Fact checking on new breeding techniques

Yves Bertheau

Fact checking on New breeding techniques

Yves Bertheau

Inra – MNHN

May, 17 2018

European Parliament's Public hearing, Brussels

Over the past 50 years are NBT disruptive techniques? on which criteria?

- As a jump in qualitative and quantitative production?
  - Part of a series of agricultural technical evolutions, often improperly described as 'revolutions' (as for middle-age, green...), with several drastic side effects (e.g. water war in India),
  - Mostly attempting to compensate previous errors (funder effect of domestication, water inefficient use of US corn varieties and see the CIMMYT traditionally bred varieties more tolerant to drought…),
  - Complex metabolic ways difficult to master (safe harbor, regulation…)
- As a competitiveness jump?
  - But all countries are participating to this Red queen's race,
- As a new technical jump?,
  - *Related techniques*: cooking recipes that have been used for several decades generating mutations and epimutations (London's October 2016 seminar),
  - Homologous recombination already available in animals,
  - A technical evolution of molecular biology: random induced mutagenesis, marker assisted selection, transgenesis issued GMOs, NBT, synthetic biology…
  - Numerous false positive and negative results, more precise?
  - Race for finding PAM in several species for extending the number of targetable sequences,
  - More than 20 years about RNAi, meganuclease,... NBT are part of the incremental innovation processes over the last 50 years,
- As a jump in allowing numerous SMEs to enter the market of "innovative seeds"?
  - Numerous patents blocking the market's access to SMEs,
  - Important initial financial investments (labs, greenhouses, sequences and software...)
  - Access to genetic background and breeding facilities,
- As a jump in risk assessment?
  - Techniques used without evaluation guidelines on certain impacts (e.g. epigenome, EFSA symposium, June 2016) nor any appropriate quality assurance scheme,
- As a jump in citizens' welfare?
  - Societal issues that led to the refusal by citizens of certain techniques and leading to regulatory questions (what is mutagenesis, GMO or not, exempted or not?) being processed at the level of the ECJ,
  - A society where technical progress and innovation are asserted as a source of happiness and wellbeing by both private and public actors (ministries in charge of the environment and agriculture, CTPS vs. Evaluation agency ...)
- As a jump in speeding up plant breeding and commercial varieties releases?
- As because numerous papers are published on that issues?
- **As for Crispr inducing less off-targets than the other NBT?**
  - **No one experimental comparison performed,**
  - **Claims of TALEN and ZFN's specialists of more off-targets are found in plants modified by Crispr-endonuclease,**
- **As some are cheaper and more rapid (more experiments performed before one succeeded)**...

Are NBT the continuation of traditional breeding?

Traditional breeding is like a very long boa of ca 10,000 years,

- Random and targeted mutagenesis of the last 50 years are like one boa's scale
- NBT are only one small part of this scale

Numerous limits of the tools for detecting unintended effects

- Relationships very poorly understood between:
  - genome, epigenome and epitranscriptome,
  - genotyping and phenotyping,
- Epigenetics study tools just emerging, no guidelines,
- Sequencing tools with numerous drastic bottlenecks:
  - **cost** of Whole Genome Sequencing, particularly for species without reference genomes, i.e. without important previous investment,
  - **systematic biases** of NGS platforms, software for assemblies, comparisons, off-target prediction, curation of databases sequences and annotations, missing reference genomes, even for « easy exome »,
  - **numerous sequences and structural differences between plants**: necessity to sequence sets of 1000 to 3000 genomes (rice, cattle...), huge differences between Elite cultivars (hampering direct transformations), reference genomes incomplete, pangenome to better describe the gaps

In front of the unexpected complexity of the genomes and epigenomes: the necessity of pangenomes and of numerous sequencing projects…

Available detection methods

- Phenotypic (ex: tolerance to a herbicide, immunology,...)
- Omics (metabolomic modification, transcriptomics, proteomics ...)
- Molecular: genomes and epigenomes / epitranscriptomes (DNA, chromatin, RNA):
  - DNA, RNA and modified or unmodified proteins
  - Simplex (PCR, LCR, OLA ...) multiplex (SNPLex, DNA chips ...),
  - From the nucleotide (SNV detected by LCR, OLA ...) to the large chromosomal rearrangement (border fragment ...),
  - Isothermal or not (LAMP, NASBA ...)
  - Combined or not (eg SNPLex = LCR + PCR + DNA chip)
  - Sequencing (Sanger, NGS, ChiSeq, RNASeq ...) with or without reference genome,
  - On isolated tissues or cells, nucleus or organelles,
  - In the laboratory or in the field (PCR, LAMP, sequencing ...)

- To be combined or not according to the aim
  - Univocal (s) or multiple targets (databases and DSS, see ENGL network works and FP6 Co-Extra program)
  - Analyzes with various software (sequences assembly and comparisons, phylogeny, statistics, gentic maps)
  - Combinable and modular according to the needs: from set of clues as used in legal identification to routine detection

The numerous tools of scientists working on genetic diversity / polymorphism and thus of plant breeders

Scars of the “related techniques”

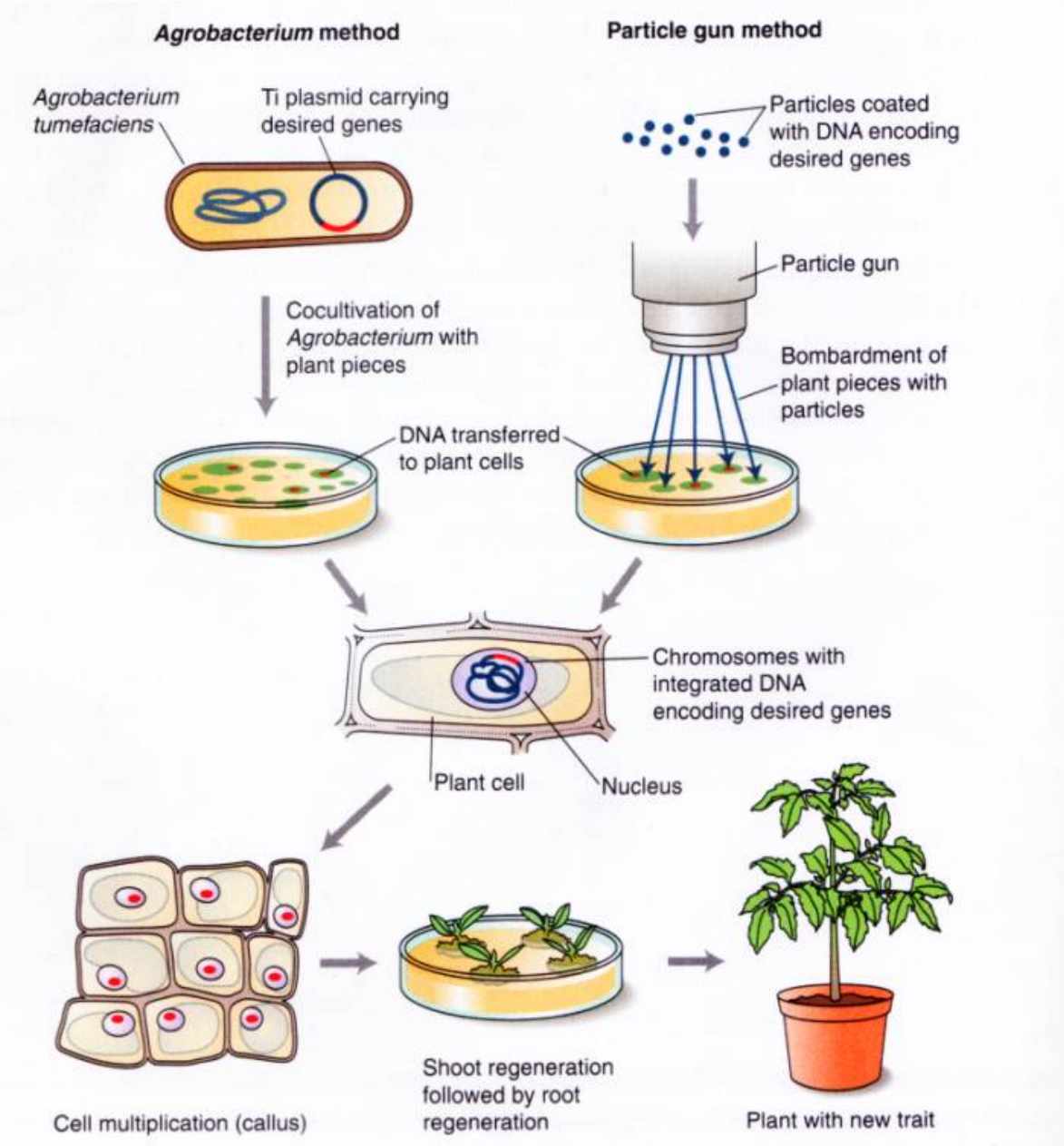
NBTs require the use of "old techniques" used for the transgenesis of already marketed GMOs:

- protoplastisation, vectorization (large proteins contaminated by DNA, remains of genome and *Agrobacterium* plasmid , genome and circular extrachromosomal DNA...), cell cultures, modified cell selection systems, regeneration of non-recalcitrant plants (hence a still limited spectrum of species),
- International meeting held in London in October 2016: laboratories are desperately looking for good, well-trained chefs and regret the lack of schools to train future "chefs"...
- All stressful techniques inducing mutations and epimutations (up to 35% for cell cultures, *somaclonal mutants*):
  - poorly identifiable (reliable software and reference genomes missing) because often point or indel mutations, especially in repeated or non-coding regions, problems of translocations and inversions...
  - phenotype detectable in some specific environment, genome stabilization needing several years,
  - Difficult to eliminate (backcrossing by insufficient firms, co-segregations according to characteristics, regions with non-Mendelian inheritance…) leaving millions of pb not "cleared" and poorly controllable (see the software and reference genomes issues)

As for transgenic GMOs the effect of these mutations and epimutations on the genes, their regulation (i.e. possible toxicity...) and environmental behavior cannot be predicted...

No risk assessment guidelines for e.g. epigenetics,

These signatures are usable to identify the NBT technique used



Agrobacterium:

- still the most efficient delivery system (*A.tumefaciens*)
- phytopathogenic bacterium with many aborted infections (*A.rhizogenes*) as observed in plants (e.g. sweet potato)...
- Numerous plant genomic scars: plasmid, chromosome and circular DNA elements

A regeneration of plants limited to a very few species:

- little hope to enlarge the circle,
- given the costs of development, late returns on investments,
- poor considerations for publications ...

A specific international meeting held in London (Oct. 2016)

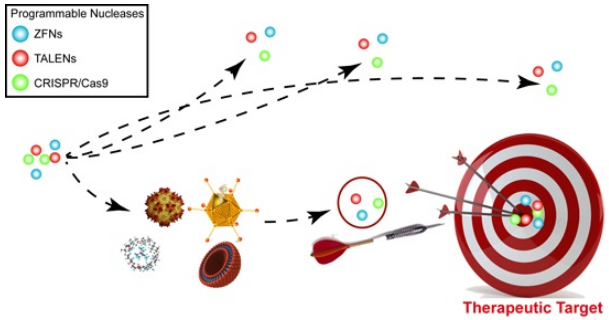
NBT: what are precision, unintended modifications and publications biases?

First training…

Then start ... for missing a lot...

- Many total or partial homologies in the rest of the genome,
- Thermodynamic considerations,
- Many recipes attempting to reduce the number of off-targets (of a **1,500 factor**...)

- As a result: occasional insertions / deletions or not, chromosomal rearrangements (inversions, translocations...), exon skipping effects, epimutations… difficult to predict and detect...
- Additionally: false positive not checked, circular reasoning accepted…



Finally, only presumed successful data are published: the usual **bias of publication**

NBT Signatures

- Scars of *in vitro* related techniques,
- Any genetic change induces epimutations,
- The need of detectable safe harbors, particularly for knockin and gene stacking,
- **PAM and on-target mutations / epimutations vicinity** (several PAM may be involved) e.g. Crispr-endonuclease,
- **Off-targets and PAM vicinities**, similar vicinities of ZFN et TALEN off-targets, seed-regions of RNAi…
- Frequency and types of mutations, such as Crispr-endonucleases due indels (high frequency…)
- **Exon skipping** and e.g. **structural mutations** when using nucleases (numerous errors prone NHEH / MMEJ… DNA reparation systems),
- Insertions of delivery systems (ex: plasmids and genome of *Agrobacterium*) for all SDN, RNAi...
- **Contaminating DNA for SDNs' RNP delivery** (see human genome insertions...) as currently observed even for purified commercial enzymes such as PCR DNA polymerases,
- Spontaneous or induced barcoding by Crispr modifications and stress recording,
- DNA, several kind of RNA, proteins and other signals circulating between rootstock and scion and influencing the detached parts, such as fruits,

As for transgenic GMOs the effect of these off-targets, exon skipping, and other mutations and epimutations on the genes, their regulation (i.e. possible toxicity...) and environmental behavior cannot be predicted...

No risk assessment avalable for e.g. epigenetics, exon skipping, single cell lineages tracking,

These NBT signatures (e.g. Crispr PAM and on- / off-targets sequences) are usable to identify the NBT technique used

“**People just don't have the time to characterize some of the very basic parameters of the system**,” says Bo Huang, a biophysicist at the University of California, San Francisco. “**There is a mentality that as long as it works, we don't have to understand how or why it works.**” (Ledford, Crispr, the disruptor. 2015. Nature)

The limits of back-crosses to discard in the progeny the unintended changes

- The **pangenome is a way to acknowledge the unexpected complexity of genomes**,
- All stresses induce numerous structural mutations (due to e.g. transposable elements mobilization) and epimutations , in genomes and organelles (generally not considered)
- Pre-breeding, insufficient number of backcrosses by companies (in theory ca 95% but…), improving hybrid populations with conserved genetic distances,
- **co-segregations between loci to be conserved and those to be discarded**
- **Linkage disequilibrium and large regions with non-Mendelian inheritance**…
- Pleiotropy, epistasis, rogue off-types, the need of safe-harbors,
- Unexpected inheritable epimutations (to be considered as a source of heterosis?),
- Particular issues with **allopolyploid** genomes, genomes with numerous repeated sequences...
- Elite varieties imposing specific crosses ways, due to their genetic distances (similar as between monkey and man) and hybrid breeding schemes, heterotic groups, population genetic structure,
- **Unavailable for (mandatory or not) vegetatively reproduced plants and perennials / trees**,
- Verifications
  - needing Whole Genome Sequencing and accurate analyses software, trained people and powerful computer on ultra-deep sequencing,
  - Needing reference genomes, missing in most of the cases,
  - Using for cost based reasons, sequencing of predicted off-target with unreliable software,
- **leaving up to several millions of base pairs not "cleared" and poorly controllable** (e.g. wheat)

"In addition, the plant genome is extremely diverse,"

says Jeffrey Sander, scientist at the Pioneer Molecular Engineering (Johnston, Iowa),

“Between two varieties of corn, there is almost the same genetic distance as between man and monkey. “

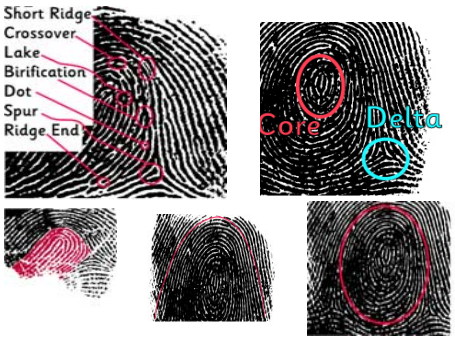
NBT proponents using
biased molecular biology figures of the 70s’,
not representative of what is currently known

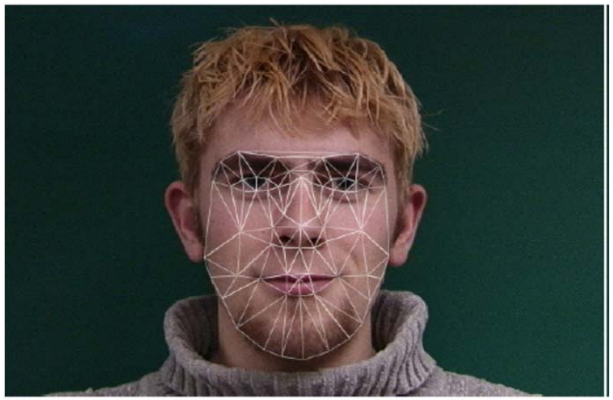
The matrix approach to identify NBT technique then trace products thereof

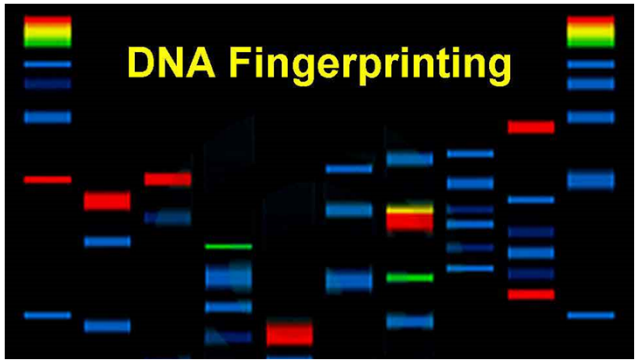
- Gathering different clues to identify an organism by using a pertinent subset of them, using databases, DSS, IA... for helping the analysts,
- Successfully formalized and used for detecting unapproved GMOs,
- The usual job
  - Of a scientist working on polymorphism / genetic diversity
  - of a plant breeder through e.g. genetic maps, Marker Assisted Selection...
  - to protect patented varieties…

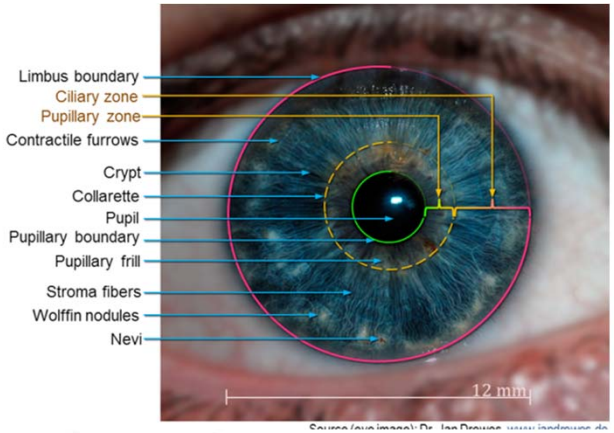
Used in several other identification domains such as…

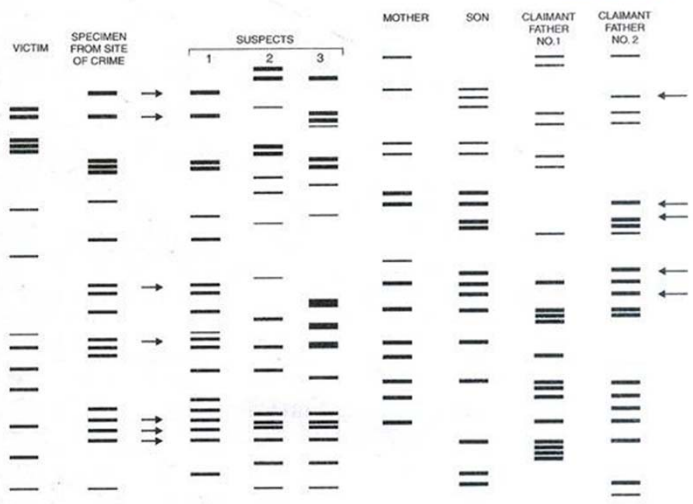
Other examples of the matrix approach







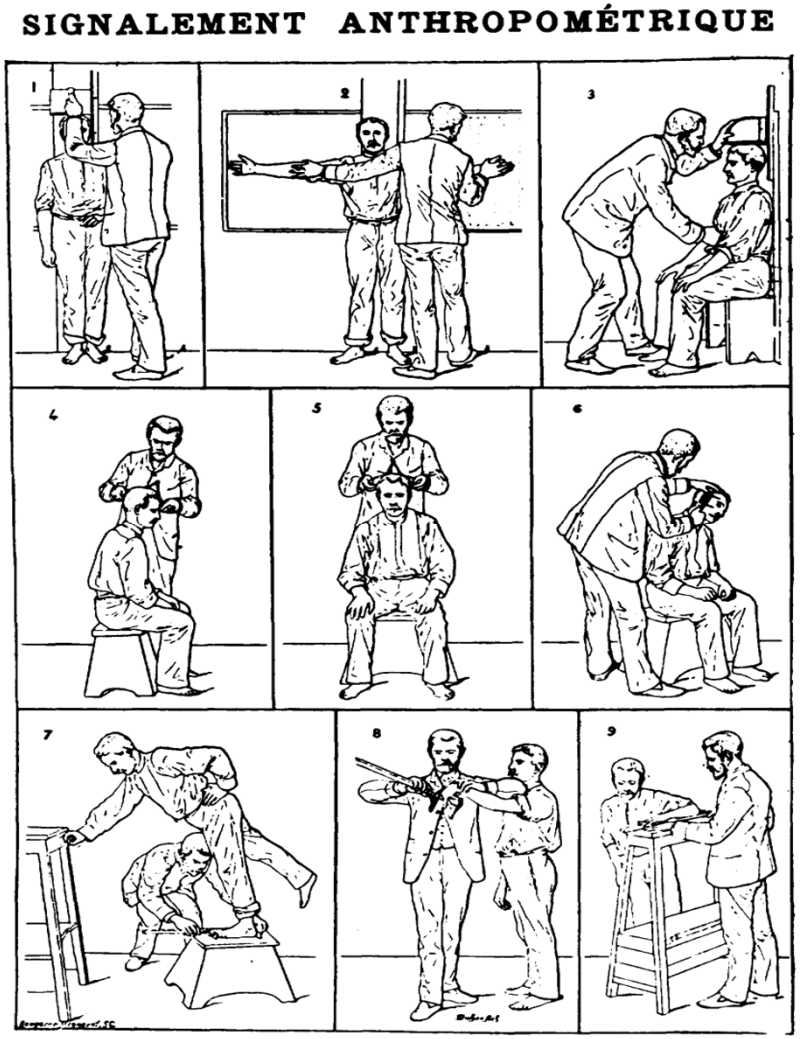




Identification of a criminal through DNA finger printing. Suspect number one is real culprit. as its VNTR bands are matching with specimen from site of crime.

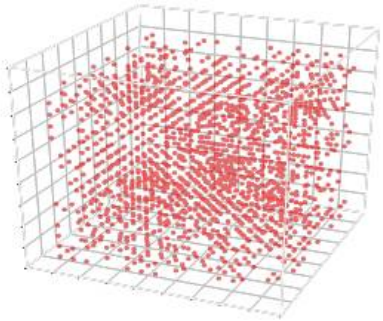
DNA fingerprints of mother under consideration, the child and two claimant fathers. Claimant father No. 2 proved to be the real/biological father.

Alphonse Bertillon's judicial anthropometry

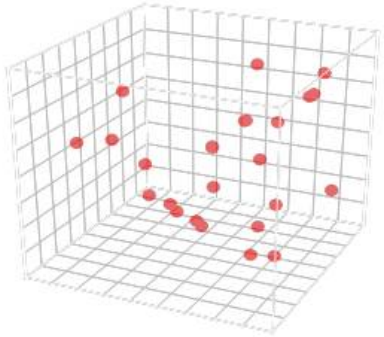


1. Taille. — 2. Envergure. — 3. Buste. —
4. Longueur de la tête. — 5. Largeur de la tête. — 6. Oreille droite. —
7. Pied gauche. — 8. Médius gauche. — 9. Coudée gauche.

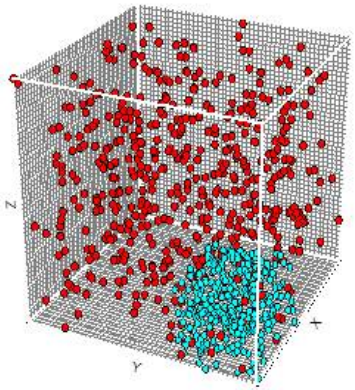
The matrix approach to identify the NBT techniques initially used and the derived products is based on the assemblies of markers of different types, for example in genomes and epigenomes



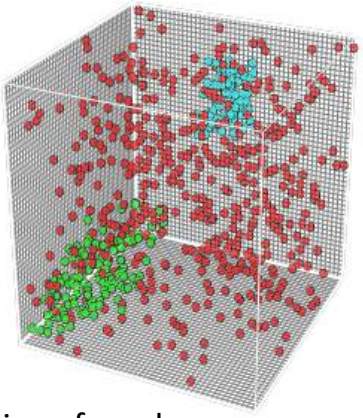
Markers in all genome and epigenomes (ex: used for a MAS, PAM, off-targets, translocation, transversion ... frequency, cartography ...)



Choice of identification markers of the species



Choice of markers differentiating products from *in vitro* techniques



Choice of markers differentiating one (of) technique(s), ex. Crispr-endonuclease (s) natural mutation (s)

- Choice of a combination for unambiguous legal identifications (subsets used for MAS…)
- Choice of one (or some) relevant marker (eg PCR on targeted mutation-PAM) for routine detections (aspects of cost, speed ...)

Do not be fooled by the tree that hides the forest…

- The matrix approach = biometrics = "weight of evidence" of EFSA's risk assessment:
  - Let certify that the targeted changes are not natural,
  - Allow to identify the NBT used.

Messages to take home

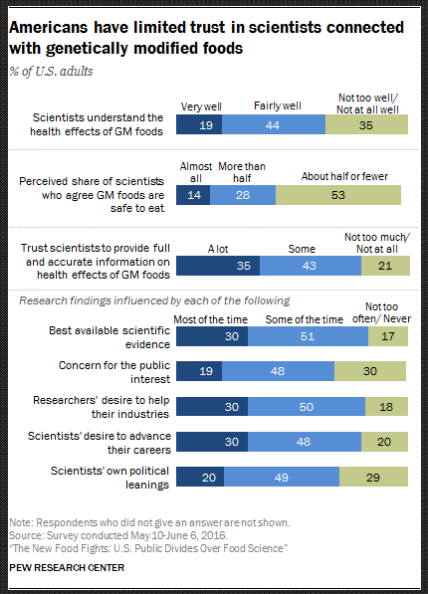
- **Worldwide traceability** and labeling of NBT products is a inexpensive tool depending only on the wishes of consumers and willingness of retailers
- There are **no technical obstacles to distinguish**:
  - *In vitro* from *in vivo* mutated organisms,
  - Induced from spontaneously appeared mutations in normal environment,
  - Identify NBT used for producing any product,
  - Detect and trace the derived product,
- The several genomic and epigenomic markers can be combined according to the aims of the tracking (numerous statistical tools, models, DSS, databases available).

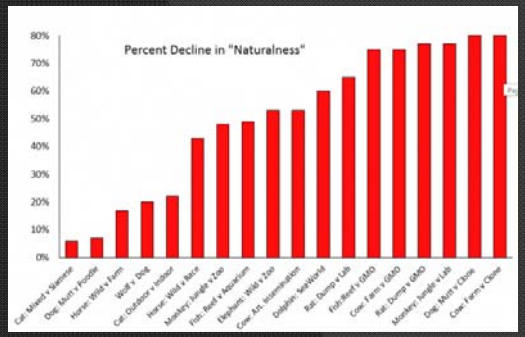
Defining the different markers to be used to identify a species, mutant, a cultivar or a NBT derived product is the usual job of a scientist working on genetic diversity, and thus of a plant breeder

Rhetoric and communication: how language matters to get citizens' consent

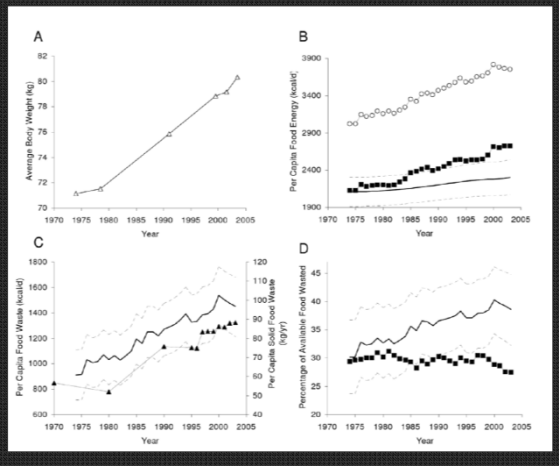
- Polls in France and USA: confidence in science but rather less on scientists, thus a communication referring only to science,
- The direct impact of using words and expressions on the reasoning and acceptance of both scientists and laypersons (e.g. "*new breeding techniques*"),
- Outrageously simplified language, omitted 'details' and metaphors (e.g. genome editing) to build up a politically correct language and change the perceptions, thinking and furthermore acting ways,
- Antiphons using 'nature', 'natural', 'traditional', 'conventional'… to facilitate the acceptance of the techniques and products…
- Feeding / nurturing the world by increasing the production and the wastes…

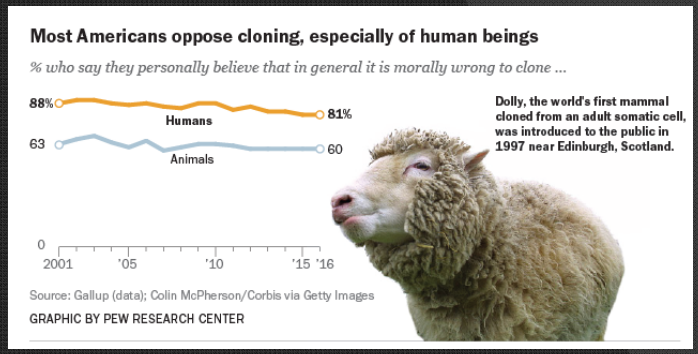
« Mal nommer un objet, c'est participer au malheur de ce monde… » (Albert Camus, 1944, Poésie 44)





The Meaning of "Natural". Process more important than content . Paul Rozin. Psychological science. 2005.





Some misleading metaphores

(which influence your perception and then your ways of thinking and finally of acting)

SDN: do not foresee an unique and precise cutting

But a series of cuttings (with numerous breaks to rapidly and accurately stick)

Editing the genome … Waiting for amending electronic and known languages?

What you have effectively to “edit”: untranslated handwritten languages…

plus

The promised precise modification?

Several 'off targets' obtained due to rebound effects from homologous sequences

On a destructed landscape due to the related techniques: everything to rebuild

It’s rather Staline’s organ shots

Targeted mutagenesis: were you thinking about a ‘one shot’?

Conclusion 1

- As for transgenic GMOs, the effect of the mutations and epimutations due to related techniques and NBT are unpredictable,
- there are no technical obstacles to
  - Differentiate products having had *in vitro* steps from only *in vivo* resulting products,
  - identify the techniques which were used to build a NBT product which can furthermore be traced and detected,
  - label NBT products which is only depending on a political wish / decision (see ENGL proposal to the EC to work on in 2013 …)
- Mutations and epimutations resulting from NBT techniques use cannot be stated as natural unless
  - accepting that a tree hides the forest (see above all “side effects”)…
  - Or you agree that the current climatic change is “natural”…

Conclusion 2/2

According to the observed facts:

- the **NBT products should be encompassed by the GMOs legislation / regulation** (due to somaclonal variation, unintended changes in several genomic and epigenomic parts of nucleus and organelles, numerus gaps in our knowledge about genetics, epigenetics and epitranscriptomics…) see e.g. https://www.efsa.europa.eu/en/supporting/pub/1129e and https://www.youtube.com/watch?v=TcD2HILOLXc
- but moreover, the **risk assessment guidelines of GMOs should be upgraded** (traces of vectors partly searched for, WGES (Whole Genome and Epigenome Sequencing) made mandatory, epigenetic and epitranscriptomic guidelines to be rapidly put in place (see EFSA 2016 meeting on the gap of knowledges on epigenetics), exon skipping to be checked, results of the unexpected very late EFSA call for proposals on RNA interference to be rapidly integrated…) as the knowledge about - and interactions between - genomes / epigenomes and epitranscriptome is very incomplete (see the “junk DNA” as called a few years ago while currently perceived as very important in organisms’ adaptation, with inheritable traits and more generally speaking gene expression.
- socio-economic and political considerations (such as observed for instance in the 2016 French OPECST hearings on Crispr-endonucleases) should not hide behind pseudo-scientific expertise…