Implementation and Evaluation of 10X Genomics Chromium technology

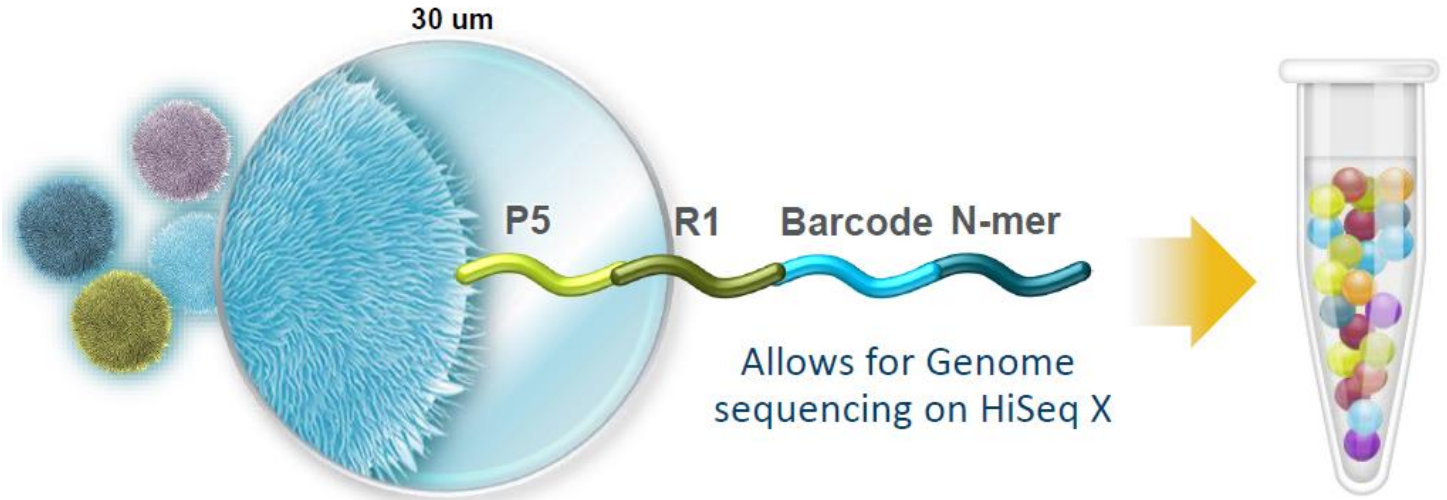
Adrien Castinel

03/10/2018

http://get.genotoul.fr
get@genotoul.fr
@get_genotoul

Chromium 10X GENOMICS

How does it work ?

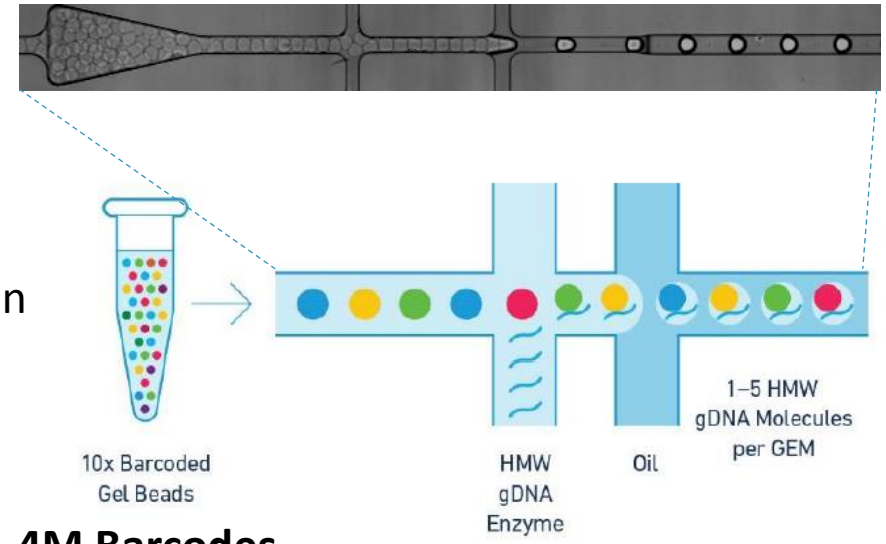


Chromium 10X GENOMICS

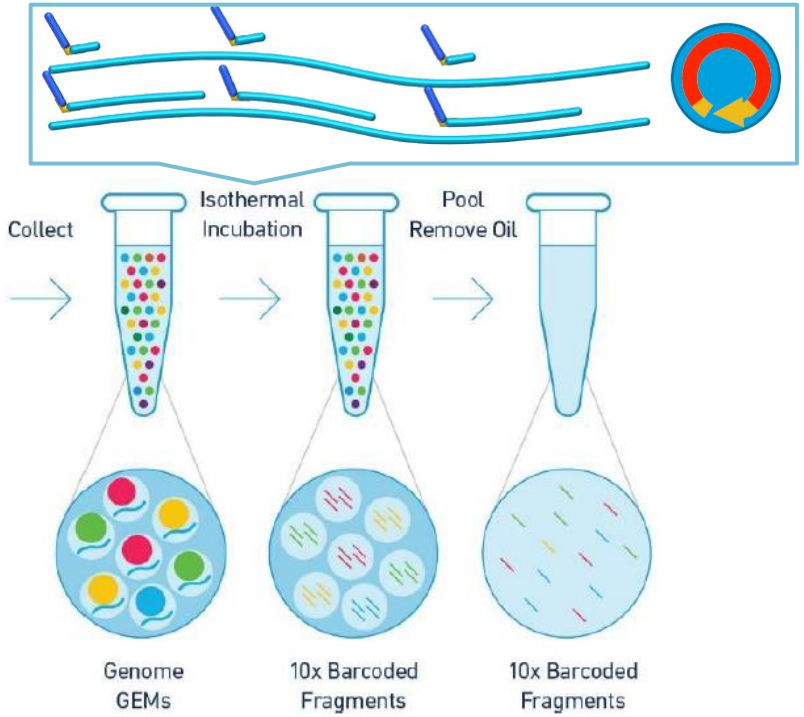
How does it work ?

Complete wetlab workflow: 2 days

GEM (Gel Bead in Emulsion)



4M Barcodes



Input: ~1 ng HMW gDNA

Linked-Reads

Sequencing on HiSeq3000 2x150pb

DNA Quality for 10X de novo sequencing

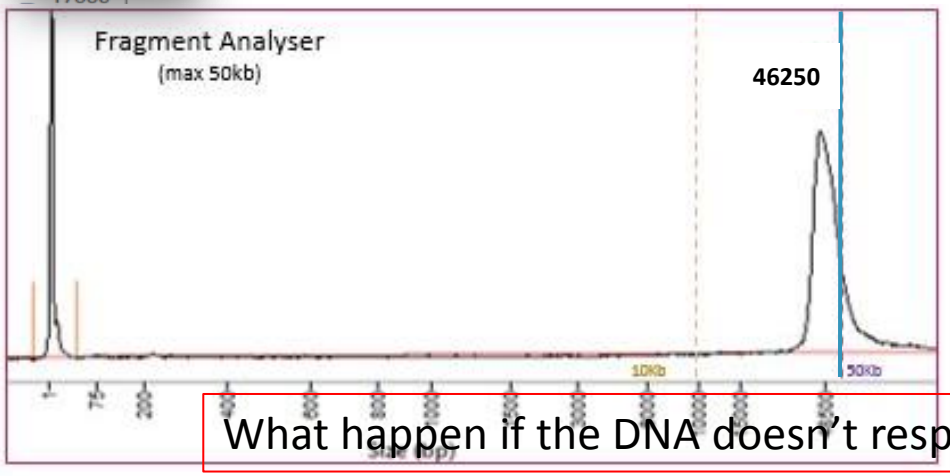
Prerequisites 10X de Novo sequencing

* Pré-requis ADN	
Qté min. de matrice (Qubit) :	5µg
concentration	environ 150 ng/µL
Volume minimum :	30 µL
Pureté 260/280 :	1,8-2
Pureté 260/230 :	2-2,2
Taille ADN minimum :	Centré autour 100 Kb

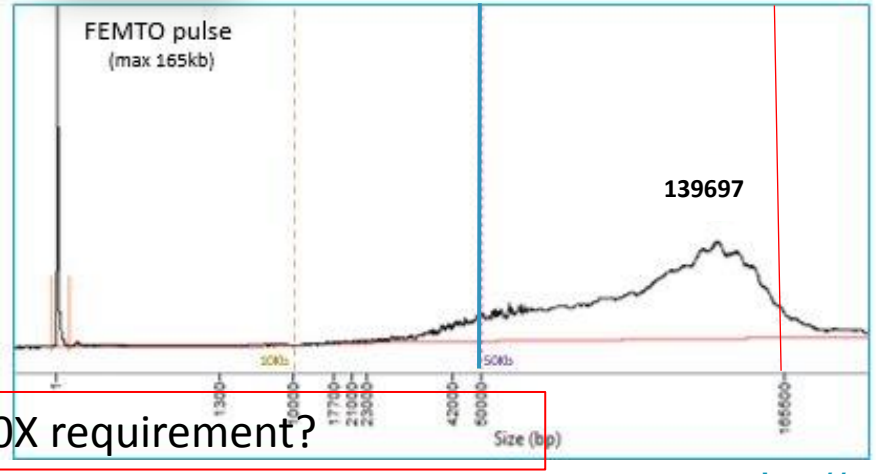
DNA QC : New Femto pulse
DNA extraction Profile on Femto (165Kb method)

138 930

2 ng/µL



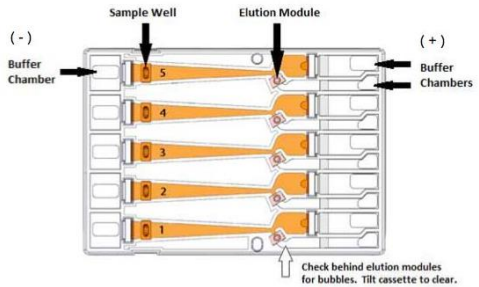
0,2 ng/µL



ets with 10X de
equirements

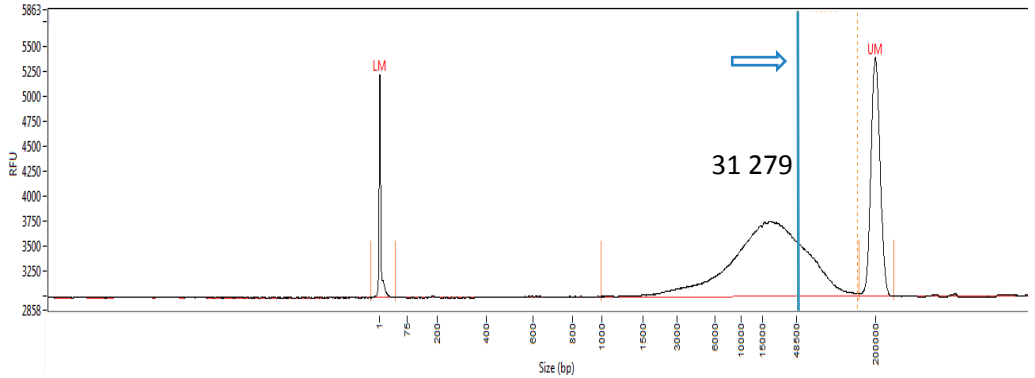
What happen if the DNA doesn't respect the 10X requirement?

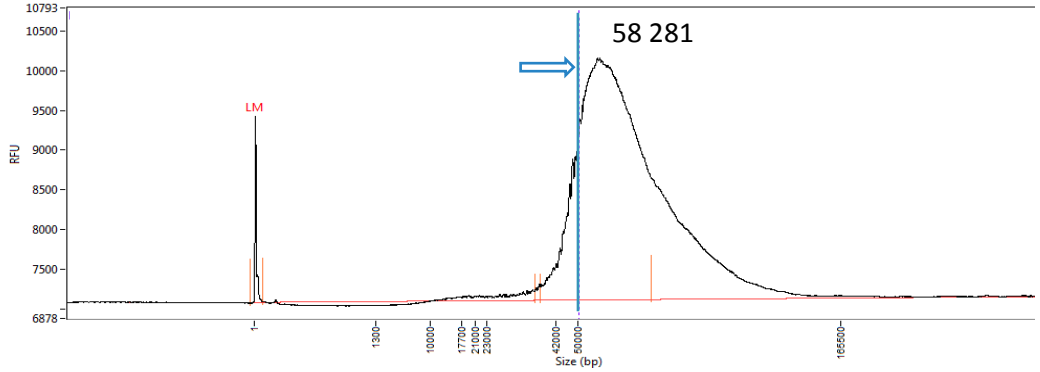
Bluepippin applications for 10X de novo sequencing



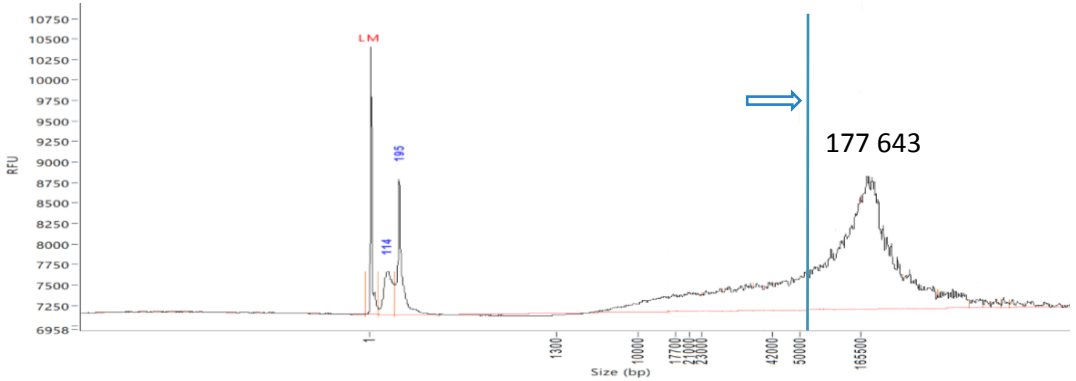
⇨ 50Kb sizing

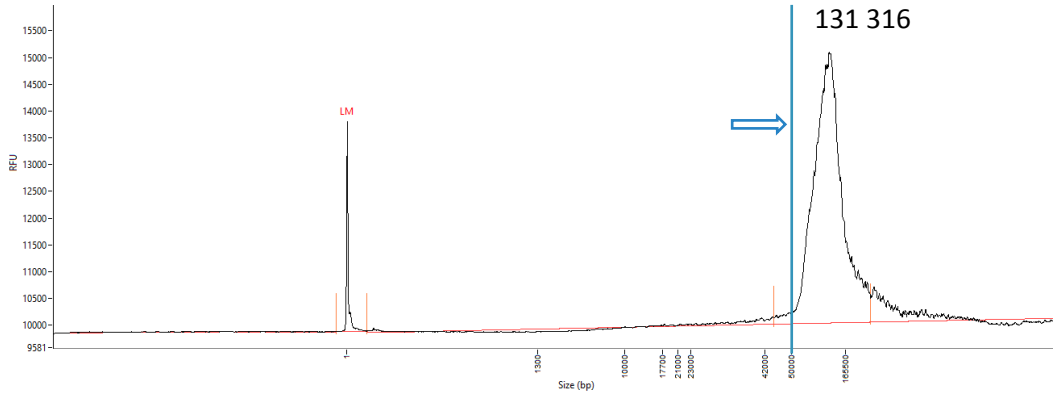
DNA fragments lower than 50Kb on FA and Femto





Degraded DNA with DNA fragments higher than 100Kb on Femto

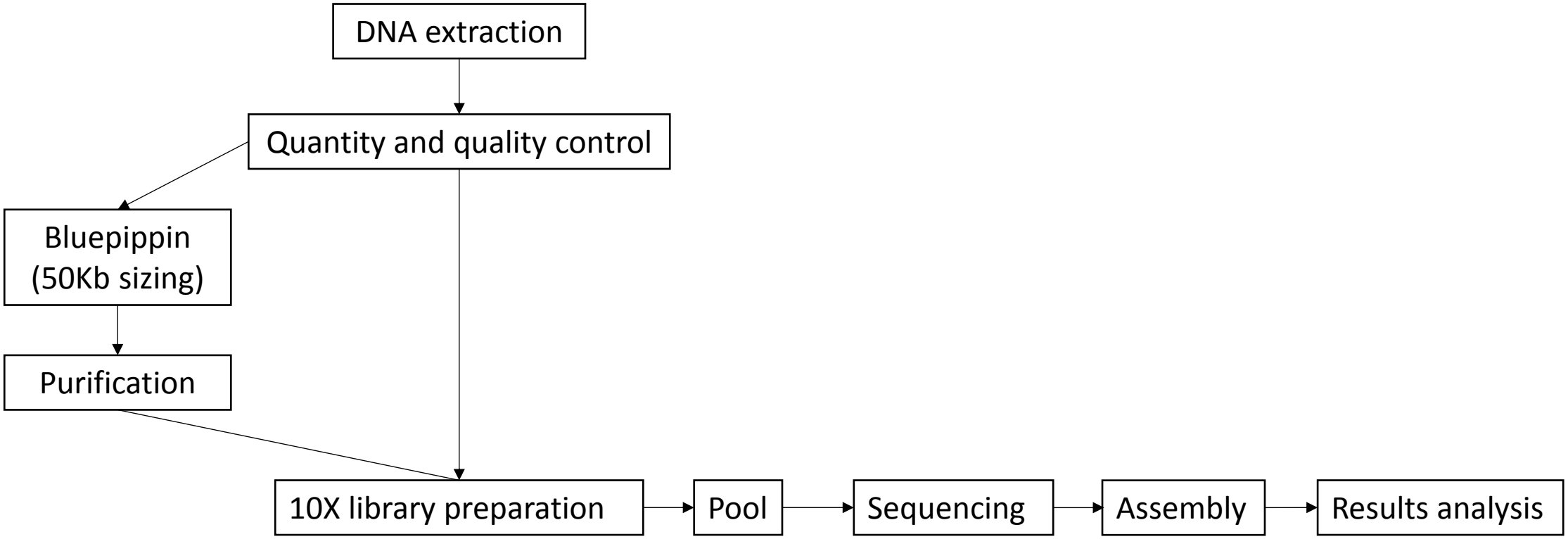




Does Bluepippin have an impact on 10X de novo sequencing assembly results?

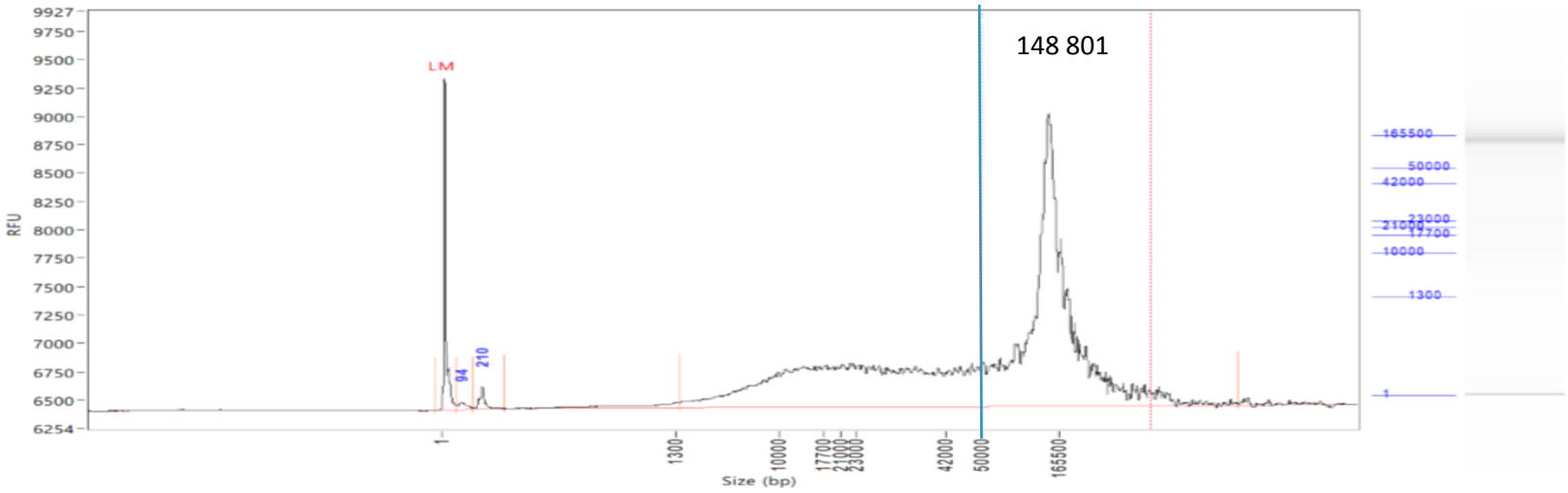
Does Bluepippin have an impact on 10X de novo sequencing assembly results?

- **Plant DNA – Genome size ~400Mb**
- **Extraction performed by William Marande of the CNRGV using Qiagen genomic tip**



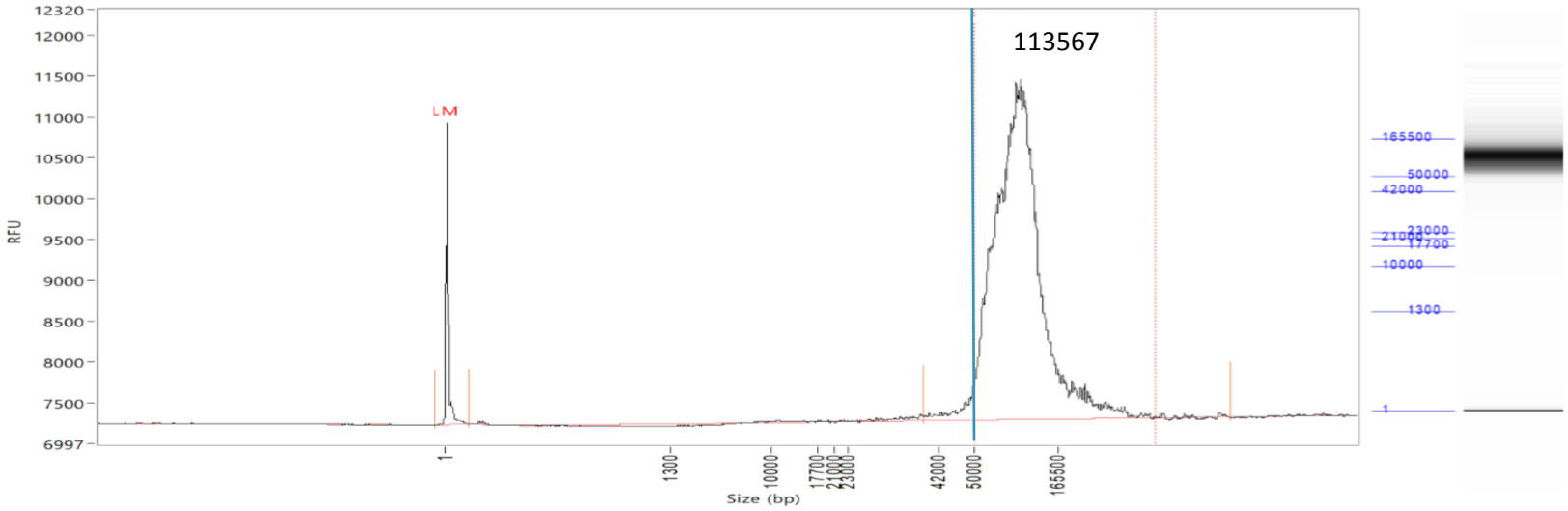
10 X de novo sequencing: Bluepippin (BPP)

DNA extraction profile on Femto



Smear Analysis 50000 bp to 300000 bp 0,0828 ng/ul 46,4 %Total 0,0009 nmole/L 152369 Avg. Size (b.p.) 31,03 %CV

50kb cut off profile on Femto



Smear Analysis 50000 bp to 300000 bp 0,1287 ng/ul 95,0 %Total 0,0019 nmole/L 112811 Avg. Size (b.p.) 32,34 %CV

Sample loading:

- Volume : 30µL
- Quantity : 3µg of DNA

Sizing by BluePippin

Sample collection :

- Volume : 120µL
- Quantity: 289ng

Big loss of DNA quantity
3µg => 289ng of DNA

10 X de novo sequencing : GEM

GEM loading

- The amount of DNA to be loaded is correlated to the genome size
- Genome Size = 0,4Gb
- Quantity : 0,625ng

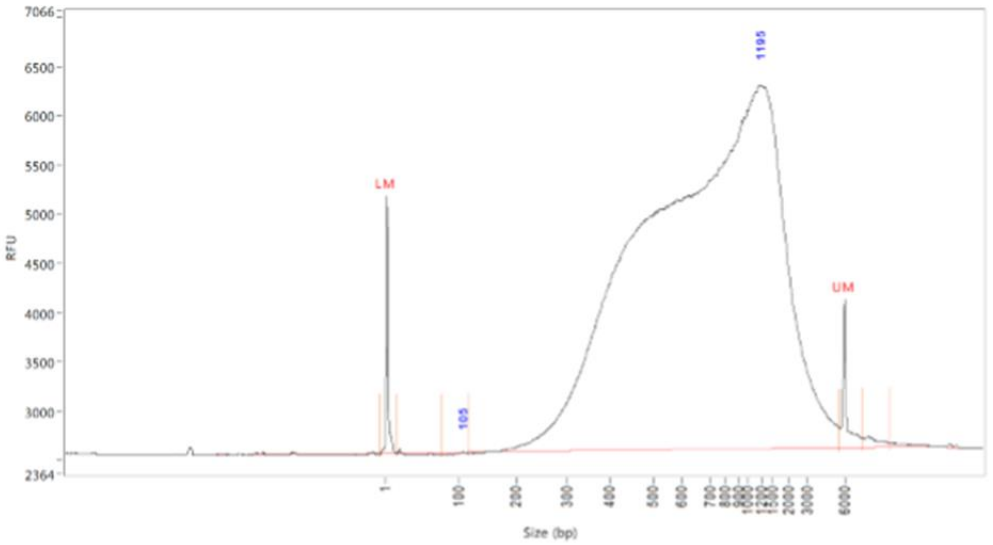
For genome 1,6 to 3,2 Gb charger entre 0,625ng et 1,25ng
For genome 0,1 to 1,6 Gb charger 0,625 ng

Rentrer la taille du génome dans la case jaune

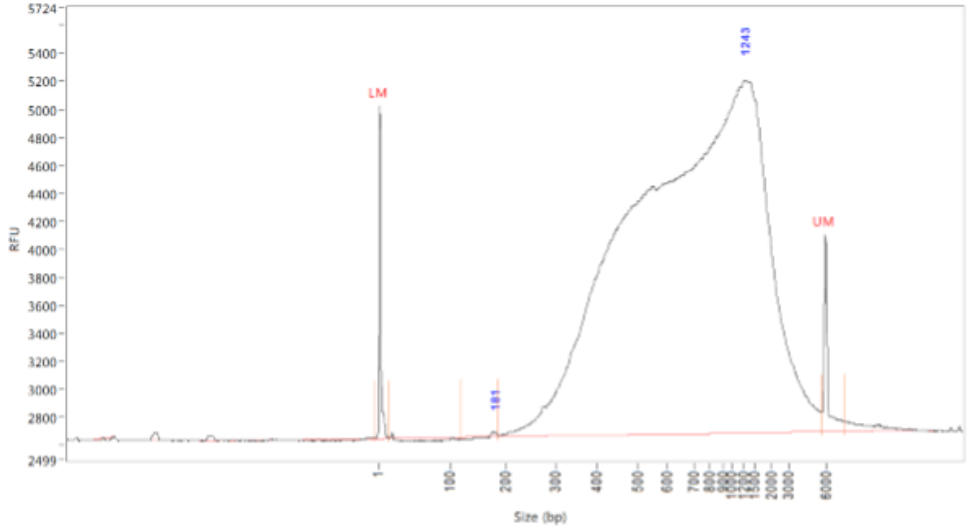
taille Genome (Gb)	quantité (ng) dans le puits
0,4	0,625

concentration de départ 0,5 ng/µl

GEM profile on FA



Plant_DNA

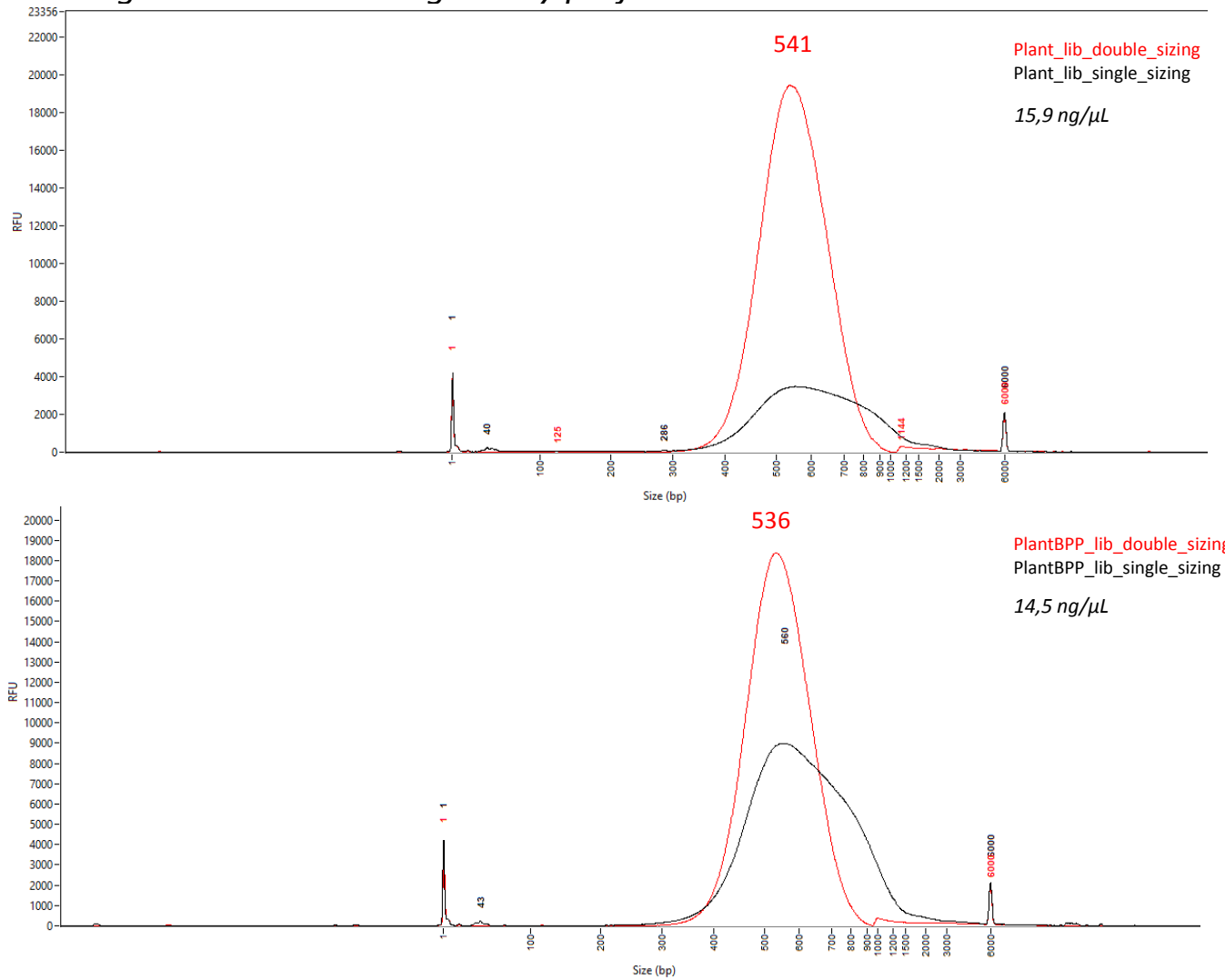


Plant_DNA + BPP

Only 0,625ng from the 3µg of DNA required for the BPP will be used for the GEM preparation

10 X de novo sequencing : library

Single and double sizing library profiles on FA



1 lane HiSeq3000
~ 100X cov

Supernova
De novo Assembly

All linked reads

No size difference between the libraries

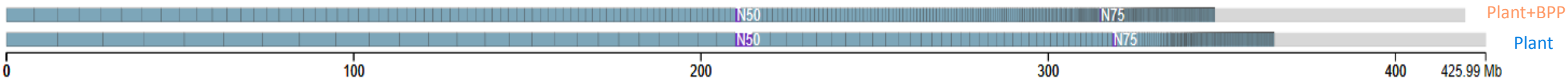
10 X de novo sequencing : Assembly

Metric Analysis

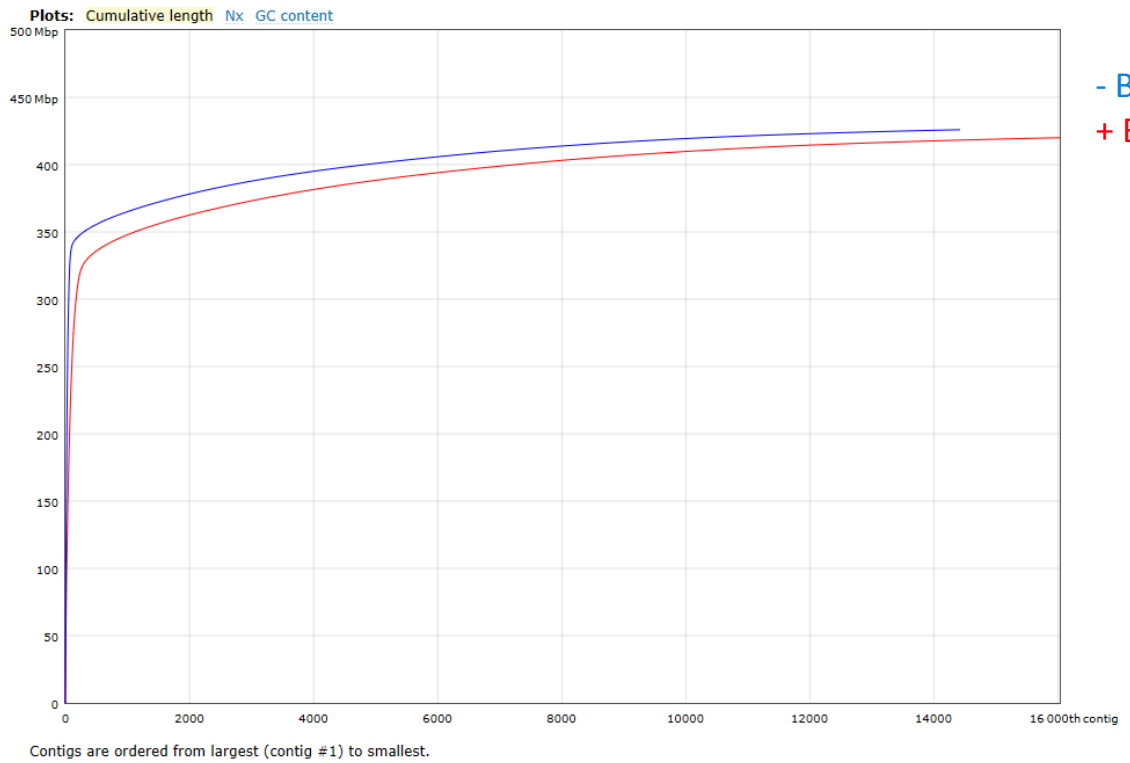
	Plant-BPP_56X	Plant_56X
Assembly size (bp) (scaffold>10Kb)	317 490	321 110
Contigs number	16 038	14 428
Largest contig (bp)	8 030 411	14 850 361
N50 (bp)	1 570 842	5 090 597
L50	74	26
Molecules mean size (bp)	48 290	74 820

Without Bluepippin:

- Better assembly
- Higher N50
- Lower L50
- Higher molecule mean size

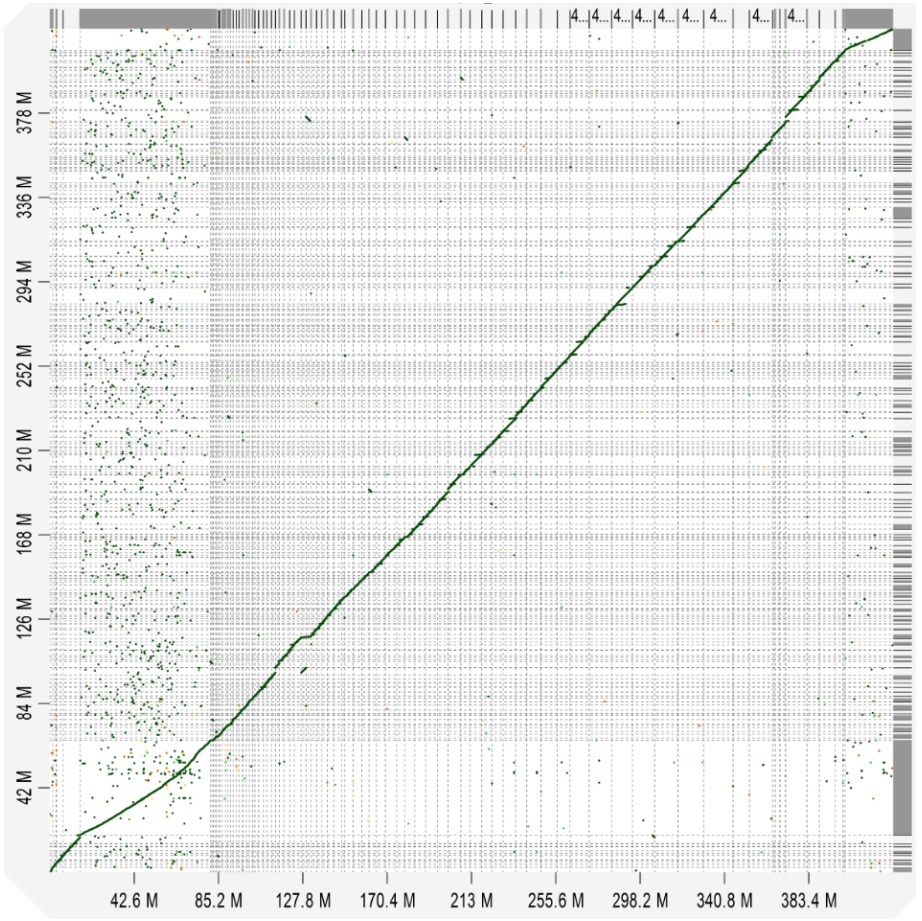


10 X de novo sequencing : Assembly



- BPP
+ BPP

Plant_56X



Plant-BPP_56X

- Contigs are larger without BPP with a higher cumulative length
- We are losing information between the two conditions

Dgenies (http://dgenies.toulouse.inra.fr/)

Conclusion

BPP Drawbacks (-)	BPP Advantage (+)
Requires high DNA quantity (3µg)	Enables to perform 10X when fragment size <50Kb
Removes large fragments (>100kb)	
Information loss for 10X de Novo assembly	
Time and cost loss	

- This test has to be repeated to confirm the results

Thanks!

NGS team:
Olivier Bouchez
ONT team:
Céline Roques
10X Genomics team:
Sophie Valière
Frédéric Martins
Bioinfo team:
Audrey Gibert
Claire Kuchly
Céline Vandecasteele
Maxime Manno

CNRGV:
William Marande

Adrien Castinel

03/10/2018

http://get.genotoul.fr
get@genotoul.fr
@get_genotoul

10 X de novo sequencing

Summary report 56X :

SUMMARY

- Tue Jul 17 15:58:48 2018
- [Nenuphar_56X]
- software release = 2.0.0(7fba7b4)
- likely sequencers = HiSeq3000/4000
- assembly checksum = 3,628,994,753,087,043,361

INPUT
- 149.34 M = READS = number of reads; ideal 800M-1200M for human
- 139.50 b = MEAN READ LEN = mean read length after trimming; ideal 140
- 47.32 x = RAW COV = raw coverage; ideal 0000098:0000056
- 37.83 x = EFFECTIVE COV = effective read coverage; ideal 0000094:0000042 for nominal 56x
- 80.40 % = READ TWO Q30 = fraction of Q30 bases in read 2; ideal 75-85
- 360.00 b = MEDIAN INSERT = median insert size; ideal 0.35-0.40
- 88.70 % = PROPER PAIRS = fraction of proper read pairs; ideal >= 75
- 1.00 = BARCODE FRACTION = fraction of barcodes used; between 0 and 1
- 476.56 Mb = EST GENOME SIZE = estimated genome size
- 17.35 % = REPETITIVE FRAC = estimated repetitive fraction
- 0.17 % = HIGH AT FRACTION = high AT index
- 74.82 Kb = MOLECULE LEN = weighted mean molecule size; ideal 50-100
- 498.18 = P10 = molecule count extending 10 kb on both sides
- 12.53 Kb = HETDIST = mean distance between heterozygous SNPs
- 6.11 % = UNBAR = fraction of reads that are not barcoded
- 122.00 = BARCODE N50 = N50 reads per barcode
- 7.68 % = DUPS = fraction of reads that are duplicates
- 70.37 % = PHASED = nonduplicate and phased reads; ideal 45-50

OUTPUT
- 1.85 K = LONG SCAFFOLDS = number of scaffolds >= 10 kb
- 13.20 Kb = EDGE N50 = N50 edge size
- 35.37 Kb = CONTIG N50 = N50 contig size
- 1.56 Kb = PHASEBLOCK N50 = N50 phase block size
- 4.92 Mb = SCAFFOLD N50 = N50 scaffold size
- 9.39 % = MISSING 10KB = % of base assembly missing from scaffolds >= 10 kb
- 321.11 Mb = ASSEMBLY SIZE = assembly size (only scaffolds >= 10 kb)

Summary report BPPP 56X :

SUMMARY

- Wed Jul 18 06:27:41 2018
- [NenupharBPPP_56X]
- software release = 2.0.0(7fba7b4)
- likely sequencers = HiSeq3000/4000
- assembly checksum = -3,162,400,894,115,605,908

INPUT
- 149.33 M = READS = number of reads; ideal 800M-1200M for human
- 139.50 b = MEAN READ LEN = mean read length after trimming; ideal 140
- 46.16 x = RAW COV = raw coverage; ideal 0000098:0000056
- 37.51 x = EFFECTIVE COV = effective read coverage; ideal 0000094:0000042 for nominal 56x
- 81.11 % = READ TWO Q30 = fraction of Q30 bases in read 2; ideal 75-85
- 355.00 b = MEDIAN INSERT = median insert size; ideal 0.35-0.40
- 86.74 % = PROPER PAIRS = fraction of proper read pairs; ideal >= 75
- 1.00 = BARCODE FRACTION = fraction of barcodes used; between 0 and 1
- 488.54 Mb = EST GENOME SIZE = estimated genome size
- 18.48 % = REPETITIVE FRAC = estimated repetitive fraction
- 0.16 % = HIGH AT FRACTION = high AT index
- 48.29 Kb = MOLECULE LEN = weighted mean molecule size; ideal 50-100
- 377.71 = P10 = molecule count extending 10 kb on both sides
- 14.20 Kb = HETDIST = mean distance between heterozygous SNPs
- 3.89 % = UNBAR = fraction of reads that are not barcoded
- 140.00 = BARCODE N50 = N50 reads per barcode
- 6.77 % = DUPS = fraction of reads that are duplicates
- 70.03 % = PHASED = nonduplicate and phased reads; ideal 45-50

OUTPUT
- 2.31 K = LONG SCAFFOLDS = number of scaffolds >= 10 kb
- 11.77 Kb = EDGE N50 = N50 edge size
- 32.26 Kb = CONTIG N50 = N50 contig size
- 1.24 Kb = PHASEBLOCK N50 = N50 phase block size
- 1.64 Mb = SCAFFOLD N50 = N50 scaffold size
- 10.94 % = MISSING 10KB = % of base assembly missing from scaffolds >= 10 kb
- 317.49 Mb = ASSEMBLY SIZE = assembly size (only scaffolds >= 10 kb)

Stats assemblage 10X supernova

N50 = 50Mb
L50= 7 séquences