

Technical Working Party on Testing Methods and Techniques

TWM/1/10

First Session

Virtual meeting, September 19 to 23, 2022

Original: English

Date: August 29, 2022


DEVELOPMENT OF STATISTICAL ANALYSIS SOFTWARE

Document prepared by an expert from China

Disclaimer: this document does not represent UPOV policies or guidance


The annex to this document contains a copy of a presentation on “Development of statistical analysis software”, prepared by an expert from China, to be made at the first session of the TWM.

[Annex follows]



IVF CAAS
Institute of Vegetables and Flowers
Chinese Academy of Agricultural Sciences

DUS
Beijing



DEVELOPMENT OF STATISTICAL ANALYSIS SOFTWARE

Yang Kun
Deputy director of Beijing Sub-center of New Plant Variety Tests, MARA, China

TWM1, Virtual meeting, September 19 to 23, 2022

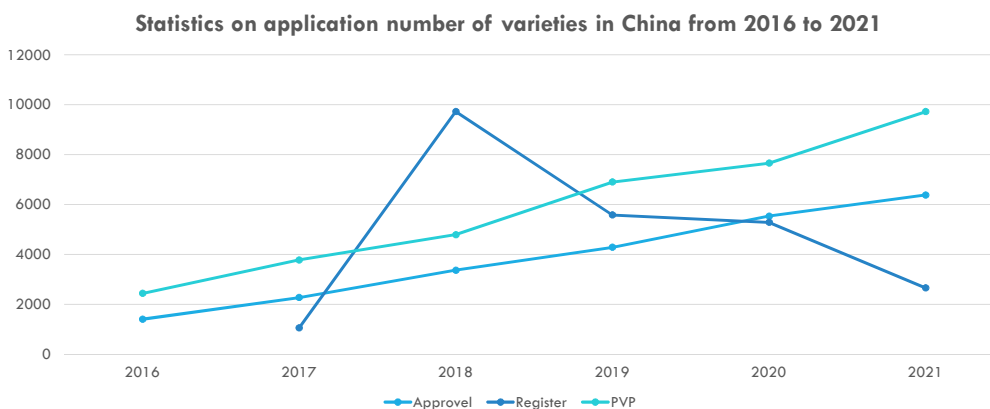
CONTENTS

- 1. Background
- 2. History of development
- 3. Innovations
- 4. Plans in future

CONTENTS

- 1. Background
- 2. History of development
- 3. Innovations
- 4. Plans in future

APPLICATIONS OF PVP AND NL



new varieties in
total :
82860
new varieties
per year :
13810
new corn
varieties in total :
22060
new corn
varieties per
year :
3677

1 CENTER, 14 OLD SUB-CENTERS, 18 NEW SUB-CENTERS OR STATIONS



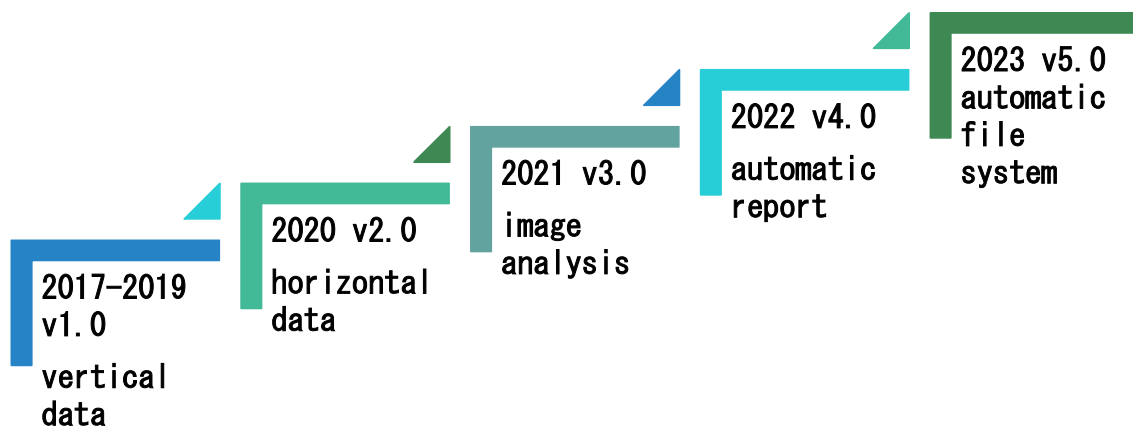
PROBLEMS OF DUS TESTING

1. Testers contribute most part of variance in data.
2. Environmental conditions contribute a significant variance in data.
3. There are no unified methods for trial design and statistical analysis.
4. There are no unified decision threshold for DUS.

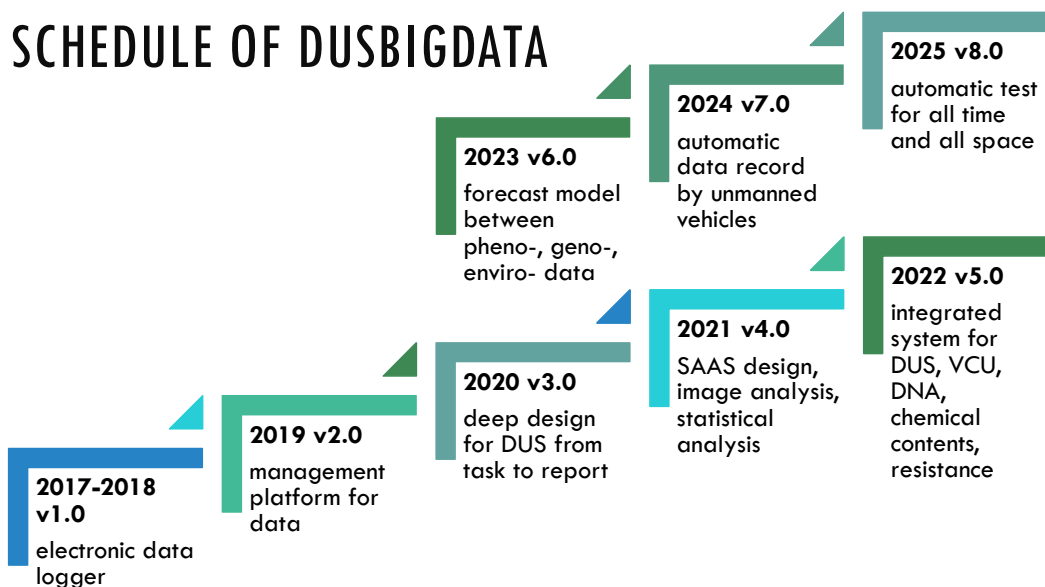
CONTENTS

- 1. Background
- 2. History of development
- 3. Innovations
- 4. Plans in future

SCHEDULE OF DUSCEL



SCHEDULE OF DUSBIGDATA



HISTORY OF REPORTS

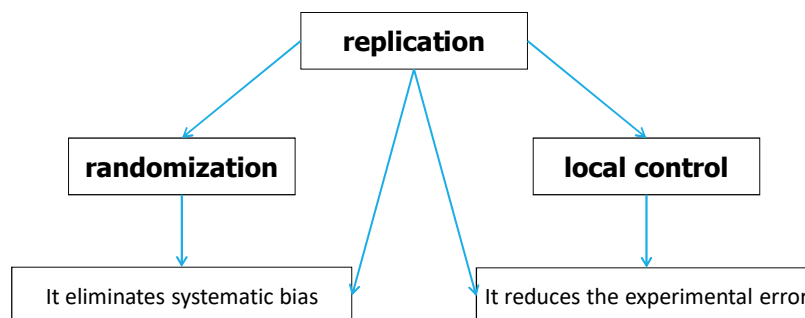
1. 2019, DUSCEL V1.0, 12 sheets and 46 functions, reported in TWC37.
2. 2020, DUSCEL V2.0, 6 sheets and 55 functions, reported in TWC38.
3. 2021, DUSCEL V2.5, 7 sheets and 52 functions, reported in TWA50.
4. 2021, DUSCEL V3.0, 5 sheets and 42 functions, reported in TWC39.
5. 2021, DUSBIGDATA V4.0, reported in TWC39.
6. 2022, DUSCEL V3.5, 8 sheets and 35 functions, reported in TWV56.
7. 2022, DUSBIGDATA V5.0 reported in TWA51.
8. 2022, Development of statistical analysis software, reported in TWM1.

CONTENTS

- 1. Background
- 2. History of development
- 3. Innovations
- 4. Plans in future

1. NEW THEORY: DUS TESTING SHOULD OBSERVE GENETIC AND STATISTICAL THEORY.

Blind test is the final solution to determine DUS. Trial design should be side-by-side, randomized plots or plants.



1. NEW THEORY: DUS TESTING SHOULD OBSERVE GENETIC AND STATISTICAL THEORY.

Sample size, mean and distance, variance or standard deviation, error I and II are main elements to normal or binominal distributions and correspondent test models.

Estimation of sample size for normal data 常规数据样本含量估计

α 显著水平

σ 标准差

Δ 均值差

☒ 错误

Default默认值

☒ Est. of Population Mean
总体均值估计

☐ Est. of two sample mean
两样本均值差估算

☐ Power est. of Population Mean
总体均值带功效估计

☐ Power Est. of two sample mean
两样本均值差带功效估算

Minimum sample size 最小样本容量

Estimation of sample size for percentage value 百分率数据样本容量估计

显著水平 α

处理1百分率 p_1

处理2百分率 p_2

百分率差 Δ

☒ 错误 β

Default默认值

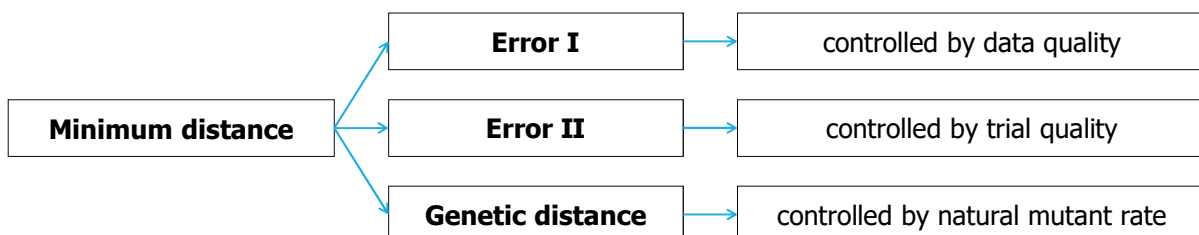
☐ Est. of total percentage
总体百分率估算

☒ Power est. of total percentage
总体百分率带功效估计

Minimum sample size 最小样本容量

1. NEW THEORY: DUS TESTING SHOULD OBSERVE GENETIC AND STATISTICAL THEORY.

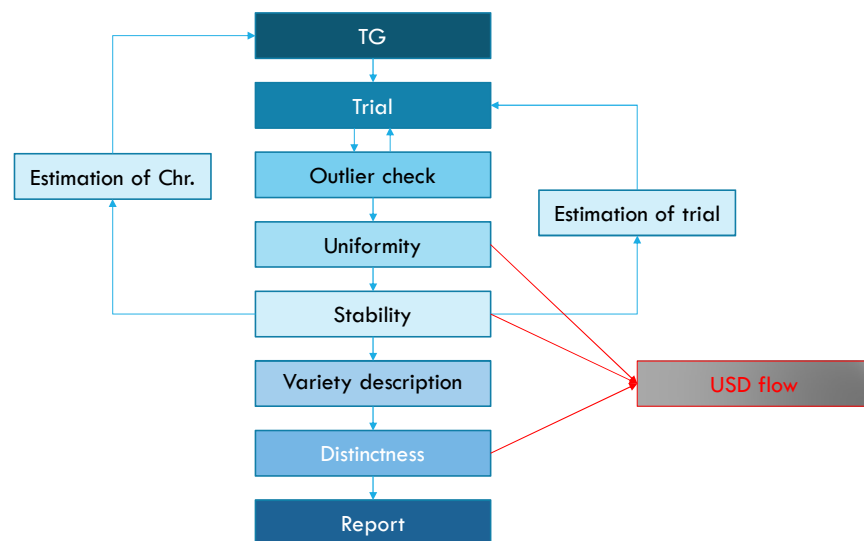
Minimum distance of DUS is only controlled by error I, error II and allowed genetic distance which could be 0.



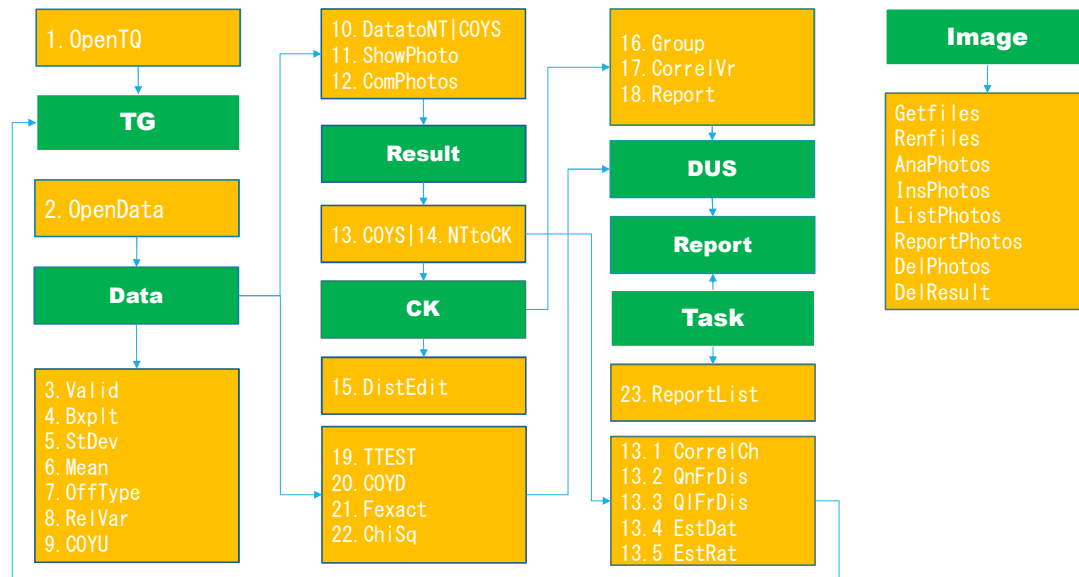
BENEFITS OF NEW THEORY

1. It provides a gold standard for all methods.
2. It harmonizes speed and quality of granting a PBR to breeders.
3. It harmonizes the theory of distinctness, uniformity, stability and EDV.

2. NEW FLOW : RECYCLING, PRECISE AND EFFICIENT ANALYSIS.



2. NEW FLOW : RECYCLING, PRECISE AND EFFICIENT ANALYSIS.



BENEFIT ONE OF NEW FLOW: UNIFIED METHODS OF OUTLIERS CHECKING, VALIDATION OF STATISTICAL ANALYSIS, UNIFORMITY ANALYSIS AND STABILITY ANALYSIS

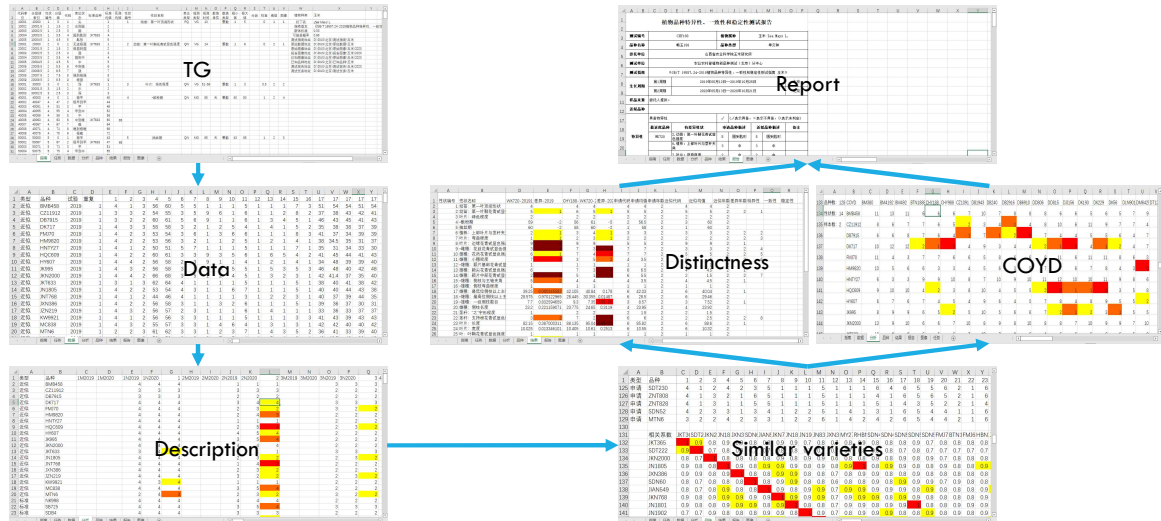
The figure consists of three parts, each illustrating a different data visualization technique for COVID-19 data.

Top Example: Heatmap
 This heatmap displays COVID-19 cases across various countries (A-Z) for the years 2019 and 2020. The color scale ranges from 0 (white) to 100 (red). A blue arrow points from the 'Data' label to this heatmap.

Middle Example: Table
 This table shows COVID-19 statistics for various countries. The columns include Country, Total, New, Recovered, and others. A blue arrow points from the 'relative variance' label to this table.

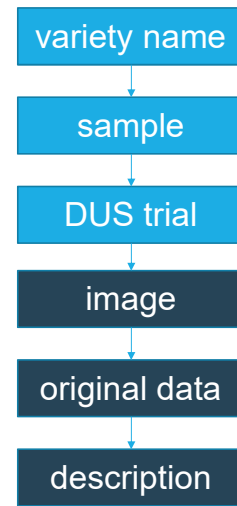
Bottom Example: Table
 This table shows COVID-19 statistics for various countries. The columns include Country, Total, New, Recovered, and others. A blue arrow points from the 'COVU' label to this table.

BENEFIT TWO OF NEW FLOW: UNIFIED TG PARAMETERS, VARIETY DESCRIPTIONS AND DISTINCTNESS ANALYSIS



3. NEW DEFINITION OF IMAGE : HARMONIZE SAMPLES, TRIALS, IMAGE, ORIGINAL DATA AND DESCRIPTION.

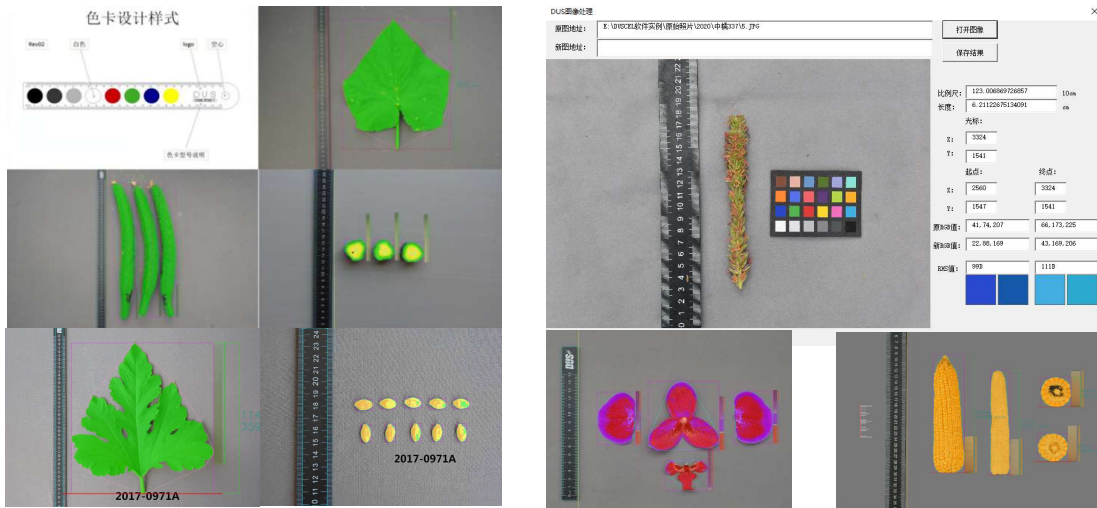
D:\DUS\squash\beijing\2020\2017-0971A\1.jpg



BENEFIT ONE OF IMAGE : IT'S USEFUL FOR CHECKING ERROR OF DATA.

	A	B	C	D	E	F		A	B	C	D	E	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	ClName	Type	Address	NavName	Photo1	Photo2		1	类型	品种	1	2	3	类型	品种	1M2015	1M2020	1N2019	1N2020	1	2M2015	2M2020	2N2019	2N2020	2	3M2015	3M2020	3N2019	3N2020	
1.jpg	文件	D:\语音库\2020\2017-0971A					5	近似	CZ11912	3		2	近似	BMB458	4	4	4	4	4	1	1	1	1	1	1	1	1	3	3	3
2.jpg	文件	D:\语音库\2020\2017-0971A					6	近似	DB7915	3		2	近似	CZ11912	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2
3.jpg	文件	D:\语音库\2020\2017-0971A					13	近似	BMB458	4		3	近似	DB7915	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2
4.jpg	文件	D:\语音库\2020\2017-0971A					14	近似	DK717	4		3	近似	DK717	4	4	4	4	4	3	4	4	4	4	4	4	2	3	2	2
							41	申请	BM4192	4		2	近似	FM370	4	4	4	4	4	2	3	2	2	2	2	2	2	2	2	2
													近似	HM9820	4	4	4	4	4	2	4	4	4	4	4	2	4	4	4	4
													近似	HNTY27	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1
													近似	HQC869	4	4	4	4	4	2	5	5	5	5	5	5	5	5	5	5
													近似	HY607	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
													近似	XK995	4	4	4	4	4	3	5	5	5	5	5	5	5	5	5	5
																			</											

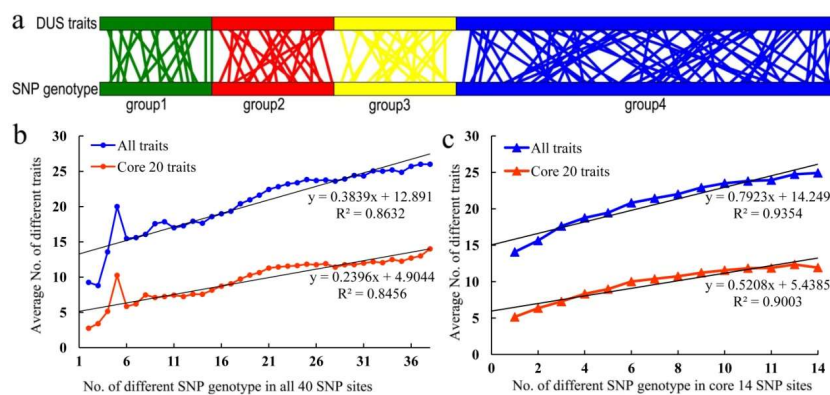
BENEFIT TWO OF IMAGE: IT'S A HIGH-THROUGHPUT WAY TO PRODUCE DATA.



BENEFIT THREE OF IMAGE: IT CREATES MANY NEW CHARACTERISTICS.



4. NEW WAY TO SELECT DNA MARKERS : SEQUENCING TECHNIQUES AND PCA ANALYSIS METHODS.



Horticultural Plant Journal
Available online 13 July 2022
In Press, Corrected Proof

Comparison of DUS testing and SNP fingerprinting for variety identification in cucumber

Jian Zhang^{a,b}, Jingjing Yang^{a,b}, Shenzao Fu^c, Jun Ren^c, Xiaofei Zhang^{a,b}, Changquan Xia^{a,b}, Hong Zhao^{a,b}, Kun Yang^c, Changlong Wen^{a,b}

134 cucumber varieties
50 DUS chr. → 20 core chr.
40 SNP sites → 14 core sites

we find a way to select SNP sites to get high relationship to morphological distance.

4. NEW WAY OF SELECTION OF DNA MARKERS : SEQUENCING TECHNIQUES AND PCA ANALYSIS METHODS.

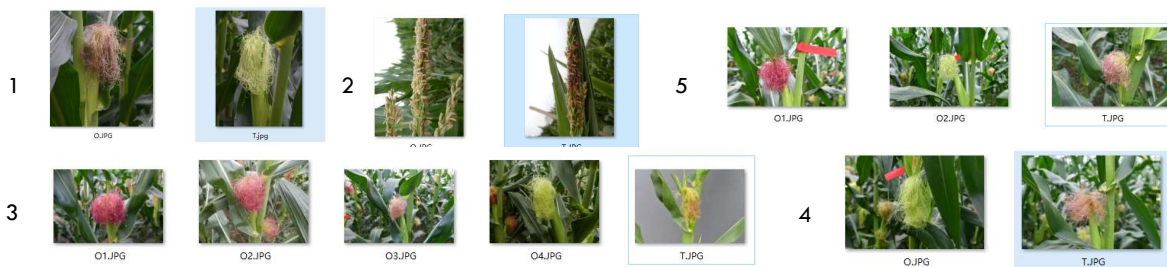
crop	variety	DNA<4	DUS=0	DNA=0	DUS=0
Cucumber	137	28	11	4	1
Cabbage	80	19	6	1	1
Tomato	287	7	3	1	0

2020, Beijing, Cucumber, Cabbage, Tomato, Distinctness

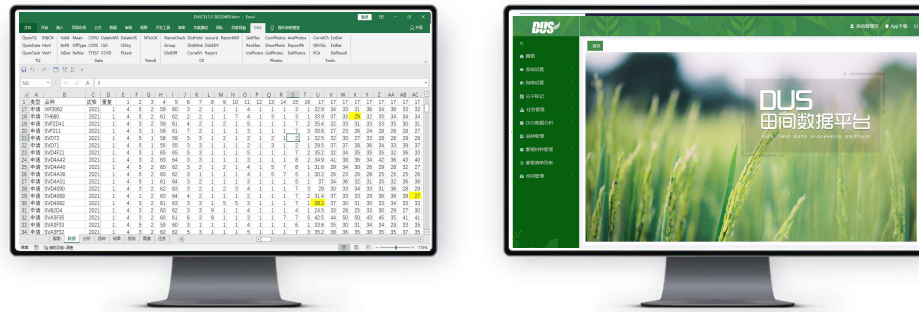
4. NEW WAY OF SELECTION OF DNA MARKERS : SEQUENCING TECHNIQUES AND PCA ANALYSIS METHODS.

variety	plant	off-types	total pairs	different at least 1 SNP	different by similarity 90%	different by similarity 95%
1	10	5	45	25	25	29
2	10	5	45	0	0	7
3	10	5	45	34	41	45
4	10	5	45	29	26	43
5	10	5	45	16	19	44

2021, Beijing, Maize, Uniformity



5. NEW TOOLS FOR MANAGEMENT AND ANALYSIS OF DUS DATA



DUSCEL=EXCEL+VBA+UI

DUSCEL3.5 20220419.xlsx - Excel

文件开始插入页面布局公式数据审阅视图开发工具帮助加载项团队百度网盘DUS帮助操作说明库

OpenTGOpenTaskENJCNValidMeanCOVUCOVDHtoVQIRFDisVtoHDistEditGroupDistMinkDistEDVReportPhGetFilesShowPhotoReporthEstDatRenFilesComPhotosAnaPhotosEstRatOpenDataCalibBxPhOffTypeCOVSDatatoMSChiSqTrialDesignCorrelChDistDiffCorrelVrInsPhotosListPhotosDelPhotosDelResultOpenCKRenewTQSTDevRefVarTTESTDatatoVSDataFExactNToCKAnalysisQIRFDisDistHoldJaccardCK

AC18

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

AA

AB

AC

1

Type

Variety

Trial

Replicate

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

17

17

17

17

17

17

17

17

17

2

Similar

BMB458

2019

1

4

1

3

56

60

5

5

1

1

5

1

1

5

1

1

7

3

51

54

54

51

54

52

53

51

48

3

Similar

CZ11912

2019

1

3

3

2

54

55

3

5

9

6

1

6

1

1

2

8

2

37

38

43

42

41

37

31

41

41

40

4

Similar

DB7915

2019

1

3

2

2

60

61

5

6

9

1

1

6

1

3

4

5

1

46

43

45

41

43

39

42

42

45

5

Similar

DK717

2019

1

4

3

3

58

58

3

2

1

2

5

4

1

4

1

5

2

35

38

38

37

39

39

36

36

36

6

Similar

FMJ70

2019

1

4

2

3

53

54

3

6

1

3

6

6

1

1

1

8

3

41

37

34

39

39

34

37

37

33

7

Similar

HM9820

2019

1

4

2

2

53

56

3

2

1

1

2

5

1

2

1

4

1

38

34.5

35

31

37

35

34

34

33

8

Similar

HNTY27

2019

1

4

1

2

50

51

5

7

1

1

1

5

1

1

1

7

1

35

31

34

33

30

33

33

33

35

9

Similar

HQC609

2019

1

4

2

2

60

61

3

3

9

3

5

6

1

6

5

4

2

41

45

44

41

43

41

42

44

41

10

Similar

HY607

2019

1

4

4

2

56

58

2

3

9

1

1

4

1

2

1

4

1

34

48

39

39

40

37

41

43

38

11

Similar

JK995

2019

1

4

3

2

56

58

3

3

1

4

5

5

1

5

3

5

3

46

48

40

42

48

45

48

45

43

12

Similar

JKN2000

2019

1

4

4

2

66

68

3

5

1

2

4

5

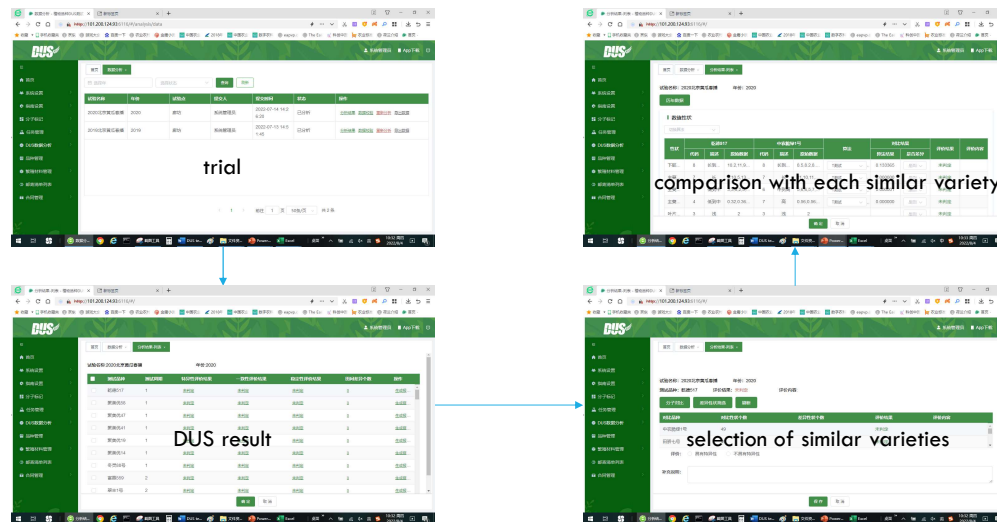
1

3

2

3

DUSBIGDATA= HADOOP+C+JAVA



CONTENTS

- 1. Background
- 2. History of development
- 3. Innovations
- 4. Plans in future

PLANS IN FUTURE

- ◆ 1. to define DUS file system and finish development project on DUSCEL in 2023.
- ◆ 2. to start development project on forecast model between morphological, molecular and environmental data in 2023.
- ◆ 3. to start development project on PVP knowledge map database in 2023.
- ◆ 4. to start development project on automatic data loggers in 2024.
- ◆ 5. to finish development project of DUSBIGDATA in 2025.
- ◆ 6. to contribute more suggestions for improvement of UPOV TGP/8 document.
- ◆ 7. to draft more general guidelines or protocols for DUS testing in China.



THANK YOU FOR YOUR
ATTENTION!

Yang Kun
yangkun@caas.cn
+86-10-8210 5623
+86-(0)1391 1259 308