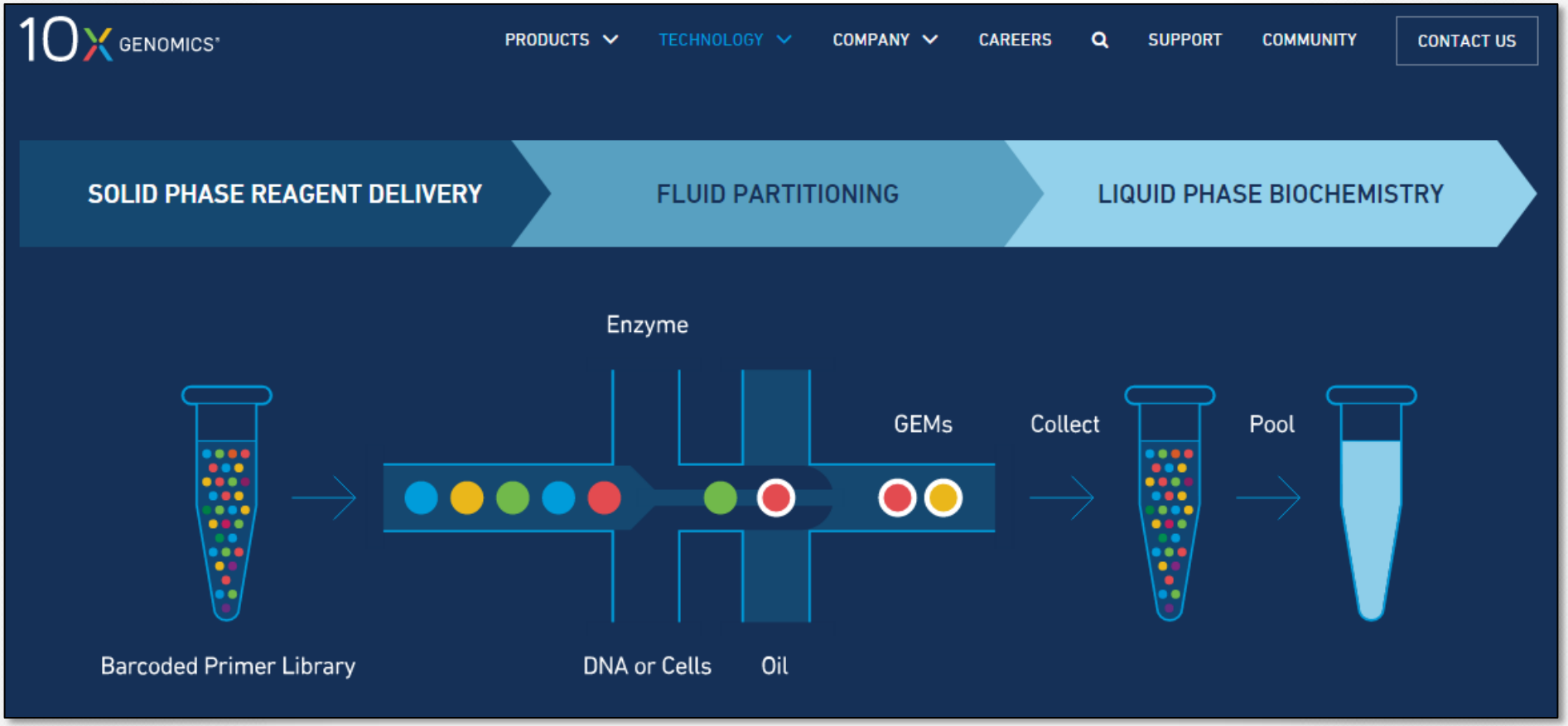
10x Chromium Synthetic Long Reads: principe et analyses

Corinne Cruaud

Benjamin Istace

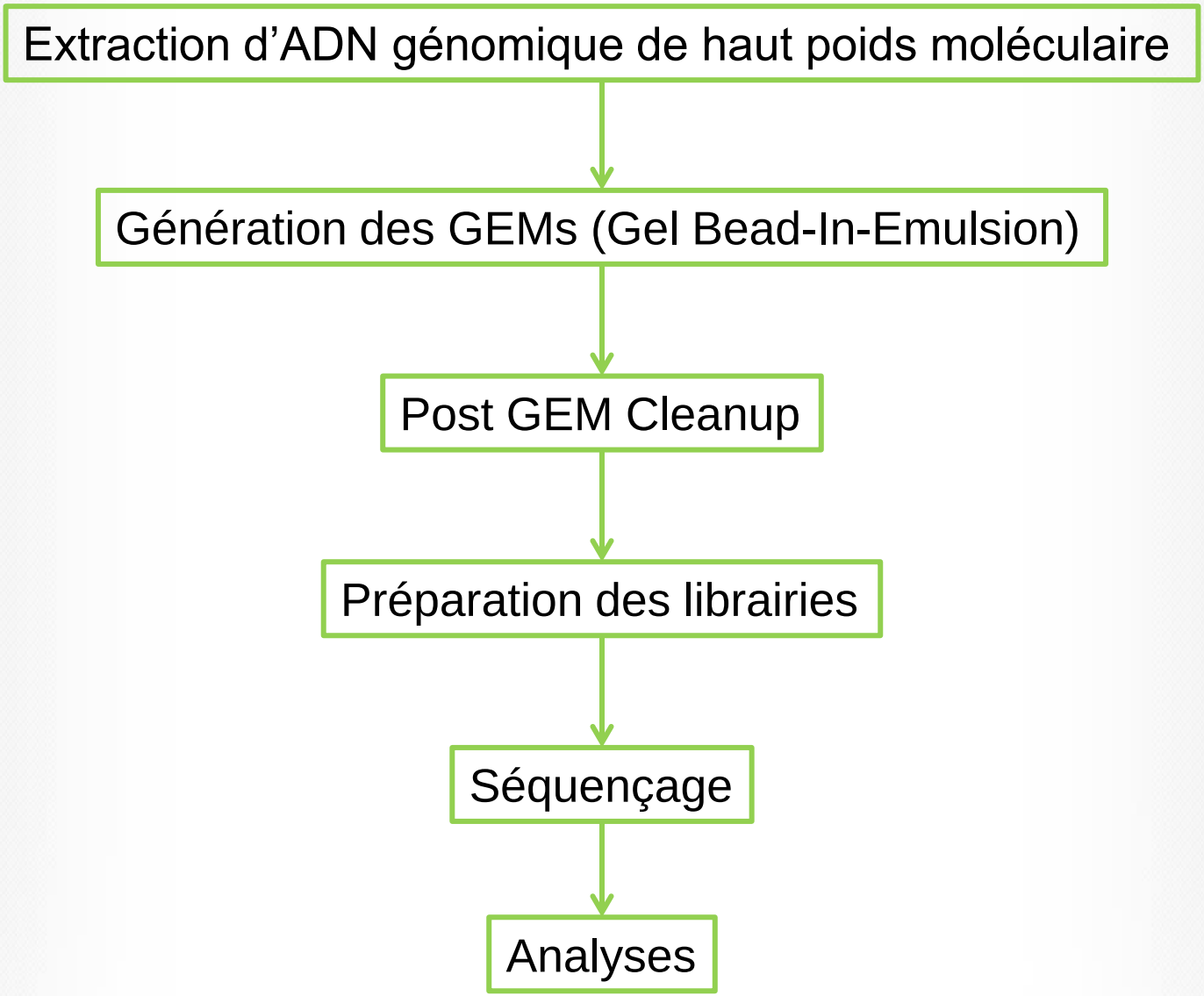
10x Genomics: principe



- Assemblage de génome de-novo
- Séparation des haplotypes
- Détection de variants structuraux

Pour des génomes de taille supérieure à 1Gb (1 à 3,2Gb)

10x Genomics: workflow



Recommandations:

Fragments de taille supérieure à 50Kb (mieux si > 80-100Kb)
Extraction avec le kit MagAttract HMW Kit (Qiagen) à partir de cellules en culture.

Différents protocoles proposés:

Salting Out Method for DNA Extraction from Cells
HMW DNA Extraction from Whole Blood
HMW DNA Extraction fron Fresh Frozen Tissue
DNA Extraction from Single Insects

Contrôle Qualité:

CHEF PFGE = la plus précise
Pippin Pulse = la plus économique
TapeStation = la plus rapide et la moins gourmande en ADN

Possibilité d'éliminer les fragments de taille inférieure à 20 ou 40Kb en utilisant le BluePippin (Sage Sciences)

Quantification de l'ADN génomique (Qubit):

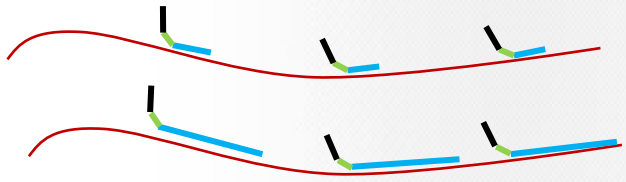
Il faut préparer idéalement 30µl d'ADN à une concentration de 5-20ng/µl.
2 dosages minimum sur 3µl d'ADN génomique
Puis 2 dosages minimum sur dilutions pour arriver à une concentration comprise entre 0,8 et 1,2ng/µl (avec une différence de concentrations inférieure à 15% entre les deux réplicats).

Dénaturation des échantillons puis **dépôt des ADNs et des réactifs sur la puce** (8 échantillons maximum par puce).

Run sur le **Chromium Controller** (20 minutes pour le programme génome) puis **amplification isothermale** (~3heures).

Puce 10x Genomics

Chromium Controller

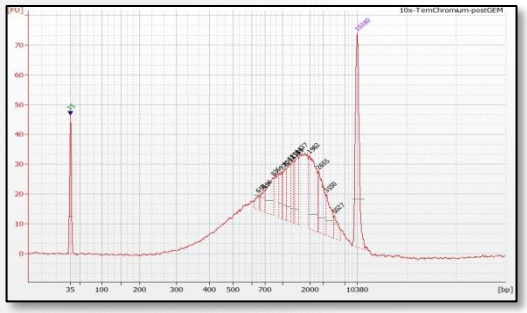


Amplification isothermale

Post GEM Cleanup

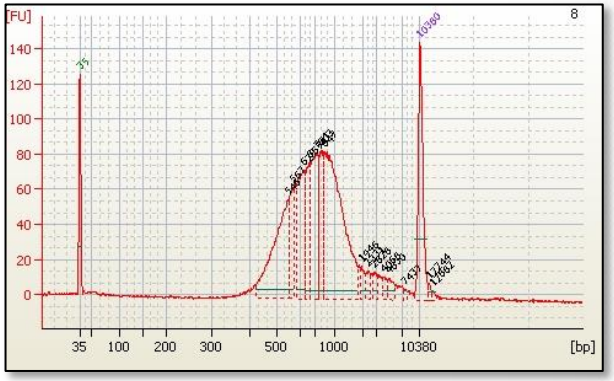
DynaBeads MyOne Silane
SPRIselect Reagent

Témoin positif kit (ADN humain)



5 – 8ng/µl

Préparation des librairies



27 – 28ng/µl
13 – 19 nM

Séquenceurs supportés par 10x Genomics:

Illumina® NovaSeq
Illumina® HiSeq X Ten/Five
Illumina® HiSeq 3000/4000
Illumina® HiSeq 2500 Rapid Run

Séquençage en **2x150bp**

Pour le génome humain: couverture de **40x** minimum.

1ng d'ADN engagé dans la manip **mais** nécessite un minimum de 10 à 20ng voire plus.

Qualité (taille des fragments) de l'ADN génomique importante (difficulté au niveau du contrôle qualité).

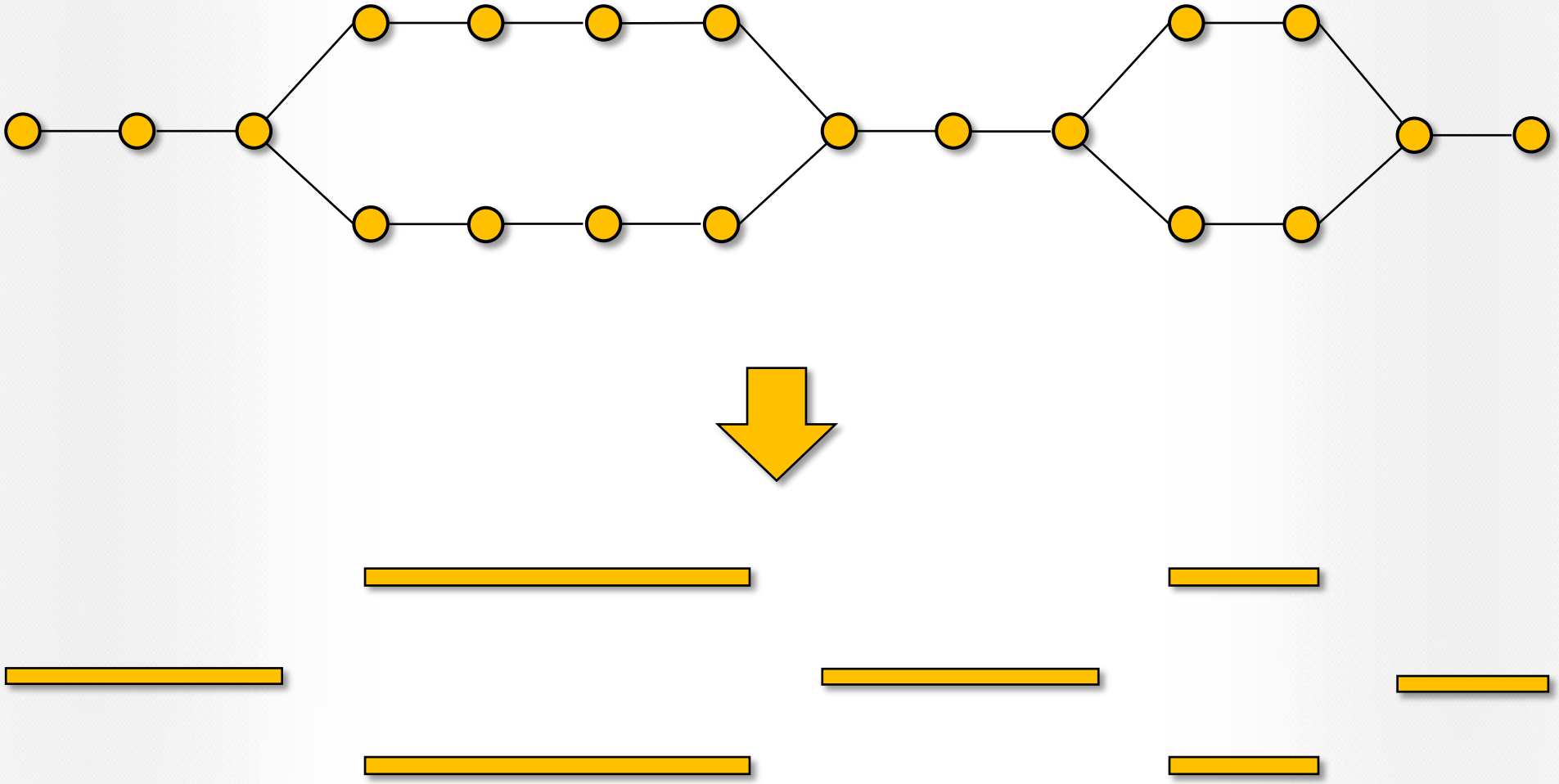
Prix pour un échantillon (réactifs 10x Genomics prix catalogue): 204 à 312 Euro en fonction du kit (96 ou 16 réactions).

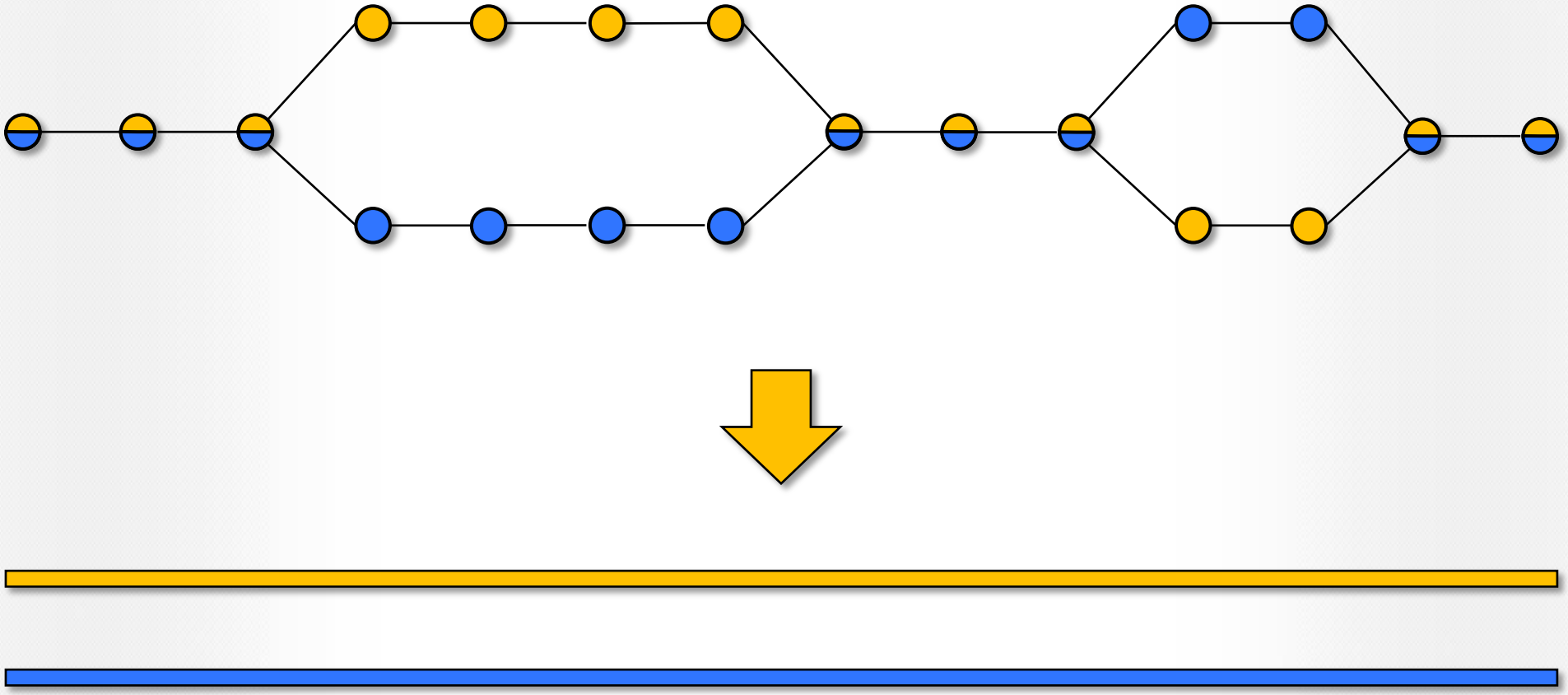
Reste la **solution de la dernière chance** pour les espèces pour lesquelles on ne peut extraire que très peu d'ADN.

Manips faites sur différentes espèces, en particulier sur l'ADN de tique et de Zanclus *cornutus*.

Genome assembly using 10X Genomics reads

10X Genomics - Haplotype phasing

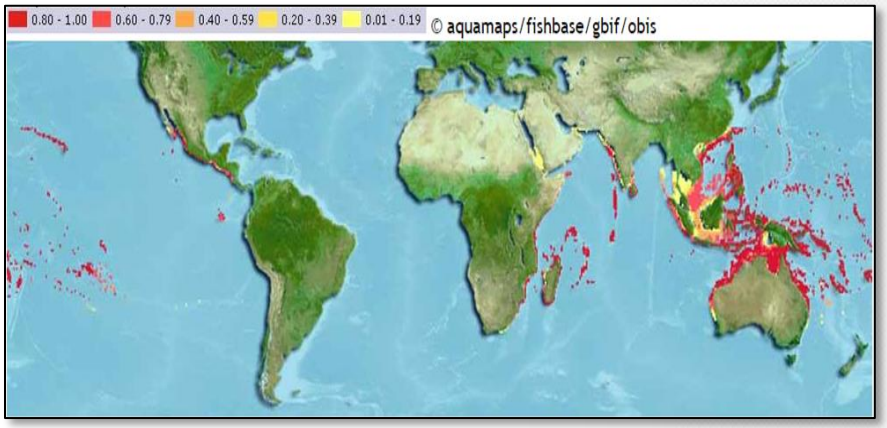




One scaffold per haplotype

Tara Pacific mission is to examine in a new way the biodiversity of coral reefs and their evolution in response to climate change and human activities.

Zanclus *cornutus*



Mostly found in the Indo-Pacific ocean

Genome is supposed to be around 700Mb in length and have a low heterozygosity.

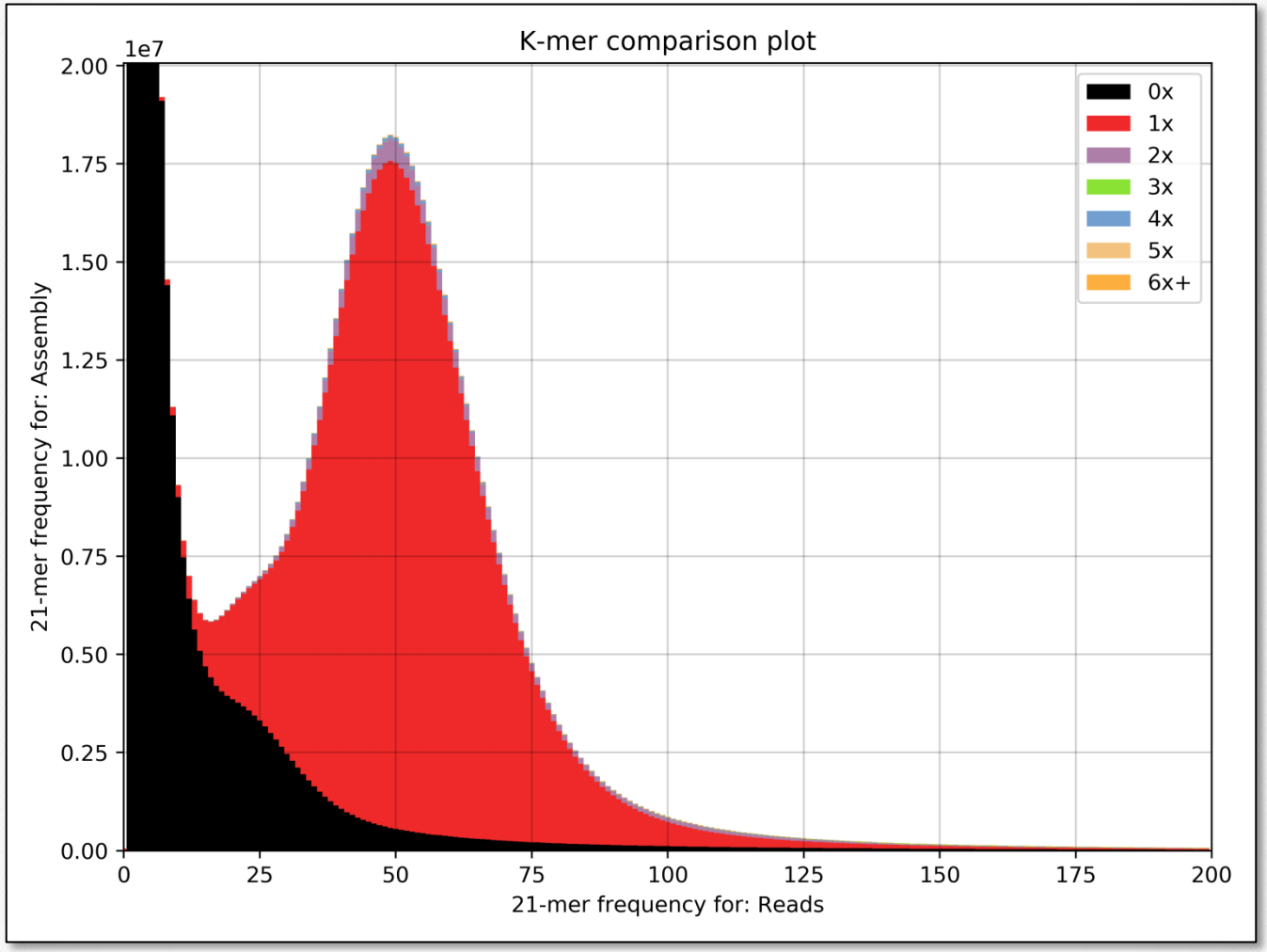
10X Genomics sequencing and assembly

Sequencing	
Number of reads (M)	568
Het. distance (b)	338
Coverage (x)	54.13
Read 2 Q30 (%)	74.32
Mean read length (b)	139
Mean molecule length (b)	23,590
Duplicates (%)	20.52

Assembly	
Cumulative size	754,762,749
Number of scaffolds	8,178
Maximum size	21,110,110
N50 (L50)	4,273,927 (51)
N90 (L90)	568,179 (223)
Number of Ns	20,321,630 (2.69 %)

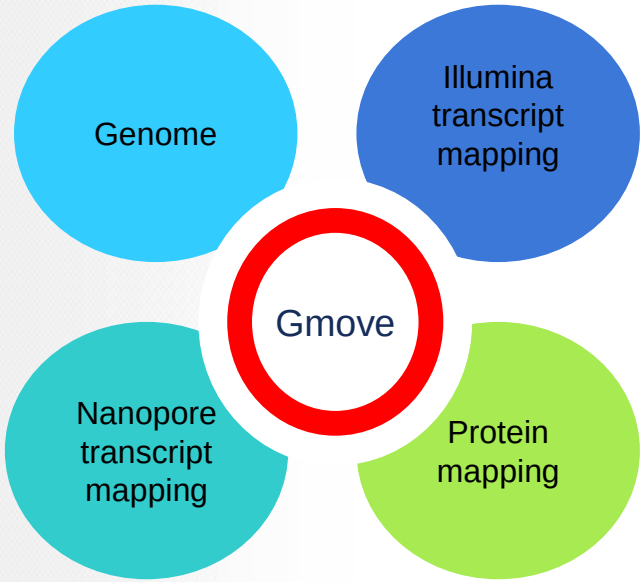
Assembly results are great even if molecule length is lower than expected

KAT analysis of the assembly



An overwhelming part of read *k*-mers are present in the assembly

Genome annotation



Annotation	
Gene number	26,440
Monoexonic genes	12.7%
Genes size (mean : med)	14,333.13 : 6,904
Number of exons/gene (mean : med)	9.18 : 6
CDS size (mean : med)	1,593.30 : 1,164

M. Dubarry talk on the 4th for the complete method

Annotation metrics seem to be on par with other available close species

Ticks harbour and transmit many different pathogens to vertebrate animals

Ixodes *ricinus*

Lyme disease is of growing concern for human health, yet much remains to be discovered about the genetic basis of tick vector competence

10X Genomics – Impact of the level of heterozygosity

Sequencing	Lab.	Wild
Number of reads (M)	1,200	1,200
Het. distance (b)	2,250	269
Coverage (x)	49.75	51.62
Read 2 Q30 (%)	74.21	82.00
Mean read length (b)	362	460
Mean molecule length (kb)	27	89
Duplicates (%)	27.33	28.76

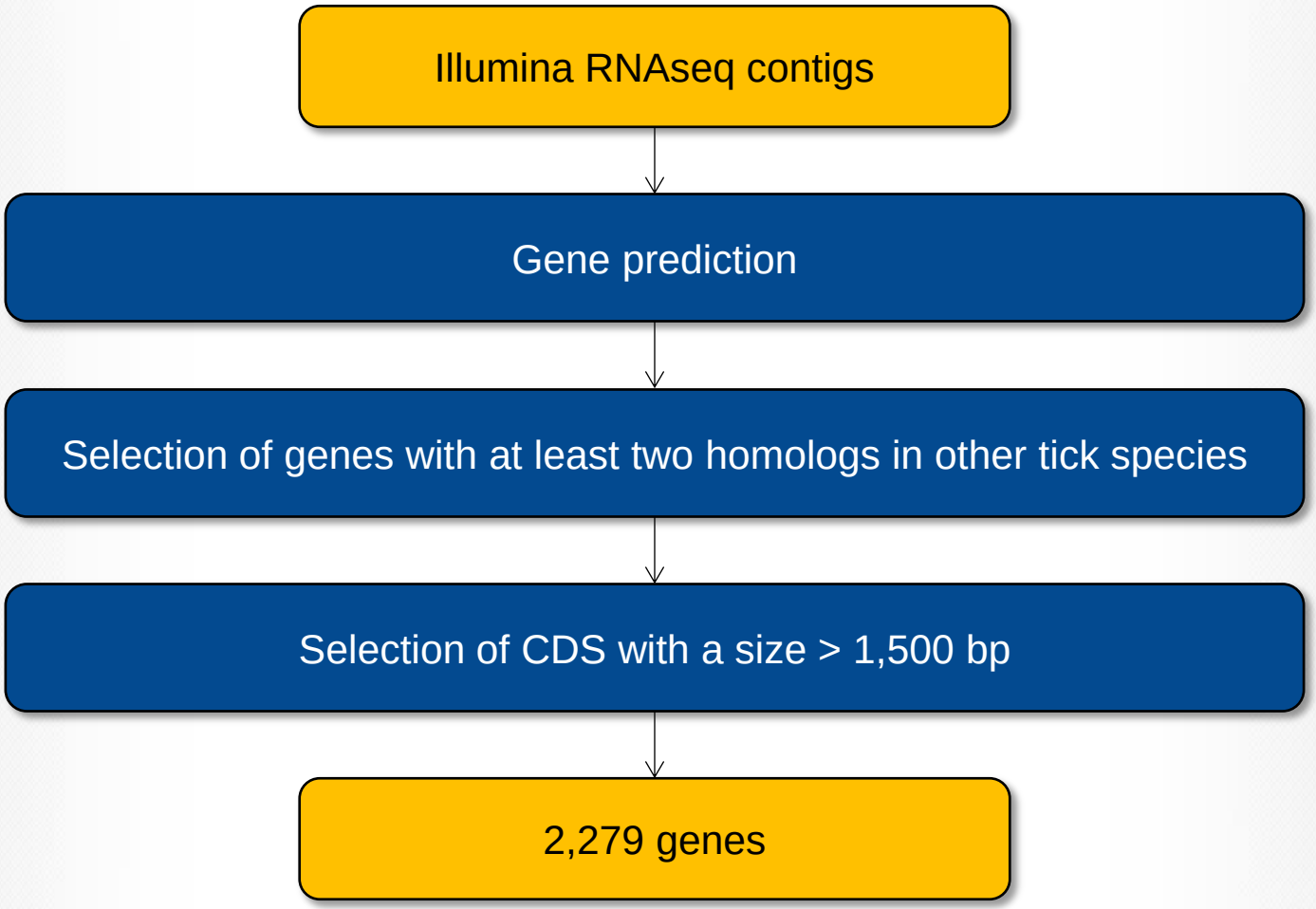
Wild individuals are much more heterozygous than their laboratory-raised counterparts

10X Genomics – Impact of the level of heterozygosity

Assembly	Lab.	Wild
Cumulative size	2,189,278,617	3,450,000,297
Number of scaffolds	183,663	165,714
Maximum size	3,104,849	4,462,173
N50 (L50)	207,468 (2,211)	99,848 (5,647)
N90 (L90)	3,098 (59,631)	6,678 (77,463)
Number of Ns	235,858,130 (10.77 %)	540,497,200 (15.66 %)

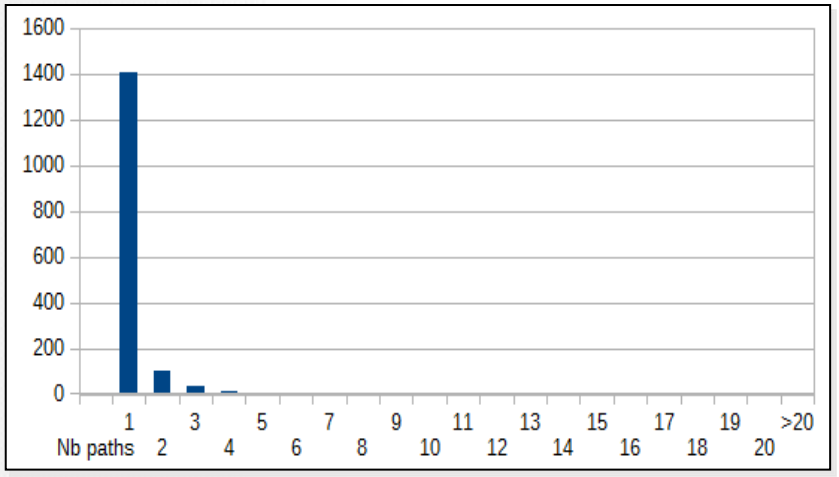
Despite longer DNA molecules, higher heterozygosity seem to lessen the quality of the assembly

Validation of the assembly

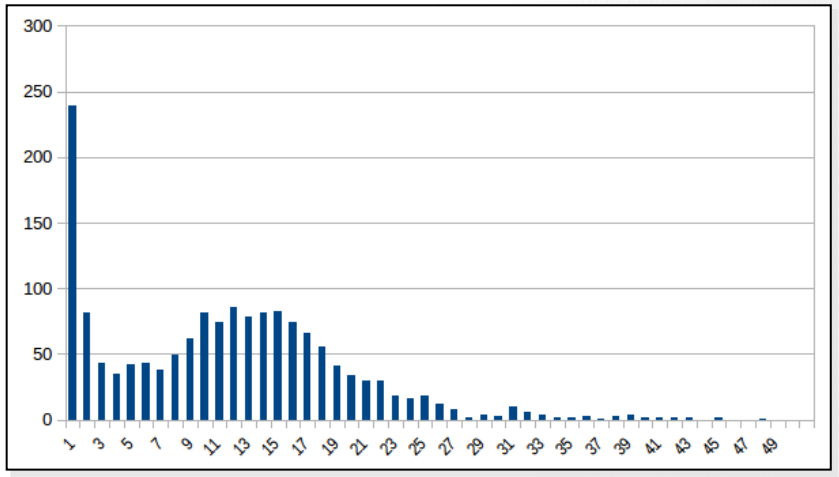


Slide courtesy of Claude Rispe, *INRA/Oniris, UMR BIOEPAR*

Only 32 genes (1.46 %) have no mapping on the genome



Repartition of the number of mappings per gene



Repartition of the number of exons per gene

Even if genes can be long and have a high number of exons, the assembly seems to be complete

Slide courtesy of Claude Rispe, *INRA/Oniris, UMR BIOEPAR*

- 10X Genomics offers tools for visualization and assembly that are easy to use. Downside is that not a lot of parameters can be tweaked.

- Genome assemblies seem to be complete and accurate

- 10X Genomics sequencing is a great approach to genome assembly if :
  - the genome is not too heterozygous
  - only low quantities of DNA can be extracted

Acknowledgements

Genoscope :

R&DBioSeq team
Sequencing laboratory
LAGE laboratory

CNRGH :

Céline Baulard
Marc Delepine

Collaborators on the Tara Pacific expedition

Collaborators on the GenIric Project