

Análise de dados de amplicons (ITS-Fungos) a partir de sequenciamento em larga escala

Autores:

Doutorando Victor Borin Centurion Biruel
Doutoranda MsC. Kelly Hidalgo Martinez



IDENTIFICAÇÃO DE FUNGOS

- Principalmente baseada em características fenotípicas e fisiológicas.
- Só pessoal com alta experiencia é capaz de identificar corretamente um fungo baseado na morfologia
- Identificação passo a passo com chaves taxonômicas
- Foram desenvolvidos métodos moleculares como FISH, hibridação de DNA, DGGE, T-RFLP e sequenciamento de DNA.
- Atualmente o DNA Barcoding é considerado o mais acurado e rápido



DNA BARCODING

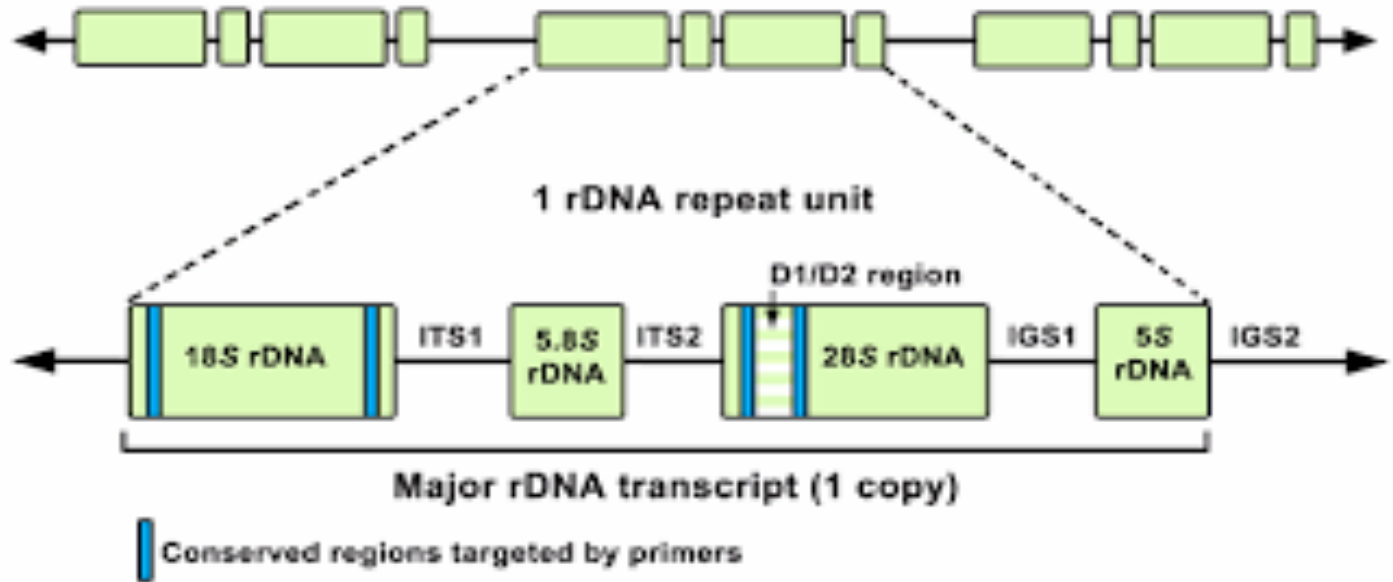
- Sequencia curta
- Deve ser suficientemente variável ao nível interespecífico (diferentes espécies), mas pouco variável ao nível intraespecífico (membros da mesma espécie)
- Padronizada
- Aproximadamente 700 bp
- Padrão único para identificar diferentes espécies
- É acurado, rápido, universalmente acessível
- Não precisa ser um experto taxonomista



ITS

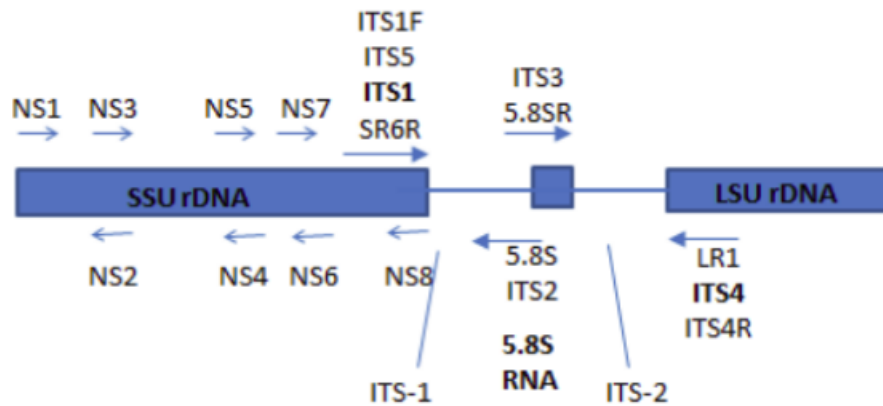
- Região intergênica do DNA ribossômico
- É a região mais sequenciada para identificar taxonomia de fungos
- É uma região não codificante
- Alto grau de variação comparado com as SSU e LSU
- Sequências disponíveis em bases de dados públicos tem aumentado

ITS



- Organizados em unidades repetitivas, múltiplas cópias ao longo do genoma
- Inicialmente são transcritos pela RNA pol I, mas depois sofrem edição e são eliminados (não codificantes)
- ITS em fungos aproximadamente 400 – 800 bp

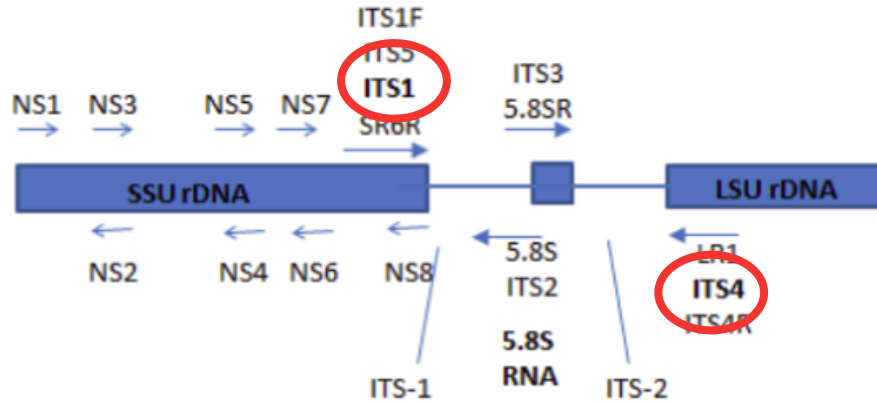
ITS



Primers

- Essenciais para o sucesso do barcode
- Devem ser tão universais para abranger um grande grupo de taxas, mas ao mesmo tempo tem que produzir aplicons que são suficientemente variados para distinguir entre espécies muito relacionadas
- Tem primers específicos para fungos

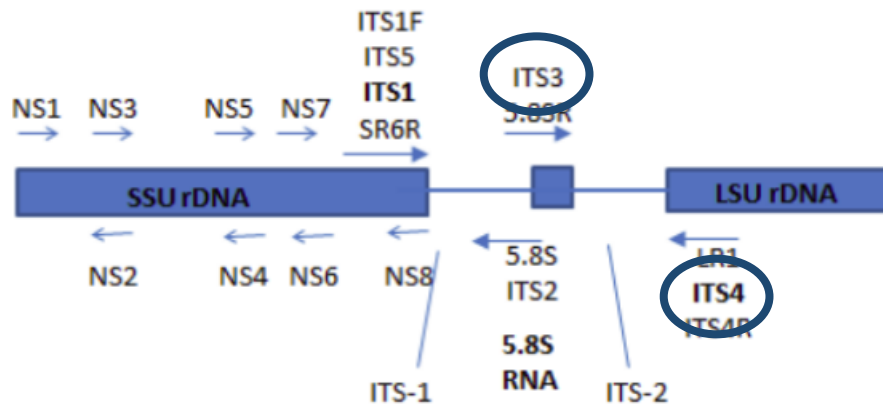
ITS



Primers

- Os mais usados
- Tem reportes de inespecificidad em amostras ambientais
- Melhor para Ascomycetes que para Basidiomycetes (ITS2)
- Resolve com primers só para a região ITS1 (Bellemain et al., 2010)
- Combinações de primers é sugerido.

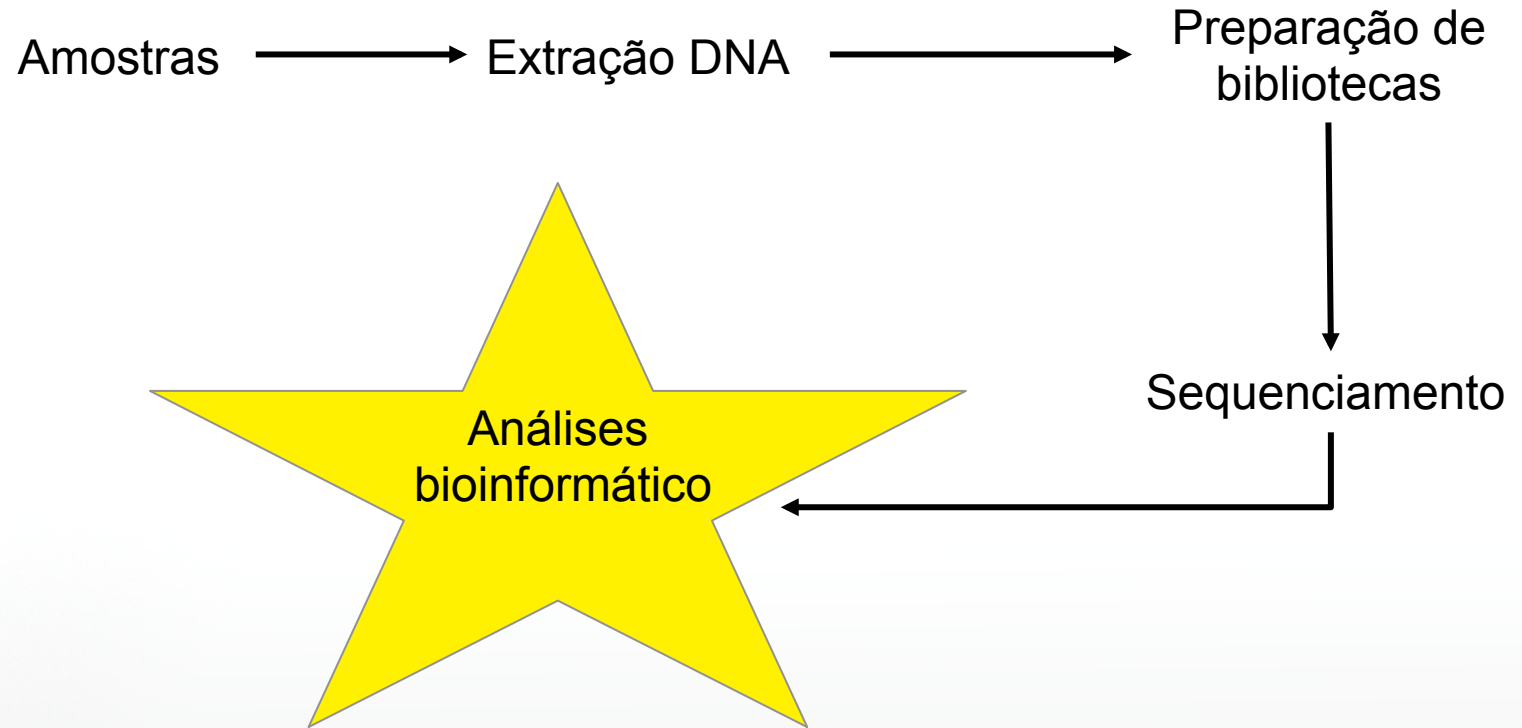
ITS



Primers

- Os mais usados
- Tem reportes de inespecificidad em amostras ambientais
- Melhor para Ascomycetes que para Basidiomycetes (ITS2)
- Resolve com primers só para a região ITS1 (Bellemain et al., 2010)
- Combinações de primers é sugerido.

Workflow



Pipeline bioinformático

Eliminação de
barcodes e primers

Controle de
qualidade

Filtragem



Afiliação taxonômica

Clusterização

Remoção de
chimeras, denoising

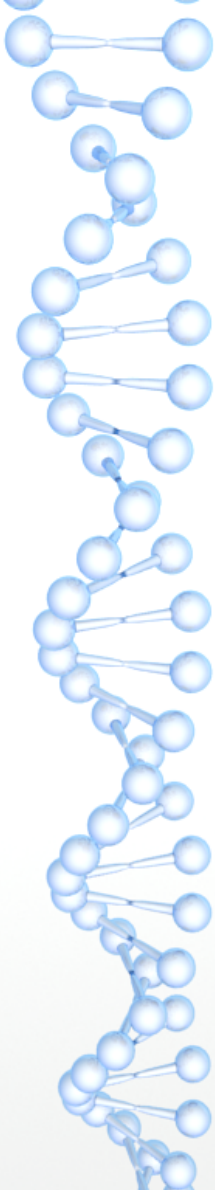
Análises

Alfa e beta diversidade

Composição da comunidade

Análises Ordenação (PCoA, PCA, NMDS)

Correlação com variáveis físico-químicas,
ambientais (CCA)



Pipeline bioinformático

Eliminação de
barcodes e primers



Sequências para diferenciar
entre amostras

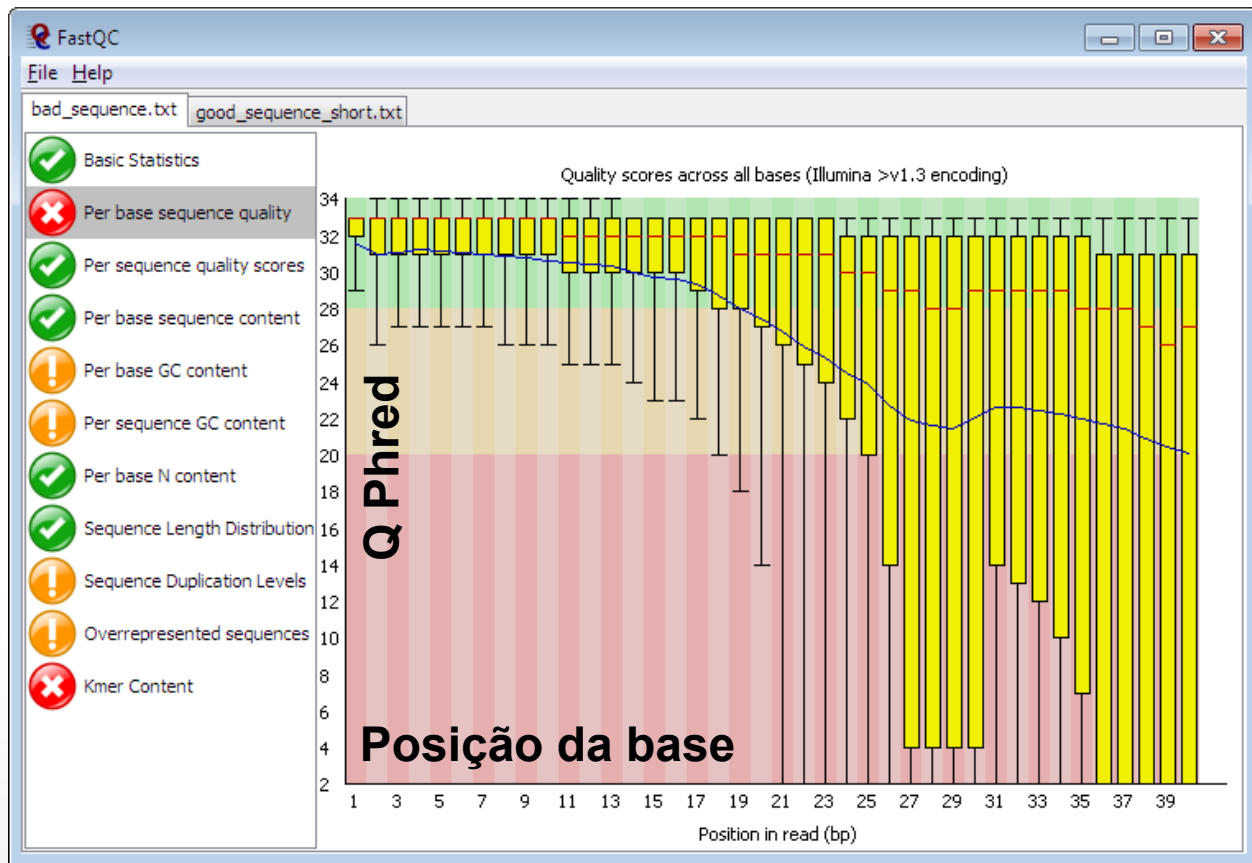


Pipeline bioinformático

Eliminação de
barcodes e primers



Controle de qualidade



Q phred: Está definido pela probabilidade que uma base este corretamente nomeada, baseado em uma série de preditores calibrados com dados reais

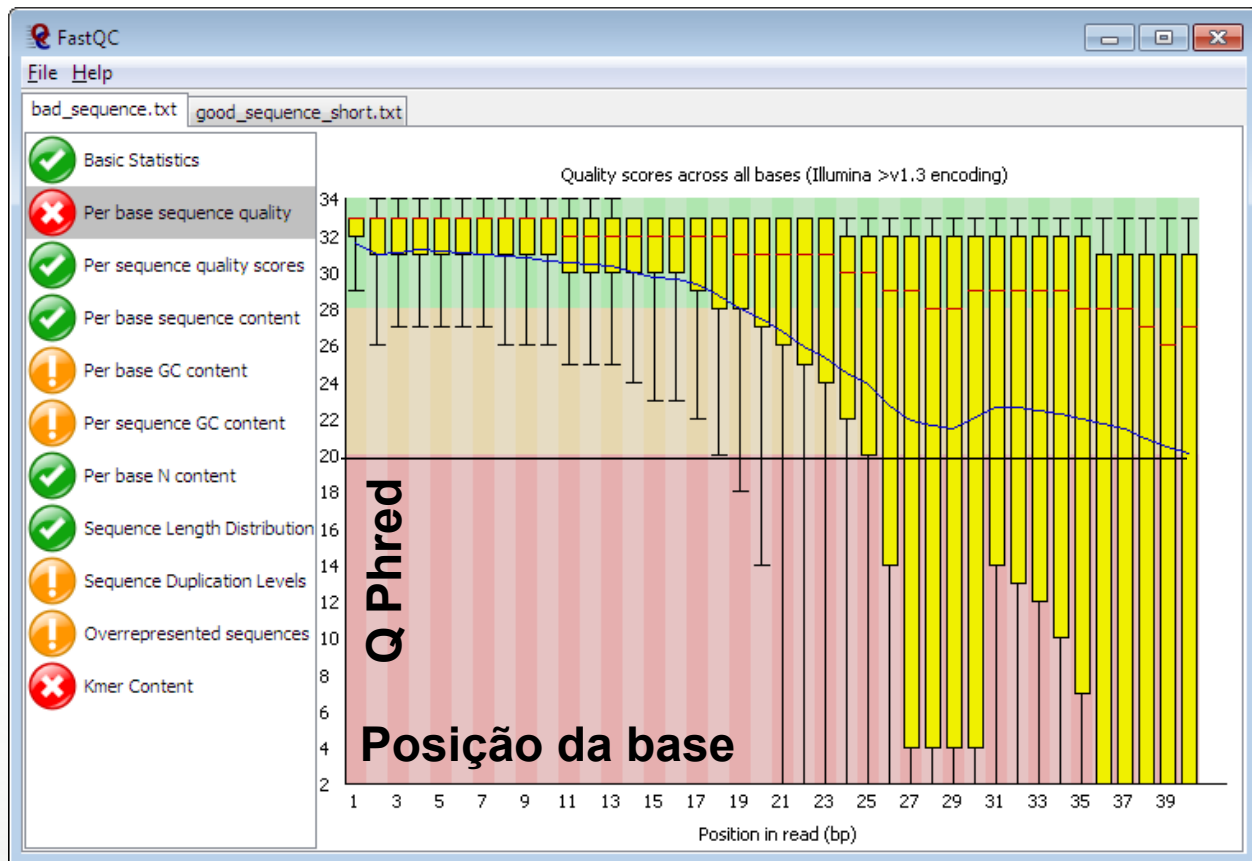
Pipeline bioinformático

Eliminação de
barcodes e primers



Controle de qualidade

Filtragem



Phred = 20 = 99%

Phred = 30 = 99,9%

Tipo de arquivos

- Fasta = .fa, .fasta, .fna



Header	●	>VIT_201s0011g03530.1
Sequence	●	AATTAAGCATAAATACTCACTCTTACCCCCTTATTTTCTTATCTCTCATCACTTTTGGTGCGAAG
	●	GACCATGAGAACAAGCTGCAATGGGTGTAGGGTTCTTCGCAAGGCATGCAGCCAAGACTGCATCA
Header	●	>VIT_201s0011g03540.1
Sequence	●	CAGGTAGCGTGAAGTTAAACCCTAGCGCTTTAGACAAACAGCTGTAGTCACCGCCCACAAACACC
	●	AGCCTCTGAGACACCACCTCAAACCTTTCCACTTAAATACACATCCCTCACACCCTTTTCAATTC
Header	●	>VIT_201s0011g03550.1
Sequence	●	CATGCAAAGCTGAACGCGATGCTGTGATTGGTGGTAAGTGGTAGTTGAGTAAATTTGACAGTGAA
	●	GCCGAAATGGTAAAAGACTAAGGCTAGAAGTAGAATACCACTGTTCTTCTCATCACGTGGGCCCA

Tipo de arquivos

- Fastq = .fq, .fastq

sequence

- Line 1 – the **name**, starts with @
- Line 2 – the **sequence**, starts at new line
- Line 3 – some **other** stuff, optional, starts with +
- Line 4 – the **quality scores**, starts at new line

base = T
score = F = 37

```
@SN1083:379:H8VA1ADXX:2:1101:1248:2144 1:N:0:12  
CCTAAATGGTGCCATGCTAGGAGGCCGTGCCCTTCTTGAAAAGTTGTATGTGAA  
+  
BBBFFFFFFFBFFIIIIIFI<FFIIIIIFIIIIIFBFIIIIIIIIFFFIIIIIFIII
```


ASCII Table

Dec	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char
0	0	0		32	20	40	[space]	64	40	100	@	96	60	140	`
1	1	1		33	21	41	!	65	41	101	A	97	61	141	a
2	2	2		34	22	42	"	66	42	102	B	98	62	142	b
3	3	3		35	23	43	#	67	43	103	C	99	63	143	c
4	4	4		36	24	44	\$	68	44	104	D	100	64	144	d
5	5	5		37	25	45	%	69	45	105	E	101	65	145	e
6	6	6		38	26	46	&	70	46	106	F	102	66	146	f
7	7	7		39	27	47	'	71	47	107	G	103	67	147	g
8	8	10		40	28	50	(72	48	110	H	104	68	150	h
9	9	11		41	29	51)	73	49	111	I	105	69	151	i
10	A	12		42	2A	52	*	74	4A	112	J	106	6A	152	j
11	B	13		43	2B	53	+	75	4B	113	K	107	6B	153	k
12	C	14		44	2C	54	,	76	4C	114	L	108	6C	154	l
13	D	15		45	2D	55	-	77	4D	115	M	109	6D	155	m
14	E	16		46	2E	56	.	78	4E	116	N	110	6E	156	n
15	F	17		47	2F	57	/	79	4F	117	O	111	6F	157	o
16	10	20		48	30	60	0	80	50	120	P	112	70	160	p
17	11	21		49	31	61	1	81	51	121	Q	113	71	161	q
18	12	22		50	32	62	2	82	52	122	R	114	72	162	r
19	13	23		51	33	63	3	83	53	123	S	115	73	163	s
20	14	24		52	34	64	4	84	54	124	T	116	74	164	t
21	15	25		53	35	65	5	85	55	125	U	117	75	165	u
22	16	26		54	36	66	6	86	56	126	V	118	76	166	v
23	17	27		55	37	67	7	87	57	127	W	119	77	167	w
24	18	30		56	38	70	8	88	58	130	X	120	78	170	x
25	19	31		57	39	71	9	89	59	131	Y	121	79	171	y
26	1A	32		58	3A	72	:	90	5A	132	Z	122	7A	172	z
27	1B	33		59	3B	73	;	91	5B	133	[123	7B	173	{
28	1C	34		60	3C	74	<	92	5C	134	\	124	7C	174	
29	1D	35		61	3D	75	=	93	5D	135]	125	7D	175	}
30	1E	36		62	3E	76	>	94	5E	136	^	126	7E	176	~
31	1F	37		63	3F	77	?	95	5F	137	_	127	7F	177	

Pipeline bioinformático

Eliminação de
barcodes e primers



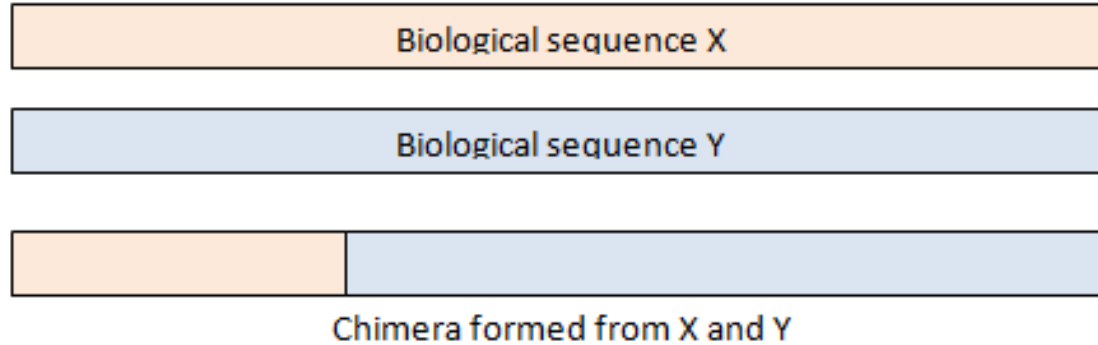
Controle de qualidade



Filtragem



Remoção de
chimeras, denoising



Híbridos formados entre duas sequências de dois
espécies diferentes

Pipeline bioinformático

Eliminação de
barcodes e primers



Controle de qualidade



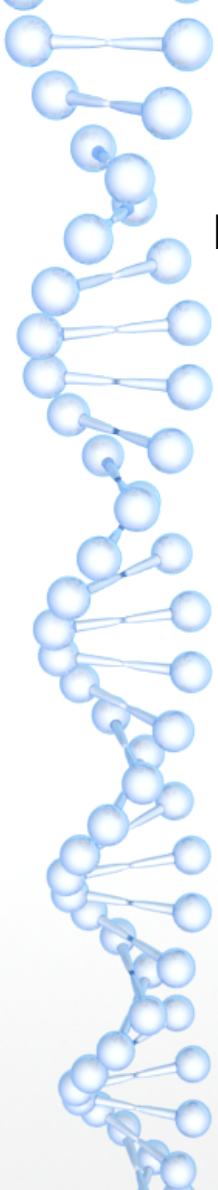
Filtragem



Remoção de
chimeras, denoising

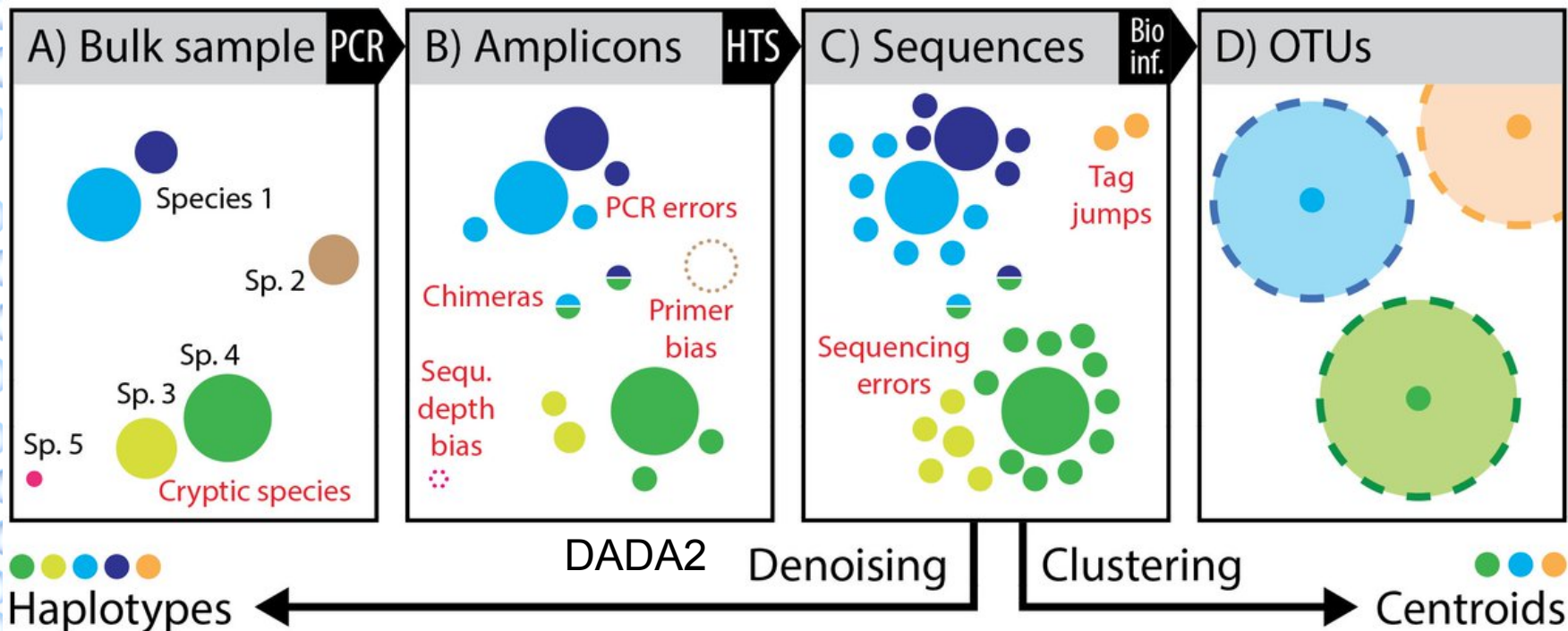


Clusterização



Pipeline bioinformático

Clusterização: OTU (qiime1) vs ASV (qiime2)



Pipeline bioinformático

Eliminação de
barcodes e primers



Controle de qualidade



Filtragem



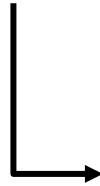
Remoção de
chimeras, denoising



Clusterização



Afiliação taxonômica



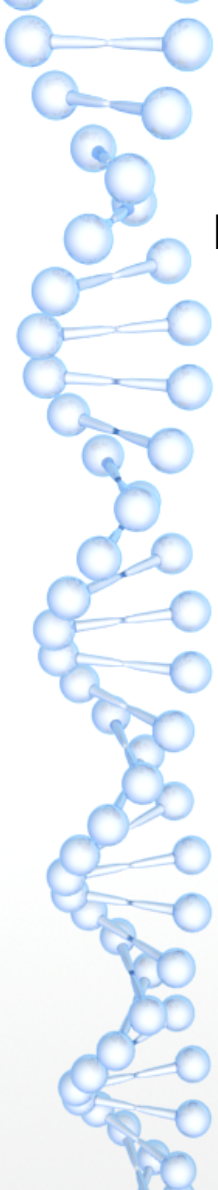
Análises

Alfa e beta diversidade

Composição da comunidade

Análises Ordenação (PCoA, PCA, NMDS)

Correlação com variáveis físico-químicas,
ambientais (CCA)



Pipeline bioinformático

Afiliação taxonômica



Classificação usando base da dados

Pipeline bioinformático

Eliminação de
barcodes e primers



Controle de qualidade



Filtragem



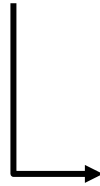
Remoção de
chimeras, denoising



Clusterização



Afiliação taxonômica



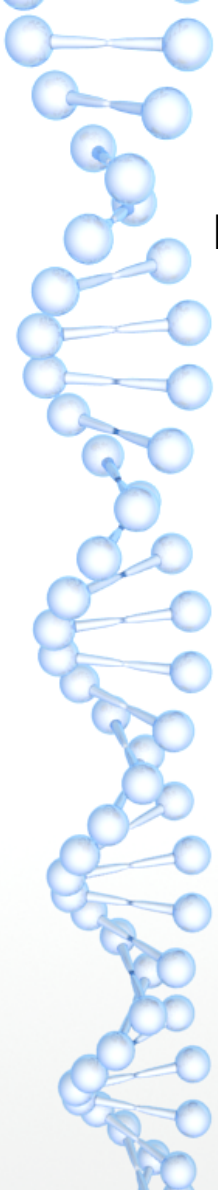
Análises

Alfa e beta diversidade

Composição da comunidade

Análises Ordenação (PCoA, PCA, NMDS)

Correlação com variáveis físico-químicas,
ambientais (CCA)





Pipeline bioinformático

Análises

Alfa e beta diversidade

Composição da comunidade

Análises Ordenação (PCoA, PCA, NMDS)

Correlação com variáveis físico-químicas, ambientais (CCA)

Alfa diversidade: Índices de diversidade (Shannon, Simpson, Dominância) e estimadores de riqueza (Chao1, ACE, espécies observadas)

Beta diversidade: Índices de similaridade
Bray-Curtis e Jaccard
Unifrac (distâncias filogenéticas)

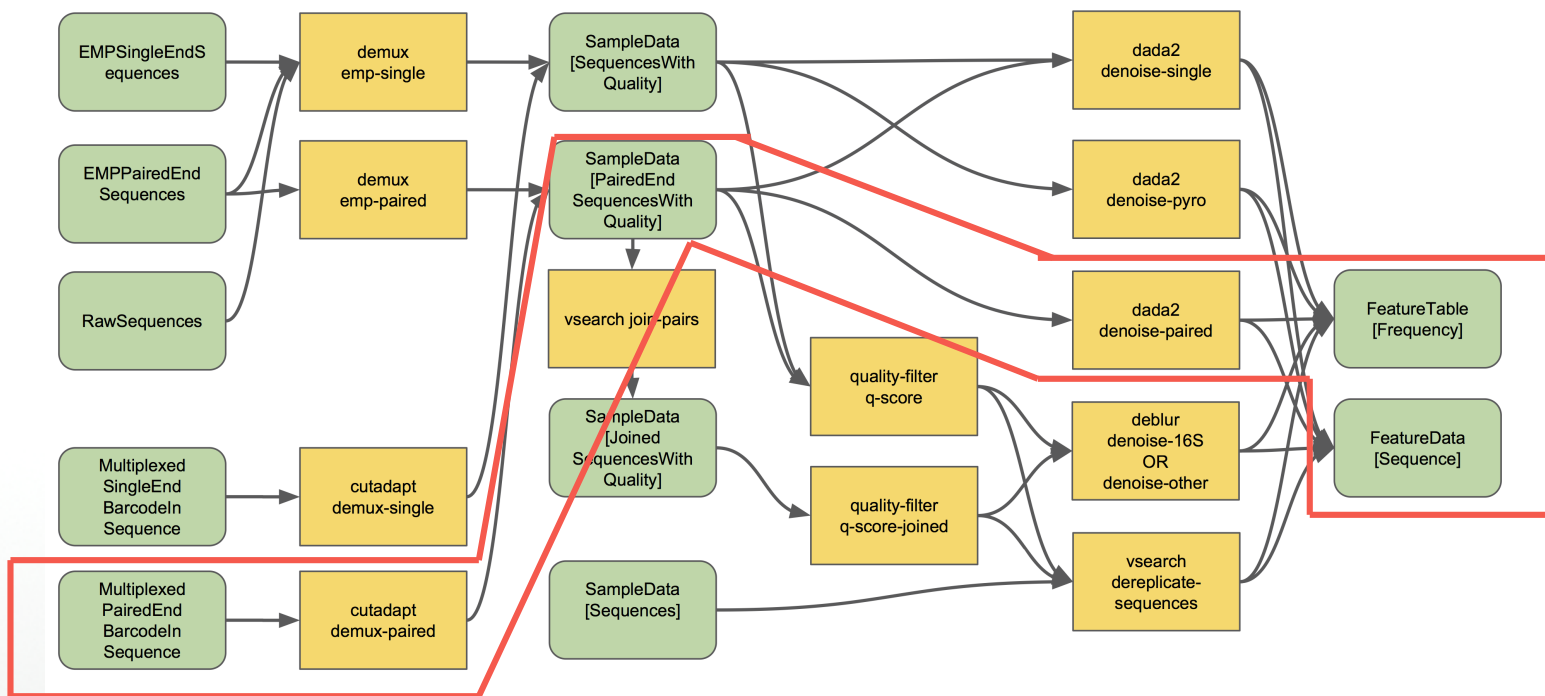


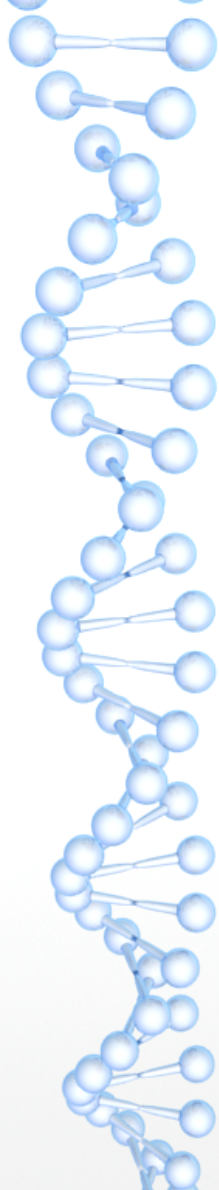
Tipos de arquivos em Qiime2

- qza (artefato): contém dados e metadados. Os metadados descrevem dados, como tipo, formato e como foram gerados (provenance). Pode ser arquivo de entrada ou saída.
- qzv (visualização): contém metadados similares aos artefatos, incluindo informações de proveniência. São saídas terminais de uma análise por exemplo uma tabela com resultados estatísticos, uma visualização interativa, gráficos. **Não** podem ser usados como entradas para outras análises no qiime2



Quantitative Insights Into Microbial Ecology







Primer comando

git clone <https://github.com/khidalgo85/qiime2.git>

- ✓ Analysis fungal ITS amplicon from Illumina sequencing by qiime2.html (*Pipeline* detalhado)
- ✓ workflow_its_qiime2.pdf (*Pipeline* em esquema)
- ✓ summary_stats.xlsx (template em excel para seguimento do número de reads em cada etapa)
- ✓ treinamento_its.pdf (estes slides)