?

NGG

RE
GBS

Plants
Next Generation Genotyping
What else ?

RAD
Seq

Rapid
Seq

Marie-Christine Le Paslier, INRA_EPGV

Analyse des variations génétiques / NGS

GBS NGG

- GBS & NGG
  - GBS = Genotyping By Sequencing
  - NGG = Next Generation Genotyping = sequence-based genotyping
- GBS = pas nouveau !
  - Génotypage par séquençage d'amplicons/ *Sanger*

GBS dans le contexte NGS

Approches déjà utilisées en génotypage par NGS

- non ciblé ou ciblé
- sur génome total
  - Re-sequençage whole genome à faible profondeur
    - envisageable pour petit génome
    - coûts librairies trop élevés pour haut débit d'échantillons
- ou sur génome réduit
  - RNAseq
  - RRL : reduced representation libraries, basées sur le polymorphisme de restriction
  - Amplicons PCR et LR-PCR
  - Capture par hybridation de régions d'intérêt

GBS

Discovery

Genotypage du connu

Genotypage par hybridation

- Besoins en génotypage
  - Panels de N individus…à répétition. N=?
  - Panels de N marqueurs …N = ? 500 / 3000 -6000?
  - **à coûts réduits**
  - **avec flexibilité**
- Bénéfices / génotypage par hybridation

  Infinium, Axiom, GoldenGate, TaqMan, KASPar…
  - Coûts réduits
  - Biais réduits
  - Augmenter la gamme de l'échelle de détection
  - **Available even with no a priori genomic information**

Quelles solutions actuelles et à venir?

- sequençage whole genome faible profondeur

- sequençage sur génome réduit
  - RE-GBS *Cornell*
  - RAD-Seq *Floragenex*
  - RAPID seq *RAPiD Genomics*

  - *Eureka Genomics*
  - OS-Seq

home-made
vs
service

- ce qui est visé ?
  - une haute couverture de séquençage
  - d'une fraction réduite et similaire du génome
  - pour une grande quantité d'échantillons

RADseq
Floragenex

- RAD = Restriction site Associated DNA

Application of RADseq to *Cedrus atlantica*

Juin 2014

Marie-Joe Karam
François Lefèvre

Ecologie des Forêts Méditerranéennes URFM
Avignon - France

Genomic Resources

Forest trees: growing genomic resources

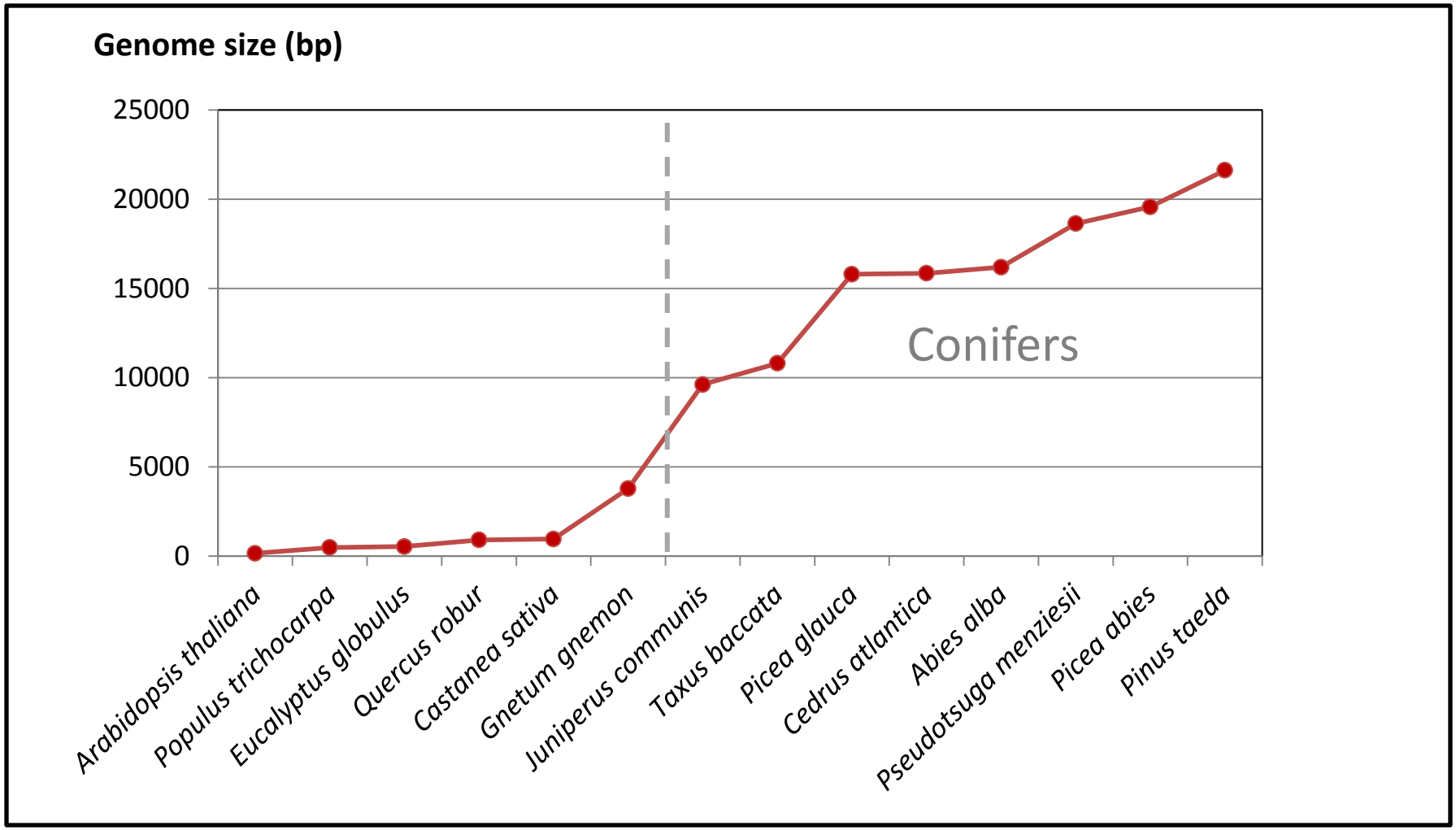
(Neale & Kremer, 2011)

Conifer trees → genomic studies slow

?

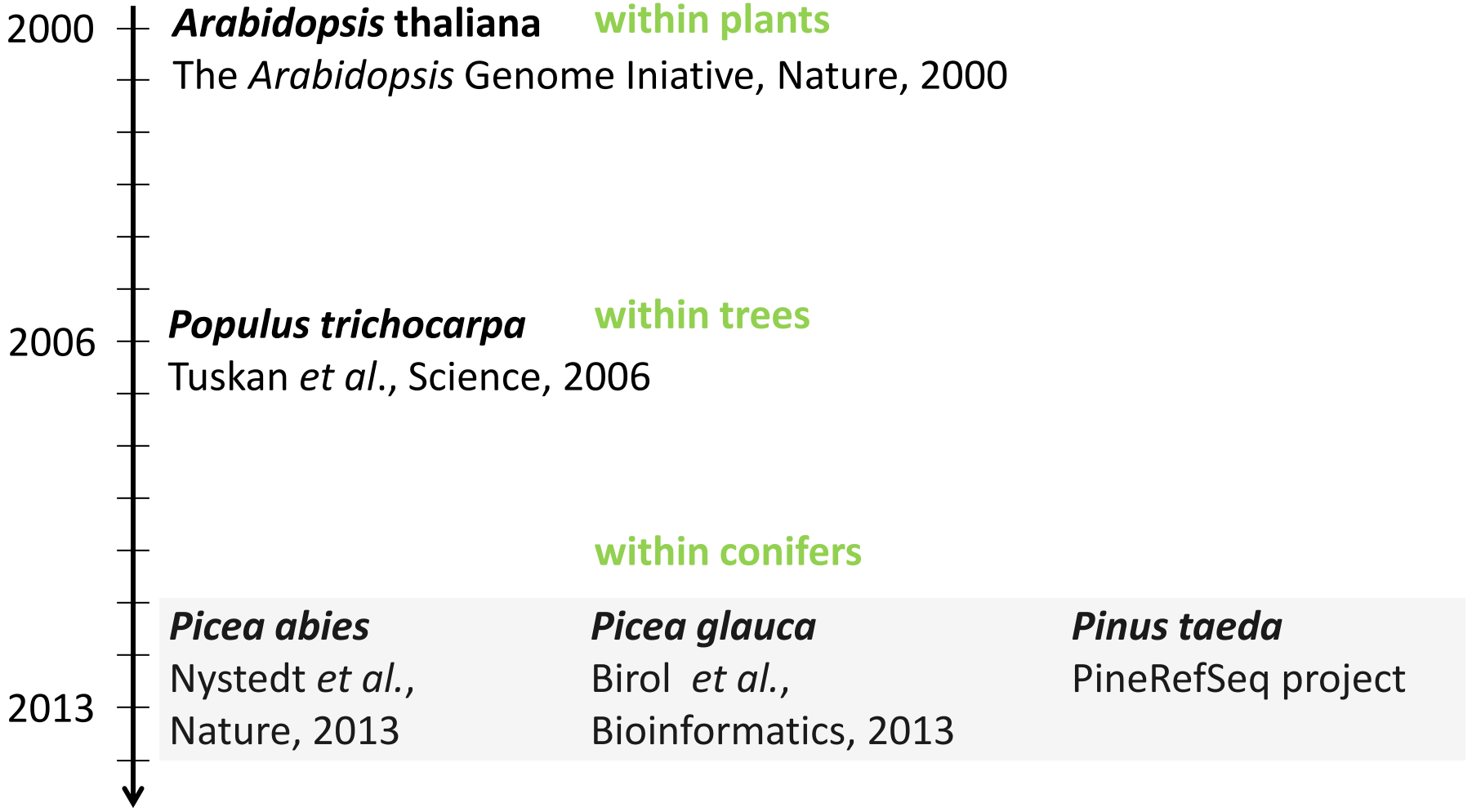
1. Large and complex genomes
2. High levels of heterozygosity
3. Lack of inbred lines

Conifers have large genome size compared to most animals and plants



Full genome sequencing:
feasible in “model” conifers thanks to the development of NGS technologies

First full genomic sequences:



Full genome sequencing:
feasible in “model” conifers thanks to the development of NGS technologies

What about the non-model conifers?

Genome complexity reduction methods * NGS
→ interesting starting point

1- mRNA extraction * NGS

2- Filtering the hypomethylated fraction of the genome * NGS

3- Isolating low-copy fraction of the genome * NGS

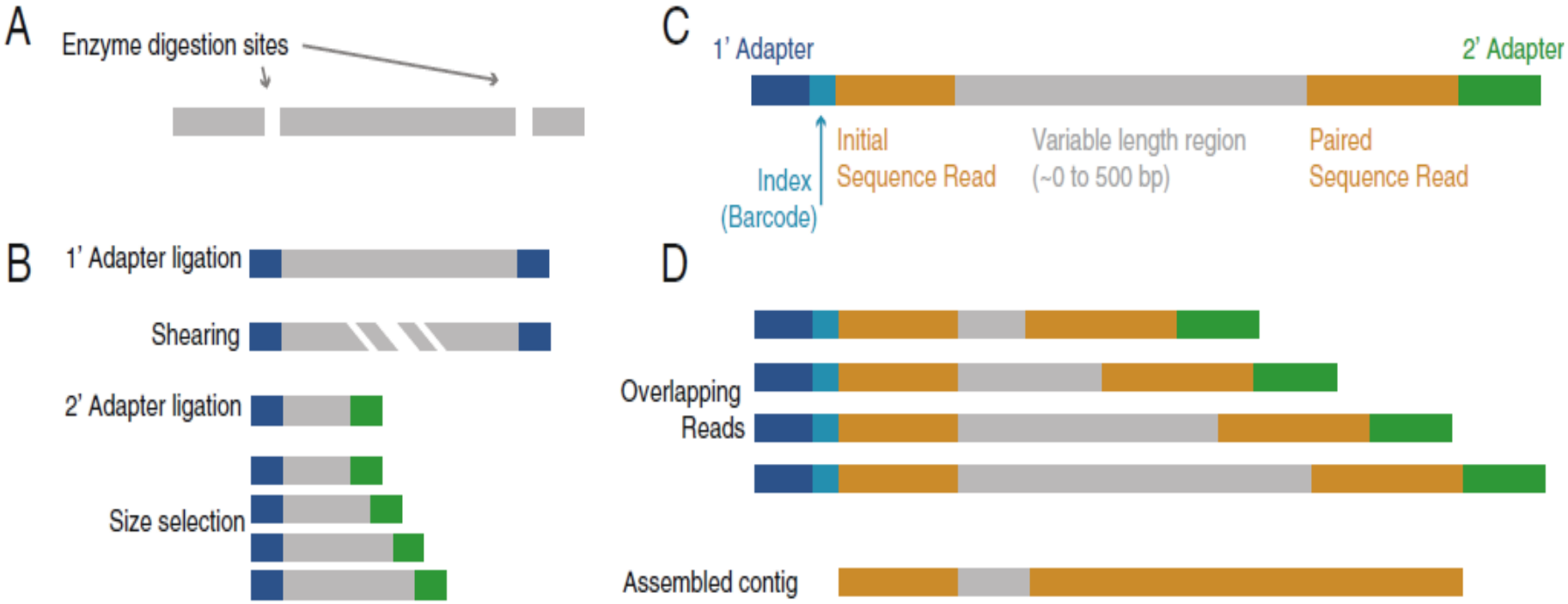
4- Exon-capturing * NGS

Objective

First genome exploration and development of genomic resources in a non-mainstream conifer species using:

Restriction Site Associated DNA sequencing (RADseq)

RADseq



Pegadaraju et al. 2013

RADseq

FLORAGENEX

Applied on:

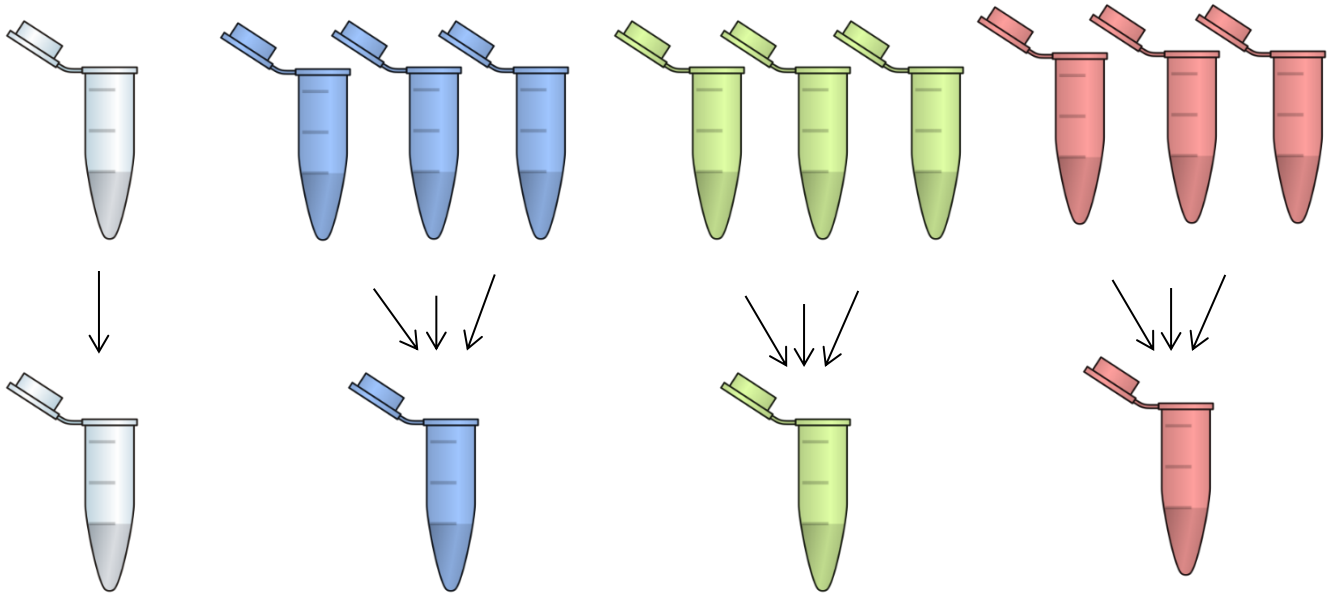
-insects (Miller *et al.*, 2007)
-fungi (Lewis *et al.*, 2007)
-fishes (Miller *et al.*, 2007a)
-Poaceae (Nipper *et al.*, 2009)
-*Solanaceae* (Barchi *et al.*, 2011)
-mammals (Peterson *et al.*, 2012)
-*Asteraceae* (Scaglione *et al.*, 2012)
-*Vitaceae* (Wang *et al.*, 2012)
-nematodes (Davey *et al.*, 2013)
-snails (Richards *et al.*, 2013)
-*Salicaceae* (Stölting *et al.*, 2013)
-*Betulaceae* (Wang *et al.*, 2013)
-Cucurbitaceae (Xu *et al.*, 2013)
-Brassicaceae (Vandepitte *et al.*, 2013)

This study is one of the first application of RADseq on a coniferous tree

Experimental designs

10 diploid individuals (1 reference + 3 individuals * 3 pools) → RADseq (Illumina sequencing)

FLORAGENEX



Main results

-Validation of capacity of RADseq to produce a genomic snapshot enriched by genes in the *C. atlantica* genome (17% of the RADseq contigs had significant hits in NR *vs.* 34% obtained with a *C. atlantica* transcriptome)

- Validation of the capacity of RADseq to reduce the proportion of transposable elements in the *C. atlantica* genome (< 4% of the RADseq contigs length)

-Development of 17,348 SNPs from which 384 were validated with Fluidigm SNP genotyping platform (conversion rate of ~58%).

RAPiD-Seq
RAPiD Genomics

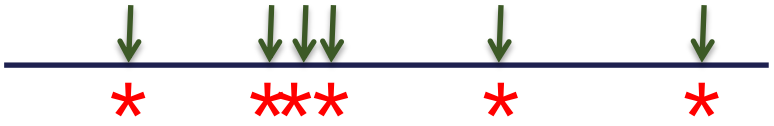
Two platforms - different uses

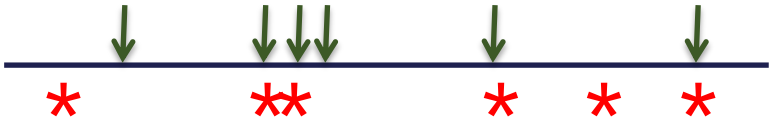
Capture-Seq

- “Capture”-based
- Targeted loci
- Service

RAPiD-Seq

- PCR-based
- “Random” loci
- Service & Kit





Matias Kirst

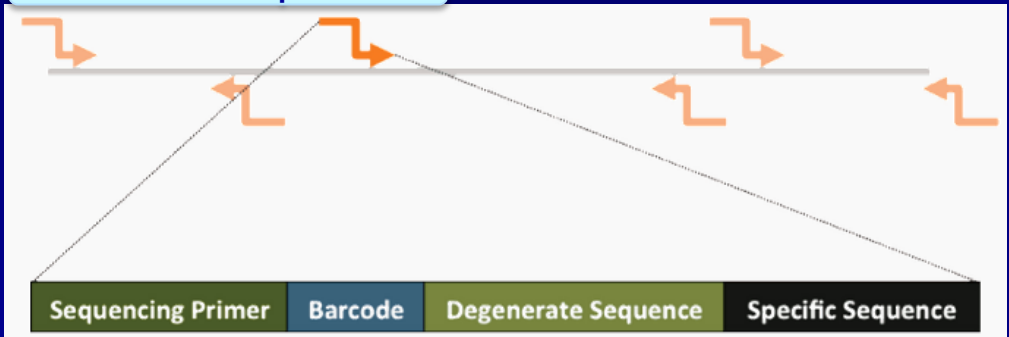
RAPiD GENOMICS

RAPiD-Seq

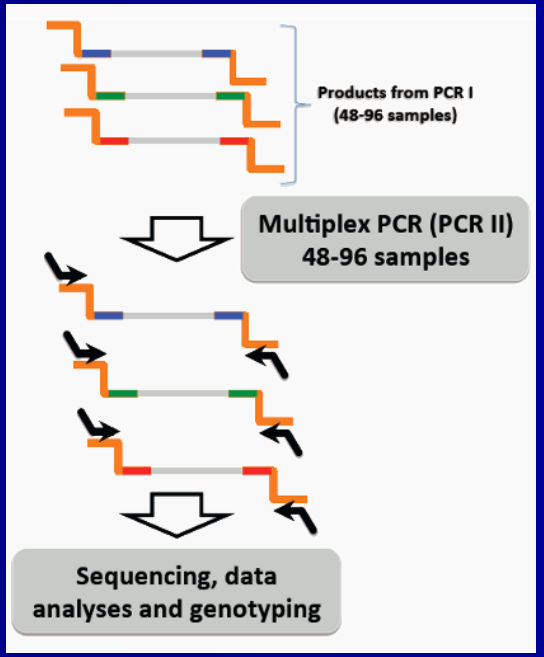
Randomly **A**mplified **P**olymorphic **D**NA **seq**uencing

- 2 PCR reactions
- to generate a reduced genome representation
- to be sequenced in any next-generation sequencing (NGS) platform and genotyped

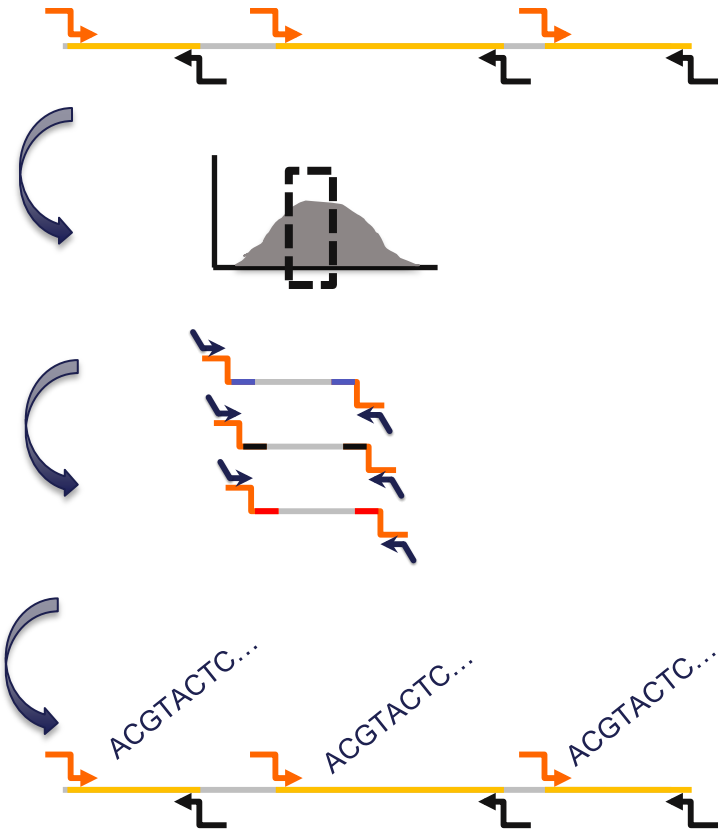
PCR I simplex



- **designed primers** →amplify the desired loci on the genome
- **barcode sequence** →identify the individuals
- **degenerate sequence** → gives stability to the primer
- **specific sequence** → gives the reaction specificity



RAPiD-Seq overview



Individual PCR (PCR I) — 1.5 h
- Reduce genome complexity
- Incorporates barcode

Purification / size selection — 0.5 h
- Reduce genome complexity

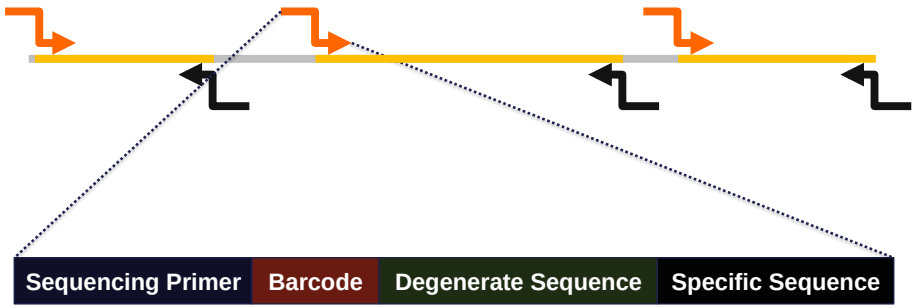
Multiplex PCR (PCR II) — 1.5 h
- 48+ individuals combined
- Sequencing compatible
- Purification / size selection 2

High-throughput Sequencing

RAPiD GENOMICS

Fast ⇔ Flexible ⇔ Accurate/Reproducible ⇔ Cost effective

RAPiD-Seq overview



In sillico primer selection

Primer A	Primer B
AAAAAA	AAAAAC
AAAAAA	AAAAAG
AAAAAA	AAAAAT
	(…)
TTTTTG	TTTTTT

Tested all combinations of 6, 7 and 8 mer primers

- Number of fragments generated (thousands to millions)
- Avoid repetitive region
- Even marker distribution
- Enrich for regions of interest

RAPiD GENOMICS

Fast ⇔ **Flexible** ⇔ Accurate/Reproducible ⇔ Cost effective

Why RAPiD-Seq?

RAPiD-Seq	RE-GBS
Flexibility in the number and position of regions amplified in the genome (> 16M primer combinations tested in sililco)	Limited by existing restriction enzymes and possible combinations among them

RAPiD GENOMICS

Why RAPiD-Seq?

RAPiD-Seq

- Very low amount of DNA (< 50ng)
- Very low cost
- Ultra high-throughput
- High data usage

RE-GBS

Higher amounts of DNA required

RAPiD GENOMICS

Kits Avaibility

RAPiD-Seq

- Maize Kit / Illumina sequencing
- Sugarcane, beans, rice in progress

C'est à vous!

- Vos Informations : autres méthodologies
- Expériences…ratées et réussies
  - les vôtres
  - les autres connues
- Vos besoins
  - N Marqueurs
  - N individus + périodicité

Hi Marie,
At this time, we currently only offer maize in our Rapid-Seq kit. Moreover, with regards to the number of SNPs—the maize kit is offered a way, whereas, you can choose your desired number of SNPs based on your sequencing multiplex (see chart below). And, the price point of the kit varies depending upon the annual volume of kits being ordered and whether or not you want to include sequencing and data analysis. With that being said, it can be anywhere from $10 per sample up to $35 per sample. And, our missing data percentages on the chart below reflect no data imputation.

We have the following other species in development: sugarcane, beans, and rice.

John McGuire

Simulated multiplexing	Subset percentage	# SNPs*	Missing data (%)	Average Depth (x)
24	1	68,352	0.14	26
48	0.5	40,279	0.18	29
96	0.25	21,630	0.17	27
144	0.17	14,853	0.16	26
192	0.125	10,540	0.16	25
384	0.0625	4,808	0.15	24

Low Density Marker Assays (LDMA) for High Throughput, Highly Multiplexed Detection, CNV, Methylation & Expression Assays

John D. Curry[1]*, Paul Dier[1], Maria Shin[1], Jessica Nguyen[1], Nadeem Bulsara[3], R. Mark Thallman[2], Viacheslav Y. Fofanov[3], Heather Koshinsky[1] @ EG[1,3], USDA[2]