E

BMT/13/10 Add.
ORIGINAL: English
DATE: December 8, 2011

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES, AND DNA-PROFILING IN PARTICULAR

Thirteenth Session
Brasilia, November 22 to 24, 2011

ADDENDUM

COMBINING MORPHOLOGICAL AND MOLECULAR DISTANCE IN THE MANAGEMENT OF THE REFERENCE COLLECTION OF POTATO

Document prepared by experts from the Netherlands

nak tuinbouw

Combining morphological and molecular distances in the management of the reference collection of potato

Henk Bonthuis

Naktuinbouw, Wageningen, The Netherlands

nak tuinbouw

UPOV – 13th BMT
Brasilia 22-24 November 2011

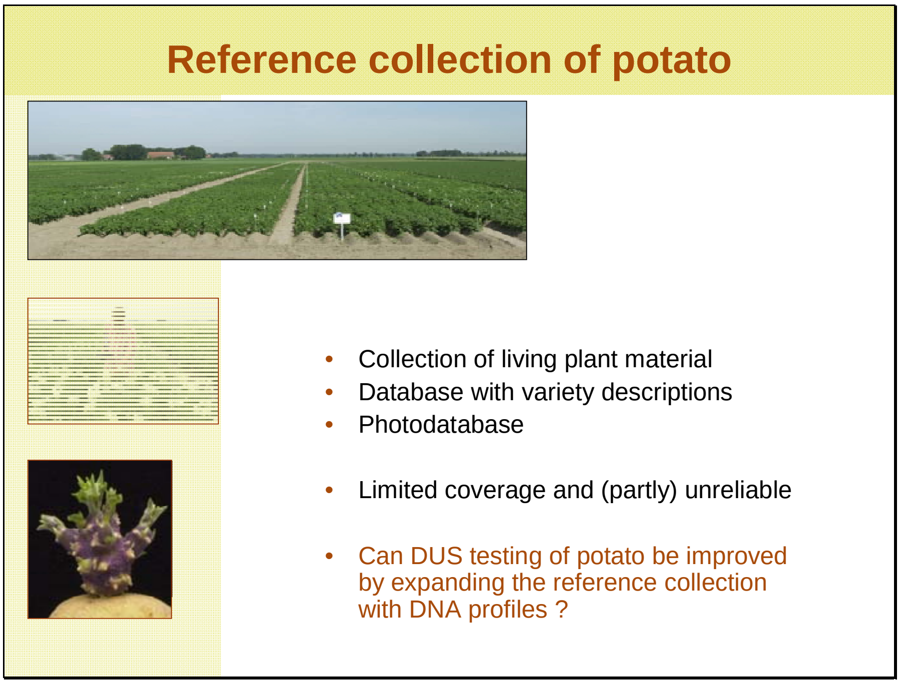
DUS testing of potato

Potato varieties are mostly distinct (and uniform):

- Tetraploid, heterozygous
- Parental cross resulting in unique seedlings
- Seedlings: vegetatively propagated (fixed genotype)
- Lightsprout test: > 80 % distinct
- Uniformity is usually no problem

Limitations in DUS potato testing

- Limited (living) reference collection:
  - World (common knowledge): over 4000 varieties
  - EU catalogue: 1400 varieties
  - NL operational collection: 350 varieties
- Limited coverage of database
  - morphological descriptions (EU coverage: 30-40 %)
  - photodatabase (EU coverage: 30-40%)
- Limitations due to:
  - quarantine regulations
  - distribution of bulk samples (tubers)
  - maintenance – annual, technical (diseases), economical
- Variation of morphological data:
  - diverse origins of variety descriptions
  - see: www.upov.org: TWA/34/13 – add 2 (2005)
  - quality and origin of tubers.
  - year-, location- and observer-effects

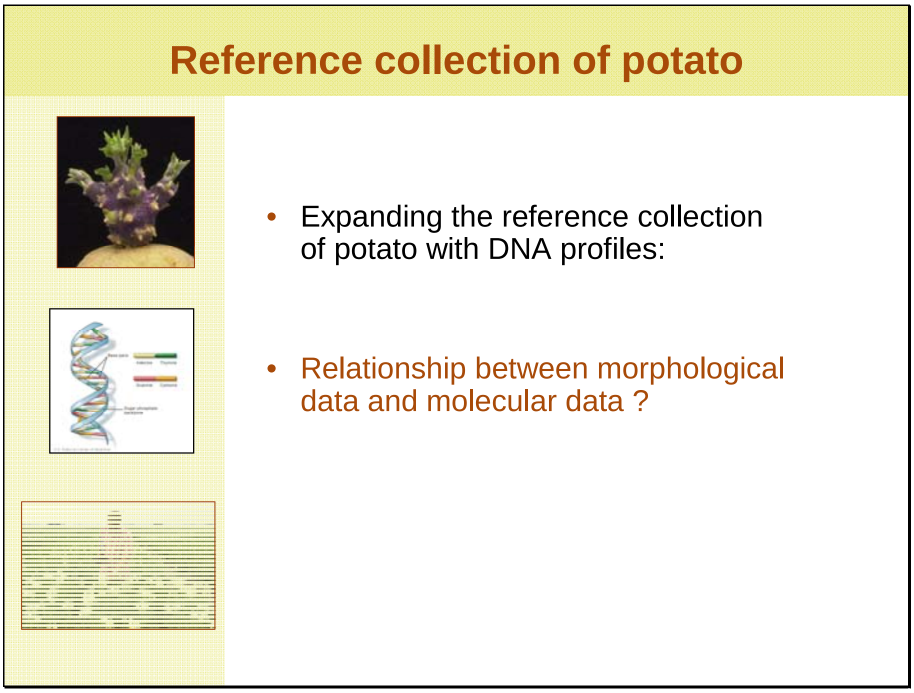


Reference collection of potato

- Collection of living plant material
- Database with variety descriptions
- Photodatabase

- Limited coverage and (partly) unreliable

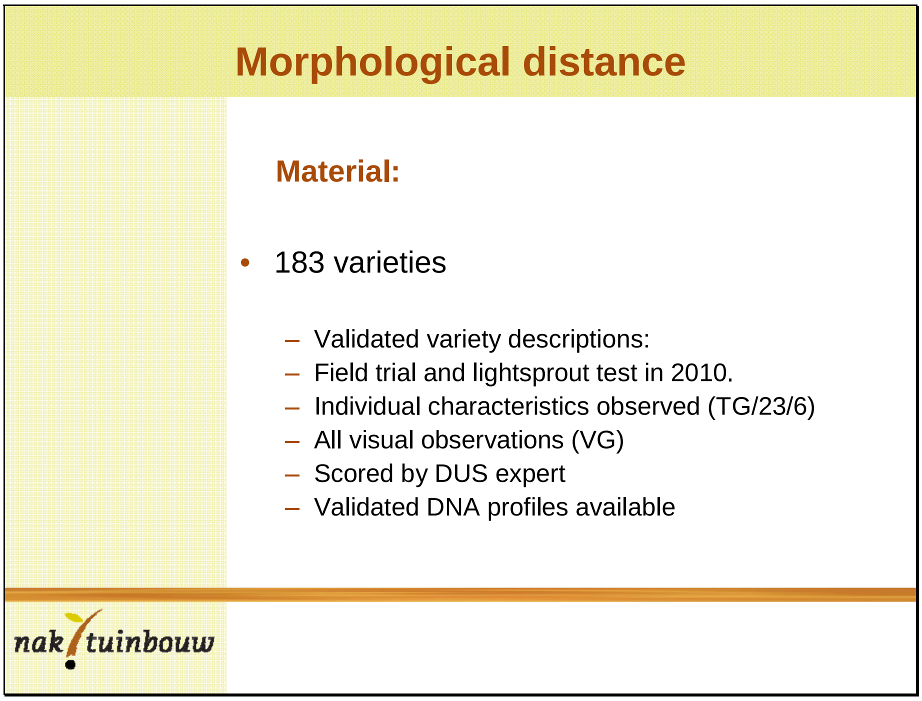
- Can DUS testing of potato be improved by expanding the reference collection with DNA profiles ?



Reference collection of potato

- Expanding the reference collection of potato with DNA profiles:

- Relationship between morphological data and molecular data ?

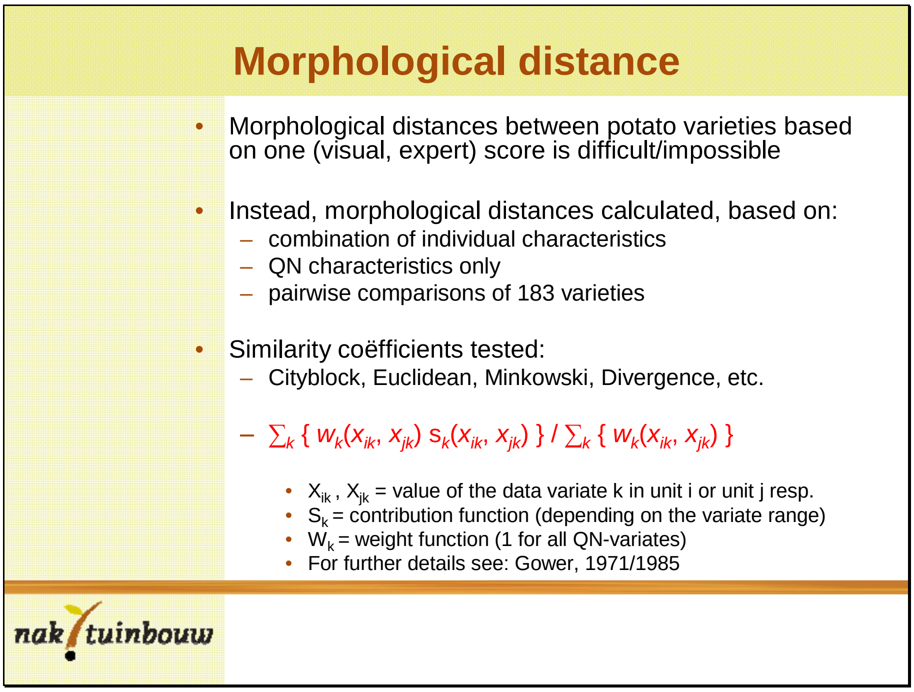


Morphological distance

Material:

- 183 varieties
  - Validated variety descriptions:
  - Field trial and lightsprout test in 2010.
  - Individual characteristics observed (TG/23/6)
  - All visual observations (VG)
  - Scored by DUS expert
  - Validated DNA profiles available

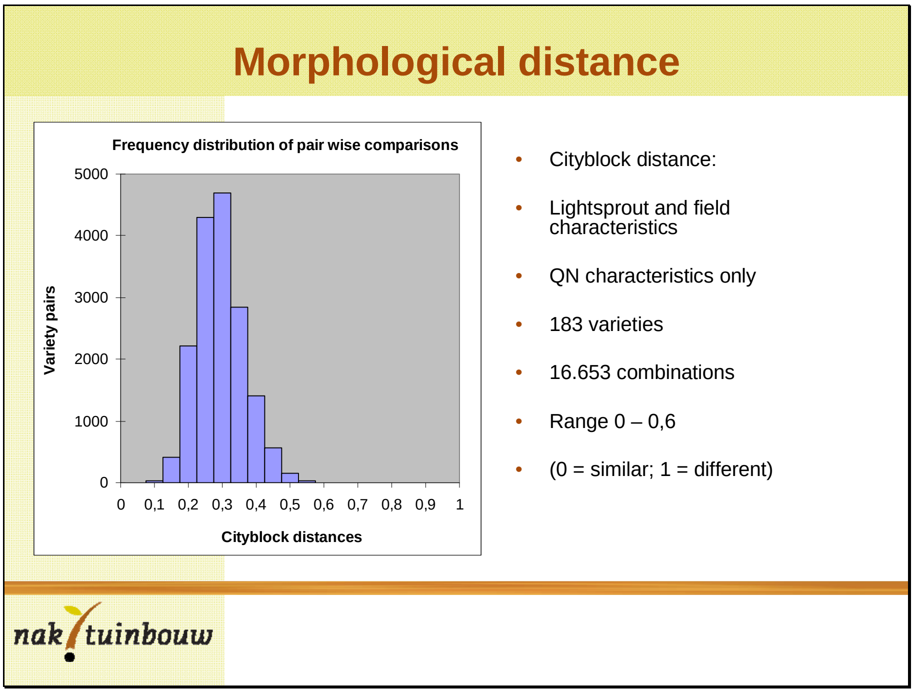
nak tuinbouw



Morphological distance

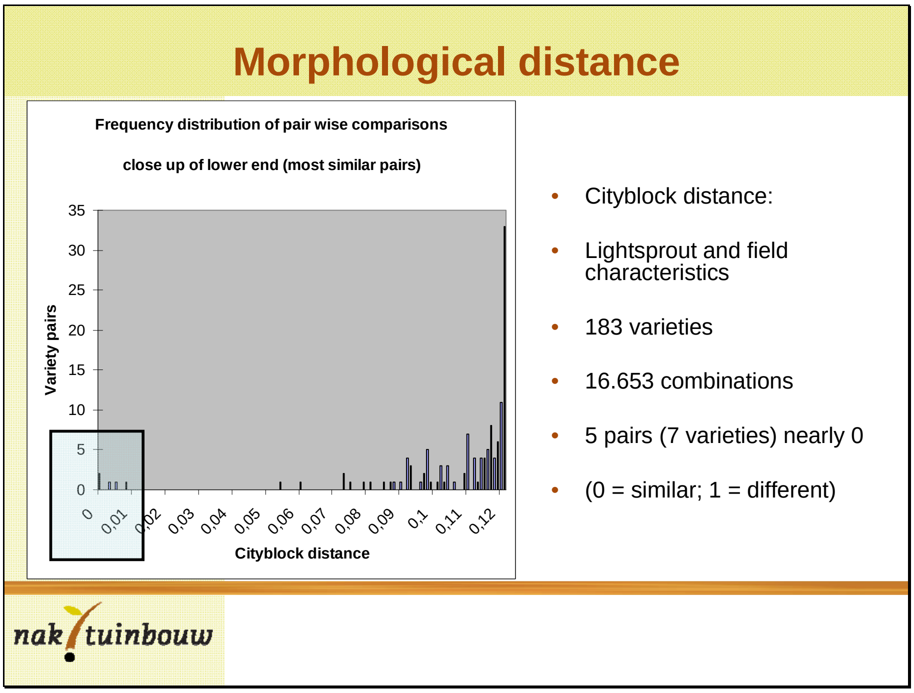
- Morphological distances between potato varieties based on one (visual, expert) score is difficult/impossible
- Instead, morphological distances calculated, based on:
  - combination of individual characteristics
  - QN characteristics only
  - pairwise comparisons of 183 varieties
- Similarity coëfficients tested:
  - Cityblock, Euclidean, Minkowski, Divergence, etc.
  - $\sum_k \{ w_k(x_{ik}, x_{jk})\, s_k(x_{ik}, x_{jk}) \} / \sum_k \{ w_k(x_{ik}, x_{jk}) \}$
    - X_{ik} , X_{jk} = value of the data variate k in unit i or unit j resp.
    - S_k = contribution function (depending on the variate range)
    - W_k = weight function (1 for all QN-variates)
    - For further details see: Gower, 1971/1985

nak tuinbouw



Morphological distance

- Cityblock distance:
- Lightsprout and field characteristics
- QN characteristics only
- 183 varieties
- 16.653 combinations
- Range 0 – 0,6
- (0 = similar; 1 = different)



Morphological distance

- Cityblock distance:
- Lightsprout and field characteristics
- 183 varieties
- 16.653 combinations
- 5 pairs (7 varieties) nearly 0
- (0 = similar; 1 = different)

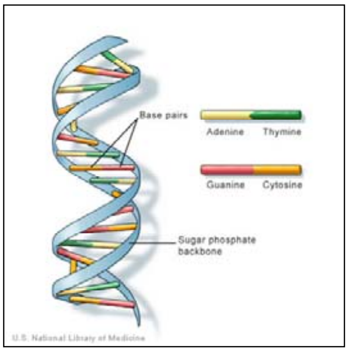
Molecular distance

- SSR database (NL/UK) of 900 varieties in total
  - samples collected from (NL/PL/DE/UK) reference collections
  - samples obtained from breeders/maintainers
- Over 200 varieties with samples from more than 1 source
- Most varieties analysed in 2 labs (NL and UK)
- Many varieties 2 profiles per lab (UK)
- In case of anomalies/doubts: variety re-sampled
- => in total almost 3000 profiles scored
- For more details, see: BMT/11/9 and BMT/11/10

Molecular profiles

Used STMS markers:

marker	chromosome
0019	VI
2005	XI
2028	XII
3009	VII
3012	IX
3012	IV
5136	I
5148	V
SSR1	VIII

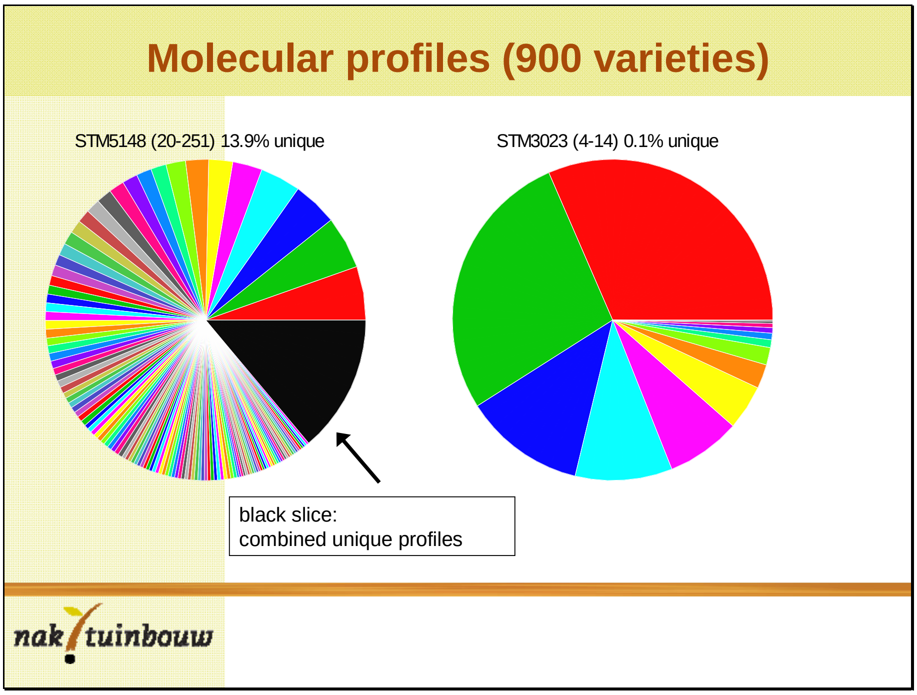


U.S. National Library of Medicine

Molecular profiles (900 varieties)

marker	# alleles	# different profiles	% unique profiles	% most common profile	PIC value
0019	10	61	1.8	17	0,92
2005	6	21	0.4	37	0,80
2028	9	62	2.2	23	0,90
3009	14	48	2.1	34	0,81
3012	7	27	0.2	19	0,87
3023	4	14	0.1	32	0,79
5136	11	54	2.8	14	0,92
5148	20	251	13.9	5	0,98
SSR1	14	119	5.5	17	0,93

nak tuinbouw



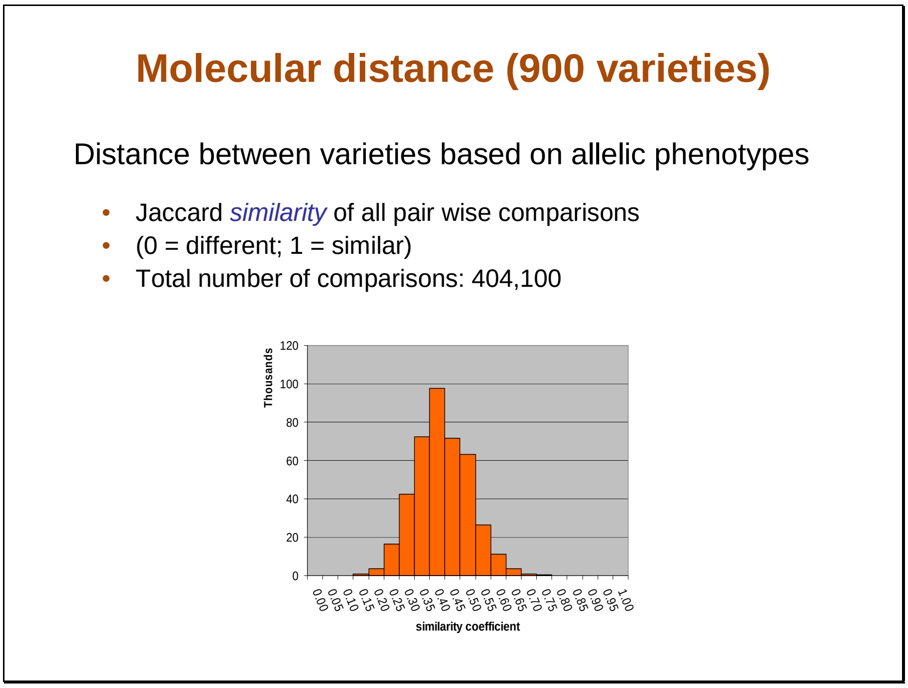
Molecular profiles (900 varieties)

STM5148 (20-251) 13.9% unique

STM3023 (4-14) 0.1% unique

black slice:
combined unique profiles

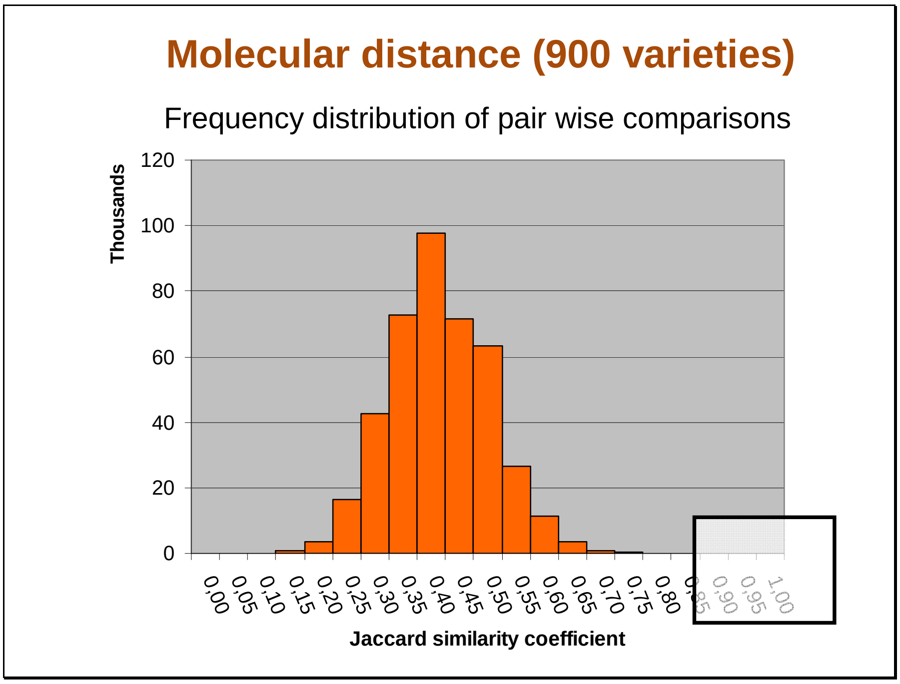
nak tuinbouw



Molecular distance (900 varieties)

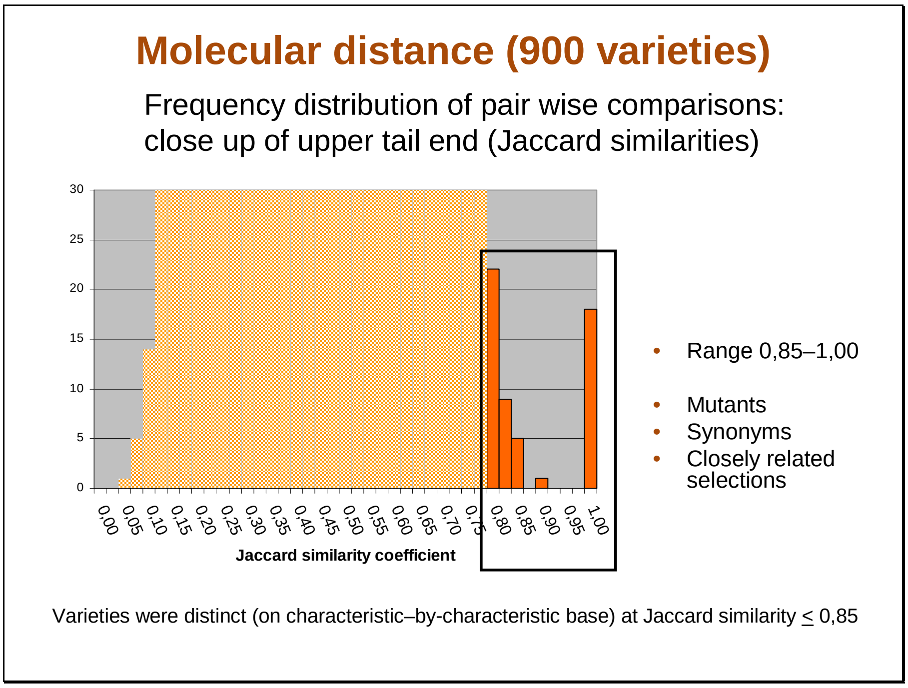
Distance between varieties based on allelic phenotypes

- Jaccard *similarity* of all pair wise comparisons
- (0 = different; 1 = similar)
- Total number of comparisons: 404,100



Molecular distance (900 varieties)

Frequency distribution of pair wise comparisons



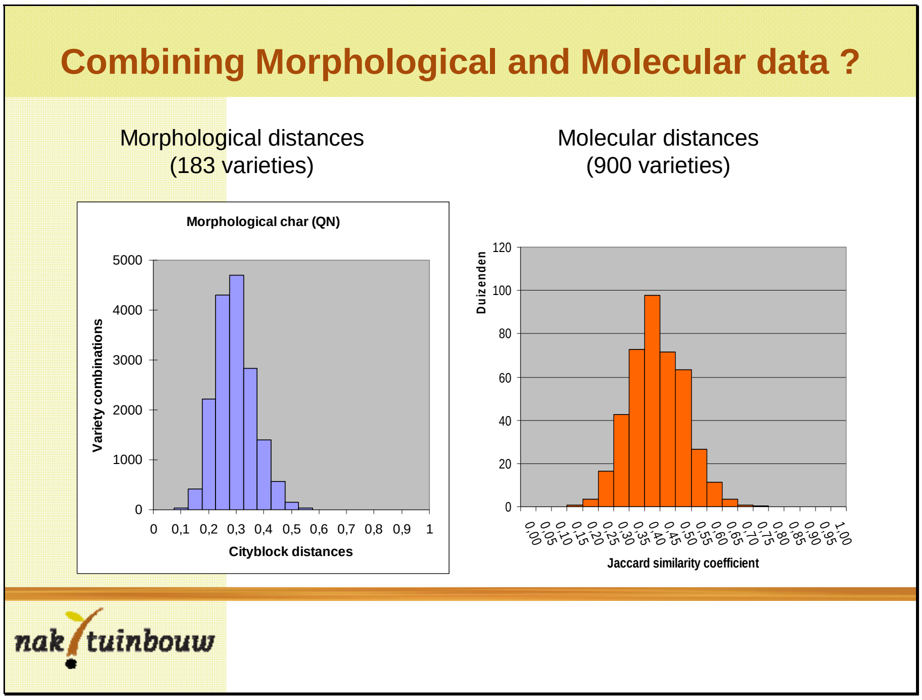
Molecular distance (900 varieties)

Frequency distribution of pair wise comparisons:
close up of upper tail end (Jaccard similarities)

- Range 0,85–1,00

- Mutants
- Synonyms
- Closely related selections

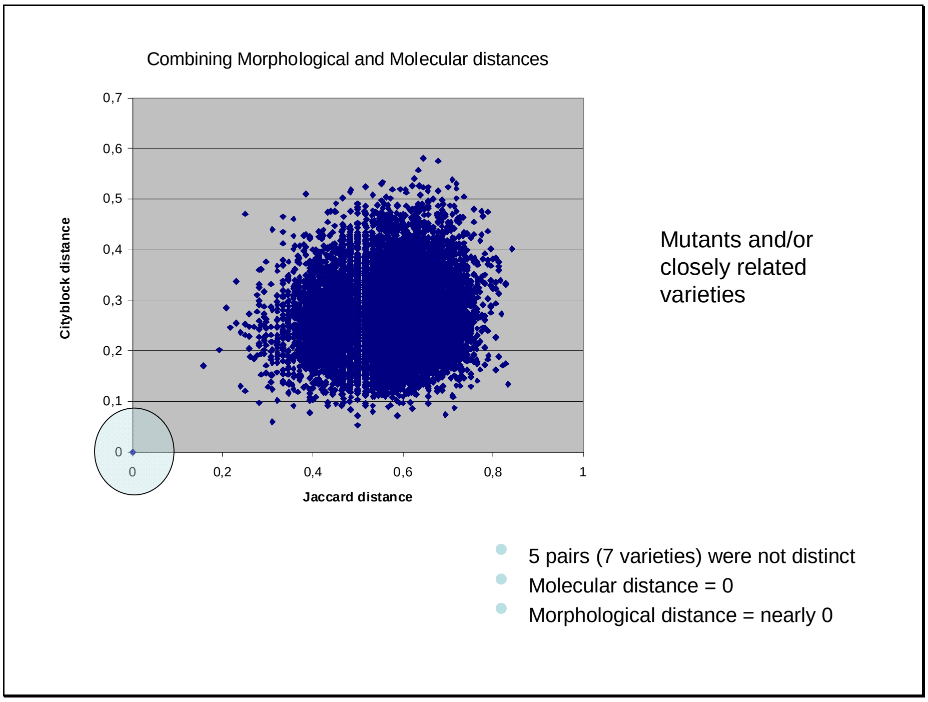
Varieties were distinct (on characteristic–by-characteristic base) at Jaccard similarity ≤ 0,85



Combining Morphological and Molecular data ?

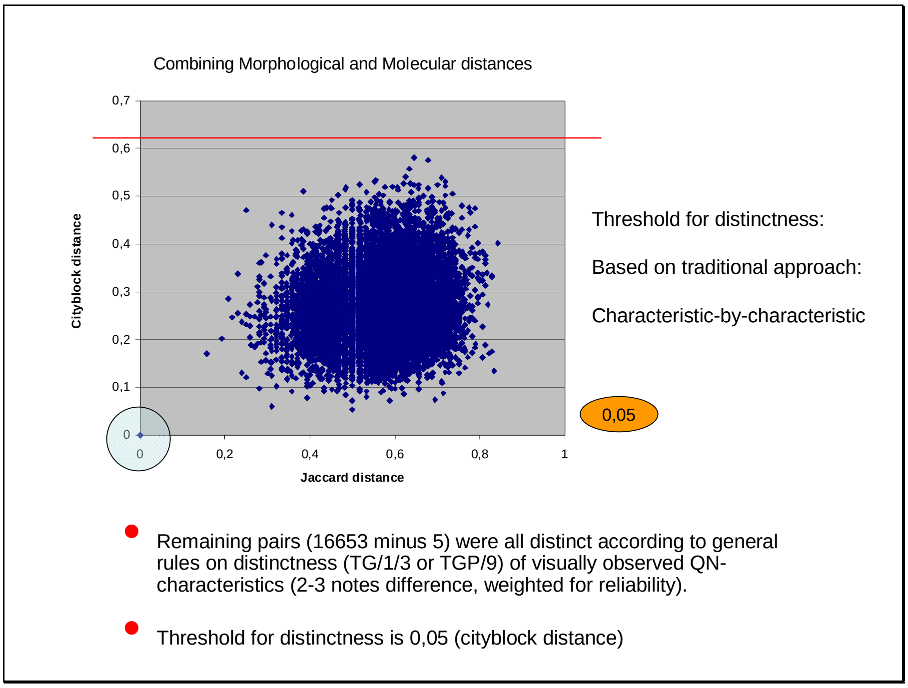
Morphological distances
(183 varieties)

Molecular distances
(900 varieties)



Mutants and/or
closely related
varieties

- 5 pairs (7 varieties) were not distinct
- Molecular distance = 0
- Morphological distance = nearly 0

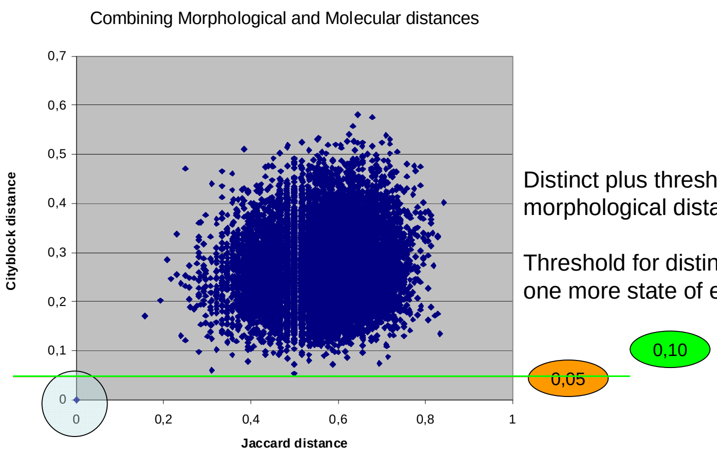


Threshold for distinctness:

Based on traditional approach:

Characteristic-by-characteristic

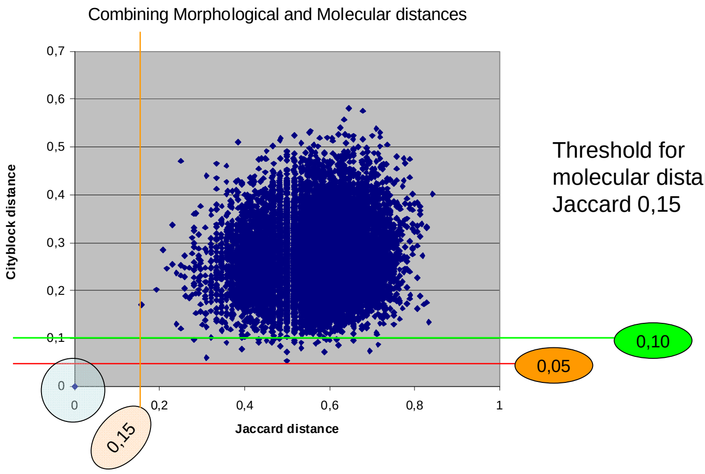
- Remaining pairs (16653 minus 5) were all distinct according to general rules on distinctness (TG/1/3 or TGP/9) of visually observed QN-characteristics (2-3 notes difference, weighted for reliability).
- Threshold for distinctness is 0,05 (cityblock distance)



Distinct plus threshold for morphological distance =

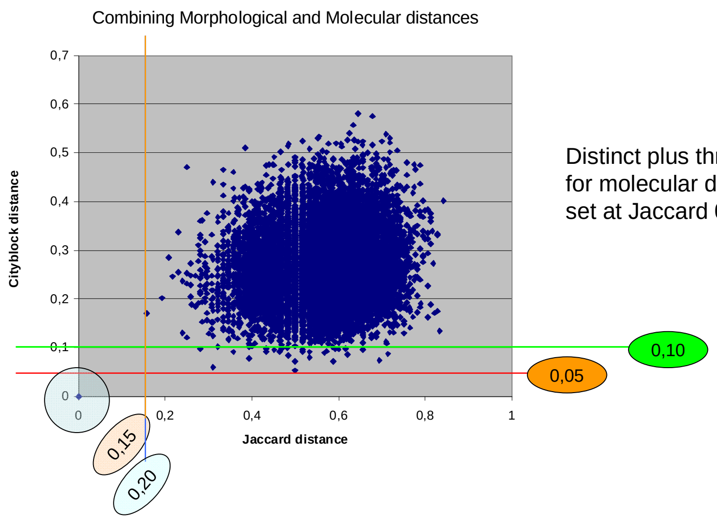
Threshold for distinctness + one more state of expression

- Reference varieties to be included in the field trial are defined by the threshold for distinctness + 1 state of expression = 3-4 notes difference, weighted for reliability.
- Threshold for distinct plus is 0,10 (cityblock distance)
- 20 pairs (out of 16653 minus 5) to be included in field trial (below distinct plus level)



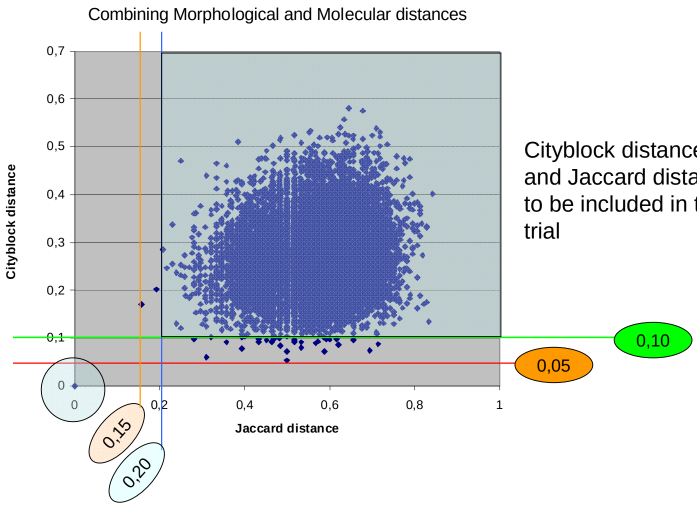
Threshold for molecular distance: Jaccard 0,15

- Threshold for molecular distance based on 16,653 pairs (minus 5) corresponds with threshold previously found in the SSR database project (900 varieties > 400,000 variety pairs)



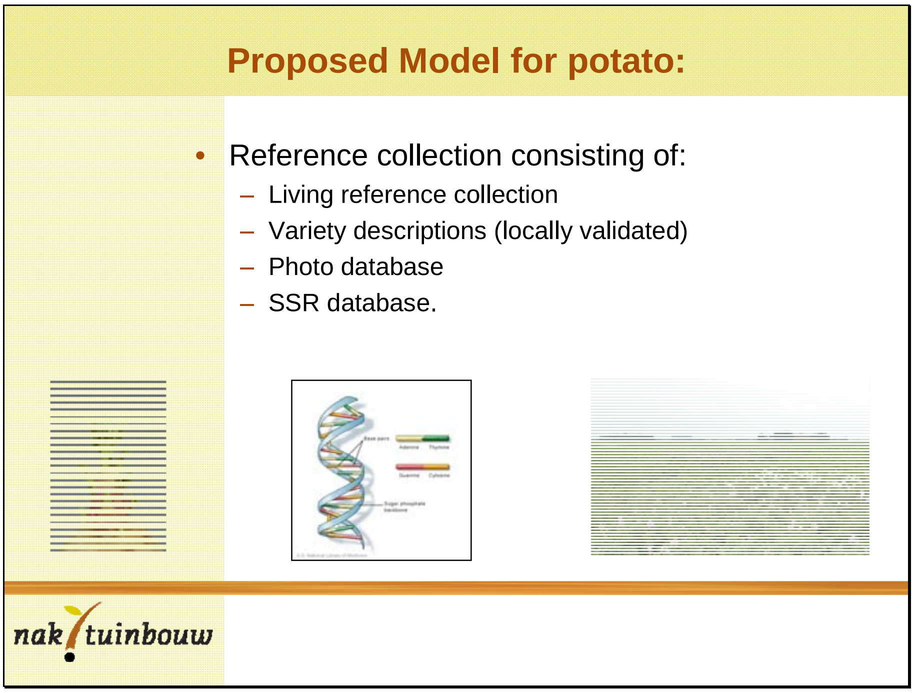
Distinct plus threshold for molecular distance: set at Jaccard 0,20

- Low risk for wrong decisions if reference varieties are excluded from the growing trial at Jaccard distance > 0,20



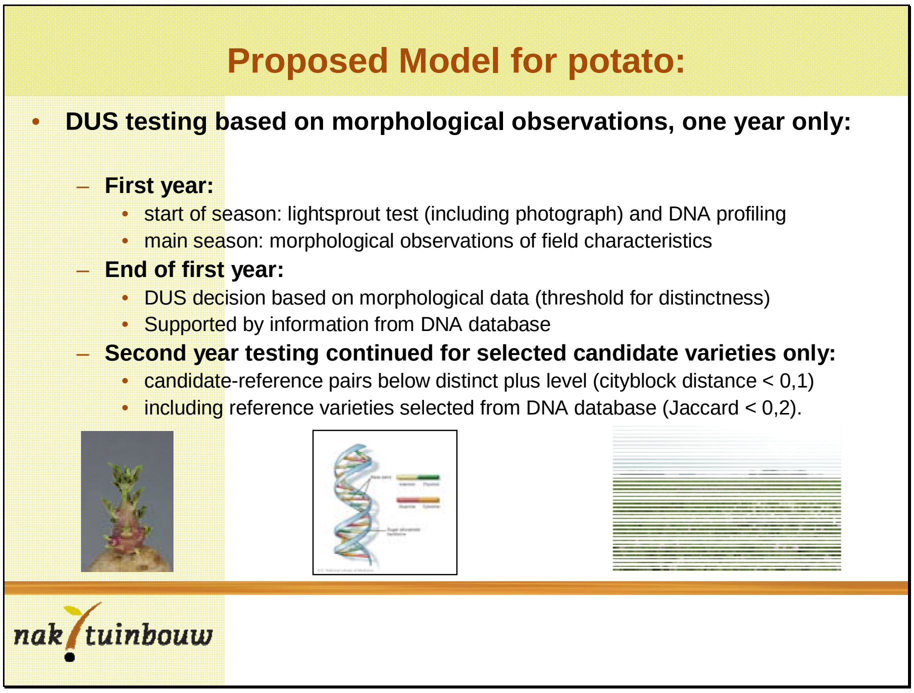
Cityblock distance < 0,10 and Jaccard distance < 0,20 to be included in the growing trial

- High-lighted area: above distinct plus thresholds
- low risk for wrong decisions on reference varieties to be excluded from the growing trial



Proposed Model for potato:

- Reference collection consisting of:
  - Living reference collection
  - Variety descriptions (locally validated)
  - Photo database
  - SSR database.

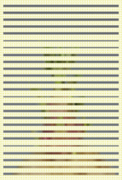


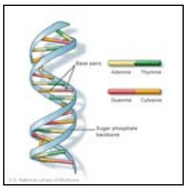
Proposed Model for potato:

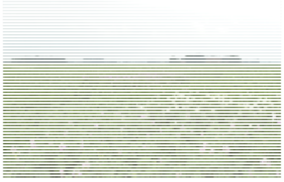
- **DUS testing based on morphological observations, one year only:**
  - **First year:**
    - start of season: lightsprout test (including photograph) and DNA profiling
    - main season: morphological observations of field characteristics
  - **End of first year:**
    - DUS decision based on morphological data (threshold for distinctness)
    - Supported by information from DNA database
  - **Second year testing continued for selected candidate varieties only:**
    - candidate-reference pairs below distinct plus level (cityblock distance < 0,1)
    - including reference varieties selected from DNA database (Jaccard < 0,2).

Proposed Model for potato:

- Reduced testing period compensated by added value molecular profile:
  - Significant expansion of reference collection (potentially expanded with SSR databases of cooperative DNA labs)
  - Detection of most similar reference varieties (Jaccard distance < 0,2)
  - Morphological observation supported by information from molecular profile.
  - Spin-off for identification purposes (based on original identity sample)







nak tuinbouw

Quality in Horticulture

[End of document]