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

Eleventh Session
Madrid, September 16 to 18, 2008

ADDENDUM
CONSTRUCTION OF AN INTEGRATED MICROSATELLITE AND KEY MORPHOLOGICAL CHARACTERISTIC DATABASE OF POTATO VARIETIES ON THE EU COMMON CATALOGUE
PART 1: DISCUSSION OF MORPHOLOGICAL AND MOLECULAR DATA

Document prepared by experts from the Netherlands
Construction of an integrated microsatellite and key morphological characteristic database of potato varieties on the EU common catalogue

Part 1: Discussion of morphological and molecular data

Lysbeth Hof
Project

- Results from a project carried out by 4 DUS testing stations, and co-funded by CPVO
- Partners:
  - SASA, United Kingdom
  - BSA, Germany
  - COBORU, Poland
  - Naktuinbouw, the Netherlands
- 2006-2008
- This presentation is part 1 of 2

Background

- DUS testing for potato in Europe is de-centralized
- >100 applications for DUS each year in Europe
- to date, no exchange of variety descriptions
- live reference collections limited
- demand for rapid identification methods
Background

- UK and NL separately already had started work on molecular markers in potato
- As of 2004 forces were joined, resulting in this project

Activities

- **Scope:**
  - Potato varieties of the EU Common Catalogue 2006 (1104 varieties)

- **End result:**
  - Morphological descriptions of 733 varieties
  - Light sprout pictures of 377 varieties
  - DNA samples + SSR profiles of 900 varieties
Morphological descriptions

- In total 856 descriptions of 733 varieties:
  - 622 from a single country
  - 99 from 2 countries
  - 11 from 3 countries
  - 1 from all 4 countries

- Characteristics:
  - light sprout char.
  - TQ char. where available

Morphological descriptions


- Both guidelines are different for some characteristics: for these the UK data were not used in comparisons

- Comparisons based on 12 varieties with descriptions of ≥ 3 countries
Morphological descriptions

Quantitative data:

<table>
<thead>
<tr>
<th>Var_xx</th>
<th>characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1  3  4  5  6  7  8  9  10  ...</td>
</tr>
<tr>
<td>Country A</td>
<td>3  7  3  2  3  8  9  2  5</td>
</tr>
<tr>
<td>Country B</td>
<td>3  6  2  5  3  8  8  5  6</td>
</tr>
<tr>
<td>Country C</td>
<td>4  5  -  4  3  -  8  3  4</td>
</tr>
<tr>
<td>difference</td>
<td>1  2  1  3  0  0  1  3  2</td>
</tr>
</tbody>
</table>

Morphological descriptions

Quantitative data (15 char.):

freq. of observed differences
Morphological descriptions

(Pseudo-)qualitative data:

- tuber eye colour
- tuber skin colour
- tuber flesh colour
- lightsprout shape

Conclusions:
- Descriptions vary between different sources (countries)
- Descriptions of the same variety from different sources would have been declared distinct in almost all cases
- Exchange of descriptions between offices is not useful for reference collections
- Comparisons should be carried out side-by-side
Molecular profiles

- In total 900 varieties
  - samples collected from own reference collections
  - samples obtained from breeders/maintainers

- More than 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

Molecular profiles

Used markers:

<table>
<thead>
<tr>
<th>marker</th>
<th>chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>0019</td>
<td>VI</td>
</tr>
<tr>
<td>2005</td>
<td>XI</td>
</tr>
<tr>
<td>2028</td>
<td>XII</td>
</tr>
<tr>
<td>3009</td>
<td>VII</td>
</tr>
<tr>
<td>3012</td>
<td>IX</td>
</tr>
<tr>
<td>3012</td>
<td>IV</td>
</tr>
<tr>
<td>5136</td>
<td>I</td>
</tr>
<tr>
<td>5148</td>
<td>V</td>
</tr>
<tr>
<td>SSR1</td>
<td>VIII</td>
</tr>
</tbody>
</table>
Molecular profiles

- Potato is tetraploid: 4 alleles present per marker
  
  \[ \text{A} \quad \text{A} \quad \text{C} \quad \text{F} \]
  
  *e.g. chromosome VI:*

- Alleles were scored as absent/present:
  - no information on number of copies per allele in each variety
  - 'allelic phenotype' (AOF, not AAOF)
  - no data on allele frequency in population

Molecular profiles, scorability

- Initially between labs a substantial number of small differences in scoring of alleles

- Often this concerned the same alleles

- Fine tuning of scoring (decision rules)

- A few alleles remain notoriously unreliable to score (depending on equipment used)
Molecular profiles, scorability

- With multiple samples/analyses consensus on correct profile can be reached
- Discarding these alleles in analyses has no effect on effectiveness of database
- When constructing a database: analyses at 2 labs enhances robustness

Molecular profiles (900 varieties)

<table>
<thead>
<tr>
<th>marker</th>
<th># alleles</th>
<th># different profiles</th>
<th>% unique profiles</th>
<th>% most common profile</th>
<th>PIC value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0019</td>
<td>10</td>
<td>61</td>
<td>1.8</td>
<td>17</td>
<td>0.92</td>
</tr>
<tr>
<td>2005</td>
<td>6</td>
<td>21</td>
<td>0.4</td>
<td>37</td>
<td>0.80</td>
</tr>
<tr>
<td>2028</td>
<td>9</td>
<td>62</td>
<td>2.2</td>
<td>23</td>
<td>0.90</td>
</tr>
<tr>
<td>3009</td>
<td>14</td>
<td>48</td>
<td>2.1</td>
<td>34</td>
<td>0.81</td>
</tr>
<tr>
<td>3012</td>
<td>7</td>
<td>27</td>
<td>0.2</td>
<td>19</td>
<td>0.87</td>
</tr>
<tr>
<td>3023</td>
<td>4</td>
<td>14</td>
<td>0.1</td>
<td>32</td>
<td>0.79</td>
</tr>
<tr>
<td>5136</td>
<td>11</td>
<td>54</td>
<td>2.8</td>
<td>14</td>
<td>0.92</td>
</tr>
<tr>
<td>5148</td>
<td>20</td>
<td>251</td>
<td>13.9</td>
<td>5</td>
<td>0.98</td>
</tr>
<tr>
<td>SSR1</td>
<td>14</td>
<td>119</td>
<td>5.5</td>
<td>17</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Molecular profiles (900 varieties)

STM5148 (20-251) 13.9% unique
STM3023 (4-14) 0.1% unique

black slice: combined unique profiles

Molecular profiles (900 varieties)

- Apart from known mutants, 8 pairs with identical profile were found:
  - 1 already suspected being the same
  - 1 already suspected being mislabeled
  - 1 already suspected being mutant, not offspring as was claimed
  - 5 unexplained
    - all pairs have very similar morphological descriptions
    - material re-sample, identical results
    - another 31 markers used, identical results
Molecular profiles, probability

Probability of 2 unrelated varieties showing identical profile by chance?

- No allele frequencies available
- Based on most common allelic phenotype per marker:

  1 in 2.8 million

- (Common ancestry not taken into account)

Molecular profiles, distance

Distance between varieties based on allelic phenotypes?

- Jaccard similarity of all pairwise comparisons
- Total number of comparisons: 404,100
Frequency distribution of pairwise comparisons

Frequency distribution of pairwise comparisons: close up of upper tail end
Frequency distribution of pairwise comparisons: close up of upper tail end

Similarity 1.0:
- 18 pairs
- 10 known mutants
- 8 unexpected pairs

Similarity 0.91:
- 1 pair
- Equivalent to 2 alleles
- Related (one is parent of the other)
Frequency distribution of pair wise comparisons:
close up of upper tail end

- Similarity 0.85 – 0.88: 5 pairs
  - equivalent to 3 alleles
  - 1 related, 4 unrelated

- > 99.99%
  - more than 4 alleles difference

> 99.99%

- more than 4 alleles difference
Molecular profiles, labeling

- 21 cases of mislabeled samples (1-2% of total)
- These cases all were brought to light and solved by the use of molecular markers

Molecular profiles, denomination

- Same variety name - different profiles?
  - variety names are re-used
  - old varieties from before regulations on denomination
  - mislabeling
  - help from pedigree-information
    - [www.plantbreeding.wur.nl/potatopedigree](http://www.plantbreeding.wur.nl/potatopedigree)
    - [www.europotato.org](http://www.europotato.org)
Molecular profiles, stability

- Only 1 variety appeared to have a genuine polymorphism:
  - samples from 2 countries consistently different for presence of 1 allele (mutation?)
    BDF - BDF

- Conclusion: molecular profiles of varieties (based on these markers) seem to be very stable
Molecular profiles, blind test

- Effectiveness of database tested by 20 blind samples (10 from PL and 10 from GE)
- 18 samples 100% match in both UK and NL
- 2 samples identical, could not be separated (already established as unexpected matching pair)

Molecular database, recommendations

- Analyse samples in duplicate
  - preferably collect samples from more than 1 source
  - in case of doubt on interpretation of alleles, re-analyse

- If implemented in DUS:
  - extract DNA samples from identity material
  - preferably store DNA at 2 locations
Use of molecular markers in DUS

1. Check identification reference varieties

2. Select reference varieties with high genetic similarity and include in trial? (in addition to other selection methods)

3. Develop adjusted guideline/protocol with morphological as well as molecular characteristics?

Co-workers

- Alex Reid
- Beate Rücker
- Swenja Tamms
- Ewa Milczynska
- Bogna Kowalczyk
- staff of PRI (NL)

THANK YOU FOR YOUR ATTENTION
Quality in Horticulture