

# GBS-RAD-sequencing:

Genómica de poblaciones en organismos no-modelo

Pablo Saenz Agudelo

# Cómo caracterizar la diversidad genómica en organismos no modelo?



# Cómo caracterizar la diversidad genómica en organismos no modelo?



Secuenciar  
genoma  
referencia



US\$muchos miles!

# Cómo caracterizar la diversidad genómica en organismos no modelo?



Secuenciar  
genoma  
referencia



US\$ muchos miles!



x 100

Re-secuencia



US\$100-500  
por  
muestra

# Que es RAD-sequencing?



RAD-seq es un método para secuenciar una pequeña fracción del genoma de varios individuos al mismo tiempo usando NGS.

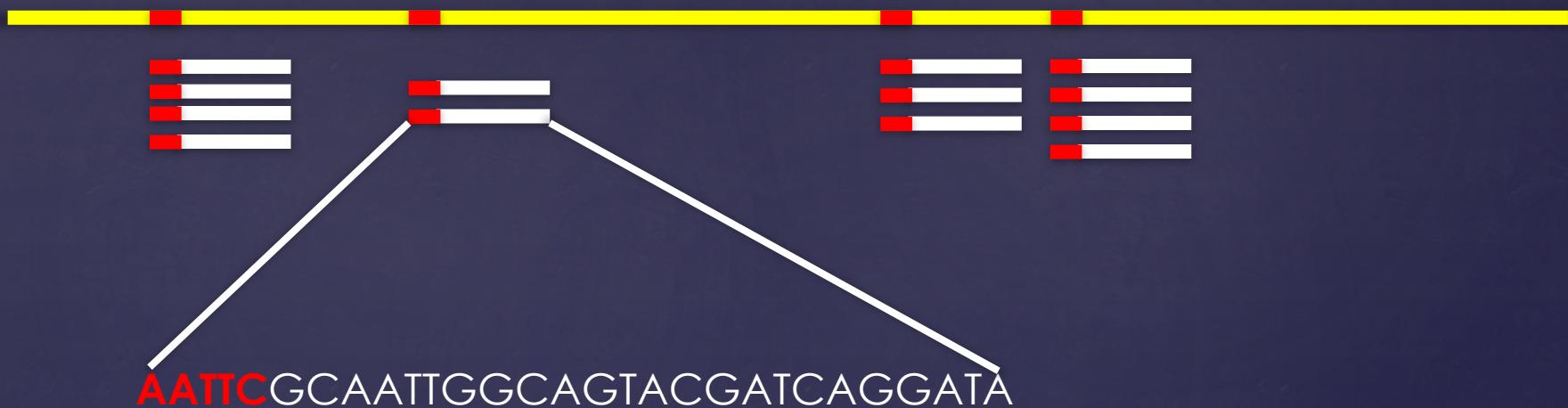
# Como funciona?

Segmento de un genoma



# Como funciona?

Segmento de un genoma



# Variaciones

- RAD-seq (Baird *et al.* 2008)
- ddRAD-seq (Peterson *et al.* 2012)
- 2bRAD-seq (Wang *et al* 2012)
- ezRADs (Toonen *et al.* 2014)

# RAD-seq



Digestion con una enzima de restricción



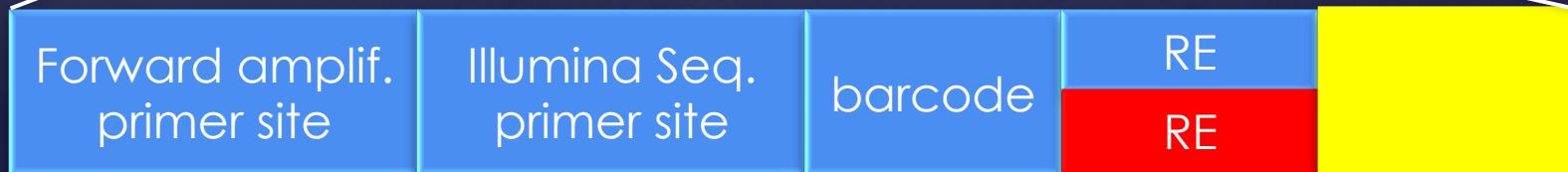
# RAD-seq



Digestion con una enzima de restricción

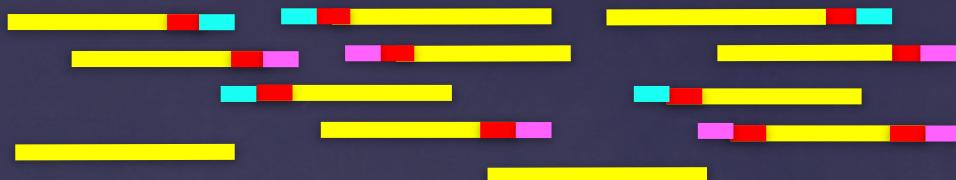


Adición de adaptadores especiales “P1”



# RAD-seq

Se combinan las muestras, se rompen los fragmentos aleatoriamente “shear”, se selecciona por tamaño.

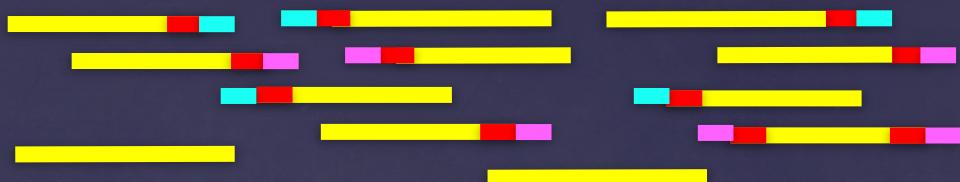


Adición de un segundo adaptador “P2”



# RAD-seq

Se combinan las muestras, se rompen los fragmentos aleatoriamente “shear”, se selecciona por tamaño.

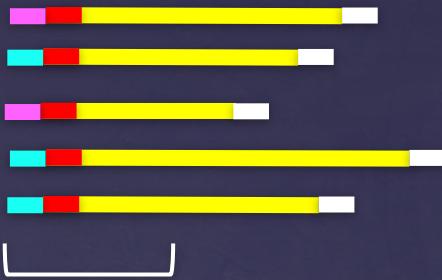


## Adición de un segundo adaptador “P2”



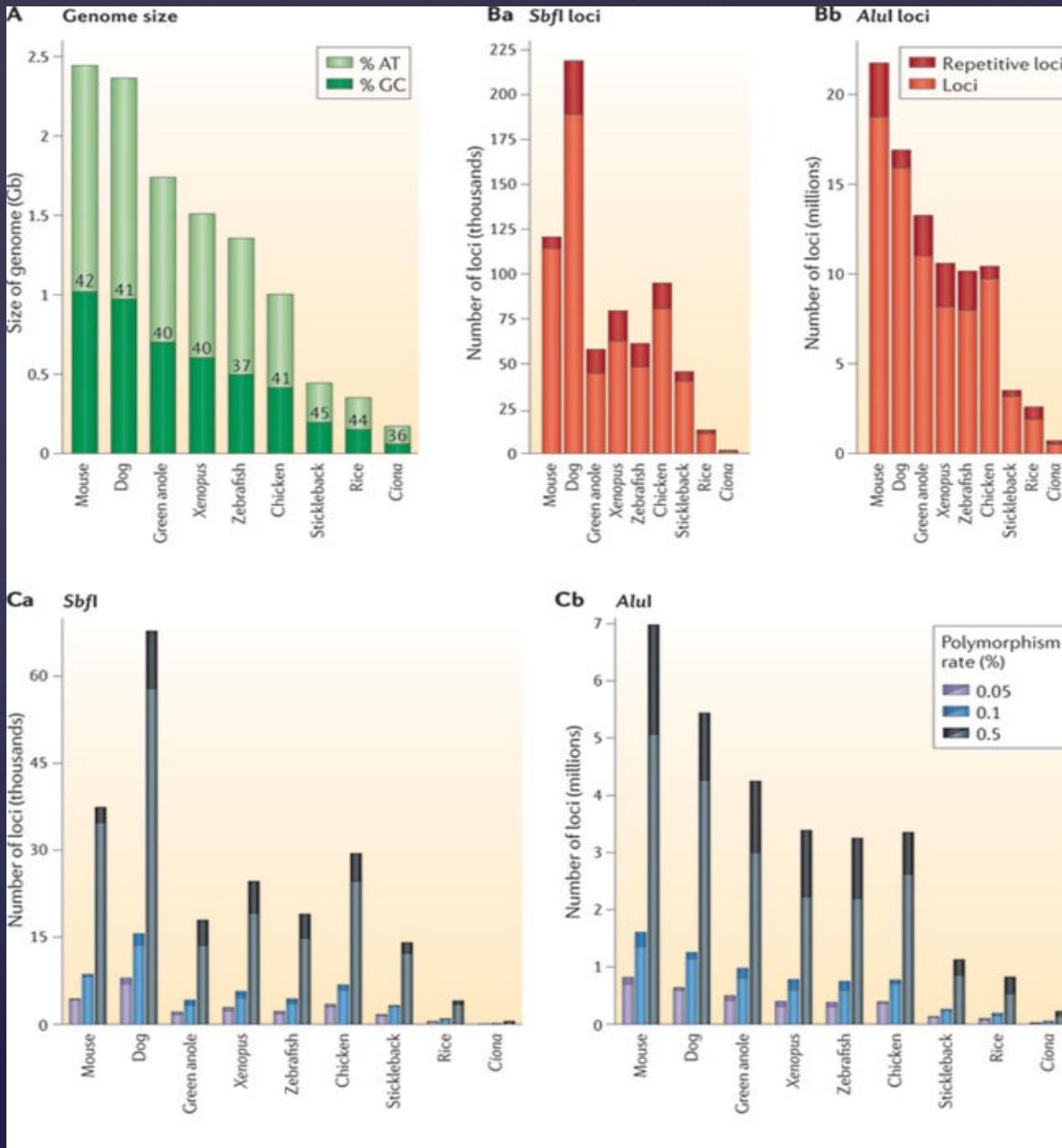
# RAD-seq

Amplificación selectiva por PCR de fragmentos RAD



Longitud de secuencias en Illumina

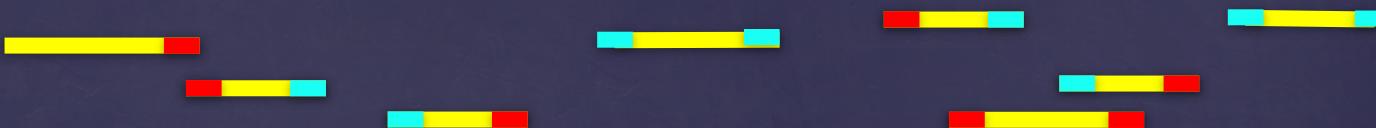
# Efecto de la enzima de restricción



# dd-RAD-seq



Digestion con dos enzimas de restricción



# dd-RAD-seq



Digestion con dos enzimas de restricción



Adición de adaptadores especiales “P1 y P2”



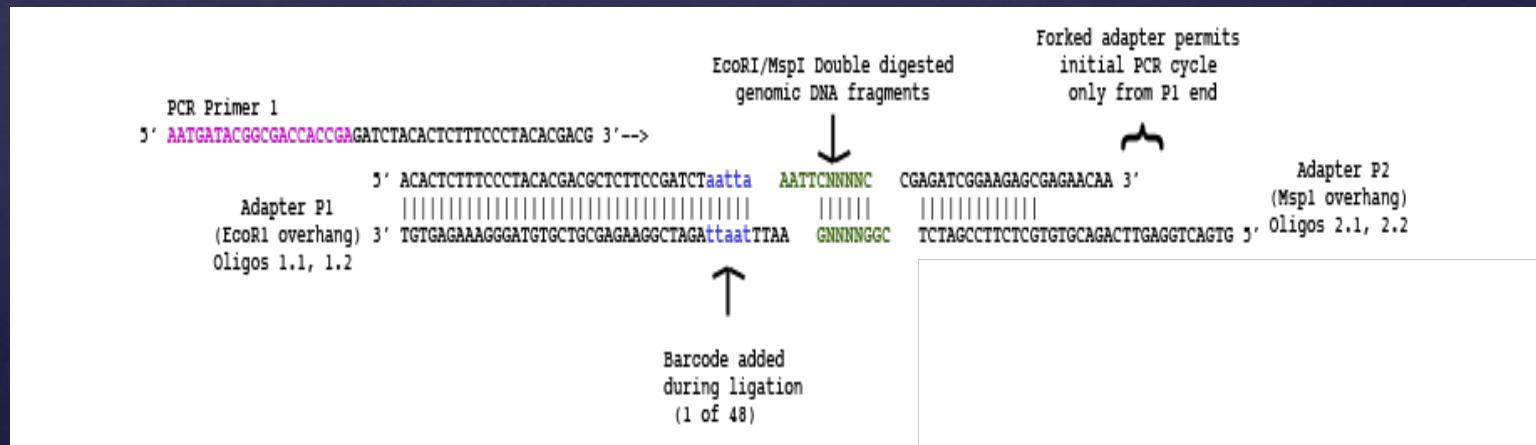
# dd-RAD-seq



Digestion con dos enzimas de restricción

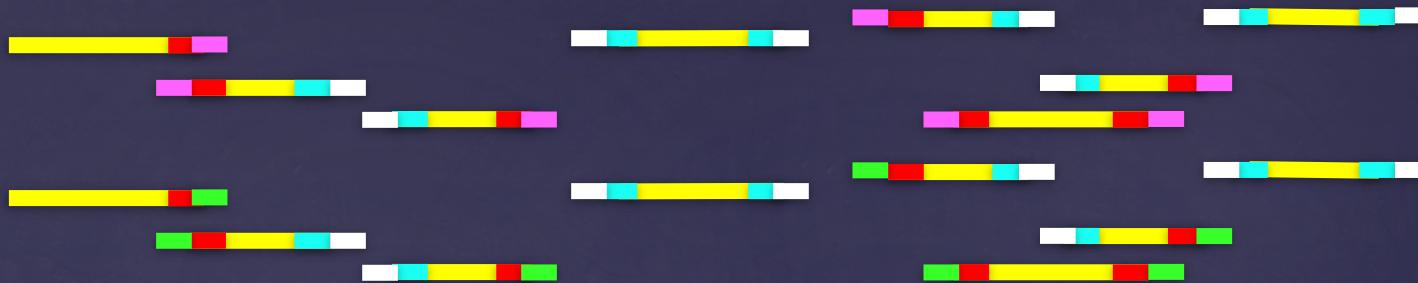


Adición de adaptadores especiales “P1 y P2”

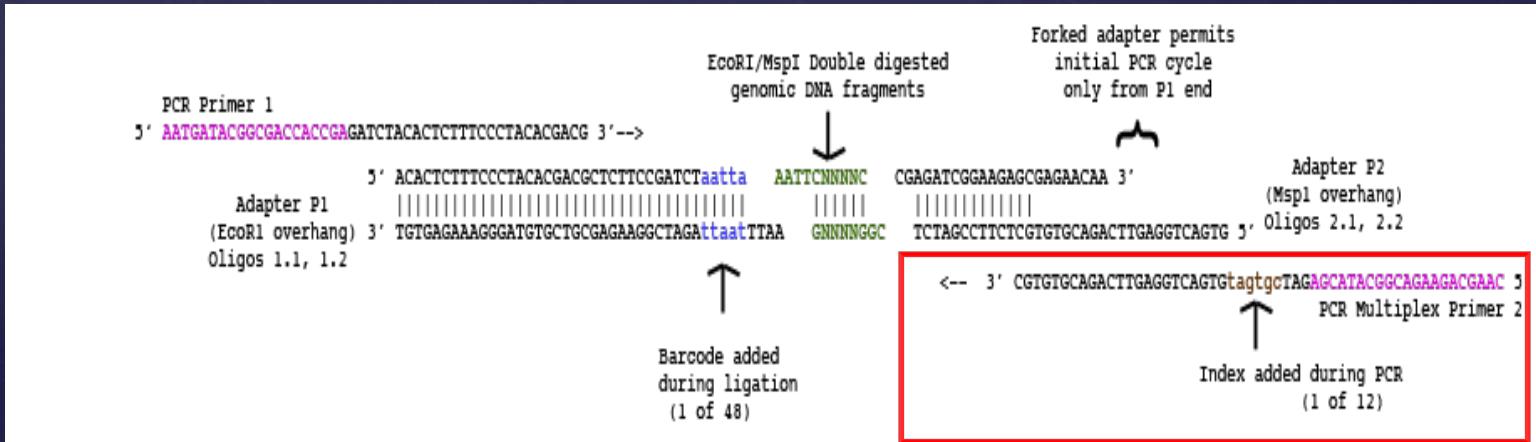


# dd-RAD-seq

Se combinan las muestras con diferentes P1, se selecciona por tamaño.

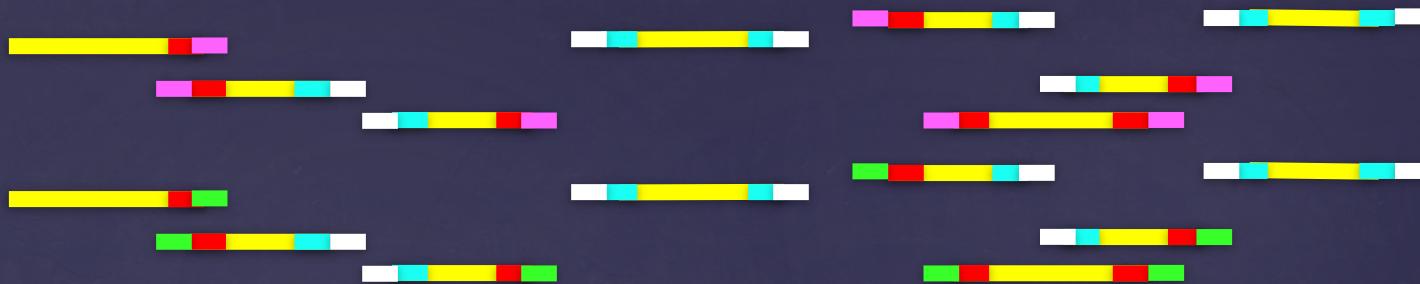


Amplificación por PCR con uno de los primers que incluye otro código de barras



# dd-RAD-seq

Se combinan las muestras con diferentes P1, se selecciona por tamaño.



Combinación de pares de P1 y P2 → multiplicación del número posible de muestras a combinar.

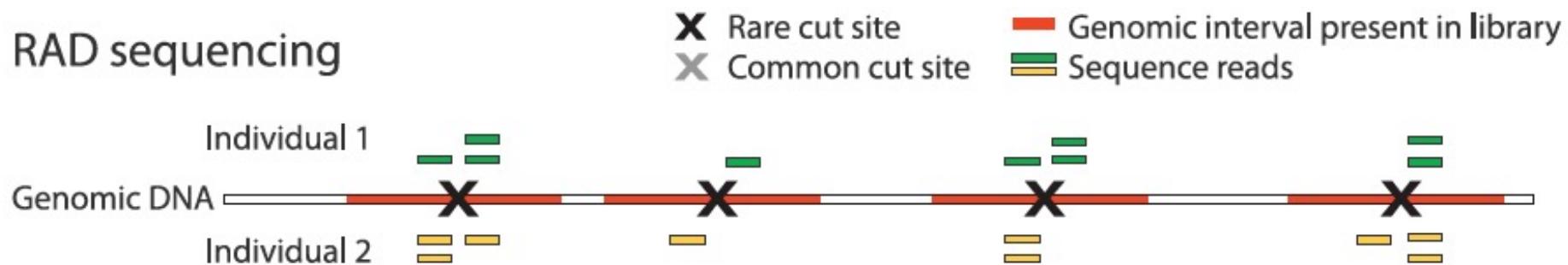
## Final sequencing library



# RADs vs ddRADs

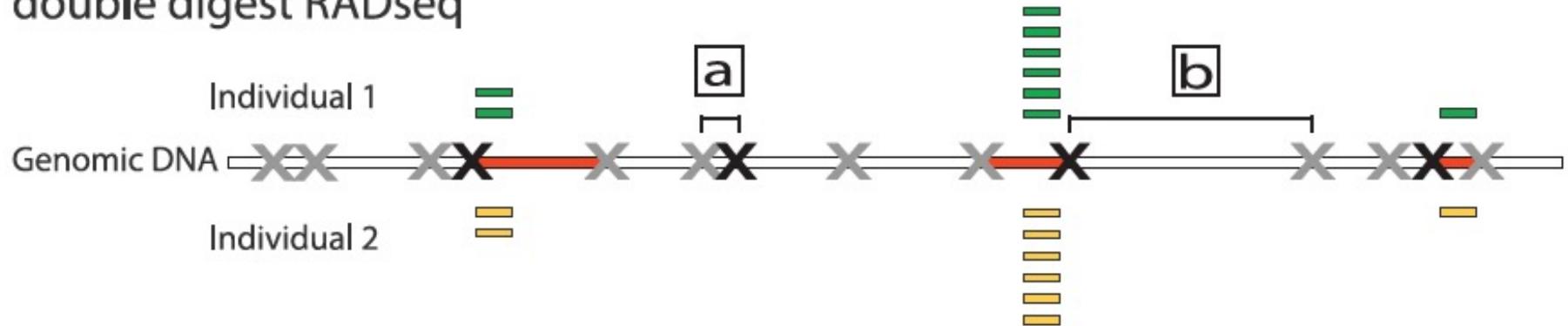
A

RAD sequencing

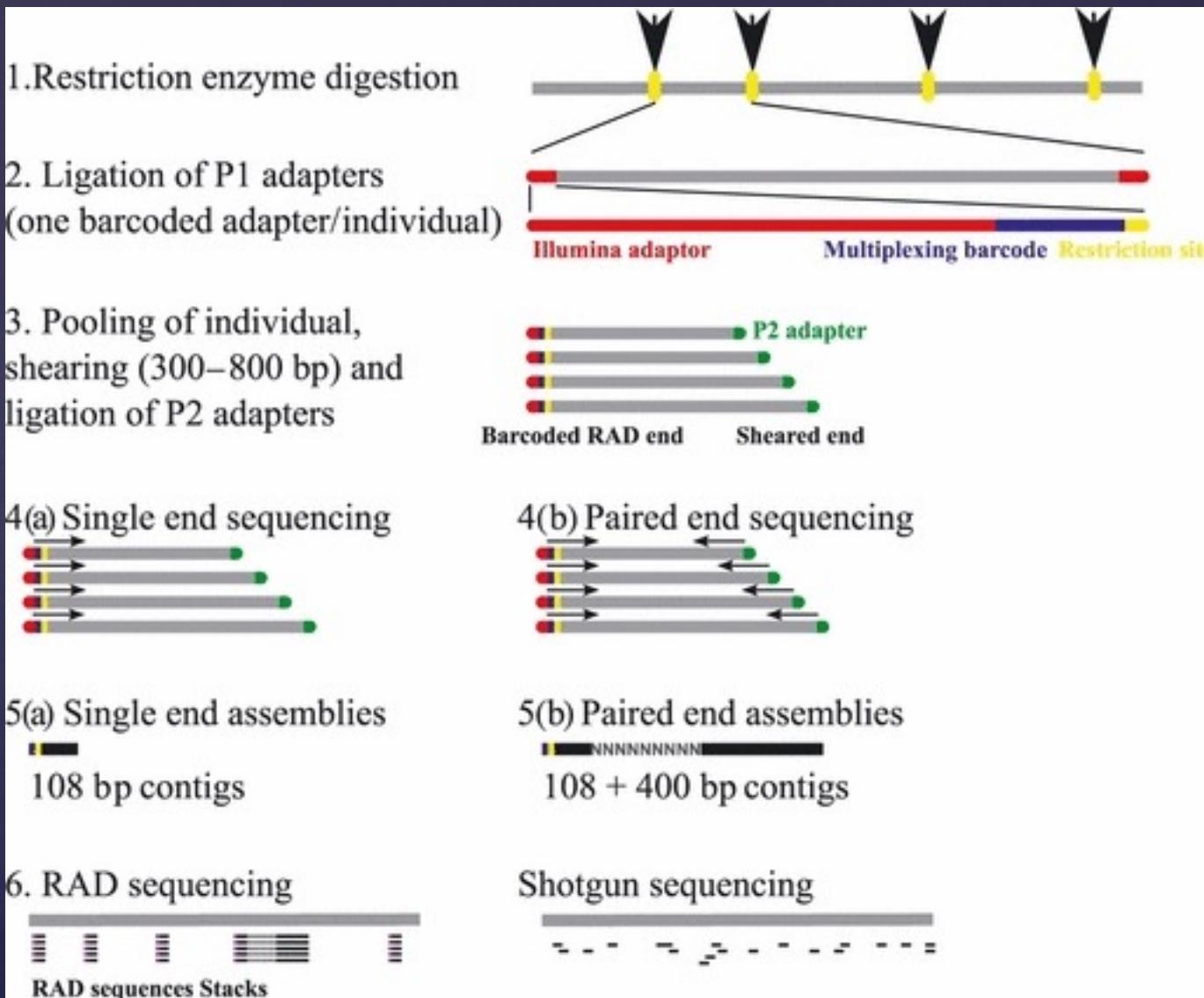


B

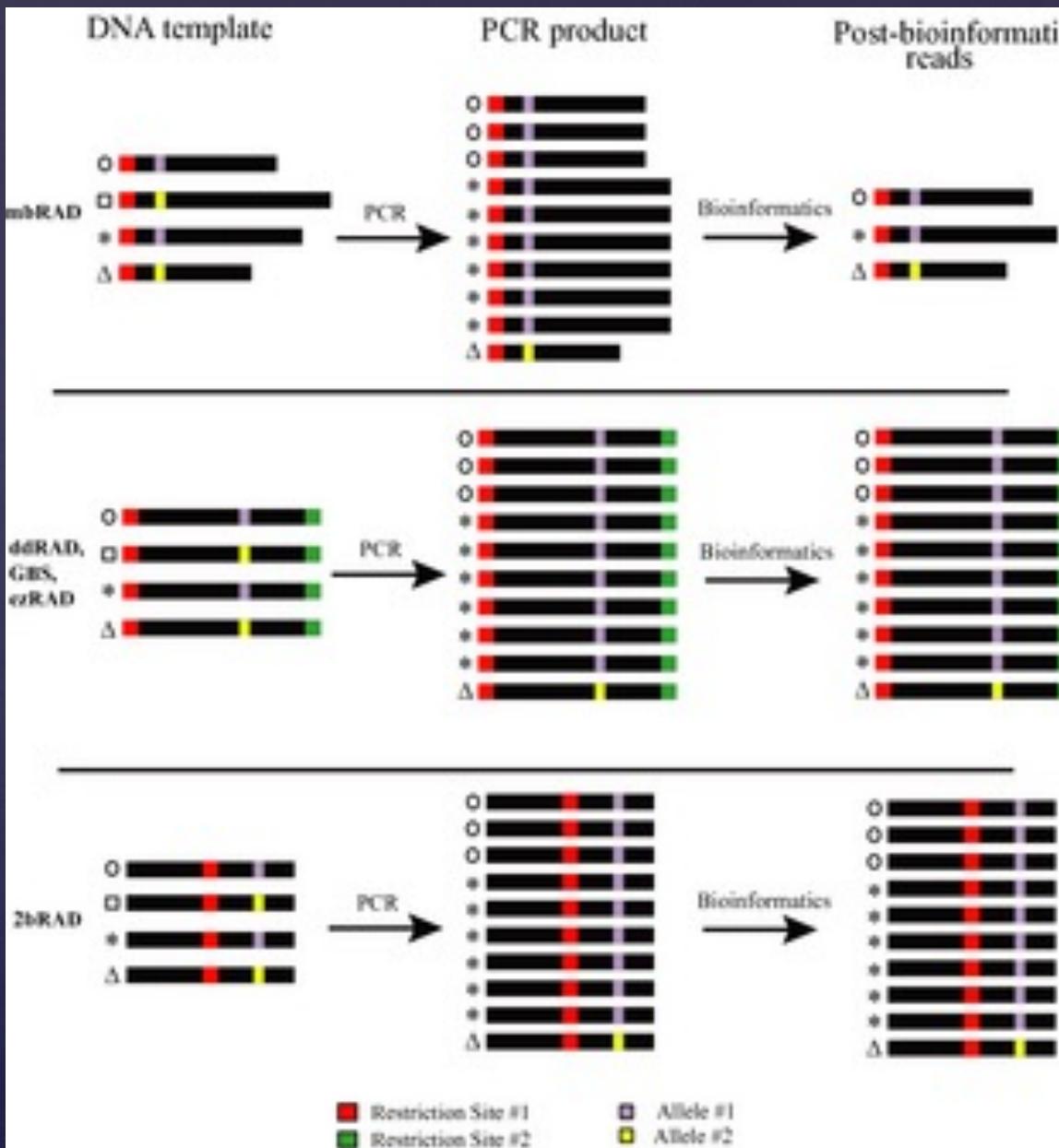
double digest RADseq



# RADs vs ddRADs



# RADs vs ddRADs



# Algunas observaciones sobre el uso de RADseq

MOLECULAR ECOLOGY  
RESOURCES

Molecular Ecology Resources (2017)

doi: 10.1111/1755-0998.12662

## Using a butterflyfish genome as a general tool for RAD-Seq studies in specialized reef fish

JOSEPH D. DIBATTISTA,\*†  PABLO SAENZ-AGUDELO,\*‡ MAREK J. PIATEK,§¶ XIN WANG,\*  
MANUEL ARANDA\* and MICHAEL L. BERUMEN\*

\*Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia, †Department of Environment and Agriculture, Curtin University, PO Box U1987, Perth WA 6845, Australia, ‡Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Valdivia 5090000, Chile, §Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia, ¶Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA



MOLECULAR ECOLOGY  
RESOURCES

Molecular Ecology Resources (2016)

doi: 10.1111/1755-0998.12588

## Draft genome of an iconic Red Sea reef fish, the blacktail butterflyfish (*Chaetodon austriacus*): current status and its characteristics

JOSEPH D. DIBATTISTA,\*† XIN WANG,\* PABLO SAENZ-AGUDELO,\*‡ MAREK J. PIATEK,§  
MANUEL ARANDA\* and MICHAEL L. BERUMEN\*

\*Division of Biological and Environmental Sciences and Engineering, Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia, †Department of Environment and Agriculture, Curtin University, PO Box U1987, Perth, WA 6845, Australia, ‡Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Valdivia 5090000, Chile, §Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

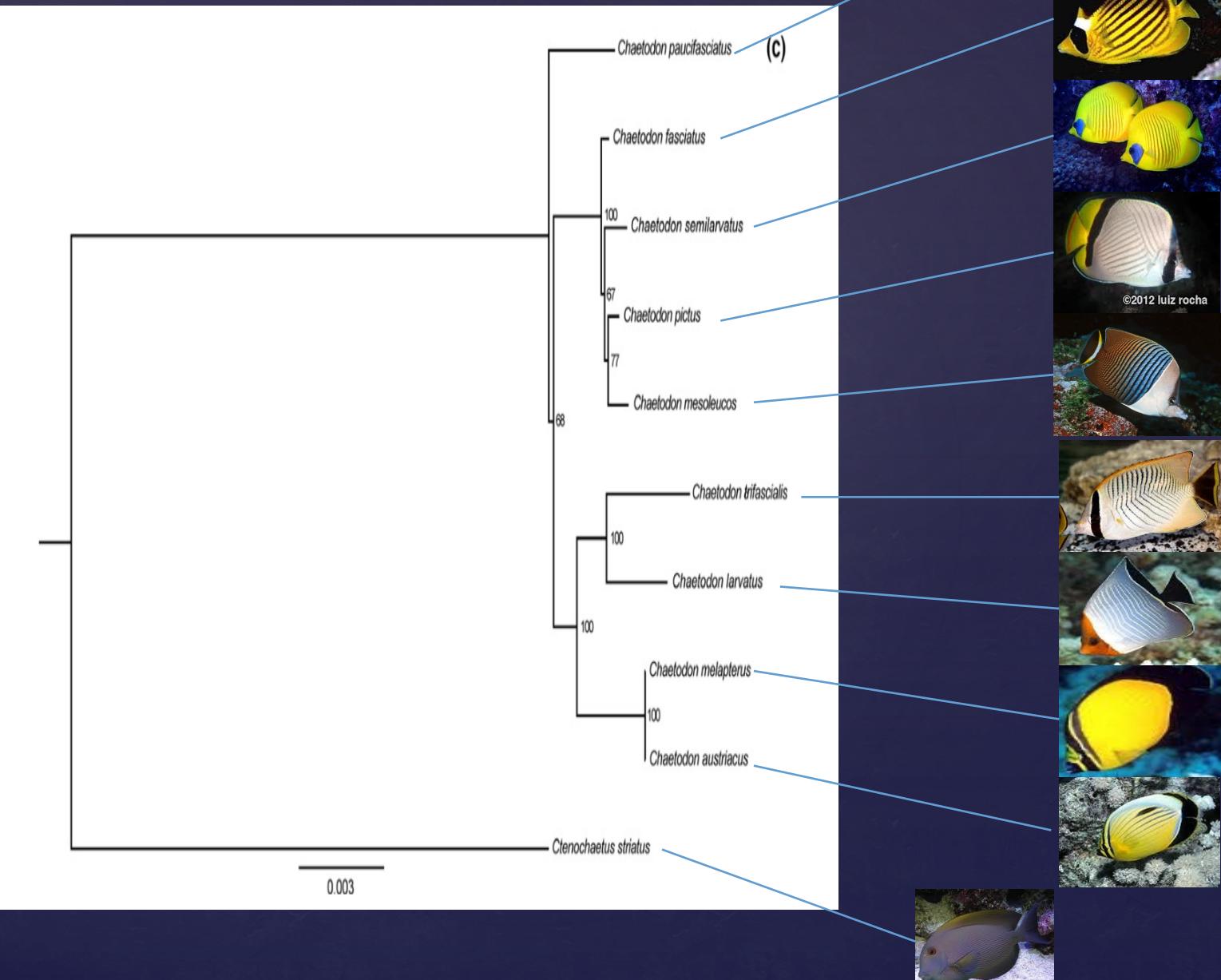
# Algunas observaciones sobre el uso de RADseq

Utilidad de un genoma de referencia para análisis RADseq en otras especies?

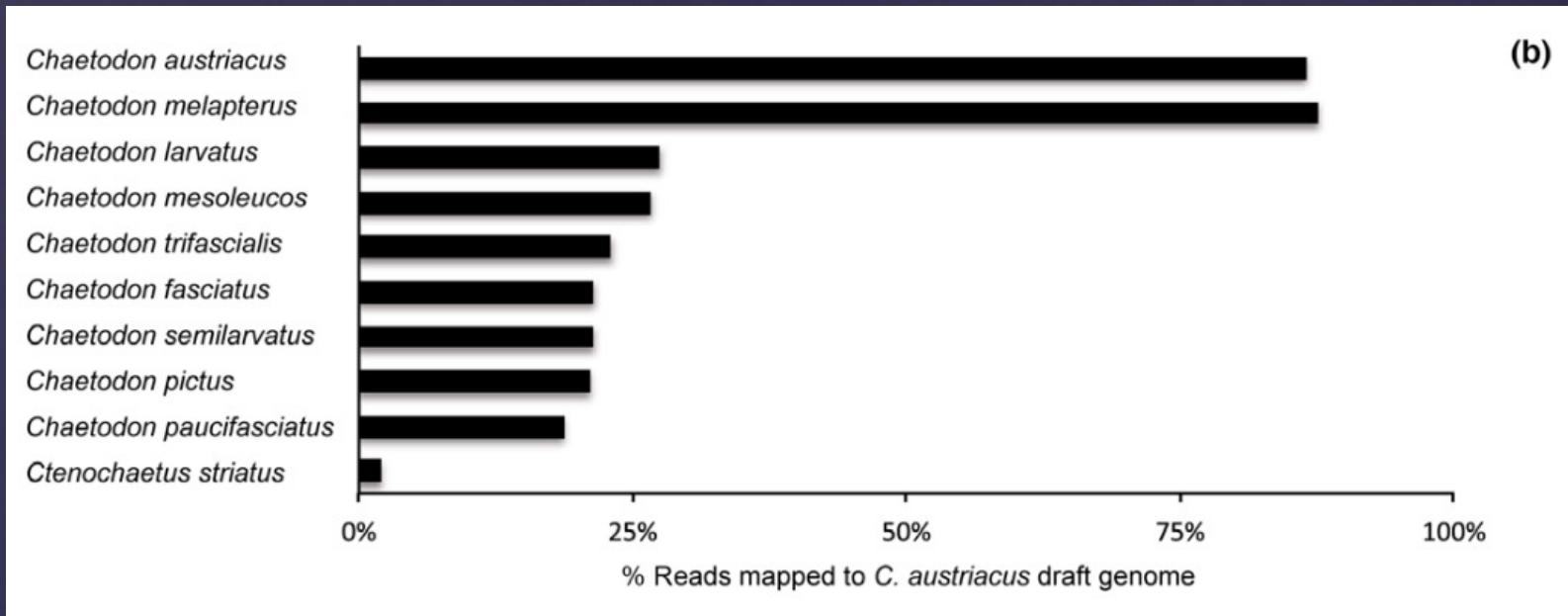
Distribución de SNPs (zonas codificantes vs no codificantes).



# Algunas observaciones sobre el uso de RADseq



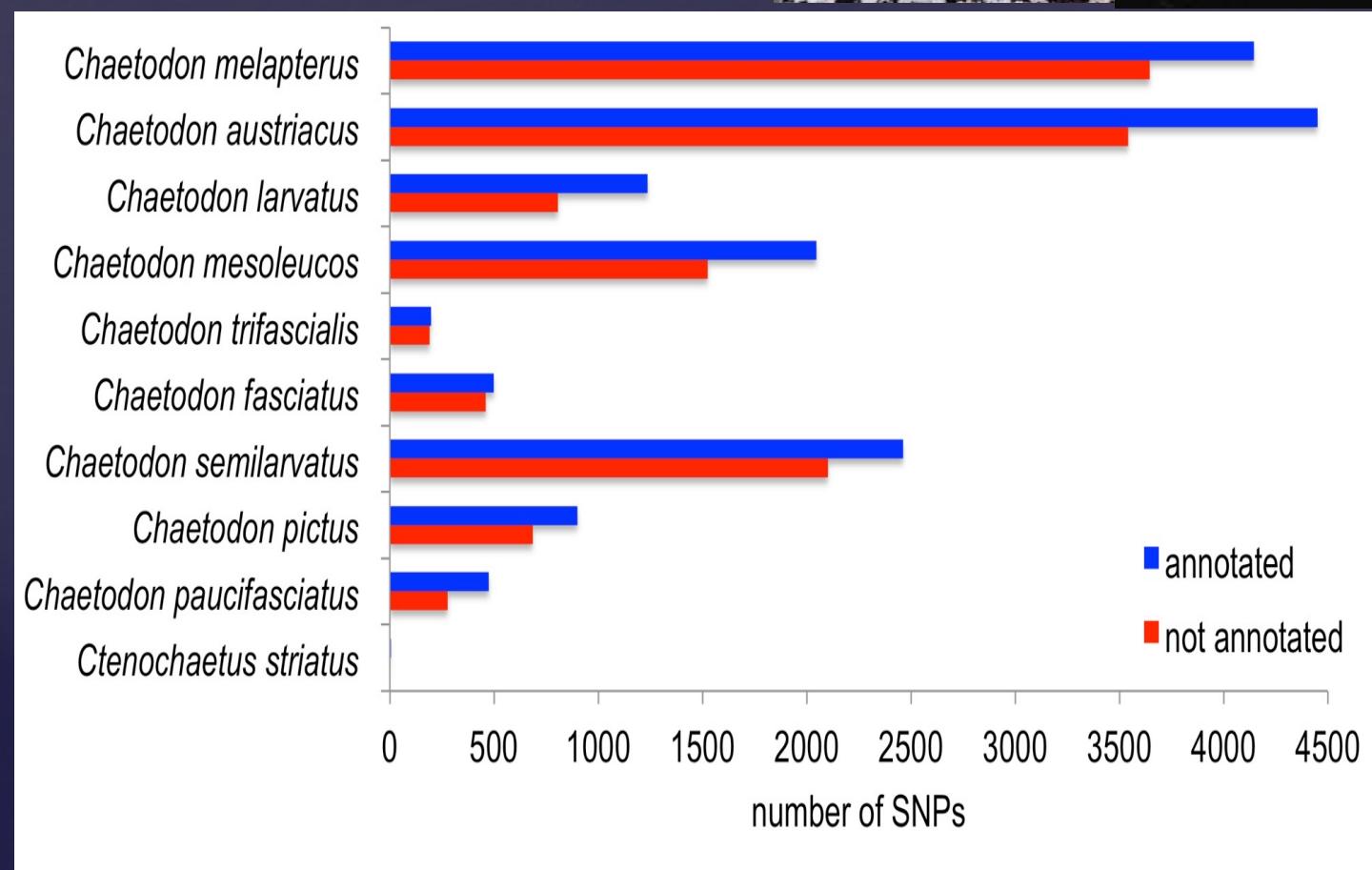
# Algunas observaciones sobre el uso de RADseq



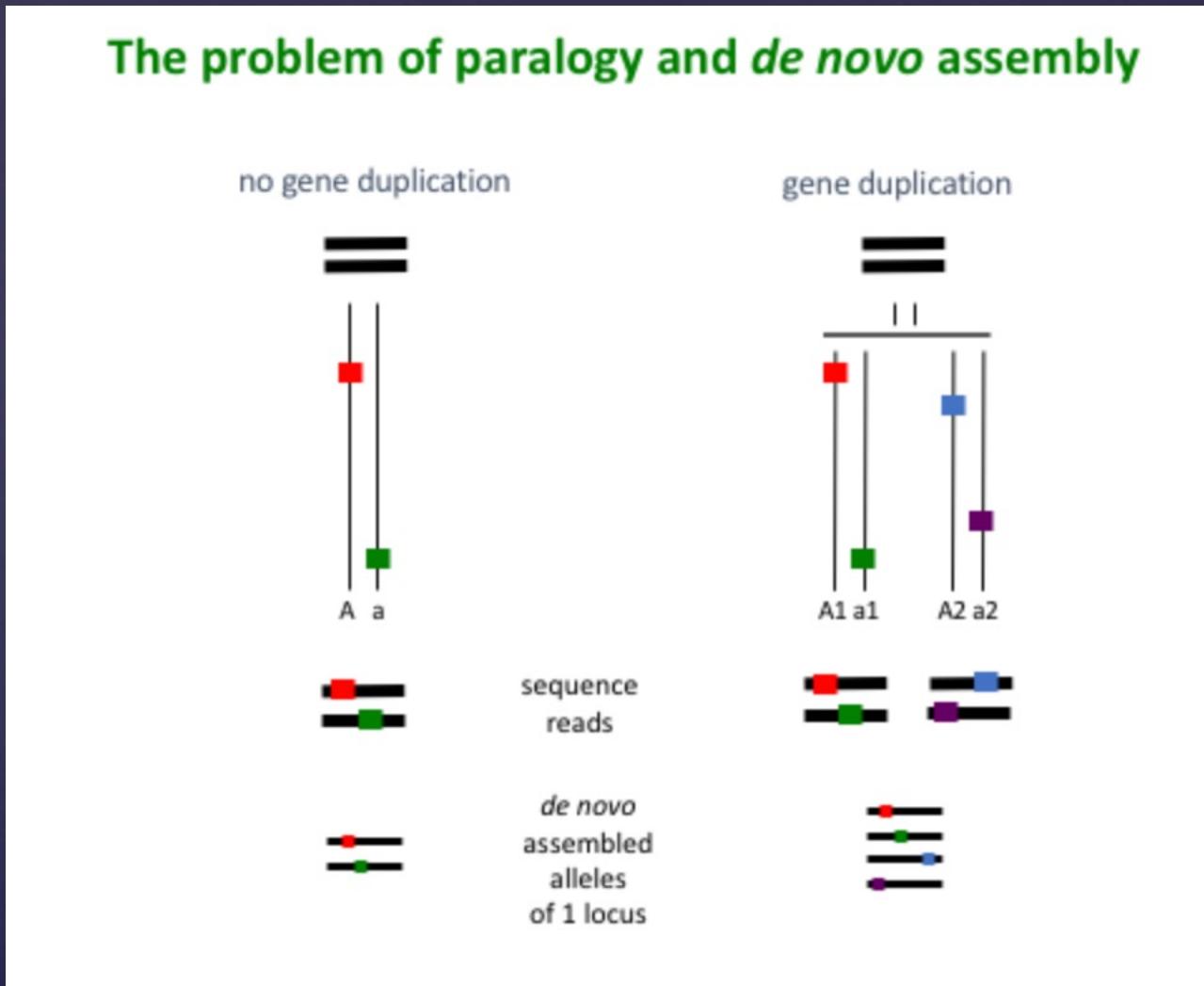
dd-RADseq (*SphI* and *MluCI*)



# Algunas observaciones sobre el uso de RADseq



# Algunas observaciones sobre el uso de RADseq



# Varias opciones para análisis de datos

	<b>Stacks</b> Catchen et al 2013	<b>ipyRAD</b> Eaton et al 2014	<b>Tassel</b> Bradbury et al 2007
<b>Better for</b>	Population genetics non-model spp.	Phylogenies non-model spp.	Population genetics GBS of crops
<b>Type of variation</b>	SNPs, indels	SNPs, indels	SNPs, indels
<b>Method</b>	<i>de novo</i> reference genome	<i>de novo</i> , reference, reference addition and reference subtraction	<i>de novo</i> (not supported anymore) reference genome
<b>Wet lab Protocols</b>	RAD-seq, ddRAD, DaRT, 2bRAD and GBS	RAD, ddRAD, GBS, NextRAD, RApture	GBS
<b>Pair-end?</b>	Yes	Yes	No in <i>de novo</i>
<b>output</b>	<b>VCF</b> , plink, Stacks custom, structure, <b>GenPop</b> , phylip, PHASE, Beagle	<b>VCF</b> , ipyrad custom, <b>phylip</b> , geno, structure, <b>nexus</b> ,	VCF, plink, h5
<b>Comments</b>	Estimates Fst, $\pi$ , He	Jupyter notebooks	Graphical interface useful to “visualize” SNPs

# Análisis de datos con

# Stacks

Stacks is a software pipeline for building loci from short-read sequences, such as those generated on the Illumina platform. Stacks was developed to work with restriction enzyme-based data, such as RAD-seq, for the purpose of building genetic maps and conducting population genomics and phylogeography.



**Download Stacks**

Version 1.24

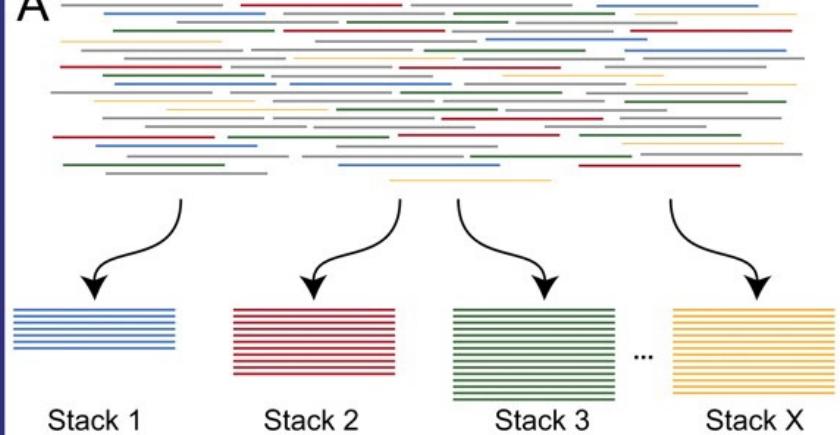
## 1. Demultiplexing

### PROCESS\_RADTAGS

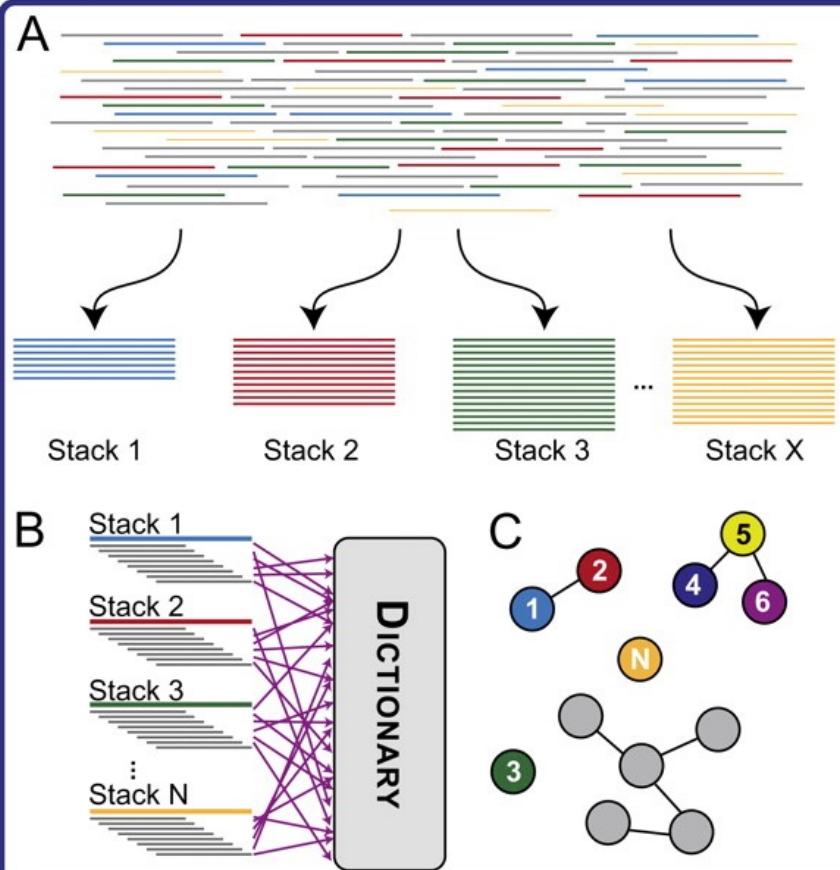


## 2. Construcción de loci

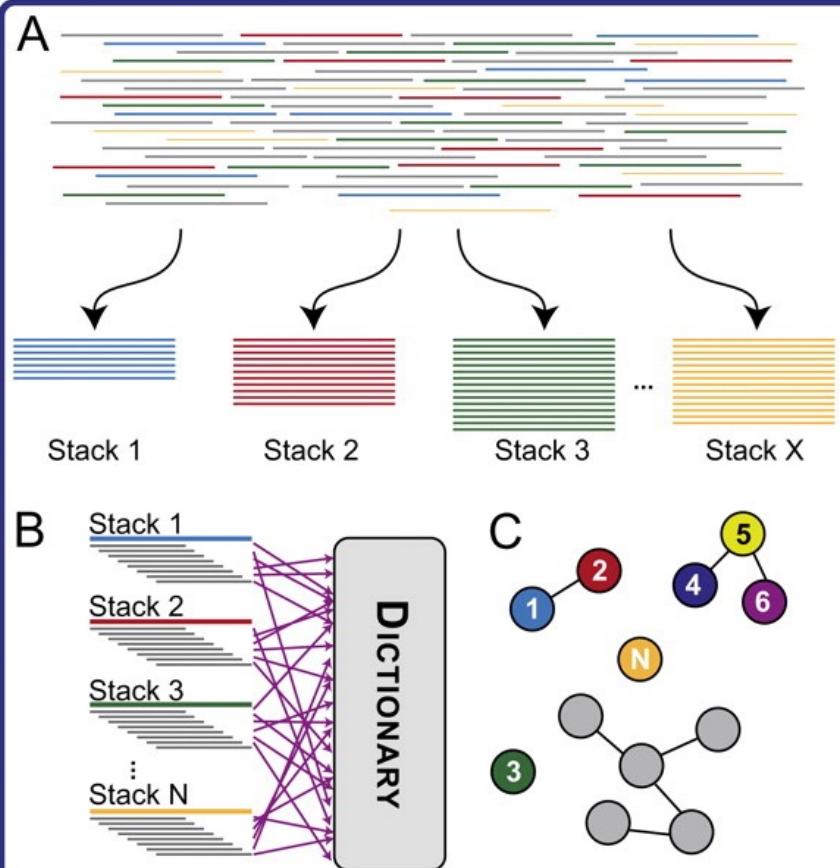
A



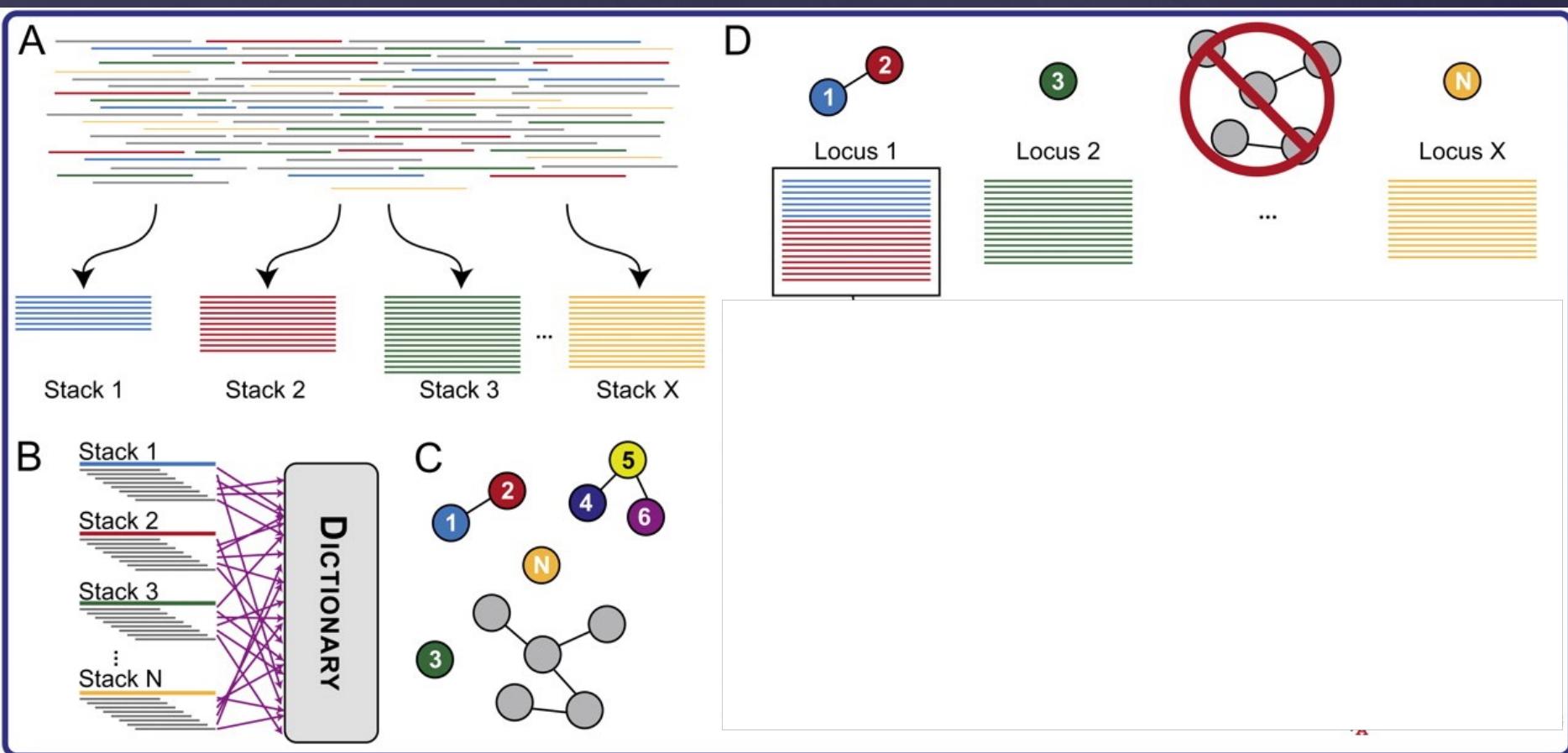
## 2. Construcción de loci



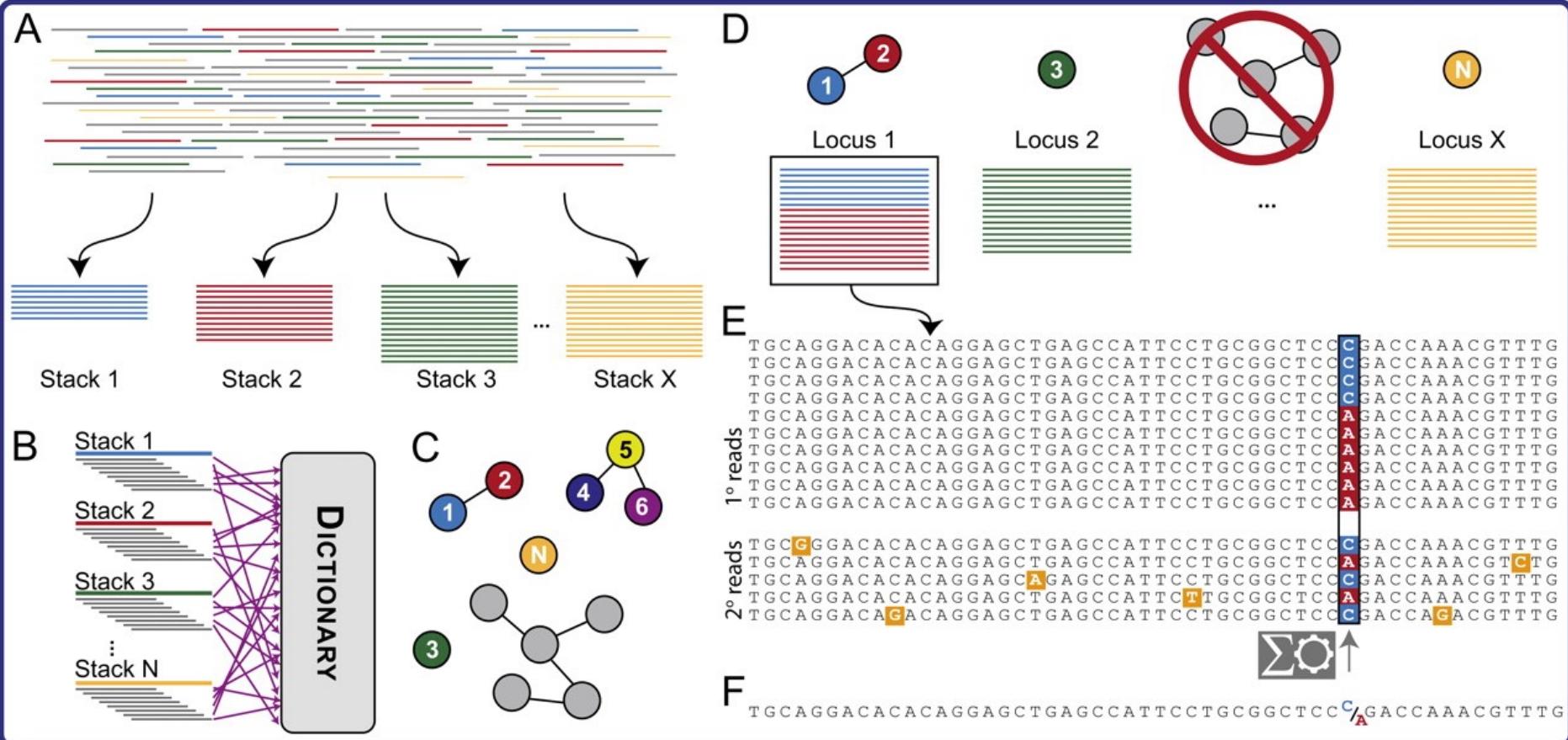
## 2. Construcción de loci



## 2. Construcción de loci

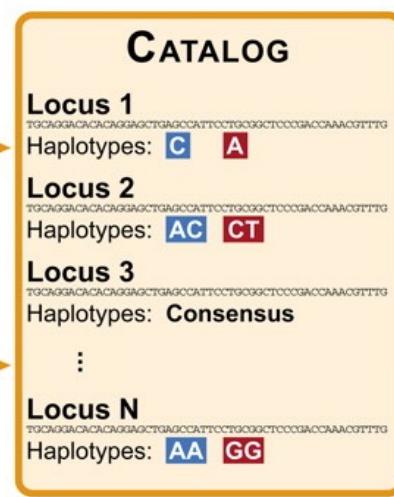
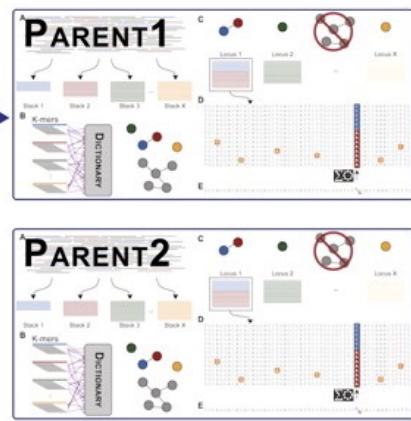


## 2. Construcción de loci

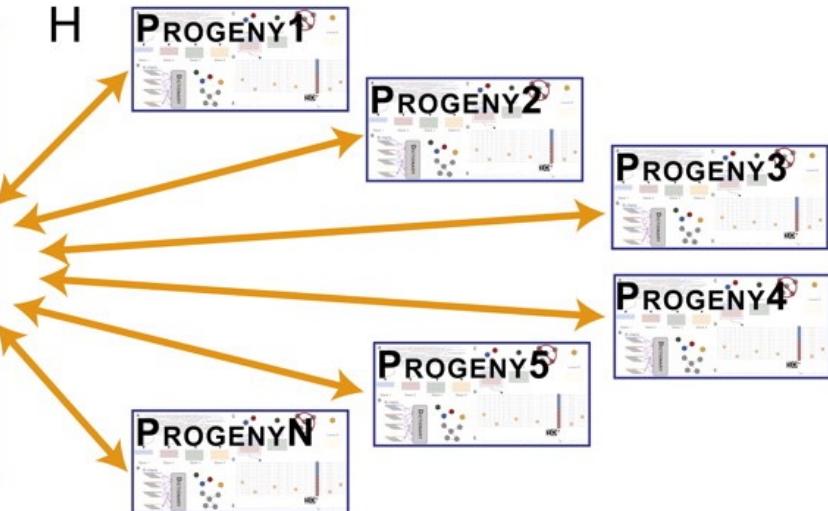


### 3. Construcción de catálogos de loci y asociar todas las muestras al catálogo

G



H



DENOVO\_MAP.PL/REF\_MAP.PL



DENOVO\_MAP.PL/REF\_MAP.PL

REFERENCE  
GENOME

USTACKS  
PSTACKS

BUILD LOCI  
DE NOVO  
BUILD LOCI  
FROM REF

CSTACKS

ASSEMBLE CATALOG

SSTACKS

MATCH TO CATALOG



## GENOTYPES

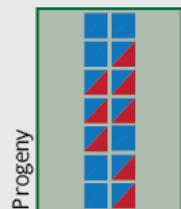
### Locus 1

TGCAGGCAGGAGCTGAGCCCCCTGC<sup>G</sup><sub>A</sub> GCTCCCGACCAA

Haplotypes: C A

Genotypes: A B

♂ AB  
♀ AA



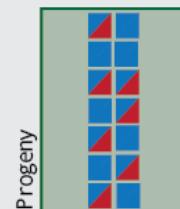
### Locus 2

TGCAGGACACGGACTG<sup>A</sup><sub>C</sub> AGCCCCCTGCGGCCACCA<sup>G</sup><sub>T</sub> AAC

Haplotypes: A C C T

Genotypes: A B

♂ AA  
♀ AB



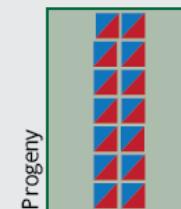
### Locus N

TGCAGGACACAGCTATTCTGCGGGCCCG<sup>G</sup><sub>T</sub> CCAAAC

Haplotypes: G T

Genotypes: A B

♂ AA  
♀ BB



DENOVO\_MAP.PL/REF\_MAP.PL

REFERENCE  
GENOME

USTACKS  
PSTACKS

BUILD LOCI  
DE NOVO

BUILD LOCI  
FROM REF

CSTACKS

ASSEMBLE CATALOG

SSTACKS

MATCH TO CATALOG

Population  
Genetic Mapping

## GENOTYPES

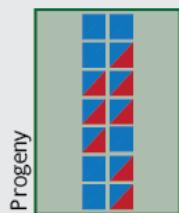
### Locus 1

TGCAGGCAGGAGCTGAGCCCCCTGCC<sup>G</sup>A GCTCCCGACCAA

Haplotypes: C A

Genotypes: A B

♂ AB  
♀ AA



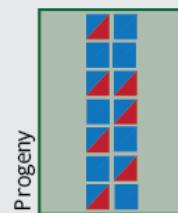
### Locus 2

TGCAGGACACGGACTG<sup>A</sup>C AGCCCCCTGCGGGACCCA<sup>G</sup>T AAC

Haplotypes: AC CT

Genotypes: A B

♂ AA  
♀ AB



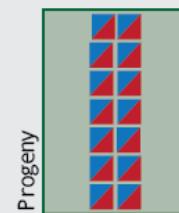
### Locus N

TGCAGGACACAGCTATTCTGCGGGCCCG<sup>G</sup>A CCAAAAC

Haplotypes: G T

Genotypes: A B

♂ AA  
♀ BB



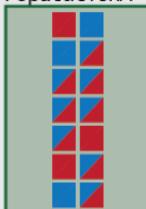
## POPULATIONS

### Locus 1

TGCAGGCAGGAGCTGAGCCCCCTGCC<sup>G</sup>A GCTCCCGACCAAAGTA

Haplotypes: C A

PopulationA PopulationB PopulationC

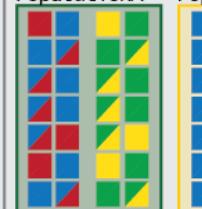


### Locus 2

TGCAGGACACGGACTG<sup>A</sup>C AGCCCCCTGCGGGACCCA<sup>G</sup>T AACGTTTG

Haplotypes: AC CT

PopulationA PopulationB PopulationC

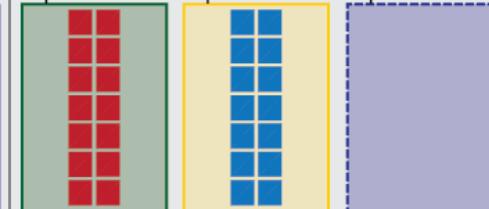


### Locus N

TGCAGGACACAGCTATTCTGCGGGCCCG<sup>G</sup>A CCAAAACGTTG

Haplotypes: G T

PopulationA PopulationB PopulationC



Preguntas?