# Diversity Trees:

## what are they, why do we need them and how can they be implemented ?

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

Annual Genebank Meeting, Sept. 2017, Meise, Belgium

# Diversity Trees – tools to describe diversity

- why do we need tools to describe and optimise the structure of PGR collections?
    - the importance of conscious collection composition
- what are diversity trees?
- how can diversity trees be implemented?
    - experience in CGN and in the CGIAR

# Diversity Trees – tools to describe diversity

- a genebank / the world has a limited capacity
  - this capacity should be used optimally to best achieve goals
    - conserve diversity for the future
    - provide access to diversity for use

- options for opmisation
  - reduction of costs of regular procedures
  - optimising the composition of the collection
    - gap / redundancy analysis

# Diversity Trees – tools to describe diversity

- a genebank can not 'just throw away material'
  - preserve material in perpetuity
    - genebanks should be able to rely on each other
    - if genebank A does it well, genebank B doesn't have to do it

what does this mean ?

  - optimisation of collections is relevant when
    - removing duplicates / reducing redundancies
    - acquiring the right new material

# Diversity Trees – tools to describe diversity

- origin of genebank collections
  - collecting missions to specific regions
  - exchanges between genebanks
  - donations from a breeding programme consisting of collected and bred material
  - advanced varieties received from breeding companies
  - various other sources incl. material donated by individuals / scientists

# Diversity Trees – tools to describe diversity

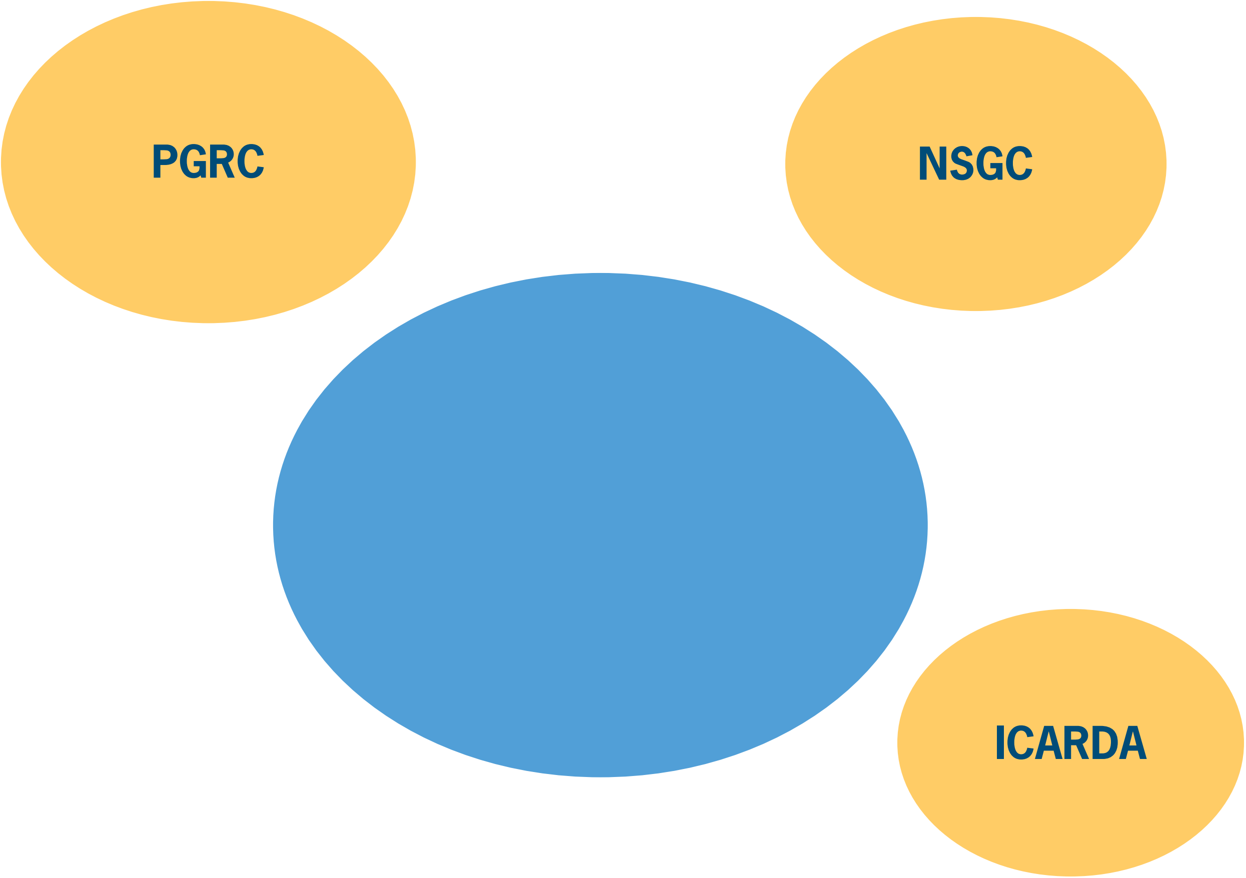
- origin of genebank collections
  - genebank collections are the results of undirected events
    - never: dedicated analysis of desired diversity followed by a targetted aquisition strategy to get the material needed
  - as a result - most genebank collections are neither balanced nor complete
    - a lot can be gained by improving this situation
  - the ‘world holding’ is even more imbalanced but slightly more complete

# Diversity Trees – tools to describe diversity

- origin of the world’s *ex situ* holding
  - genebanks often have material in common
    - as a result of exchange
  - most genebanks have some unique material

  - combined PGR collections will have
    - better coverage of the genepool
      - more total diversity
    - higher redundancy due to duplication

# Diversity Trees – tools to describe diversity

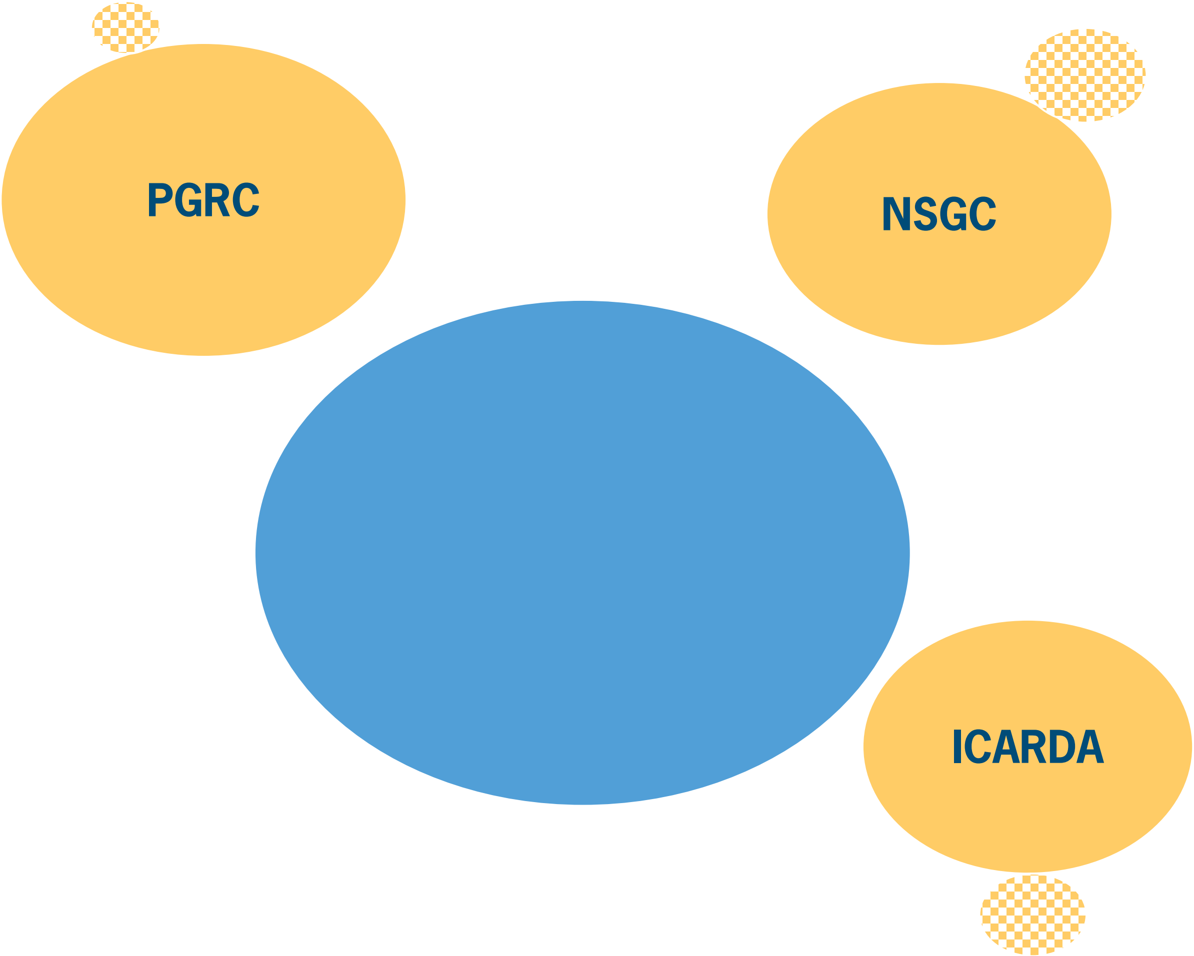


- accessions of cultivated barley in 3 largest collections

Source: Hintum & Menting 2003

Centre for Genetic Resources, the Netherlands

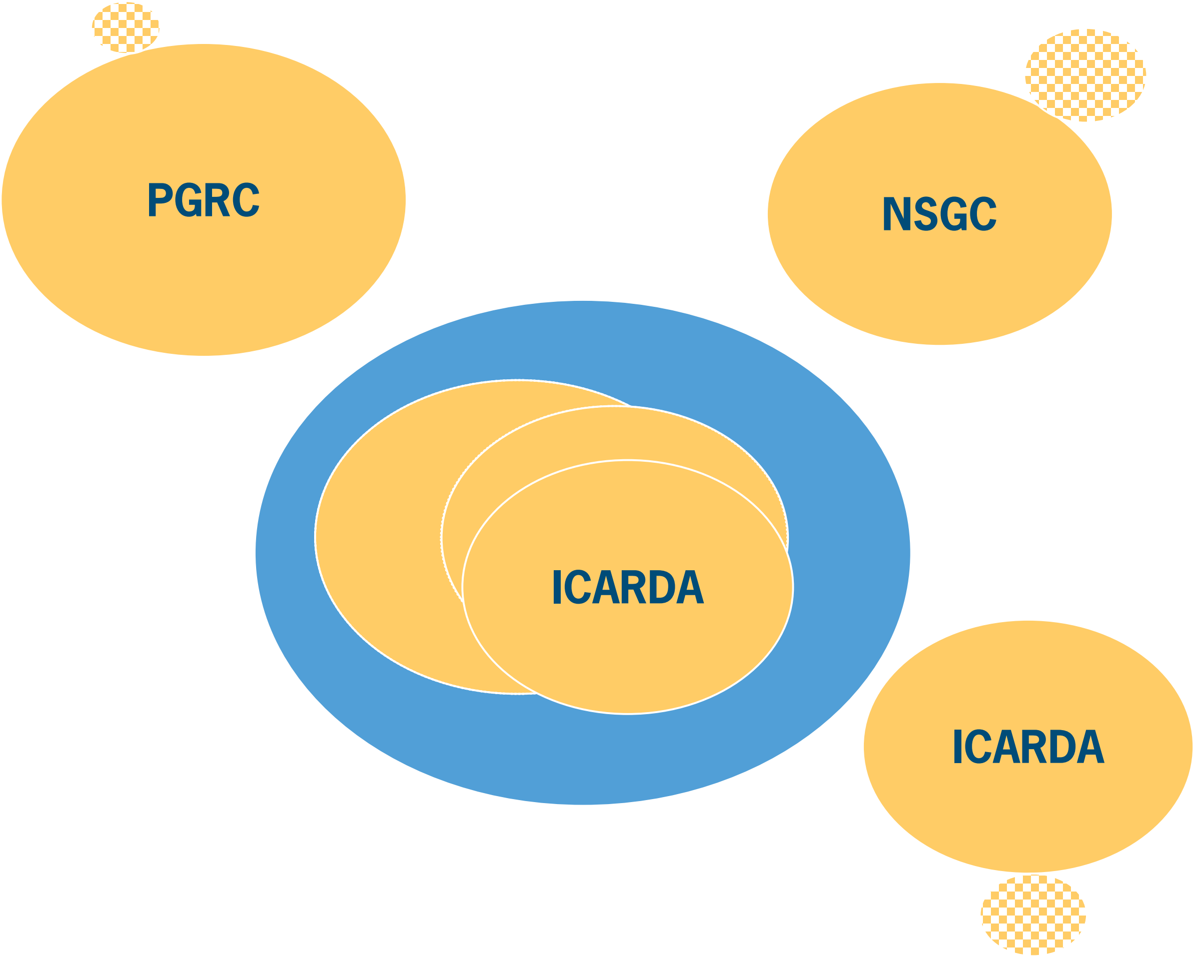
# Diversity Trees – tools to describe diversity



- accessions of cultivated barley in 3 largest collections
- internal redundancy

Source: Hintum & Menting 2003

# Diversity Trees – tools to describe diversity



- accessions of cultivated barley in 3 largest collections
- internal redundancy
- overlap: 46% redundant accessions

Source: Hintum & Menting 2003

Centre for Genetic Resources, the Netherlands

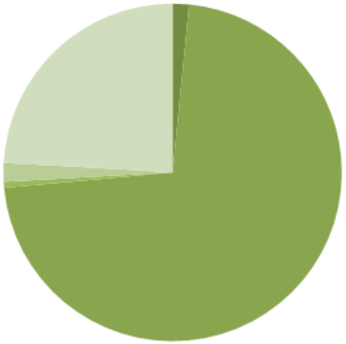
# Diversity Trees – tools to describe diversity

- two aspects of optimisation collection composition
  - reflect importance of material for use
    - composition is function of importance
      - arbitrary
      - visionary
  - diversity is maximised
    - composition contains minimal redundancy
      - no duplication
      - core methodology for maximization of diversity

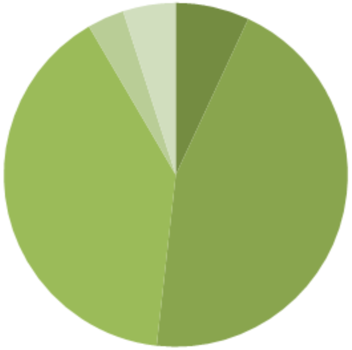
# Diversity Trees – tools to describe diversity

- optimal collection will reflect importance of material for current and future use
  - ease of breeding
    - phenotyping is easier if distance to crop is small
      - some exotic material requires test crosses
    - breeding is easier if genetic distance is small
      - marker assisted breeding increased the popularity of crop wild relatives
  - diversity in genepools
    - diversity is wider if distance to crop is large
      - all CWR combined will have higher diversity than the primary genepool

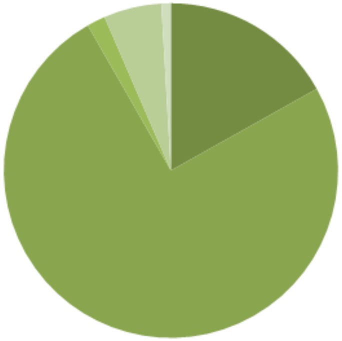
# Diversity Trees – tools to describe diversity



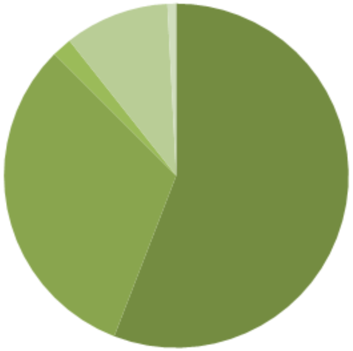
finger millet
n=8916 (100% of accs in Genesys)



rice
n=145598 (69% of accs in Genesys)



beans
n=73782 (98% of accs in Genesys)



lettuce
n=9896 (91% of accs in Genesys)

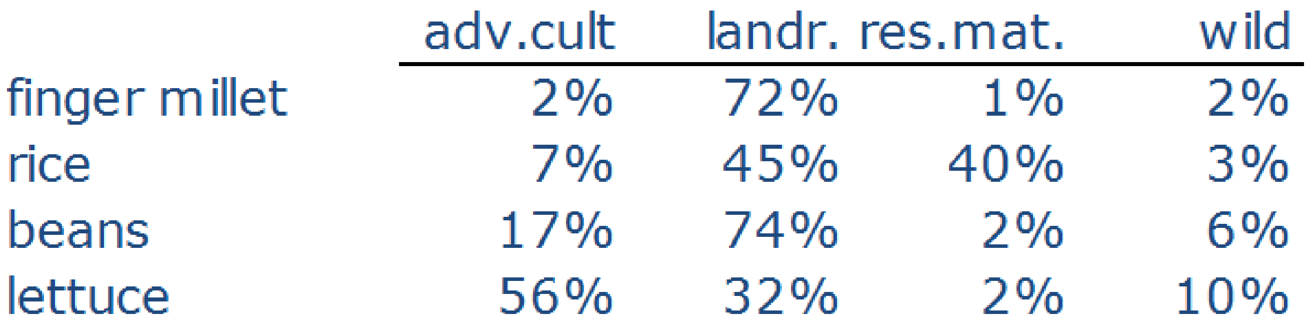
🟩 Advanced/improved cultivar 🟩 Traditional cultivar/Landrace 🟩 Breeding/Research Material 🟩 Wild/Weedy 🟩 Other

data downloaded from Genesys 22/3/2017
Centre for Genetic Resources, the Netherlands

WAGENINGEN
UNIVERSITY & RESEARCH

# Diversity Trees – tools to describe diversity

- composition of collections differ strongly



| | adv.cult | landr. | res.mat. | wild |
|---|---|---|---|---|
| finger millet | 2% | 72% | 1% | 2% |
| rice | 7% | 45% | 40% | 3% |
| beans | 17% | 74% | 2% | 6% |
| lettuce | 56% | 32% | 2% | 10% |

(duplication is not visible in the numbers)

data downloaded from Genesys 22/3/2017

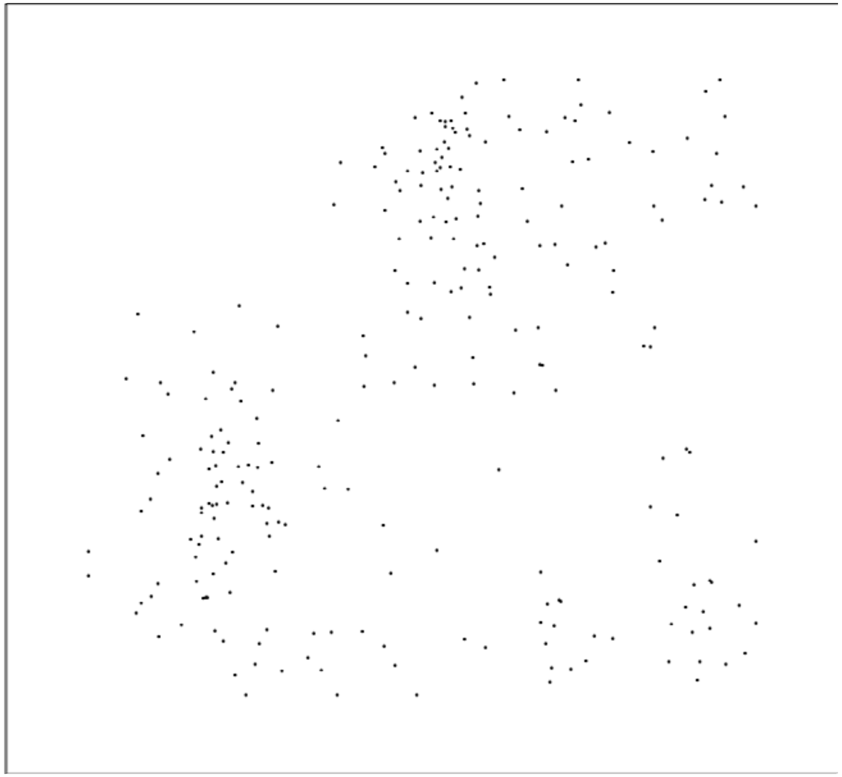
Centre for Genetic Resources, the Netherlands

# Diversity Trees – tools to describe diversity

- optimal collection will have a minimum of redundancy
  - avoiding duplication
    - easier said than done
  - maximising diversity within groups
    - core collection methodologies can be applied

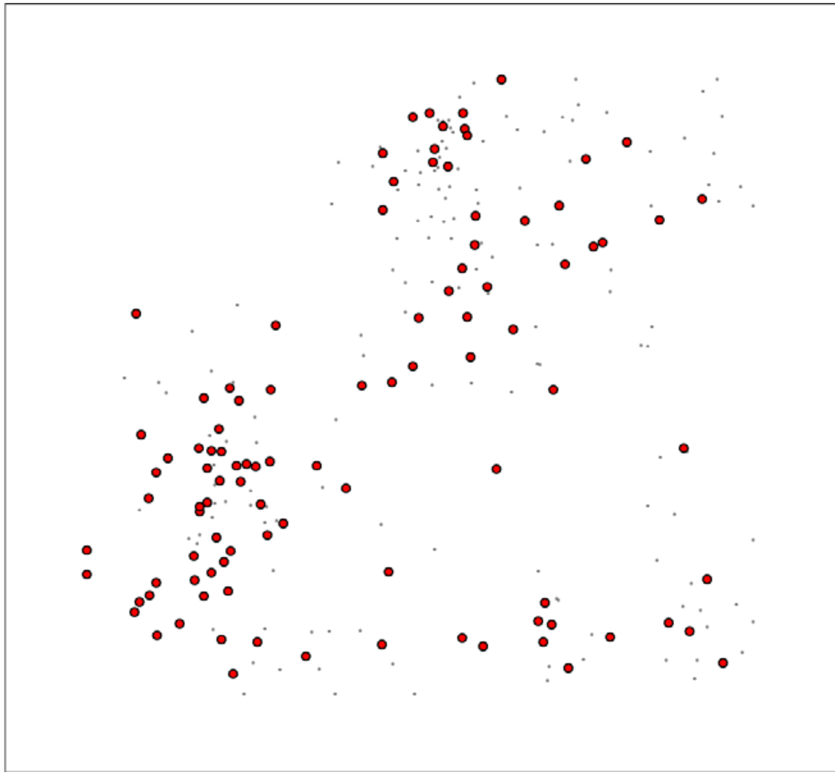
data downloaded from Genesys 22/3/2017

# PGR Collection Composition

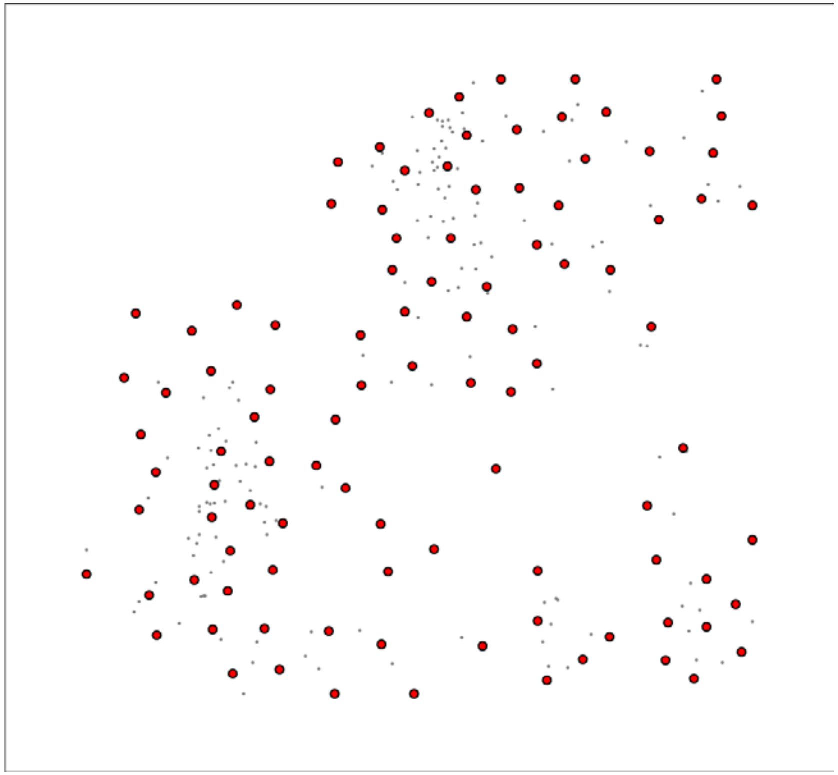


250 accessions

# PGR Collection Composition



40% random selection



40% core selection

# Diversity Trees – tools to describe diversity

in conclusion

- composition of collections is determined by
  - importance of material for current (and future) use
  - efforts to avoid redundancy and maximise diversity
  - availability of material
    - pure chance: focus of past breeding or research programs, collaboration with other institutes, opportunities for collecting, etc.

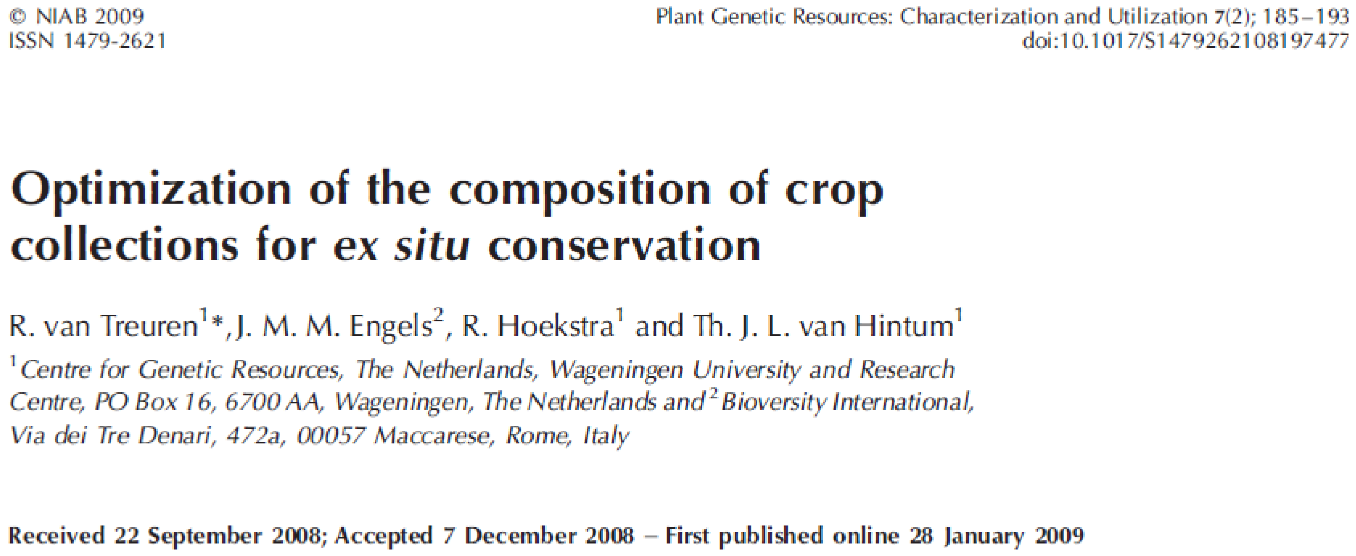
# Diversity Trees – tools to describe diversity

## 2 questions for a PGR curator:

- what do you know about the composition of your collection?
    - proportion wild vs. cultivated, various crop types, various origin continents/countries?
    - redundancies in specific groups?
- do you have clear objectives in terms of what should be added (or should be rationalized)?

# Diversity Trees – tools to describe diversity

- diversity tree
  - tool to describe diversity in a PGR collection
  - tool to define optimal collection given available capacity



© NIAB 2009
ISSN 1479-2621

Plant Genetic Resources: Characterization and Utilization 7(2); 185–193
doi:10.1017/S1479262108197477

**Optimization of the composition of crop collections for ex *situ* conservation**

R. van Treuren[1]*, J. M. M. Engels[2], R. Hoekstra[1] and Th. J. L. van Hintum[1]

*[1]Centre for Genetic Resources, The Netherlands, Wageningen University and Research Centre, PO Box 16, 6700 AA, Wageningen, The Netherlands and [2]Bioversity International, Via dei Tre Denari, 472a, 00057 Maccarese, Rome, Italy*

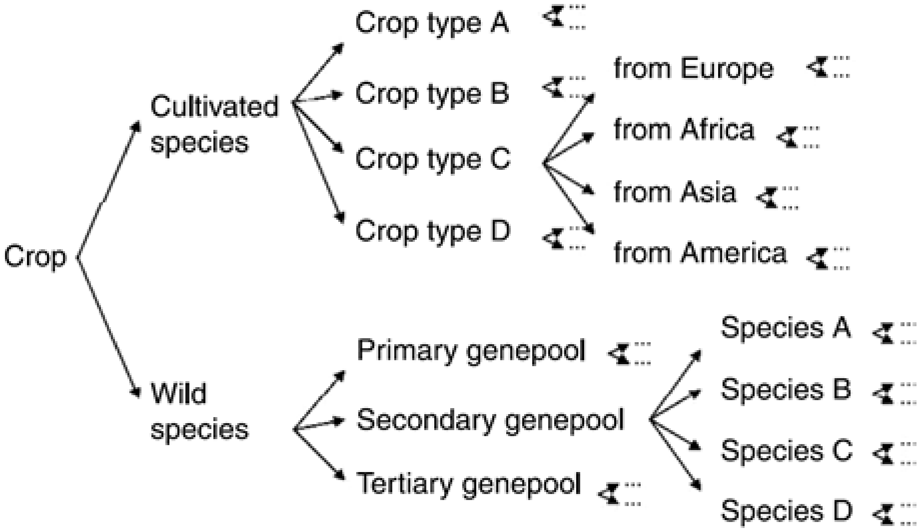
**Received 22 September 2008; Accepted 7 December 2008 – First published online 28 January 2009**

# Diversity Trees – tools to describe diversity

- approach for construction diversity trees
  - determine structure of genetic diversity of crop
    - define stepwise division of group in sub-groups
      - hierarchical structure
    - undividable group = end-group
  - determine relative importance of sub-groups
    - stepwise process
    - arbitrary decisions
      - based on importance for breeding
      - available diversity

# Diversity Trees – tools to describe diversity

- structure of genetic diversity



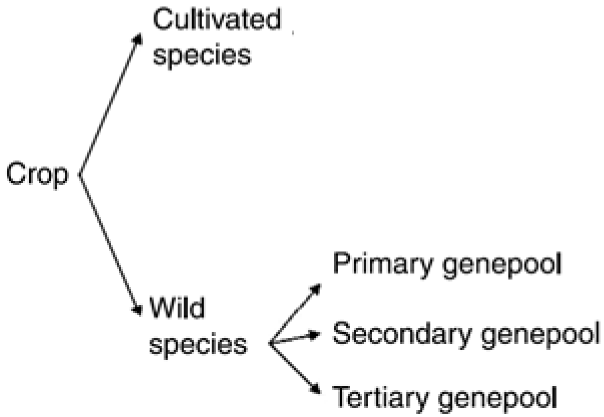
  - examples end-groups
    - subgroup ‘Sucrier’ of the AA diploids of the cultivated material in the section Eumusa of banana’s
    - Bolivian cultivated triploid`potato of the taxon *S. tuberosum* subsp. *andigenum*

# Diversity Trees – tools to describe diversity

- structure of genetic diversity
  - divisions should be meaningful
    - quantitative approach doesn’t result in meaningful divisions
    - tools to find best divisions could be developed
      - find division based on passport/phenotypic data with highest $F_s$
  - based on expert knowledge
  - always based on the question “*how can we divide this group in meaningful subgroups that are as different as possible*”
    - example: the ‘wild rice from the primary genepool’ can be divided in African, Asian, American and Australian species

# Diversity Trees – tools to describe diversity

- relative importance of end-groups
  - examples stepwise approach
    - the cultivarted species should have four times more accessions than the wild species: 80% and 20%
    - primary vs secundary vs tertiatry genepools of the wild species should have the ratios 60%, 30% and 10%



2% of the accessions will be from the tertiary genepool

# Diversity Trees – tools to describe diversity

- relative importance of end-groups
  - always based on the question "*how important are these sub-groups relative to each other in terms of diversity and relevance to future breeding*"
    - very difficult question
    - the basic quastion for any genebank curator composing her collection

# Diversity Trees – tools to describe diversity

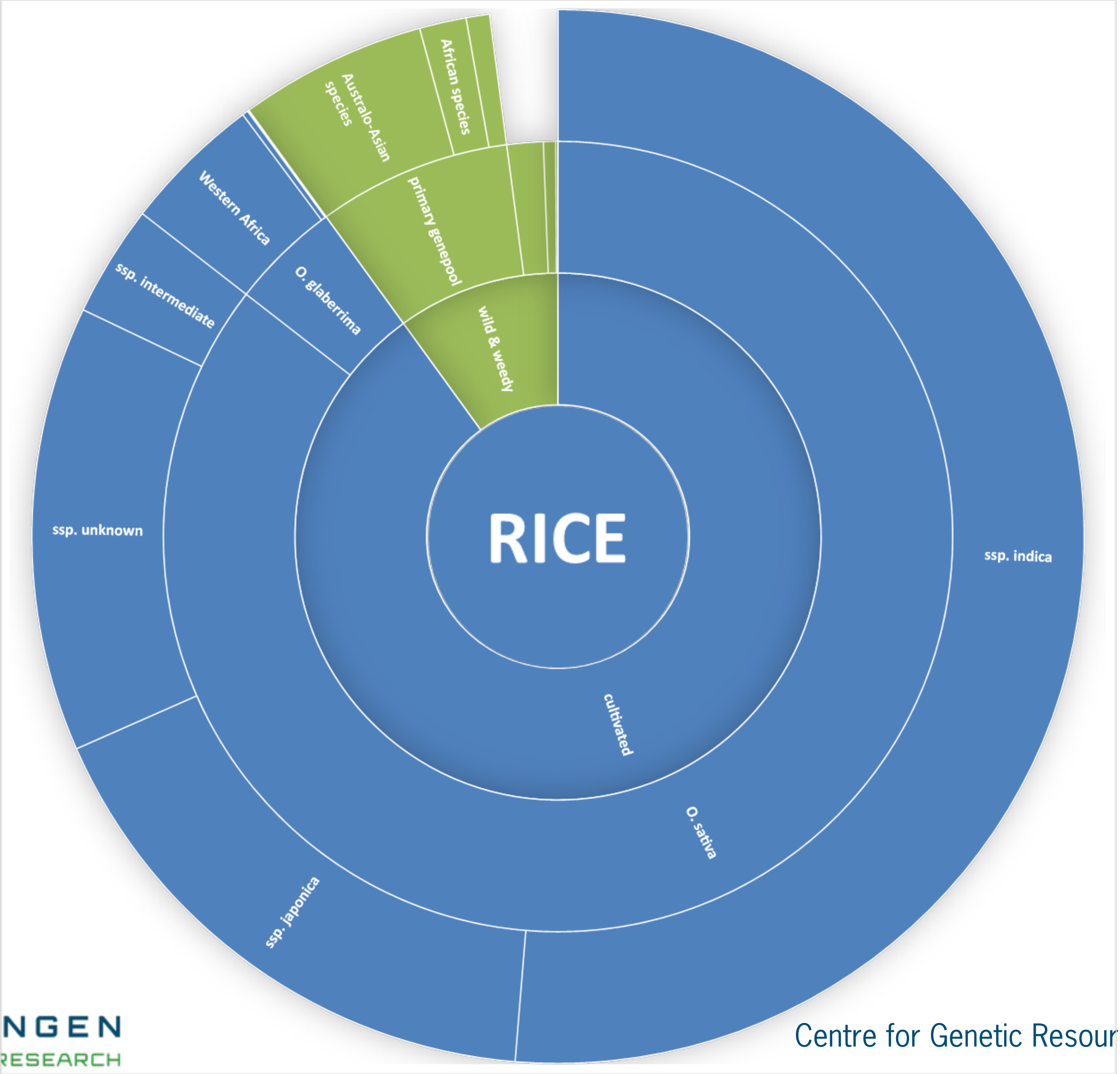
- implementation in CGIAR Genebank Platform
  - make manual and Excel-tool for creating diversity trees
  - for selected crops:
    1. use bilateral contacts with expert to explain the initiative
    2. expert provides overview structure genepool
    3. draft tree is made based on this structure and literature
    4. expert improves draft tree
    5. expert suggests other experts to ask for feed-back
    6. other experts provide feedback on topology and weights
    7. consensus tree is created
    8. collections are mapped on the tree
    9. gaps are identified and priorities are defined

# Diversity Trees – tools to describe diversity

- rice
  - curator IRRI collection (RSH) was approached
  - structure genepool was represented by a table with species / variety group / origin country / origin province / number of accessions
  - tree was created and improved by curator
    - 45 divided groups & 162 end-groups
    - largest end-group 14.4%
  - overview for experts was created
  - next steps: identification of experts and request for feedback

# Diversity Trees – tools to describe diversity

- rice



# Diversity Trees – tools to describe diversity

- banana
  - experts Bioversity (MR, JS, MR) were approached
  - structure genepool was presented in the taxonomy browser of the Musa Germplasm Information System

# Diversity Trees – tools to describe diversity

- banana
  - tree was created and improved by experts (with additional experts NR and IvdH)
    - 23 divided groups & 100 end-groups
    - largest end-group with 8.0%
  - overview for experts was created
  - experts with their expertise were identified and approached
  - constructive feedback from experts is coming in
  - next steps: creation consensus tree, mapping collection on tree, analysis

# Diversity Trees – tools to describe diversity

- potato
  - curator CIP collection (DE) was approached
  - Skype conversation with CIP staff was held
  - draft tree was created based on CGN potato tree, CIP expert (RG) improved it – added the cultivated part of the tree
    - still some issues regarding topology
    - 61 divided groups & 261 end-groups
    - largest end-group 5.9%
  - overview for experts was created
  - next steps: identification of experts and request for feedback

# Diversity Trees – tools to describe diversity

- barley
  - no experts approached – draft tree in-house
  - draft tree was created based on CGN barley tree
    - 25 divided groups & 102 end-groups
    - largest end-group 6.9%
  - overview for experts was created
  - next steps: identification of experts and request for feedback

# Diversity Trees – tools to describe diversity

- observations
  - creating tree topology is relatively simple
    - getting time / attention of experts is sometimes difficult
    - concept of ‘genetically distinct groups’ is not always clear
  - weights tend to be based on current collection
    - current compositions ≠ ideal situation
    - stepwise decisions make it relatively easy

# Diversity Trees – tools to describe diversity

- issues
  - issues to be resolved
    - how can consensus weights be determined?
    - what collections should be mapped on the trees, how to handle duplication?
    - how to determine the ultimate collection size?
  - useful new tools
    - visualisation of trees
    - validation of topology with markers / NGS data

# Diversity Trees – tools to describe diversity

- applications of the trees
  - initial purpose: gap analysis
  - tree is required for core selector
    - possibility to select on-line core selection of user-defined domain
  - description / analysis of distribution of traits in collection
    - phenotypic traits → FIGS-type selection of material
    - use → prediction of future use for stock management

| | |
|---|---|
| CGN monocots | 0.10 sample / accession-year |
| CGN dicots | 0.35 sample / accession-year |
| CGN cultivated lettuce | 0.38 sample / accession-year |
| CGN wild lettuce | 0.71 sample / accession-year |

# Diversity Trees – tools to describe diversity

- concluding remarks
  - diversity trees are tools to describe diversity in a (combination of) collection(s)
  - creating diversity trees is easy
  - diversity trees serve many purposes – all increasing insight in and control over the collection(s)

it's strange not all genebanks managers use them !

# Thank you for your attention

WAGENINGEN UR
*For quality of life*

Centre for Genetic Resources, The Netherlands (CGN)