Locus allele count : a tool to estimate the number of contributors in a DNA mixture

Marie-Pier Thibault, Forensic biologist, LSJML

Green Mountain DNA Conference – July 2023

1

About our lab

- Serves a population of about 8,5 million
- Caseload of >8000/year; 35 000 exhibits
- Team of 84, including 45 forensic biologists

2

About our lab

- Serves a population of about 8,5 million
- Caseload of >8000/year; 35 000 exhibits
- Team of 84, including 45 forensic biologists

- Amplification and CE : AmpFLSTR™ Identifiler™ Plus (ID+; 15 sites), 3500 XL, Genemapper ID-X
- STRMix 2.5, validated up to 5C
- Support the inclusion with an LR > 1000 (0.001 to 1000 = inconclusive)

3

Plan

- Introduction:
  - Importance of making an informed estimation of the number of contributors (NOC) in a DNA mixture
  - Highlights of our previous work on the Total Allele Count (TAC)
- How can locus allele count (LAC) be used to refine NOC estimation
  - DNA mixtures generation
  - How to use the tool
  - Let's put the tool to the test

4

Acknowledgments

Tommy Harding
tommy.harding@msp.gouv.qc.ca

Nicolas Tremblay
nicolas.tremblay@msp.gouv.qc.ca

5

Acknowledgments

Tommy Harding
tommy.harding@msp.gouv.qc.ca

Nicolas Tremblay
nicolas.tremblay@msp.gouv.qc.ca

Don't shoot the messenger

6

DNA evidence

- DNA evidence is widely used as a circumstantial evidence in crime investigation.

7

DNA evidence

- DNA evidence is widely used as a circumstantial evidence in crime investigation.
- … but DNA evidence doesn't speak on its own

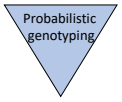
8

DNA evidence

- DNA evidence is widely used as a circumstantial evidence in crime investigation.
- … but DNA evidence doesn't speak on its own
- Complex DNA mixtures now represent a significant portion of DNA casework results and approaches to mixture **interpretation** are of great interest to the forensic community.

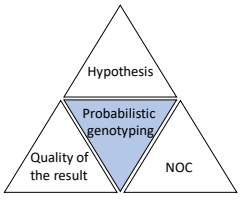
9

Interpretation is required for probabilistic genotyping



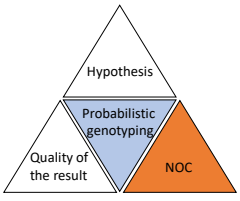
10

NOC estimation is a core component of the weight of evidence



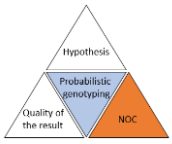
11

NOC estimation is a core component of the weight of evidence



12

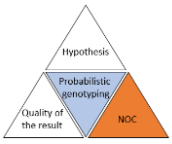
NOC estimation is a core component of the weight of evidence



NOC
- the true NOC is never known in forensic samples
- an essential variable of the deconvolution
- uncertainty remains about the NOC even after careful analysis (> 3C)
- Variability between analysts (**DNAmix 2021**)

13

NOC estimation is a core component of the weight of evidence

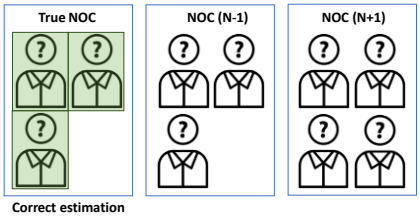


NOC
- the true NOC is never known in forensic samples
- an essential variable of the deconvolution
- uncertainty remains about the NOC even after careful analysis (> 3C)
- Variability between analysts

What are the consequences of an incorrect NOC estimate?

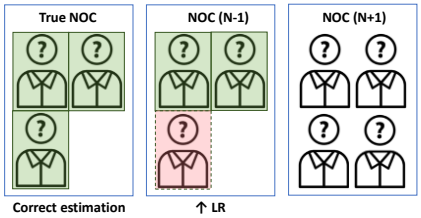
14

Incorrect NOC estimate can lead to false exclusion



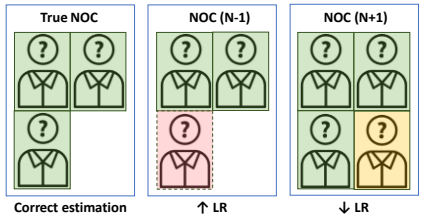
15

Incorrect NOC estimate can lead to false exclusion



16

Incorrect NOC estimate can lead to false exclusion



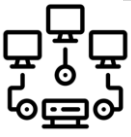
17

Tools used to estimate the NOC

Maximum allele count (MAC)
Allelic balance
Inter-loci consistency

18

Tools used to estimate the NOC



Maximum allele count (MAC)
Allelic balance
Inter-loci consistency

Maximum likelihood estimation (MLE)

NOCit (Bayesian method)
PACE (AI classification)
FaSTR (AI classification)

19

Tools used to estimate the NOC

Maximum allele count (MAC)
Allelic balance
Inter-loci consistency

Maximum likelihood estimation (MLE)

NOCit (Bayesian method)
PACE (AI classification)
FaSTR (AI classification)

Need time to learn & validate

20

Total allele count distribution (TAC curves) improves NOC estimation

CANADIAN SOCIETY OF FORENSIC SCIENCE JOURNAL
https://doi.org/10.1080/00085030.2022.2028359
Taylor & Francis
RESEARCH ARTICLE
Check for updates

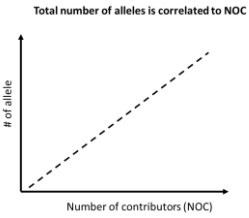
Total allele count distribution (TAC curves) improves number of contributor estimation for complex DNA mixtures

Josée Noël[a], Sarah Noël[a], France Mailly[a], Dominic Granger[a], Jean-François Lefebvre[a], Emmanuel Milot[b] and Diane Séguin[a]

[a]Laboratoire de Sciences Judiciaires et de Médecine Légale, Montréal, Québec, Canada; [b]Laboratoire de Recherche en Criminalistique, Department of Chemistry, Biochemistry and Physics and Centre International de Criminologie Comparée, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada

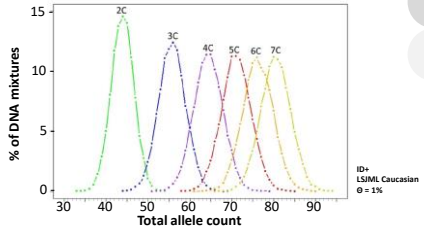
21

Total allele count (TAC) is proportional to NOC



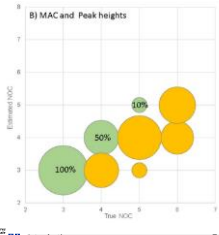
22

Total allele count (TAC) is proportional to NOC



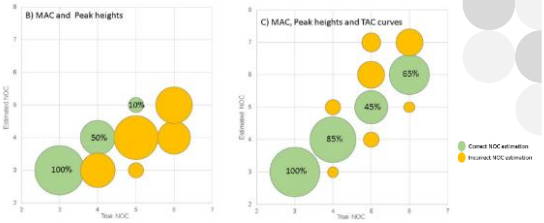
23

TAC curves improves NOC estimation



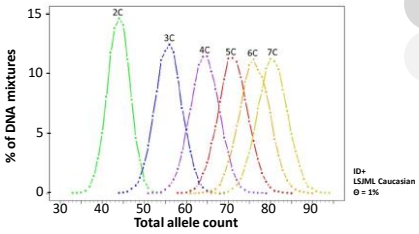
24

TAC curves improves NOC estimation



25

Can we learn more from TAC curves?

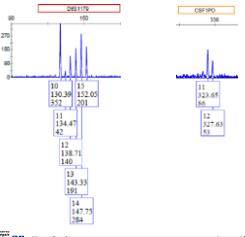


26

Can we learn more from TAC curves?

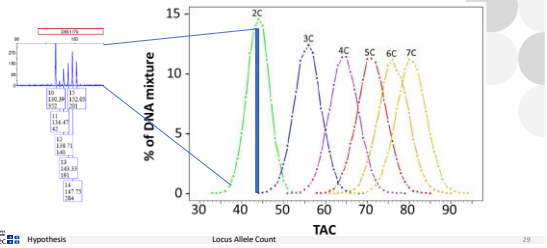
27

Can we learn more from TAC curves?



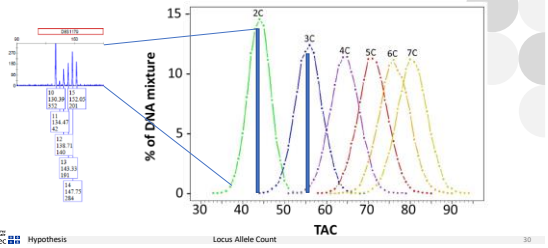
28

Can we learn more from TAC curves?



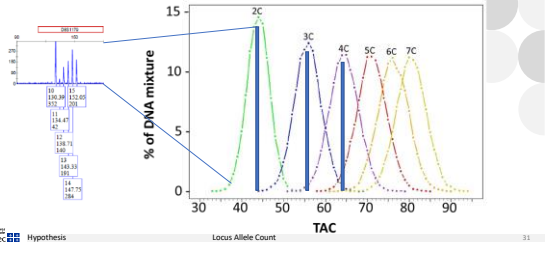
29

Can we learn more from TAC curves?



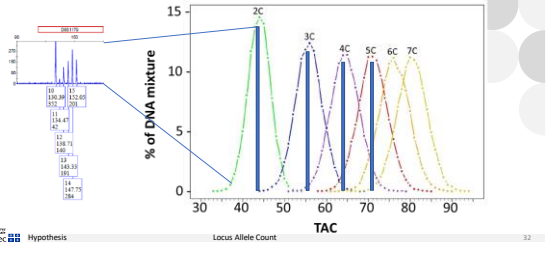
30

Can we learn more from TAC curves?



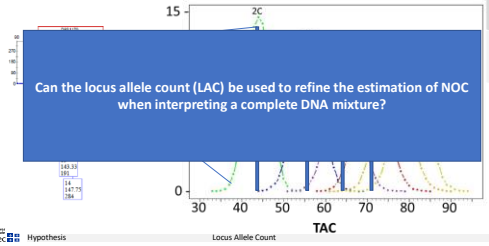
31

Can we learn more from TAC curves?



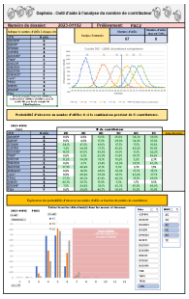
32

Can we learn more from TAC curves?



Can the locus allele count (LAC) be used to refine the estimation of NOC when interpreting a complete DNA mixture?

33



34

Generate DNA mixtures

- Dataset generated in Excel
- 2,7 millions profiles were generated
- 100 000 DNA mixtures were generated for each NOC
- LSJML-Caucasian allele frequencies; θ = 1%; ID+

35

Step 1 : Draw genotypes

- Example for 2C DNA mixtures

	D8S1179	
	C1	C2
Mix 1	1067	1014
Mix 2	1002	1017
Mix 3	1043	1095
…		
Mix 100 000	1046	1013

36

Step 2 : Convert genotypes to alleles

- Example for 2C DNA mixtures

	D8S1179			
	C1	C2	C1	C2
Mix 1	1067	1014	14, 15	8, 10
Mix 2	1002	1017	9, 12	9, 15
Mix 3	1043	1095	12, 11	17, 13
…				
Mix 100 000	1046	1013	9, 12	8, 9

37

Step 3 : Count the number of alleles

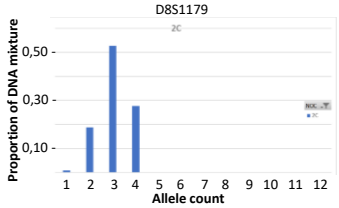
- Example for 2C DNA mixtures

	D8S1179				
	C1	C2	C1	C2	
Mix 1	1067	1014	14, 15	8, 10	4
Mix 2	1002	1017	9, 12	9, 15	3
Mix 3	1043	1095	12, 11	17, 13	4
…					
Mix 100 000	1046	1013	9, 12	8, 9	3

38

Graphical data analysis

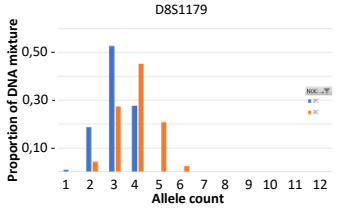
- From the data generated, we can compute:



39

Graphical data analysis

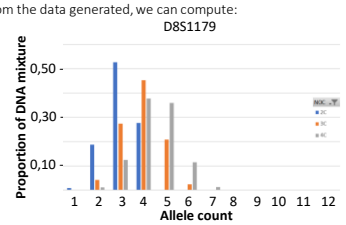
- From the data generated, we can compute:



40

Graphical data analysis

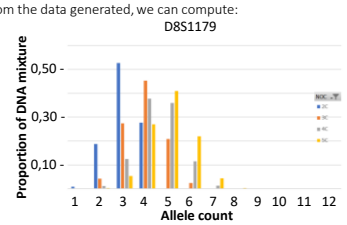
- From the data generated, we can compute:



41

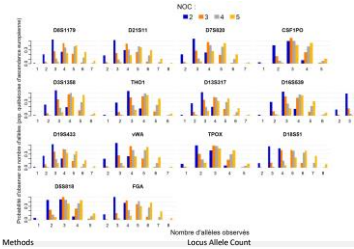
Graphical data analysis

- From the data generated, we can compute:



42

Graphical data analysis

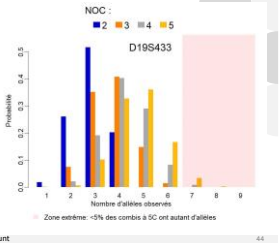


43

Graphical data analysis

Extreme zone ⚠

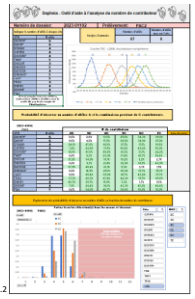
Upper region of the probability distribution where <5% of the 5C mixtures have such a high number of alleles



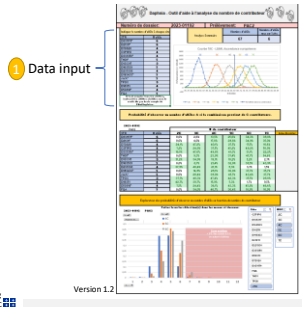
44

How to use the tool

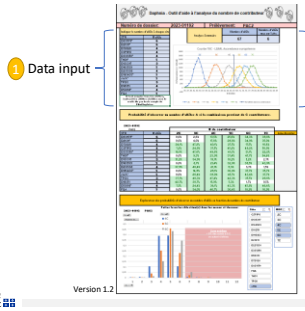
45



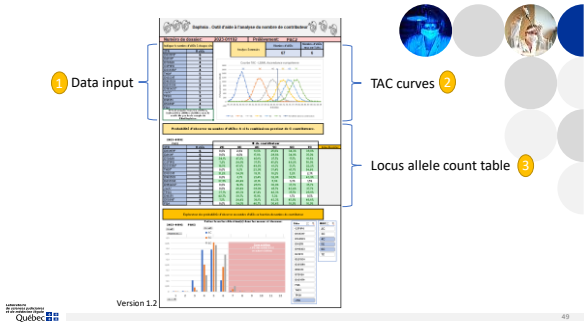
46



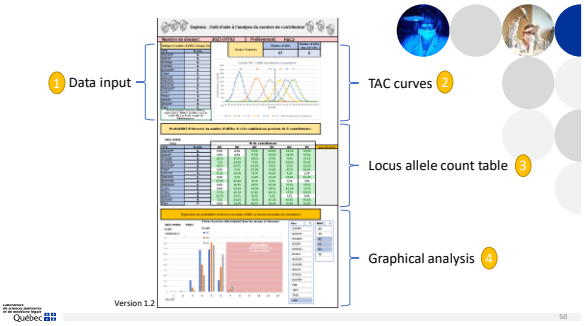
47



48



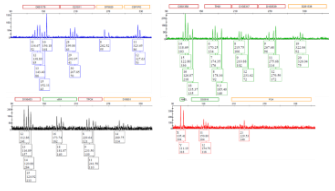
49



50

How to use the tool

- Step 1: Count every allele, including those that are visible below the analytical threshold.

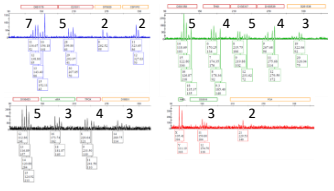


Results Locus Allele Count 51

51

How to use the tool

- Step 1: Count every allele, including those that are visible below the analytical threshold.

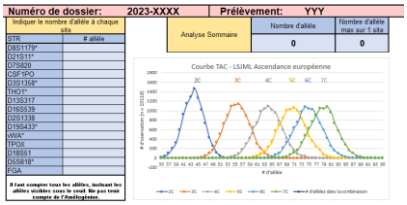


Results Locus Allele Count 52

52

How to use the tool

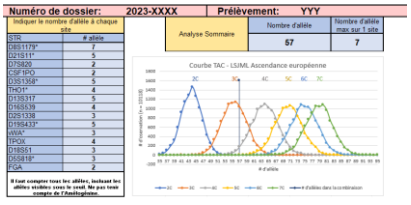
- Step 2: Input the # of allele at each locus



53

How to use the tool

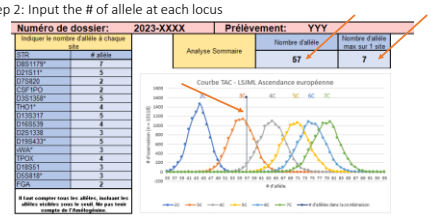
- Step 2: Input the # of allele at each locus



54

How to use the tool

- Step 2: Input the # of allele at each locus



55

How to use the tool

- Locus allele count table

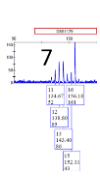
2023-XXXX
YYY

STR	# allèle	# de contributeur 2C	3C	4C	5C	6C	7C
D8S1179*	7	0,0%	0,0%	1,6%	6,0%	12,4%	19,2%
D21S11*	5	0,0%	25,6%	38,3%	36,1%	28,2%	21,7%
D7S820	2	20,3%	3,6%	0,7%	0,1%	0,0%	0,0%
CSF1PO	2	40,2%	16,3%	6,3%	2,5%	1,0%	0,4%
D3S1358*	5	0,0%	11,6%	29,8%	45,3%	56,2%	63,2%
THO1*	4	15,8%	42,3%	49,1%	44,9%	37,7%	30,4%
D13S317	5	0,0%	14,5%	29,8%	37,5%	39,4%	37,3%
D16S539	4	14,8%	38,9%	46,5%	44,6%	39,3%	33,0%
D2S1338	3	48,3%	15,8%	4,4%	1,2%	0,3%	0,1%
D19S433*	5	0,0%	14,9%	28,1%	36,0%	37,7%	35,7%
vWA*	3	54,5%	27,3%	10,6%	3,8%	1,4%	0,6%
TPOX	4	4,9%	17,6%	29,3%	38,1%	44,0%	47,3%
D18S51	3	47,2%	11,5%	2,2%	0,4%	0,1%	0,0%
D5S818*	3	44,8%	49,9%	43,9%	35,6%	28,2%	21,8%
FGA	2	12,2%	1,2%	0,1%	0,0%	0,0%	0,0%

56

How to use the tool

- Locus allele count table



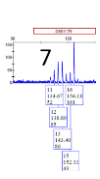
2023-XXXX
YYY

STR	# allèle	# de contributeur					
		2C	3C	4C	5C	6C	7C
D8S1179*	7	0,0%	0,0%	1,6%	6,0%	12,4%	19,2%
D21S11*	5	0,0%	25,6%	38,3%	36,1%	29,2%	21,7%
D7S820	2	20,3%	3,6%	0,7%	0,1%	0,0%	0,0%
CSF1PO	2	40,2%	16,3%	6,3%	2,5%	1,0%	0,4%
D3S1358*	5	0,0%	11,6%	29,8%	45,3%	56,2%	63,2%
THO1*	4	15,8%	42,3%	49,1%	44,9%	37,7%	30,4%
D13S317	5	0,0%	14,5%	29,8%	37,5%	39,4%	37,3%
D16S539	4	14,8%	38,9%	46,5%	44,6%	39,3%	33,0%
D2S1338	3	48,3%	15,8%	4,4%	1,2%	0,3%	0,1%
D19S433*	5	0,0%	14,9%	29,1%	36,0%	37,7%	35,7%
vWA*	3	54,5%	27,3%	10,6%	3,8%	1,4%	0,6%
TPOX	4	4,9%	17,6%	29,3%	38,1%	44,0%	47,3%
D18S51	3	47,2%	11,5%	2,2%	0,4%	0,1%	0,0%
D5S818*	3	44,8%	49,9%	45,9%	35,6%	28,2%	21,8%
FGA	2	12,2%	1,2%	0,1%	0,0%	0,0%	0,0%

57

How to use the tool

- Locus allele count table



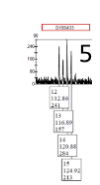
2023-XXXX
YYY

STR	# allèle	# de contributeur					
		2C	3C	4C	5C	6C	7C
D8S1179*	7	0,0%	0,0%	1,6%	6,0%	12,4%	19,2%
D21S11*	5	0,0%	25,6%	38,3%	36,1%	29,2%	21,7%
D7S820	2	20,3%	3,6%	0,7%	0,1%	0,0%	0,0%
CSF1PO	2	40,2%	16,3%	6,3%	2,5%	1,0%	0,4%
D3S1358*	5	0,0%	11,6%	29,8%	45,3%	56,2%	63,2%
THO1*	4	15,8%	42,3%	49,1%	44,9%	37,7%	30,4%
D13S317	5	0,0%	14,5%	29,8%	37,5%	39,4%	37,3%
D16S539	4	14,8%	38,9%	46,5%	44,6%	39,3%	33,0%
D2S1338	3	48,3%	15,8%	4,4%	1,2%	0,3%	0,1%
D19S433*	5	0,0%	14,9%	29,1%	36,0%	37,7%	35,7%
vWA*	3	54,5%	27,3%	10,6%	3,8%	1,4%	0,6%
TPOX	4	4,9%	17,6%	29,3%	38,1%	44,0%	47,3%
D18S51	3	47,2%	11,5%	2,2%	0,4%	0,1%	0,0%
D5S818*	3	44,8%	49,9%	45,9%	35,6%	28,2%	21,8%
FGA	2	12,2%	1,2%	0,1%	0,0%	0,0%	0,0%

58

How to use the tool

- Locus allele count table



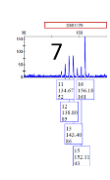
2023-XXXX
YYY

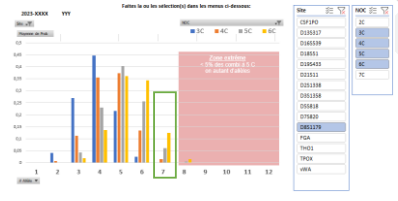
STR	# allèle	# de contributeur					
		2C	3C	4C	5C	6C	7C
D8S1179*	7	0,0%	0,0%	1,6%	6,0%	12,4%	19,2%
D21S11*	5	0,0%	25,6%	38,3%	36,1%	29,2%	21,7%
D7S820	2	20,3%	3,6%	0,7%	0,1%	0,0%	0,0%
CSF1PO	2	40,2%	16,3%	6,3%	2,5%	1,0%	0,4%
D3S1358*	5	0,0%	11,6%	29,8%	45,3%	56,2%	63,2%
THO1*	4	15,8%	42,3%	49,1%	44,9%	37,7%	30,4%
D13S317	5	0,0%	14,5%	29,8%	37,5%	39,4%	37,3%
D16S539	4	14,8%	38,9%	46,5%	44,6%	39,3%	33,0%
D2S1338	3	48,3%	15,8%	4,4%	1,2%	0,3%	0,1%
D19S433*	5	0,0%	14,9%	29,1%	36,0%	37,7%	35,7%
vWA*	3	54,5%	27,3%	10,6%	3,8%	1,4%	0,6%
TPOX	4	4,9%	17,6%	29,3%	38,1%	44,0%	47,3%
D18S51	3	47,2%	11,5%	2,2%	0,4%	0,1%	0,0%
D5S818*	3	44,8%	49,9%	43,9%	35,6%	28,2%	21,8%
FGA	2	12,2%	1,2%	0,1%	0,0%	0,0%	0,0%

59

How to use the tool

- Graphical analysis

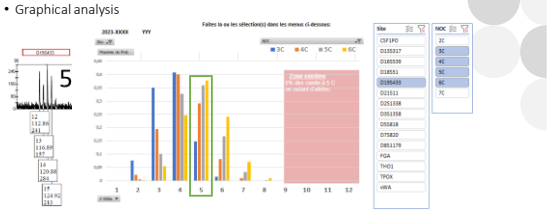




60

How to use the tool

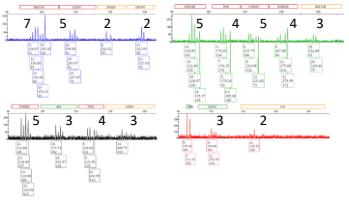
- Graphical analysis



61

How to use the tool

- MAC: ≥4C
- TAC: 3 or 4C
- LAC: >4C



62

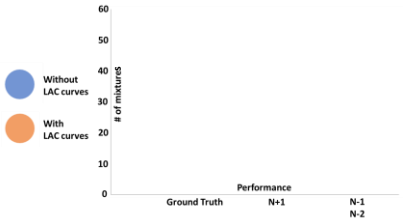
Does it help to improve NOC estimation?

- 3 DNA analysts
- 100 DNA mixtures from the Proved-IT degradation data set (> 3C)
- Each analyst:
  1. Used all the tools at his disposal to estimate the NOC
  2. Re-analyzed the DNA mixture, using all the tools **+ LAC curves**

*A single NOC was recorded for each DNA mixture.

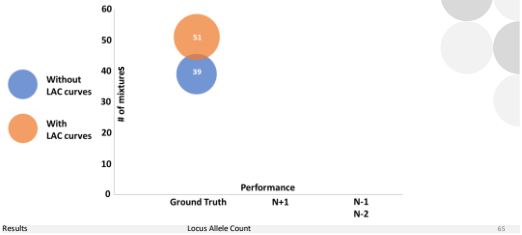
63

Does it help to improve NOC estimation?



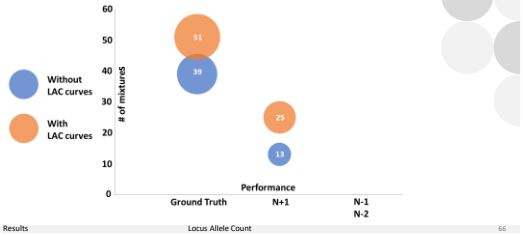
64

Does it help to improve NOC estimation?



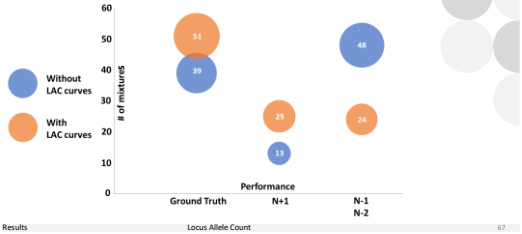
65

Does it help to improve NOC estimation?



66

Does it help to improve NOC estimation?



67

Limitations

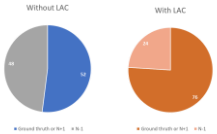
- Ground truth is not known for casework sample
  - Forced to choose a NOC
- Performed well using mixtures from the Proved-IT dataset
  - Compared to our caucasian population
    - Differences in allelic distribution
    - Higher diversity index
  - Tested the tool with ID+ (15 sites)
- We support the inclusion for LR > 1000

68

Conclusion

- We have developped a tool that is
  - ✓ Easy to use
  - ✓ Doesn't require any advanced software or coding skills to implement
    - Excel templates are available!
  - ✓ Leverage from the Total Allele Count method
- Version 2.0 coming soon!

Updating...



69

Acknowledgements

Tommy Harding

Nicolas Tremblay

Josée Noël

Sarah Noël

Valérie Clermont-Beaudoin

Diane Séguin

70

« N'avance rien
que tu ne sois capable de prouver »
Dr Wilfrid Derome

« Don't claim anything
you can't prove» Dr Wilfrid Derome

To communicate with me:

Email ……..marie-pier.thibault@msp.gouv.qc.ca

Web site …….. www.quebec.ca/lsjml

71