



**BMT/11/10 Rev. Add.**

**ORIGINAL:** English

**DATE:** October 9, 2008

**INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS**  
GENEVA




**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 2: THE DATABASE (REVISED)**

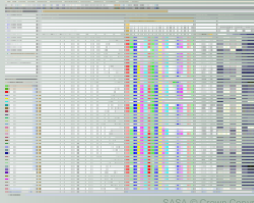
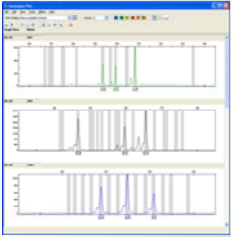

*Document prepared by experts from the United Kingdom*





**CONSTRUCTION OF AN INTEGRATED  
MICROSATELLITE AND KEY  
MORPHOLOGICAL CHARACTERISTIC  
DATABASE OF POTATO VARIETIES ON  
THE EU COMMON CATALOGUE**

**PART 2: THE DATABASE**

**Alex Reid & Lysbeth Hof**



SASA © Crown Copyright



## Introduction

During the course of the project a large amount of different types of data were generated

- Binary data for the 9 SSR markers
- Multi-state data for the morphological characteristics
- Textual data for the information about each sample
- Lightsprout photographs

SASA © Crown Copyright



## Requirements for the software

The software had to be able to store and analyse different types of data

It had to be able to analyse these data in different combinations

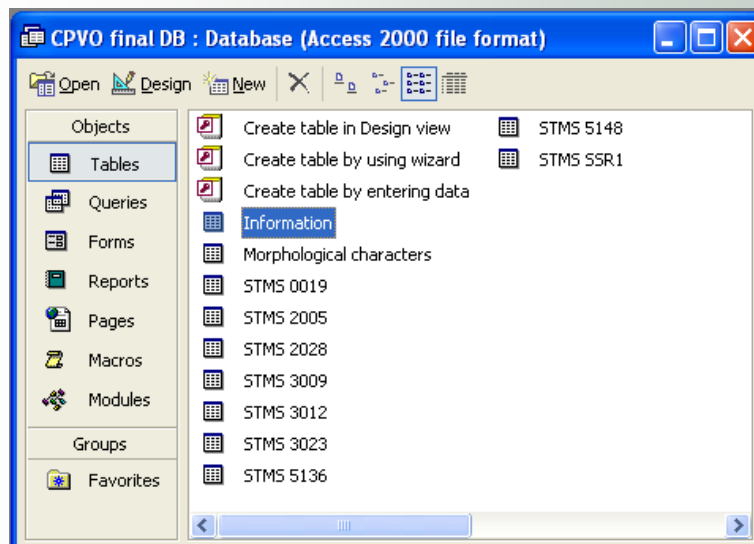
A highly desirable feature was the ability to use the database to identify unknown samples rapidly and accurately

The solution was to store the data in an Access database which was linked to BioNumerics via an ODBC (Open DataBase Connectivity) for analysis

SASA © Crown Copyright



## Access database structure



SASA © Crown Copyright



## The Information Table

The structure of the Access database was specifically designed so that it could serve BioNumerics

BioNumerics requires each entry to have a unique identifier or Key

The variety name could not be used for the DNA samples as some varieties had samples from numerous sources

For example cv. Ditta had samples submitted for SSR analysis from the Netherlands, Poland and the UK and therefore the DNA sample codes were used as the Key (in this case NL-089, PL-021 and UK-0598)

There is only one entry per country for the morphological data for a variety and in these cases the Key is the variety name followed by the country of origin (e.g. Ditta\_DE, Ditta\_NL, Ditta\_PL and Ditta\_UK)

SASA © Crown Copyright



## The Information Table Fields

Key	Unique identifier for use in BioNumerics
Variety Denomination	The name of the variety
Origin of sample e.g. Breeder (B), office (O) or other (T)	Where the sample was obtained from either the breeder or from one of the partners collections or from another source
Submitting office (S SASA, B BSA, C COBORU, N Naktuinbouw)	The office which submitted the sample for DNA analysis or submitted the morphological description
Harvest year	The year the sample analyzed was harvested

SASA © Crown Copyright



## The Information Table Fields

DNA extraction laboratory (S SASA, N Naktuinbouw)	The laboratory the DNA sample was extracted (either UK or Netherlands)
Extraction year	The year the DNA extraction was made
Place of storage of DNA sample (S SASA, N Naktuinbouw)	The place where the DNA sample is kept in long term storage
SSR analysis performed at (S SASA, N Naktuinbouw)	The laboratory where the SSR analysis was performed
SSR analysis year	The year the SSR analysis was performed

SASA © Crown Copyright



## The Information Table Fields

Technical protocol used for morphological description	The technical protocol used for the morphological description (if known)
Description year (either before 1995 or actual year)	The year that the latest description was carried out, N.B. not necessarily from official description
Place description carried out	The office the description was carried out at
Photograph availability (and link to photograph)	The file name for the photograph (if available) and hyperlink

SASA © Crown Copyright



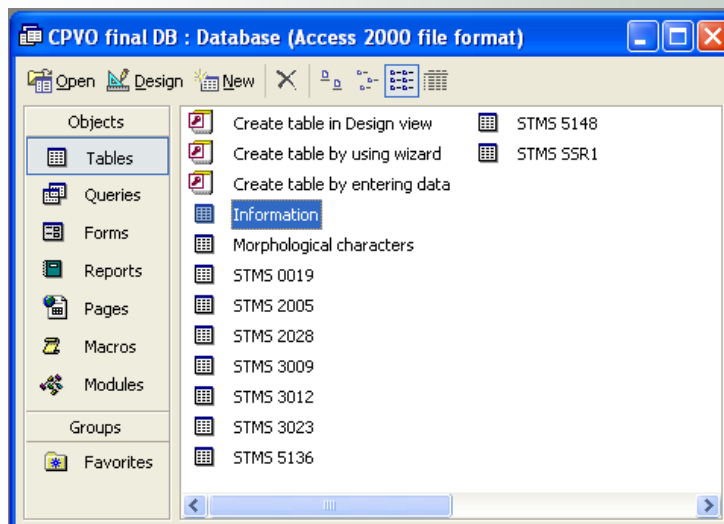
## The Information Table Fields

Place photograph taken (office)	The office the photograph was taken at
Photograph year (if known)	The year the photograph was taken
National Reference	The national reference number of the sample (if there is one)
Comments	Any other comments (e.g. is the variety a mutant of another variety)

SASA © Crown Copyright



## Access database structure



SASA © Crown Copyright



## Examples of morphological and SSR data

Key	Variety	1	2	3	4	5	6	7	8	9	10	11	24	28	29	31	32	34	35	36
Av_NL	Av	5	2	5	1	5	4	4	2	5	4	4	5	4	1	5	3	2	2	3
Ave_NL	Ave	7	2	7	3	4	6	3	3	5	4	5	4	1		8	2	2	2	1
Ax_DE	Ax	5	4	5	1	3	4	3	1	2	6	3								
Ba_DE	Ba	3	3	4	2	5	5	3	2	3	5	4								
Bal_NL	Bal	6	2	4	1	3	5	3	3	3	3	3	5	1		7	4	2	2	3

Key	Variety	A	B	C	D	E	F	G	H	I	J
UK-0523	Arran B	0	0	0	0	0	1	1	0	0	0
UK-0524	Arran C	0	1	0	0	0	1	1	0	0	0
UK-0525	Arran C	0	0	0	0	0	1	1	0	0	0
UK-0526	Arran P	0	1	0	1	0	1	0	0	0	0
UK-0527	Arran V	0	0	0	0	0	1	1	0	0	0

SASA © Crown Copyright



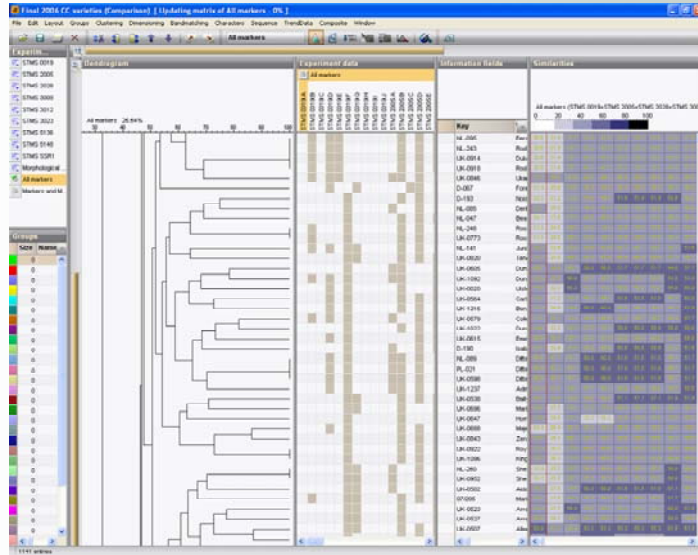
## BioNumerics database

The screenshot displays the BioNumerics software interface. The main window shows a table of database entries with columns for Key, Variety denomination, Or., Sub., Harv., DNA, Extr., Pla., SSR, and SSR a. The table lists various varieties such as Anya, Argos, Arhala, Amora, Arran Banner, Arran Comet, Arran Consul, Arran Blue, Arran Victory, Aspergera, Astafe, Azzurri, Alice, Aurora, Avondale, Ballydoon, Balmoral, Barba, Baraka, Barina, Barina, Belle de Fontenay, Berber, Bright, BF15, Bonarda, Bonella, Bonika, Bright, British Queen, Brodia, and Brora. The interface also includes a menu bar, a toolbar, and a right-hand panel with sections for Experiments, Files, Comparisons, and Entry relations.

copyright



## SSR data tree



SA © Crown Copyright




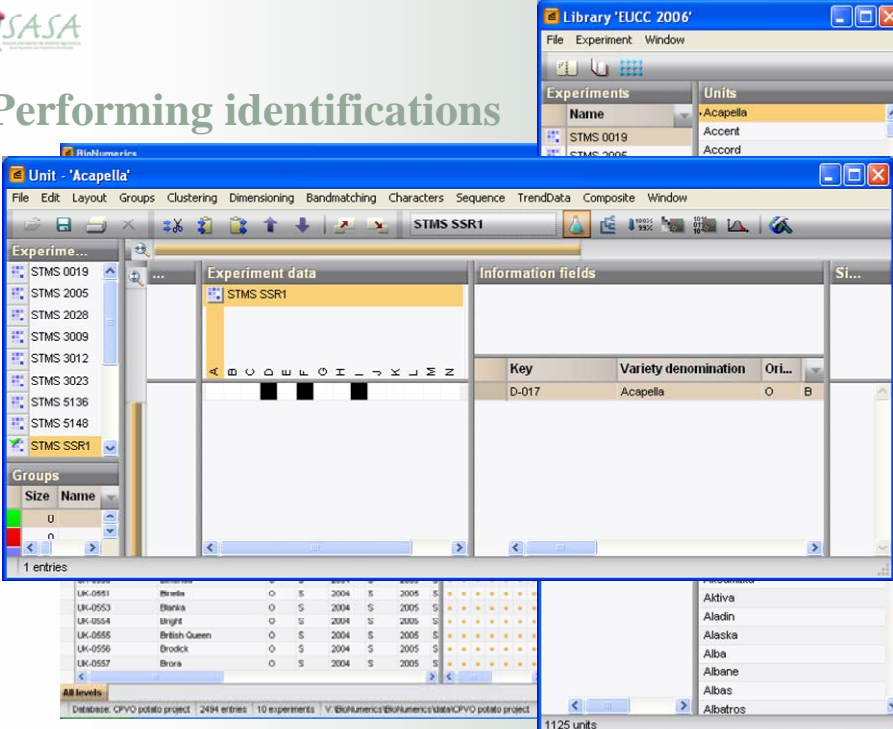
## Morphological data tree



SA © Crown Copyright



 **Performing identifications**



Library 'EUCC 2006'

Experiments

Name	Units
STMS 0019	Acapella
STMS 2005	Accent
STMS 2005	Accord

Unit - 'Acapella'

Experiment data

Key	Variety denomination	Ori...
D-017	Acapella	O B

Information fields


Key: D-017, Variety denomination: Acapella, Ori...: O B

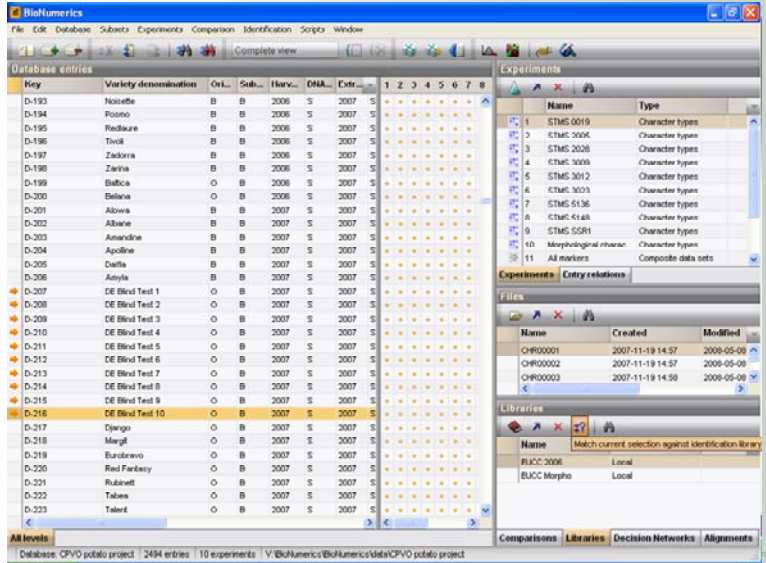
1 entries

Key	Variety denomination	Ori...	2004	2005	2006	2007
UK-0591	Birella	O S	2004 S	2005 S		
UK-0553	Bianka	O S	2004 S	2005 S		
UK-0554	blight	O S	2004 S	2005 S		
UK-0555	British Queen	O S	2004 S	2005 S		
UK-0556	Brodick	O S	2004 S	2005 S		
UK-0557	Brona	O S	2004 S	2005 S		

Database: CPVO potato project | 2494 entries | 10 experiments | V: BioNumerics\BioNumerics\data\CPVO potato project

1125 units

 **Performing identifications**



BioNumerics

Database entries

Key	Variety denomination	Ori...	Sub...	Harv...	DNA...	Exp...
D-193	Nouette	B	B	2006	S	2007 S
D-194	Poisino	B	B	2006	S	2007 S
D-195	Redeare	B	B	2006	S	2007 S
D-196	Tivol	B	B	2006	S	2007 S
D-197	Zakorra	B	B	2006	S	2007 S
D-198	Zarina	B	B	2006	S	2007 S
D-199	Belfica	O	B	2006	S	2007 S
D-200	Belfica	O	B	2006	S	2007 S
D-201	Albana	B	B	2007	S	2007 S
D-202	Albane	B	B	2007	S	2007 S
D-203	Amandine	B	B	2007	S	2007 S
D-204	Apoline	B	B	2007	S	2007 S
D-205	Daila	B	B	2007	S	2007 S
D-206	Amyla	B	B	2007	S	2007 S
D-207	DE Blind Test 1	O	B	2007	S	2007 S
D-208	DE Blind Test 2	O	B	2007	S	2007 S
D-209	DE Blind Test 3	O	B	2007	S	2007 S
D-210	DE Blind Test 4	O	B	2007	S	2007 S
D-211	DE Blind Test 5	O	B	2007	S	2007 S
D-212	DE Blind Test 6	O	B	2007	S	2007 S
D-213	DE Blind Test 7	O	B	2007	S	2007 S
D-214	DE Blind Test 8	O	B	2007	S	2007 S
D-215	DE Blind Test 9	O	B	2007	S	2007 S
D-216	DE Blind Test 10	O	B	2007	S	2007 S
D-217	Djanga	O	B	2007	S	2007 S
D-218	Margl	O	B	2007	S	2007 S
D-219	Eurobravo	O	B	2007	S	2007 S
D-220	Real Fantasy	O	B	2007	S	2007 S
D-221	Rubinet	O	B	2007	S	2007 S
D-222	Tabea	O	B	2007	S	2007 S
D-223	Talent	O	B	2007	S	2007 S

Database: CPVO potato project | 2494 entries | 10 experiments | V: BioNumerics\BioNumerics\data\CPVO potato project

Experiments

Name	Type
1 STMS 0019	Character types
2 EUCC 2006	Character types
3 STMS 2008	Character types
4 EUCC 2009	Character types
5 STMS 3012	Character types
6 STMS 5136	Character types
7 EUCC 4148	Character types
8 STMS SSR1	Character types
9 Morphological markers	Character types
10 All markers	Composite data sets

Experiments Entry relations



Name	Created	Modified
CHR00001	2007-11-19 14:57	2009-05-08
CHR00002	2007-11-19 14:57	2009-05-08
CHR00003	2007-11-19 14:58	2009-05-08

Libraries

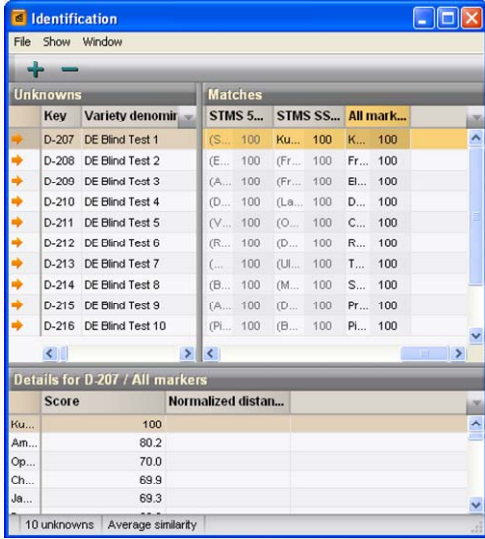
Name	Match current selection against identification library
EUCC 2006	Local
EUCC Morpho	Local

Database: CPVO potato project | 2494 entries | 10 experiments | V: BioNumerics\BioNumerics\data\CPVO potato project

own Copyright



## Library screen identification report





The screenshot shows a software window titled 'Identification' with a menu bar (File, Show, Window) and a toolbar. The main area is divided into two panes: 'Unknowns' and 'Matches'. The 'Unknowns' pane contains a table with columns 'Key' and 'Variety denomin...'. The 'Matches' pane contains a table with columns 'STMS 5...', 'STMS SS...', and 'All mark...'. Below these panes is a 'Details for D-207 / All markers' section with a table showing 'Score' and 'Normalized distan...'. At the bottom, it says '10 unknowns Average similarity'.

Key	Variety denomin...	STMS 5...	STMS SS...	All mark...
D-207	DE Blind Test 1	(S... 100	Ku... 100	K... 100
D-208	DE Blind Test 2	(E... 100	(Fr... 100	Fr... 100
D-209	DE Blind Test 3	(A... 100	(Fr... 100	El... 100
D-210	DE Blind Test 4	(D... 100	(La... 100	D... 100
D-211	DE Blind Test 5	(V... 100	(O... 100	C... 100
D-212	DE Blind Test 6	(R... 100	(D... 100	R... 100
D-213	DE Blind Test 7	(... 100	(U... 100	T... 100
D-214	DE Blind Test 8	(B... 100	(M... 100	S... 100
D-215	DE Blind Test 9	(A... 100	(D... 100	Pr... 100
D-216	DE Blind Test 10	(Pl... 100	(B... 100	Pl... 100

Score	Normalized distan...
Ku...	100
Am...	80.2
Op...	70.0
Ch...	69.9
Ja...	69.3

10 unknowns Average similarity

SASA © Crown Copyright



## Practical uses

At SASA we now perform 100s of variety tests every year for

- Official field inspectors
- Breeders (national and international)
- Regulatory bodies

Cases have included

- Verification of varieties for export (when there have been queries)
- Labelling infringements in supermarkets
- Illegal dumping of potatoes in river systems
- Cannabis smuggling!

SASA © Crown Copyright



## Summary

During the course of the project a large amount of different types of data were generated

These data are held in an Access database

Which is linked to BioNumerics via ODBC

Clustering analysis can be performed on single SSR markers or the morphological data or any combination of these

A library was created against which unknown samples can be screened resulting in rapid and accurate identifications

SASA © Crown Copyright



## Without the help of...

Project partners

BSA - Beate Rücker and Swenja Tams

COBORU - Ewa Milczynska and Julia Borys

CPVO - Anne Weitz

SASA – Heather Campbell, Arlene Cameron and Sandra Goodfellow

**None of this would have been possible!**

SASA © Crown Copyright