The Digital Genebank

Access to Diversity through Access to Information

Theo van Hintum, Centre for Genetic Resources, The Netherlands (CGN)

Meeting 'Commissie Onderzoek Voedingstuinbouw Zaden'
April 26th, 2016, De Lier

The Digital Genebank

- three points will be made
  - the potential of crop diversity has so far been underutilised
    - most of it was hidden in (exotic) genebank accessions
  - genomics and other high-throughput screening techniques allow exploitation of this potential
    - increasingly available and affordable
    - happening as-we-speak
  - remaining weak spot: the connection of different data types to each other and to the germplasm
    - access to, and analysis of newly generated data
    - translation of findings into advise to breeders

... and the Digital Genebank will be introduced

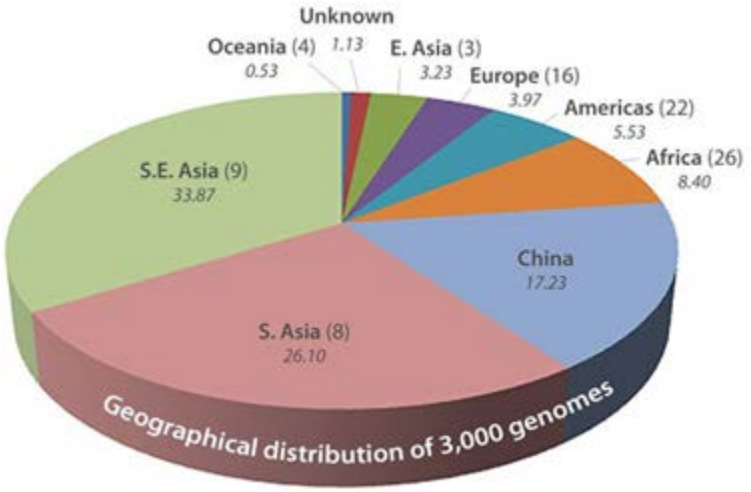
The Digital Genebank

- the potential of crop diversity has so far been underutilised
  - the use of crop wild relatives so far limited to qualitative easy observable traits
    - disease resistances
  - Tanksley *et al*. showed that exotics can provide QTLs for advanced material
    - cumbersome procedure
    - narrow genetic basis (one exotic parent)

(Source: Tanksley & McCouch 1997)

The Digital Genebank

- genomics and other high-throughput screening techniques allow exploitation of the potential
  - genebank material is being (re-)sequenced at a high rate
    - 3,000 Rice Genomes Project (3K RGP) has sequenced 3,024 rice varieties from 89 countries – and continues
      average sequencing depth of 14×
      average genome coverages of 94.0%
      18.9 million SNP's (ref Nipponbare)



The Digital Genebank

- genomics and other high-throughput screening techniques allow exploitation of the potential
  - genebank material is being (re-)sequenced at a high rate
    - CGN's collections are being sequenced
      - 150 Tomato Genome ReSequencing project
      - 100 Melon Genome project
      - International *Lactuca* Genomics Consortium: exome sequencing 500 CGN *Lactuca* accessions
    - initial steps towards metabolomic analysis of CGN material
      - metabolite composition of 150 ILGC samples both targeted and untargeted by ZonMw 'Plant Metabolomics Hotel' project

The Digital Genebank

- genomics and other high-throughput screening techniques allow exploitation of the potential
  - genebank users are increasingly using marker-assisted selection (MAS) and genomic selection (GS)
    - need for access to all available molecular data + phenotypic data + any other relevant data
    - need for selection tools that allow browsing, selection, combination and analysis of these data
    - access and tools are not available yet because the data are
      - not sufficiently annotated
      - not connected
      - not semantically annotated
      - tools do not exist

The Digital Genebank

The International Center for Tropical Agriculture in Colombia holds 65,000 crop samples from 141 countries.

Feeding the future

We must mine the biodiversity in seed banks to help to overcome food shortages, urge **Susan McCouch** and colleagues.

Humanity depends on fewer than a dozen of the approximately 300,000 species of flowering plants for 80% of its caloric intake. And we capitalize on only a fraction of the genetic diversity that resides within each of these species. This is not enough to support our food system in the future. Food availability must double in the next 25 years to keep pace with population and income growth around the world. Already, food-production systems are precarious in the face of intensifying demand, climate change, soil degradation and water and land shortages.

Farmers have saved the seeds of hundreds of crop species and hundreds of thousands of 'primitive' varieties (local domesticates called landraces), as well as the wild relatives of crop species and modern varieties no longer in use. These are stored in more than 1,700 gene banks worldwide. Maintaining the 11 international gene-bank collections alone costs about US$18 million a year.

The biodiversity stored in gene banks fuels advances in plant breeding, generates billions of dollars in profits, and saves many lives. For example, crossbreeding a single wild species of rice, *Oryza nivara*, which was found after screening more than 6,000 seed-bank accessions, has provided protection against grassy stunt virus disease in almost all tropical rice varieties in Asia for the past 36 years[1]. During the green revolution, high-yielding rice and wheat varieties turned India into a net food exporter. By 1997, the world economy had accrued annual benefits of approximately $115 billion from the use of crop wild relatives[2] as sources of environmental resilience and resistance to pests and diseases.

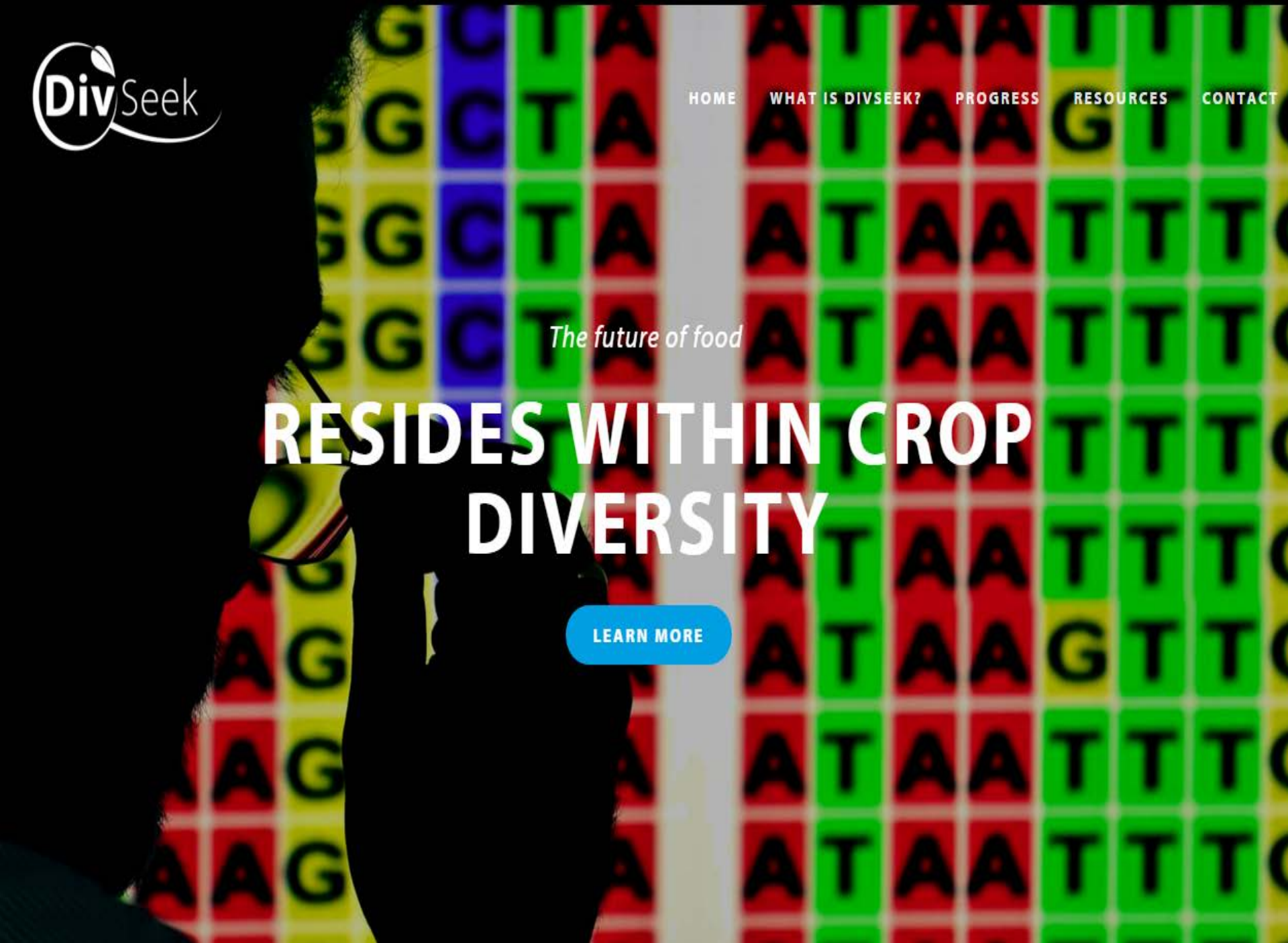
The time is ripe for an effort to harness the full power of biodiversity to feed the world. Plant scientists must efficiently and systematically domesticate new crops and increase the productivity and sustainability of current crop-production systems.

Why does plant breeding need a boost? Because new, high-yielding seeds that are adapted for future conditions are a cornerstone of sustainable, intensified food production[3]. Since the mid-1990s, progress in conventional plant breeding has ▶

The Digital Genebank

- bridge the gap between genebanks and -omics community
  - addressed by DivSeek

> Using the emerging deluge of omics data along with mathematical models and systems biology approaches, we can now ... reveal the many ways ... how humans have shaped the genomes of crop plants through domestication and breeding. This information will provide unprecedented insights into the intricate and finely tuned genetic networks that enable plants to respond to subtle or catastrophic perturbations in the environment. It will also present novel solutions and ingenious templates that can serve as a blueprint for accelerating plant breeding to provide sustainable food security for humanity.
>
> McCouch et al. (2013) Feeding the future. *Nature* 499:23-24.


DivSeek
HOME
WHAT IS DIVSEEK?
PROGRESS
RESOURCES
CONTACT
The future of food
RESIDES WITHIN CROP DIVERSITY
LEARN MORE

DivSeek will help to bridge the gap between the information requirements of genebank curators, plant breeders and more targeted upstream biological researchers, to support applied germplasm curation, forward-looking breeding programs and strategic research.

k the potential of crop
ebanks around the
ilable to all so that it
ance the productivity
ience of crops and

DivSeek will help to bridge the gap between the information requirements of genebank curators, plant breeders and more targeted upstream biological researchers, to support applied germplasm curation, forward-looking breeding programs and strategic research.

"

DivSeek will focus a powerful beam of light into the depths of gene pools essential for human survival.

The DivSeek initiative will work with existing, emerging and future initiatives to characterize crop diversity and develop a unified, coordinated and cohesive

The Digital Genebank

- principles CGN Digital Genebank
  - digital bridge between genebank users and genebank data
    - –omics data + classical data
  - developed to be generalised
    - functionalities defined by users
    - functionalities created based on existing software
      - new code will be open source
  - functional prototype developed in 2017-2019
    - based on tomato data (available) and lettuce data (now generated)

The Digital Genebank

- timeliness
  - wide support for DivSeek
    - need for solution in genebank community
    - need to use opportunities in breeding community
  - passport and some phenotypic data are available at CGN
  - sequence data and metabolomics data are increasingly available to CGN
  - technology is largely available in WUR-PSG, some needs to be developed or applied to the Digital Genebank

The Digital Genebank

- feasibility
  - data
    - all data types for lettuce and tomato cases, and thus the Digital Genebank, are under WUR control
    - links to related US initiative are good – data will be public
  - technology
    - the required technology (largely) available, rest can be developed at WUR
  - functionallity
    - needs to be defined in contact with the users

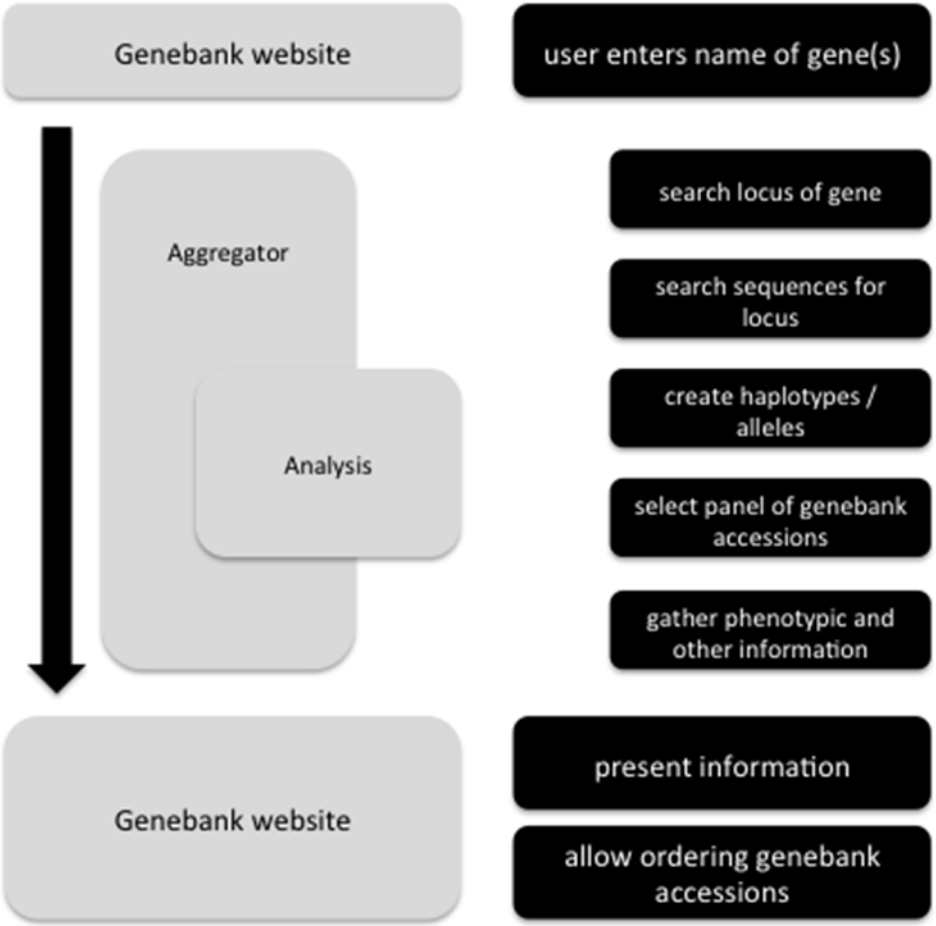
The Digital Genebank

- approach
  - make data sources available
    - use lettuce and tomato as model
    - use FAIR principles (Findable, Accessible, Interoperable, and Re-usable) - to allow generalisation
  - interact with future users to define functionalities
    - breeders (large, medium, small)
    - crop researchers
    - genebank curators
  - develop user-interfaces and underlying software
    - open source algorithms and code - to allow generalisation
    - testing during development by future users

The Digital Genebank

- content
  - to be determined by potential users
  - functionalities for high-end users
    - access to -omics data linked to CGN material
      - similar to the ‘Sol Genomics Network’ or ‘Compositdb’
      - combining / linking the resources
      - downloads, browsers, query interfaces
  - functionalities for low-end users
    - query interface
      - for example: enter name of gene and get panel of CGN accessions with all alleles for that gene

The Digital Genebank



(Source: Finkers R, Chibon PY, van Treuren R, Visser R, van Hintum TJL (2015) Genebanks and genomics: how to interconnect data from both communities? Plant Gen Res: Char & Util 13:90-93)

The Digital Genebank

- why CGN ?
  - CGN is high quality genebank (material, procedures, services)
    - first genebank with ISO 9001:2008 QMS
  - CGN has possibly the best documentation system and public user-interface (web site)
    - all data searchable and downloadable, click-wrap SMTA, on-line core-selector
  - CGN functions in high-tech WUR and NL-breeding environment
    - collaboration in several sequencing projects and PPP's
  - CGN has strong position in genebank documentation community
    - chair of ECPGR Doc/Info WG, member Genesys AC, chair EURISCO AC, co-chair ITPGRFA WG on the Global Information System
    - active participant DivSeek

The Digital Genebank

- ambition
  - provide best possible service to CGN-users
    - create mechanism for mining –omics and other data sets related to CGN material
      - add value to newly generated ILGC datasets
  - take initiative in ‘DivSeek’ related activities
    - CGN becomes leader in the field
      - partner companies sit in first row
    - CGN/WUR becomes/stays partner of choice in new projects

Thank you for your attention !

Genetic resources are useful, pretty and tasteful

Genetic resources are the food on your plate

Thank you for your attention !