Introdução ao uso de dados genômicos em Sistemática

Sandra Reinales & Luana J. Sauthier
spreinalesl@gmail.com; sauthier@ib.usp.br

Departamento de Botânica, Universidade de São Paulo

2024

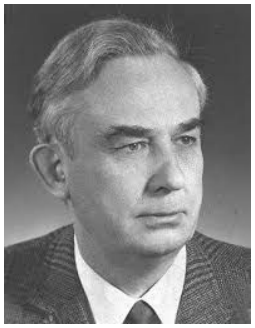
Plano de aula

- Contextualização
- Técnicas de sequenciamento: Sanger vs. HTS
- Panorama geral dos métodos mais usados na sistemática hoje:
  - RAD-Seq
  - Genome skimming
  - Target capture
- Fluxo de trabalho no target capture
- Considerações finais
- Para saber mais

Contextualização

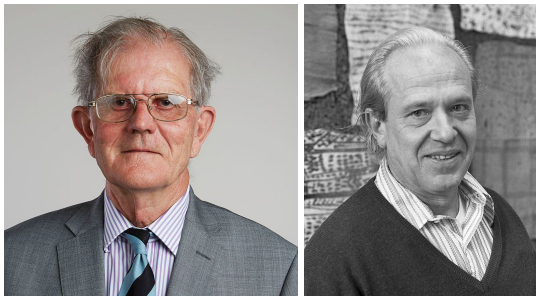
Fontes de dados para inferência filogenética: MORFOLOGIA

Inicios da sistemática

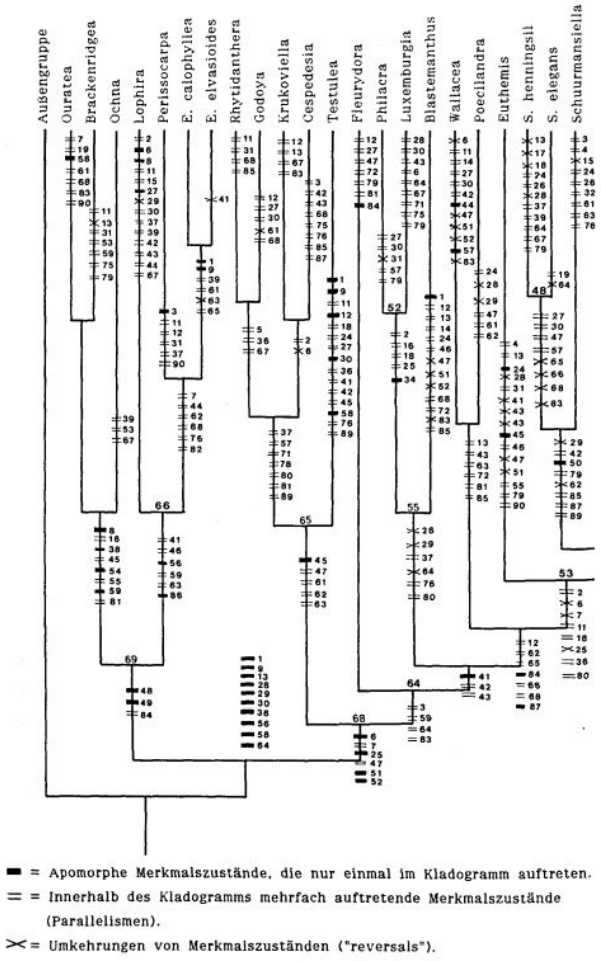


Cladística
Willi Hennig (1950-1966)

Modelos probabilísticos
Joseph Felsenstein (1981)



Modelos probabilísticos
Anthony Edwards & Lucas Cavalli-Sforza (1964)



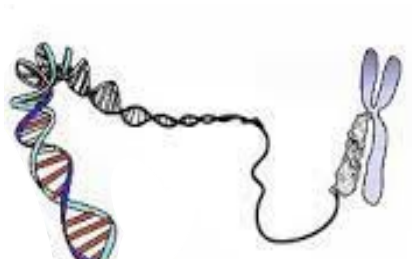
Amaral. Bot. Jahrb. Syst. 113 (1991)

Contextualização

Fontes de dados para inferência filogenética: DNA

- Sanger sequencing (d. 1977 - c. 1986)
  - *rbcL* (inícios 90's)
  - ITS, *ndhF*, *matK* (1995)

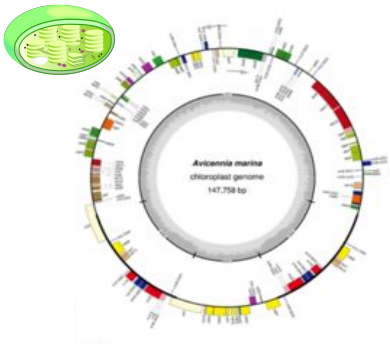
Nuclear



60 Mbp - 150 Gbp

(Heslop-Harrison & Schmidt 2012)

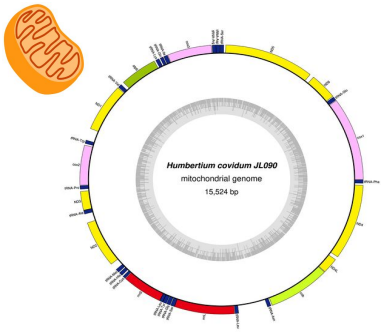
Plastidial



120 - 160 kbp

(Kress et al. 2022)

Mitocondrial



200 - 2000 kbp

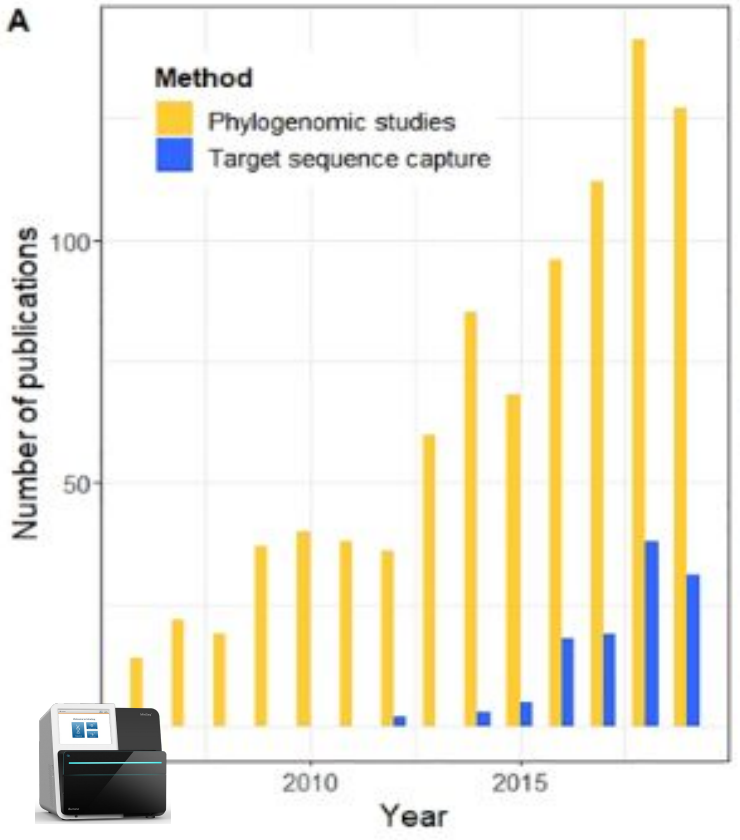
(Morley & Nielsen 2017)

O tamanho dos genomas varia consideravelmente nas plantas

Contextualização

Fontes de dados para inferência filogenética: DNA

- Desenvolvimento do NGS iniciou na década dos 90’s
- Solexa - Cambridge Chemistry Department (1998)
- Primeiro genoma seq. (2005)
- Genome Analyzer (2006)
- Solexa - Illumina (2007)



Andermann *et al.* Front. Genet. 10 (2020)

Contextualização

Fontes de dados para inferência filogenética: DNA

Next Generation Sequencing (NGS)

ou

High Throughput Sequencing (HTS)

Contextualização

Fontes de dados para inferência filogenética: DNA

2º Geração

Short reads

DNA fragmentado

- Genome skimming
- RAD-Seq
- Target sequence capture

3º Geração

Long reads

DNA **não** fragmentado

Técnicas de sequenciamento

Sanger (1977)

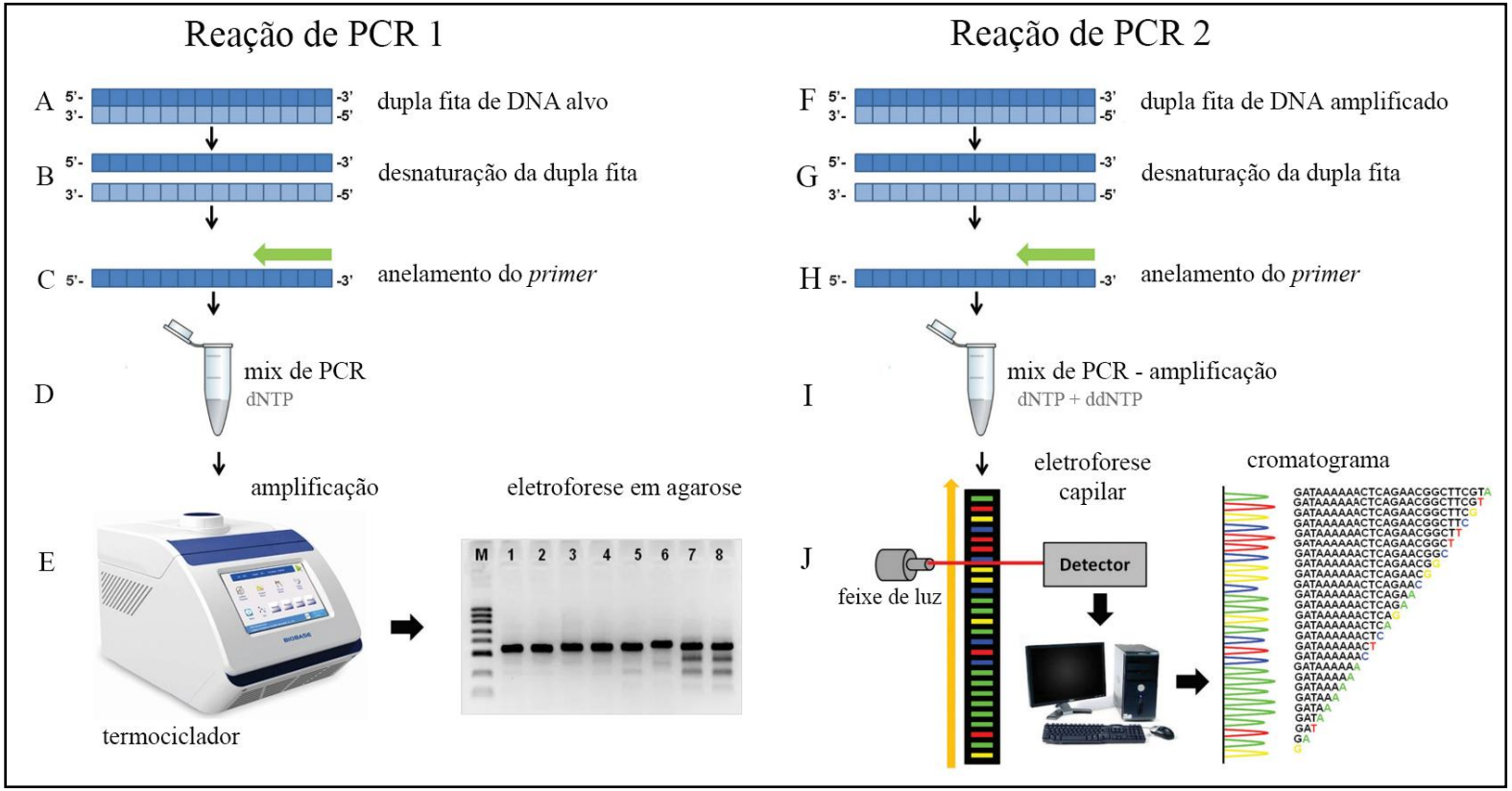
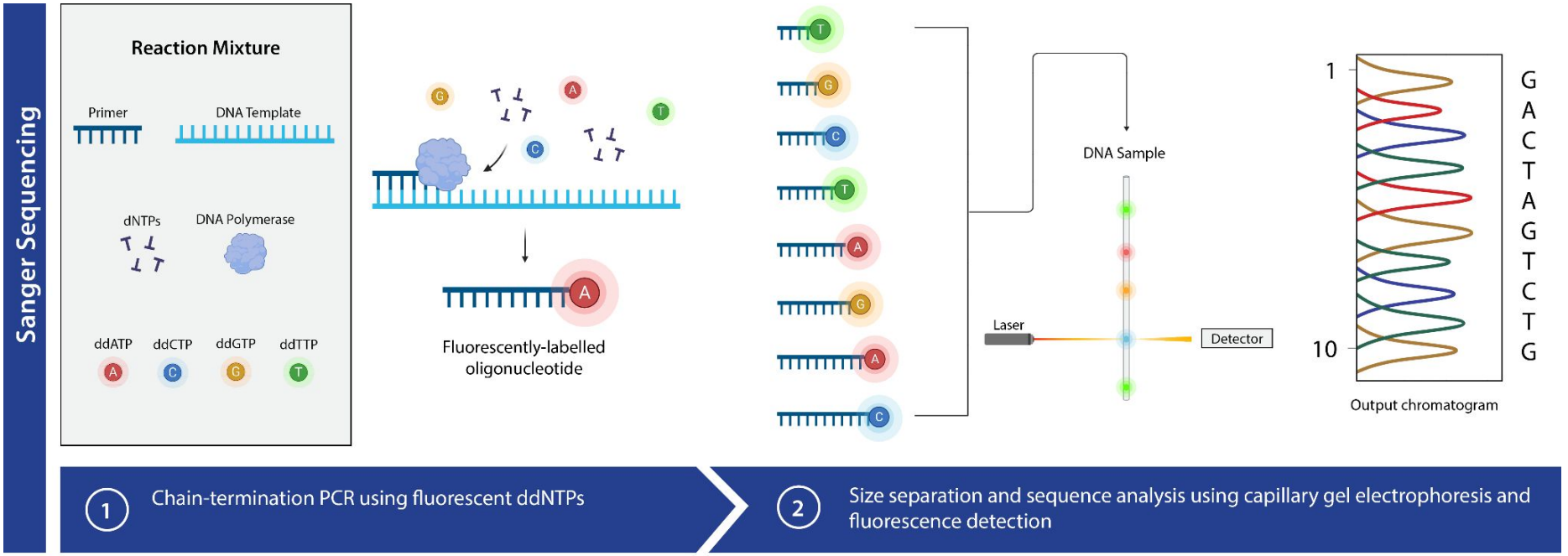


Figura 1: Sequenciamento Sanger diferenciando as etapas um (A-E) e dois (F-J) de reações de PCR . (A-B; F-J) desnaturação da dupla fita do DNA alvo; (C; G) anelamento dos *primers;* (D-E; I-J) transcrição da nova fita; (J) sequenciamento. Fonte: adaptado de Fietto & Maciel (2015).

Técnicas de sequenciamento

Sanger (1977)



https://www.aatbio.com/catalog/sanger-sequencing

Técnicas de sequenciamento

Comparação entre tecnologias de sequenciamento

	Sanger	NGS short reads
Input (amostra)	Clones de uma região específica do DNA (produtos de PCR)	Biblioteca de DNA do genoma completo ou regiões selecionadas
Número de amostras	96-384	1-384 por canaleta (8 canaletas)
Duração da preparação das amostras para o sequenciamento	1 dia	3-4 dias
Quantidade de dados gerados	1 fragmento (2 seqs) /amostra	Milhões de sequencias (reads) /amostra
Tamanho do fragmento/sequencia	300-1500 bp	100-500 bp
Processamento dos dados	Simples, rápido, não requer programação	Mais complexa, algumas horas/semanas requer HPC e programação

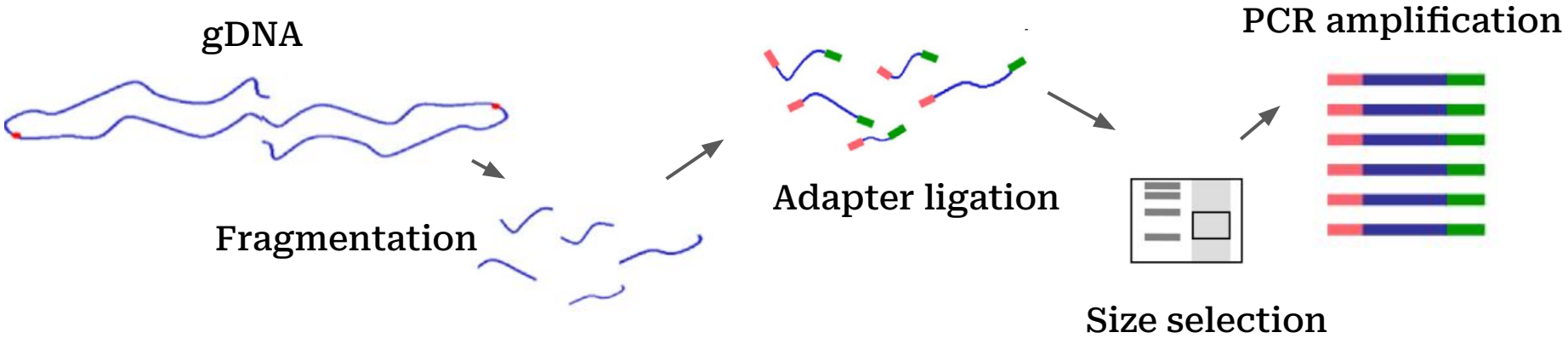
Adaptado de: https://github.com/carol-siniscalchi

O que é uma biblioteca genômica?

Definição

Uma biblioteca para NGS é uma coleção de fragmentos de DNA de tamanhos semelhantes que contém sequências adaptadoras conhecidas nas extremidades 5' e 3'.

- Uma biblioteca corresponde a uma única amostra.
- Podem ser agrupadas e sequenciadas na mesma corrida.



Adaptado de: Dijk *et al.* Exp. Cell Res. 322 (2014)

O que é uma biblioteca genômica?

Kits comerciais

NEBNext® Ultra™ II for DNA Library Prep

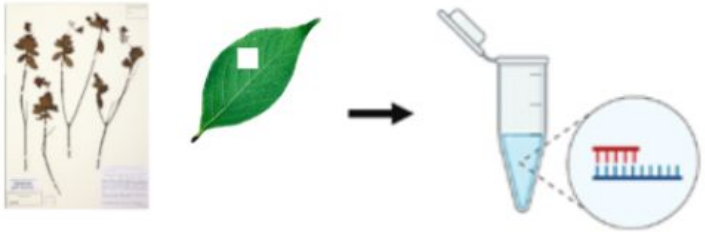
Nextera XT DNA Library Preparation Kit

KAPA HyperPrep Kits

O que é uma biblioteca genômica?

Etapas: extração do DNA

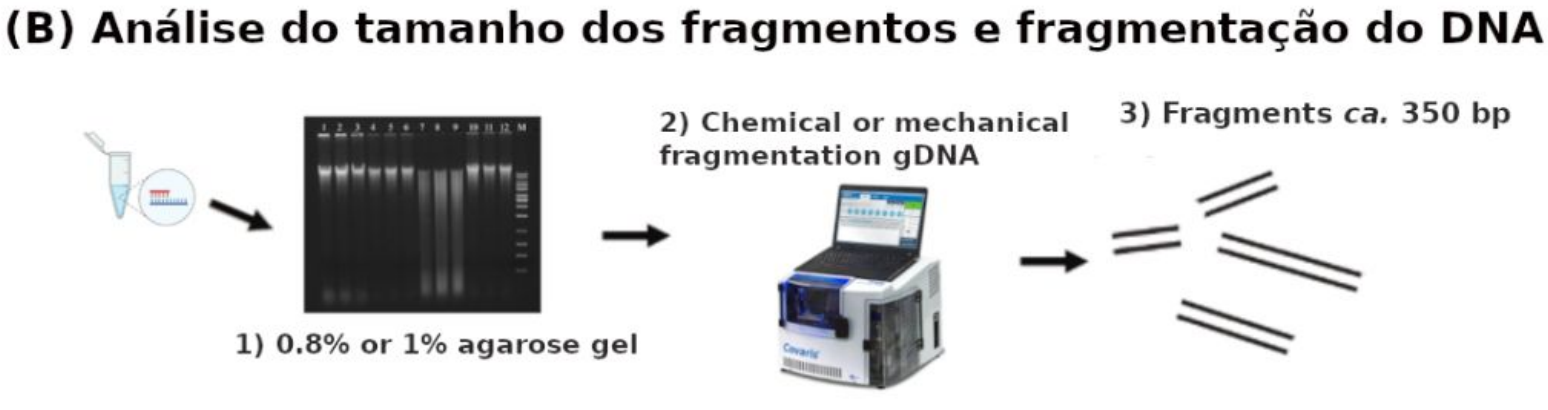
(A) Extração de DNA com o protocolo CTAB



RECOMMENDED INPUT AMOUNTS	TOTAL DNA
Ultra II DNA Library Prep Kit (NEB #E7645)*	500 pg–1 µg

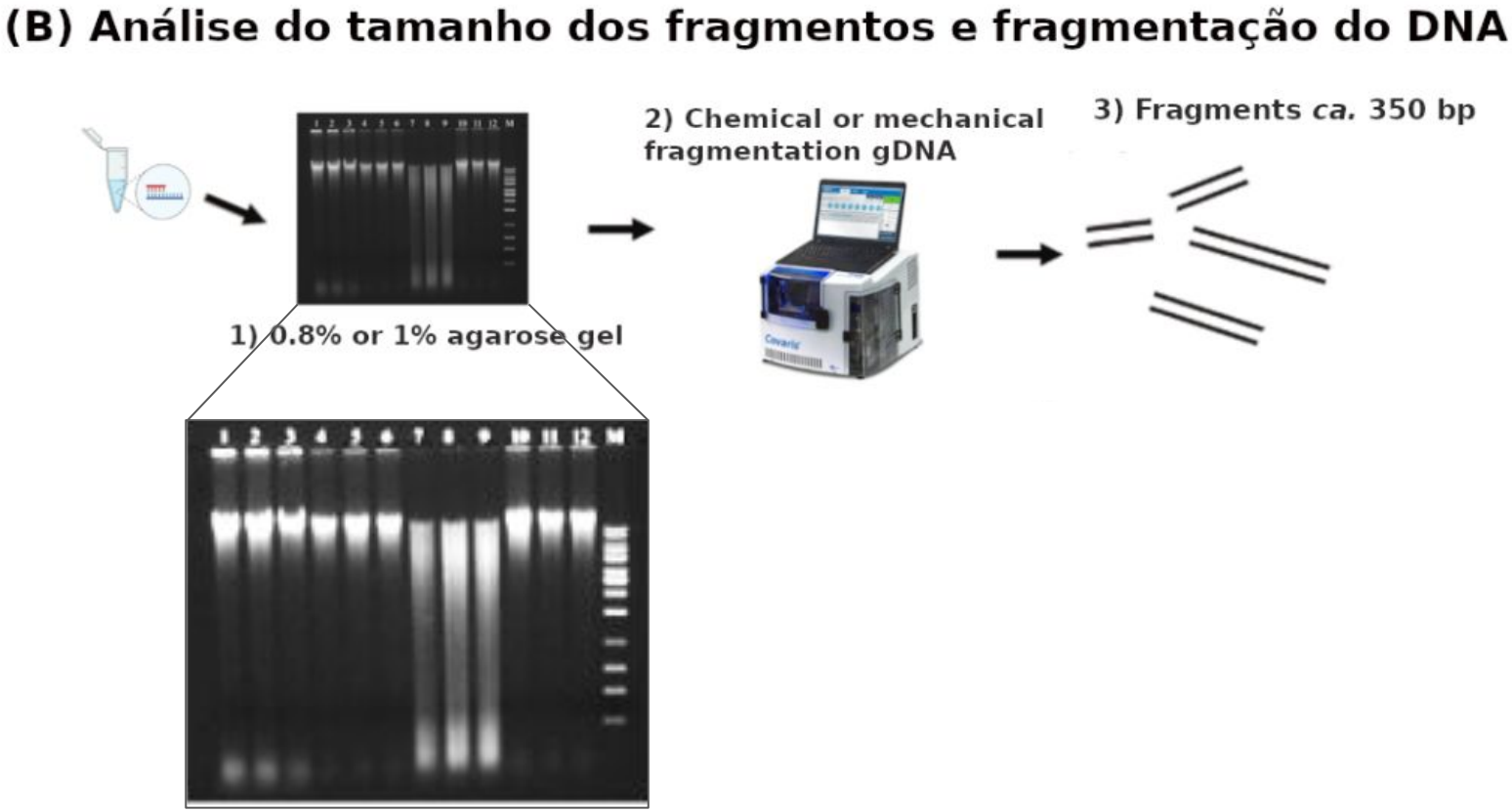
O que é uma biblioteca genômica?

Etapas: análise do DNA extraido



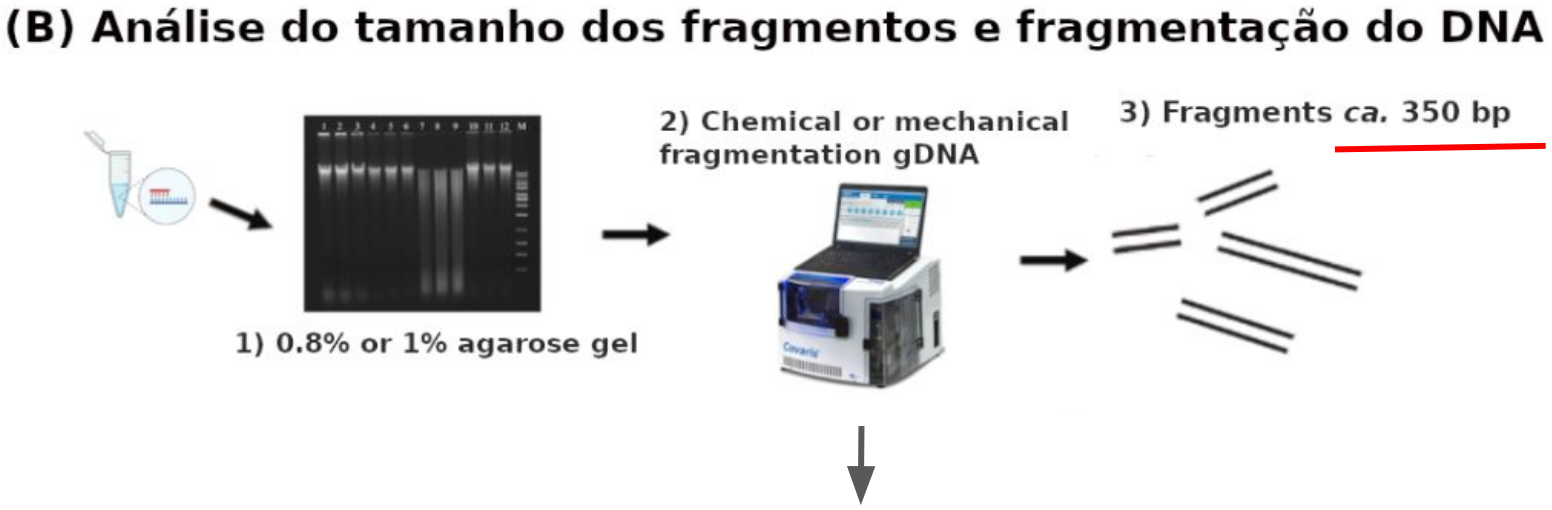
O que é uma biblioteca genômica?

Etapas: análise do DNA extraido



O que é uma biblioteca genômica?

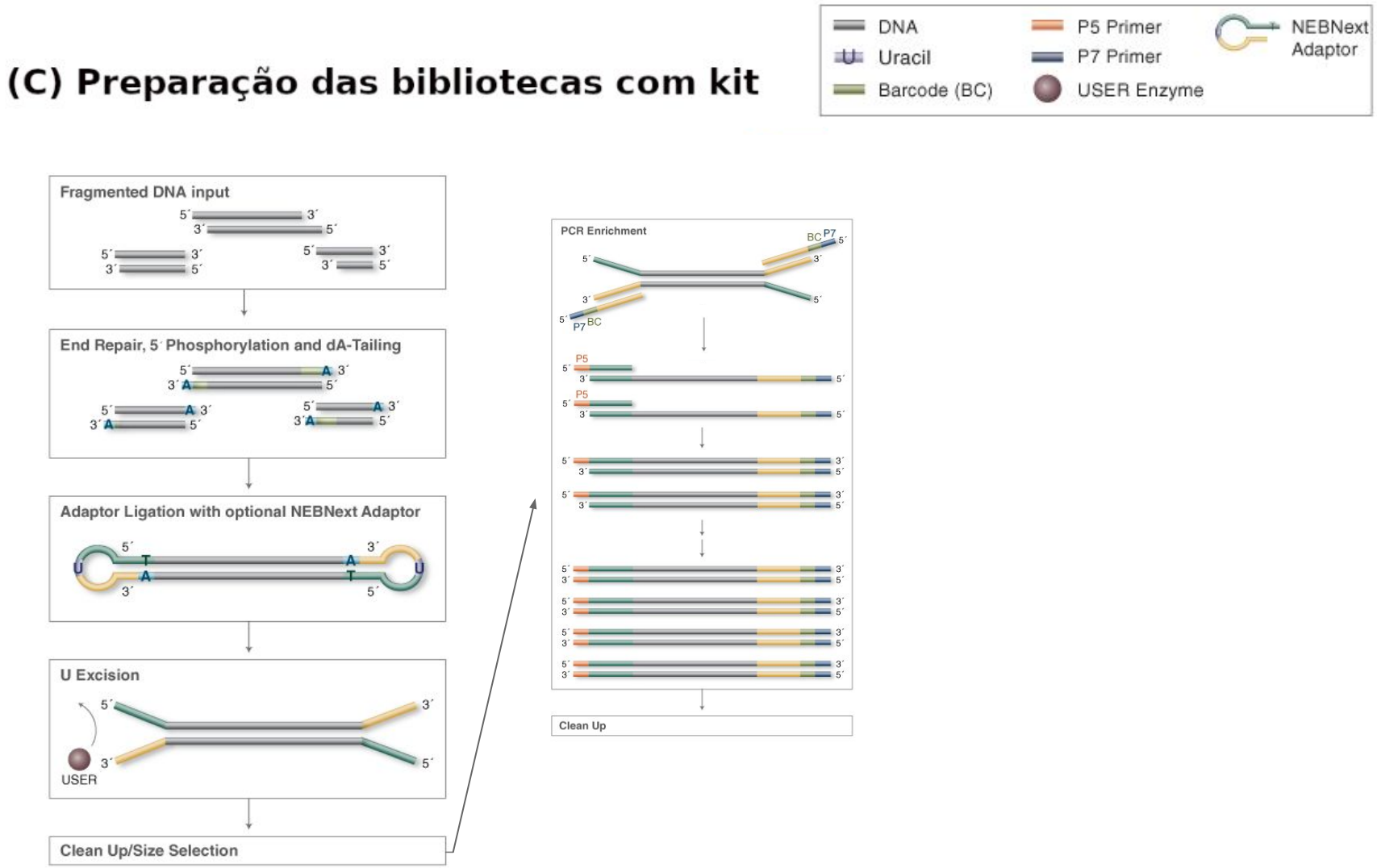
Etapas: fragmentação do DNA



- Covaris - sonicador UDS $5.28 c/u
- Enzimas de restrição

O que é uma biblioteca genômica?

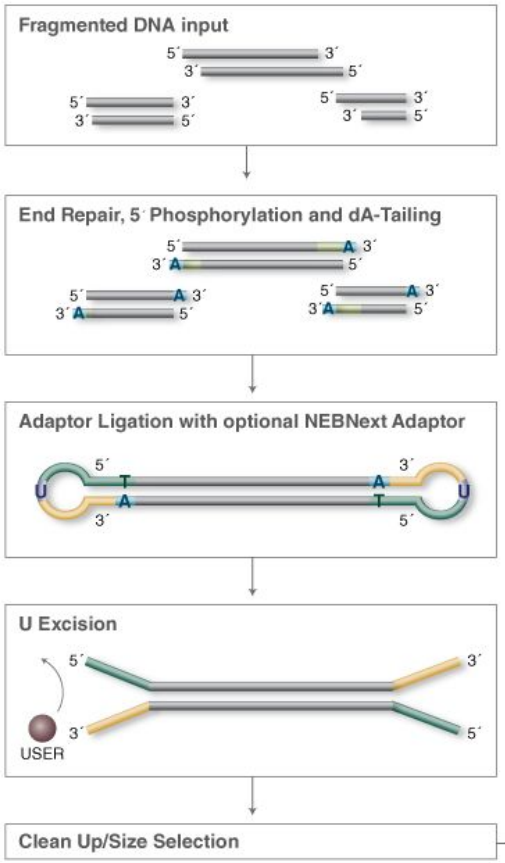
Etapas: ligação de adaptadores

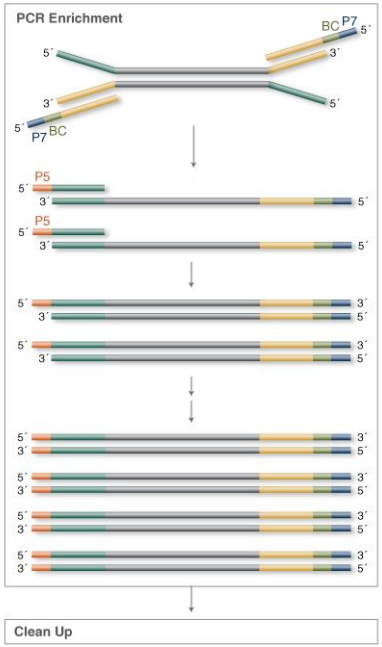


O que é uma biblioteca genômica?

Etapas: ligação de adaptadores

(C) Preparação das bibliotecas com kit





P5 Primer

AATGATACGGCGACCACCGAGATCTACAC

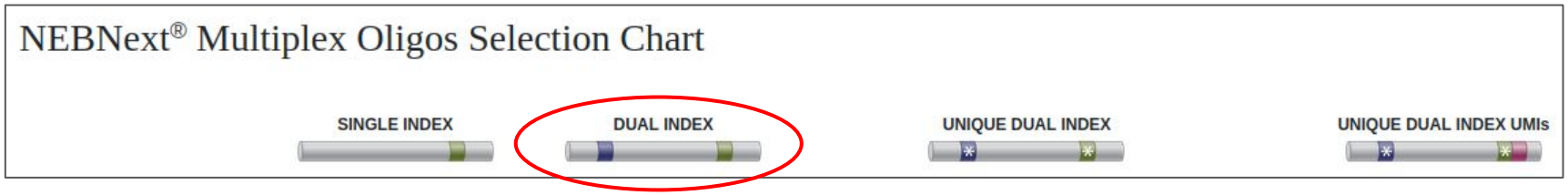
P7 Primer

ATCTCGTATGCCGTCTTCTGCTTG

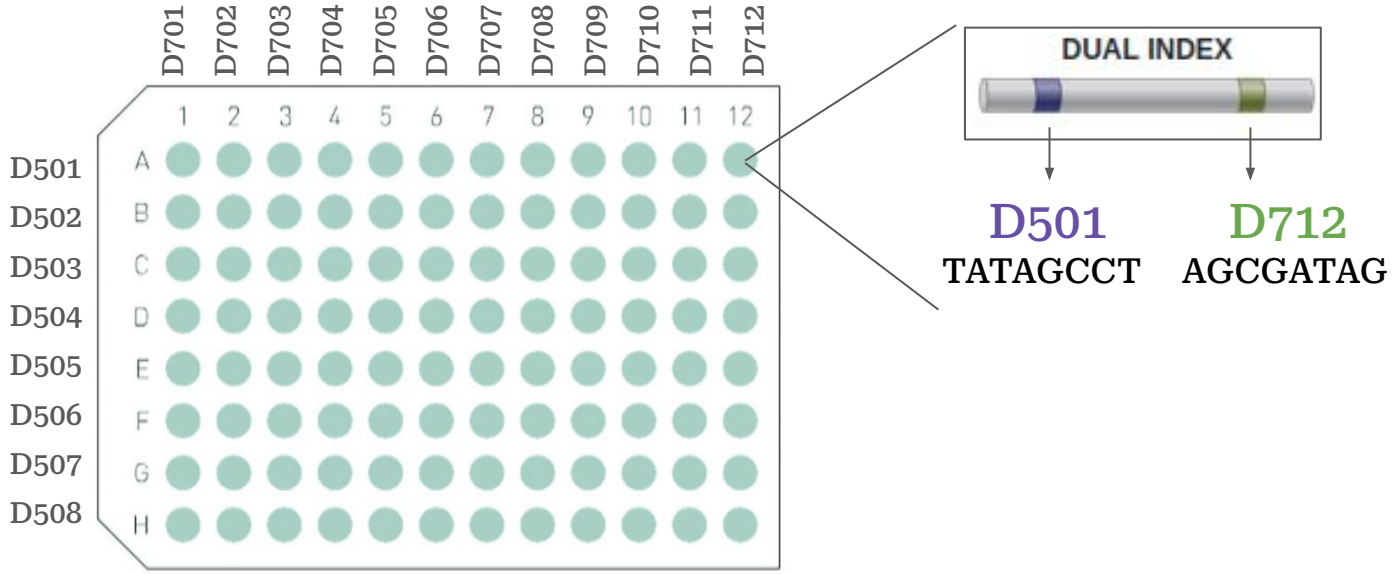
O que é uma biblioteca genômica?

Etapas: ligação de adaptadores

(C) Preparação das bibliotecas com kit

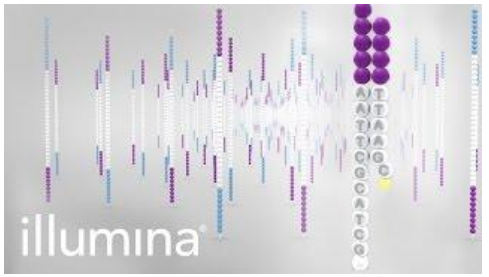


Multiplexing



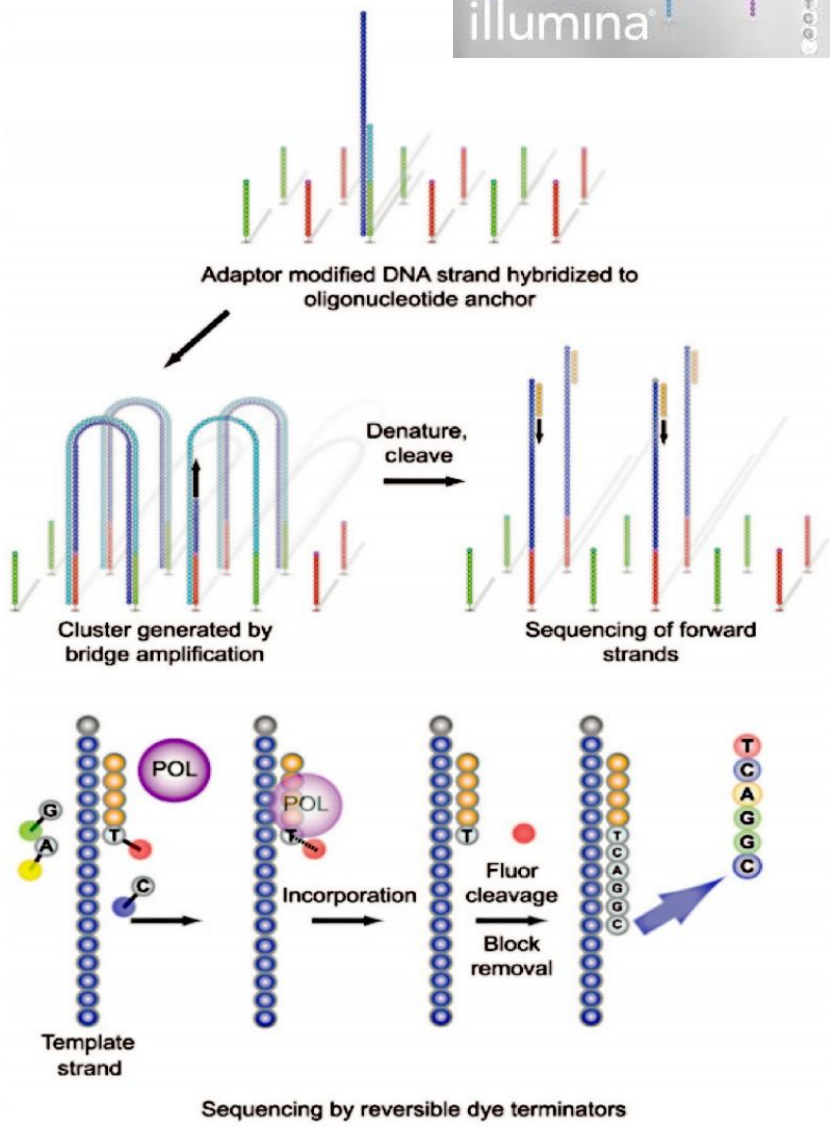
Técnicas de sequenciamento

Sequenciamento Illumina



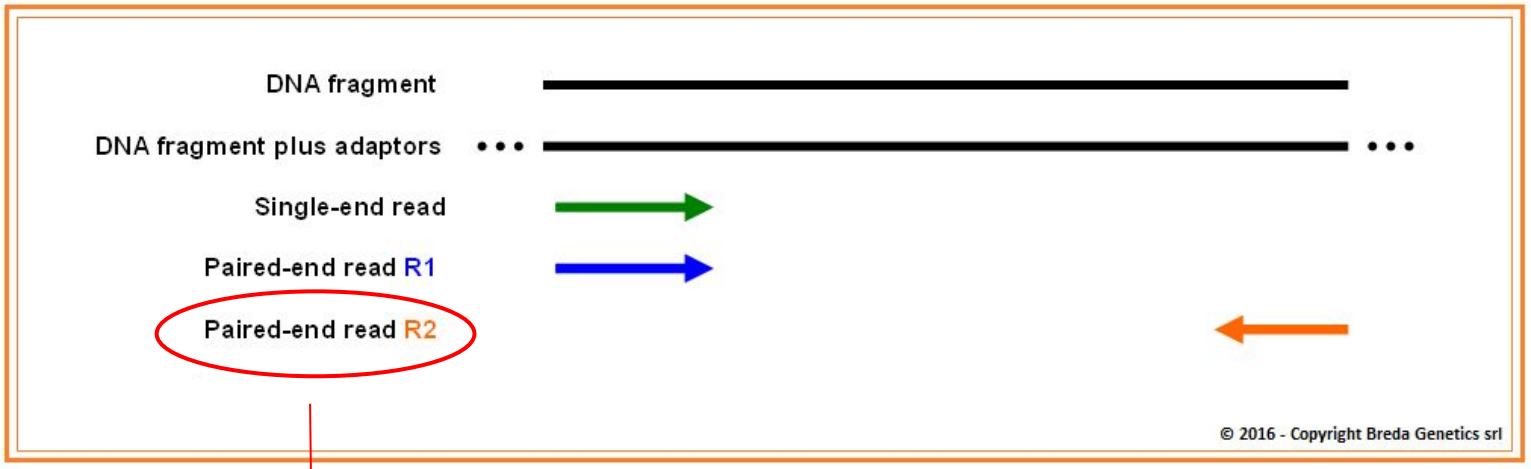
Três etapas principais:

1. Ligação das bibliotecas na plataforma
2. PCR fase sólida (Bridge amplification)
3. Sequenciamento por síntese



Técnicas de sequenciamento

Sequenciamento Illumina



https://bredagenetics.com/single-end-reads-e-paired-end-reads/?lang=it

SRR10951301_1.fastq.gz

SRR10951301_2.fastq.gz

Técnicas de sequenciamento

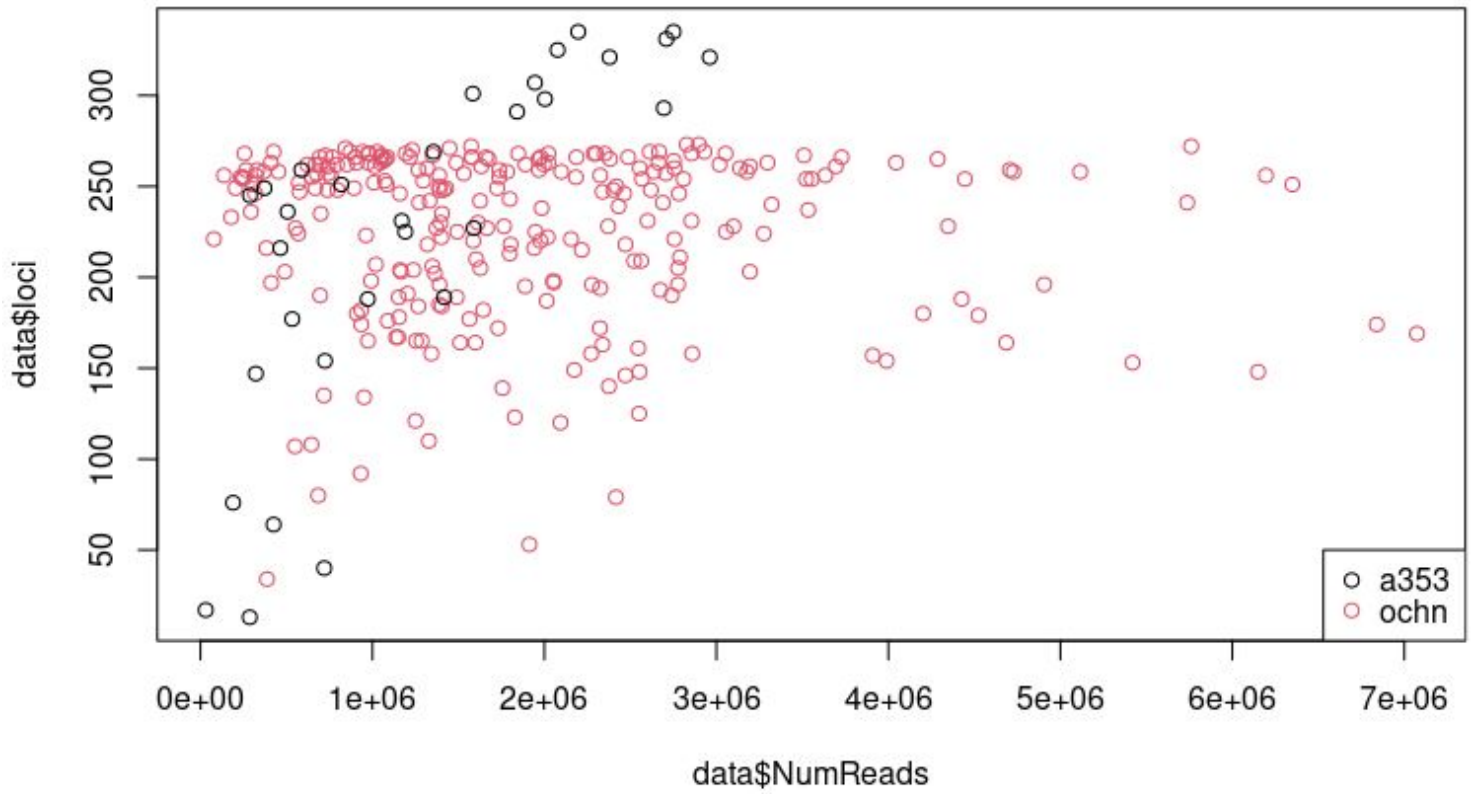
Plataformas Illumina

https://www.illumina.com/systems/sequencing-platforms.html

6 bilhões / 8 canaletas = 750 milhões reads ⟶ 750 milhões / 96 amostras = ~ 8 milhões de reads por amostra

Técnicas de sequenciamento

Sequenciamento Illumina



Os dados chegaram, e agora?

O que é um pipeline?

Conjunto de ferramentas (pacotes, programas e scripts):

- Instalação conjunta
- Estão interligadas: saída de uma é a entrada da outra
- Facilita análises padrão

O que é um pipeline?

Conjunto de ferramentas (pacotes, programas e scripts):

- Instalação conjunta
- Estão interligadas: saída de uma é a entrada da outra
- Facilita análises padrão

- Pouco flexíveis: análises mais complexas precisam de programação
- Desconhecimento do procedimento

O que é um pipeline?

	Read cleaning	Sequence engineering	Intron recovery	MSA generation	Allele phasing	SNP extraction	Ease of installation
Hybpiper (Johnson et al., 2016)	○	●	●	○	○	○	◒
Phyluce (Faircloth, 2016)	◒	○	○	●	●	◒	●
Secapr (Andermann et al., 2018)	●	◒	◒	●	●	◒	●
Hybphaser (Nauheimer, et al., 2021)	○	●	●	●	●	●	●
Captus (Ortiz, et al., 2023)	●	●	●	●	○	○	●

Adaptado de : Andermann *et al.* Front. Genet. 10 (2020)

Comparação dos métodos

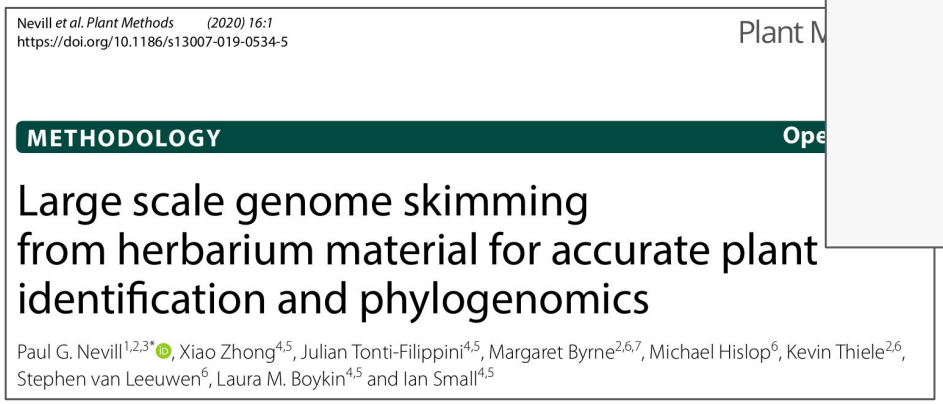
Genome skimming, RAD-Seq, Target capture

	Genome skimming	RAD-seq	Hyb-seq
Demand for plant materials	Fresh, silica-gel dried plant tissues and specimen	Fresh, silica-gel dried plant tissues	Fresh, silica-gel dried plant tissues and specimen
Demand for DNA template quality	Low	Medium	Low
Applicable to specimen	Yes	No	Yes
Material for sequencing	Total genomic DNA	Restriction fragments	Captured loci using probes
Genome data obtained	Complete plastid genome, nrDNA, partial mitochondrial genome, coding and non-coding genes	Loci with single nucleotide polymorphism (SNP) mainly from nuclear genome; coding and non-coding genes	Targeted nuclear, plastid and/or mitochondrial loci; coding and non-coding genes
Targeted loci sequenced	Yes	No	Yes
Identification of orthologs	Easy	Difficult	Easy
Missing data among species	No	Yes	No
Taxonomic levels for phylogenetic relationships	All levels from shallow to deep	Shallow levels, below inter-generic	All levels from shallow to deep, above intraspecific

Yu *et al.* Plant Diversity 40 (2018)

Comparação dos métodos

Genome skimming

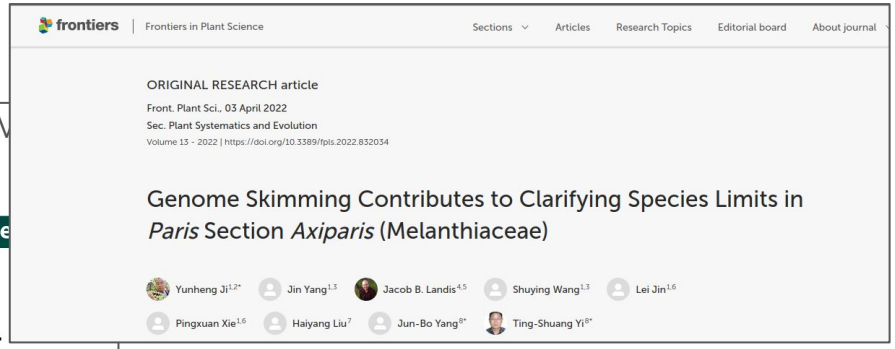

Nevill *et al. Plant Methods* (2020) 16:1
https://doi.org/10.1186/s13007-019-0534-5

Plant M

METHODOLOGY Ope

Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics

Paul G. Nevill[1,2,3*], Xiao Zhong[4,5], Julian Tonti-Filippini[4,5], Margaret Byrne[2,6,7], Michael Hislop[6], Kevin Thiele[2,6], Stephen van Leeuwen[6], Laura M. Boykin[4,5] and Ian Small[4,5]


frontiers | Frontiers in Plant Science

Sections Articles Research Topics Editorial board About journal

ORIGINAL RESEARCH article

Front. Plant Sci., 03 April 2022

Sec. Plant Systematics and Evolution

Volume 13 - 2022 | https://doi.org/10.3389/fpls.2022.832034

Genome Skimming Contributes to Clarifying Species Limits in *Paris* Section *Axiparis* (Melanthiaceae)

Yunheng Ji[1,2*] Jin Yang[1,3] Jacob B. Landis[4,5] Shuying Wang[1,3] Lei Jin[1,6]

Pingxuan Xie[1,6] Haiyang Liu[7] Jun-Bo Yang[8*] Ting-Shuang Yi[8*]

"… we employed a genome skimming approach to recover the plastid genomes (plastomes) and nuclear ribosomal DNA (nrDNA)"

Chloroplast genome - 96.1% of samples
Complete or near-complete nuclear ribosomal RNA gene - 93.3% of samples


KeAi CHINESE ROOTS GLOBAL IMPACT

Plant Diversity

Volume 46, Issue 3, May 2024, Pages 344-352

Research paper

Deep genome skimming reveals the hybrid origin of *Pseudosasa gracilis* (Poaceae: Bambusoideae)

Xiang-Zhou Hu[a b 1], Cen Guo[a 1], Sheng-Yuan Qin[a b], De-Zhu Li[a b], Zhen-Hua Guo[a b]


A pipeline for assembling low copy nuclear markers from plant genome skimming data for phylogenetic use

Marcelo Reginato

PeerJ

"By integrating … chloroplast genome data … and subgenome data … together with morphological and geographic information, we provide an empirical example of bamboo hybridization

Comparação dos métodos

RAD-seq

JOURNAL ARTICLE

Integrating restriction site-associated DNA sequencing (RAD-seq) with morphological cladistic analysis clarifies evolutionary relationships among major species groups of bee orchids

Richard M Bateman, Gábor Sramkó, Ovidiu Paun

Annals of Botany, Volume 121, Issue 1, January 2018, Pages 85–105, https://doi.org/10.1093/aob/mcx129

Published: 09 January 2018 **Article history**

scientific reports

Explore content · About the journal · Publish with us

nature > scientific reports > articles > article

Article | Open access | Published: 23 June 2022

Genetic diversity of loquat (*Eriobotrya japonica*) revealed using RAD-Seq SNP markers

Yukio Nagano, Hiroaki Tashiro, Sayoko Nishi, Naofumi Hiehata, Atsushi J. Nagano & Shinji Fukuda

Scientific Reports 12, Article number: 10200 (2022) | Cite this article

scientific reports

Explore content · About the journal · Publish with us

nature > scientific reports > articles > article

Article | Open access | Published: 23 March 2015

Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method

Ling Zhou, Shi-Bo Wang, Jianbo Jian, Qing-Chun Geng, Jia Wen, Qijian Song, Zhenzhen Wu, Guang-Jun Li, Yu-Qin Liu, Jim M. Dunwell, Jin Zhang, Jian-Ying Feng, Yuan Niu, Li Zhang, Wen-Long Ren & Yuan-Ming Zhang

Systematic Entomology

Royal Entomological Society

Original Article

RAD-seq phylogenomics recovers a well-resolved phylogeny of a rapid radiation of mutualistic and antagonistic yucca moths

CLIVE T. DARWELL, DAVID M. RIVERS, DAVID M. ALTHOFF

First published: 02 May 2016 | https://doi.org/10.1111/syen.12185 | Citations: 17

ELSEVIER

Molecular Phylogenetics and Evolution

Volume 126, September 2018, Pages 1-16

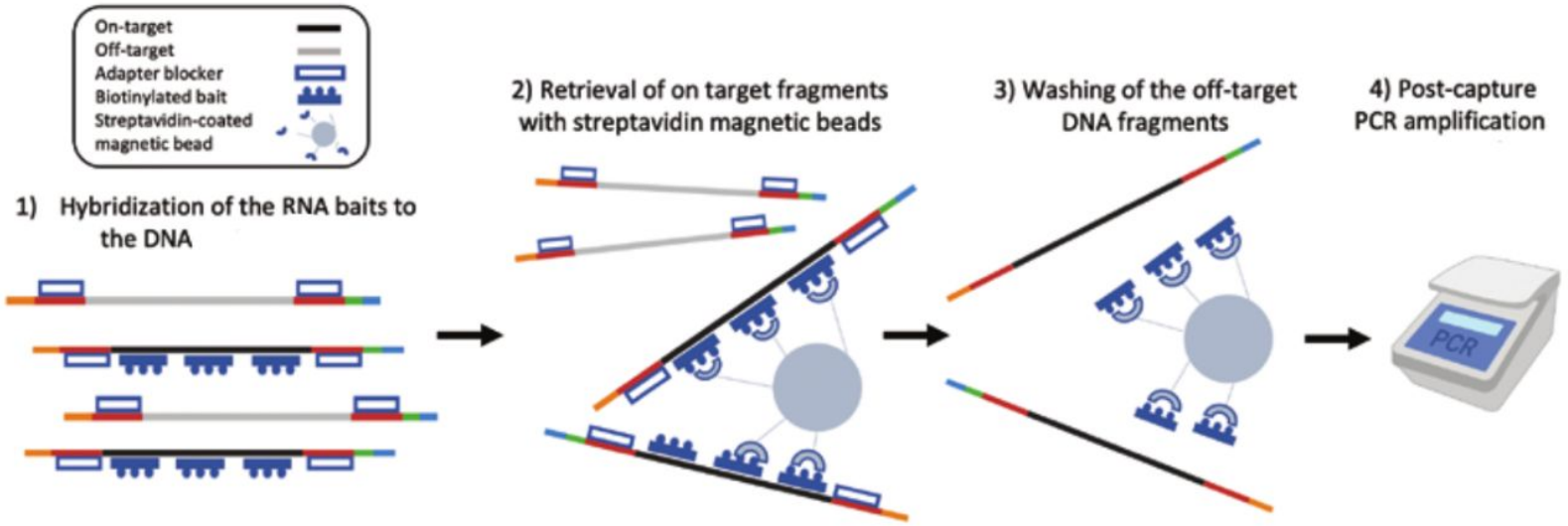
Resolving relationships and phylogeographic history of the *Nyssa sylvatica* complex using data from RAD-seq and species distribution modeling

Wenbin Zhou [a], Xiang Ji [b c], Shihori Obata [a], Andrew Pais [a], Yibo Dong [a], Robert Peet [d], Qiu-Yun (Jenny) Xiang [a]

O que são as probes?

Um fragmento de oligonucleotídeos usualmente de RNA de 80-120 bp, desenhados para capturar fragmentos alvo numa biblioteca de DNA genômico.

(E) Captura dos fragmentos de interesse com RNA *baits* mediante hibridização (enriquecimento das bibliotecas)



O que são as probes?

- Geralmente capturam regiões de baixa cópia do genoma nuclear
- Podem ser universais ou específicas

Angiosperms353
(Johnson et al. 2019)

CompCOS
(Mandel et al. 2017)

Ochnaceae
(Schneider et al. 2021)

Probes universais vs. específicas

	Universais	Específicas
Poder de resolução a nível infraespecífico	Baixo*	Alto
Especificidade das probes	Baixa	Alta
Custo do desenho das probes	Baixo	Alto
Compatibilidade entre diferentes estudos	Alta	Baixa*

Probes universais vs. específicas

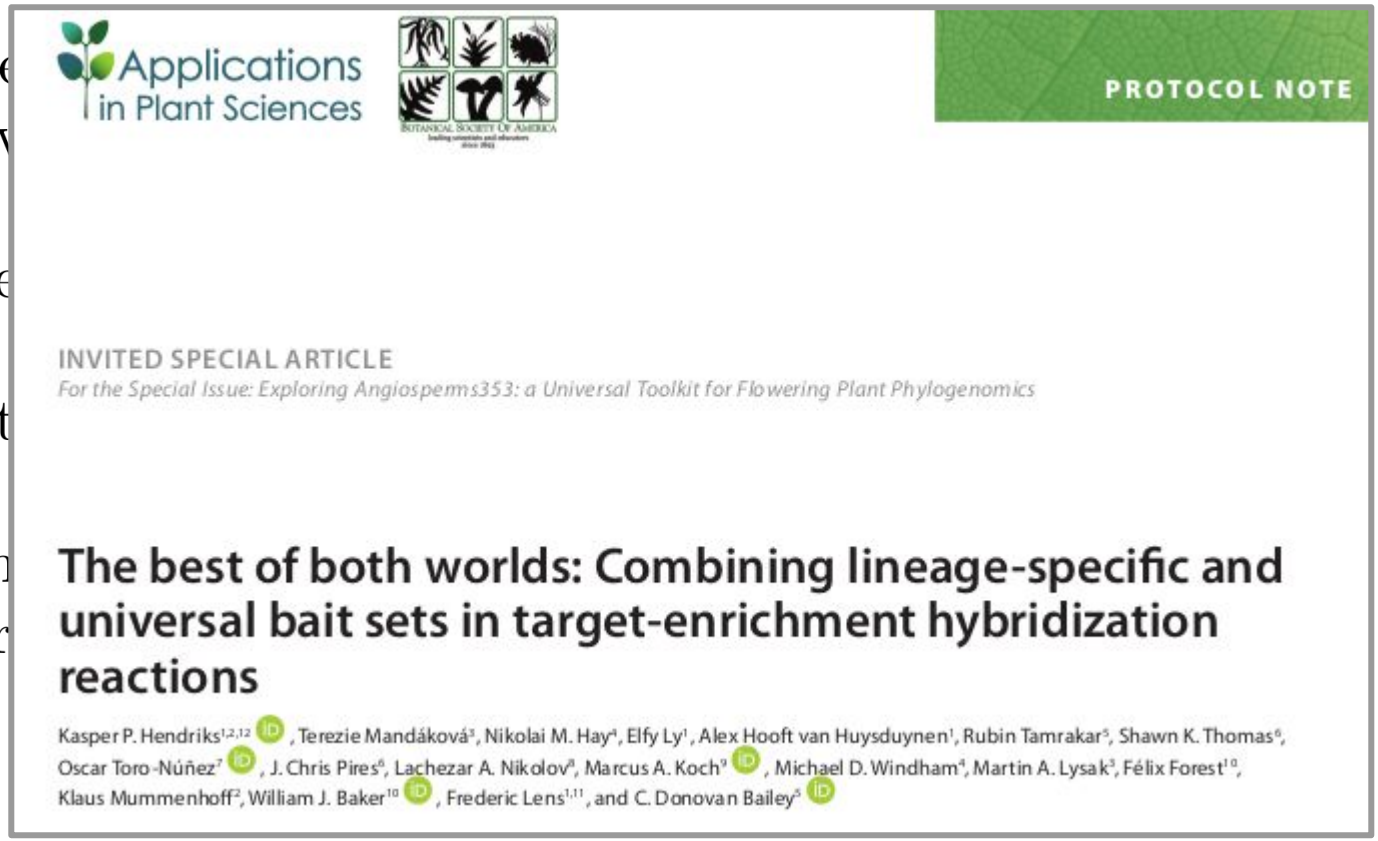
Universais	Específicas

Pode
a nív

Espe

Cust

Com
difer



Applications in Plant Sciences

PROTOCOL NOTE

INVITED SPECIAL ARTICLE

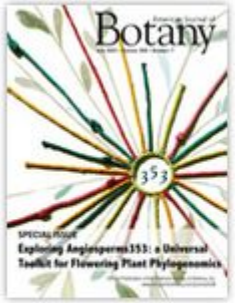
For the Special Issue: Exploring Angiosperms353: a Universal Toolkit for Flowering Plant Phylogenomics

The best of both worlds: Combining lineage-specific and universal bait sets in target-enrichment hybridization reactions

Kasper P. Hendriks[1,2,12], Terezie Mandáková[3], Nikolai M. Hay[4], Elfy Ly[1], Alex Hooft van Huysduynen[1], Rubin Tamrakar[5], Shawn K. Thomas[6], Oscar Toro-Núñez[7], J. Chris Pires[6], Lachezar A. Nikolov[8], Marcus A. Koch[9], Michael D. Windham[4], Martin A. Lysak[3], Félix Forest[10], Klaus Mummenhoff[2], William J. Baker[10], Frederic Lens[1,11], and C. Donovan Bailey[5]

Comparação dos métodos

Target capture



American Journal of Botany: Volume 108, Issue 7

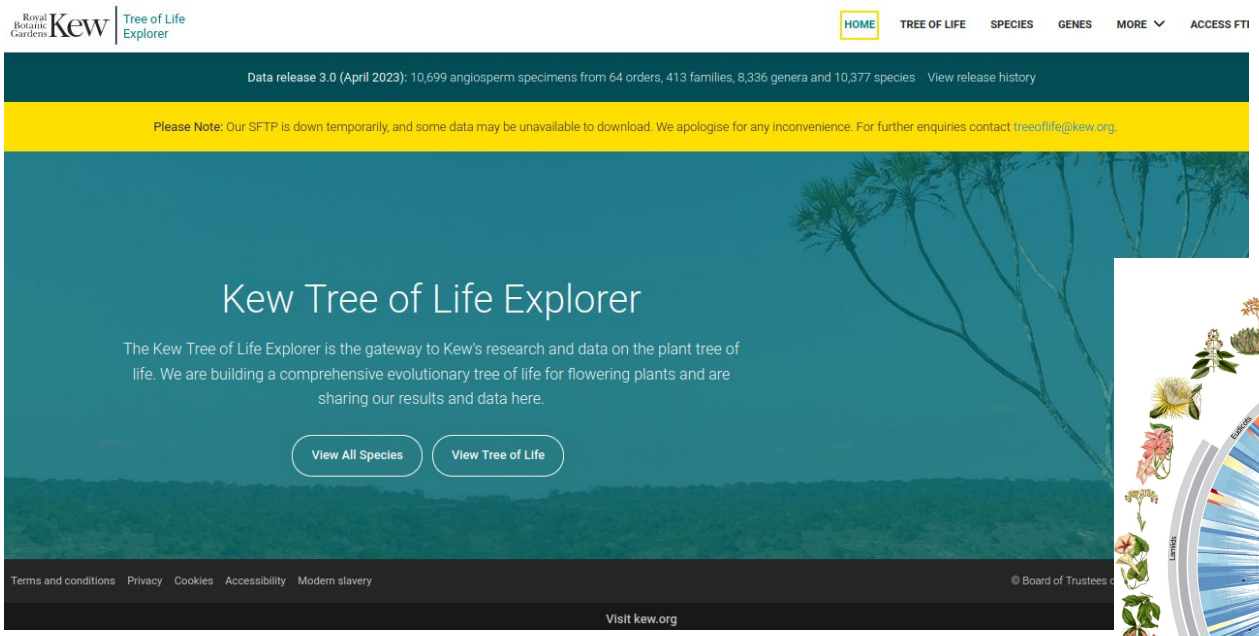
Special Issue: Exploring Angiosperms353: a Universal Toolkit for Flowering Plant Phylogenomics

Pages: i-v, 1059-1306

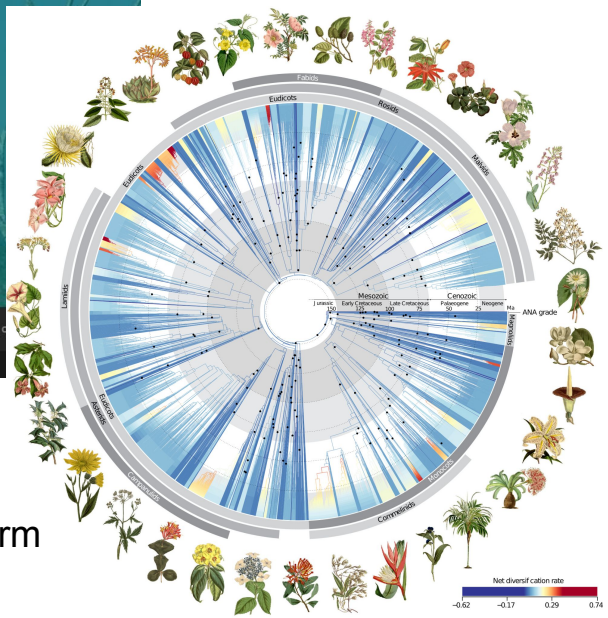
July 2021

Comparação dos métodos

Target capture - Angiosperms353



https://treeoflife.kew.org/

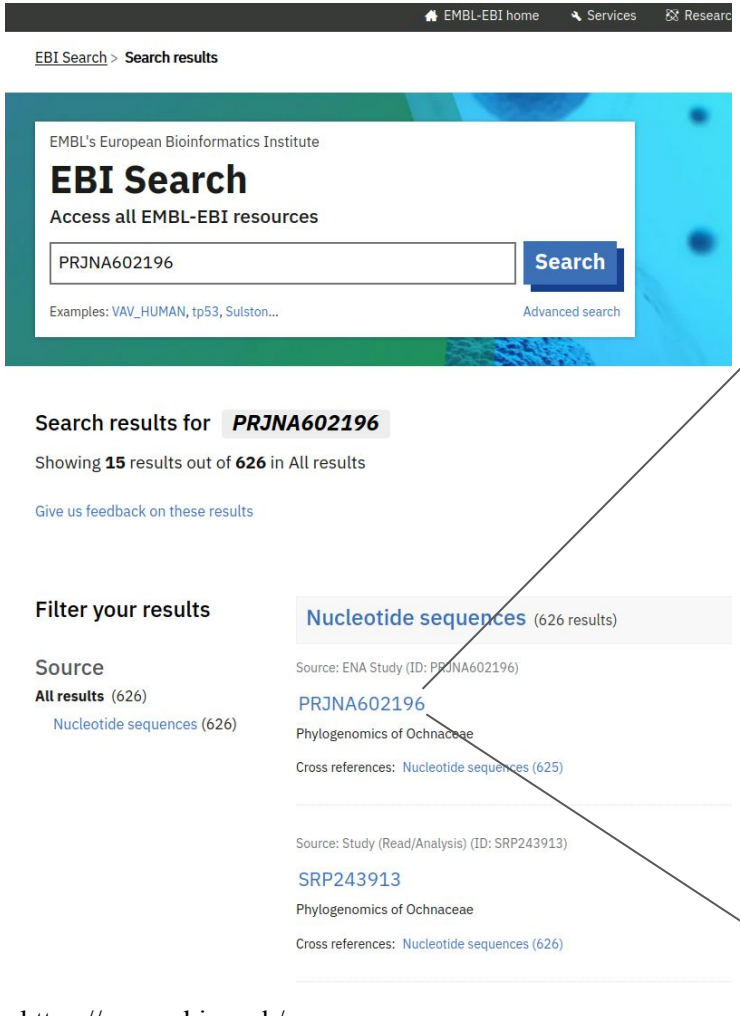


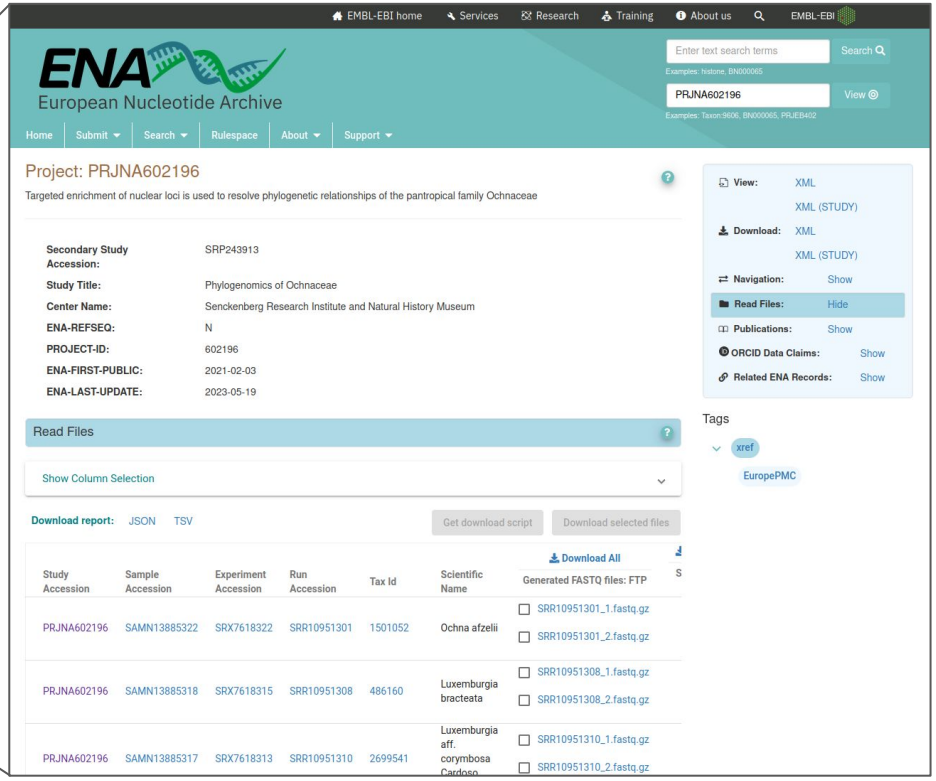
8,000 (about 60%) angiosperm genera

Zuntini *et al.* Nature 629. (2024)

Comparação dos métodos

Target capture - dados disponíveis em bases de dados





https://www.ebi.ac.uk/

Target capture - *Angiosperms353* - Hybpiper

Target capture - *Angiosperms353* - Hybpiper

Fluxo de trabalho após sequenciamento Illumina

Controle qualidade

Montagem das reads

Filtragem das sequências

Inferência filogenética

SRR10951301_1.fastq.gz

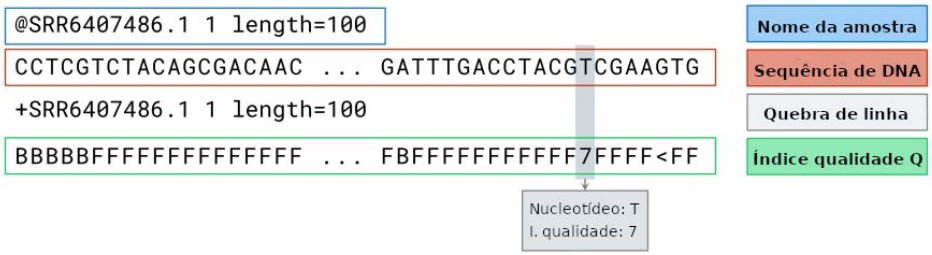
SRR10951301_2.fastq.gz

Target capture - Hybpiper

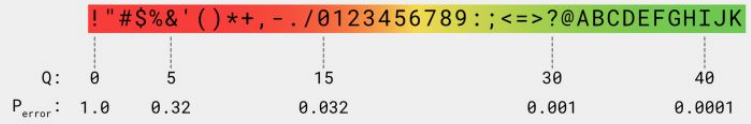
Formato FASTQ

Arquivo FASTQ:

```
@SRR6407486.1 1 length=100
CCTCGTCTACAGCGACAACGTCCAGACCCGCGAACGGGTGATGCGGGCCCTGGGCAAACGGTTGCACCCGGATCTGCCCGATTTGACCTACGTCGAAGTG
+SRR6407486.1 1 length=100
BBBBBFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF<FFFFFFFFFFFFFFFFFBFFFFFFFFFFFFFFFFFFFFBFFFFFFFFFFF7FFFF<FF
```



Índice de qualidade Q associado a cada um dos caracteres do ASCII:



$$Q = -10\log_{10} P_{error}$$

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%

SRR10951301_1.fastq.gz

SRR10951301_2.fastq.gz

Target capture - Hybpiper

Controle de qualidade das reads

MultiQC

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

General Statistics

Copy table | Configure Columns | Plot | Showing 192/192 rows and 3/6 columns.

Sample Name	% Dups	% GC	M Seqs
SPRL_184_2	82.0%	41%	2.3
SPRL_186_1	68.9%	41%	7.0
SPRL_186_2	64.3%	41%	7.0
SPRL_187_1	80.5%	44%	6.0
SPRL_187_2	77.7%	45%	6.0
SPRL_190_1	72.5%	42%	14.9
SPRL_190_2	66.4%	42%	14.9
SPRL_192_1	76.1%	42%	11.4
SPRL_192_2	71.2%	42%	11.4

SRR10951301_1.fastq.gz

SRR10951301_2.fastq.gz

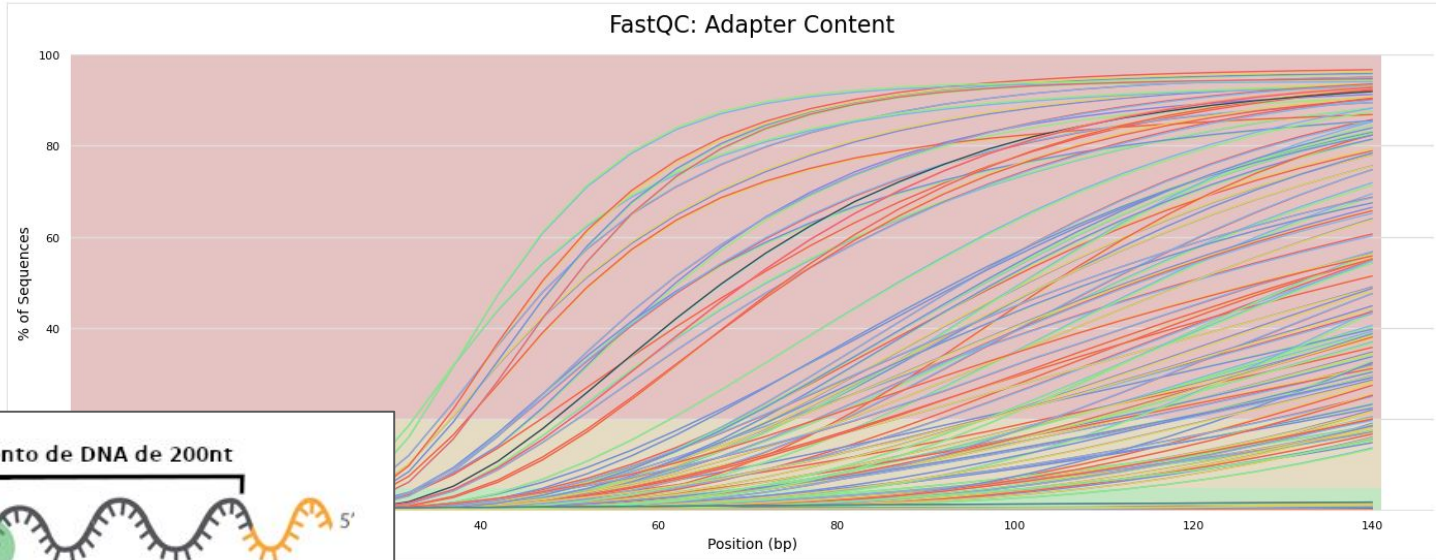
Target capture - Hybpiper

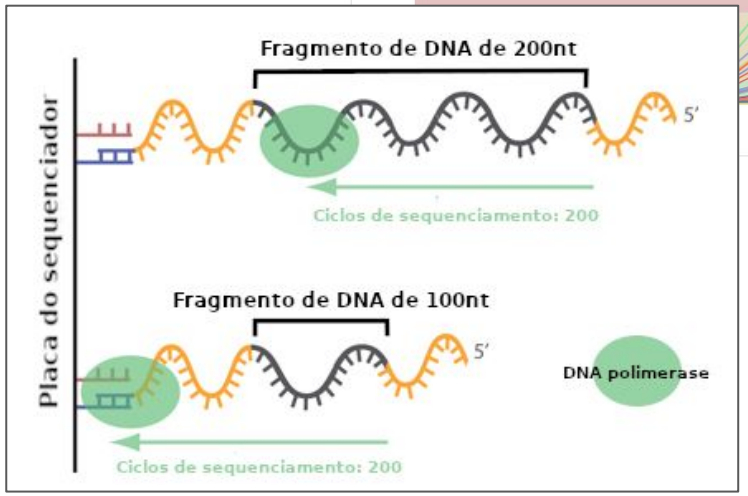
Controle de qualidade das reads

Adapter Content 192

The cumulative percentage count of the proportion of your library which has seen each of the adapter sequences at each position.

Flat image plot. Toolbox functions such as highlighting / hiding samples will not work (see the docs).

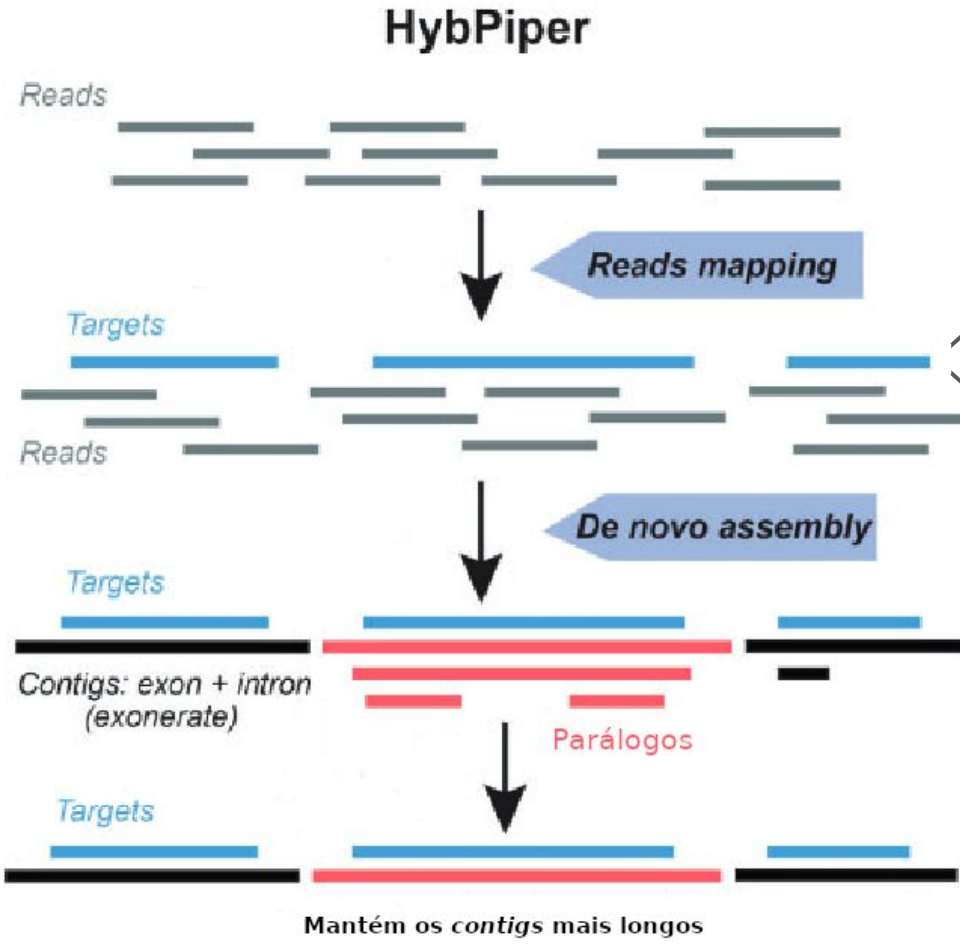


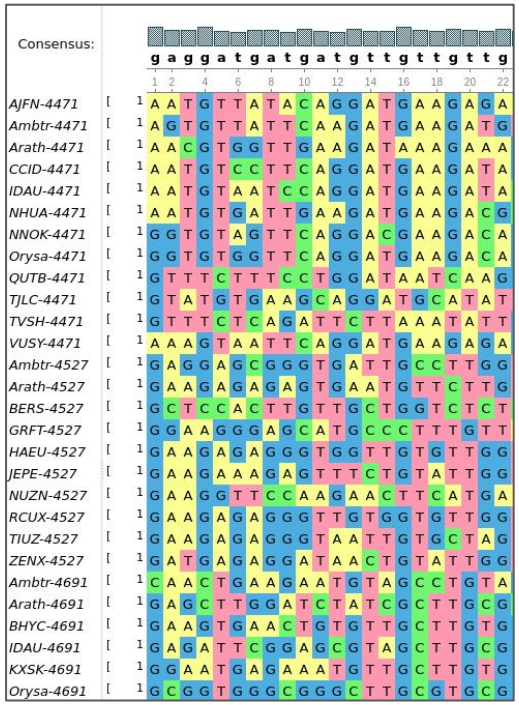


Trimmomatic Manual: V0.32

Target capture - Hybpiper

Montagem das sequencias



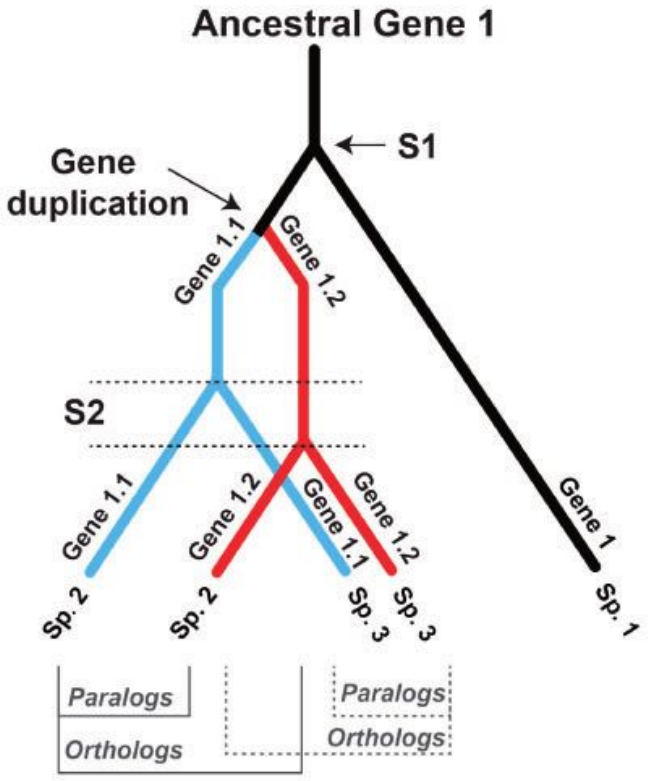


Target file

Target capture - Hybpiper

Filtragem das sequências - Parálogos

- Alelos
- Verdadeiros parálogos
- Contaminação



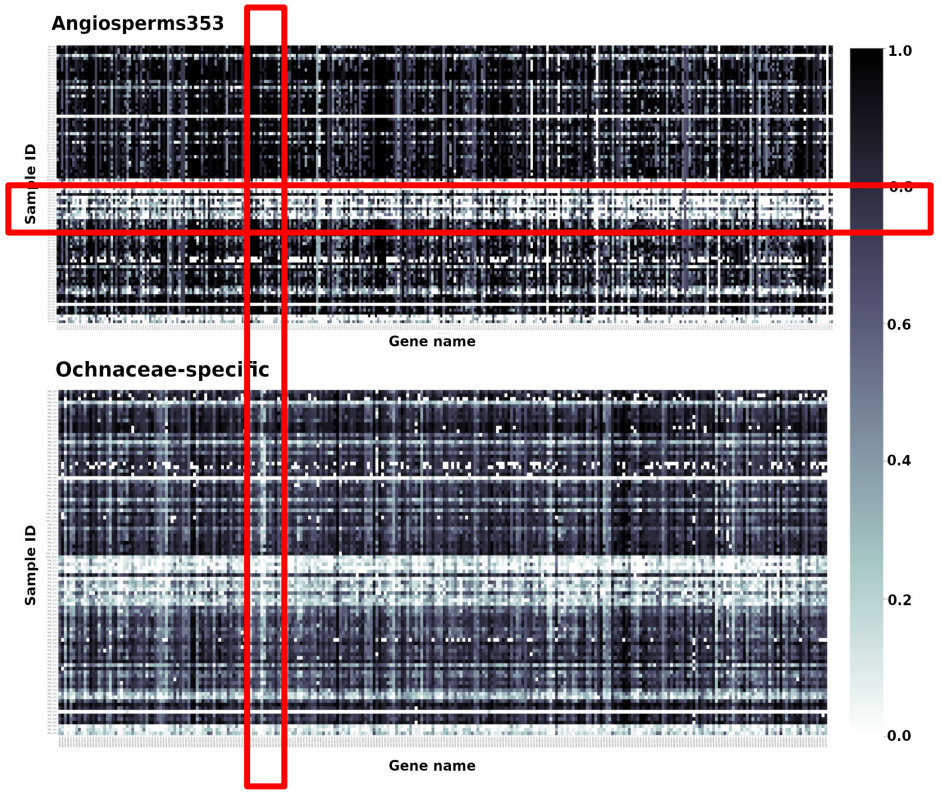
Zhou *et al.* Sys. Bio. 71: 410–425 (2022)

Target capture - Hybpiper

Filtragem das sequências - Dados faltantes

Baixa recuperação:

- Regiões/loci com poucas amostras
- Amostras com sequências para poucos loci
- Sequências muito curtas



Inferência filogenética

Material para outro curso

- Sumarização de árvores de genes
- Concatenação
- Redes filogenéticas

Accurate Species Tree EstimatoR (ASTER⁎)

A family of optimatization algorithms for species tree inference:

1. ASTRAL-IV (from unrooted gene tree ***topologies*** with integrated CASTLES-II for branch lengths)
2. ASTRAL-Pro3 (from unrooted gene ***family*** tree topologies with integrated CASTLES-Pro)
3. Weighted ASTRAL (from unrooted gene trees with branch ***lengths*** and/or ***supports***)
4. CASTER-site (from ***whole genome alignments*** or aligned sequences)
5. CASTER-pair (from ***whole genome alignments*** or aligned sequences)
6. WASTER-site (from ***raw reads***)
7. SISTER (from optical-map-like distance data or shape data)
8. MONSTER

Considerações finais

- NGS ou HTS tem o melhor custo benefício para estudos em sistemática hoje.
- Target sequence é uma ótima opção para inferência filogenética a diferentes níveis, funciona muito bem em material de herbário (e.g. tipos).
- Mais dados não representam maior resolução mas sim a possibilidade de explorar fenômenos de evolução molecular interessantes (e.g. hibridização, duplicação genica).
- Estratégias de limpeza e filtragem dos dados são fundamentais para uma ‘boa’ reconstrução filogenética.

Para saber mais:

Curso completo de Introdução a Next Generation Sequencing (NGS).

Profa. Dr. Carolina M. Siniscalchi
Mississippi State University

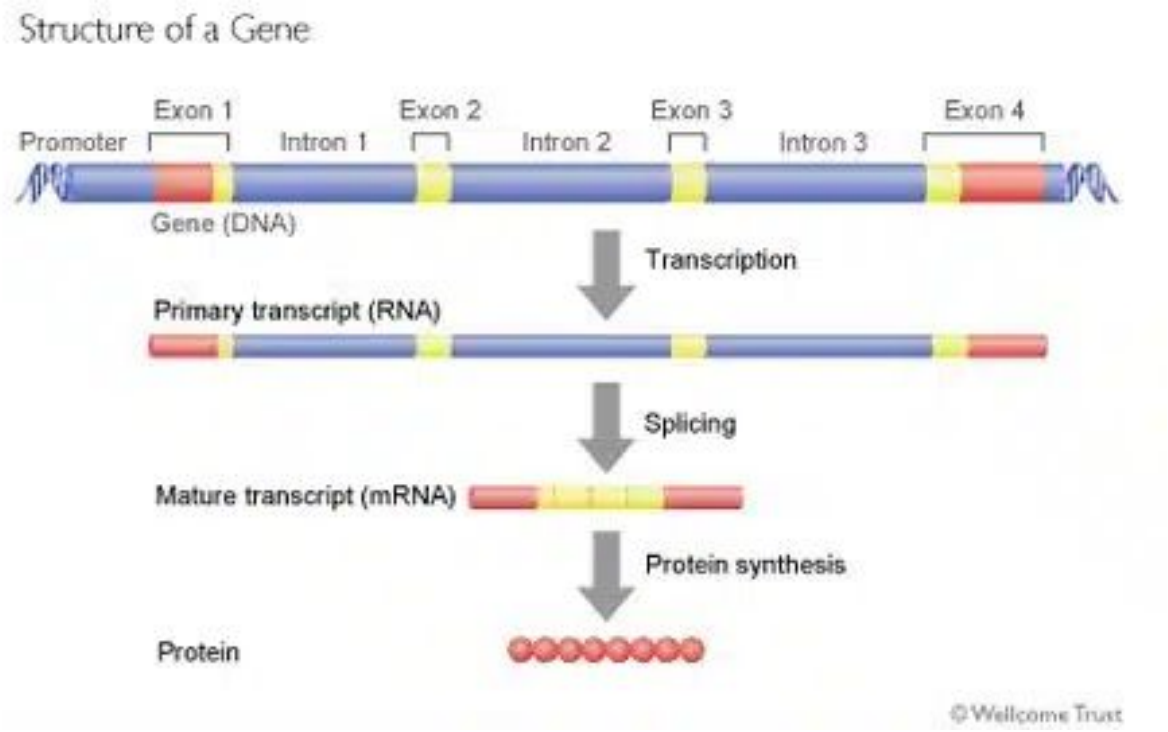
Slides, lista com materiais de apoio:

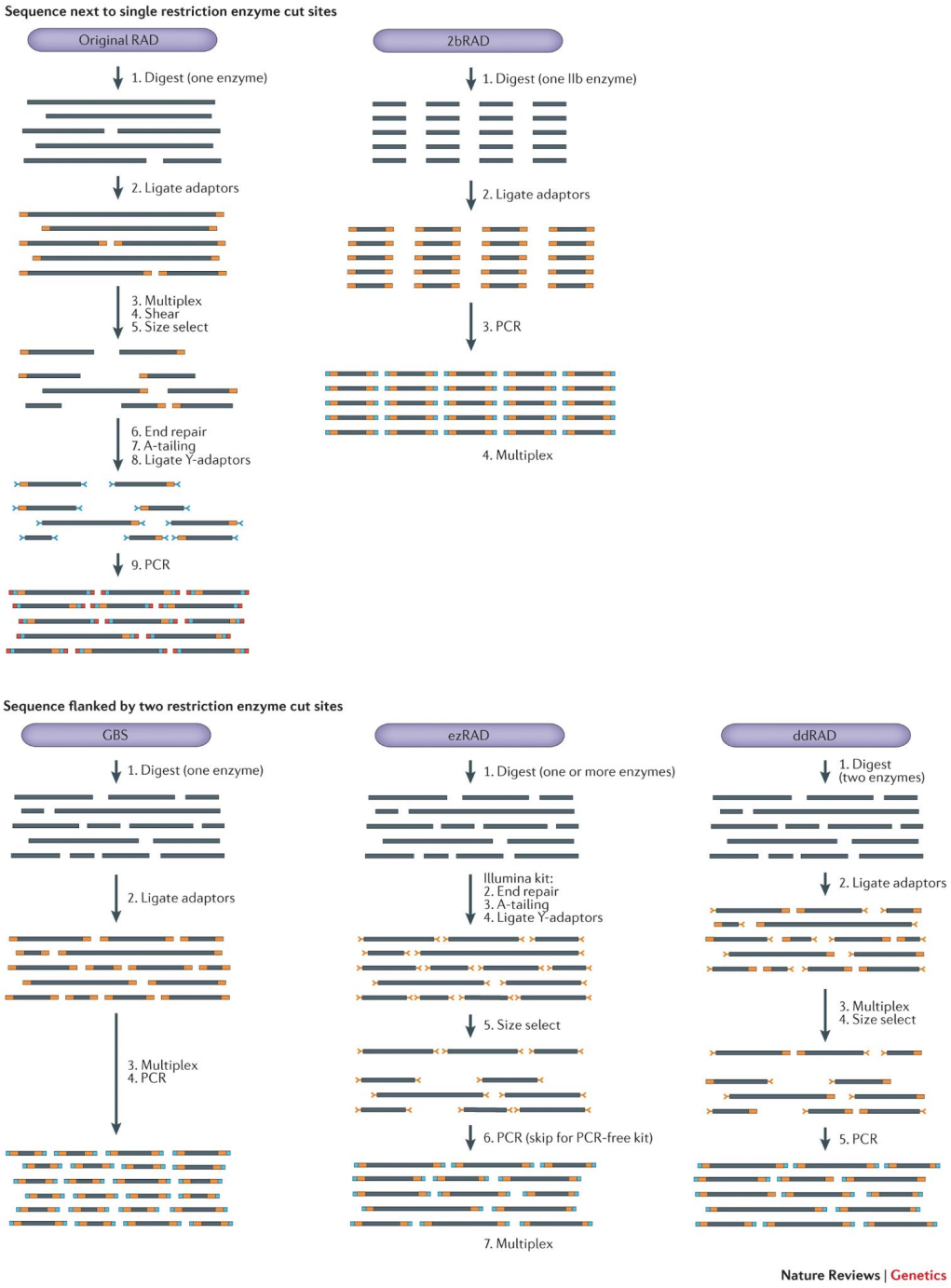
- Bash, R e Python para iniciantes
- Tutoriais
- Livros

Dúvidas?

Target capture - Hybpiper

Montagem das sequencias





All protocols begin by digesting high-molecular-weight genomic DNA with one or more restriction enzymes. For most protocols, the sequencing adaptors (oligonucleotides) are added in two stages, with one set of oligonucleotides added during a ligation step early in the protocol, and a second set of oligonucleotides incorporated during a final PCR step. The second set of oligonucleotides extends the length of the total fragment to produce the entire Illumina adaptor sequences. By contrast, the original RADseq adds adaptors in three stages. For Illumina sequencing, the adaptors on either end of each DNA fragment must differ, and therefore some protocols (for example, original RADseq, double digest RAD (ddRAD) and ezRAD) use Y-adaptors that are structured to ensure that only fragments with different adaptors at either end are PCR-amplified (illustrated here as Y-shaped adaptors). Other protocols (for example, genotyping by sequencing (GBS)) simply rely on the fact that fragments without the correct adaptors will not be sequenced. To generate fragments of an ideal length for sequencing, most methods use common-cutter enzymes (for example, 4–6 bp cutters) to generate a wide range of fragment sizes, followed by a direct size selection (gel-cutting or magnetic beads, for example, ezRAD and ddRAD) or an indirect size selection (as a consequence of PCR amplification or sequencing efficiency, for example, GBS).

https://www.nature.com/articles/nrg.2015.28/figures/1

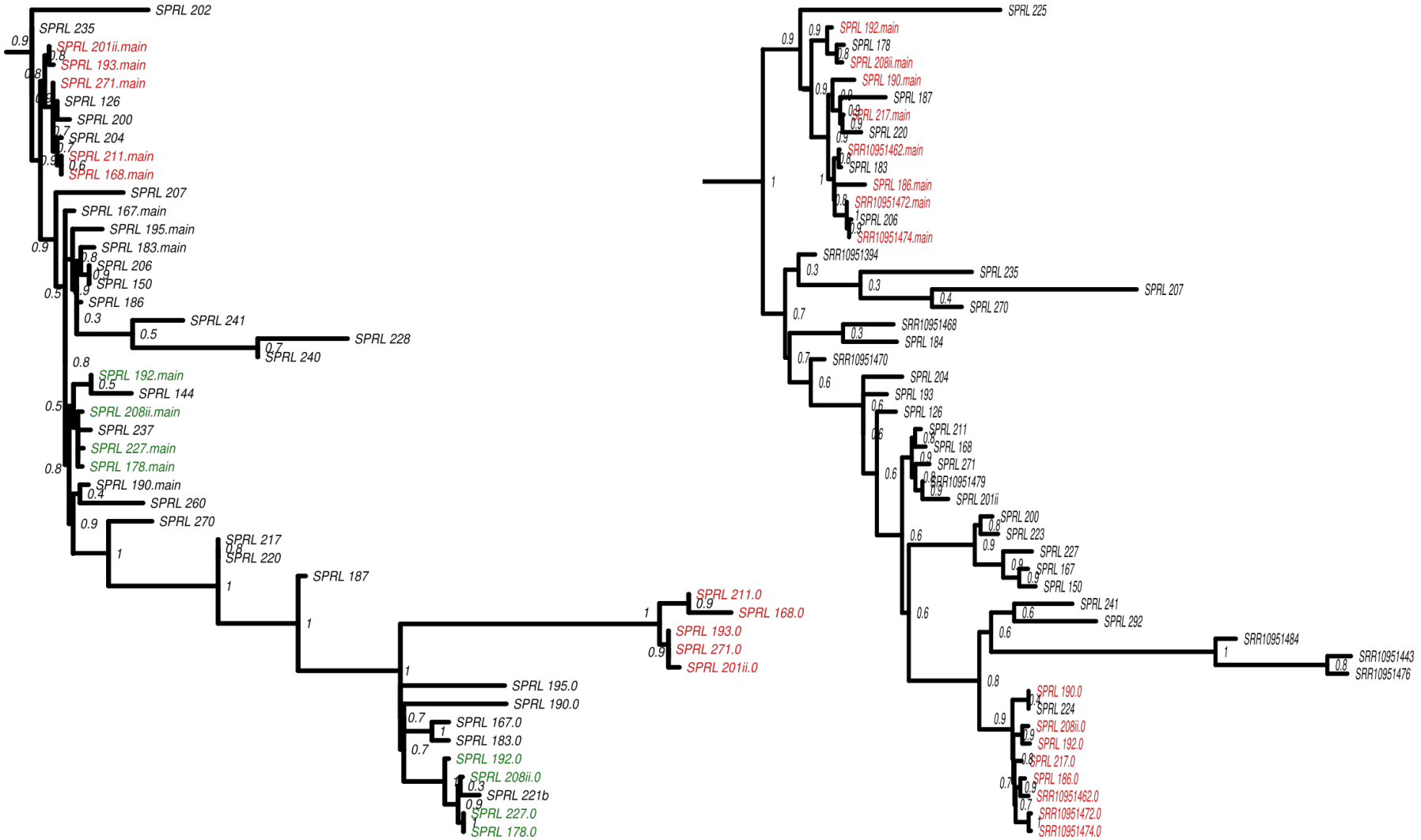
An example of allele dropout for a restriction site-associated DNA sequencing (RADseq) protocol that uses size selection to reduce the number of loci to be sequenced. Grey lines represent chromosomes within one individual, red squares represent restriction cut sites, coloured squares represent heterozygous SNPs, and square brackets represent genomic regions that are sequenced. Mutation in cut site B for haplotype 1 makes the post-digestion fragment containing the SNP too long to be retained during size selection for haplotype 1, eliminating the possibility of sequencing of any loci on that fragment, and causing the individual to appear homozygous at the heterozygous SNP. **b** | An example of fragments produced after PCR for one heterozygous locus for different RADseq protocols, and the reads retained after bioinformatic analyses. PCR duplicates are shown with the same symbol (circle, square, asterisk or triangle) as the parent fragment from the original template DNA. By chance, some alleles will amplify more than others during PCR. For all protocols, PCR duplicates will be identical in sequence composition and length to the original template molecule. For the original RADseq, this feature (that is, identical length) can be used to identify and remove PCR duplicates bioinformatically, because original template molecules for a given locus will not be identical in length. For alternative RADseq methods, this feature cannot be used to identify PCR duplicates, because all original template molecules for a given locus are identical in length. High frequencies of PCR duplicates can cause heterozygotes to appear as homozygotes or can cause PCR errors to appear as true diversity. Part **b** is adapted with permission from Ref. 37, Wiley.

https://www.nature.com/articles/nrg.2015.28

A


SPRL_187 multi_hit_stitched_contig_comprising_3_
SPRL_168.main NODE_1_length_2704_cov_30
SPRL_168.0 NODE_2_length_2204_cov_71.248
SPRL_190.0 NODE_2_length_2408_cov_449.02
SPRL_190.main NODE_1_length_2824_cov_54
SPRL_201ii.main NODE_1_length_2542_cov_1
SPRL_201ii.0 NODE_2_length_2054_cov_49.41

B


SPRL 202
SPRL 235
SPRL 201ii.main
SPRL 193.main
SPRL 271.main
SPRL 126
SPRL 200
SPRL 204
SPRL 211.main
SPRL 168.main
SPRL 207
SPRL 167.main
SPRL 195.main
SPRL 183.main
SPRL 206
SPRL 150
SPRL 186
SPRL 241
SPRL 228
SPRL 240
SPRL 192.main
SPRL 144
SPRL 208ii.main
SPRL 237
SPRL 227.main
SPRL 178.main
SPRL 190.main
SPRL 260
SPRL 270
SPRL 217
SPRL 220
SPRL 187
SPRL 211.0
SPRL 168.0
SPRL 193.0
SPRL 271.0
SPRL 201ii.0
SPRL 195.0
SPRL 190.0
SPRL 167.0
SPRL 183.0
SPRL 192.0
SPRL 208ii.0
SPRL 221b
SPRL 227.0
SPRL 178.0
SPRL 225
SPRL 192.main
SPRL 178
SPRL 208ii.main
SPRL 190.main
SPRL 187
SPRL 217.main
SPRL 220
SRR10951462.main
SPRL 183
SPRL 186.main
SRR10951472.main
SPRL 206
SRR10951474.main
SRR10951394
SPRL 235
SPRL 207
SPRL 270
SRR10951468
SPRL 184
SRR10951470
SPRL 204
SPRL 193
SPRL 126
SPRL 211
SPRL 168
SPRL 271
SRR10951479
SPRL 201ii
SPRL 200
SPRL 223
SPRL 227
SPRL 167
SPRL 150
SPRL 241
SPRL 292
SRR10951484
SRR10951443
SRR10951476
SPRL 190.0
SPRL 224
SPRL 208ii.0
SPRL 192.0
SPRL 217.0
SPRL 186.0
SRR10951462.0
SRR10951472.0
SRR10951474.0

What is long-read sequencing?

Long-read sequencing is a DNA sequencing technique that enables the sequencing of much longer DNA fragments than traditional short-read sequencing methods. While short reads can capture the majority of genetic variation, long-read sequencing allows the detection of complex structural variants that may be difficult to detect with short reads. These include large inversions, deletions, or translocations, some of which have been implicated in areas like genetic disease.

Long-read technology can help resolve challenging regions of the genome by sequencing thousands of bases to:

- Resolve traditionally difficult to map genes or regions of the genome, such as those containing high variable or highly repetitive elements
- Perform phased sequencing to identify co-inherited alleles, haplotype information, and phase *de novo* mutations
- Generate long reads for *de novo* assembly and genome finishing applications

Técnicas de sequenciamento

Sequenciamento de alto rendimento - HTS (2005 - presente)

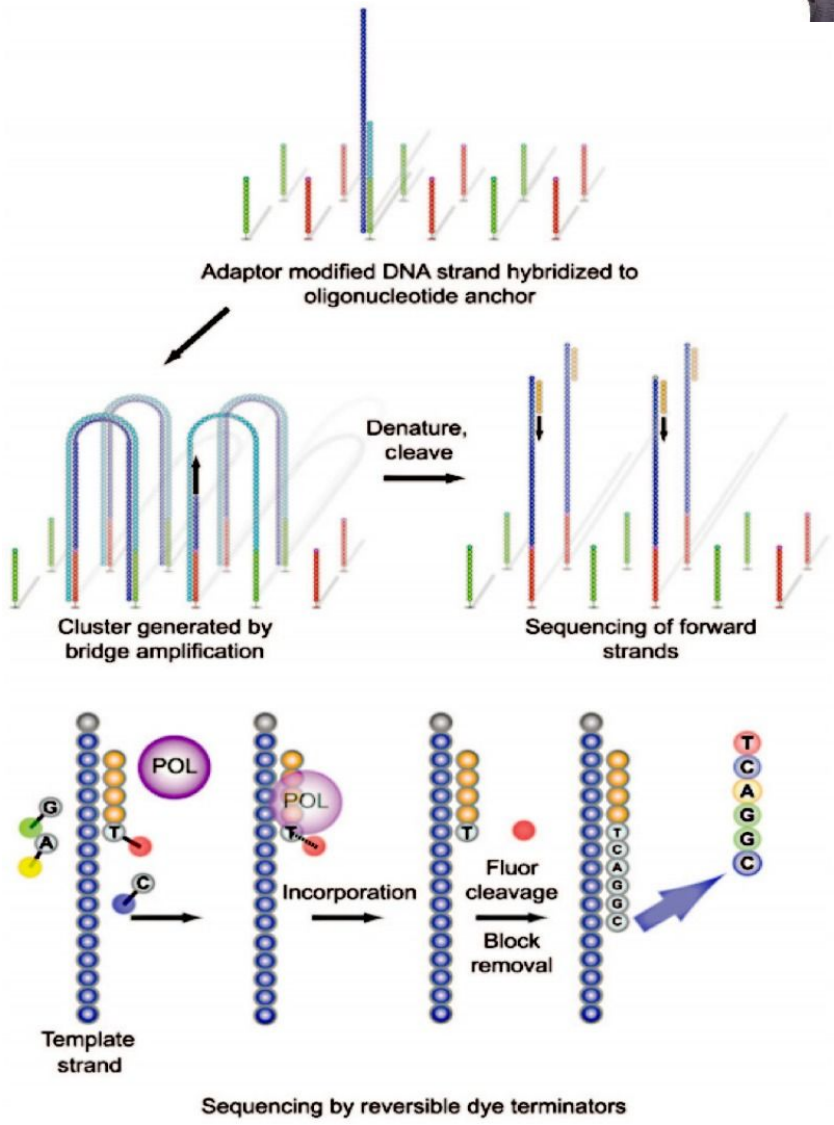
- Sequencia milhões de fragmentos simultaneamente (em paralelo) em cada corrida (Número dependendo da plataforma)
- Custo por nucleotídeo sequenciado é muito menor quando comparado com Sanger
- Precisa de preparação de bibliotecas genômicas

Técnicas de sequenciamento

Sequenciamento Illumina

Três etapas principais:

1. Ligação das bibliotecas na plataforma
2. PCR fase sólida (Bridge amplification)
3. Sequenciamento por síntese



https://www.youtube.com/watch?v=fCd6B5HRaZ8