E

BMT/9/10 Add.
ORIGINAL: English
DATE: July 29, 2005

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

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

Ninth Session

Washington, D.C., June 21 to 23, 2005

ESSENTIAL DERIVATION AND DIVERSITY ISSUES IN WINTER WHEAT

Document prepared by experts from the United Kingdom

1. The BMT agreed that, where agreed by the relevant experts, the presentations made at the meeting should be made available in the BMT document section of the UPOV website, as addenda to the relevant documents. This document contains a copy of the presentation made by Mr. John Law (United Kingdom), based on document BMT/9/10.

ESSENTIAL DERIVATION & DIVERSITY ISSUES IN WINTER WHEAT.

John R Law and Robert J Cooke

NIAB, Cambridge, UK

There are few very large data sets with both extensive phenotype and genotype data - as they are costly to put together – but some do exist.

One such is the FP6 QL project Gediflux coordinated by NIAB. (4 crops Winter Wheat, Potato, Maize, Barley). Partners in the project from UK, NL, F and D.

These data may be exploited for other, secondary purposes, which will illustrated in this paper based on Winter Wheat.

For winter wheat, over 270 varieties were selected to represent established commercial varieties grown in Europe over the past 5 decades and taken as an estimate of **genetic diversity exploited in commercial terms and are designated the "RL" set.**

A further 229 varieties were analysed that had been National Listed in the 1980s and 1990s to represent the **wider genetic diversity available, but not necessarily exploited and are designated the "NL" set.**

Three marker systems are utilised.
- SSR
- S-SAP multi-locus AFLP like marker
- NBS *"expressed regions and linked to disease resistance genes"*

Data from 40 SSR loci were analysed as pattern profiles, and pair-wise similarity coefficients computed by Euclidean algorithm and the method known as “City block” or Manhattan.

In addition, 70 S-SAP bands and 70 NBS bands were available and, as the bands are present/absent, the pair-wise similarities were generated by the Jaccards’ method.

There are a number of possible approaches to the calculation and analysis of such data.

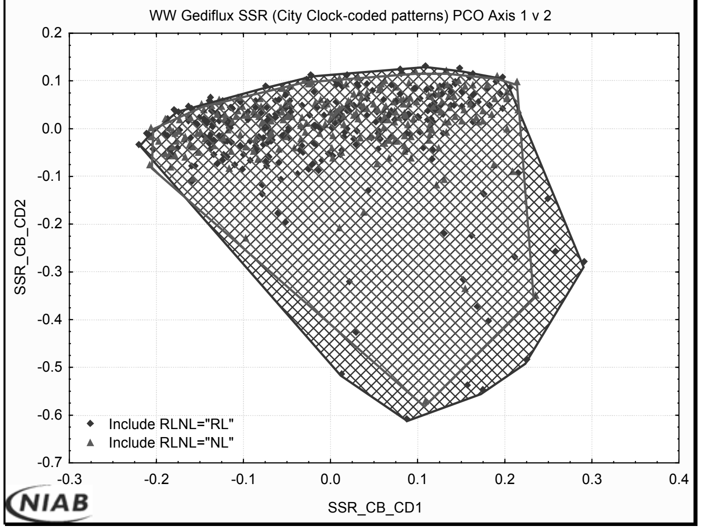
We here have restricted the presentation to PCO’s and “convex hulls” which we have used in other publications and presentations to illustrate temporal flux in genetic diversity.

NIAB

Following diagram shows the Genetic Diversity based on the SSR marker data; coded to show the data from the RL and the NL portions of the data.

PCO on SSR, City Block similarity calculation with “convex hulls” added to illustrate the most extreme RL and the NL varieties in the data set.

NIAB

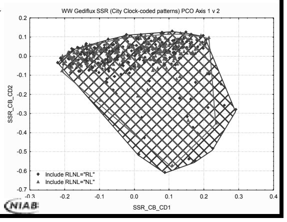


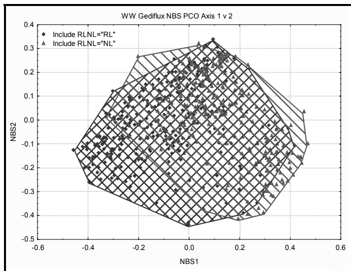
NIAB

Based on the analysis and presentational approach for SSR data:-

No major partitioning of the RL and the NL portions of the data – high degree of overlap.
In fact, RL > NL exploitation of "useful genes".

Some "structure" within both the RL and the NL portions of the data

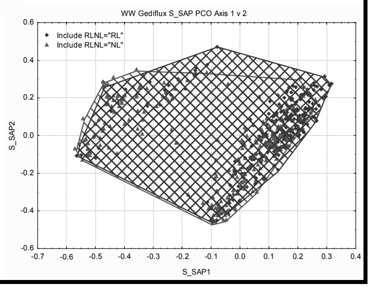




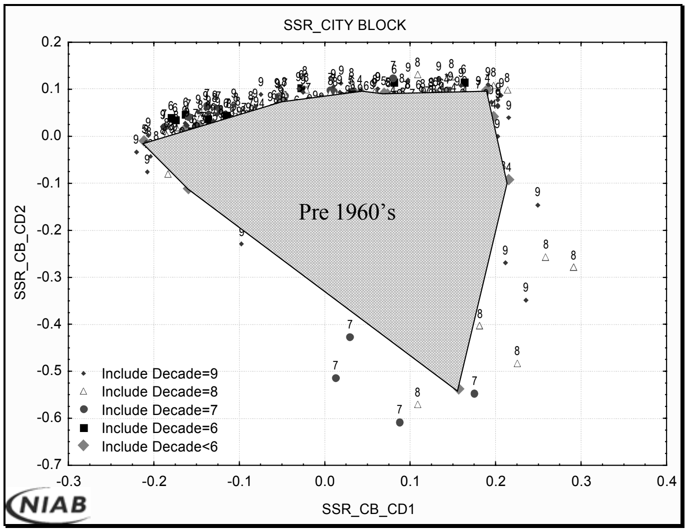
Similar "patterns" with NBS and S-SAP so results with SSR may apply generally

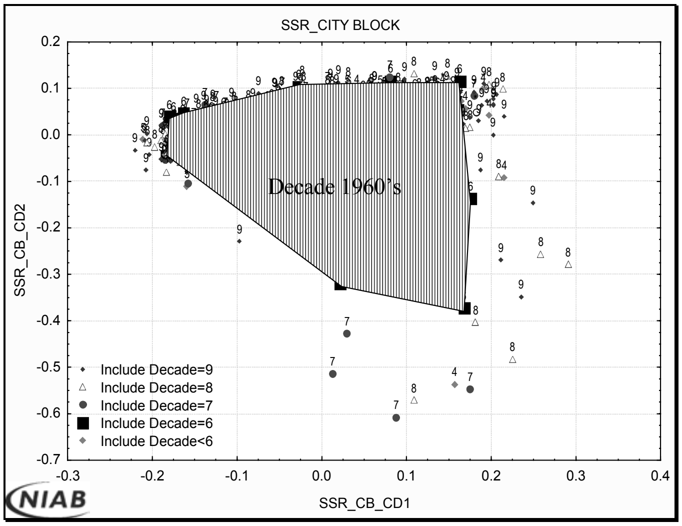
WW S-SAP

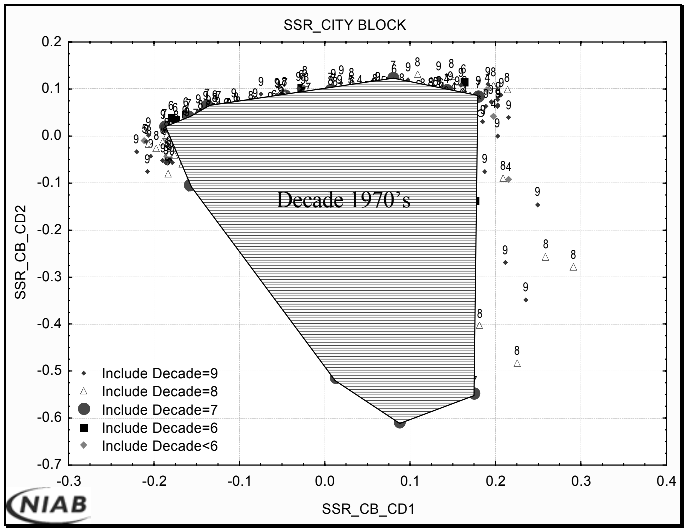
WW NBS

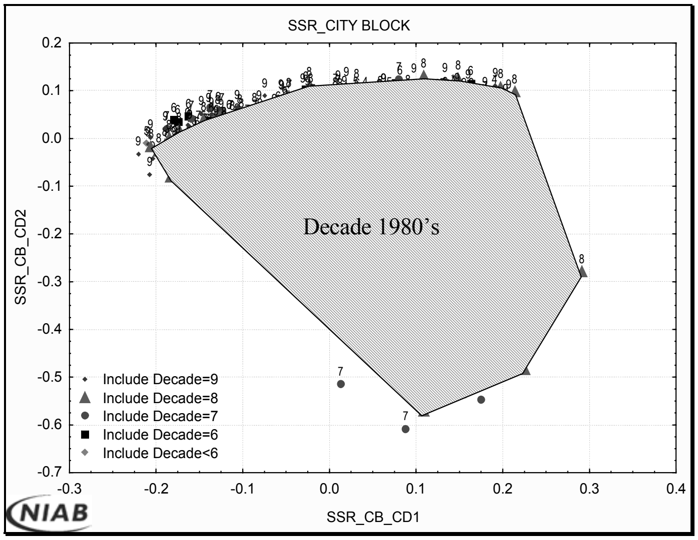


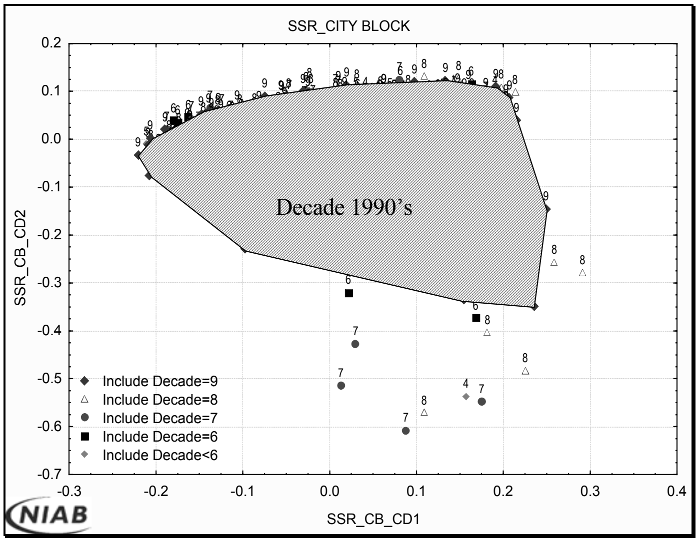
Winter Wheat Decadal Diversity Flux

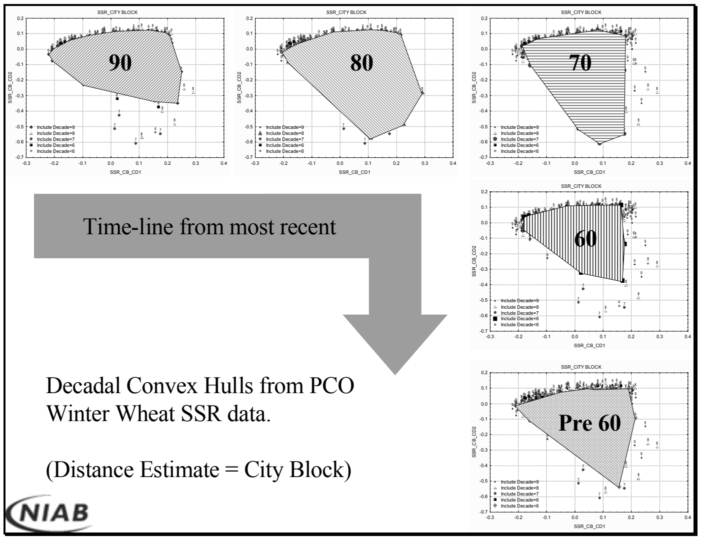

SSR_CITY BLOCK
Pre 1960's
SSR_CB_CD2
SSR_CB_CD1
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
Include Decade=9
Include Decade=8
Include Decade=7
Include Decade=6
Include Decade<6
NIAB


SSR_CITY BLOCK
Decade 1960's
SSR_CB_CD2
SSR_CB_CD1
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
Include Decade=9
Include Decade=8
Include Decade=7
Include Decade=6
Include Decade<6
NIAB


SSR_CITY BLOCK
Decade 1970's
SSR_CB_CD2
SSR_CB_CD1
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
Include Decade=9
Include Decade=8
Include Decade=7
Include Decade=6
Include Decade<6
NIAB

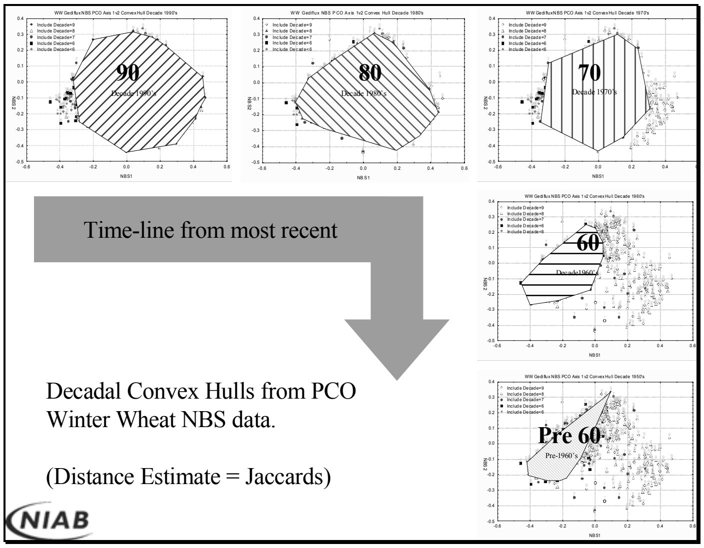






Decadal Convex Hulls from PCO Winter Wheat SSR data.

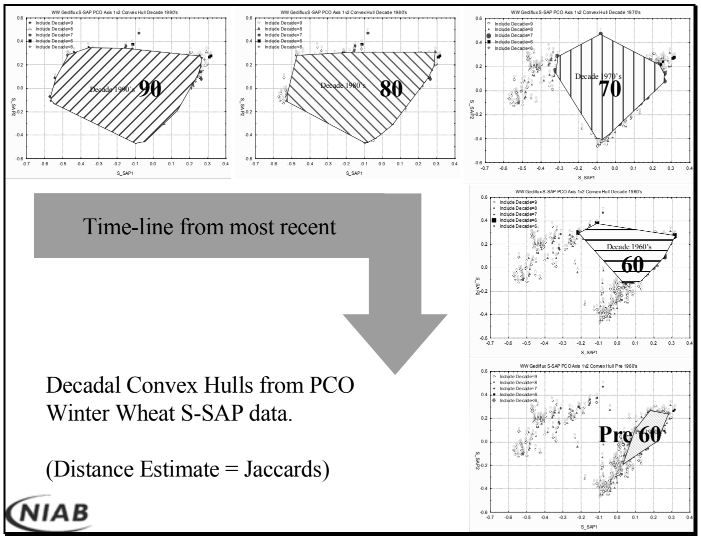
(Distance Estimate = City Block)



Time-line from most recent

Decadal Convex Hulls from PCO
Winter Wheat NBS data.

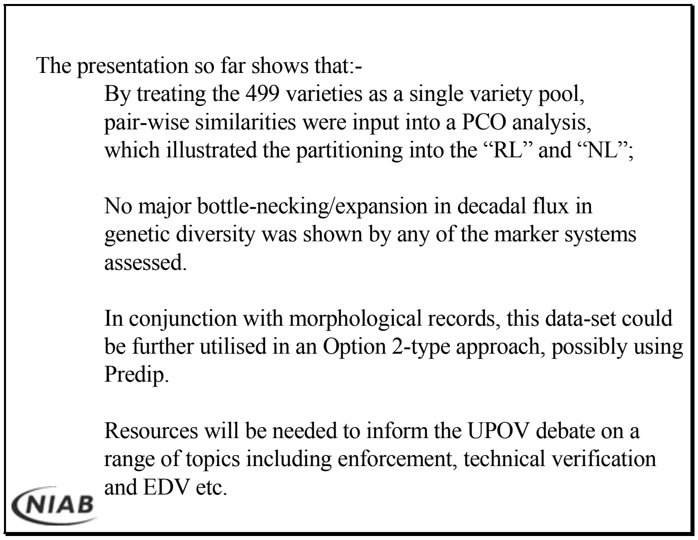
(Distance Estimate = Jaccards)



Time-line from most recent

Decadal Convex Hulls from PCO
Winter Wheat S-SAP data.

(Distance Estimate = Jaccards)



The presentation so far shows that:-

By treating the 499 varieties as a single variety pool, pair-wise similarities were input into a PCO analysis, which illustrated the partitioning into the “RL” and “NL”;

No major bottle-necking/expansion in decadal flux in genetic diversity was shown by any of the marker systems assessed.

In conjunction with morphological records, this data-set could be further utilised in an Option 2-type approach, possibly using Predip.

Resources will be needed to inform the UPOV debate on a range of topics including enforcement, technical verification and EDV etc.

The presentation so far shows that:-

By treating the 499 varieties as a single variety pool, pair-wise similarities were input into a PCO analysis, which illustrated the partitioning into the “RL” and “NL”;

No major bottle-necking/expansion in decadal flux in genetic diversity was shown by any of the marker systems assessed.

In conjunction with morphological records, this data-set could be further utilised in an Option 2-type approach, possibly using Predip.

Resources will be needed to inform the UPOV debate on a range of topics including enforcement, technical verification and EDV etc.

The presentation so far shows that:-

By treating the 499 varieties as a single variety pool, pair-wise similarities were input into a PCO analysis, which illustrated the partitioning into the “RL” and “NL”;

No major bottle-necking/expansion in decadal flux in genetic diversity was shown by any of the marker systems assessed.

In conjunction with morphological records, this data-set could be further utilised in an Option 2-type approach, possibly using Predip.

Resources will be needed to inform the UPOV debate on a range of topics including enforcement, technical verification and EDV etc.

The presentation so far shows that:-

By treating the 499 varieties as a single variety pool, pair-wise similarities were input into a PCO analysis, which illustrated the partitioning into the “RL” and “NL”;

No major bottle-necking/expansion in decadal flux in genetic diversity was shown by any of the marker systems assessed.

In conjunction with morphological records, this data-set could be further utilised in an Option 2-type approach, possibly using Predip.

Resources will be needed to inform the UPOV debate on a range of topics including enforcement, technical verification and EDV etc.

In particular …
Winter Wheat - Essential Derivation.

Evidence of “structure” ,previous work has shown, is vital in the establishment of any potential EDV system.

Here “structure” is natural classification; grouping; Clustering etc

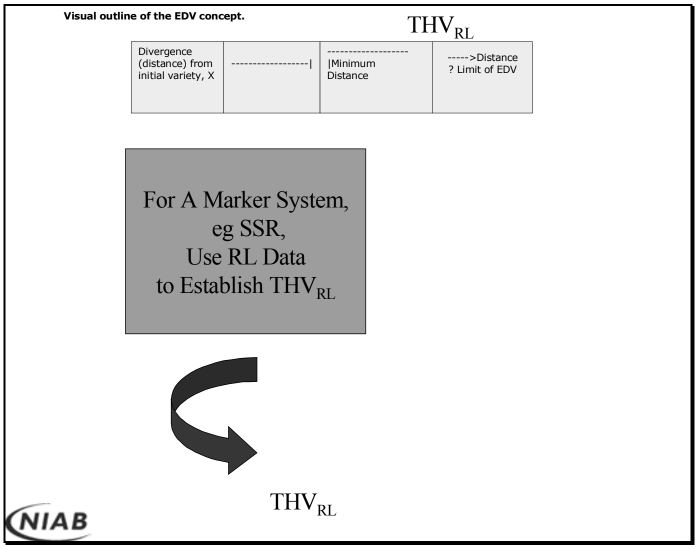
---

Visual outline of the EDV concept.

Divergence (distance) from initial variety, X	------------------|	------------------- |Minimum Distance	----->Distance ? Limit of EDV
Type of variety:	X*plagiarised X	X' new, EDV (dependent)	Y new (independent)
Differentiation from X	insignificant 'cosmetic' changes	clearly distinct, but retains 'essential characteristics'	Clearly distinct
Eligibility for PBR	No	Yes (essentially derived)	Yes

Visual outline of the EDV concept.

Divergence (distance) from initial variety, X	------------------|	------------------ |Minimum Distance	----->Distance ? Limit of EDV



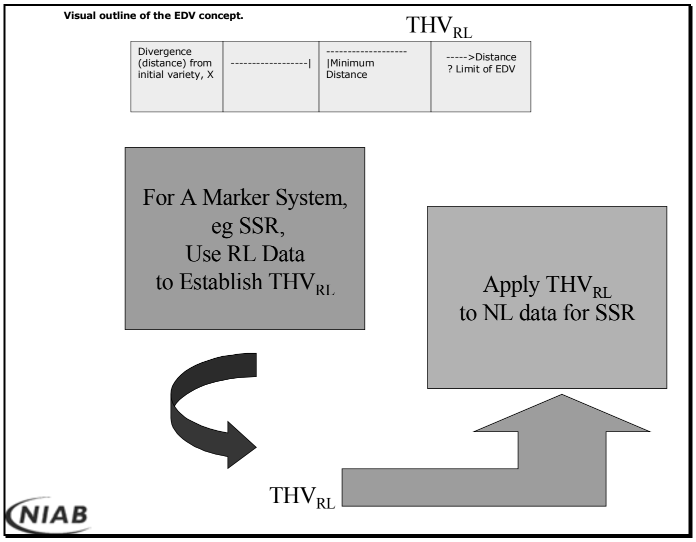
Visual outline of the EDV concept.

THV_{RL}

Divergence (distance) from initial variety, X	------------------|	------------------ |Minimum Distance	----->Distance ? Limit of EDV

For A Marker System,
eg SSR,
Use RL Data
to Establish THV_{RL}

THV_{RL}



Visual outline of the EDV concept.

THV_{RL}

Divergence (distance) from initial variety, X	------------------|	------------------ |Minimum Distance	----->Distance ? Limit of EDV

For A Marker System,
eg SSR,
Use RL Data
to Establish THV_{RL}

Apply THV_{RL}
to NL data for SSR

THV_{RL}

Essential Derivation:
To simulate an EDV situation from the computed matrices of pair-wise similarities/distances, the following steps were taken.

Firstly the RL material only was utilised to establish a framework whereby varieties are distinct, but potential EDV cases may occur. All pair-wise similarity coefficients were collated, ranked and the value corresponding to a threshold value (THV_{RL}) of the top 5-percentile determined (this potential threshold was based on previous work with barley). This was repeated for matrices from each marker system.

Secondly, the set of NL pair-wise similarities were then compared to the THV established by the RL varieties as defined above.

Statistics were then computed for the percentage agreement between the marker systems in terms of the identification of possible EDV cases.

Essential Derivation:
To simulate an EDV situation from the computed matrices of pair-wise similarities/distances, the following steps were taken.

Firstly the RL material only was utilised to establish a framework whereby varieties are distinct, but potential EDV cases may occur.
All pair-wise similarity coefficients were collated, ranked and the value corresponding to a threshold value ($\mathrm{THV}_{\mathrm{RL}}$) of the top 5-percentile determined (this potential threshold was based on previous work with barley). This was repeated for matrices from each marker system.

Secondly, the set of NL pair-wise similarities were then compared to the THV established by the RL varieties as defined above.

Statistics were then computed for the percentage agreement between the marker systems in terms of the identification of possible EDV cases.

Essential Derivation:
To simulate an EDV situation from the computed matrices of pair-wise similarities/distances, the following steps were taken.

Firstly the RL material only was utilised to establish a framework whereby varieties are distinct, but potential EDV cases may occur.
All pair-wise similarity coefficients were collated, ranked and the value corresponding to a threshold value ($\mathrm{THV}_{\mathrm{RL}}$) of the top 5-percentile determined (this potential threshold was based on previous work with barley). This was repeated for matrices from each marker system.

Secondly, the set of NL pair-wise similarities were then compared to the THV established by the RL varieties as defined above.

Statistics were then computed for the percentage agreement between the marker systems in terms of the identification of possible EDV cases.

Essential Derivation:
To simulate an EDV situation from the computed matrices of pair-wise similarities/distances, the following steps were taken.

Firstly the RL material only was utilised to establish a framework whereby varieties are distinct, but potential EDV cases may occur.
All pair-wise similarity coefficients were collated, ranked and the value corresponding to a threshold value ($\mathrm{THV}_{\mathrm{RL}}$) of the top 5-percentile determined (this potential threshold was based on previous work with barley). This was repeated for matrices from each marker system.

Secondly, the set of NL pair-wise similarities were then compared to the THV established by the RL varieties as defined above.

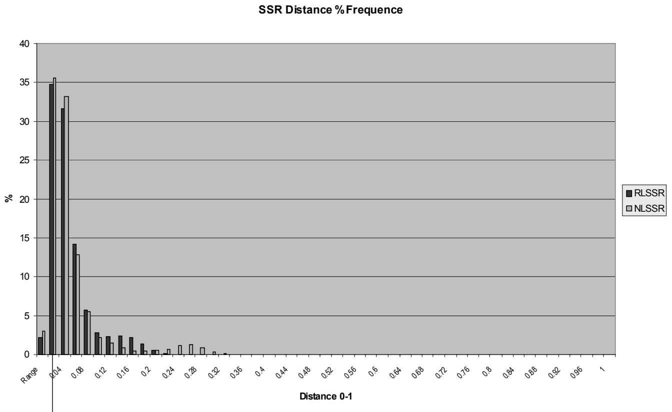
Statistics were then computed for the percentage agreement between the marker systems in terms of the identification of possible EDV cases.

Essential Derivation:
To simulate an EDV situation from the computed matrices of pair-wise similarities/distances, the following steps were taken.

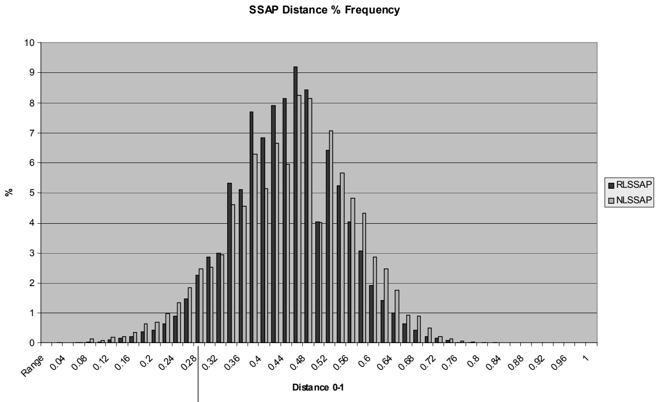
Firstly the RL material only was utilised to establish a framework whereby varieties are distinct, but potential EDV cases may occur. All pair-wise similarity coefficients were collated, ranked and the value corresponding to a threshold value (THV_{RL}) of the top 5-percentile determined (this potential threshold was based on previous work with barley). This was repeated for matrices from each marker system.

Secondly, the set of NL pair-wise similarities were then compared to the THV established by the RL varieties as defined above.

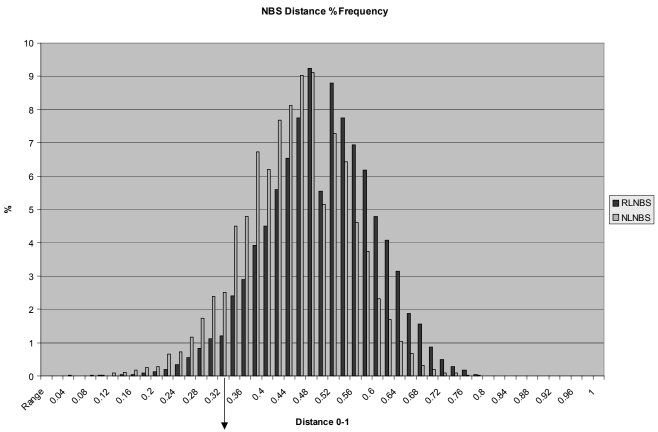
Statistics were then computed for the percentage agreement between the marker systems in terms of the identification of possible EDV cases.



5th Percentile
(THV_{RL}=0.024)



5th Percentile
(THV_{RL}=0.029)



5th Percentile
(THV_{RL}=0.34)

Application of NL Data to Values of THV_{RL}

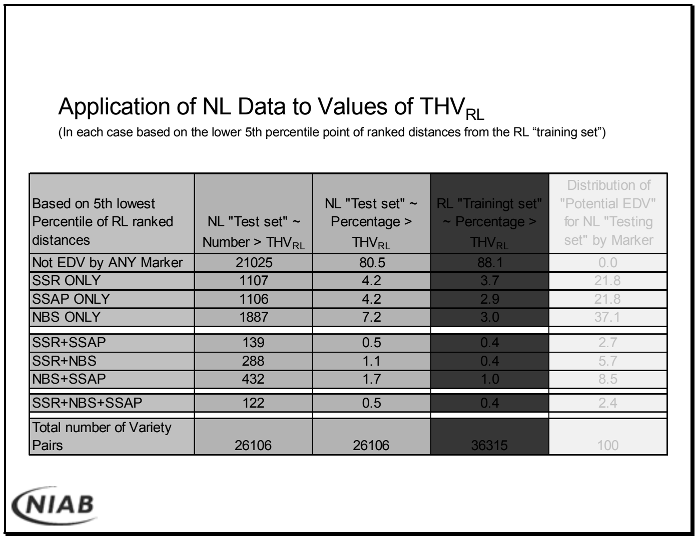
(In each case based on the lower 5th percentile point of ranked distances from the RL “training set”)

Marker	Distance THV_{RL}	Number < THV_{RL}	Percentage < THV_{RL}
SSR	0.0236	1656	6.3
S_SAP	0.2857	1799	6.9
NBS	0.3409	2729	10.5

Application of RL Data to Values of THV_{RL}

(In each case based on the lower 5th percentile point of ranked distances from the RL “training set”)

Based on 5th lowest Percentile of RL ranked distances	NL "Test set" ~ Number > THV_{RL}	NL "Test set" ~ Percentage > THV_{RL}	RL "Trainingt set" ~ Percentage > THV_{RL}	Distribution of "Potential EDV" for NL "Testing set" by Marker
Not EDV by ANY Marker	21025	80.5	88.1	0.0
SSR ONLY	1107	4.2	3.7	21.8
SSAP ONLY	1106	4.2	2.9	21.8
NBS ONLY	1887	7.2	3.0	37.1
SSR+SSAP	139	0.5	0.4	2.7
SSR+NBS	288	1.1	0.4	5.7
NBS+SSAP	432	1.7	1.0	8.5
SSR+NBS+SSAP	122	0.5	0.4	2.4
Total number of Variety Pairs	26106	26106	36315	100

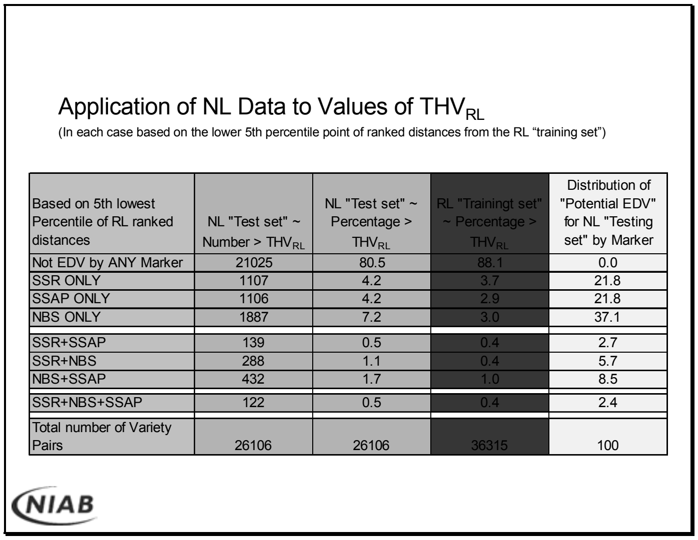


Application of NL Data to Values of THV_{RL}

(In each case based on the lower 5th percentile point of ranked distances from the RL “training set”)

Based on 5th lowest Percentile of RL ranked distances	NL "Test set" ~ Number > THV_{RL}	NL "Test set" ~ Percentage > THV_{RL}	RL "Trainingt set" ~ Percentage > THV_{RL}	Distribution of "Potential EDV" for NL "Testing set" by Marker
Not EDV by ANY Marker	21025	80.5	88.1	0.0
SSR ONLY	1107	4.2	3.7	21.8
SSAP ONLY	1106	4.2	2.9	21.8
NBS ONLY	1887	7.2	3.0	37.1
SSR+SSAP	139	0.5	0.4	2.7
SSR+NBS	288	1.1	0.4	5.7
NBS+SSAP	432	1.7	1.0	8.5
SSR+NBS+SSAP	122	0.5	0.4	2.4
Total number of Variety Pairs	26106	26106	36315	100

NIAB

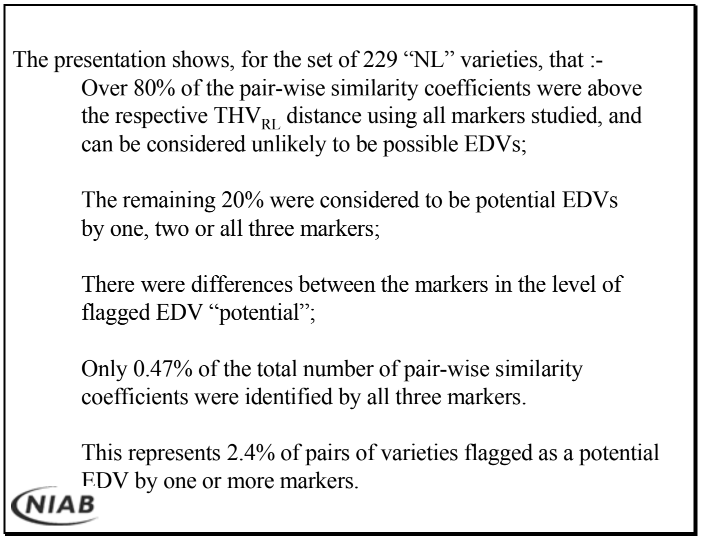


Application of NL Data to Values of THV_{RL}

(In each case based on the lower 5th percentile point of ranked distances from the RL “training set”)

Based on 5th lowest Percentile of RL ranked distances	NL "Test set" ~ Number > THV_{RL}	NL "Test set" ~ Percentage > THV_{RL}	RL "Trainingt set" ~ Percentage > THV_{RL}	Distribution of "Potential EDV" for NL "Testing set" by Marker
Not EDV by ANY Marker	21025	80.5	88.1	0.0
SSR ONLY	1107	4.2	3.7	21.8
SSAP ONLY	1106	4.2	2.9	21.8
NBS ONLY	1887	7.2	3.0	37.1
SSR+SSAP	139	0.5	0.4	2.7
SSR+NBS	288	1.1	0.4	5.7
NBS+SSAP	432	1.7	1.0	8.5
SSR+NBS+SSAP	122	0.5	0.4	2.4
Total number of Variety Pairs	26106	26106	36315	100

NIAB



The presentation shows, for the set of 229 “NL” varieties, that :-

Over 80% of the pair-wise similarity coefficients were above the respective THV_{RL} distance using all markers studied, and can be considered unlikely to be possible EDVs;

The remaining 20% were considered to be potential EDVs by one, two or all three markers;

There were differences between the markers in the level of flagged EDV “potential”;

Only 0.47% of the total number of pair-wise similarity coefficients were identified by all three markers.

This represents 2.4% of pairs of varieties flagged as a potential EDV by one or more markers.

NIAB

The presentation shows, for the set of 229 “NL” varieties, that :-

Over 80% of the pair-wise similarity coefficients were above the respective THV_{RL} distance using all markers studied, and can be considered unlikely to be possible EDVs;

The remaining 20% were considered to be potential EDVs by one, two or all three markers;

There were differences between the markers in the level of flagged EDV “potential”;

Only 0.47% of the total number of pair-wise similarity coefficients were identified by all three markers.

This represents 2.4% of pairs of varieties flagged as a potential EDV by one or more markers.

The presentation shows, for the set of 229 “NL” varieties, that :-

Over 80% of the pair-wise similarity coefficients were above the respective THV_{RL} distance using all markers studied, and can be considered unlikely to be possible EDVs;

The remaining 20% were considered to be potential EDVs by one, two or all three markers;

There were differences between the markers in the level of flagged EDV “potential”;

Only 0.47% of the total number of pair-wise similarity coefficients were identified by all three markers.

This represents 2.4% of pairs of varieties flagged as a potential EDV by one or more markers.

The presentation shows, for the set of 229 “NL” varieties, that :-

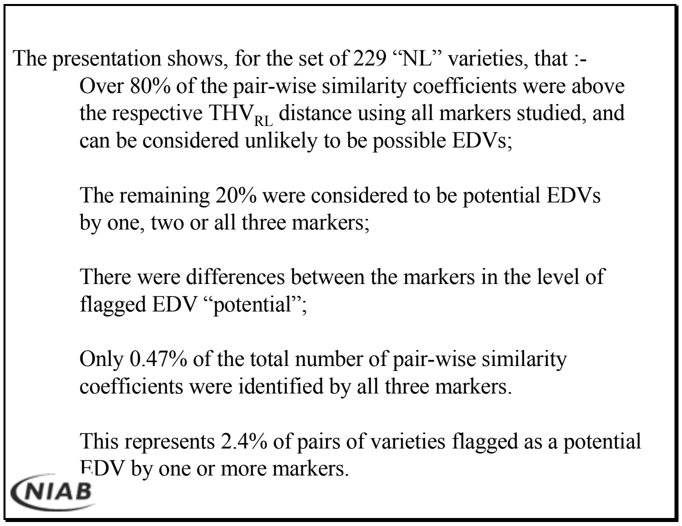
Over 80% of the pair-wise similarity coefficients were above the respective THV_{RL} distance using all markers studied, and can be considered unlikely to be possible EDVs;

The remaining 20% were considered to be potential EDVs by one, two or all three markers;

There were differences between the markers in the level of flagged EDV “potential”;

Only 0.47% of the total number of pair-wise similarity coefficients were identified by all three markers.

This represents 2.4% of pairs of varieties flagged as a potential EDV by one or more markers.



The presentation shows, for the set of 229 “NL” varieties, that :-

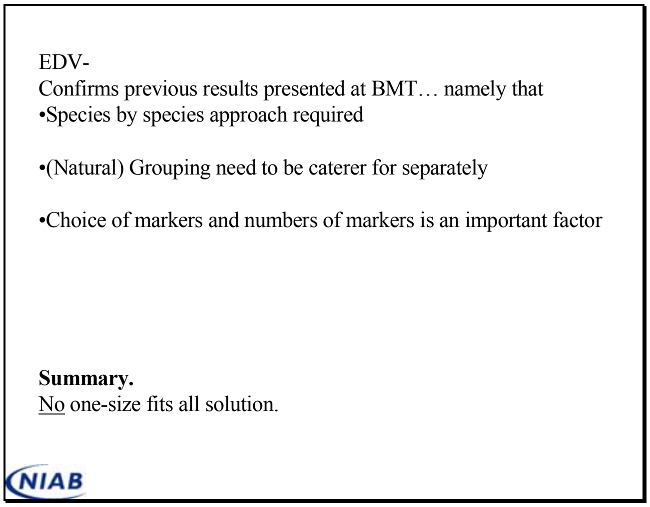
Over 80% of the pair-wise similarity coefficients were above the respective THV_{RL} distance using all markers studied, and can be considered unlikely to be possible EDVs;

The remaining 20% were considered to be potential EDVs by one, two or all three markers;

There were differences between the markers in the level of flagged EDV “potential”;

Only 0.47% of the total number of pair-wise similarity coefficients were identified by all three markers.

This represents 2.4% of pairs of varieties flagged as a potential EDV by one or more markers.

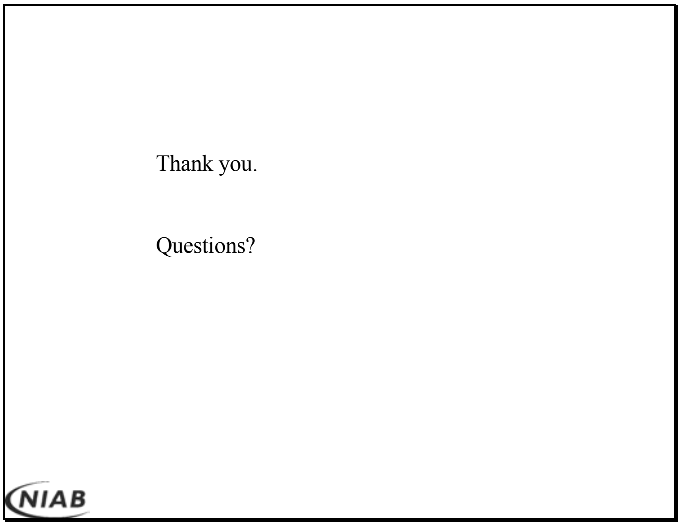


EDV-
Confirms previous results presented at BMT… namely that
•Species by species approach required

•(Natural) Grouping need to be caterer for separately

•Choice of markers and numbers of markers is an important factor

Summary.
No one-size fits all solution.



Thank you.

Questions?

[End of document]