues that led to the realization of this work because the main question that is asked to the community is to know whether to regulate this data in the same way as the genetic resources from which they are derived: is DSI, itself, comparable to a resource that would thus enter into the definition of the concept of genetic resource and which would benefit from the same regulation?

CONTRIBUTORS

COORDINATION AND REDACTION
Charlotte Navarro, French Foundation for Research on Biodiversity
Jean Lanotte, French Ministry of Agriculture and Food
Steering Committee of the study
Ana Deligny, translation

QUOTE
French Foundation for Research on Biodiversity (2019), Executive Summary of the Report of the Study on the Use of Sequencing Data on Genetic Resources for Food and Agriculture, Paris, France: FRB, 26p.

Study carried out as part of the study project on the use of digital sequence information on genetic resources for agriculture and food commissioned by the Ministry of Agriculture and Food in May 2018.
Directorate: Ministry of Agriculture and Food
Coordination: Robin Goffaux (FRB)
Conduct of the study: Charlotte Navarro (FRB)
Editorial: Charlotte Navarro (FRB), Robin Goffaux (FRB)

ACKNOWLEDGEMENTS
The Biodiversity Research Foundation would like to thank the entire Steering Committee for their expertise and guidance, as well as all those who took the time to complete the questionnaire.

Summary reviewers: Robin Goffaux (FRB), Jean Lanotte (MAA), Hélène Soubelet (FRB).

Table of contents

I. HISTORY OF MOLECULAR BIOLOGY AND THE RISE OF GENOMICS

The evolution of life sciences is tightly linked to the technological progress. Digital sequence information is becoming an important element for scientific analysis and the scientific community is considering enabling sharing this type of data via open access databases.

A. OBSERVING THE GENOME

A living cell contains a series of instructions (consisting of individual "genes") called the "genome." Each individual instruction is encoded as a chemical sequence representing a molecule of four elements also called nucleobases (A, C, G, T for DNA[4] and A, C, G, U for RNA[5]). A specific combination of these four nucleobases, called "sequence," allows encoding a particular instruction as the byte succession encodes information for computer programs. In this particular case, the encoding is of biochemical nature. A sequence is therefore a combination of the building blocks of a DNA molecule (basis) and is depicted using the letters A, C, G, and T[6].

Molecular biology is a standalone scientific branch that studies, among others things, biological macromolecules as nucleic acids, e.g. DNA, and proteins. This discipline is at the heart of the scientific activities of a large number of researchers studying the expression of genetic information and its regulations (see Appendix 6). Molecular biology tools are at the brink of life sciences and computer science[7].

Genomics is the study of genes that define a particular species and their genome. In recent years, genomics was market by significant developments[8]: from a descriptive phase to a functional experimentation phase. Genome analysis is essential for the study of living beings and their functions.

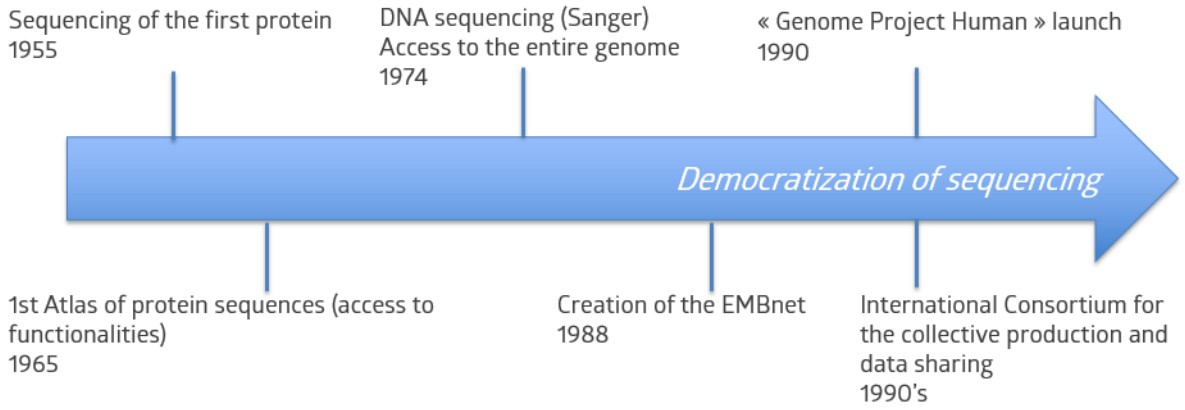

Figure 2: Main historical stages in the evolution of genomics and of sequencing techniques, in particular (Christine Gaspin, 2015)

[4] DNA: Deoxyribonucleic acid. Large molecules that contain instructions (genes) and make up the chromosomes (J. Weissenbach, 2000).

[5] RNA: Ribonucleic acid. Molecule consisting of a sequence of nucleotides, considered as intermediate support for genes to synthesize proteins, and which has other functions. A for adenine, C for cytosine, G for guanine, T for thymine and U for uracil.

[6] Weissenbach J. (2000). Texte de la 27ème conférence de l'Université de tous les savoirs réalisée le 27 janvier 2000, Le séquençage du génome humain : comment et pourquoi.

[7] Gallezot G. (2002)."La recherche in silico", In : Chartron G. (sous la dir.) "Les chercheurs et la documentation numérique : nouveaux serviceset usages", Edition du cercle de la Librairie, Collections.

[8] Gaspin Christine (2015). « Les données de la recherche dans le domaine des sciences du vivant : évolution et perspectives à la lumière des nouvelles technologies du numérique et d'exploration du vivant », Présentation à Toulouse.

Sequencing is the set of procedures for determining the sequence of a DNA or RNA molecule, or that of a protein. Bioinformaticians have to handle thus three alphabets: DNA, RNA, and proteins. The study of the genome can be compared to reading a large volume of text and inferring words and functions to understand the mechanics of living beings (see Figure 2). The size of a genome varies considerably and depends of the given species.

Bioinformatics is the automatic treatment of biological information. It applies computer science to genomics by : acquiring, organizing, analyzing, visualizing, and modelling information. Today, computer scientists, mathematicians, biologists and researchers in the field of information science join the teams of life sciences researchers.

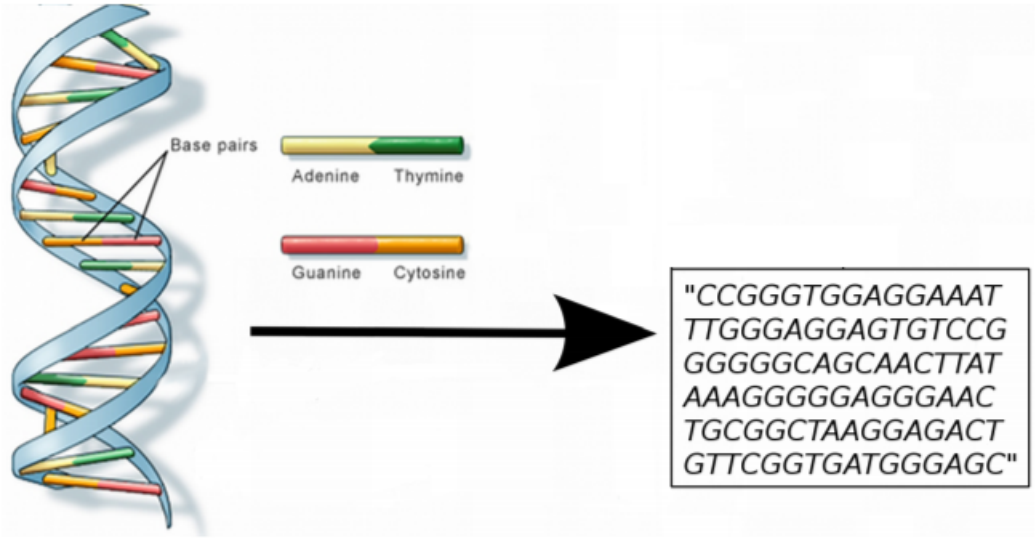

Figure 3: DNA biological and information image (P. Leleux, 2014)

B. THE PRODUCTION OF SEQUENCING DATA AND THEIR DATABASE SETTING

The work of researchers consists of developing different ways of retrieving DSI. These include the collection of information that can be carried out using different media (databases, websites, laboratory experiments, field collection); the processing of information, which is usually done by bioinformaticians or scientists with the knowledge of bioinformatics; the dissemination of knowledge that corresponds to standardized operations[9]. These activities form a cycle that begins with the acquisition of information from databases and is followed by their processing using new computerized sequencing techniques (see Figure 3).

[9] Gallezot G. (2002)."La recherche in silico", In : Chartron G. (sous la dir.) "Les chercheurs et la documentation numérique : nouveaux services et usages", Edition du cercle de la Librairie, Collections.

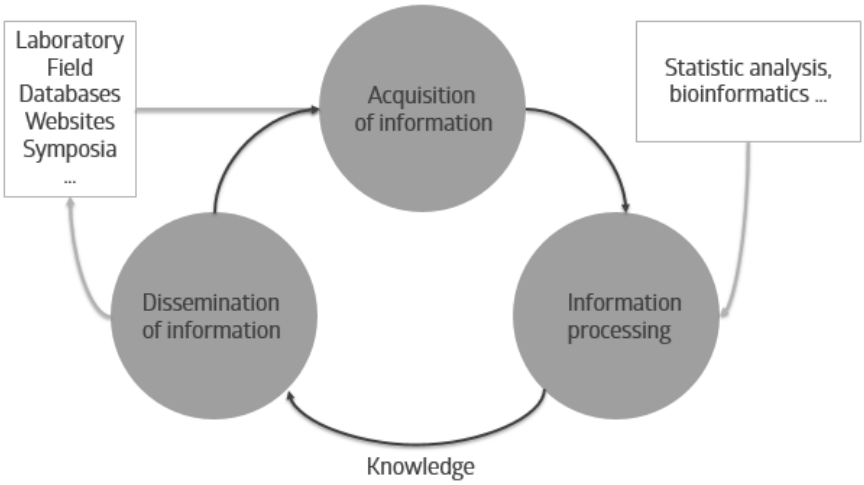

Figure 4: The cycle data collection, processing and reporting (Gallezot, 2002)

Since the first digital sequencing of a 3 billion bases sequence, performed within the framework of the Human Genome Project in 2003 at the cost of 300 million dollars and over a 10 years' period, international databanks are recording a rapid increase in the number of stored sequences. Today, digital sequencing can be performed in a few days and only for a few thousand dollars. The Human Genome Project introduced research to the era of genomics. The results of this project paved the way for a new generation of research programs aimed at decrypting the function of newly detected genes in all living species. Functional genetics now combines biochemical and physiological approaches with a complete genome analysis.

In order to facilitate sharing digital sequence information, the International Nucleotide Sequence Database Collaboration (INSDC) was created. It represents a major initiative between the National Centre for Biotechnology Information (NCBI) in the United States, the European Molecular Biology Laboratory (EMBL) and the Japanese Bioinformation and DDBJ Center. This collaboration covers the spectrum of raw data reads, functional annotations, and contextual information about samples and experimental setups. These databases include:

- raw digital sequence information obtained as the output of a measuring equipment (see Figure 5):

```
>gnl|ti|1586495440 name:1047100384971 mate:1902487597
TTGCAAGCTTAGTATTACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTCTTGCC
AAAGACAATGCACCGCGGGACATTGCTGTACCAATCACCTTTTGATCCACTTCCTACCGAATGGATGCAA
AAATCAGTTTTAAATAGACAAAGGCATGTGGGAGAGGCGATCTTAGGGTTCCCTCTAGATCTACAGGGTG
ACCTAGTTGATGCGAATGGAGAGACTTGTAGAAGTTTTGATCGGTCAGGTTTATGTACTAGTTTCTCTAA
ATCTGCATCTACAGGTAATGATCTTTTACTTGGTAAAAAAAAAAAAAAAAAAAAAAAAAAGTACTCTGCGT
TGATACCACTGCTTAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACAA
TTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCA
GCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGC
GCAGCCTATACGTACGGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGA
TGTACAGAGTGATATTATTGACACGCCGGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTG
TCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCA
CCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGA
CATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTCAGCATGAGATTATCAAAAAGGATCT
TCACCTAGATCCTTTTCACGTAGAAAGCCAGTCCGCAGAAACGTGCTGACCCCTGATGAATGTCAGCTAC
TGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGTAGC
```

Figure 5: Raw digital sequence information of the Arabidopsis thaliana (Sequence Read Archive, 2019)

- readings of sequences resulting from capillary electrophoresis which are the chromatograms[10] of a DNA sequence (see Figure 6):

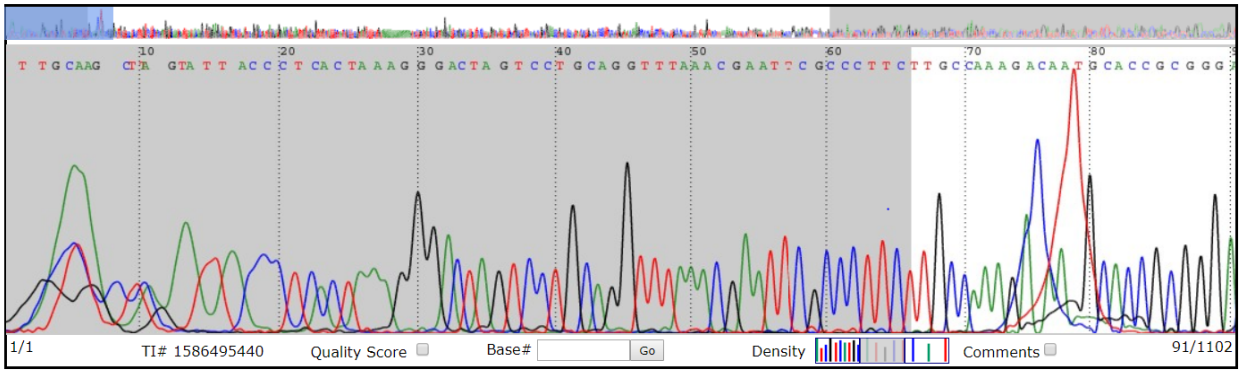

Figure 6: capillary reading or DNA sequence chromatograms of Arabidopsis thaliana (Trace Archive, 2019)

- annotated sequences also known as identified functional regions, these are often the genes that encode the proteins (see Figure 7):

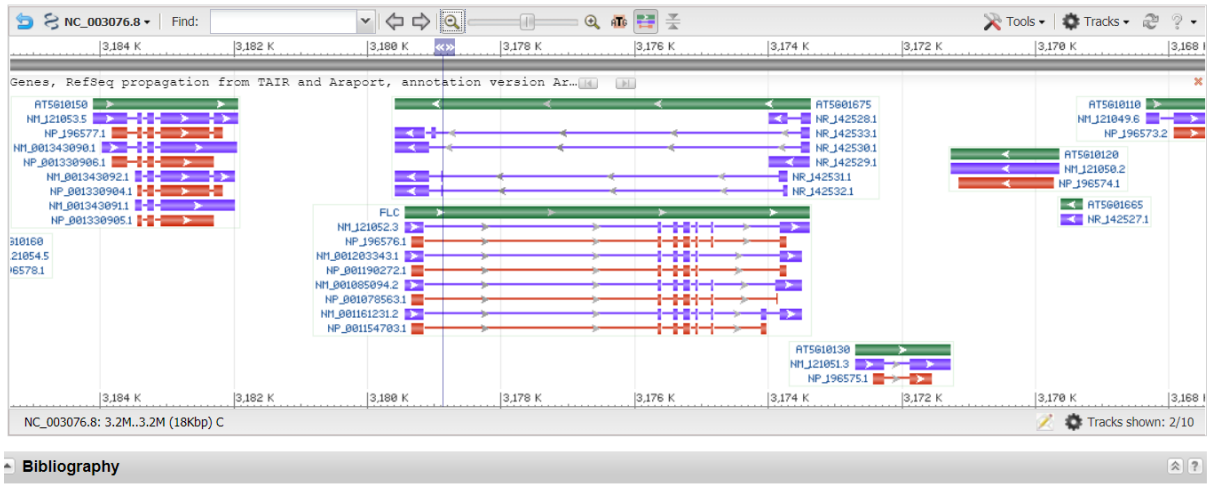

Figure 7: Annotated sequences of the Flowering Locus C gene encoding the MADS-BOX (flowering) protein of Arabidopsis thaliana (GenBank, 2018)

Some databases combine digital sequence information with additional explanatory data, e.g. BioSample adds the description of biological materials for experimental trials (identifier, organism, title, description, link, etc.) and the BioProject collection includes project related biological data (field, methodology, objectives).

[10] Diagram resulting from a chromatography, technique for separating the chemical components present in a mixture. In the case of FIG. 6, the different fragments resulting from the degradation of *Arabidopsis thaliana* DNA show variable lengths and different nucleic acids at the end of the chain. By identifying these different lengths and the nucleobase that ends them, we retrace the sequence of these bases.

II. DEFINING DIGITAL SEQUENCE INFORMATION

Digital Sequence Information includes data outputs from activities in the field of genomics, that is to say the discipline bringing together different techniques to study genetic information. These can be of different nature[11]:

A. FACTUAL DATA VS TEXTUAL DATA

The factual data materialize in the form of nucleotide sequences (A, T, C, G, U) to which are associated annotations filled in by the repositories of the sequences. Textual data is associated with scientific publications.

Factual data or representations of nucleotide sequences are derived from experiments ("benchtop") or retrieved from international sequences databanks. Their description follows a standardized notice where the custodians can specify fields for the information of their data:

- Biological identity, a kind of civil status: name, type of molecule, biological affiliation, date of entry (LOCUS field), access number (ACCESSION field) as an identifier of the record in the databank, definitions and keywords, origin of the sequence (SOURCE field), etc. (see Figure 8);
- Sequence-related bibliographic references;
- Properties of the sequence (FEATURES field): annotations describing the sequence, i.e. functions of subsequences as well as their specific position and attributes (see Figure 9);
- DNA sequence text (ORIGIN field): a representation of the nucleotide sequence (ATGC symbols) (see Figure 10).

```
LOCUS       SCU49845     5028 bp    DNA             PLN       21-JUN-1999
DEFINITION  Saccharomyces cerevisiae TCP1-beta gene, partial cds, and Axl2p
            (AXL2) and Rev7p (REV7) genes, complete cds.
ACCESSION   U49845
VERSION     U49845.1  GI:1293613
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 5028)
  AUTHORS   Torpey,L.E., Gibbs,P.E., Nelson,J. and Lawrence,C.W.
  TITLE     Cloning and sequence of REV7, a gene whose function is required for
            DNA damage-induced mutagenesis in Saccharomyces cerevisiae
  JOURNAL   Yeast 10 (11), 1503-1509 (1994)
  PUBMED    7871890
REFERENCE   2  (bases 1 to 5028)
  AUTHORS   Roemer,T., Madden,K., Chang,J. and Snyder,M.
  TITLE     Selection of axial growth sites in yeast requires Axl2p, a novel
            plasma membrane glycoprotein
  JOURNAL   Genes Dev. 10 (7), 777-793 (1996)
  PUBMED    8846915
REFERENCE   3  (bases 1 to 5028)
  AUTHORS   Roemer,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Terry Roemer, Biology, Yale University, New
            Haven, CT, USA
```

Figure 8: Sample recording of the "Locus", "Accession" and "Reference" fields of an annotated sample in Flat file format in the GenBank database

[11] Gallezot G. (2002)."La recherche in silico", In : Chartron G. (sous la dir.) "Les chercheurs et la documentation numérique : nouveaux services et usages", Edition du cercle de la Librairie, Collections.

```
FEATURES             Location/Qualifiers
     source          1..5028
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="IX"
                     /map="9"
     CDS             <1..206
                     /codon_start=3
                     /product="TCP1-beta"
                     /protein_id="AAA98665.1"
                     /db_xref="GI:1293614"
                     /translation="SSIYNGISTSGLDLNNGTIADMRQLGIVESYKLKRAVVSSASEA
                     AEVLLRVDNIIRARPRTANRQHM"
     gene            687..3158
                     /gene="AXL2"
     CDS             687..3158
                     /gene="AXL2"
                     /note="plasma membrane glycoprotein"
                     /codon_start=1
                     /function="required for axial budding pattern of S.
                     cerevisiae"
                     /product="Axl2p"
                     /protein_id="AAA98666.1"
                     /db_xref="GI:1293615"
                     /translation="MTQLQISLLLTATISLLHLVVATPYEAYPIGKQYPPVARVNESF
                     TFQISNDTYKSSVDKTAQITYNCFDLPSWLSFDSSSRTFSGEPSSDLLSDANTTLYFN
                     VILEGTDSADSTSLNNTYQFVVTNRPSISLSSDFNLLALLKNYGYTNGKNALKLDPNE
```

Figure 9: Sample recording of the "Feature" field of an annotated sample in Flat file format in the GenBank database

```
ORIGIN
        1 gatcctccat atacaacggt atctccacct caggtttaga tctcaacaac ggaaccattg
       61 ccgacatgag acagttaggt atcgtcgaga gttacaagct aaaacgagca gtagtcagct
      121 ctgcatctga agccgctgaa gttctactaa gggtggataa catcatccgt gcaagaccaa
      181 gaaccgccaa tagacaacat atgtaacata tttaggatat acctcgaaaa taataaaccg
      241 ccacactgtc attattataa ttagaaacag aacgcaaaaa ttatccacta tataattcaa
      301 agacgcgaaa aaaaaagaac aacgcgtcat agaacttttg gcaattcgcg tcacaaataa
      361 attttggcaa cttatgtttc ctcttcgagc agtactcgag ccctgtctca agaatgtaat
      421 aatacccatc gtaggtatgg ttaaagatag catctccaca acctcaaagc tccttgccga
      481 gagtcgccct cctttgtcga gtaattttca cttttcatat gagaacttat tttcttattc
      541 tttactctca catcctgtag tgattgacac tgcaacagcc accatcacta gaagaacaga
      601 acaattactt aatagaaaaa ttatatcttc ctcgaaacga tttcctgctt ccaacatcta
      661 cgtatatcaa gaagcattca cttaccatga cacagcttca gatttcatta ttgctgacag
      721 ctactatatc actactccat ctagtagtgg ccacgcccta tgaggcatat cctatcggaa
      781 aacaataccc cccagtggca agagtcaatg aatcgtttac atttcaaatt tccaatgata
      841 cctataaatc gtctgtagac aagacagctc aaataacata caattgcttc gacttaccga
      901 gctggctttc gtttgactct agttctagaa cgttctcagg tgaacctttct tctgacttac
      961 tatctgatgc gaacaccacg ttgtatttca atgtaatact cgagggtacg gactctgccg
```

Figure 10: Sample recording of the "Origin" field of an annotated sample in Flat file format in the GenBank database

Textual data refer to the literature in the broad sense of the term (journal articles, scientific works, conference proceedings, etc.) which is based on the use of these data, and where these are analyzed, interpreted and discussed. Their description follows a catalographic system (authors, title, summary, review, date, etc.) allowing their referencing and facilitating their dissemination. Bioinformatics uses this classification to extract information automatically from databanks. However, biological information is often better described in scientific articles than in databanks, therefore a certain challenge in exploiting this information in its wider framework is evident.

B. PROPOSED TYPOLOGY

First, raw data represents sequencer data outputs, which are not saved for future operations. Second, computer processed "clean" data[12], which represents text files in the fast.q format, also called a "sequence." Last, the assembly process[13] produced new text files, which include the analysis of the previously obtained "clean" data. This typology follows the program chaining which represents the bioinformatics protocol (pipeline) for the analysis of data outputs from a sequencing broadband. The files that are generated along different steps of the program differ in terms of size and usefulness to the user[14].

Figure 11: Pipeline or the Bioinformatic Protocol for sequencer data processing

The French National Institute for Research (Inra) proposes, in more detail, the following diagram to characterize different types of data. It identifies the raw data that, after processing, becomes the so-called "clean" data, then the final - analyzed - data that communicates scientific knowledge and which is found in the scientific "publications."

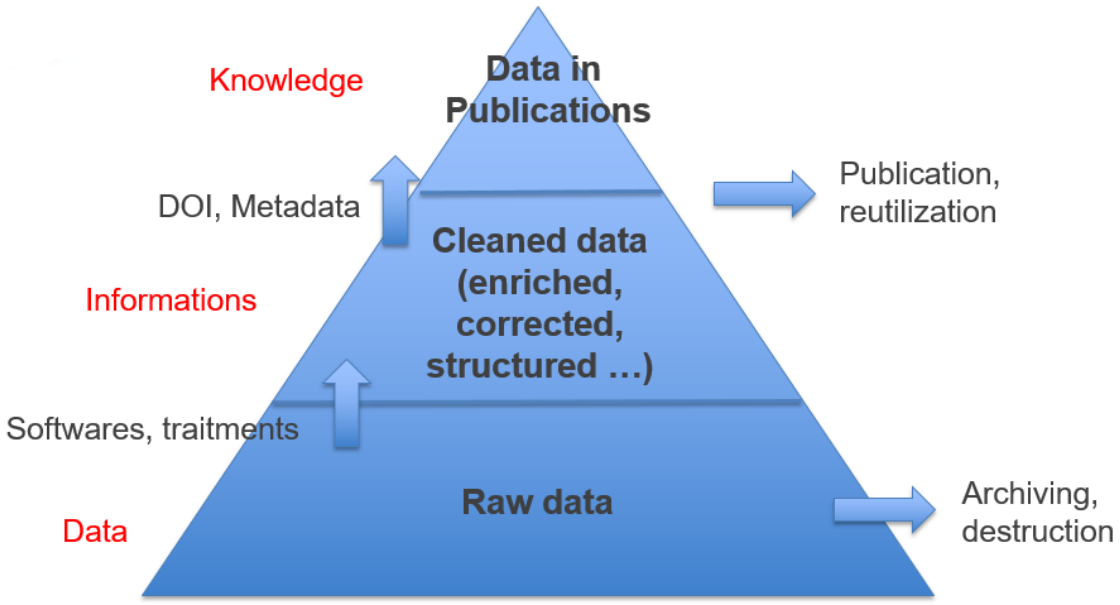

Figure 12: Data classification, Inra

[12] The FASTQ file (extension. fastq ou .fq) is a standard text file used for sequence and qualitative data exchange by all types of sequencers, including Sanger. It includes the names of the sequences, the sequences themselves, and the quality value of nucleotides. This file should be compressed for storing.

[13] The assembly process of aligning or fusing the sequence fragments for reconstructing the original sequence. It can be compared to the reconstruction of a textbook that was shredded into small pieces (Rayan Chikhi, 2012, in Leleux, 2014).

[14] Working Group « IT, bio-analysis/bioinformatics, databases mutations specifications within the framework of the NGS-Diagnosis Network, General Recommendations for Broadband Sequence Data Management and Analysis for Molecular Diagnosis of Genetic Diseases Laboratories, May 2016.

Nonetheless, as technology evolves rapidly the cost of analysis, management and storage of data increases faster than the cost of data production. Today, there is no business model that considers addressing this issue.

III. THE USE OF DIGITAL SEQUENCE INFORMATION ON GENETIC RESOURCES FOR FOOD AND AGRICULTURE: IMPLICATIONS AND PRACTICAL APPLICATION

A. THE TYPOLOGY OF DIFFERENT USAGES FOR DIGITAL SEQUENCE INFORMATION ON GRFAS

Microorganism sequencing projects are the most popular given the small scale of microorganism genomes, inferior to that of plants or animals and therefore requiring less time to sequence and analyze. Nevertheless, as sequencing technologies advance, animal and plant genetic resources (with a complex genome) are becoming the subject of an ever-increasing number of projects.

Through the cases encountered during this work, several objectives motivated the production of DSI on genetic resources: fundamental knowledge of the composition and functioning of the genome, the identification and exploration of genetic diversity, genetic improvement by different selection techniques based on the study of DSI or in vitro modification of the genome.

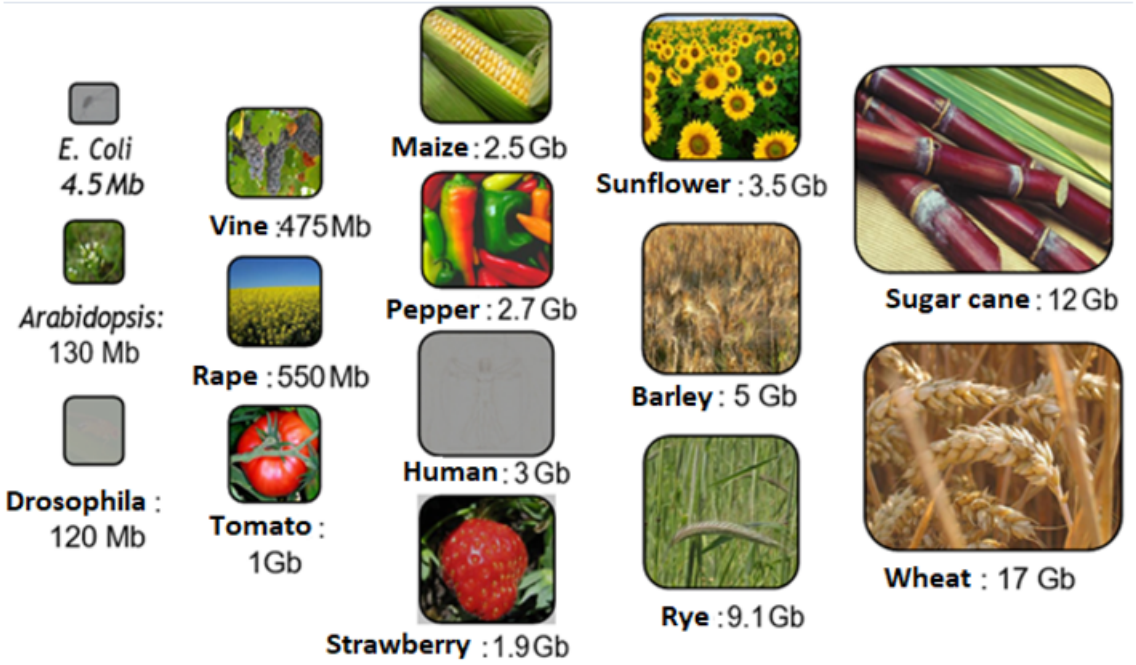

Figure 13: Genome size, Centre of National Plant Genetic Resources

The use of DSI on GRFAs is also subject to commercial application. In particular, with regard to animal genetic resources, cattle have been traditionally the subject of selection and breeding programs.

For example, some of these programs consisted in identifying genetic markers responsible for disease and performing male counter-selection, as a result of sequencing the genome of the animals.

These genetic selection programs are gradually integrating the entire spectrum of GRFAs. The discovery of markers of interest (sex, resistance to a parasite, etc.) allows for an early selection of offsprings. For example, for sturgeons, understanding the major determinant of sex allows earlier selection of caviar producing females.

The process of sequencing provides a better understanding of the molecular functioning of organisms and refines genetic selection programs.

As an observation, for every type of genetic resource analyzed, e.g. plant, animal, aquatic, forest, microorganisms and invertebrates, for species of lower economic interest, genetic diversity studies are generally more frequent. These studies measure the degree of variety in the genes within the same species. Mapping involves locating known DNA sequences along chromosomes. It permits the representation of a genome as tags.

Others technics are utilized such as marker-assisted selection to track genes and early a sorting of genetic resources of interest after natural crossbreeding. Genome editing has allowed, more recently, new possibilities for targeted modification of agronomic character. However, these procedures concern to a larger degree microorganisms (bacteria) and in a small proportion plants, as explained by the size of their genomes. The more complex a genome is, the more difficult it is to understand its functioning.

Main uses of DSI on GR:
• Analysis of allelic diversity • Development of tools (DNA chips) capable of tracking gene activation under certain conditions • Characterization of genetic resources • Location and identification of genes present in certain genome areas (genetic mapping) • Identification of most effective / interesting alleles • Marker-assisted selection: selection of breeding individuals, sorting of descendants obtained • Facilitation of gene cloning for better-targeted, more efficient genetic transformations using specific genes or seeking to extinguish or disable the expression of certain unwanted genes

Table: Typology of main uses of DSI on GR, summary findings of the study

Interviews conducted during the investigation revealed various uses of DSI on GRFA. A summary table of these examples is available in the appendix (see annex).

B. EXAMPLES OF DIGITAL SEQUENCING INFORMATION USES

i. BioDivA Project: genetic characterization for the conservation of local poultry breeds

Figure 14: Gray chicken of Vercors (Association Ouantia Grise du Vercors)

Objective: The low numbers in individuals observed in traditional local breeds and / or old French chicken breeds pose a threat to their genetic diversity and therefore, ultimately, to their existence. In order to address this issue, the BioDivA project aims to characterize the genetic diversity of French local chicken breeds and to help foster the implementation of appropriate conservation programs. The BioDivA project is funded by the Ministry of Agriculture and Food and the Rhône-Alpes Region and began in 2013 for a period of three years.

Tools: Molecular analysis proves to be the perfect tool to characterize genetic diversity. Between 2013 and 2016, 1,517 animals were genotyped, revealing a great genetic diversity within French local breeds

(Restoux et al., 2017). This diversity study is a precursor to implementing *in vivo* and/or *in vitro* preservation programs.

Results: The genetic characterization of endangered breeds and the safeguard of pedigrees in databases have allowed to develop and implement targeted preservation programs (Chiron et al., 2018). The development of genetic management tools for breeds that include a small number of individuals renders possible, among other things, to select breeding candidates according to predefined objectives and to establish mating plans for local breeds. For example, dedicated software developments allow the French professional association of selection enterprises to propose breeding lists of males and females for the conservation of intra-breed diversity and for creating of genetic progress while controlling consanguinity. Selected data is henceforth transferred, controlled and analyzed step-by-step allowing the evaluation of lineages by the members of the association in order to ensure long-term selection nuclei and variability within the herds.

ii. Vivaldi Project: control of diseases impacting the shellfish industry through the epidemiological monitoring of species

Context: European shellfish farming holds a privileged place on the world scale. European shellfish production is mainly based on mussels, oysters and clams. In recent years, the sector has been weakened by high mortality associated with various viruses (e.g. OsHV-1), bacteria (e.g. *Vibrio aestuarianus*) and parasites (e.g. *Marteilia cochillia*), which induced a significant economic loss to the sector.

Oyster farming

Objective: The VIVALDI Horizon H2020 European project began in 2016 for a period of 4 years. It aims to increase the sustainability and competitiveness of the European shellfish sector, which brings together different shellfish cultures, by developing tools and approaches to prevent and control bivalves'[15] diseases. The project aims to study the genetic resources of molluscs and their pathogens, from samples taken in Europe, Israel and Norway. To meet these needs, VIVALDI must not only provide new insights into complex interactions between shellfish, environment and pathogenic organisms, but also focus on the development of practical tools and approaches to prevent and control diseases affecting marine bivalves. As diseases know no borders, an international network bringing together experts from major shellfish producing countries such as China, Japan, Korea, Australia, New Zealand, the United States and Canada will be established. At the heart of this network, VIVALDI will contribute to sharing information and experiences on shellfish mortality for improved control over associated diseases.

Tools: The sequencing of the complete genome of bivalves.

Results: Many samples have already been collected and are being analyzed for the study of diversity of pathogens affecting bivalve molluscs. An innovative result has shown that it is possible to detect virus DNA in oyster beds using immersed plastic strips that act as sensors[16].

[15] Marine bivalves are a class of seawater molluscs, also known as *Pelecypoda* (pelecypods) or *Lamellibranchia* (lamellibranchs). This class includes clams, oysters, mussels, scallops and many other shellfish families.

[16] Ifremer (2017). La recherche européenne pour une conchyliculture durable et compétitive., premier bilan un an après le démarrage du projet Vivaldi.

iii. Future Investment Program (PIA) SUNRISE: from complete sequencing of genomes to the varietal improvement program

Context: Sunflowers, through their low water intake, represent one of the solutions to facilitate the adaptation of the plant sector to the effects of climate change. Improving sunflower drought resistance and its agronomic characteristics is therefore a major environmental issue. The world production of oilseeds, and especially sunflower seeds, must also meet a growing demand for human nutrition (diversification of oils), animal feed (protein richness of its cakes) and for the development of biofuels and green chemistry.

Objective: The aim of the SUNRISE Project (2012-2020) is to carry out complete genome sequencing, which will serve as a basis for new genetic selection projects and for resequencing the genomes of 300 sunflower varieties for the identification of agronomic markers.
The project is governed by a consortium agreement that also establishes its data sharing rules. It foresees the construction of an initial database to encourage new variety improvement projects.

Results: One of the results of the SUNRISE Project was to highlight the genetic variability of sunflower for photosynthesis processes and leaf transpiration of the plant in a context of water deficit. These results can be integrated into agricultural crop models. The identification of drought tolerance genes will improve genetic selection programs and bring to market new varieties adapted to climate change.

iv. Genius Project: "Cellular engineering: Technological improvement and innovation for sustainable agriculture plants," tools for targeted modification of agronomic traits

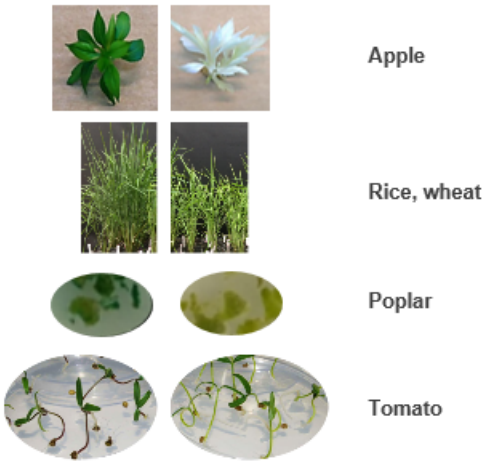

Figure 15: Cultivated plants with improved crop and quality traits for food, feed and other uses, Peter Rogowsky, 2018.

Objective: The Genius project (2012-2019) aims to respond to current challenges of sustainable agriculture, through the study of several traits of 12 different species (nine cultivated and three models) in order to reduce inputs - fertilizers and pesticides - by increasing resistance to pathogens in tomatoes, apples, poplars, rapeseeds, to facilitate adaptation to climate change by enhancing the tolerance to salinity in rice, to improve the use of plant biomass by improving the quality of starch in potatoes, etc. This project is financed by public funds (governmental future investment program).

Tools: The project involves applying targeted gene modification tools to abovementioned species. When designing the Genius Project in 2011, the project partners focused on meganucleases and TALENs, in full development at the time. Since then, they have been able to adapt the work program to take into account the appearance of the Cas9-CRISPR technology in 2012. It is important to note that these changes were implemented ever since the preliminary stage of observation of the characters of interest in the phenotype of field and laboratory plants. The characters of interest observed are, for example, the color of leaves for apple trees or the length of stalks for rice.

Results: The main research was focused on developing selection tools. For corn, this allowed obtaining diploid gametes for asexual propagation of crops. Another example of the Genius Project research is the work on enhancing the quality of final products, for potatoes, gene(s) modification allowed producing starch composed only of amylopectin, of primary use for the food industry and as glue. The following table details the research results for the Genius Project.

Research theme	Gene modification	Results
Selection tool	Corn: diploid gametes	Asexual propagation of cultures
Product quality	Potato: starch composed solely of amylopectin	Food industry and glue
Flowering time	Apple: very early flowering	Shortened life cycle and adaptation to climate change
Adaptation to abiotic stress	Rice: tolerance to salinity	Cultivation on marginal lands and adaptation to climate change
Disease resistance	Tomato: potyvirus resistance	Plant protection and pesticide reduction

v. Precompetitive projects in the dairy industry: study of microbial diversity

Context: The characteristics of cheese (colour, acidification, texture, flavour, etc.) are partly determined by their microbial ecosystems, which therefore contribute to the quality of cheese itself. The microbiological component in cheese production and the particular characteristics of outputs are of significant importance but they represent complex elements and vary from one type of cheese to another. DSI is used to better understand microbial ecosystems, their diversity and their functionality to improve production and milk processing. In France, the National Interprofessional Center for Dairy Economics (CNIEL) crated thus a microbial strains bank (the "MIL" collection) composed of flora of interest, pathogenic flora, and bacterial viruses. The collection is managed by Actalia, an Agro-Industrial Technical Institutes (ITAI[17]).

Objective: The aim of the CNIEL project is to establish a catalogue of microbial communities present in all cheeses protected by a Protected Designation of Origin (PDO[18]) in France, and resulting from the combination of various dairy production and cheese processing practices.

Tools: The project mobilized the metagenomic approach associated with the use of new broadband sequencing techniques.
Results: This project will enhance the knowledge on the diversity of natural microbial communities that are lost progressively in milk and cheese production as a result of health reforms.

[17] Agro-Industrial Technical Institutes (ITAI) are private organizations for technological research, expertise, technical assistance and training, and providing services to companies, in particular to SMEs. Positioned at the crossroads of the research communities, companies and professional organizations, they play a major role in the dissemination, transfer and exploitation of research results to small and medium-sized enterprises.

[18] Protected Designation of Origin (PDO) means a product for which all production stages are carried out according to recognised know-how in the same geographical area, which gives its characteristics to the product. It is a European logo that protects the name of the product throughout the European Union (Website of the National Institute of Origin and Quality, Inao).

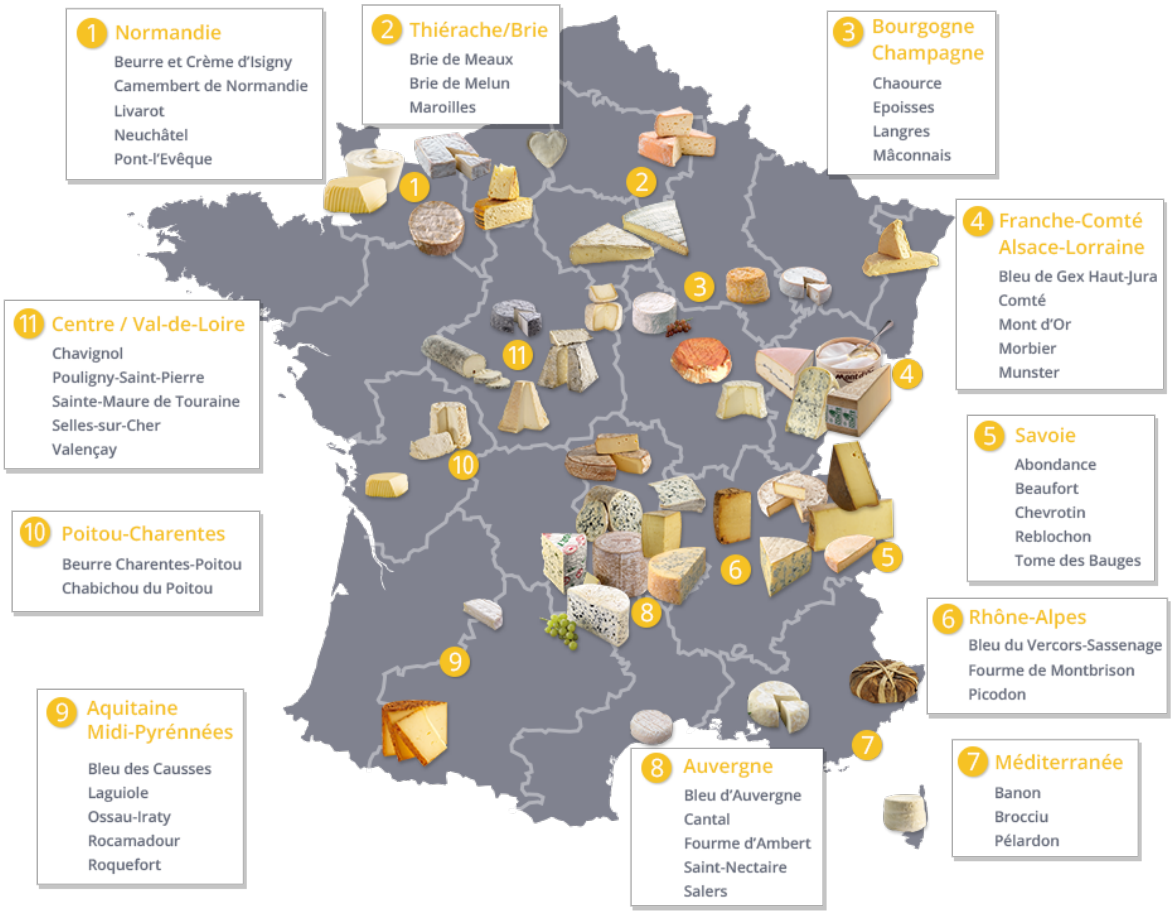

Figure 15: 45 French PDO cheeses, presentation "Acquisition and use of sequencing data in CNIEL-supported projects," Frédéric Gaucheron

vi. Bakery Project: the domestication of yeast for the food industry

Context: The Bakery Project (2014-2018) aims to study the diversity of yeast and the interactions of a low-input wheat / human / sourdough agro-food ecosystem for a better understanding of the means for reaching sustainability in the bakery industry. This project is financed by public funds (governmental future investment program).

Objective: This multidisciplinary and participatory research project aims to (i) describe the socio-cultural diversity of bakery practices and the perception of consumers (ii) study the effects of wheat varieties, local peculiarities and bakers' practices on the diversity of microbial yeast, the sensory and nutritional quality of bread as well as on the preferences of consumers (iii) analyze microbial interactions within yeast and their consequences on the yeast performance and the quality of bread (iv) integrate all data to identify determinants of biological and socio-cultural diversity in the bakery chain, (v) consider strategies for conservation of biodiversity and socio-cultural diversity in bakery.

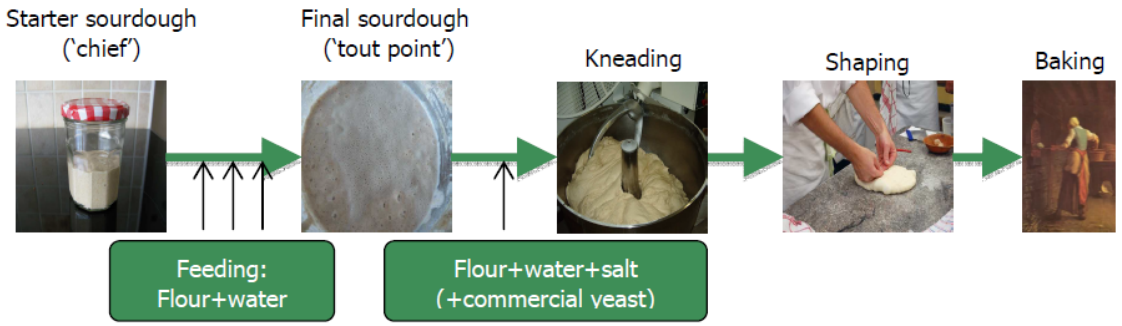

Figure 16: The stages of bread production (Sustainable Food Systems: Bakery Project, ANR 2013)

Tools: Preliminary surveys were conducted with 30 bakers and farmers, mainly with French bakers who make leavened bread, using flour resulting from agro-ecological practices. These surveys provided information on bakers' practices and on the origin of wheat seeds used by bakers and farmers, as well as allowed collecting samples of flour, yeast and bread. During a second step, a survey was conducted among consumers. In the laboratory, microbiome analyses of seeds, flour and yeast mobilized metagenomic sequencing techniques and phylogeny. A biochemical characterization as well as sensory analysis will be further performed on yeast and bread.

Results: The first results show that the low-input bakeries in France host a large and original diversity of microbial species compared to the diversity observed elsewhere in the world. New yeast species have been discovered. Several species of lactic acid bacteria were detected in bakeries for the first time. Different types of bakers (peasant bakers, artisan bakers and small and medium-sized enterprises) host different microbial communities, which shows the importance of maintaining a socio-cultural diversity.

vii. Maritime pine trees: molecular markers for genetic improvement

Context: The genetic improvement of the maritime pine trees began in the 1960s with the selection of forest trees showing superior traits of interest for forestry (growth period, trunk straightness, branching, and resistance to pathogens). These "elite" trees have been preserved by grafting in clone parks; and form the base population of the improvement program. Candidate trees are first evaluated by the performance of their offspring and then the best candidates are crossbred to generate genetic variability for the next generation. In parallel, the best individuals are grafted to establish seed orchards that will provide, after 8 to 10 years, seeds for future plantations. Today, the development of high-performance, low-cost genotyping methods allows developing strategies (for new variety selection and production) that can be implemented in a much shorter period of time.

Tools: The adoption of high-throughput genotyping and high-throughput sequencing technologies led to scientific breakthroughs applicable to forest trees and opened up potential applications in terms of FGR management and conservation.

Results: Based on the nucleotide diversity, it was possible to identify the markers that differentiate different geographical origins (in terms of allelic frequency) which allowed, in its turn, to identify and analyze the geographical origin of a stand or a sample of seeds.

Further on, the ability to store the markers of interest and compare the latter to other markers of interest, which was made possible by the development of new biochip technologies, allowed to improve the researchers' understanding of the reproduction cycle of maritime pine tree orchards. These advancements will contribute to optimizing the design and management of orchards (obtained from planting genetically modified seeds) to maximize the genetic improvements.

A set of 80 molecular markers has been developed to estimate more accurately the genetic value of each tree in order to increase the genetic improvements of future varieties.

Researchers discovered molecular markers that provide a unique identification for each individual and allow tracing its pedigree. It then becomes possible to simplify the selection cycles by replacing the two-parent crossbreed with polycross crossbreed where a mother is crossed with a mixture of several pollens. This strategy also has the advantage of promoting genetic mixing in the breeding population. The pedigree of trees, which is indispensable to evaluating their genetic value with precision, is then reconstituted, *a posteriori*, with the use of molecular markers.

Another approach made possible by the use of molecular markers, and applied to maritime pine trees, is constructing a calibrated prediction model for a population genotyped for a large number of molecular markers (several thousands) and characterized finely for its performance (growth period, straightness of the trunk, etc.) This statistical model then renders possible to predict the genetic value of a tree from molecular markers without waiting for its performance to be measured in adulthood, i.e. leading to a considerable gain in time, of the order of ten years.

viii. BEEHOPE Project to fight bee collapse syndrome and foster sustainable management of beekeeping

Chizé black bee (Biodiversa)

The current extinction rates of species in the biosphere is comparable to that of the latest massive extinctions[19]. The reduction in species richness and genetic diversity is accompanied by the deterioration of a large number of ecosystem services such as pollination by animals (zoogamy). Several biotic factors (e.g. pathogens, alien species) and abiotic factors (habitat loss and fragmentation, agrochemicals, climate change, etc.) are likely to be involved in pollination disturbance and in the decline of pollinating species leading ultimately to a loss of genetic diversity.

The honeybee case is particularly illustrative of these issues: honeybees are of paramount importance for ecology and agriculture; however, losses of honeybee Colonies have been recently reported around the world and at alarming rates. The honeybee is an insect of agri-environmental importance. Honeybee feeding perimeter can extend up to 12 km from the hive, which puts the species in contact with a wide variety of pollutants, including pesticides. For about 20 years, it has been observed that the honeybee populations are in constant decline and pesticides and pathogens emerge as the main contributors to this decline. However, recent studies suggest that current honeybee populations' decline in European apiaries might also be caused by commercial and European honeybee trade through (i) the introduction of unsuitable and artificially maintained colonies (ii) spread of invasive pathogens carried by non-native bees[20].

[19] Franck P., L. Garnery, A. Loiseau B.P. Oldroyd, H.R. Hepburn,M. Solignac, J.M. Cornuet (2001) Genetic diversity of the Honey bee in Africa: microsatellite and mitochondrial data Heredity 86 : 420-430

[20] Résumé à l'intention des décideurs de l'évaluation de la plateforme intergouvernementale scientifique et politique sur la biodiversité et les services écosystémiques (IPBES) des pollinisateurs, de la pollinisation et de la production alimentaire. Copyright © 2016, Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) ISBN: 978-92-807-3568-0 Job Number: DEW/1990/NA

Objective: The BEEHOPE Project by the National Research Agency (ANR) began in 2013 and aims to understand better the ecology of the black bee (*Apis mellifera mellifera*) in order to establish a sustainable management of beekeeping. The black bee is a species that has been abandoned by beekeepers in favor of more productive species, although it is perfectly adapted to the climate and landscape of northern Europe. The project aims to collect data on the black bee with the aim of studying its adaptive traits.

Tools: The evaluation of genetic diversity is carried out using molecular markers. Sequencing is applied to i) creating a new genetic system based on molecular markers, (ii) creating an exclusive molecular marker profile for the bee population of every conservation center, and which can be used for assigning origin to bees (iii) creating a set of genomic fragments including the traits associated with local adaptation.

Results: Despite extensive efforts to protect the black bee, it still registers high levels of hybridization (8% versus 30% for unprotected populations). Some protected populations require further management strategies[21] adjustments in order to eliminate foreign alleles identified with the use of molecular markers (Pint et al., 2014).

[21] In protected areas, selected breeding stocks are mated within isolated breeding stations to prevent gene flow from unwanted sources.

CONCLUSION

The report highlights the multiple identity of sequencing data, which ultimately increases in value as it is processed, analyzed, and cross-referenced with other data.

Two important results stem from the interviews conducted as part of the study. On one hand, a new terminology is proposed as a substitute to the acronym "DSI" "digital sequence information" used by the CBD. This new terminology would be “digital sequence data" or "digital data on genetic resource sequences". On the other hand, a simple typology is proposed and following the bioinformatics protocol of data processing directly from the sequencer: raw data, cleaned data, and analyzed data.

The development and exchange of new sequencing techniques has revolutionized the tools used in the field of molecular biology. Today, other issues arise, and mainly, those related to data processing, data transfer, storage and the legal status of international databanks. Nevertheless, in all, computer sciences have acquired a fundamental place within life sciences research teams.

The interest in DSI lies beyond contributing to improved food production and agriculture. It equally concerns other sectors as cosmetics and pharmaceuticals and, therefore, the regulation of DSI deserves a coordinated approach in technical and legal terms.

Although, generally, an open access to data is largely advocated by European and French research programs, certain legal provisions are already in place to regulate access to data and databases. For example, there is a contradiction between a desire to promote open access to data and a desire to control access to information contained in data sets created as part of partnership research project. Moreover, in France, private research is not required to make the data it produces available if the acquisition of such data has not been obtained thanks to mainly public funding. Here, private research benefits from access to public databases without any restriction and without having participated in its financing thus creating an imbalance in terms of costs and benefits from open access databanks.

The survey of research organizations and networks and enterprises in the agriculture and food sector highlighted various uses of digital sequence information with majority cases such as genetic diversity studies or genome characterization. Other techniques use DSI on GRFA such as marker-assisted selection and more recently the new breeding technics (NBT) such as genome editing. Different genetic resources are used for different purposes.
Historically, animal and aquatic genetic resources have largely benefited from genetic selection projects involving genomics, motivated by economic and technical needs. The genetic resources of microorganisms have equally been the subject of early research programs in genomics, in this case, the explanation lies in the size of their genome that is considerably smaller and therefore easier to analyses. Concerning plant genetic resources, they are included today in a variety of programs ranging from genetic characterization to the addition of traits of interest for agriculture. Recently, research integrating forest genetic resources is entering the genomics era, an opportunity rendered possible by the accessibility of DSI tools and their diminishing cost.

Large-scale and interdisciplinary projects are planned to begin in the near future to search for "lost" genes of the ancestors of domesticated and genetically selected species, that is key to understanding the evolution and adaptation of plant life on earth.

This analytic report raises, however, many questions inferred from interviews and by different research fields that are yet to be explored (cf. annex 18 of the report).

ANNEX:

Summary table of examples of the use of digital sequence information on genetic resources for food and agriculture.

	Name of the project/initiative	Partner	Type of GR (plant genetic, zoogenetic, aquatic, forest, microorganism and invertebrates)	Purpose
1	1011 génomes de levures 2013-2019	University of Strasbourg, IRCAN, Genocscope	Microorganisms GR: yeasts in natural environment and in the diet	Highly detailed genetic map of yeast Saccharomyces cerevisiae Genetic and phenotypic diversity
2	AATTOL 2011-2016	Cirad, Cidres	GR of cattle and microorganisms (parasite)	Characterization of molecular bases of trypanotolerance in cattle
3	Alive 2018-2020	Public and private: AFB, University of Montpellier, WWF, etc.	All type of GR	Creation of a database for DSI and GR of environmental samples
4	Bakery 2014-2018	CIRM-levures, CIRM-BIA,ITAB, universities	GR of yeasts	Genetic diversity of microbial communities
5	BEEHOPE	Six european partners inlcuding : CNRS de Chizé	GR of invertebrates	Genetic analysis; Protection of the bees of the territory (black bee)
6	BiodivA 2012-2016	l'UMR Gabi de l'Inra, le Sysaaf, Itavi, Selection center of Béchanne et Labogena	Poultry GR	Characterization of genetic diversity
7	Catch My Interest	Institut Carnot Plante2Pro, FEDER, Inra UMR LIPM, CNRGV 2016-	GR of Sunflower Resistant and Non-Resistant	Characterize zones of agronomic interest on the genome "diagnostic markers" - region that gives Sunflower resistance to the parasite Orobranche
8	Divseek 2016	68 partners : Africa Rice, Ag Research, AAC, ACPFG, AIT CATIE, CIAT, CGIAR, etc.	GR of plants	Facilitate the generation, integration and sharing of data and information related to plant genetic resources
9	ECOBIOPRO 2010-2013	ADIV, ADRIA, AERIAL, BIOCEANE, IFIP, IFREMER, Inra, ONIRIS, PFI	GR of bacteria, yeasts, molds	Description and evolution of microbial ecosystems of meat products and the sea; Food bioprotection (development of protective crops)
10	EMBARC YEASTIP	Inra, CBS, DSMZ, CABI, etc.	Yeasts : about 5000 sequences of GR of microorganisms	Obtain ten markers from reference strain to facilitate identification

				and phylogeny; taxonomy
11	FISHBOOST 2014-2017	14 european partners inlcuding : Inra, Ifremer, Sysaaf	Aquatic GR	Genomic selection
12	Food Microbiomes levure laitière *Geotrichum candidum* within Saccharomycotina	ANR, CNIEL, French and foreign milk producers	French reference yeast (6000 genes) cheese the bishop bridge	Get to know the genome better to understand possible adaptation to the cheese environment
13	Generation Challenge program 2004-2013 (JC Glaszmann)	200 partners	GR of plants	Crop improvement (drought tolerance)
14	GeneRice : Generation and deployment of Genome-Edited Nitrogen-use-Efficient Rice Varieties 2017-2019	Inra, Cirad, FOFIFA, CIAT, UC Chile	GR of rice, Nepali variety	Marker assisted selection for genetic amelioration of a complex agronomic character (the efficiency of nitrogen utilization); Socio-economic assessment of new plant improvement techniques
15	Genius 2012-2019	Inra, Cirad, Lyon3, Biogemma, Gemricopa, Société nouvelle Pépinières&Roseraies Georges Delbard and Vilmorin	GR of plant	Targeted modification of the genome for adaptation to climate change
16	GnpIS 2002 – today	Inra, Génoplante, Transplant, ELIXIR-Excelerate.	Plant species and their pathogenic fungi	Create a multi-specific integrative information system dedicated to plant and fungus parasites. Identification of links between structure of genetic material and agronomic traits
17	International Wheat Genome Sequencing Consortium (IWGSC) 2005-aujourd'hui	1500 Public-private members; 60 countries	GR of plant of wheat "Gold" from the CNRGV (National center pf vegetal genetic resources)	Fundamental knowledge and characterization of regions of interest; Make a genomic sequence of high quality of soft wheat

18	IRIC (International Rice Informatics Consortium) Projet Genomes riz 3000	IRD, Cirad, CIAT (Colombia), AfricaRice	GR of vegetal varieties of rice	Genetic diversity. Varietal selection
19	MétaPDOcheese Precompetitive project of the dairy industry	CNIEL, France génomique	GR of microorganisms GR	Genetic diversity of microbial communities
20	Project ANR PEAKYEAST 2015-2018	Inra (plusieurs UMR dont STLO à Rennes, l'institut MICALIS de Jouy en Josas, SPO de Montpellier)	GR of microorganisms (yeasts Saccharomyces cerevisiae)	Taxonomic identification; Evolution of the wine yeast Saccharomyces cerevisiae towards its adaptive peak; Characterization of bacteria and yeast relations
21	Project Emissage 2018-2022	ACTALIA, Ifip, Anses, UMT ASIICS	Microbial GR (three serovars of *Salmonella enterica*)	Health surveillance of Salmonella by the chain operators
22	Project Gaïa (has not started yet)	International	Vegetal GR	Exploring hot spots on the surface of the planet biodiversity (genes of interest other than yield)
23	Project IMAGE (Innovative Management of Animal Genetic Resources) 2016-2023	28 Partners : 3 entreprises, 3 NGOs, la FAO, 9 research infrastructure, etc.	Animal GR	Improve animal gene banks for varietal breeding New harmonization of databases; Adaptive character research
24	Project investissement d'Avenir SUNRISE 2012-2019	16 partners (6 seed companies, a biotechnology company, Inra, UPMC)	RG Sunflower (heliantus) + Orobanche parasite species (vegetal GR)	Deciphering the complete genome to "accelerate varietal breeding programs and new varieties that are responsive to changes and respectful of the environment"
25	Private project: Identification of microorganisms	ACTALIA Laboratory, Safety and Food Center, Private Partners	RG of microorganisms (bacteria, yeasts, molds)	Identification of microorganisms responsible for an error in the expected food product (yoghurt that swells, mold development)
26	RETHINk Tomate et Bio agressors 2018-2020	SYNGENTA, Inra	GR of Tomato (200 lines of the wild species *Solanum pimpinellifolium*), + GR of 100 bacterial strains (*Ralstonia solanacearum)*	Identification of genetic bases of mechanisms of sustainable resistance to pathogens; Generate new sustainable

				varieties; Genetic basis of coevolution Tomato / Ralstonia under abiotic stress
27	VIVALDI 2016-2020	21 partners	GR of molluscs and bacteria	Study of the impact of diseases in bivalves (class of molluscs); Future Economic Valuation