Montpellier SupAgro

www.supagro.fr

Amplicon sequencing by NGS methods

Jean-François Martin

Centre de Biologie pour la Gestion des Populations

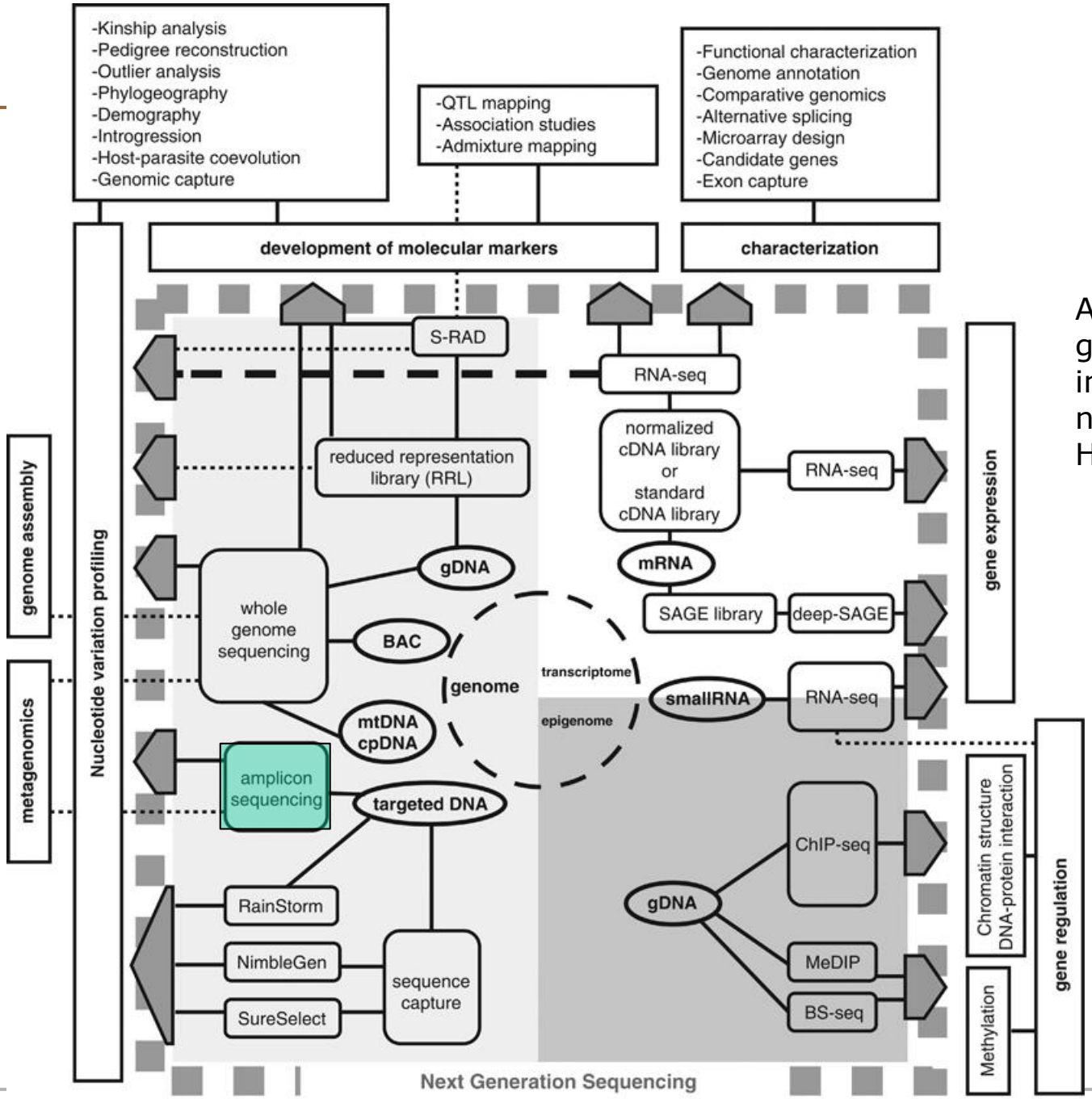
Centre international d'études supérieures en sciences agronomiques

Goals and expectations

The aim of this discussion today : review the different options to massively sequence amplicons through NGS technology and show some current work on adapting protocols.

Amplicon sequencing

Why sequence amplicons ?



Applications of next generation sequencing in molecular ecology of non-model organisms, Heredity, 2011

Amplicon sequencing

Most commonly used applications:

- Barcoding
- Metagenomics (including metabarcoding)
- Phylogeny
- Microsatellites genotyping (seriously?!)

Amplicon sequencing

Most of those applications could also be addressed by capture methods

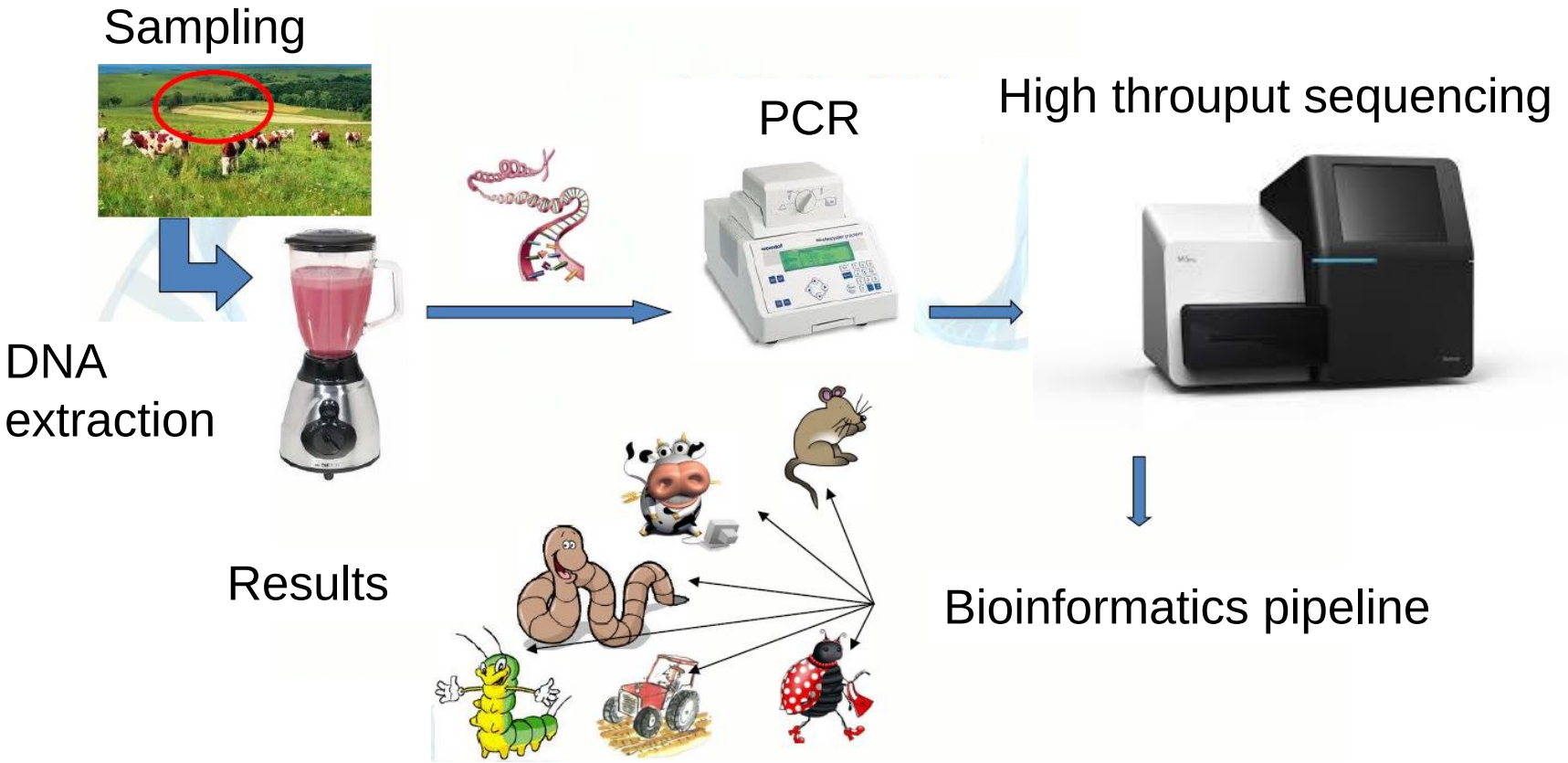
Amplicon sequencing

The applications will require different analysis methods. This is important in designing the project.

Think about the output information and format you will use for further analysis

Amplicon sequencing

Example among others : environmental barcoding



Amplicon sequencing

Goals and applications

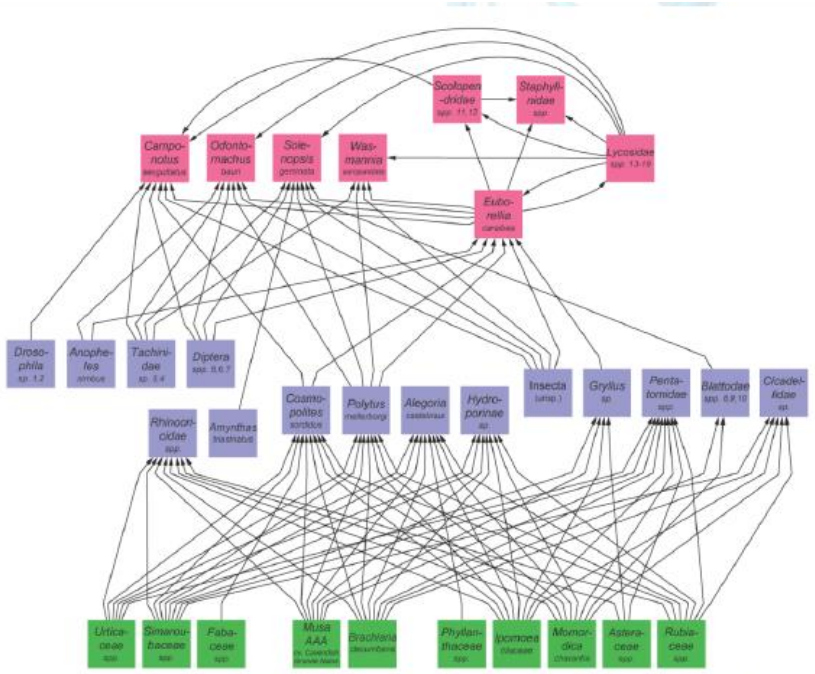
Characterization of the environmental species diversity

Characterization of parasites

Characterization of diet

Trophic network

Microbiome analysis



Amplicon sequencing

It is required to setup a simple and efficient methodology to acquire data so it makes amplicon sequencing accessible to anyone

Amplicon sequencing

What technology should we choose for sequencing amplicons ?

It depends !

but in any case error rate should be as low as possible

Amplicon sequencing

Preliminary testing of pacbio RS on long amplicons

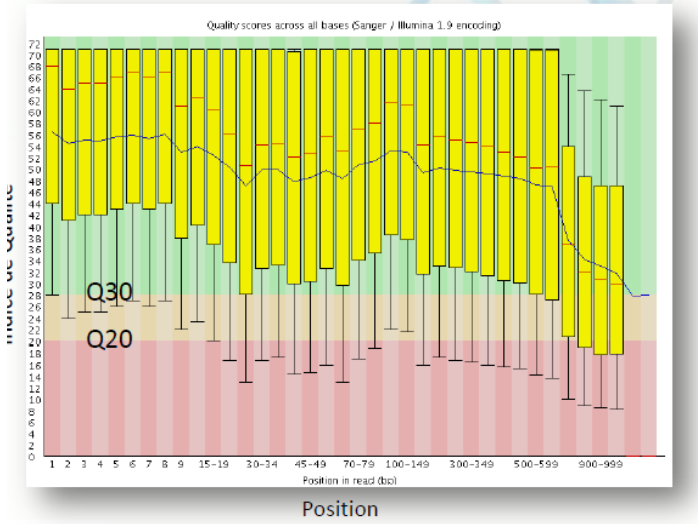
By Jean-François Martin

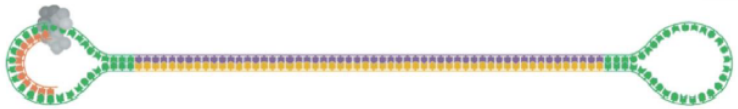
Amplicon sequencing

Pacbio RS sequencer

Preliminary testing of pacbio RS on long amplicons

- 100k sequences, including 19-21k ccs
- Up to 17k bases / sequence at the time (march 2014)
- Highly variable quality from one run to the next
- 15% eror rate on controls





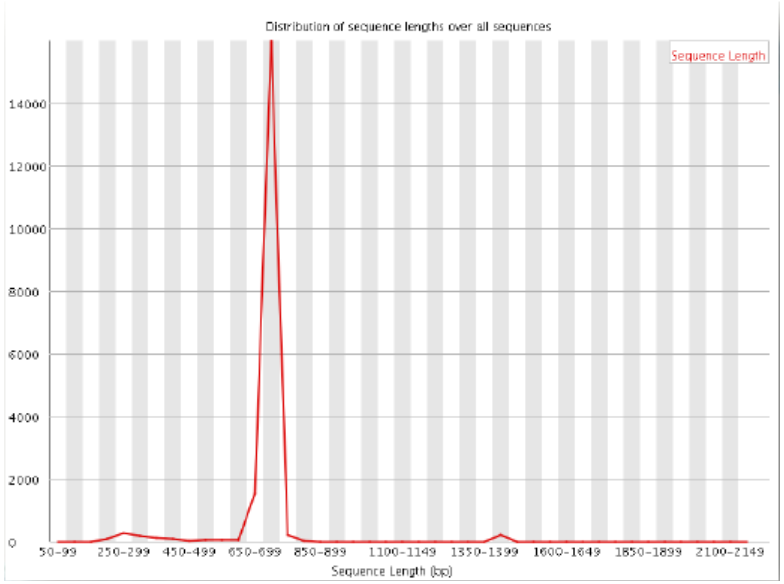
Circular consensus sequence (CCS)

Today on a Pacbio RS II, 15kb median, 40kb max

Amplicon sequencing

Preliminary testing of pacbio RS on long amplicons

98% of the css fragments are of the correct size (658pb)



Amplicon sequencing

Preliminary testing of pacbio RS on long amplicons

- 42% of quantitative variation between technical replicates
- Majoritary sequence (17% in average) always correspond to the expected
- Random error based correction always recovers the correct sequence
- The number of errors/sequence never exceeds 2 (out of 658 bases)
- When mixed samples occur, the reproducibility of ratios is very low

Amplicon sequencing

Preliminary testing of pacbio RS on long amplicons

Conclusion : SMRT is very useful for barcoding long amplicons (658pb) but not usable in environmental applications.

Although compared to sanger it is still expensive

Amplicon sequencing

Designing amplicon sequencing strategy with Miseq sequencing

By Jean-François Martin

Amplicon sequencing

Designing amplicon sequencing strategy with Miseq sequencing

Goal : simultaneously sequence up to thousands of amplicons

- How many sequences for each amplicon ?
- How do I multiplex the amplicons ?
- What are the limits of the technology ?

Amplicon sequencing

Designing amplicon sequencing strategy with Miseq sequencing

How many sequences for each amplicon ?

- What is the expected variation for an amplicon ?
- How many sequences to validate a variant ?
- Also means that all amplicons are equitably represented in the sequencing run and all sequences are usable

Amplicon sequencing

Designing amplicon sequencing strategy with Miseq sequencing

How do I multiplex the amplicons?

What multiplexing level do I need ?

What are the built-in multiplex limits for Illumina ?

How do I combine amplification and sequencing ?

Amplicon sequencing

Designing amplicon sequencing strategy with Miseq sequencing

What are the limits of the technology?

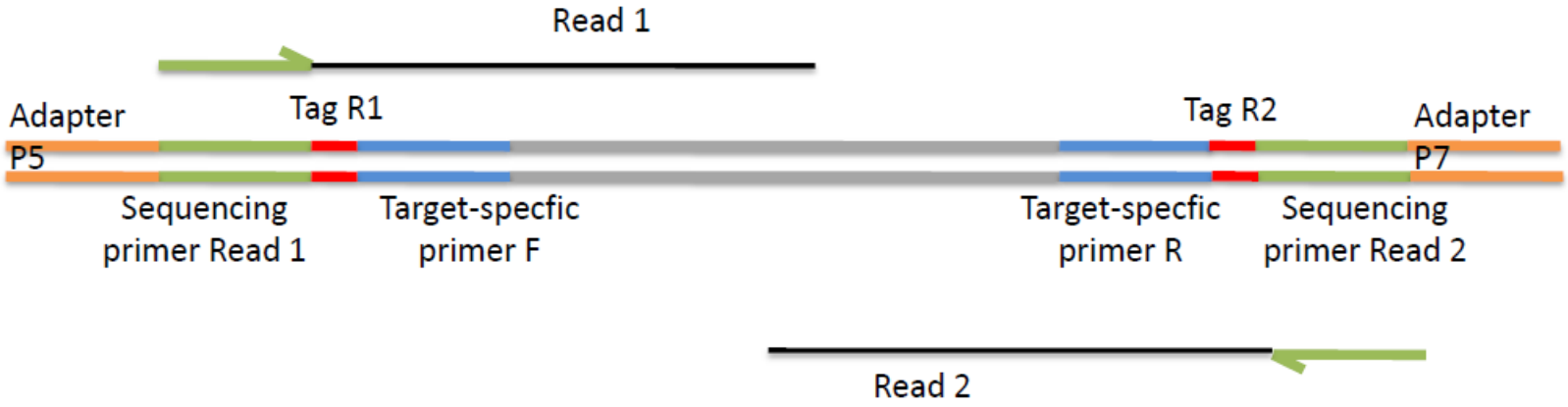
Are there artefacts that matter for my application ?

How many amplicons can I sequence in a given period of time ?

What are the costs (direct and indirect) for the different strategies ?

Amplicon sequencing

The 1-step tagged PCR + library strategy



Step 1: PCR + Tag

Step 2: Pooling + Ligation (kit TrueSeq)

Amplicon sequencing

The 1-step tagged PCR + library strategy

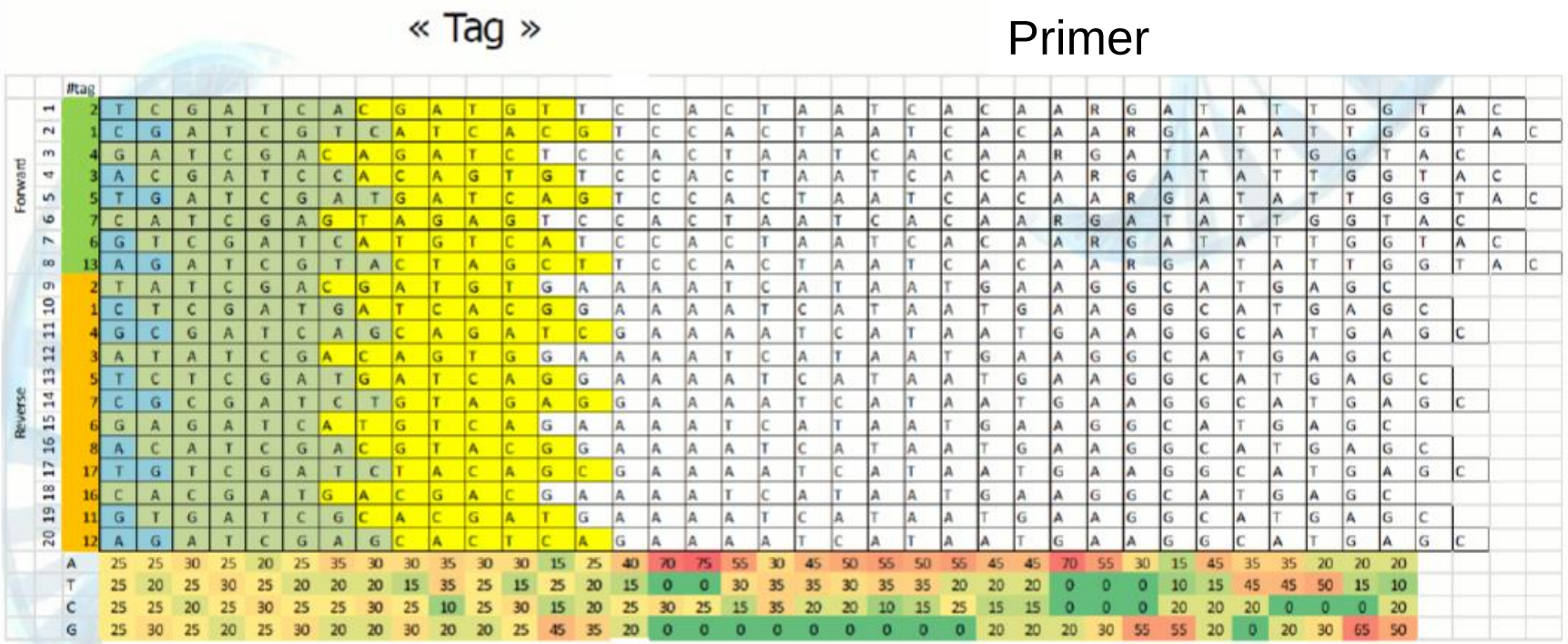
- Amplify the target with tagged primers
- Pool the amplicons by marker / size ...
- Prepare a library for each pool
- Sequence (Miseq v3)

Amplicon sequencing

Step 1 : amplification process

« Tag »

Primer



Percentage for each base

Amplicon sequencing

The 1-step tagged PCR + library strategy

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
F1	F1R1	F1R2	...									
F2												
F3												
F4												
F5												
F6												
F7												
F8												

Amplicon sequencing

The 1-step tagged PCR + library strategy

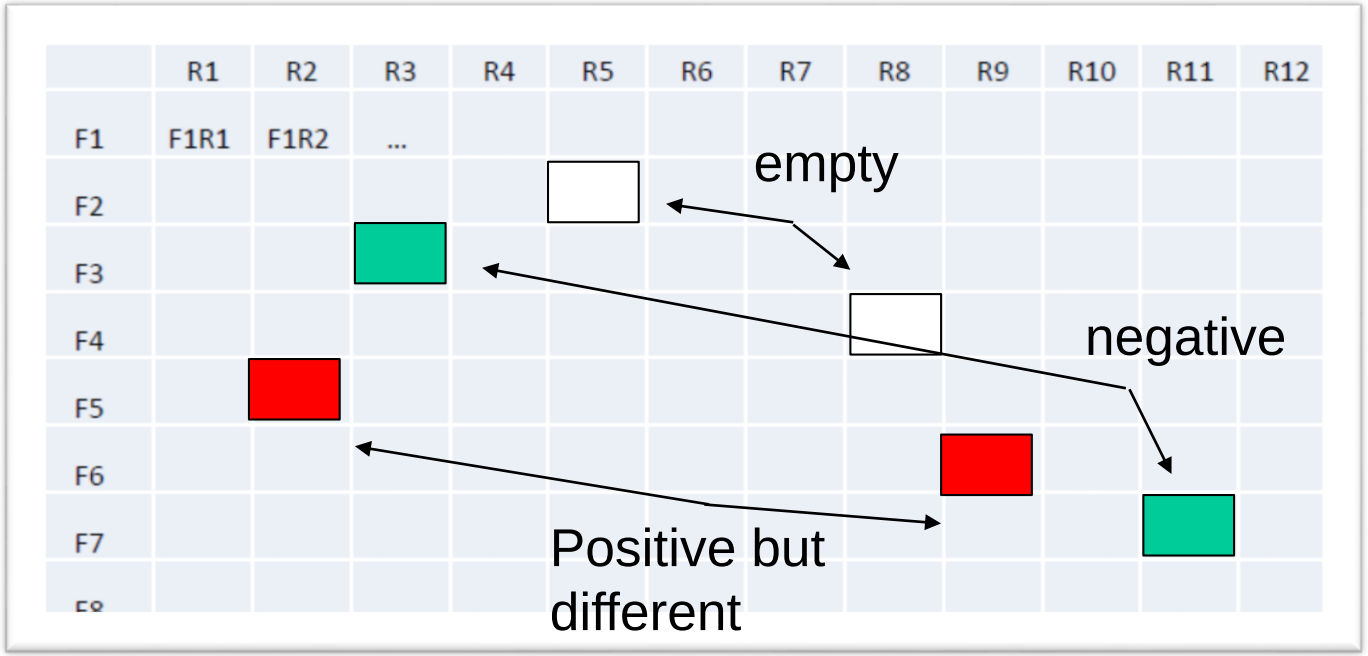


Plate design

Amplicon sequencing

The 1-step tagged PCR + library strategy

How many tag combinations and how many indexes ?

In other words should I maximise the number of PCR ?

For reference, 1 index / library -> 90€ + handling effort

Amplicon sequencing

Step 2 : combining amplicons as pools

How do I normalize the amplicons before pooling ?

old-school method : they all look the same on my gel, I pool them as is.

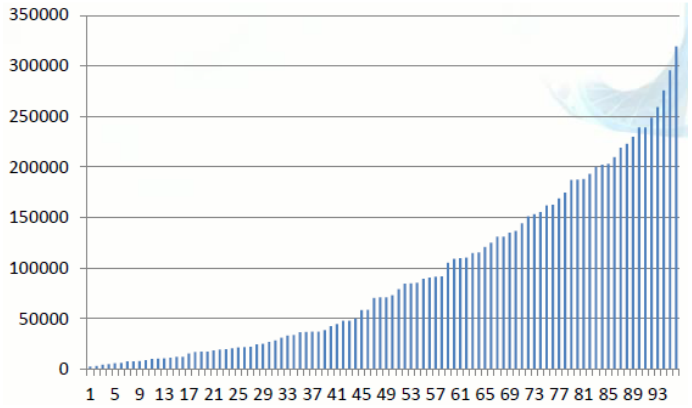
refined old-school : I create a couple of categories according to intensities on gel and pool different volumes accordingly

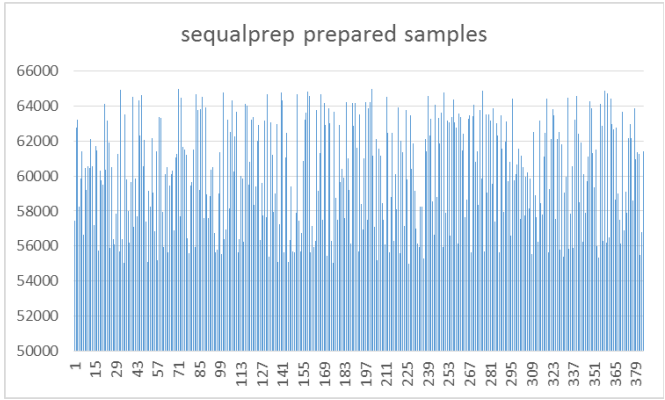
Amplicon sequencing

Step 2 : combining amplicons as pools

How do I normalize the amplicons before pooling ?

the « what I can I do best for my precious ?» strategy, uses a sequalprep kit to retain known quantity of amplicon (25ng) then pool them





Amplicon sequencing

Step 2 : combining amplicons as pools

It is advised not to mix different size amplicons as it would favor the shortest ones in the process (PCR if any during the library and sequencing).

1 pool = 1 library

Should I do replicates ?

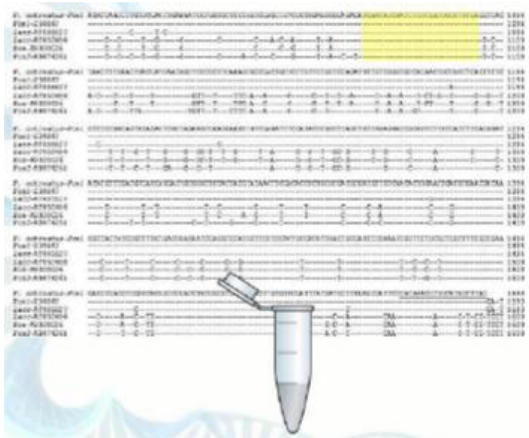
Amplicon sequencing

Step 2 : combining amplicons as pools

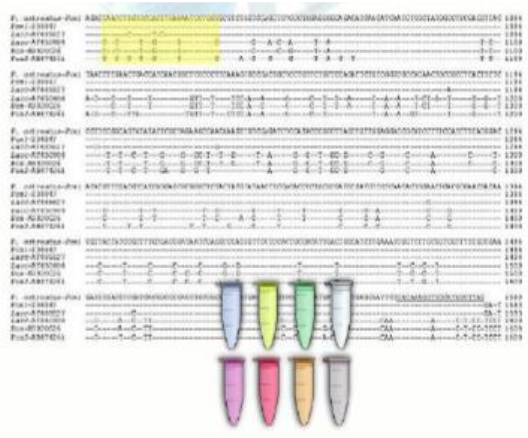
Common problems are linked to successfully amplify a wide range of species with an equal efficiency

Improve efficiency

Universal primers



Primers cocktail



Amplicon sequencing

Step 2 : combining amplicons as pools

Gel purification of the pools

Qualifying the pools on bioanalyzer

Quantifiying the pools with what is convenient (nanodrop)

Normalize pools to 0.4ng/µL (20 ng in 50µL)

Amplicon sequencing

Step 3 : building libraries for each pool

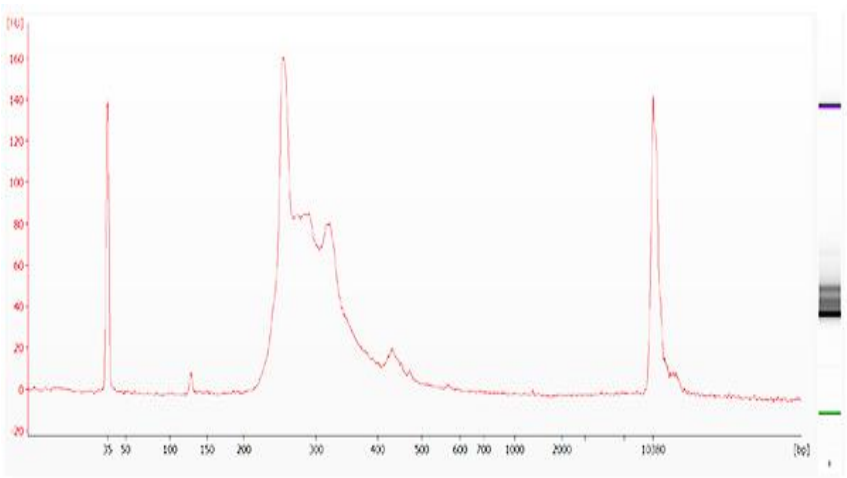
This protocol uses part of the Trueseq nano LT kit (4h)

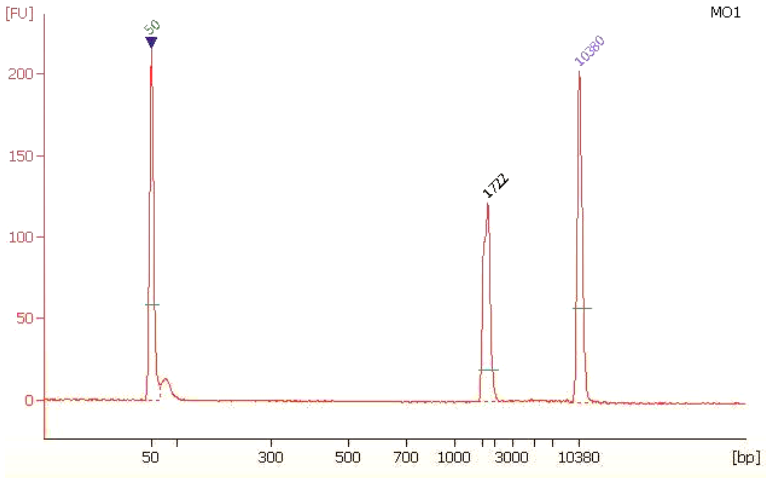
1. End-repair to make the amplicons blunt-ended.
2. A-tailing to allow further ligation
3. Ligate the illumina adapters (holding indexes)
4. Enrich in P5-P7 libraries through amplification

Amplicon sequencing

Step 3 : building libraries for each pool

When the libraries are built, I need to qualify them using the bioanalyzer

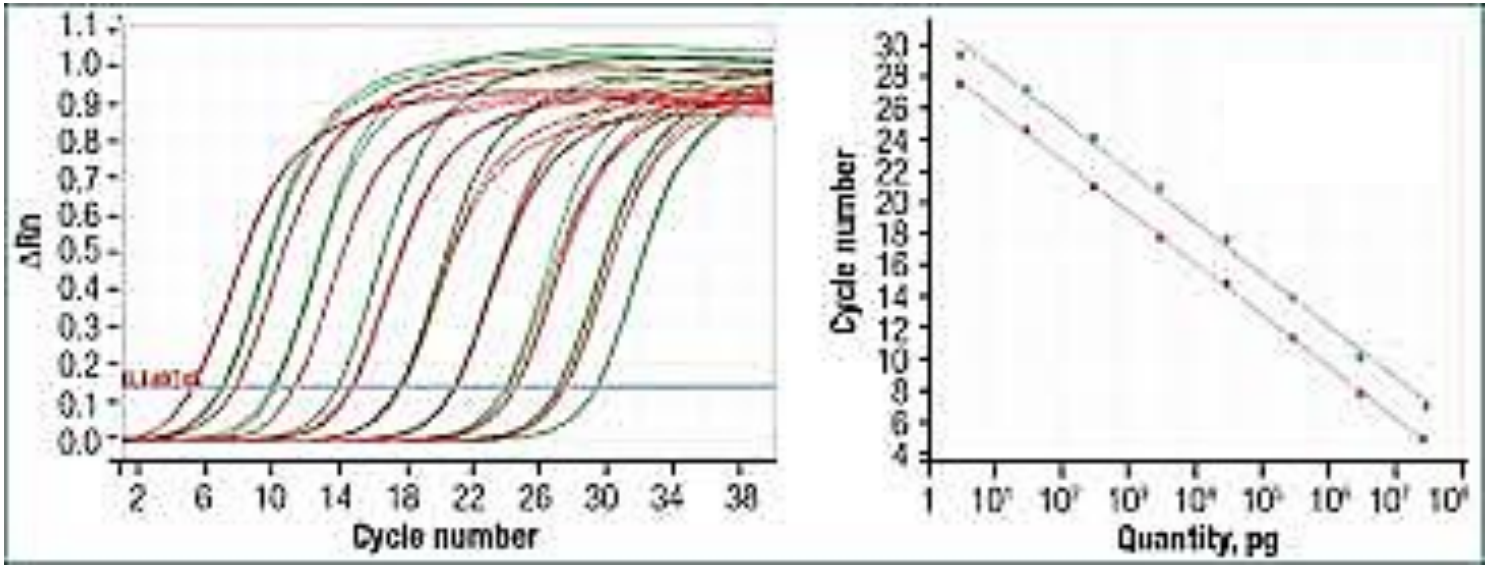




Amplicon sequencing

Step 3 : building libraries for each pool

The last step is to quantify the libraries with qPCR using adapter sequences as primers.



Amplicon sequencing

Step 4 : Miseq sequencing

Preparing the sample (pooled libraries) involves dilution and denaturation in NaOH.

Setup of the sequencer takes around an hour (warming)

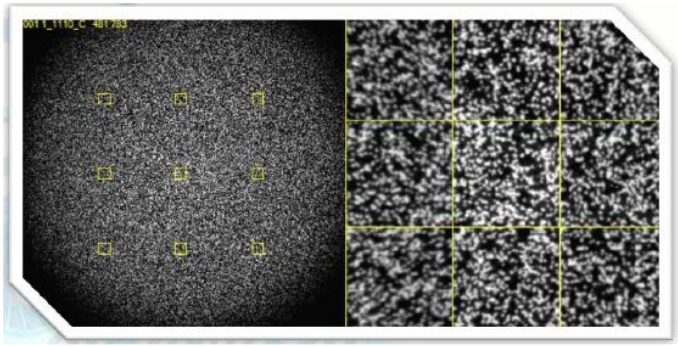
The run is 46 hours long

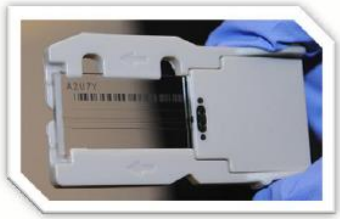
Amplicon sequencing

Step 4 : Miseq sequencing

What can go wrong at the sequencing step?

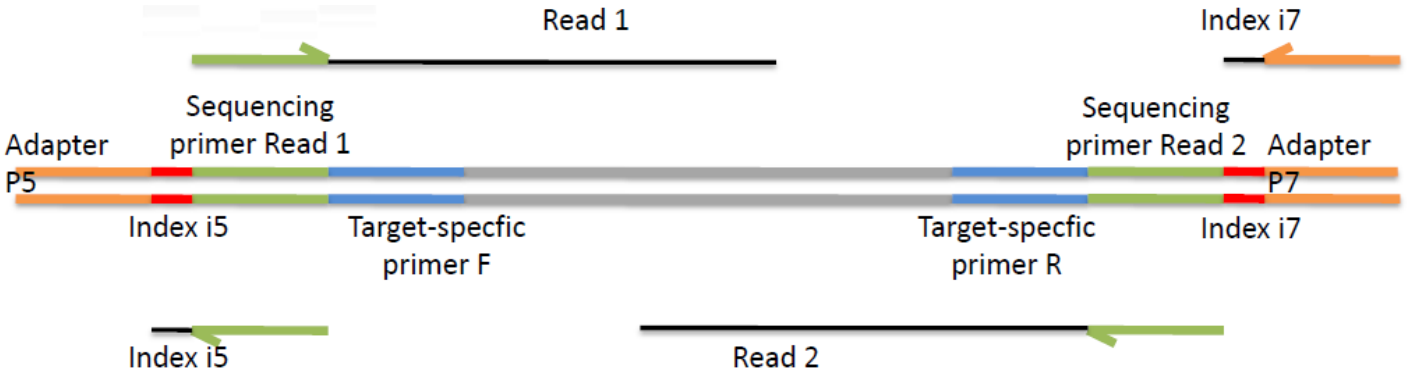
- Overclusterization
- Bad quality
- Door opened..
- Road construction…..





Amplicon sequencing

Two other wet lab strategies 2-steps PCR



Step 1: PCR to amplify region of interest

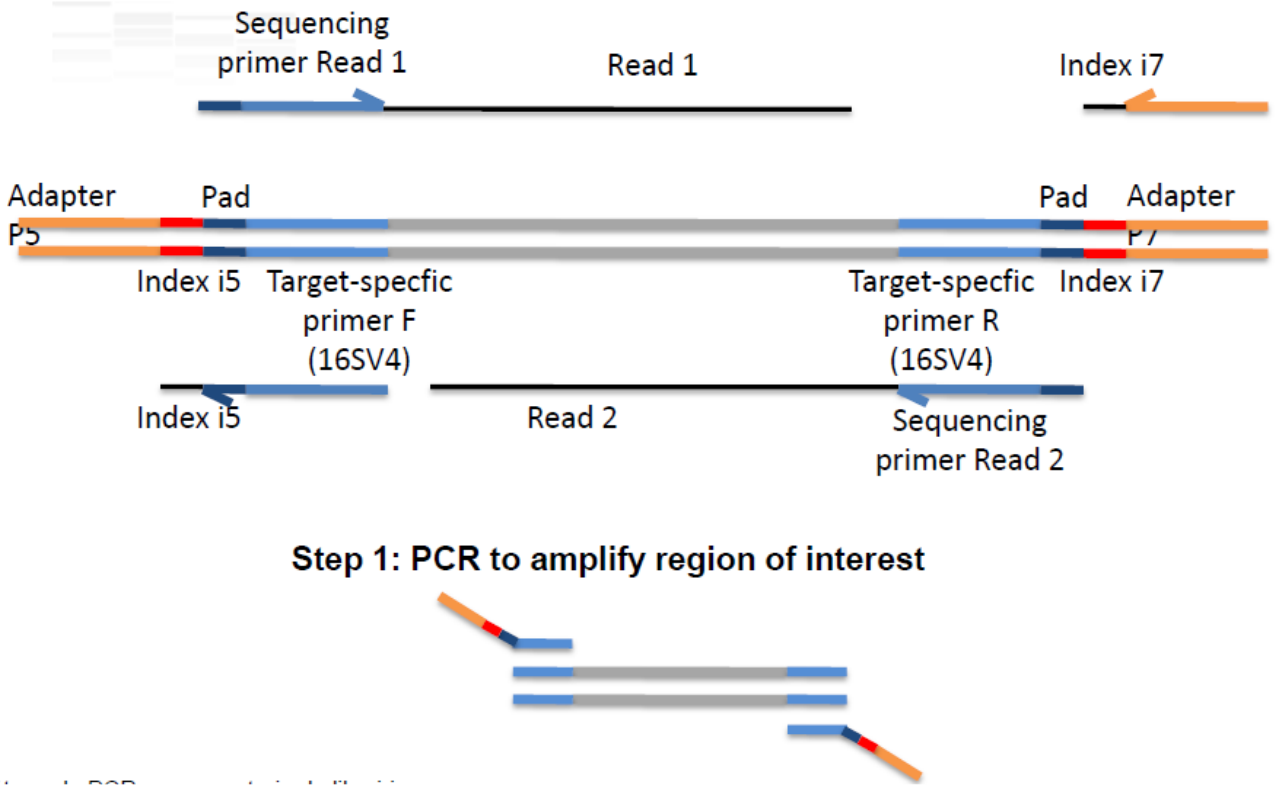
Step 2: PCR to add indeces & adapters (kit Nextera)

Advantages : multiple markers, universal and easy

Drawbacks : 2 steps PCR costs, biases and contamination issue

Amplicon sequencing

Two other wet lab strategies : 1-step PCR



Advantages : 1 step PCR, no primer sequencing

Drawbacks : add custom primers into the sequencer cartridge, only 16S for now

Amplicon sequencing

comparing wet lab strategies

	Protocole 2-step PCR	Protocole 1-step PCR	Protocole tag PCR - Truseq
Few markers and < 1728 samples	+++	---	
Metabarcoding 16S (bacteria)	+	+++	
Multiple projects, > 1728 samples		---	+++

Amplicon sequencing

Amplicon sequencing

Data analysis, from raw data to usable data

General steps:

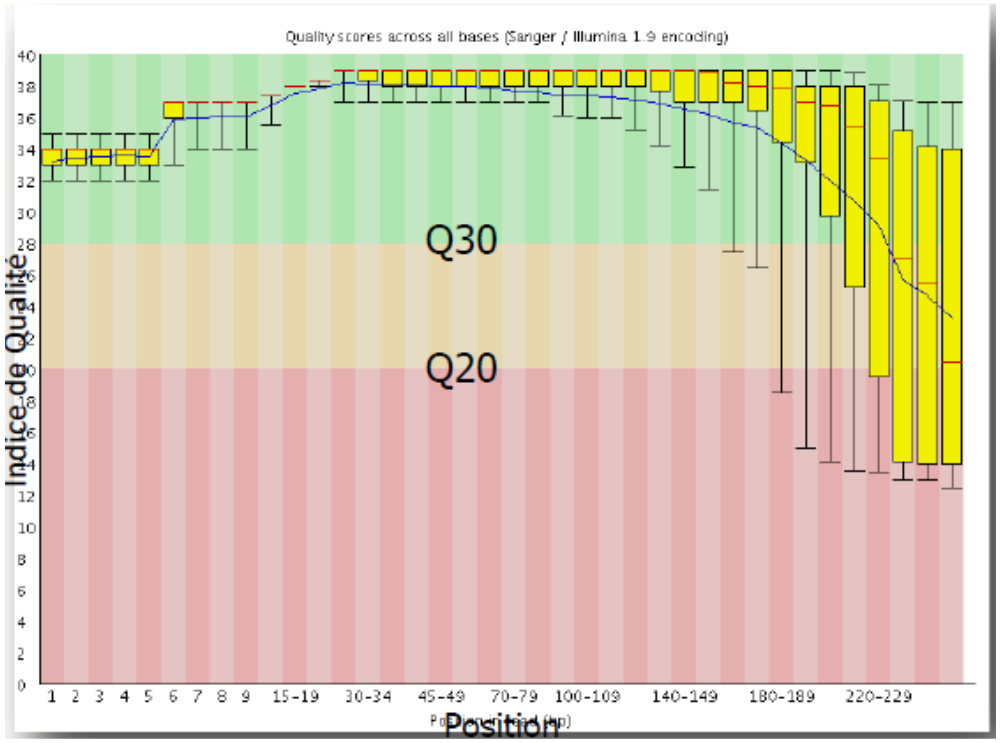
1. quality assessment
2. filtering using quality
3. contig of paired reads

4. demultiplexing the markers if relevant
5. demultiplexing the amplicons
6. characterizing variants
7. Aligning variants if necessary
8. assigning to a reference library

Amplicon sequencing

Data analysis, from raw data to usable data

1. quality assessment (Galaxy)



Amplicon sequencing

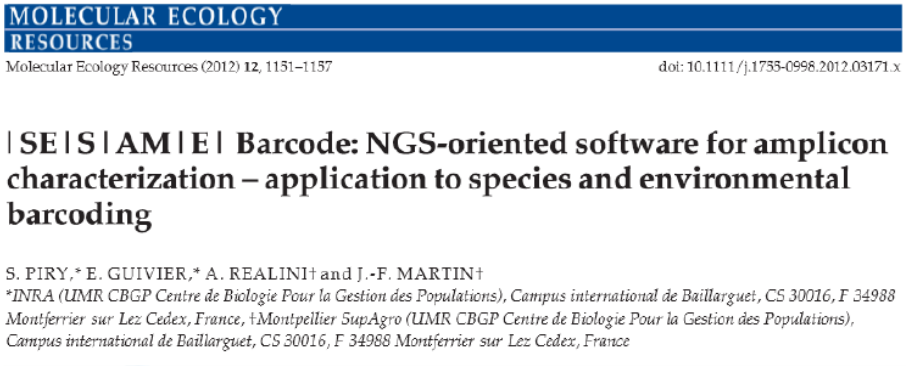
Data analysis, from raw data to usable data

2. Filtering the reads (Galaxy)

3. Contig the paired reads (Mothur) and prepare a fasta file

Amplicon sequencing

From now, everything is done with



MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2012) **12**, 1151–1157 doi: 10.1111/j.1755-0998.2012.03171.x

| SE | S | AM | E | Barcode: NGS-oriented software for amplicon characterization – application to species and environmental barcoding

S. PIRY,* E. GUIVIER,* A. REALINI† and J.-F. MARTIN†
*INRA (UMR CBGP Centre de Biologie Pour la Gestion des Populations), Campus international de Baillarguet, CS 30016, F 34988 Montferrier sur Lez Cedex, France, †Montpellier SupAgro (UMR CBGP Centre de Biologie Pour la Gestion des Populations), Campus international de Baillarguet, CS 30016, F 34988 Montferrier sur Lez Cedex, France

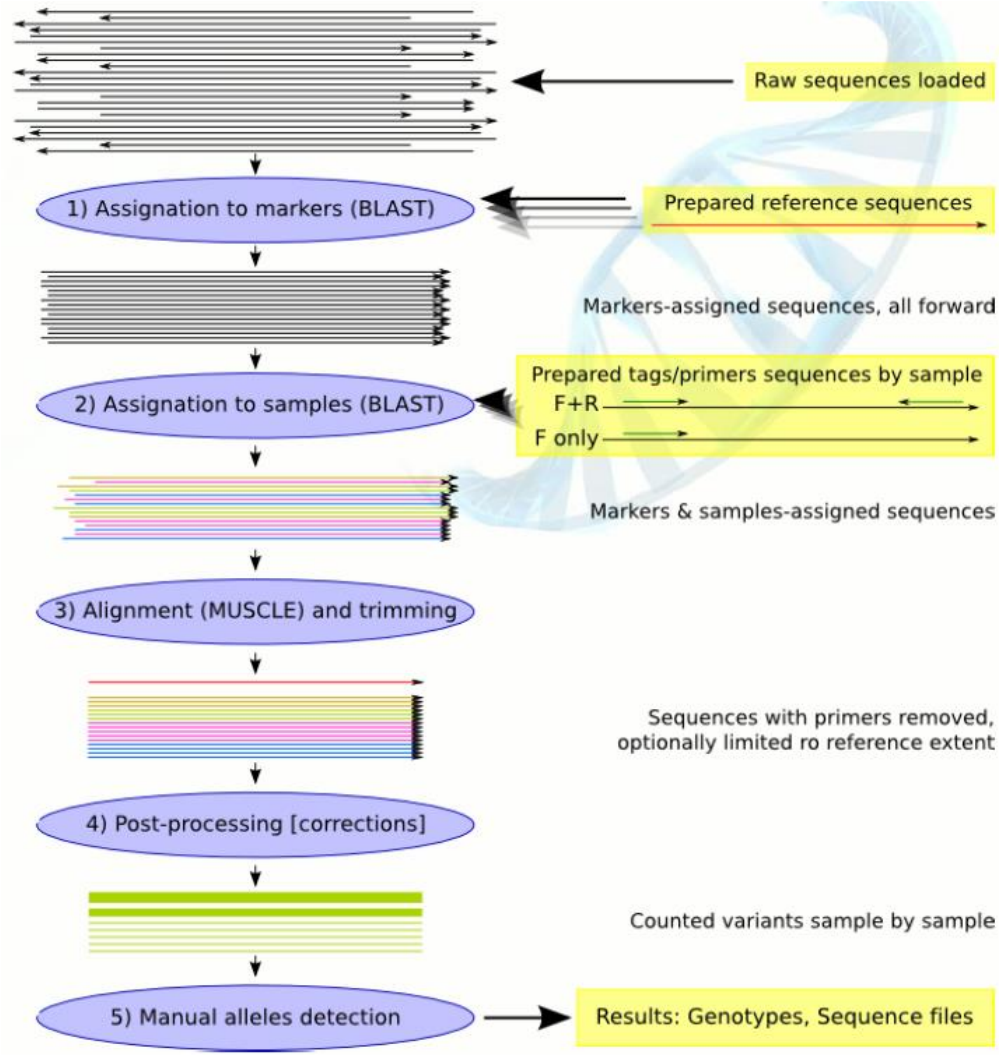
Could also be done with

mothur

Quantitative Insights Into Microbial Ecology
qiime

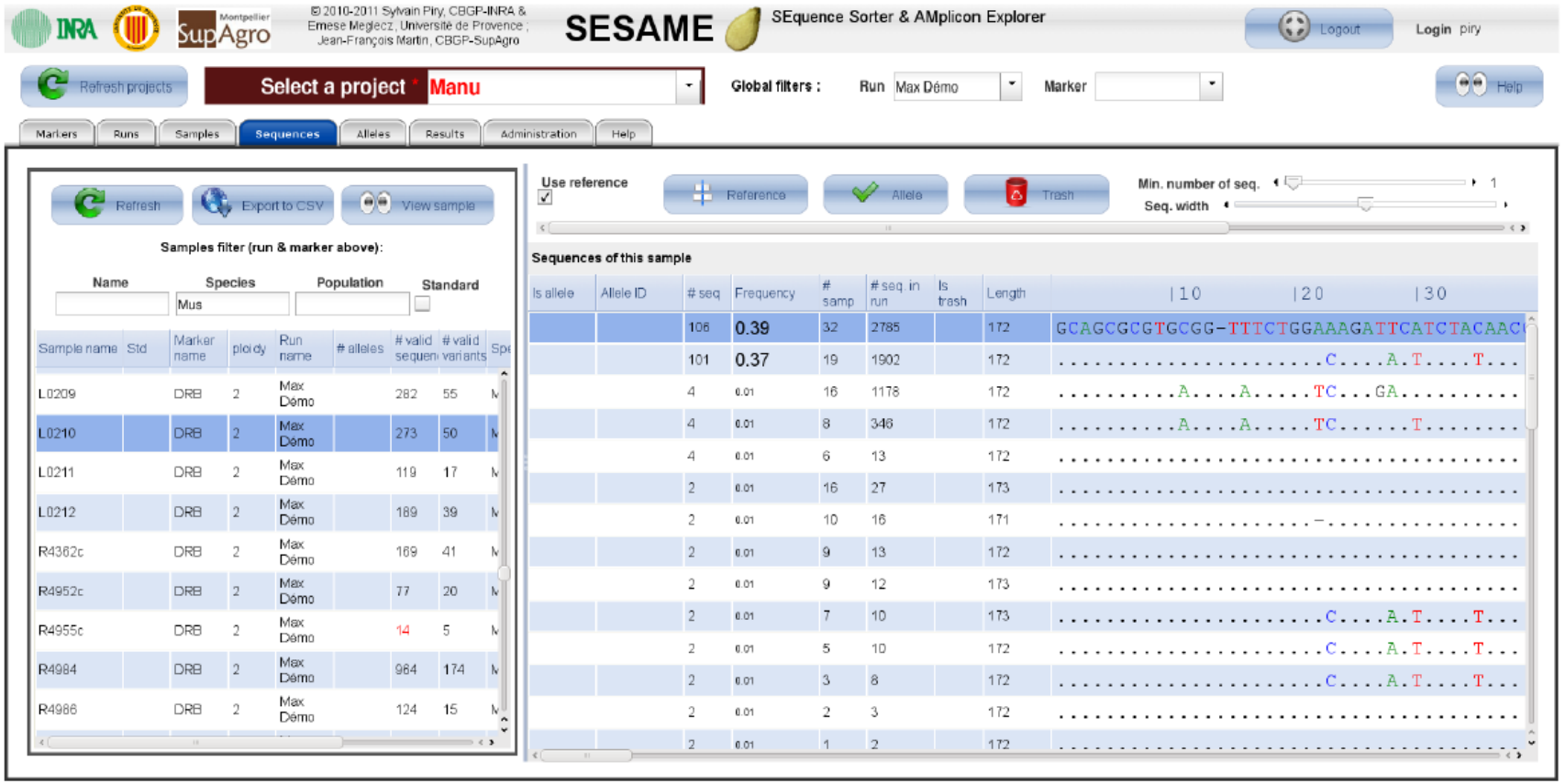
Amplicon sequencing

The pipeline



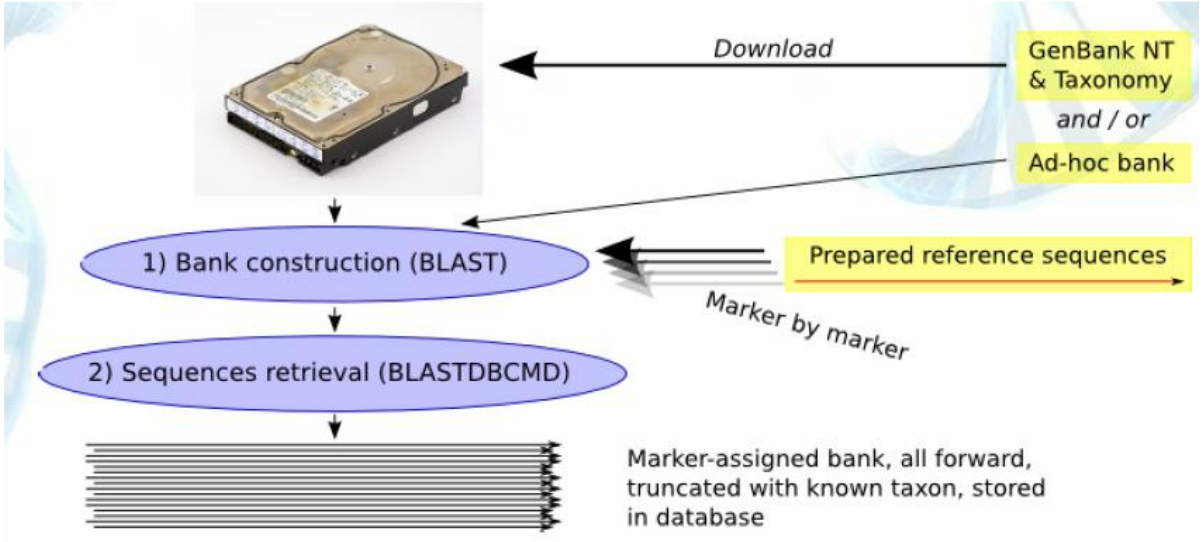
Amplicon sequencing

The sequences alignment interface (if needed)



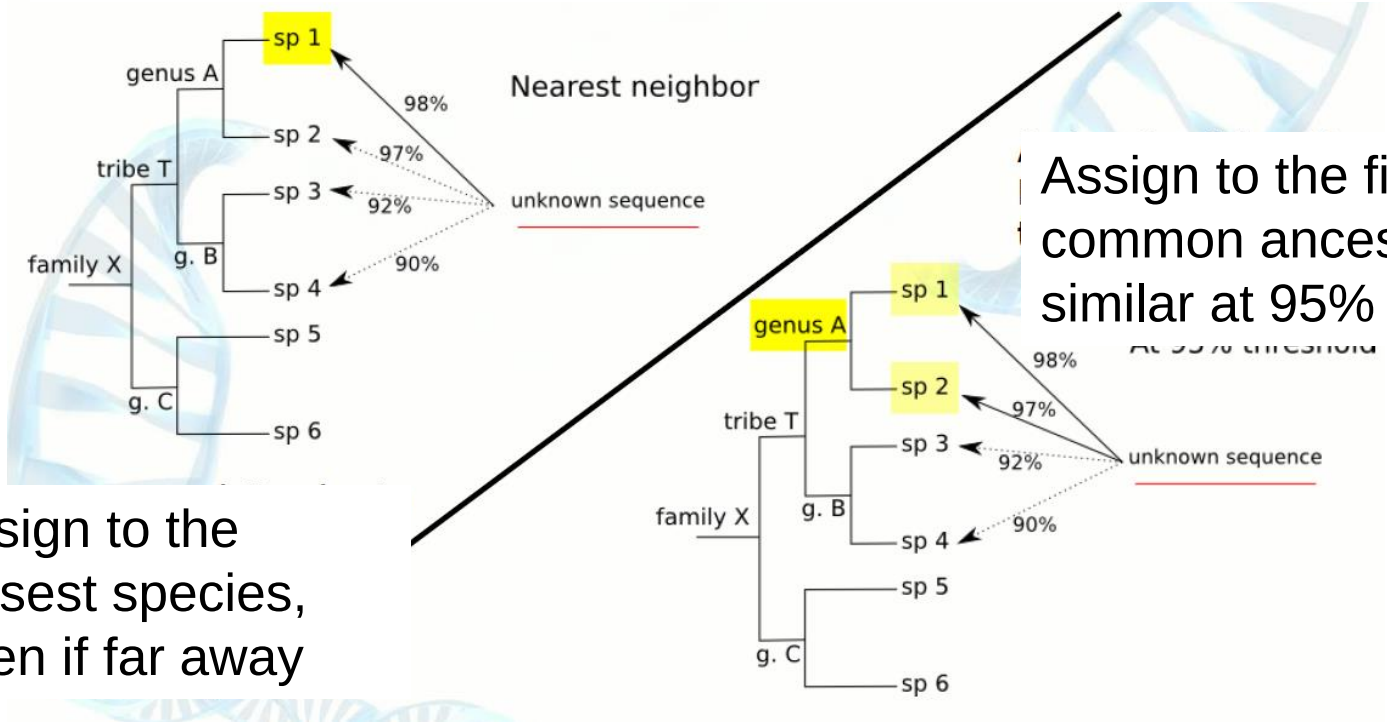
Amplicon sequencing

Preparing reference libray



Amplicon sequencing

Assignment to the library

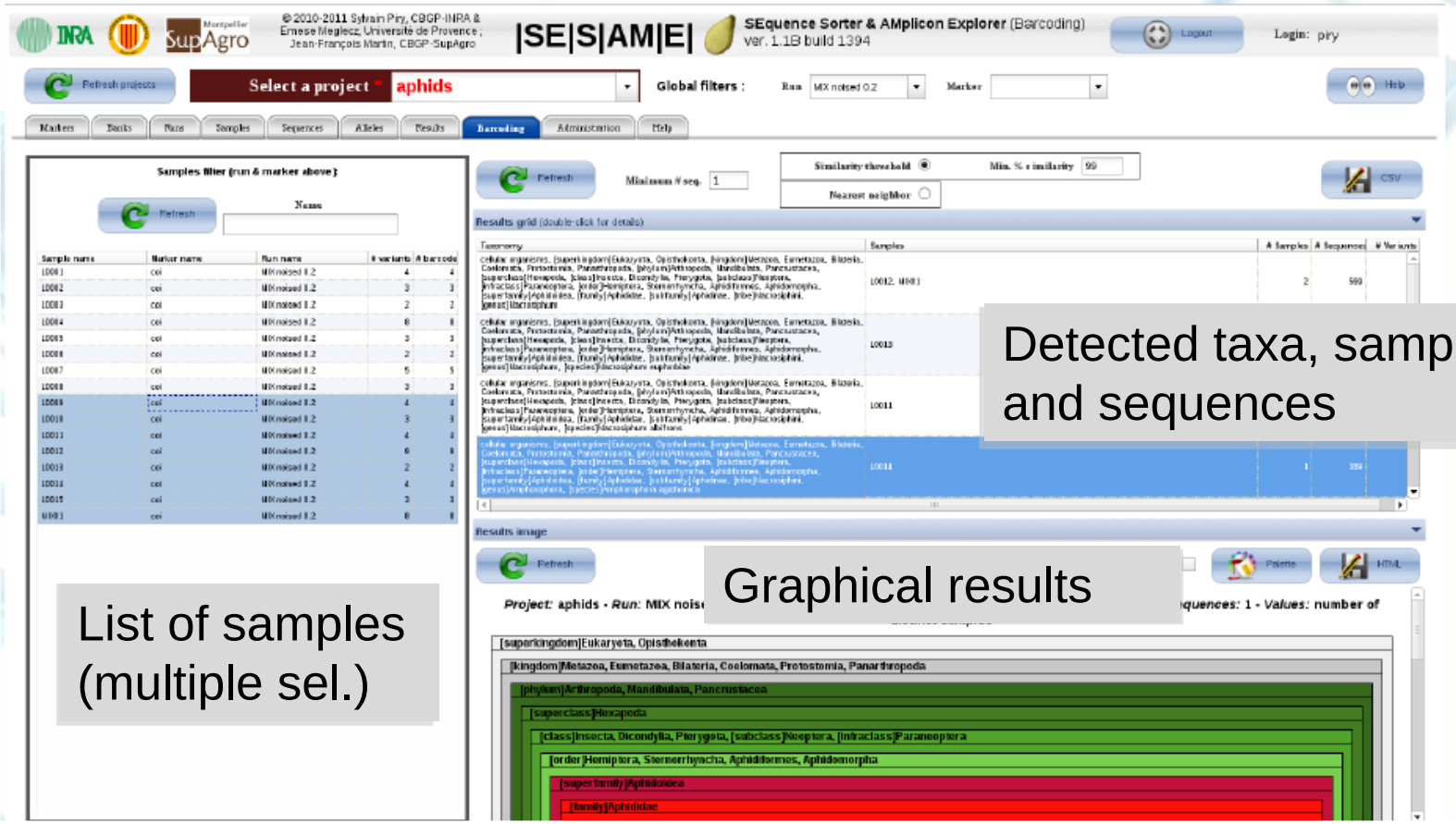


Assign to the first common ancestor if similar at 95% threshold

Assign to the closest species, even if far away

Amplicon sequencing

Exploring results



Concluding remark about future directions

Improving PCR of
alternative approaches

Be aware of the
continuous evolution
of thechnologies

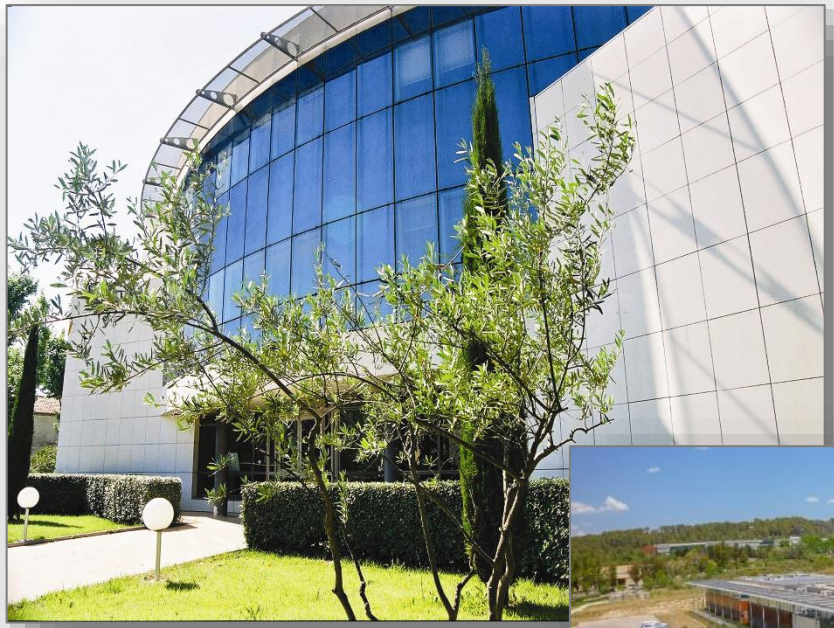
Work on the scalability of
bioinformatics solutions

Acknowledgements

- Morgane Ardisson
- Anne-Laure Clamens
- Armelle Cœur d'Acier
- Emmanuel Corse
- Vincent Dubut
- Philippe Gauthier
- André Gilles
- Emmanuel Guivier
- Emese Meglecz
- Grégory Mollot
- Sylvain Piry
- Audrey Réalini

Montpellier SupAgro

www.supagro.fr



Centre de Biologie pour la Gestion des Populations

Centre international d'études supérieures en sciences agronomiques