E

BMT/13/23 Add.
ORIGINAL: English
DATE: December 8, 2011

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

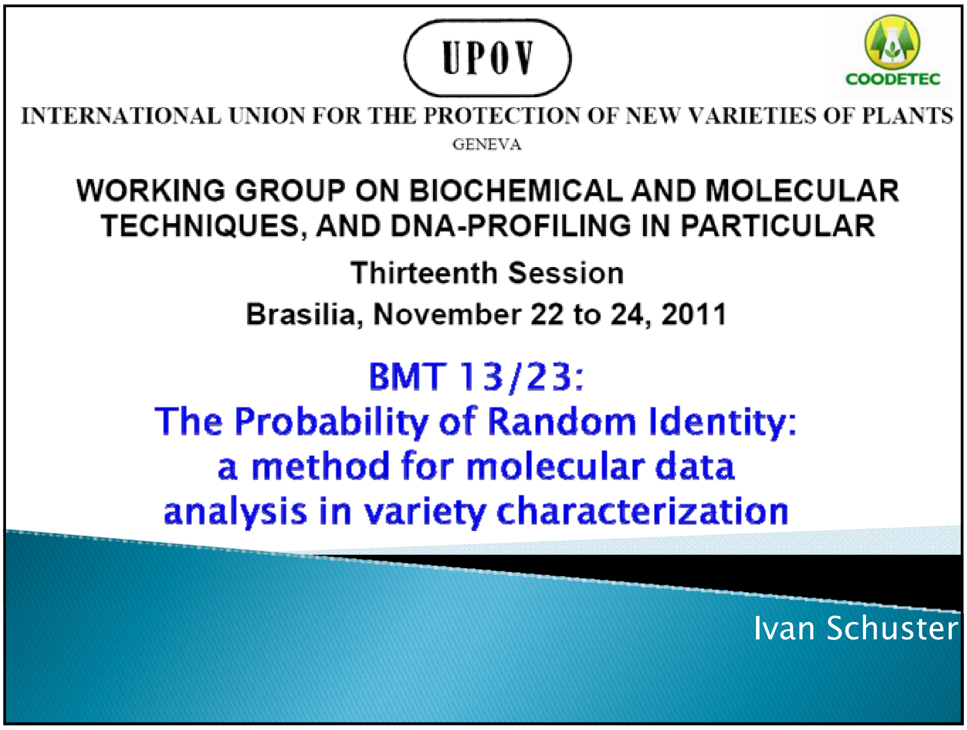
WORKING GROUP ON BIOCHEMICAL AND MOLECULR TECHNIQUES, AND DNA-PROFILING IN PARTICULAR

Thirteenth Session
Brasilia, November 22 to 24, 2011

ADDENDUM

THE PROBABILITY OF RANDOM IDENTITY: A METHOD FOR MOLECULAR DATA ANALYSIS IN VARIETY CHARACTERIZATION

Document prepared by experts from Brazil



UPOV

COODETEC

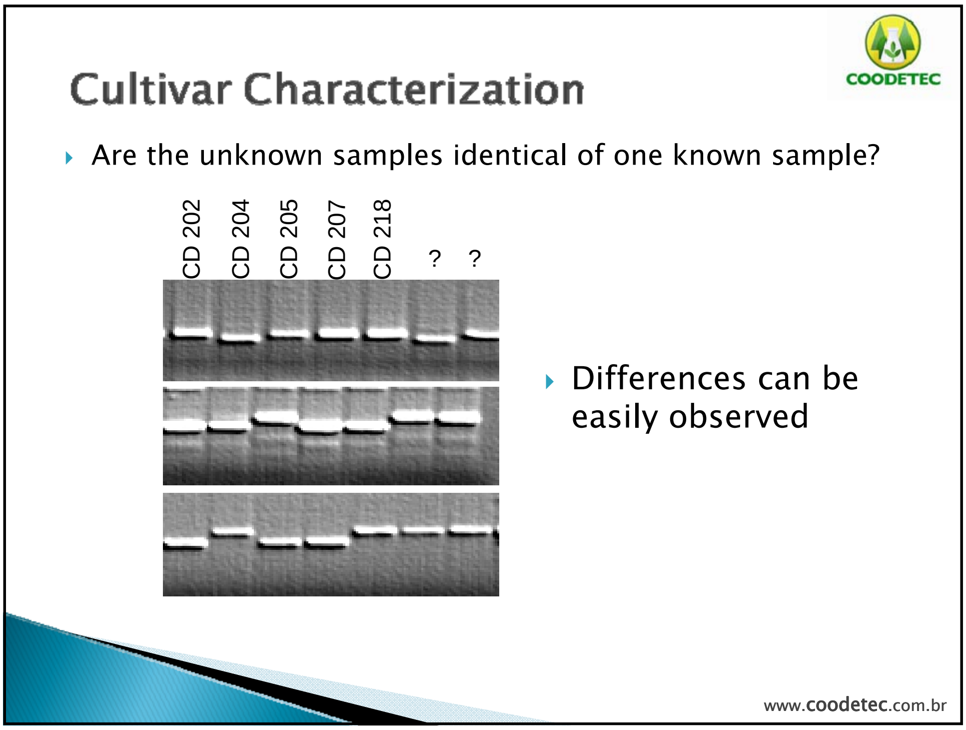
INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

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

Thirteenth Session
Brasilia, November 22 to 24, 2011

BMT 13/23:
The Probability of Random Identity: a method for molecular data analysis in variety characterization

Ivan Schuster



COODETEC

Cultivar Characterization

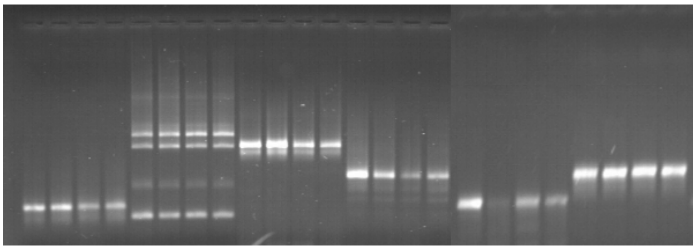
- Are the unknown samples identical of one known sample?

- Differences can be easily observed

www.coodetec.com.br

Cultivar Characterization

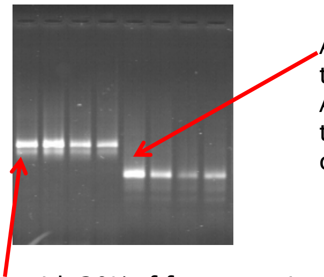
- Are the unknown samples identical of the known sample?



- With no differences:
  - How many markers should I evaluate to conclude for identity?

Cultivar Characterization

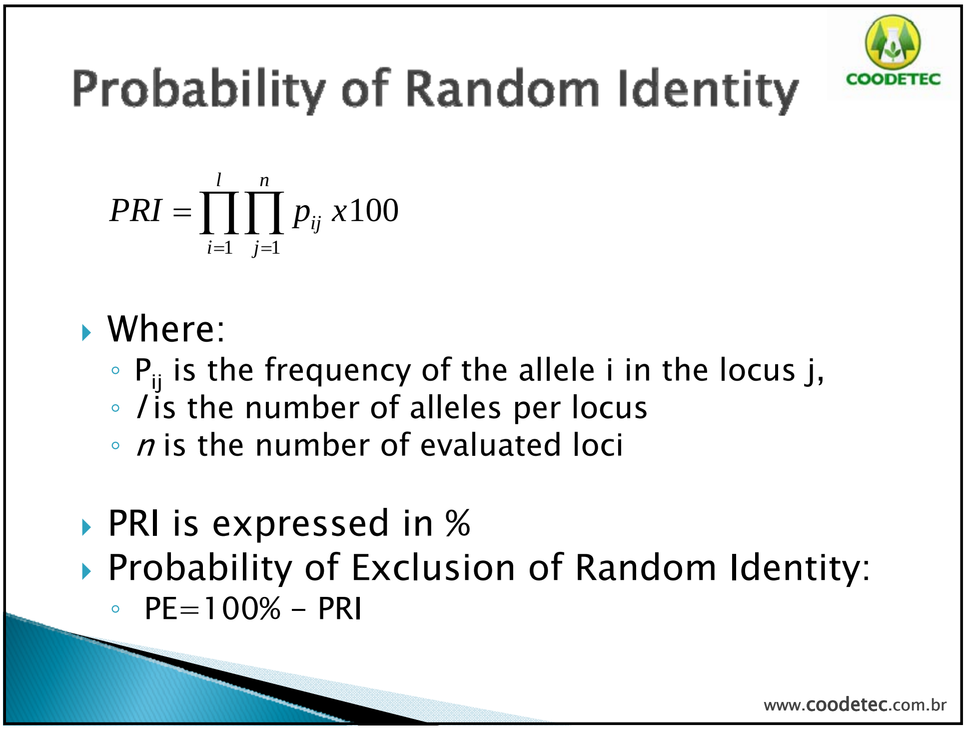
- Exemple:



Allele with 40% of frequency in the population of varieties. Any variety taken at random in the population has 40% chance of having this allele.

Allele with 30% of frequency in the population of varieties. Any variety taken at random in the population has 30% chance of having this allele.

Any variety taken at random in the population has 12% chance of having these two alleles.
30% x 40% = 12%

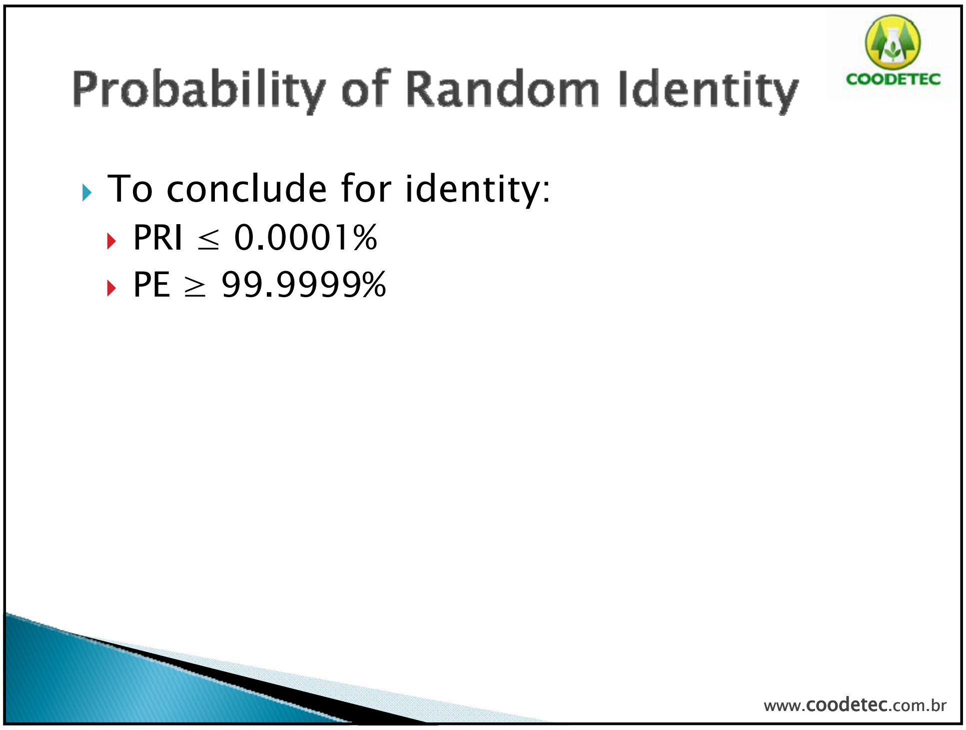


Probability of Random Identity

$$PRI = \prod_{i=1}^{l}\prod_{j=1}^{n} p_{ij}\, x100$$

- Where:
  - P_{ij} is the frequency of the allele i in the locus j,
  - *l* is the number of alleles per locus
  - *n* is the number of evaluated loci
- PRI is expressed in %
- Probability of Exclusion of Random Identity:
  - PE=100% – PRI

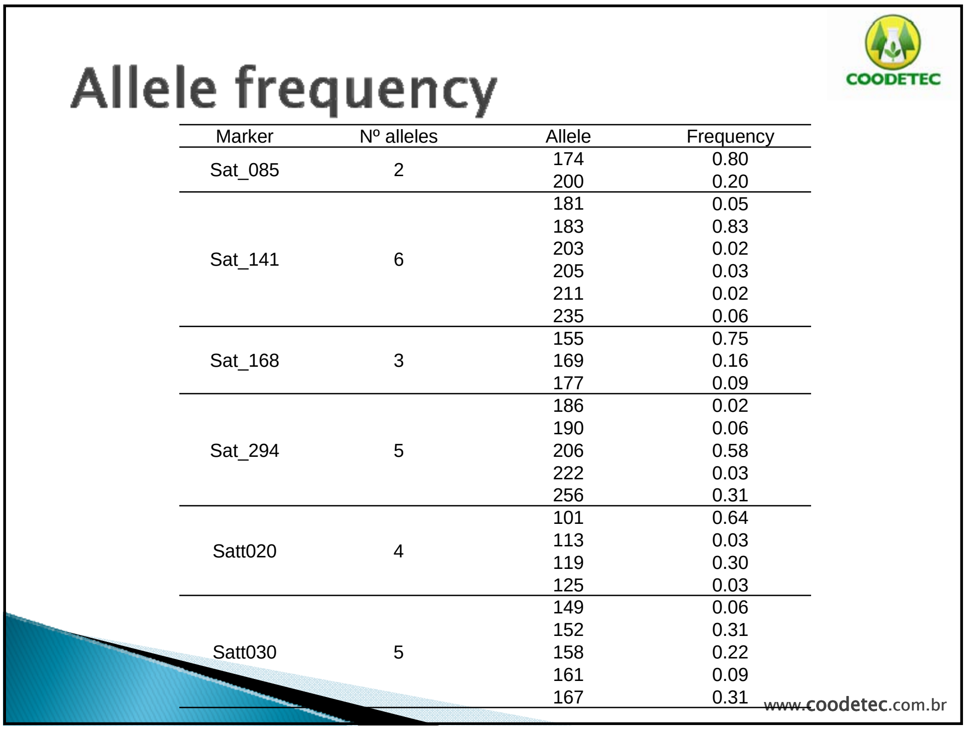
www.coodetec.com.br



Probability of Random Identity

- To conclude for identity:
  - PRI $\leq$ 0.0001%
  - PE $\geq$ 99.9999%

www.coodetec.com.br

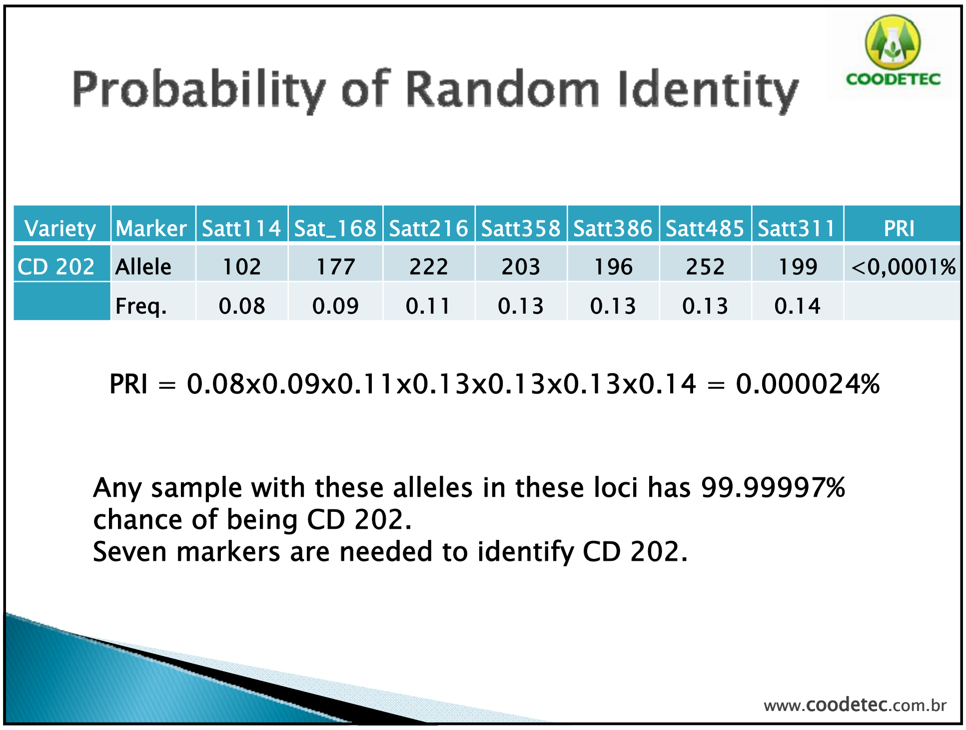


Allele frequency

COODETEC

Marker	Nº alleles	Allele	Frequency
Sat_085	2	174	0.80
		200	0.20
Sat_141	6	181	0.05
		183	0.83
		203	0.02
		205	0.03
		211	0.02
		235	0.06
Sat_168	3	155	0.75
		169	0.16
		177	0.09
Sat_294	5	186	0.02
		190	0.06
		206	0.58
		222	0.03
		256	0.31
Satt020	4	101	0.64
		113	0.03
		119	0.30
		125	0.03
Satt030	5	149	0.06
		152	0.31
		158	0.22
		161	0.09
		167	0.31

www.coodetec.com.br



Probability of Random Identity

COODETEC

Variety	Marker	Satt114	Sat_168	Satt216	Satt358	Satt386	Satt485	Satt311	PRI
CD 202	Allele	102	177	222	203	196	252	199	<0,0001%
	Freq.	0.08	0.09	0.11	0.13	0.13	0.13	0.14	

PRI = 0.08x0.09x0.11x0.13x0.13x0.13x0.14 = 0.000024%

Any sample with these alleles in these loci has 99.99997% chance of being CD 202.
Seven markers are needed to identify CD 202.

www.coodetec.com.br

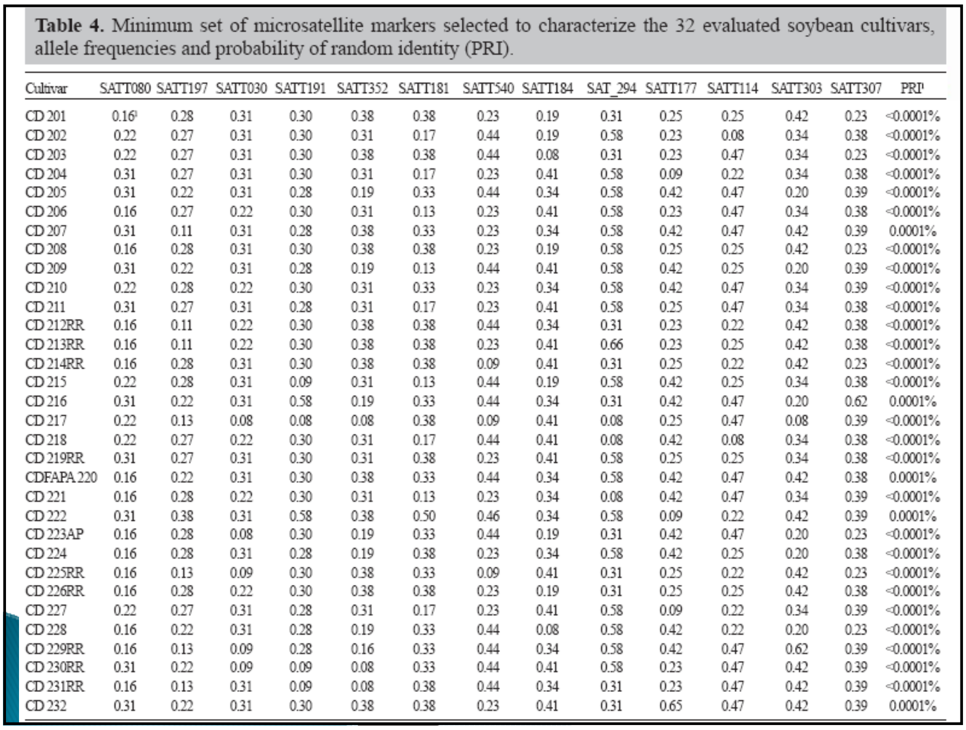
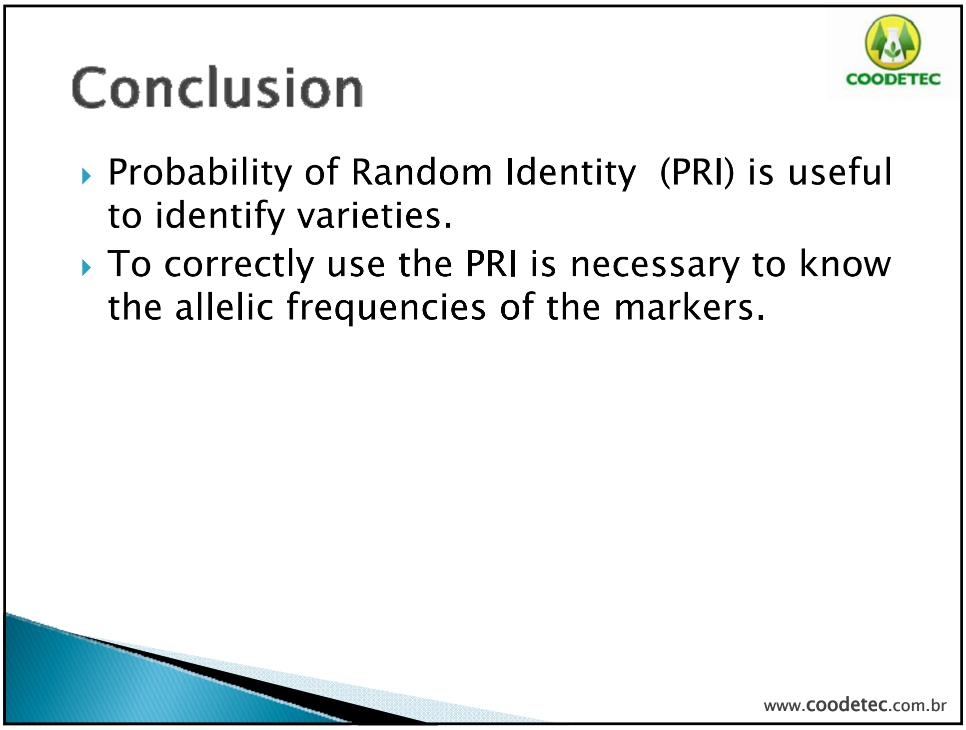


Table 4. Minimum set of microsatellite markers selected to characterize the 32 evaluated soybean cultivars, allele frequencies and probability of random identity (PRI).

Cultivar	SATT080	SATT197	SATT030	SATT191	SATT352	SATT181	SATT540	SATT184	SAT_294	SATT177	SATT114	SATT303	SATT307	PRI[1]
CD 201	0.16[1]	0.28	0.31	0.30	0.38	0.38	0.23	0.19	0.31	0.25	0.25	0.42	0.23	<0.0001%
CD 202	0.22	0.27	0.31	0.30	0.31	0.17	0.44	0.19	0.58	0.23	0.08	0.34	0.38	<0.0001%
CD 203	0.22	0.27	0.31	0.30	0.38	0.38	0.44	0.08	0.31	0.23	0.47	0.34	0.23	<0.0001%
CD 204	0.31	0.27	0.31	0.30	0.31	0.17	0.23	0.41	0.58	0.09	0.22	0.34	0.38	<0.0001%
CD 205	0.31	0.22	0.31	0.28	0.19	0.33	0.44	0.34	0.58	0.42	0.47	0.20	0.39	<0.0001%
CD 206	0.16	0.27	0.22	0.30	0.31	0.13	0.23	0.41	0.58	0.23	0.47	0.34	0.38	<0.0001%
CD 207	0.31	0.11	0.31	0.28	0.38	0.33	0.23	0.34	0.58	0.42	0.47	0.42	0.39	0.0001%
CD 208	0.16	0.28	0.31	0.30	0.38	0.38	0.23	0.19	0.58	0.25	0.25	0.42	0.23	<0.0001%
CD 209	0.31	0.22	0.31	0.28	0.19	0.13	0.44	0.41	0.58	0.42	0.25	0.20	0.39	<0.0001%
CD 210	0.22	0.28	0.22	0.30	0.31	0.33	0.23	0.34	0.58	0.42	0.47	0.34	0.39	<0.0001%
CD 211	0.31	0.27	0.31	0.28	0.31	0.17	0.23	0.41	0.58	0.25	0.47	0.34	0.38	<0.0001%
CD 212RR	0.16	0.11	0.22	0.30	0.38	0.38	0.44	0.34	0.31	0.23	0.22	0.42	0.38	<0.0001%
CD 213RR	0.16	0.11	0.22	0.30	0.38	0.38	0.23	0.41	0.66	0.23	0.25	0.42	0.38	<0.0001%
CD 214RR	0.16	0.28	0.31	0.30	0.38	0.38	0.09	0.41	0.31	0.25	0.22	0.42	0.23	<0.0001%
CD 215	0.22	0.28	0.31	0.09	0.31	0.13	0.44	0.19	0.58	0.42	0.25	0.34	0.38	<0.0001%
CD 216	0.31	0.22	0.31	0.58	0.19	0.33	0.44	0.34	0.31	0.42	0.47	0.20	0.62	0.0001%
CD 217	0.22	0.13	0.08	0.08	0.08	0.38	0.09	0.41	0.08	0.25	0.47	0.08	0.39	<0.0001%
CD 218	0.22	0.27	0.22	0.30	0.31	0.17	0.44	0.41	0.08	0.42	0.08	0.34	0.38	<0.0001%
CD 219RR	0.31	0.27	0.31	0.30	0.31	0.38	0.23	0.41	0.58	0.25	0.25	0.34	0.38	<0.0001%
CDFAPA 220	0.16	0.22	0.31	0.30	0.38	0.33	0.44	0.34	0.58	0.42	0.47	0.42	0.38	0.0001%
CD 221	0.16	0.28	0.22	0.30	0.31	0.13	0.23	0.34	0.08	0.42	0.47	0.34	0.39	<0.0001%
CD 222	0.31	0.38	0.31	0.58	0.38	0.50	0.46	0.34	0.58	0.09	0.22	0.42	0.39	0.0001%
CD 223AP	0.16	0.28	0.08	0.30	0.19	0.33	0.44	0.19	0.31	0.42	0.47	0.20	0.23	<0.0001%
CD 224	0.16	0.28	0.31	0.28	0.19	0.38	0.23	0.34	0.58	0.42	0.25	0.20	0.38	<0.0001%
CD 225RR	0.16	0.13	0.09	0.30	0.38	0.33	0.09	0.41	0.31	0.25	0.22	0.42	0.23	<0.0001%
CD 226RR	0.16	0.28	0.22	0.30	0.38	0.38	0.23	0.19	0.31	0.25	0.25	0.42	0.38	<0.0001%
CD 227	0.22	0.27	0.31	0.28	0.31	0.17	0.23	0.41	0.58	0.09	0.22	0.34	0.39	<0.0001%
CD 228	0.16	0.22	0.31	0.28	0.19	0.33	0.44	0.08	0.58	0.42	0.22	0.20	0.23	<0.0001%
CD 229RR	0.16	0.13	0.09	0.28	0.16	0.33	0.44	0.34	0.58	0.42	0.47	0.62	0.39	<0.0001%
CD 230RR	0.31	0.22	0.09	0.09	0.08	0.33	0.44	0.41	0.58	0.23	0.47	0.42	0.39	<0.0001%
CD 231RR	0.16	0.13	0.31	0.09	0.08	0.38	0.44	0.34	0.31	0.23	0.47	0.42	0.39	<0.0001%
CD 232	0.31	0.22	0.31	0.30	0.38	0.38	0.23	0.41	0.31	0.65	0.47	0.42	0.39	0.0001%



Conclusion

COODETEC

- Probability of Random Identity (PRI) is useful to identify varieties.
- To correctly use the PRI is necessary to know the allelic frequencies of the markers.

www.coodetec.com.br

[End of document]