EURISCO and CCDBs
coverage and overlap

Theo van Hintum

Centre for Genetic Resources, The Netherlands (CGN)

EURISCO and CCDBs: coverage and overlap

- objective of the study
  - compare the coverage of EURISCO and the CCDBs
    - estimate the number of accessions in Europe
      - determine what is missing in EURISCO
    - see what the CCDBs add to EURISCO

EURISCO and CCDBs: coverage and overlap

- definition of the domain
  - PGR accessions maintained *ex situ* in Europe
    - accessions of PGRFA maintained in public or other collections for public use
      - no working collections of companies
      - no research collections (mapping populations etc.)
    - CCDBs apply various definitions
      - some require availability of the material

EURISCO and CCDBs: coverage and overlap

- approach
  - data collection
    - list of CCDBs extracted from ECPGR website (11-11-2013)
      - 54 databases (48 CCDBs)
        - 33 CCDBs downloaded + 13 received on request
    - EURISCO dump received on request (19-11-2013)
    - FAO-INST codes & GRIN-Tax data downloaded (11-11-2013)

EURISCO and CCDBs: coverage and overlap

- approach
  - data preparation - same approach EURISCO and CCDBs
    - data copied in spreadsheet
    - records counted per INSTCODE, GENUS, SPECIES & SUBTAXA
    - INSTCODEs checked and added or replaced
    - non-European records removed
    - GENUS, SPECIES & SUBTAXA matched to GRIN-Tax and valid taxon was recorded
      - non-matches were inspected and corrected
    - records summed per institute/taxon combination

EURISCO and CCDBs: coverage and overlap

- approach - comparison
  - data analysis
    - EURISCO sums were matched with CCDB sums
      - if institute/taxon combination in more than one CCDB, manual assignment
        - mainly *Medicago* and *Solanum*
        - e.g. '*Medicago radiata'* occurred in both 'Annual Medicago' and 'Minor Forage Legumes' CCDBs
        - e.g. '*Solanum* spp.' occurred in 'Eggplant', 'Pepino', 'Tree Tomato' and 'Wild Potato' CCDBs
      - if >20 records were available from a country/taxon but different institutes this was checked
        - e.g. material transferred in Germany from Braunschweig DEU001 to Gatersleben DEU146 after the merger of the two

EURISCO and CCDBs: coverage and overlap

- approach - number existing accessions
  - number of records per institute/taxon combination = sum in EURISCO or, if not EURISCO, sum in CCDB
    - see analysis of age of database records
  - gap per institute / taxon combination = difference between sum in CCDB and sum in EURISCO (only if this is positive)
  - some additional calculations

EURISCO and CCDBs: coverage and overlap

- approach
  - data analysis
    - part of resulting data sheet

DB	INST	TAXON	CCDBs	EURISCO	PROBA	GAP
3	FRA179	Solanum tuberosum	883	0	883	883
10	GBR004	Stylosanthes spp.	2	2	2	0
19	SVK001	Aegilops triuncialis	2	2	2	0
5	HRV041	Linum usitatissimum	0	2	2	0
	ESP048	Musa spp.	0	10	10	0
3	FRA179	Solanum tuberosum	883	0	883	883
21	DEU146	Allium akaka	3	3	3	0
	CZE122	Elymus drobovii	0	1	1	0
10	GBR004	Vicia ervilia	2	2	2	0
36	BEL094	Phleum pratense	5	6	6	0
	GBR004	Stipa clarazii	0	8	8	0
	ROM009	Vaccinium spp.	0	29	29	0
36	POL003	Phleum pratense	2570	2621	2621	0
10	ESP004	Vicia villosa	141	137	137	4

EURISCO and CCDBs: coverage and overlap

- results
  - number of accessions in the data sources
    - EURISCO: 1,065,766 accessions
      - 844,266 (79%) could be mapped on CCDB taxa
    - 46 CCDBs: 813,631 accessions
      - 635,918 (78%) could be matched to EURISCO on an institute / taxon basis
    - either in EURISCO or, if not in EURISCO, the CCDBs: 1,243,479 accessions

EURISCO and CCDBs: coverage and overlap

- results
  - age of records
    - EURISCO data included an 'upload date' descriptor
      - not *per se* the date of last update in NI
      - average age: 1.16 years, 'oldest' 10%: 4.96 years
      - oldest NIs: Slovenia (5.35 yr.), Turkey (5.23 yr.), Bulgaria (5.19 yr.), Georgia (5.05 yr.) and Lithuania (5.05 yr.)
      - half of the NIs (20 out of the 39 in EURISCO = 84% of accessions) were less than one year old
    - some CCDBs were very old
      - 9 CCDBs had data from genebank in Braunschweig (DEU001)
      - 2 largest CCDBs (wheat & barley = 40.9% of accessions) both used YUG and DEU001 codes
      - in some cases MS Access versions that are not supported by MicroSoft anymore were used

EURISCO and CCDBs: coverage and overlap

- results
  - number of accessions in Europe

	accessions in EURISCO	accessions not in EURISCO	total
taxa covered by the CCDBs	n_{11}	n_{12}	$n_{1.}$
taxa not covered by the CCDBs	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

EURISCO and CCDBs: coverage and overlap

- results
  - number of accessions in Europe

	accessions in EURISCO	accessions not in EURISCO	total
taxa covered by the CCDBs	844,266	177,713	1,021,979
taxa not covered by the CCDBs	221,500	n_{22}	$n_{2.}$
Total	1,065,766	$n_{.2}$	$n_{..}$

    - EURISCO contained 82.6% of the total number of accessions of the taxa covered by the CCDBs
      - assumption: same is true for the other taxa

EURISCO and CCDBs: coverage and overlap

- results
  - number of accessions in Europe

	accessions in EURISCO	accessions not in EURISCO	total
taxa covered by the CCDBs	844,266	177,713	1,021,979
taxa not covered by the CCDBs	221,500	46,624	268,124
Total	1,065,766	224,337	1,290,103

  - there are ~ 1.3 million PGRFA accessions in Europe

EURISCO and CCDBs: coverage and overlap

- results
  - number of accessions in Europe
    - weak spots in calculation:
      - CCDBs & EURISCO will not cover all PGRFA for the CCDB taxa
      - old CCDBs overestimate number of accessions
      - (less important) taxa not in CCDBs might also be missing in NIs
      - estimation of overlap contains errors
    - there are ~ 1.3 million PGRFA accessions in Europe

EURISCO and CCDBs: coverage and overlap

- results
  - gaps in EURISCO
    - 1.3 mio accessions in Europe, 1.1 mio accessions in EURISCO
    - where are the 224k missing accessions?
      - in the analysis of the gaps a total gap of 279k accessions was used due to the definition of gap

EURISCO and CCDBs: coverage and overlap

- results
  - gaps in EURISCO
    - 1.3 mio accessions in Europe, 1.1 mio accessions in EURISCO
    - largest gaps per country

Country	CCDBs	EURISCO	Estim.	Gap	%tot
Russia	141,920	123,430	174,228	90,598	33%
France	36,561	3,589	39,946	36,357	13%
UK	68,108	104,284	111,129	24,500	9%
Germany	135,515	155,784	164,906	22,309	8%
Italy	31,527	41,438	59,966	18,858	7%
Poland	58,169	67,915	77,496	14,585	5%
Israel	10,717	3,127	12,444	9,615	3%
Spain	43,473	71,694	73,754	8,834	3%
Turkey	13,934	12,996	19,654	6,843	2%
Swiss	16,257	39,787	44,758	5,653	2%

EURISCO and CCDBs: coverage and overlap

- results
  - gaps in EURISCO
    - 1.3 mio accessions in Europe, 1.1 mio accessions in EURISCO
    - largest gaps per institute

FAO-Inst	CCDBs	EURISCO	Estim.	Gap	%tot
RUS001	141,420	123,430	173,728	90,098	32%
DEU146	113,251	127,688	131,474	16,134	6%
GBR011	37,507	23,582	23,584	15,672	6%
FRA139	7,502		7,502	7,502	3%
FRA097	7,161		7,161	7,161	3%
ITA004	7,061		7,061	7,061	3%
TUR001	13,934	12,996	19,654	6,843	2%
ISR003	6,753		6,753	6,753	2%
FRA040	5,987		5,987	5,987	2%
POL030	5,522	227	4,870	5,348	2%

EURISCO and CCDBs: coverage and overlap

- results
  - gaps in EURISCO
    - 1.3 mio accessions in Europe, 1.1 mio accessions in EURISCO
    - largest gaps per crop

Crop	CCDBs	EURISCO	Estim.	Gap	%tot
Barley	115,885	86,837	120,996	57,009	20%
Wheat	168,278	147,988	154,653	46,331	17%
Vitis	36,662	29,244	50,380	24,525	9%
Avena	34,482	23,167	40,003	21,895	8%
Malus	21,013	26,123	41,263	16,479	6%
Pisum	32,503	26,630	35,900	11,860	4%
Cucurbits	27,486	26,440	33,157	8,986	3%
Pyrus	10,303	13,301	21,333	8,672	3%
Brassica	25,901	23,547	30,249	8,024	3%
Tomato	21,332	23,129	26,596	5,575	2%

EURISCO and CCDBs: coverage and overlap

- results
  - gaps in EURISCO
    - 1.3 mio accessions in Europe, 1.1 mio accessions in EURISCO
    - most missing accessions are
      - in VIR in Russia
      - in France
      - in institutions not in the NI of Italy and Israel

EURISCO and CCDBs: coverage and overlap

- conclusions
  - EURISCO has a pretty good coverage of the PGRFA accessions in Europe
    - 83% of accessions are covered
    - CCDB taxa: EURISCO has better coverage than CCDBs
      - 813,631 accessions in the CCDBs
      - 844,266 accessions in EURISCO
  - gaps in EURISCO can be identified
    - and, with a targeted approach, be closed !

Thank you for your attention !

Genetic resources are useful, pretty and tasteful

Genetic resources are the food on your plate

Thank you for your attention !