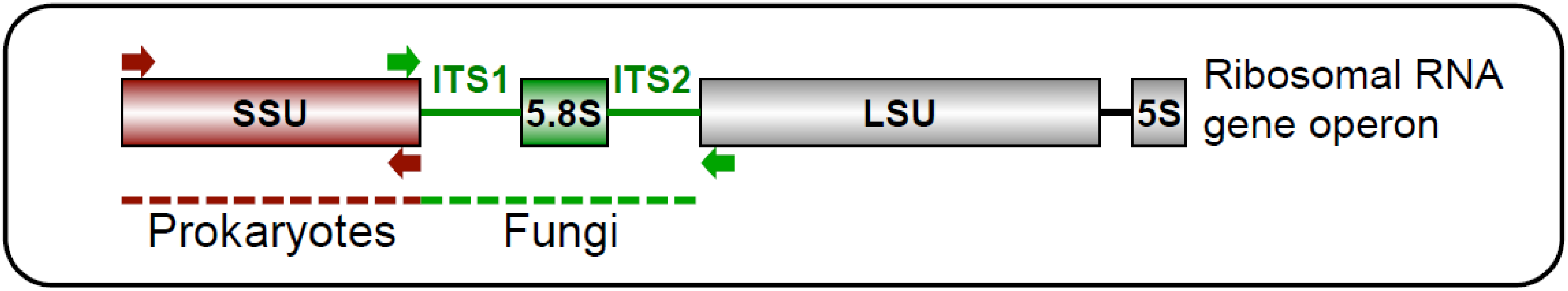
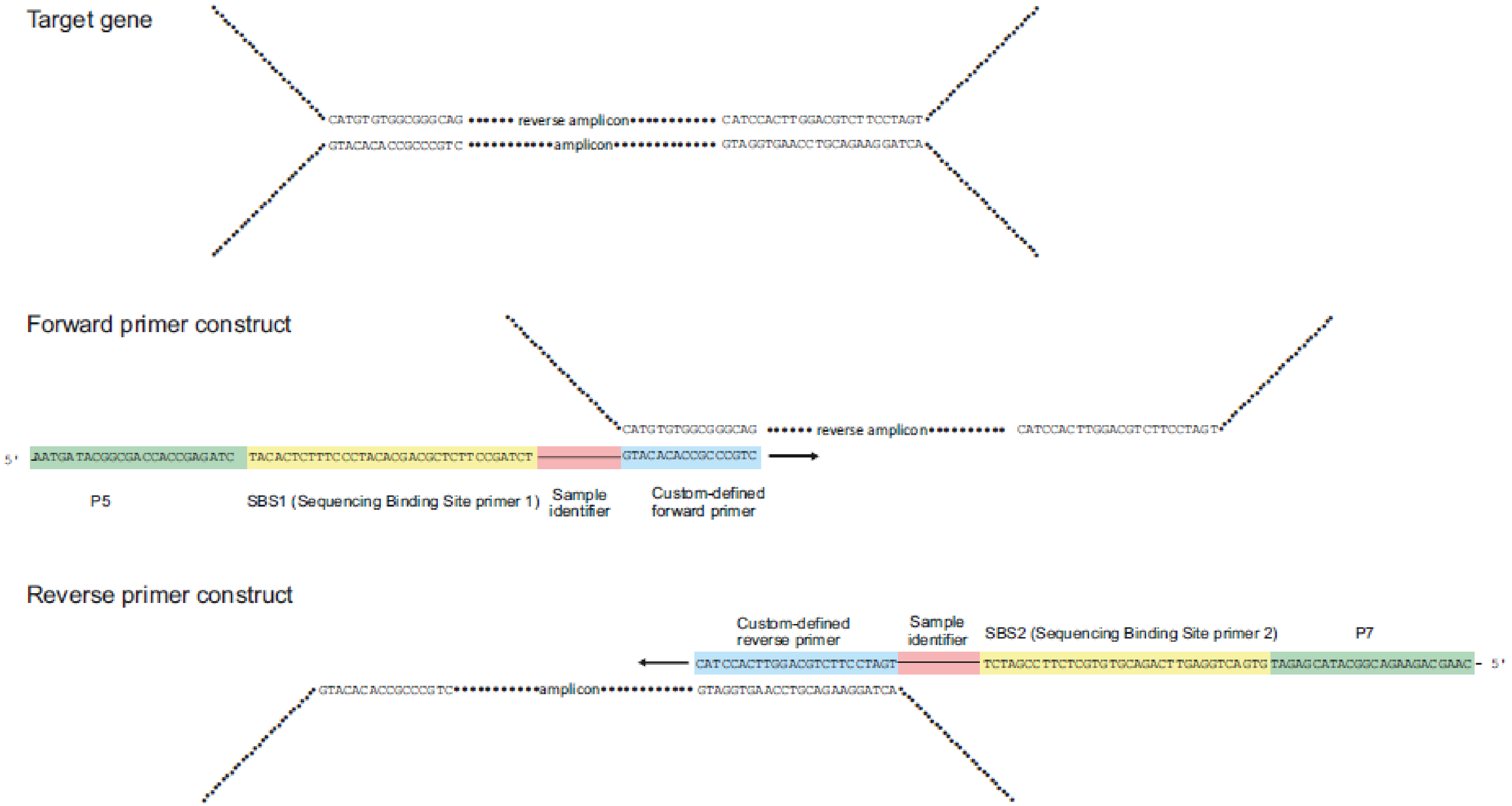
Illumina MiSeq processing Pipeline

Illumina Miseq



Ribosomal RNA gene operon



Target gene

CATGTGTGGCGGGCAG •••••• reverse amplicon•••••••••••• CATCCACTTGGACGTCTTCCTAGT

GTACACACCGCCCGTC ••••••••••••amplicon•••••••••••••• GTAGGTGAACCTGCAGAAGGATCA

Forward primer construct

CATGTGTGGCGGGCAG •••••• reverse amplicon•••••••••• CATCCACTTGGACGTCTTCCTAGT

5' AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTTCCGATCT GTACACACCGCCCGTC →

P5

SBS1 (Sequencing Binding Site primer 1)

Sample identifier

Custom-defined forward primer

Reverse primer construct

Custom-defined reverse primer

Sample identifier

SBS2 (Sequencing Binding Site primer 2)

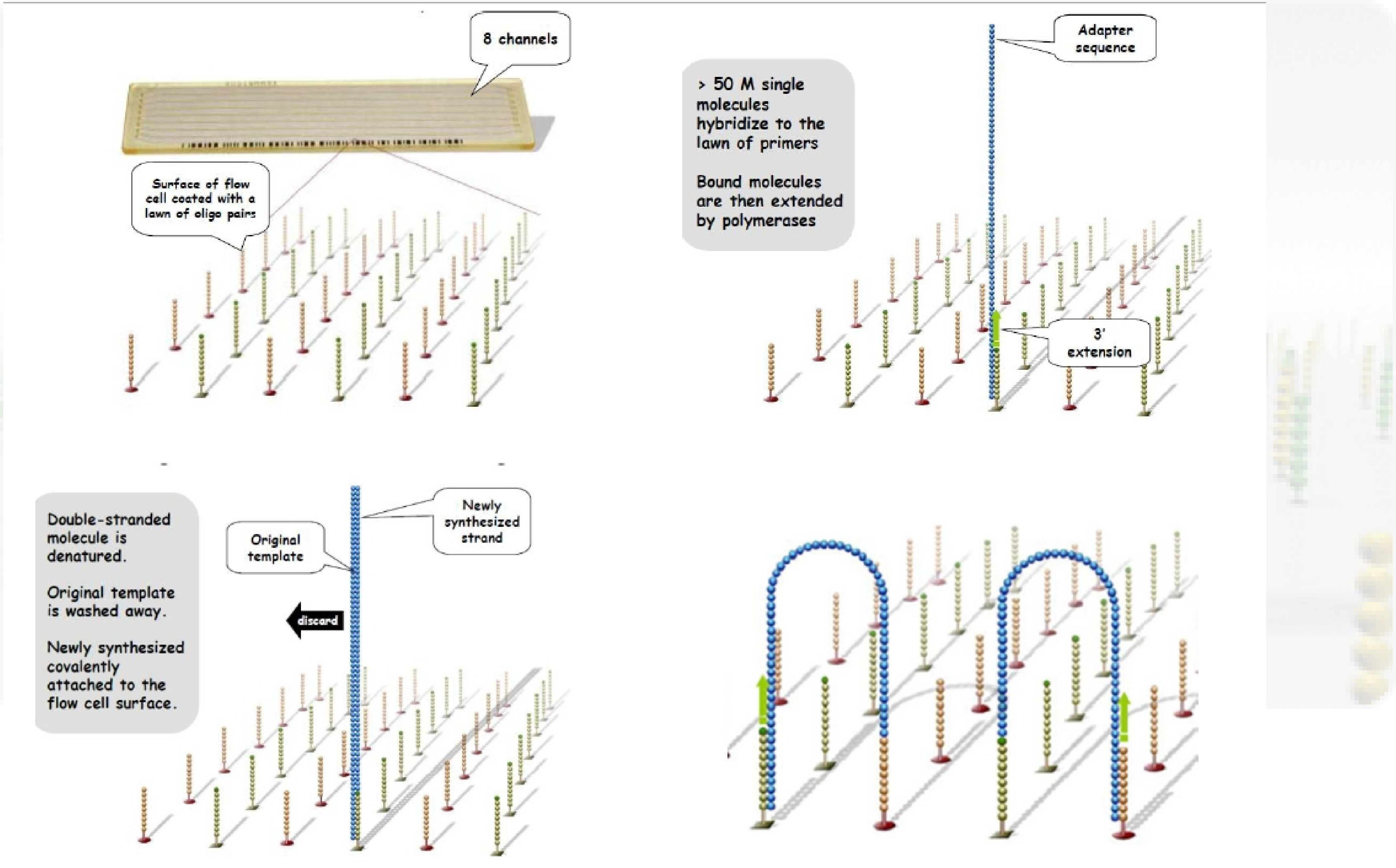
P7

← CATCCACTTGGACGTCTTCCTAGT TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG TAGAGCATACGGCAGAAGACGAAC - 5'

GTACACACCGCCCGTC•••••••••••amplicon••••••••••• GTAGGTGAACCTGCAGAAGGATCA

Illumina MiSeq

bridge amplification



> 50 M single molecules hybridize to the lawn of primers

Bound molecules are then extended by polymerases

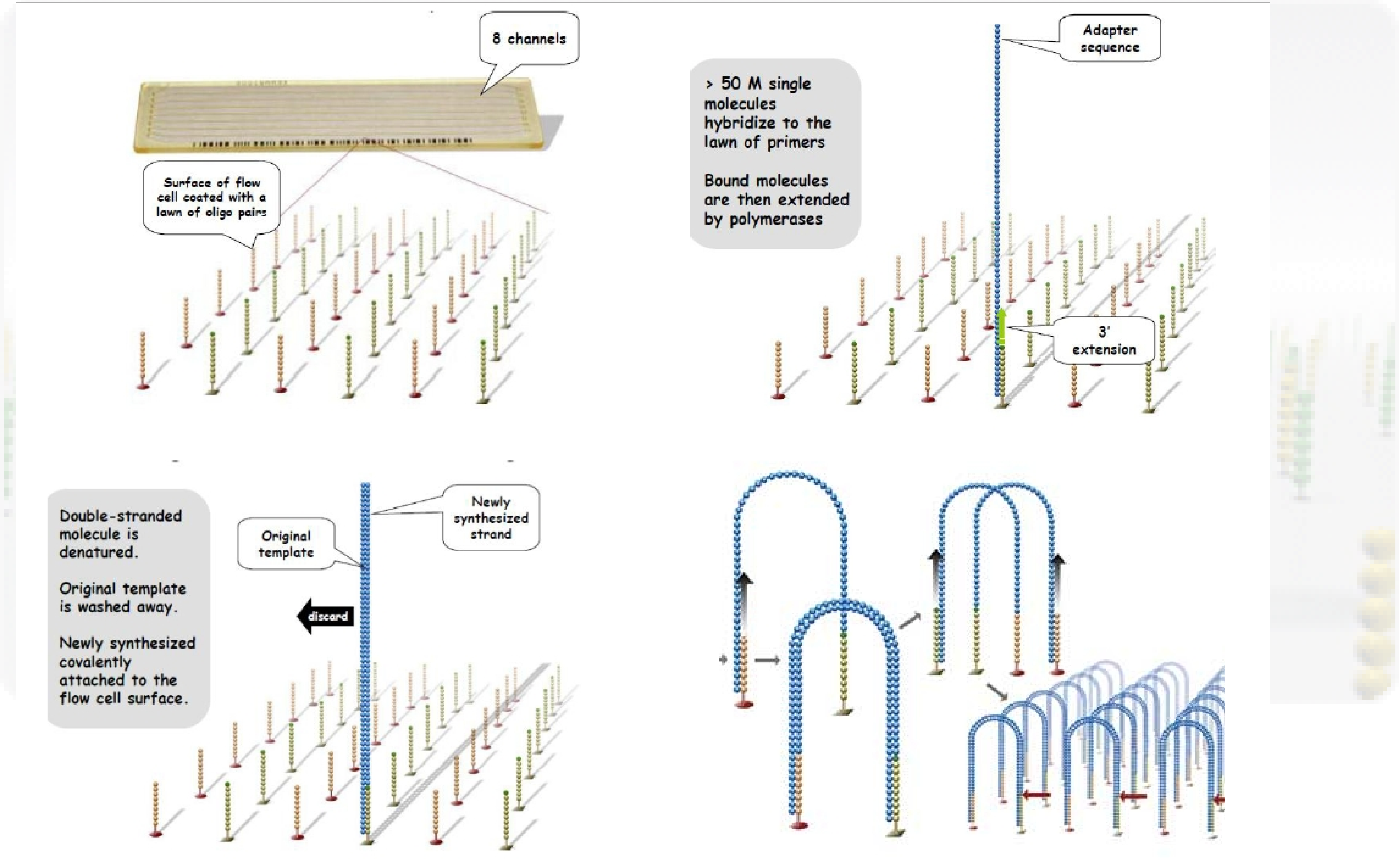
Double-stranded molecule is denatured.

Original template is washed away.

Newly synthesized covalently attached to the flow cell surface.

Illumina MiSeq

bridge amplification



> 50 M single molecules hybridize to the lawn of primers

Bound molecules are then extended by polymerases

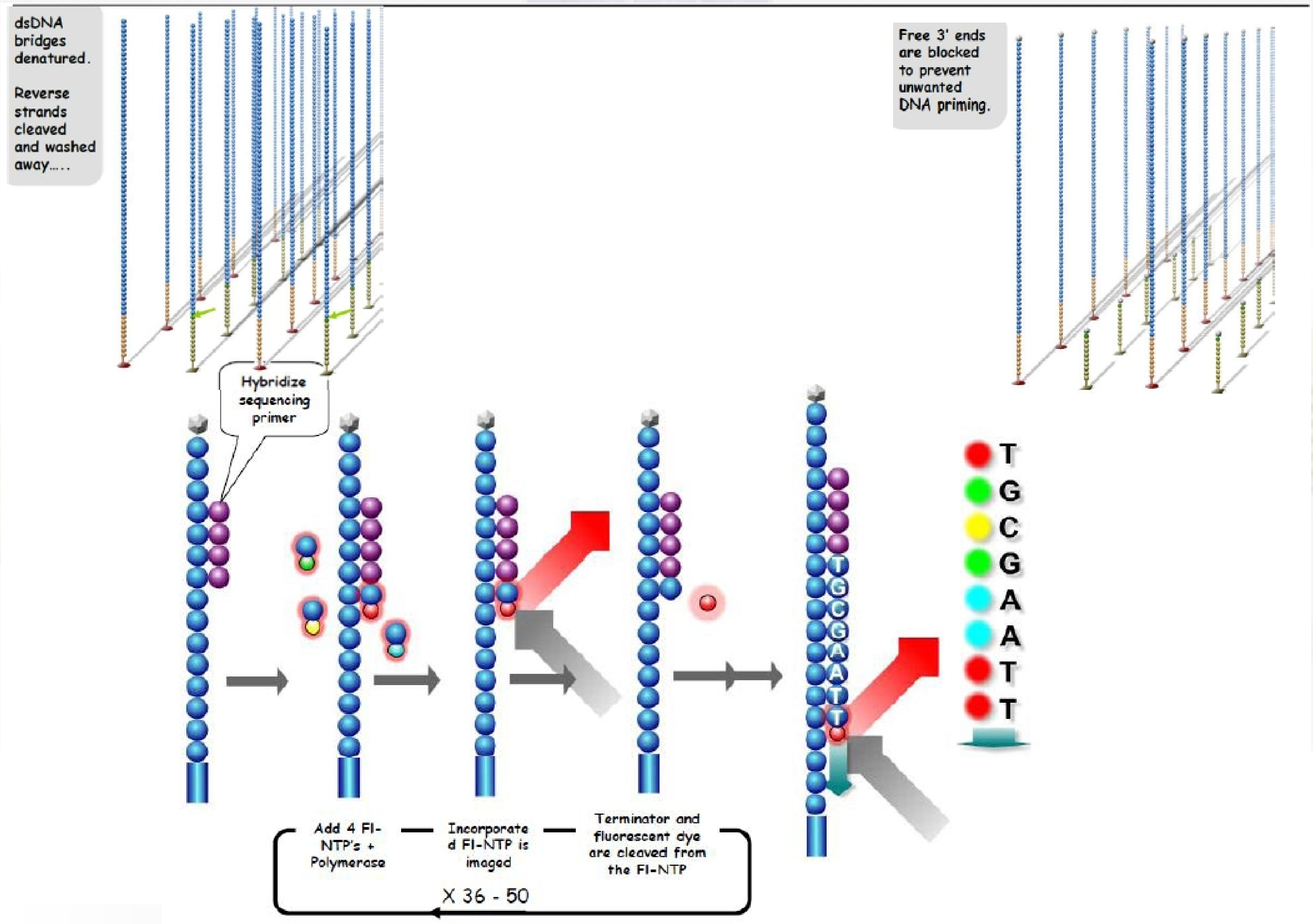
Double-stranded molecule is denatured.

Original template is washed away.

Newly synthesized covalently attached to the flow cell surface.

Illumina MiSeq

Sequencing by synthesis



dsDNA bridges denatured.

Reverse strands cleaved and washed away.....

Free 3' ends are blocked to prevent unwanted DNA priming.

Hybridize sequencing primer

Add 4 Fl-NTP's + Polymerase

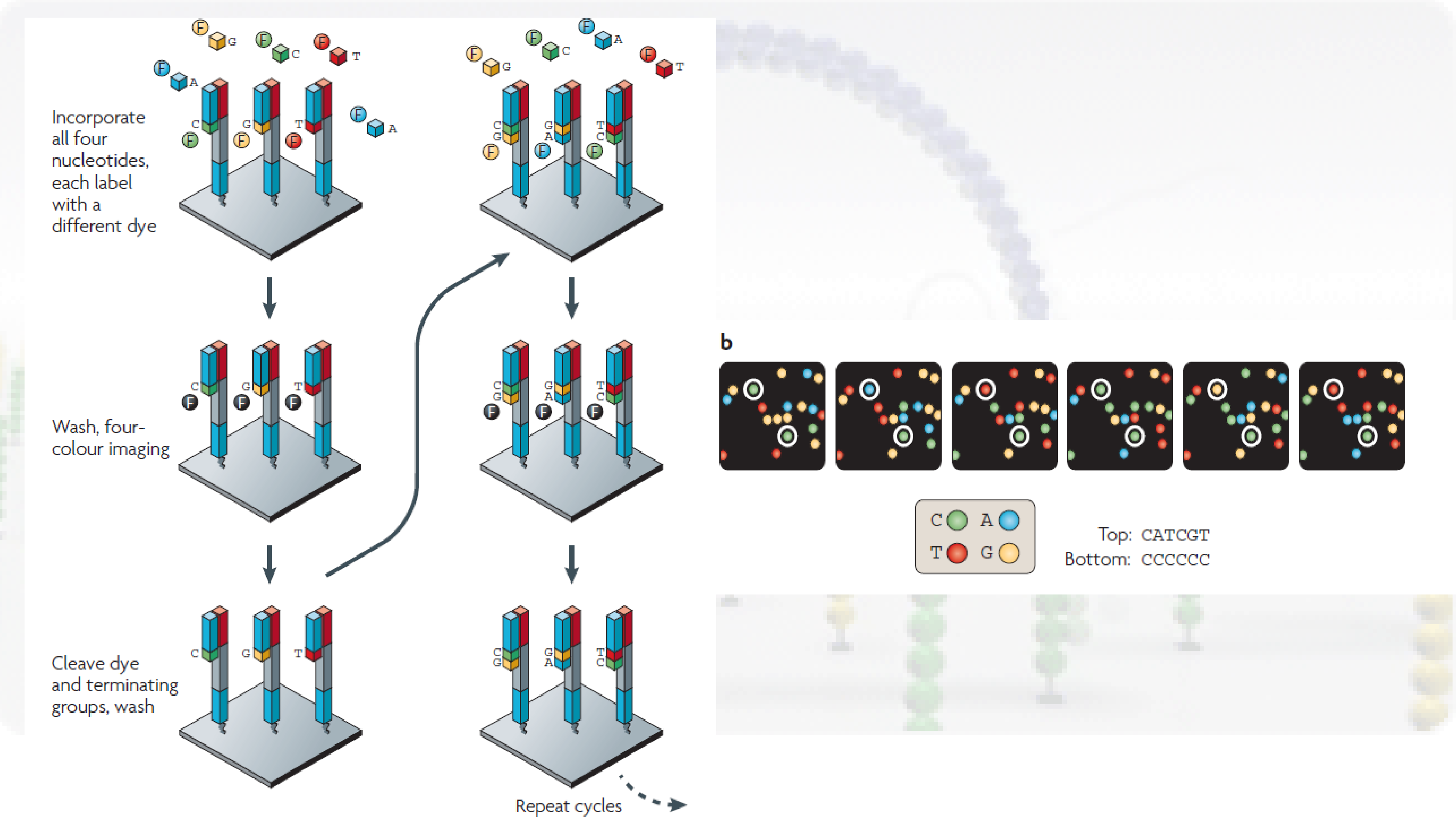
Incorporated Fl-NTP is imaged

Terminator and fluorescent dye are cleaved from the Fl-NTP

X 36 - 50

Illumina MiSeq

Sequencing by synthesis



Incorporate all four nucleotides, each label with a different dye

Wash, four-colour imaging

Cleave dye and terminating groups, wash

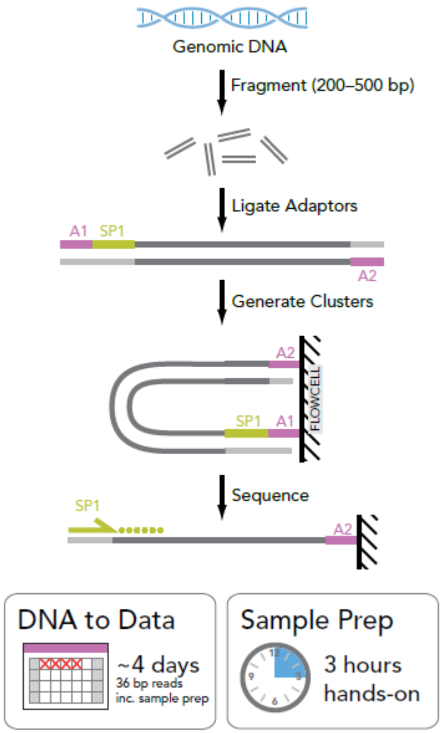
Repeat cycles

Top: CATCGT
Bottom: CCCCCC

NATuRe RevIewS | Genetics vOLume 11 | jANuARy 2010

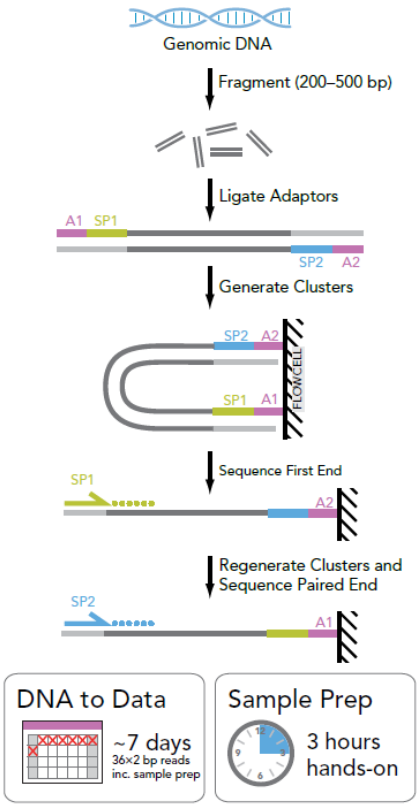
Illumina Miseq

Figure 6A: Single-Read Sequencing



Fragmented sample DNA is size-selected and adaptors are ligated to the ends. Adaptors (A1 and A2) are used to attach fragments to the flow cell, and A1 includes the sequencing primer site (SP1). Libraries are deposited on a flow cell and clusters are generated in the Illumina Cluster Station. Flow cells prepared with template clusters are sequenced in the Genome Analyzer.

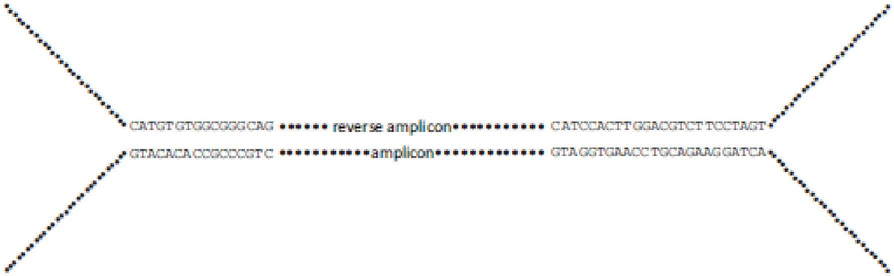
Figure 6B: Paired-End Sequencing



Adapters containing attachment sequences (A1 & A2) and sequencing primer sites (SP1 & SP2) are ligated onto DNA fragments (e.g., genomic DNA). The resulting library of single molecules is attached to a flow cell. Each end of every template is read sequentially.

Primer constructs

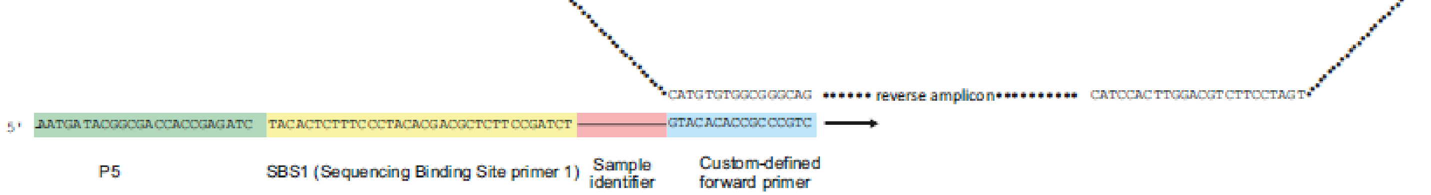
Target gene



CATGTGTGGCGGGCAG •••••• reverse amplicon••••••••••• CATCCACTTGGACGTCTTCCTAGT

GTACACACCGCCCGTC •••••••••••amplicon••••••••••••• GTAGGTGAACCTGCAGAAGGATCA

Forward primer construct



CATGTGTGGCGGGCAG •••••• reverse amplicon•••••••••• CATCCACTTGGACGTCTTCCTAGT

5' AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTTCCGATCT GTACACACCGCCCGTC

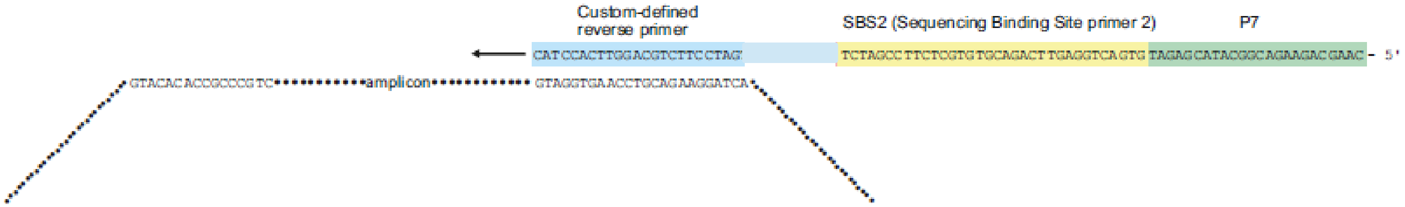
P5

SBS1 (Sequencing Binding Site primer 1)

Sample identifier

Custom-defined forward primer

Reverse primer construct



Custom-defined reverse primer

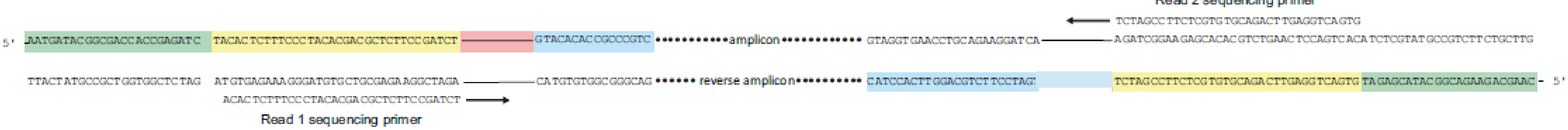
SBS2 (Sequencing Binding Site primer 2)

P7

CATCCACTTGGACGTCTTCCTAG TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG TAGAGCATACGGCAGAAGACGAAC - 5'

GTACACACCGCCCGTC•••••••••••amplicon••••••••••• GTAGGTGAACCTGCAGAAGGATCA

Amplification products



Read 2 sequencing primer

TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG

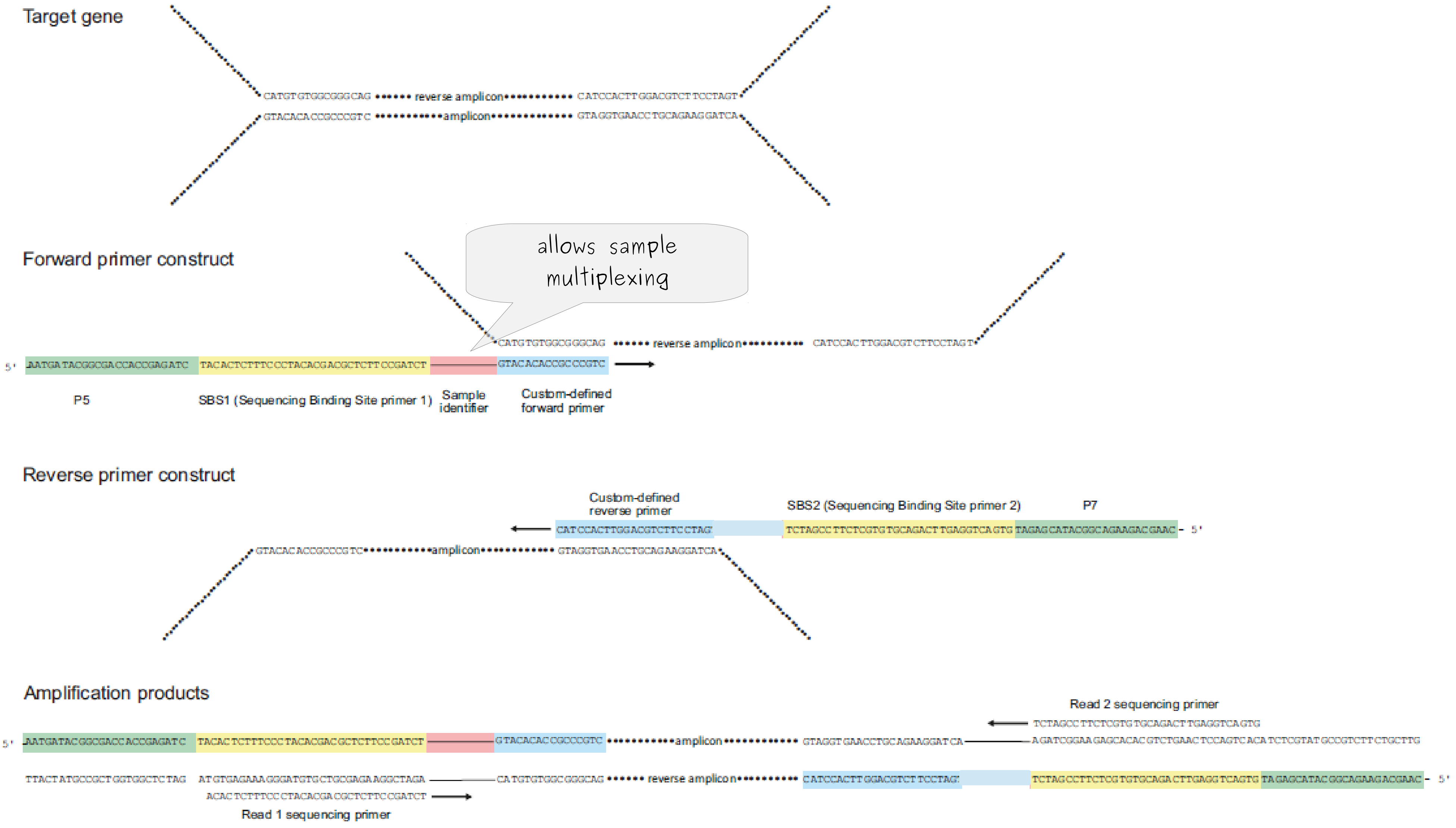
5' AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTTCCGATCT GTACACACCGCCCGTC •••••••••••amplicon••••••••••••• GTAGGTGAACCTGCAGAAGGATCA AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCTCGTATGCCGTCTTCTGCTTG

TTACTATGCCGCTGGTGGCTCTAG ATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA CATGTGTGGCGGGCAG •••••• reverse amplicon•••••••••• CATCCACTTGGACGTCTTCCTAG TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG TAGAGCATACGGCAGAAGACGAAC - 5'

ACACTCTTTCCCTACACGACGCTCTTCCGATCT

Read 1 sequencing primer

Primer constructs



Target gene

CATGTGTGGCGGGCAG •••••• reverse amplicon•••••••••••• CATCCACTTGGACGTCTTCCTAGT

GTACACACCGCCCGTC •••••••••••amplicon••••••••••••• GTAGGTGAACCTGCAGAAGGATCA

Forward primer construct

allows sample multiplexing

CATGTGTGGCGGGCAG •••••• reverse amplicon•••••••••• CATCCACTTGGACGTCTTCCTAGT

5' AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTTCCGATCT GTACACACCGCCCGTC

P5

SBS1 (Sequencing Binding Site primer 1)

Sample identifier

Custom-defined forward primer

Reverse primer construct

Custom-defined reverse primer

SBS2 (Sequencing Binding Site primer 2)

P7

CATCCACTTGGACGTCTTCCTAG TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG TAGAGCATACGGCAGAAGACGAAC - 5'

GTACACACCGCCCGTC•••••••••••amplicon•••••••••• GTAGGTGAACCTGCAGAAGGATCA

Amplification products

Read 2 sequencing primer

TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG

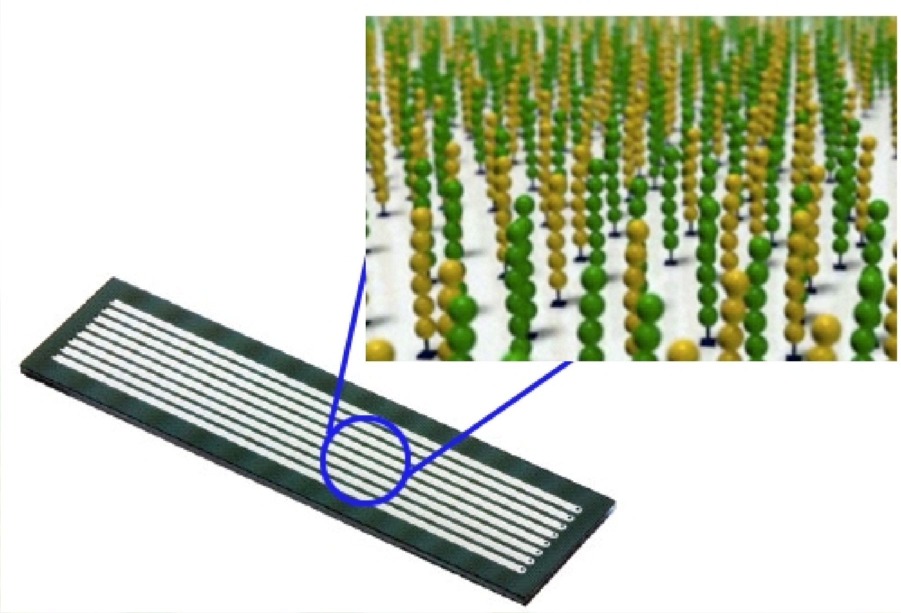
5' AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTTCCGATCT GTACACACCGCCCGTC •••••••••••amplicon•••••••••••• GTAGGTGAACCTGCAGAAGGATCA AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCTCGTATGCCGTCTTCTGCTTG

TTACTATGCCGCTGGTGGCTCTAG ATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA CATGTGTGGCGGGCAG •••••• reverse amplicon•••••••••• CATCCACTTGGACGTCTTCCTAG TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG TAGAGCATACGGCAGAAGACGAAC - 5'

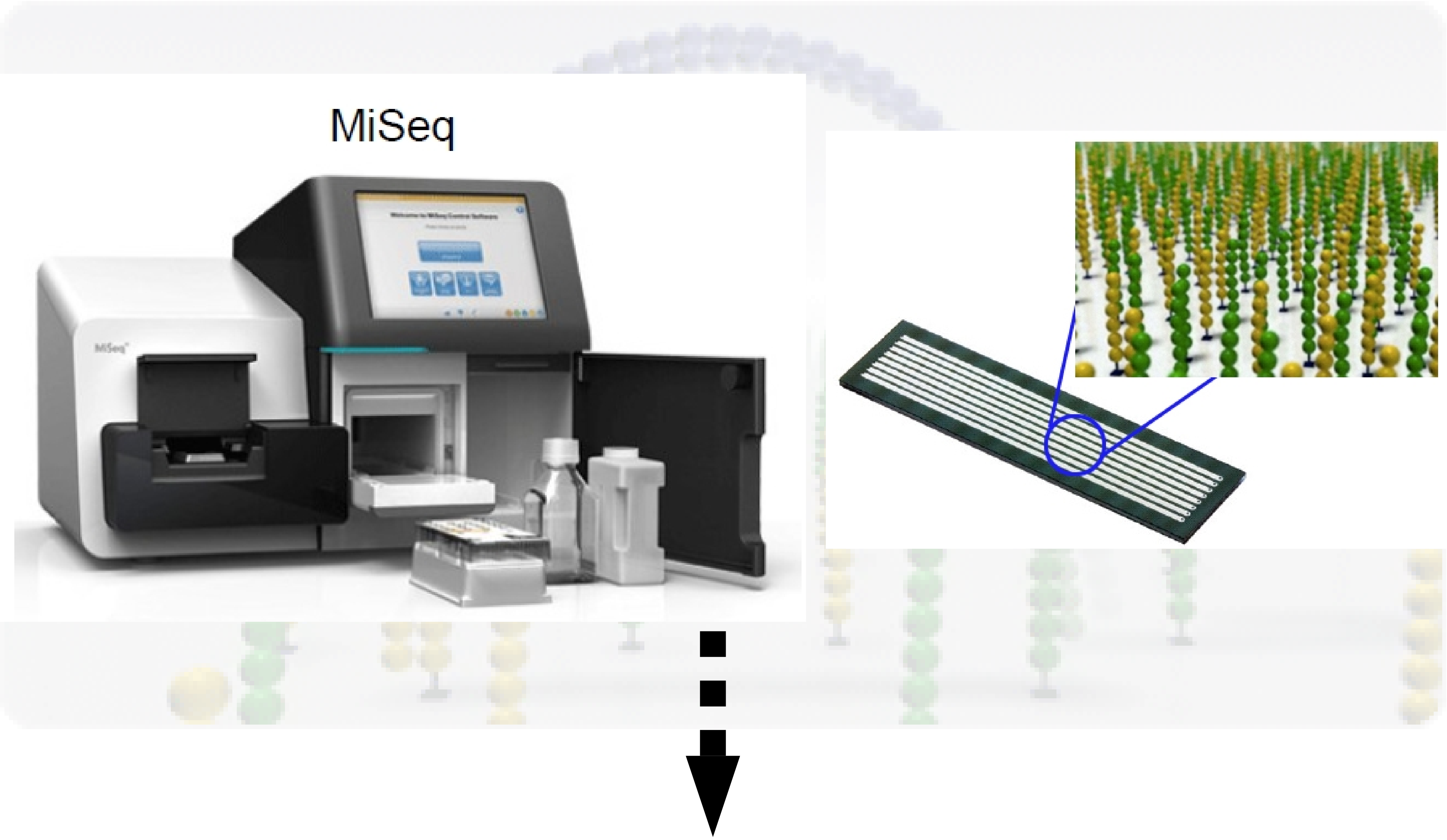
ACACTCTTTCCCTACACGACGCTCTTCCGATCT

Read 1 sequencing primer

Illumina MiSeq

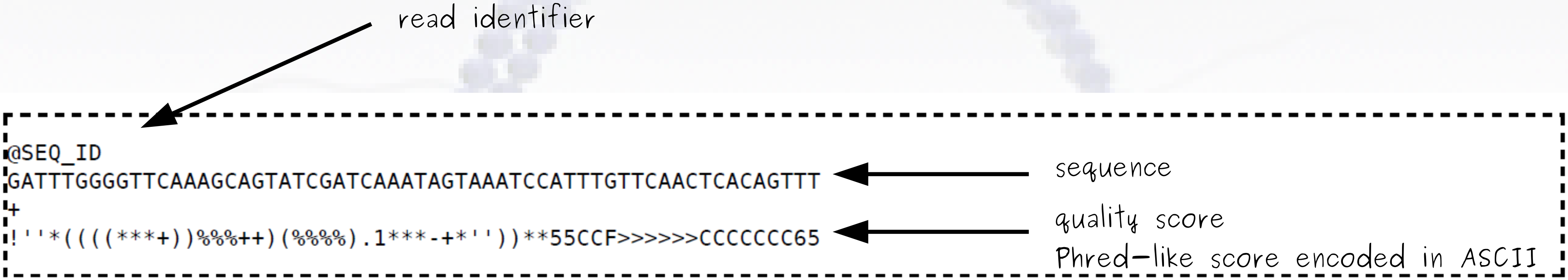


Illumina MiSeq



fastq files with a lot of reads

Fastq

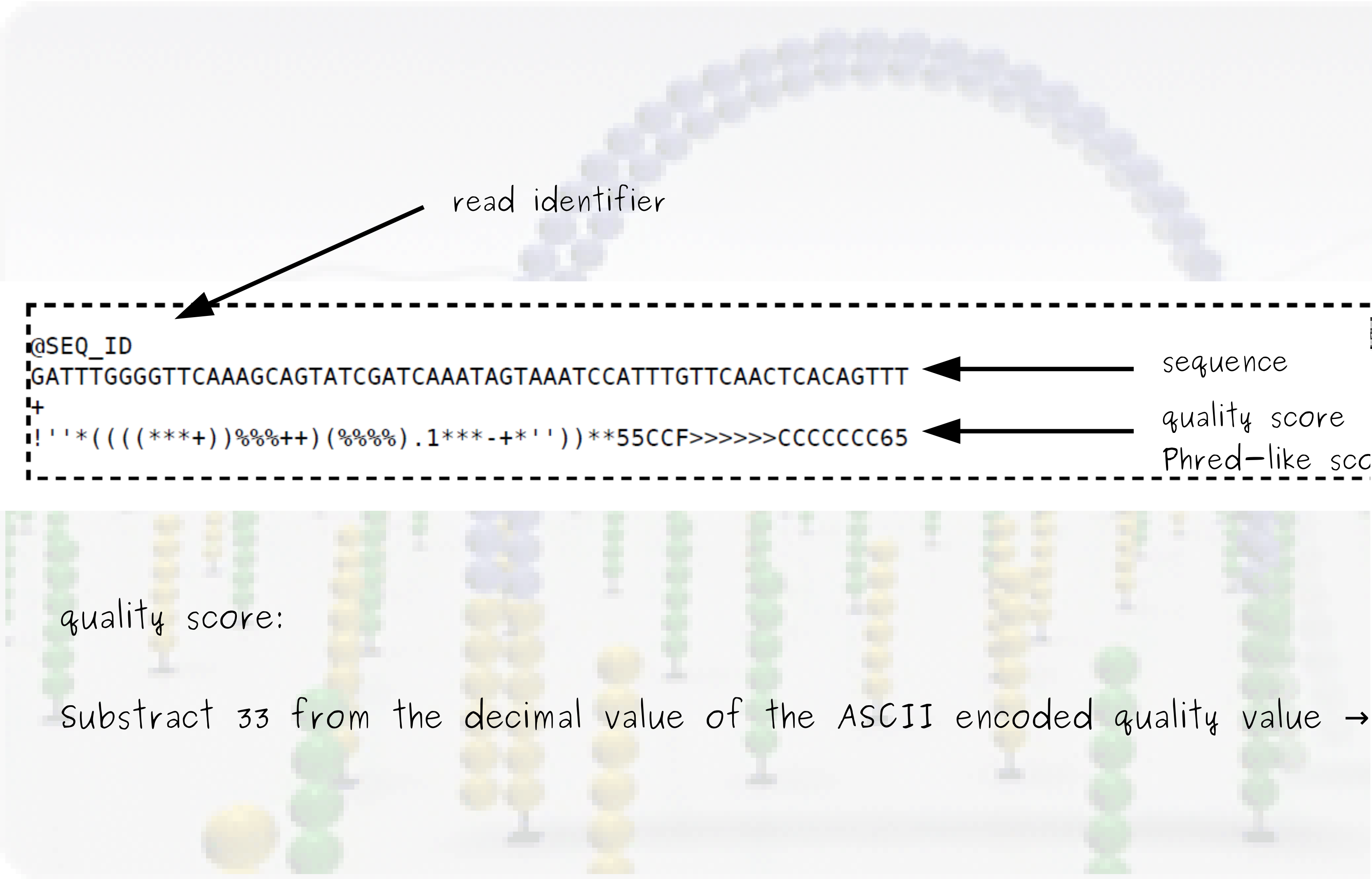


```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC65
```

quality score:

Substract 33 from the decimal value of the ASCII encoded quality value → Phred quality value

Fastq



read identifier

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC65
```

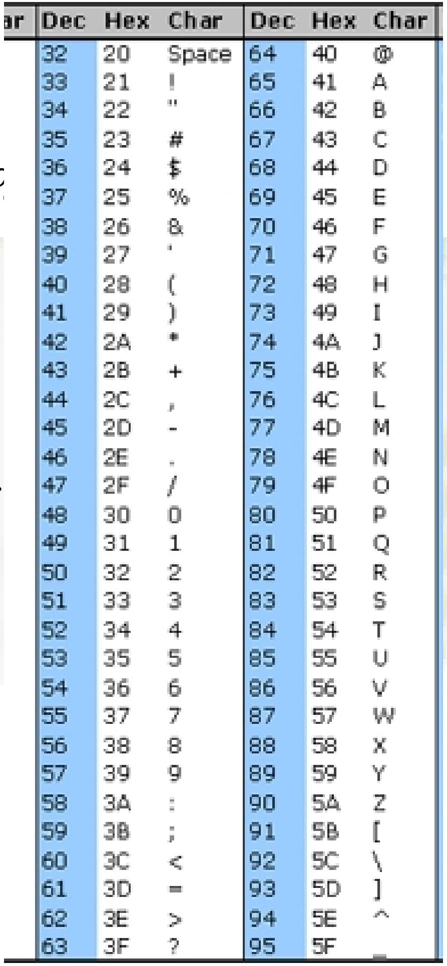
sequence

quality score

Phred-like scc

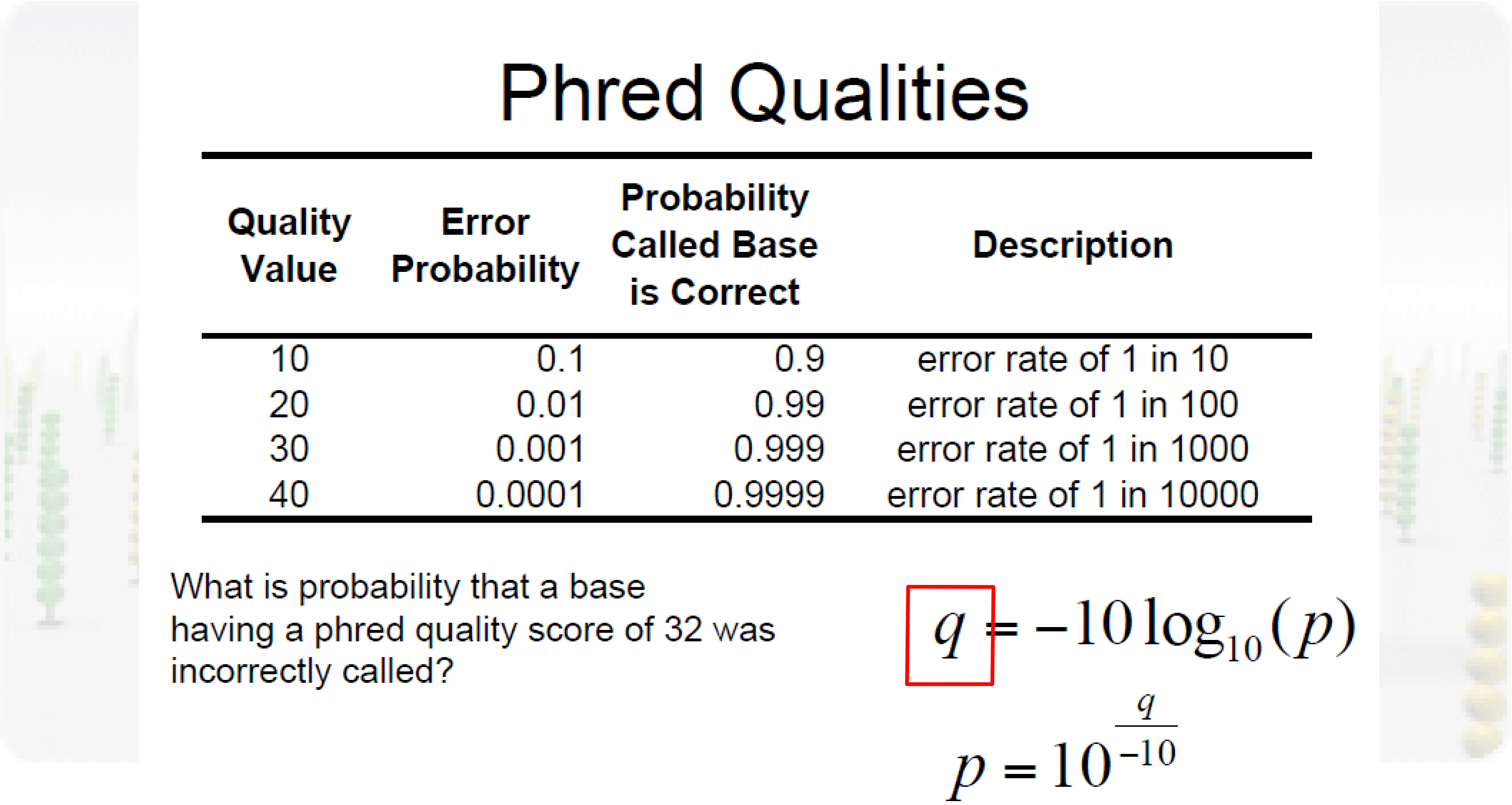
quality score:

Substract 33 from the decimal value of the ASCII encoded quality value →



ar	Dec	Hex	Char	Dec	Hex	Char
	32	20	Space	64	40	@
	33	21	!	65	41	A
	34	22	"	66	42	B
	35	23	#	67	43	C
	36	24	$	68	44	D
	37	25	%	69	45	E
	38	26	&	70	46	F
	39	27	'	71	47	G
	40	28	(	72	48	H
	41	29	)	73	49	I
	42	2A	*	74	4A	J
	43	2B	+	75	4B	K
	44	2C	,	76	4C	L
	45	2D	-	77	4D	M
	46	2E	.	78	4E	N
	47	2F	/	79	4F	O
	48	30	0	80	50	P
	49	31	1	81	51	Q
	50	32	2	82	52	R
	51	33	3	83	53	S
	52	34	4	84	54	T
	53	35	5	85	55	U
	54	36	6	86	56	V
	55	37	7	87	57	W
	56	38	8	88	58	X
	57	39	9	89	59	Y
	58	3A	:	90	5A	Z
	59	3B	;	91	5B	[
	60	3C	<	92	5C	\
	61	3D	=	93	5D	]
	62	3E	>	94	5E	^
	63	3F	?	95	5F	_

Phred quality



Phred Qualities

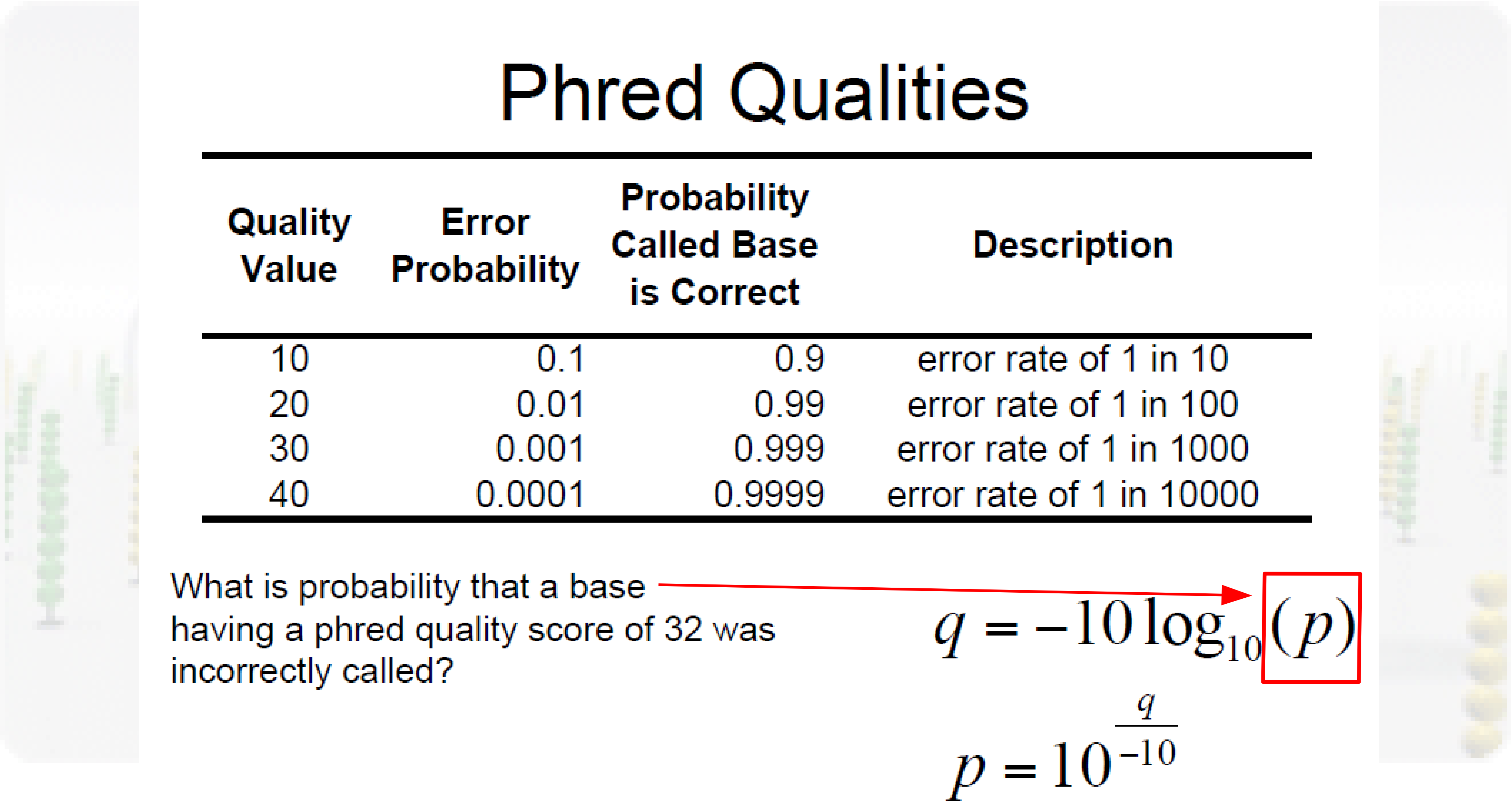
Quality Value	Error Probability	Probability Called Base is Correct	Description
10	0.1	0.9	error rate of 1 in 10
20	0.01	0.99	error rate of 1 in 100
30	0.001	0.999	error rate of 1 in 1000
40	0.0001	0.9999	error rate of 1 in 10000

What is probability that a base having a phred quality score of 32 was incorrectly called?

$$\boxed{q} = -10\log_{10}(p)$$

$$p = 10^{\frac{q}{-10}}$$

Phred quality



Phred Qualities

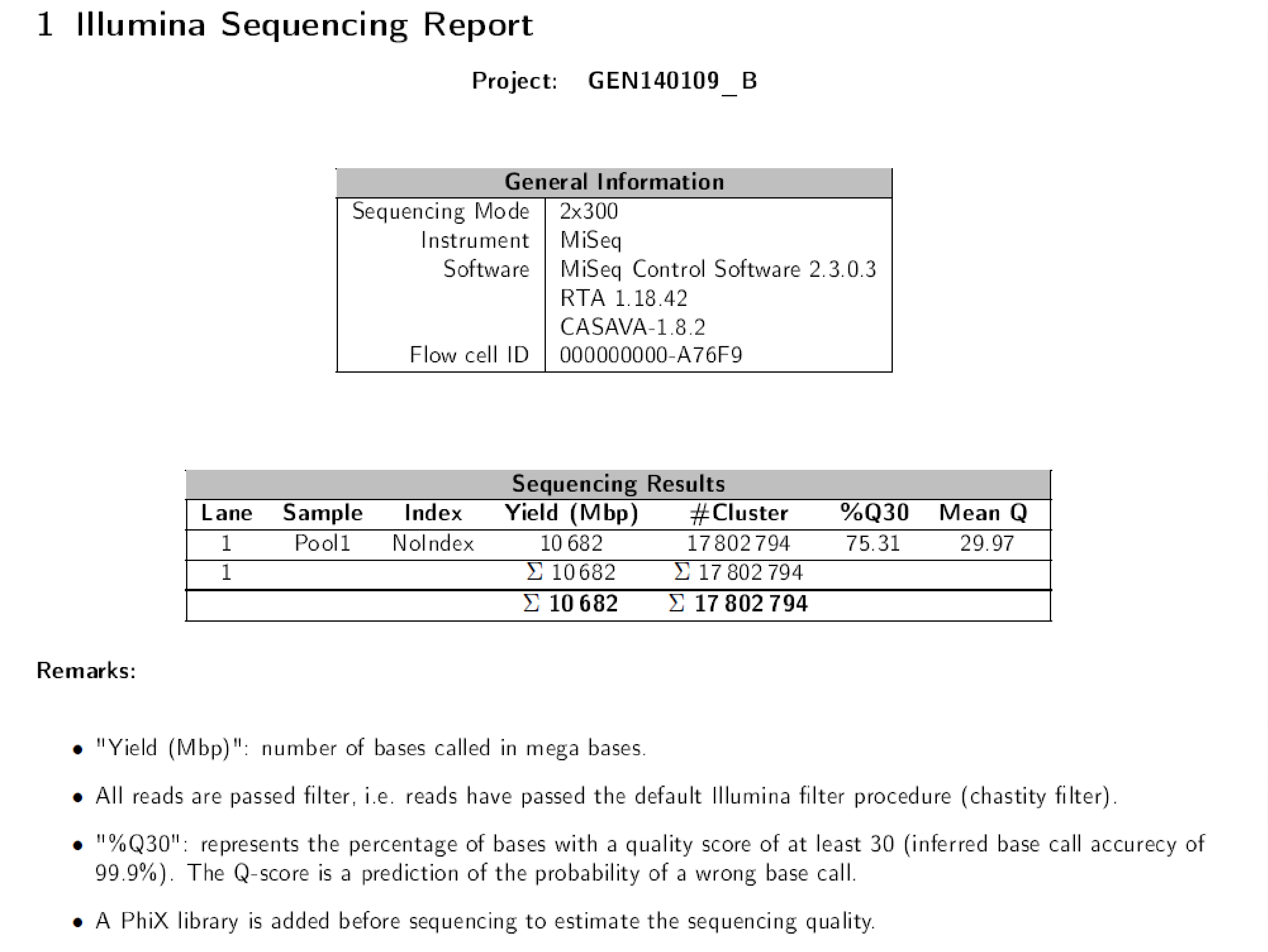
Quality Value	Error Probability	Probability Called Base is Correct	Description
10	0.1	0.9	error rate of 1 in 10
20	0.01	0.99	error rate of 1 in 100
30	0.001	0.999	error rate of 1 in 1000
40	0.0001	0.9999	error rate of 1 in 10000

What is probability that a base having a phred quality score of 32 was incorrectly called?

$$q = -10\log_{10}\boxed{(p)}$$

$$p = 10^{\frac{q}{-10}}$$

Illumina data from eurofins



1 Illumina Sequencing Report

Project: GEN140109_B

General Information	
Sequencing Mode	2x300
Instrument	MiSeq
Software	MiSeq Control Software 2.3.0.3 RTA 1.18.42 CASAVA-1.8.2
Flow cell ID	000000000-A76F9

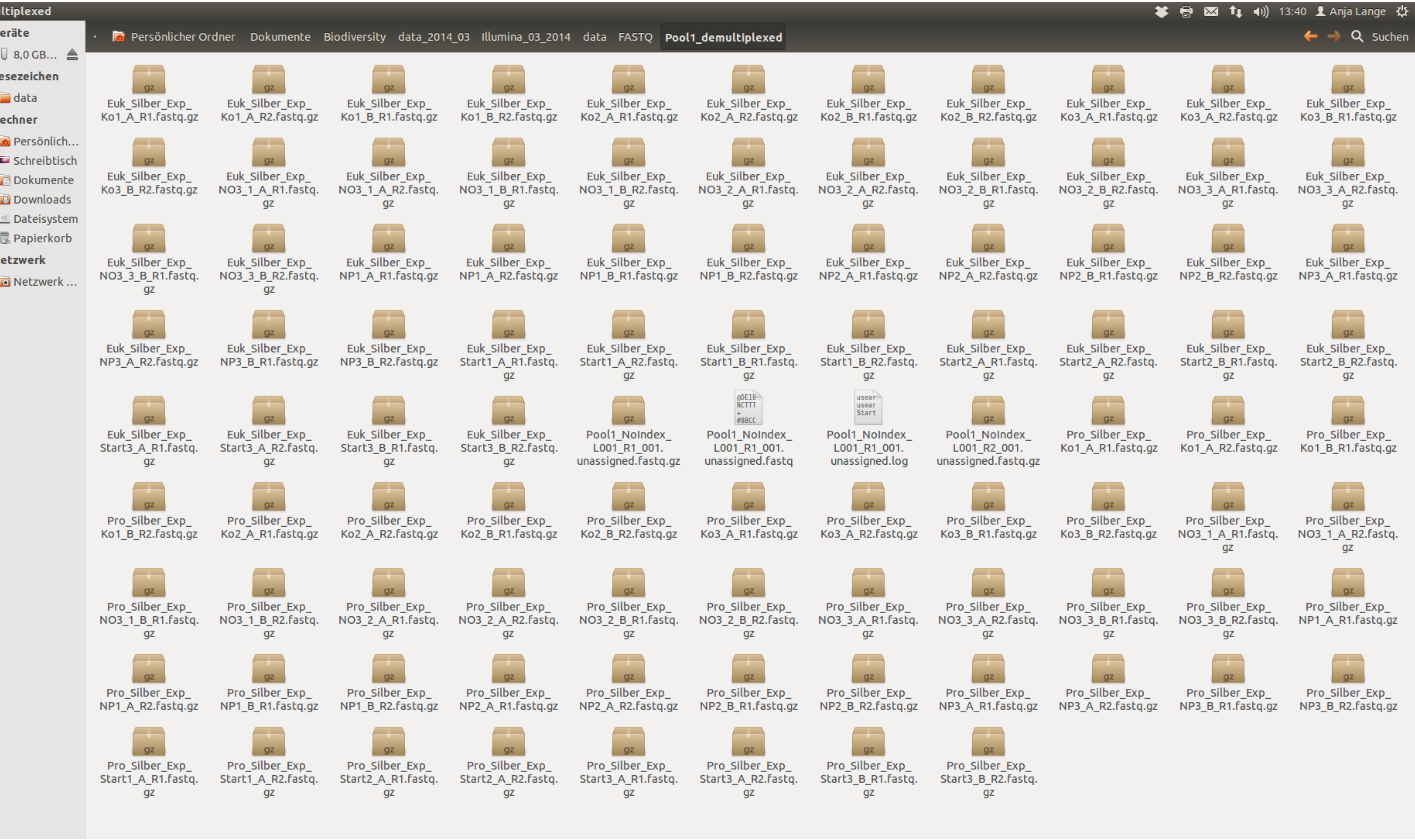
Sequencing Results

Lane	Sample	Index	Yield (Mbp)	#Cluster	%Q30	Mean Q
1	Pool1	NoIndex	10 682	17 802 794	75.31	29.97
1			Σ 10 682	Σ 17 802 794		
			Σ 10 682	**Σ 17 802 794**		

Remarks:

- "Yield (Mbp)": number of bases called in mega bases.
- All reads are passed filter, i.e. reads have passed the default Illumina filter procedure (chastity filter).
- "%Q30": represents the percentage of bases with a quality score of at least 30 (inferred base call accurecy of 99.9%). The Q-score is a prediction of the probability of a wrong base call.
- A PhiX library is added before sequencing to estimate the sequencing quality.

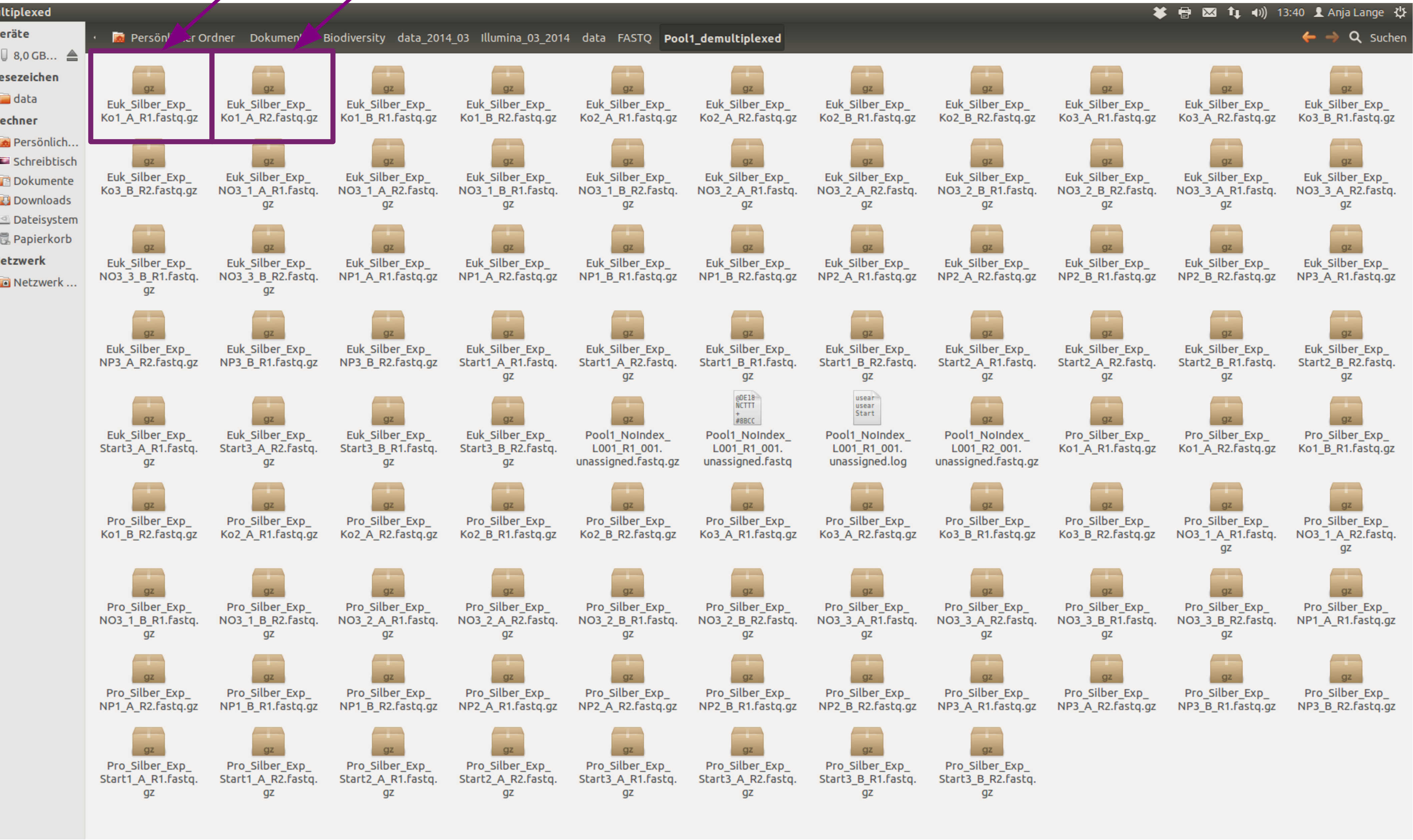
Illumina data from eurofins



Euk_Silber_Exp_Ko1_A_R1.fastq.gz
Euk_Silber_Exp_Ko1_A_R2.fastq.gz
Euk_Silber_Exp_Ko1_B_R1.fastq.gz
Euk_Silber_Exp_Ko1_B_R2.fastq.gz
Euk_Silber_Exp_Ko2_A_R1.fastq.gz
Euk_Silber_Exp_Ko2_A_R2.fastq.gz
Euk_Silber_Exp_Ko2_B_R1.fastq.gz
Euk_Silber_Exp_Ko2_B_R2.fastq.gz
Euk_Silber_Exp_Ko3_A_R1.fastq.gz
Euk_Silber_Exp_Ko3_A_R2.fastq.gz
Euk_Silber_Exp_Ko3_B_R1.fastq.gz
Euk_Silber_Exp_Ko3_B_R2.fastq.gz
Euk_Silber_Exp_NO3_1_A_R1.fastq.gz
Euk_Silber_Exp_NO3_1_A_R2.fastq.gz
Euk_Silber_Exp_NO3_1_B_R1.fastq.gz
Euk_Silber_Exp_NO3_1_B_R2.fastq.gz
Euk_Silber_Exp_NO3_2_A_R1.fastq.gz
Euk_Silber_Exp_NO3_2_A_R2.fastq.gz
Euk_Silber_Exp_NO3_2_B_R1.fastq.gz
Euk_Silber_Exp_NO3_2_B_R2.fastq.gz
Euk_Silber_Exp_NO3_3_A_R1.fastq.gz
Euk_Silber_Exp_NO3_3_A_R2.fastq.gz
Euk_Silber_Exp_NO3_3_B_R1.fastq.gz
Euk_Silber_Exp_NO3_3_B_R2.fastq.gz
Euk_Silber_Exp_NP1_A_R1.fastq.gz
Euk_Silber_Exp_NP1_A_R2.fastq.gz
Euk_Silber_Exp_NP1_B_R1.fastq.gz
Euk_Silber_Exp_NP1_B_R2.fastq.gz
Euk_Silber_Exp_NP2_A_R1.fastq.gz
Euk_Silber_Exp_NP2_A_R2.fastq.gz
Euk_Silber_Exp_NP2_B_R1.fastq.gz
Euk_Silber_Exp_NP2_B_R2.fastq.gz
Euk_Silber_Exp_NP3_A_R1.fastq.gz
Euk_Silber_Exp_NP3_A_R2.fastq.gz
Euk_Silber_Exp_NP3_B_R1.fastq.gz
Euk_Silber_Exp_NP3_B_R2.fastq.gz
Euk_Silber_Exp_Start1_A_R1.fastq.gz
Euk_Silber_Exp_Start1_A_R2.fastq.gz
Euk_Silber_Exp_Start1_B_R1.fastq.gz
Euk_Silber_Exp_Start1_B_R2.fastq.gz
Euk_Silber_Exp_Start2_A_R1.fastq.gz
Euk_Silber_Exp_Start2_A_R2.fastq.gz
Euk_Silber_Exp_Start2_B_R1.fastq.gz
Euk_Silber_Exp_Start2_B_R2.fastq.gz
Euk_Silber_Exp_Start3_A_R1.fastq.gz
Euk_Silber_Exp_Start3_A_R2.fastq.gz
Euk_Silber_Exp_Start3_B_R1.fastq.gz
Euk_Silber_Exp_Start3_B_R2.fastq.gz
Pool1_NoIndex_L001_R1_001.unassigned.fastq.gz
Pool1_NoIndex_L001_R1_001.unassigned.fastq
Pool1_NoIndex_L001_R1_001.unassigned.log
Pool1_NoIndex_L001_R2_001.unassigned.fastq.gz
Pro_Silber_Exp_Ko1_A_R1.fastq.gz
Pro_Silber_Exp_Ko1_A_R2.fastq.gz
Pro_Silber_Exp_Ko1_B_R1.fastq.gz
Pro_Silber_Exp_Ko1_B_R2.fastq.gz
Pro_Silber_Exp_Ko2_A_R1.fastq.gz
Pro_Silber_Exp_Ko2_A_R2.fastq.gz
Pro_Silber_Exp_Ko2_B_R1.fastq.gz
Pro_Silber_Exp_Ko2_B_R2.fastq.gz
Pro_Silber_Exp_Ko3_A_R1.fastq.gz
Pro_Silber_Exp_Ko3_A_R2.fastq.gz
Pro_Silber_Exp_Ko3_B_R1.fastq.gz
Pro_Silber_Exp_Ko3_B_R2.fastq.gz
Pro_Silber_Exp_NO3_1_A_R1.fastq.gz
Pro_Silber_Exp_NO3_1_A_R2.fastq.gz
Pro_Silber_Exp_NO3_1_B_R1.fastq.gz
Pro_Silber_Exp_NO3_1_B_R2.fastq.gz
Pro_Silber_Exp_NO3_2_A_R1.fastq.gz
Pro_Silber_Exp_NO3_2_A_R2.fastq.gz
Pro_Silber_Exp_NO3_2_B_R1.fastq.gz
Pro_Silber_Exp_NO3_2_B_R2.fastq.gz
Pro_Silber_Exp_NO3_3_A_R1.fastq.gz
Pro_Silber_Exp_NO3_3_A_R2.fastq.gz
Pro_Silber_Exp_NO3_3_B_R1.fastq.gz
Pro_Silber_Exp_NO3_3_B_R2.fastq.gz
Pro_Silber_Exp_NP1_A_R1.fastq.gz
Pro_Silber_Exp_NP1_A_R2.fastq.gz
Pro_Silber_Exp_NP1_B_R1.fastq.gz
Pro_Silber_Exp_NP1_B_R2.fastq.gz
Pro_Silber_Exp_NP2_A_R1.fastq.gz
Pro_Silber_Exp_NP2_A_R2.fastq.gz
Pro_Silber_Exp_NP2_B_R1.fastq.gz
Pro_Silber_Exp_NP2_B_R2.fastq.gz
Pro_Silber_Exp_NP3_A_R1.fastq.gz
Pro_Silber_Exp_NP3_A_R2.fastq.gz
Pro_Silber_Exp_NP3_B_R1.fastq.gz
Pro_Silber_Exp_NP3_B_R2.fastq.gz
Pro_Silber_Exp_Start1_A_R1.fastq.gz
Pro_Silber_Exp_Start1_A_R2.fastq.gz
Pro_Silber_Exp_Start2_A_R1.fastq.gz
Pro_Silber_Exp_Start2_A_R2.fastq.gz
Pro_Silber_Exp_Start3_A_R1.fastq.gz
Pro_Silber_Exp_Start3_A_R2.fastq.gz
Pro_Silber_Exp_Start3_B_R1.fastq.gz
Pro_Silber_Exp_Start3_B_R2.fastq.gz

Illumina data from eurofins

forward reads   reverse reads

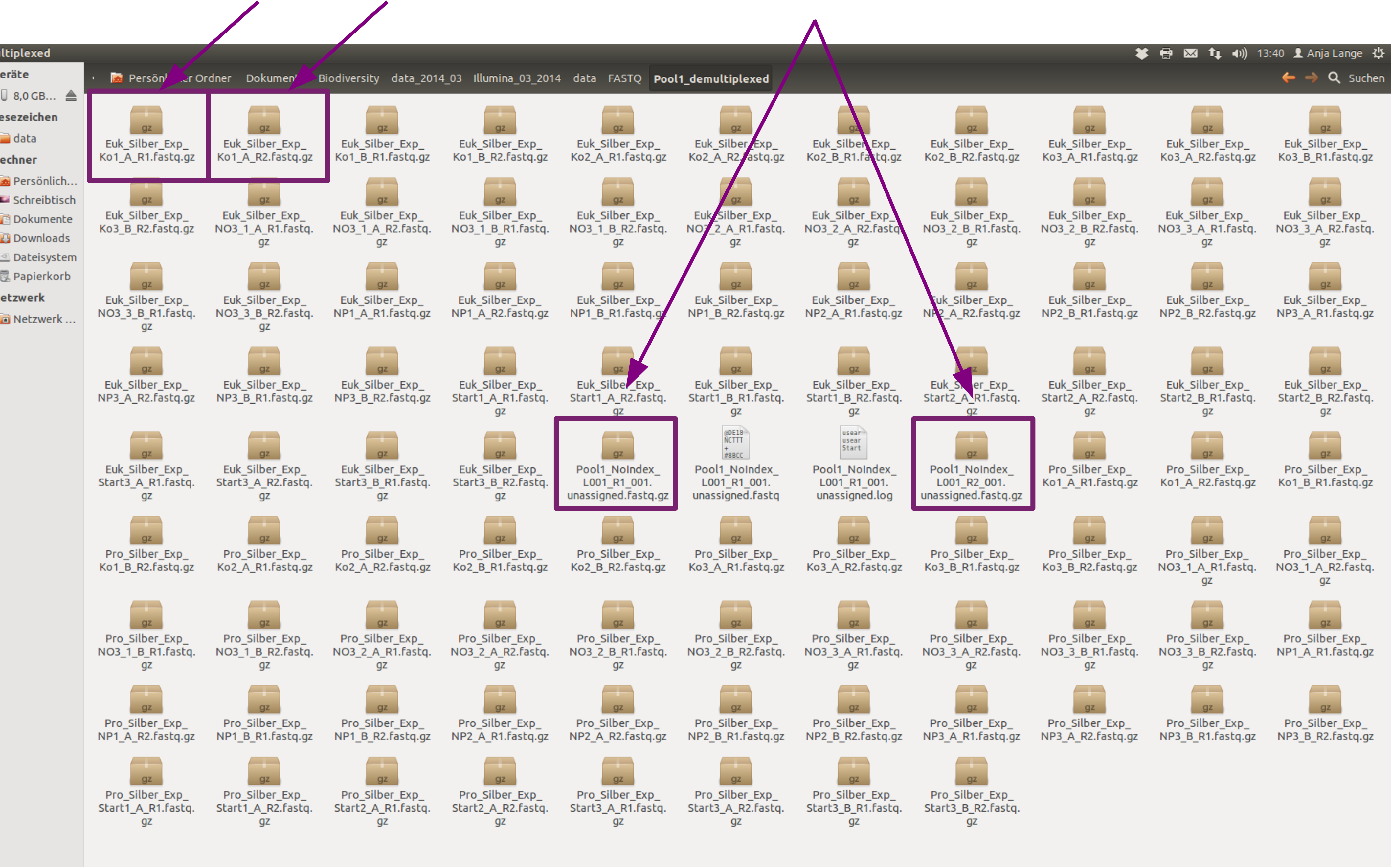


```
Euk_Silber_Exp_Ko1_A_R1.fastq.gz
Euk_Silber_Exp_Ko1_A_R2.fastq.gz
Euk_Silber_Exp_Ko1_B_R1.fastq.gz
Euk_Silber_Exp_Ko1_B_R2.fastq.gz
Euk_Silber_Exp_Ko2_A_R1.fastq.gz
Euk_Silber_Exp_Ko2_A_R2.fastq.gz
Euk_Silber_Exp_Ko2_B_R1.fastq.gz
Euk_Silber_Exp_Ko2_B_R2.fastq.gz
Euk_Silber_Exp_Ko3_A_R1.fastq.gz
Euk_Silber_Exp_Ko3_A_R2.fastq.gz
Euk_Silber_Exp_Ko3_B_R1.fastq.gz
Euk_Silber_Exp_Ko3_B_R2.fastq.gz
Euk_Silber_Exp_NO3_1_A_R1.fastq.gz
Euk_Silber_Exp_NO3_1_A_R2.fastq.gz
Euk_Silber_Exp_NO3_1_B_R1.fastq.gz
Euk_Silber_Exp_NO3_1_B_R2.fastq.gz
Euk_Silber_Exp_NO3_2_A_R1.fastq.gz
Euk_Silber_Exp_NO3_2_A_R2.fastq.gz
Euk_Silber_Exp_NO3_2_B_R1.fastq.gz
Euk_Silber_Exp_NO3_2_B_R2.fastq.gz
Euk_Silber_Exp_NO3_3_A_R1.fastq.gz
Euk_Silber_Exp_NO3_3_A_R2.fastq.gz
Euk_Silber_Exp_NO3_3_B_R1.fastq.gz
Euk_Silber_Exp_NO3_3_B_R2.fastq.gz
Euk_Silber_Exp_NP1_A_R1.fastq.gz
Euk_Silber_Exp_NP1_A_R2.fastq.gz
Euk_Silber_Exp_NP1_B_R1.fastq.gz
Euk_Silber_Exp_NP1_B_R2.fastq.gz
Euk_Silber_Exp_NP2_A_R1.fastq.gz
Euk_Silber_Exp_NP2_A_R2.fastq.gz
Euk_Silber_Exp_NP2_B_R1.fastq.gz
Euk_Silber_Exp_NP2_B_R2.fastq.gz
Euk_Silber_Exp_NP3_A_R1.fastq.gz
Euk_Silber_Exp_NP3_A_R2.fastq.gz
Euk_Silber_Exp_NP3_B_R1.fastq.gz
Euk_Silber_Exp_NP3_B_R2.fastq.gz
Euk_Silber_Exp_Start1_A_R1.fastq.gz
Euk_Silber_Exp_Start1_A_R2.fastq.gz
Euk_Silber_Exp_Start1_B_R1.fastq.gz
Euk_Silber_Exp_Start1_B_R2.fastq.gz
Euk_Silber_Exp_Start2_A_R1.fastq.gz
Euk_Silber_Exp_Start2_A_R2.fastq.gz
Euk_Silber_Exp_Start2_B_R1.fastq.gz
Euk_Silber_Exp_Start2_B_R2.fastq.gz
Euk_Silber_Exp_Start3_A_R1.fastq.gz
Euk_Silber_Exp_Start3_A_R2.fastq.gz
Euk_Silber_Exp_Start3_B_R1.fastq.gz
Euk_Silber_Exp_Start3_B_R2.fastq.gz
Pool1_NoIndex_L001_R1_001.unassigned.fastq.gz
Pool1_NoIndex_L001_R1_001.unassigned.fastq
Pool1_NoIndex_L001_R1_001.unassigned.log
Pool1_NoIndex_L001_R2_001.unassigned.fastq.gz
Pro_Silber_Exp_Ko1_A_R1.fastq.gz
Pro_Silber_Exp_Ko1_A_R2.fastq.gz
Pro_Silber_Exp_Ko1_B_R1.fastq.gz
Pro_Silber_Exp_Ko1_B_R2.fastq.gz
Pro_Silber_Exp_Ko2_A_R1.fastq.gz
Pro_Silber_Exp_Ko2_A_R2.fastq.gz
Pro_Silber_Exp_Ko2_B_R1.fastq.gz
Pro_Silber_Exp_Ko2_B_R2.fastq.gz
Pro_Silber_Exp_Ko3_A_R1.fastq.gz
Pro_Silber_Exp_Ko3_A_R2.fastq.gz
Pro_Silber_Exp_Ko3_B_R1.fastq.gz
Pro_Silber_Exp_Ko3_B_R2.fastq.gz
Pro_Silber_Exp_NO3_1_A_R1.fastq.gz
Pro_Silber_Exp_NO3_1_A_R2.fastq.gz
Pro_Silber_Exp_NO3_1_B_R1.fastq.gz
Pro_Silber_Exp_NO3_1_B_R2.fastq.gz
Pro_Silber_Exp_NO3_2_A_R1.fastq.gz
Pro_Silber_Exp_NO3_2_A_R2.fastq.gz
Pro_Silber_Exp_NO3_2_B_R1.fastq.gz
Pro_Silber_Exp_NO3_2_B_R2.fastq.gz
Pro_Silber_Exp_NO3_3_A_R1.fastq.gz
Pro_Silber_Exp_NO3_3_A_R2.fastq.gz
Pro_Silber_Exp_NO3_3_B_R1.fastq.gz
Pro_Silber_Exp_NO3_3_B_R2.fastq.gz
Pro_Silber_Exp_NP1_A_R1.fastq.gz
Pro_Silber_Exp_NP1_A_R2.fastq.gz
Pro_Silber_Exp_NP1_B_R1.fastq.gz
Pro_Silber_Exp_NP1_B_R2.fastq.gz
Pro_Silber_Exp_NP2_A_R1.fastq.gz
Pro_Silber_Exp_NP2_A_R2.fastq.gz
Pro_Silber_Exp_NP2_B_R1.fastq.gz
Pro_Silber_Exp_NP2_B_R2.fastq.gz
Pro_Silber_Exp_NP3_A_R1.fastq.gz
Pro_Silber_Exp_NP3_A_R2.fastq.gz
Pro_Silber_Exp_NP3_B_R1.fastq.gz
Pro_Silber_Exp_NP3_B_R2.fastq.gz
Pro_Silber_Exp_Start1_A_R1.fastq.gz
Pro_Silber_Exp_Start1_A_R2.fastq.gz
Pro_Silber_Exp_Start2_A_R1.fastq.gz
Pro_Silber_Exp_Start2_A_R2.fastq.gz
Pro_Silber_Exp_Start3_A_R1.fastq.gz
Pro_Silber_Exp_Start3_A_R2.fastq.gz
Pro_Silber_Exp_Start3_B_R1.fastq.gz
Pro_Silber_Exp_Start3_B_R2.fastq.gz
```

Illumina data from eurofins

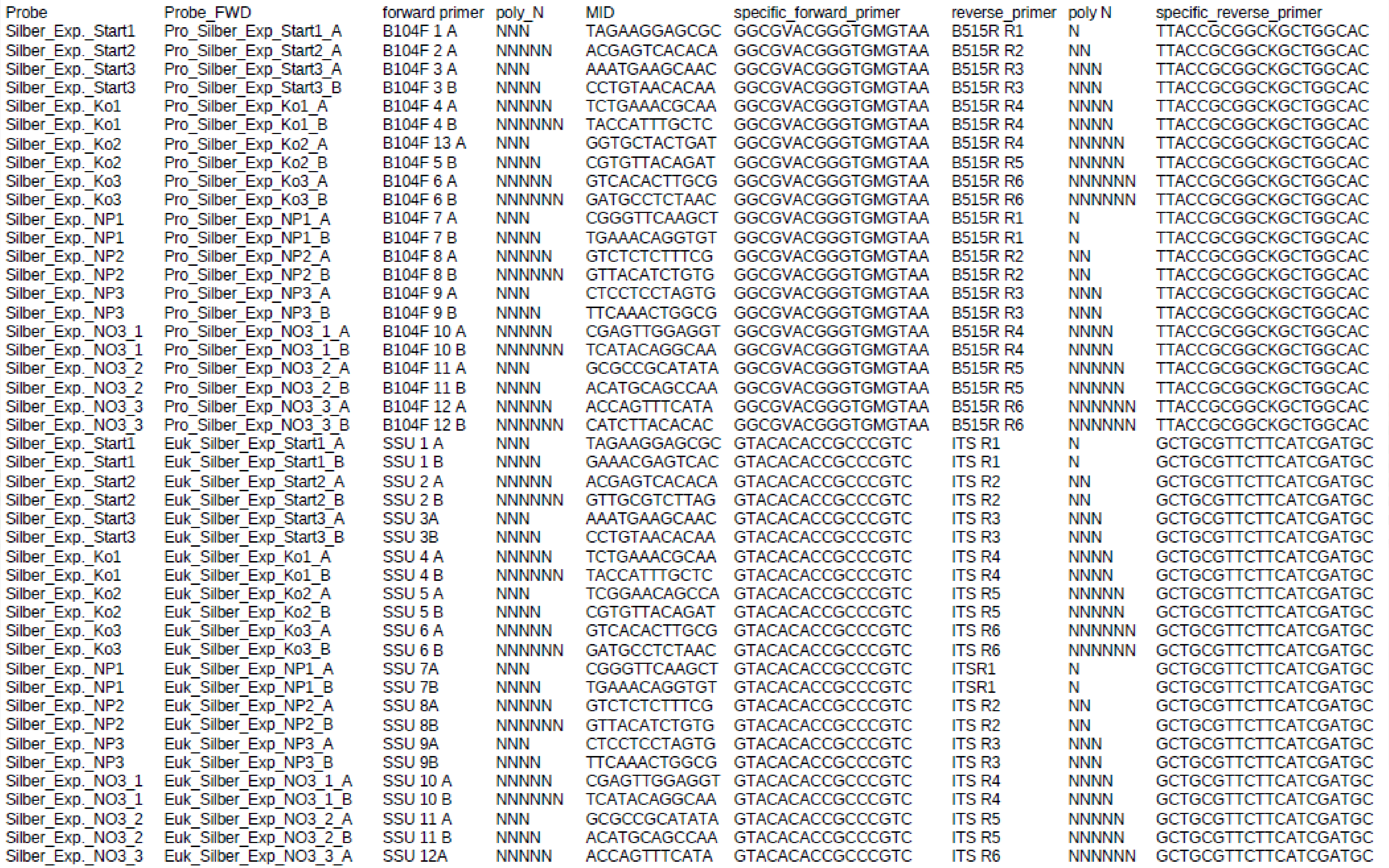
forward reads reverse reads

unassigned reads



Euk_Silber_Exp_Ko1_A_R1.fastq.gz
Euk_Silber_Exp_Ko1_A_R2.fastq.gz
Euk_Silber_Exp_Ko1_B_R1.fastq.gz
Euk_Silber_Exp_Ko1_B_R2.fastq.gz
Euk_Silber_Exp_Ko2_A_R1.fastq.gz
Euk_Silber_Exp_Ko2_A_R2.fastq.gz
Euk_Silber_Exp_Ko2_B_R1.fastq.gz
Euk_Silber_Exp_Ko2_B_R2.fastq.gz
Euk_Silber_Exp_Ko3_A_R1.fastq.gz
Euk_Silber_Exp_Ko3_A_R2.fastq.gz
Euk_Silber_Exp_Ko3_B_R1.fastq.gz
Euk_Silber_Exp_Ko3_B_R2.fastq.gz
Euk_Silber_Exp_NO3_1_A_R1.fastq.gz
Euk_Silber_Exp_NO3_1_A_R2.fastq.gz
Euk_Silber_Exp_NO3_1_B_R1.fastq.gz
Euk_Silber_Exp_NO3_1_B_R2.fastq.gz
Euk_Silber_Exp_NO3_2_A_R1.fastq.gz
Euk_Silber_Exp_NO3_2_A_R2.fastq.gz
Euk_Silber_Exp_NO3_2_B_R1.fastq.gz
Euk_Silber_Exp_NO3_2_B_R2.fastq.gz
Euk_Silber_Exp_NO3_3_A_R1.fastq.gz
Euk_Silber_Exp_NO3_3_A_R2.fastq.gz
Euk_Silber_Exp_NO3_3_B_R1.fastq.gz
Euk_Silber_Exp_NO3_3_B_R2.fastq.gz
Euk_Silber_Exp_NP1_A_R1.fastq.gz
Euk_Silber_Exp_NP1_A_R2.fastq.gz
Euk_Silber_Exp_NP1_B_R1.fastq.gz
Euk_Silber_Exp_NP1_B_R2.fastq.gz
Euk_Silber_Exp_NP2_A_R1.fastq.gz
Euk_Silber_Exp_NP2_A_R2.fastq.gz
Euk_Silber_Exp_NP2_B_R1.fastq.gz
Euk_Silber_Exp_NP2_B_R2.fastq.gz
Euk_Silber_Exp_NP3_A_R1.fastq.gz
Euk_Silber_Exp_NP3_A_R2.fastq.gz
Euk_Silber_Exp_NP3_B_R1.fastq.gz
Euk_Silber_Exp_NP3_B_R2.fastq.gz
Euk_Silber_Exp_Start1_A_R1.fastq.gz
Euk_Silber_Exp_Start1_A_R2.fastq.gz
Euk_Silber_Exp_Start1_B_R1.fastq.gz
Euk_Silber_Exp_Start1_B_R2.fastq.gz
Euk_Silber_Exp_Start2_A_R1.fastq.gz
Euk_Silber_Exp_Start2_A_R2.fastq.gz
Euk_Silber_Exp_Start2_B_R1.fastq.gz
Euk_Silber_Exp_Start2_B_R2.fastq.gz
Euk_Silber_Exp_Start3_A_R1.fastq.gz
Euk_Silber_Exp_Start3_A_R2.fastq.gz
Euk_Silber_Exp_Start3_B_R1.fastq.gz
Euk_Silber_Exp_Start3_B_R2.fastq.gz
Pool1_NoIndex_L001_R1_001.unassigned.fastq.gz
Pool1_NoIndex_L001_R1_001.unassigned.fastq
Pool1_NoIndex_L001_R1_001.unassigned.log
Pool1_NoIndex_L001_R2_001.unassigned.fastq.gz
Pro_Silber_Exp_Ko1_A_R1.fastq.gz
Pro_Silber_Exp_Ko1_A_R2.fastq.gz
Pro_Silber_Exp_Ko1_B_R1.fastq.gz
Pro_Silber_Exp_Ko1_B_R2.fastq.gz
Pro_Silber_Exp_Ko2_A_R1.fastq.gz
Pro_Silber_Exp_Ko2_A_R2.fastq.gz
Pro_Silber_Exp_Ko2_B_R1.fastq.gz
Pro_Silber_Exp_Ko2_B_R2.fastq.gz
Pro_Silber_Exp_Ko3_A_R1.fastq.gz
Pro_Silber_Exp_Ko3_A_R2.fastq.gz
Pro_Silber_Exp_Ko3_B_R1.fastq.gz
Pro_Silber_Exp_Ko3_B_R2.fastq.gz
Pro_Silber_Exp_NO3_1_A_R1.fastq.gz
Pro_Silber_Exp_NO3_1_A_R2.fastq.gz
Pro_Silber_Exp_NO3_1_B_R1.fastq.gz
Pro_Silber_Exp_NO3_1_B_R2.fastq.gz
Pro_Silber_Exp_NO3_2_A_R1.fastq.gz
Pro_Silber_Exp_NO3_2_A_R2.fastq.gz
Pro_Silber_Exp_NO3_2_B_R1.fastq.gz
Pro_Silber_Exp_NO3_2_B_R2.fastq.gz
Pro_Silber_Exp_NO3_3_A_R1.fastq.gz
Pro_Silber_Exp_NO3_3_A_R2.fastq.gz
Pro_Silber_Exp_NO3_3_B_R1.fastq.gz
Pro_Silber_Exp_NO3_3_B_R2.fastq.gz
Pro_Silber_Exp_NP1_A_R1.fastq.gz
Pro_Silber_Exp_NP1_A_R2.fastq.gz
Pro_Silber_Exp_NP1_B_R1.fastq.gz
Pro_Silber_Exp_NP1_B_R2.fastq.gz
Pro_Silber_Exp_NP2_A_R1.fastq.gz
Pro_Silber_Exp_NP2_A_R2.fastq.gz
Pro_Silber_Exp_NP2_B_R1.fastq.gz
Pro_Silber_Exp_NP2_B_R2.fastq.gz
Pro_Silber_Exp_NP3_A_R1.fastq.gz
Pro_Silber_Exp_NP3_A_R2.fastq.gz
Pro_Silber_Exp_NP3_B_R1.fastq.gz
Pro_Silber_Exp_NP3_B_R2.fastq.gz
Pro_Silber_Exp_Start1_A_R1.fastq.gz
Pro_Silber_Exp_Start1_A_R2.fastq.gz
Pro_Silber_Exp_Start2_A_R1.fastq.gz
Pro_Silber_Exp_Start2_A_R2.fastq.gz
Pro_Silber_Exp_Start3_A_R1.fastq.gz
Pro_Silber_Exp_Start3_A_R2.fastq.gz
Pro_Silber_Exp_Start3_B_R1.fastq.gz
Pro_Silber_Exp_Start3_B_R2.fastq.gz

Primer constructs

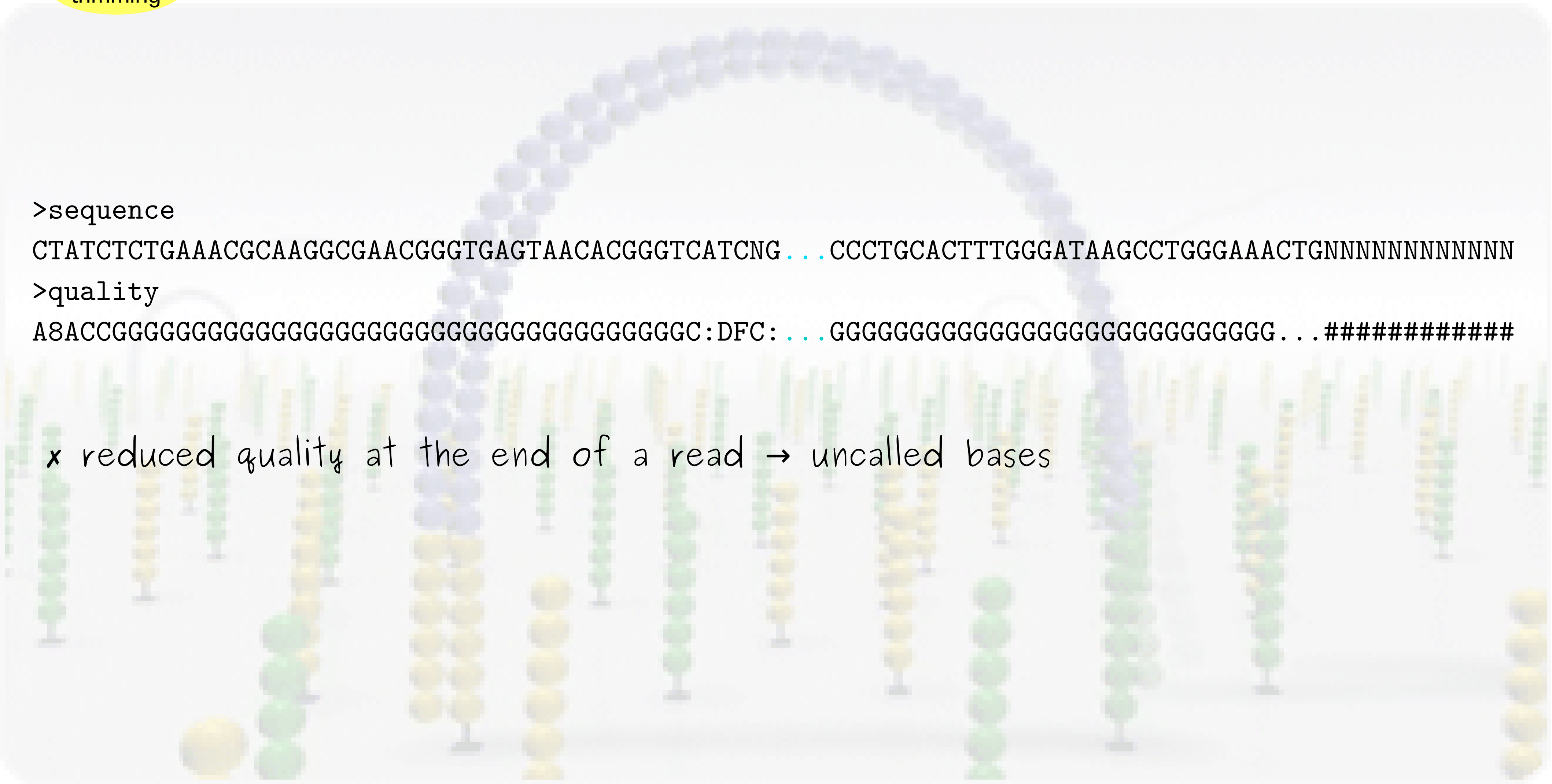


Probe	Probe_FWD	forward primer	poly_N	MID	specific_forward_primer	reverse_primer	poly N	specific_reverse_primer
Silber_Exp._Start1	Pro_Silber_Exp_Start1_A	B104F 1 A	NNN	TAGAAGGAGCGC	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start2	Pro_Silber_Exp_Start2_A	B104F 2 A	NNNNN	ACGAGTCACACA	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_A	B104F 3 A	NNN	AAATGAAGCAAC	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_B	B104F 3 B	NNNN	CCTGTAACACAA	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_A	B104F 4 A	NNNNN	TCTGAAACGCAA	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_B	B104F 4 B	NNNNNN	TACCATTTGCTC	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_A	B104F 13 A	NNN	GGTGCTACTGAT	GGCGVACGGGTGMGTAA	B515R R4	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_B	B104F 5 B	NNNN	CGTGTTACAGAT	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko3	Pro_Silber_Exp_Ko3_A	B104F 6 A	NNNNN	GTCACACTTGCG	GGCGVACGGGTGMGTAA	B515R R6	NNNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko3	Pro_Silber_Exp_Ko3_B	B104F 6 B	NNNNNN	GATGCCTCTAAC	GGCGVACGGGTGMGTAA	B515R R6	NNNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NP1	Pro_Silber_Exp_NP1_A	B104F 7 A	NNN	CGGGTTCAAGCT	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._NP1	Pro_Silber_Exp_NP1_B	B104F 7 B	NNNN	TGAAACAGGTGT	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._NP2	Pro_Silber_Exp_NP2_A	B104F 8 A	NNNNN	GTCTCTCTTTCG	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NP2	Pro_Silber_Exp_NP2_B	B104F 8 B	NNNNNN	GTTACATCTGTG	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NP3	Pro_Silber_Exp_NP3_A	B104F 9 A	NNN	CTCCTCCTAGTG	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NP3	Pro_Silber_Exp_NP3_B	B104F 9 B	NNNN	TTCAAACTGGCG	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NO3_1	Pro_Silber_Exp_NO3_1_A	B104F 10 A	NNNNN	CGAGTTGGAGGT	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NO3_1	Pro_Silber_Exp_NO3_1_B	B104F 10 B	NNNNNN	TCATACAGGCAA	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NO3_2	Pro_Silber_Exp_NO3_2_A	B104F 11 A	NNN	GCGCCGCATATA	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NO3_2	Pro_Silber_Exp_NO3_2_B	B104F 11 B	NNNN	ACATGCAGCCAA	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NO3_3	Pro_Silber_Exp_NO3_3_A	B104F 12 A	NNNNN	ACCAGTTTCATA	GGCGVACGGGTGMGTAA	B515R R6	NNNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NO3_3	Pro_Silber_Exp_NO3_3_B	B104F 12 B	NNNNNN	CATCTTACACAC	GGCGVACGGGTGMGTAA	B515R R6	NNNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start1	Euk_Silber_Exp_Start1_A	SSU 1 A	NNN	TAGAAGGAGCGC	GTACACACCGCCCGTC	ITS R1	N	GCTGCGTTCTTCATCGATGC
Silber_Exp._Start1	Euk_Silber_Exp_Start1_B	SSU 1 B	NNNN	GAAACGAGTCAC	GTACACACCGCCCGTC	ITS R1	N	GCTGCGTTCTTCATCGATGC
Silber_Exp._Start2	Euk_Silber_Exp_Start2_A	SSU 2 A	NNNNN	ACGAGTCACACA	GTACACACCGCCCGTC	ITS R2	NN	GCTGCGTTCTTCATCGATGC
Silber_Exp._Start2	Euk_Silber_Exp_Start2_B	SSU 2 B	NNNNNN	GTTGCGTCTTAG	GTACACACCGCCCGTC	ITS R2	NN	GCTGCGTTCTTCATCGATGC
Silber_Exp._Start3	Euk_Silber_Exp_Start3_A	SSU 3A	NNN	AAATGAAGCAAC	GTACACACCGCCCGTC	ITS R3	NNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._Start3	Euk_Silber_Exp_Start3_B	SSU 3B	NNNN	CCTGTAACACAA	GTACACACCGCCCGTC	ITS R3	NNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._Ko1	Euk_Silber_Exp_Ko1_A	SSU 4 A	NNNNN	TCTGAAACGCAA	GTACACACCGCCCGTC	ITS R4	NNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._Ko1	Euk_Silber_Exp_Ko1_B	SSU 4 B	NNNNNN	TACCATTTGCTC	GTACACACCGCCCGTC	ITS R4	NNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._Ko2	Euk_Silber_Exp_Ko2_A	SSU 5 A	NNN	TCGGAACAGCCA	GTACACACCGCCCGTC	ITS R5	NNNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._Ko2	Euk_Silber_Exp_Ko2_B	SSU 5 B	NNNN	CGTGTTACAGAT	GTACACACCGCCCGTC	ITS R5	NNNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._Ko3	Euk_Silber_Exp_Ko3_A	SSU 6 A	NNNNN	GTCACACTTGCG	GTACACACCGCCCGTC	ITS R6	NNNNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._Ko3	Euk_Silber_Exp_Ko3_B	SSU 6 B	NNNNNN	GATGCCTCTAAC	GTACACACCGCCCGTC	ITS R6	NNNNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._NP1	Euk_Silber_Exp_NP1_A	SSU 7A	NNN	CGGGTTCAAGCT	GTACACACCGCCCGTC	ITSR1	N	GCTGCGTTCTTCATCGATGC
Silber_Exp._NP1	Euk_Silber_Exp_NP1_B	SSU 7B	NNNN	TGAAACAGGTGT	GTACACACCGCCCGTC	ITSR1	N	GCTGCGTTCTTCATCGATGC
Silber_Exp._NP2	Euk_Silber_Exp_NP2_A	SSU 8A	NNNNN	GTCTCTCTTTCG	GTACACACCGCCCGTC	ITS R2	NN	GCTGCGTTCTTCATCGATGC
Silber_Exp._NP2	Euk_Silber_Exp_NP2_B	SSU 8B	NNNNNN	GTTACATCTGTG	GTACACACCGCCCGTC	ITS R2	NN	GCTGCGTTCTTCATCGATGC
Silber_Exp._NP3	Euk_Silber_Exp_NP3_A	SSU 9A	NNN	CTCCTCCTAGTG	GTACACACCGCCCGTC	ITS R3	NNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._NP3	Euk_Silber_Exp_NP3_B	SSU 9B	NNNN	TTCAAACTGGCG	GTACACACCGCCCGTC	ITS R3	NNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._NO3_1	Euk_Silber_Exp_NO3_1_A	SSU 10 A	NNNNN	CGAGTTGGAGGT	GTACACACCGCCCGTC	ITS R4	NNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._NO3_1	Euk_Silber_Exp_NO3_1_B	SSU 10 B	NNNNNN	TCATACAGGCAA	GTACACACCGCCCGTC	ITS R4	NNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._NO3_2	Euk_Silber_Exp_NO3_2_A	SSU 11 A	NNN	GCGCCGCATATA	GTACACACCGCCCGTC	ITS R5	NNNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._NO3_2	Euk_Silber_Exp_NO3_2_B	SSU 11 B	NNNN	ACATGCAGCCAA	GTACACACCGCCCGTC	ITS R5	NNNNN	GCTGCGTTCTTCATCGATGC
Silber_Exp._NO3_3	Euk_Silber_Exp_NO3_3_A	SSU 12A	NNNNN	ACCAGTTTCATA	GTACACACCGCCCGTC	ITS R6	NNNNNN	GCTGCGTTCTTCATCGATGC

The pipeline

single-read mode

poly-N trimming



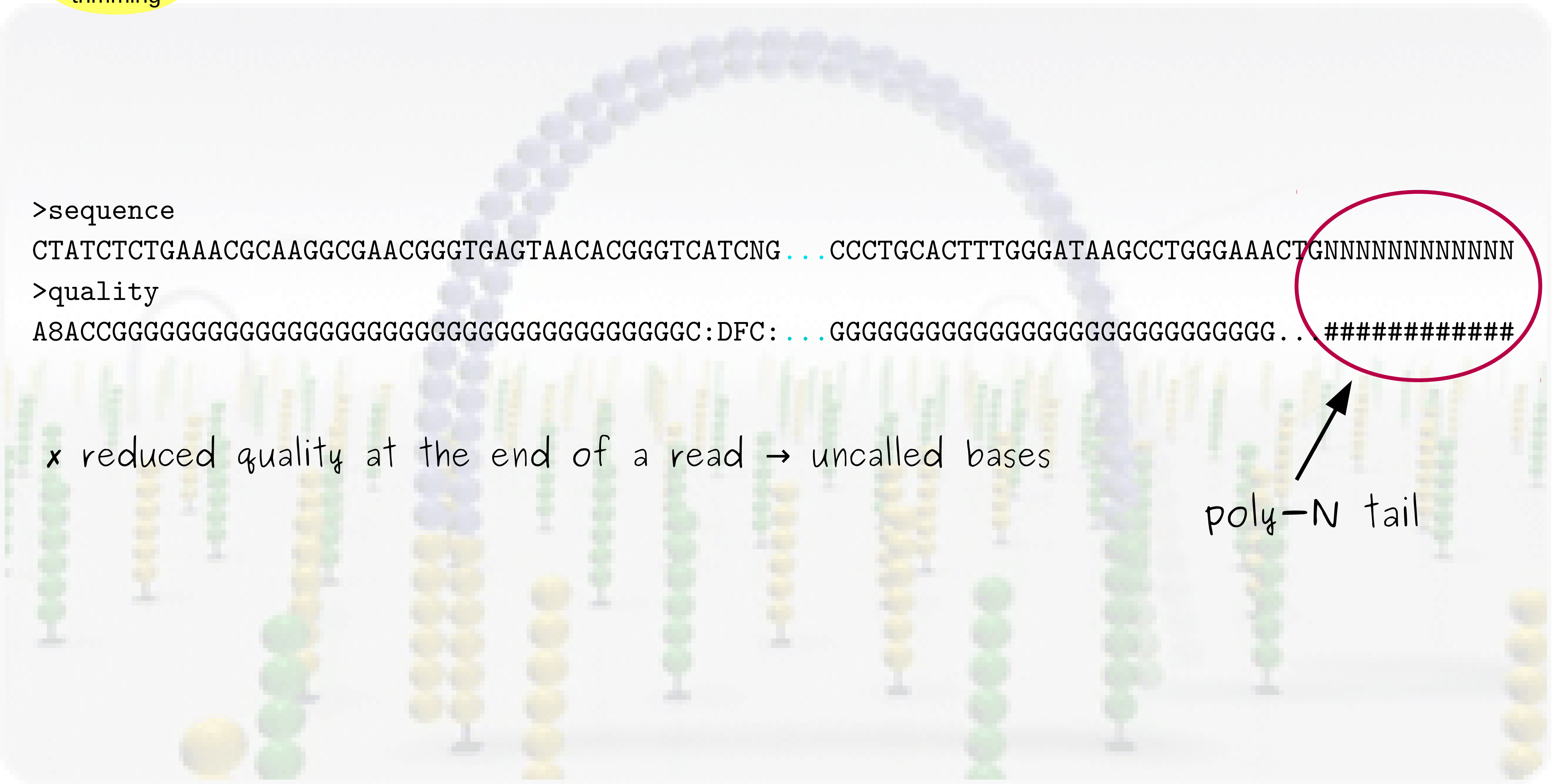
```
>sequence
CTATCTCTGAAACGCAAGGCGAACGGGTGAGTAACACGGGTCATCNG...CCCTGCACTTTGGGATAAGCCTGGGAAACTGNNNNNNNNNNNN
>quality
A8ACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC:DFC:...GGGGGGGGGGGGGGGGGGGGGGGGGGGG...############
```

✗ reduced quality at the end of a read → uncalled bases

The pipeline

single-read mode

poly-N trimming



```
>sequence
CTATCTCTGAAACGCAAGGCGAACGGGTGAGTAACACGGGTCATCNG...CCCTGCACTTTGGGATAAGCCTGGGAAACTGNNNNNNNNNNNN
>quality
A8ACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC:DFC:...GGGGGGGGGGGGGGGGGGGGGGGGGGGG...############
```

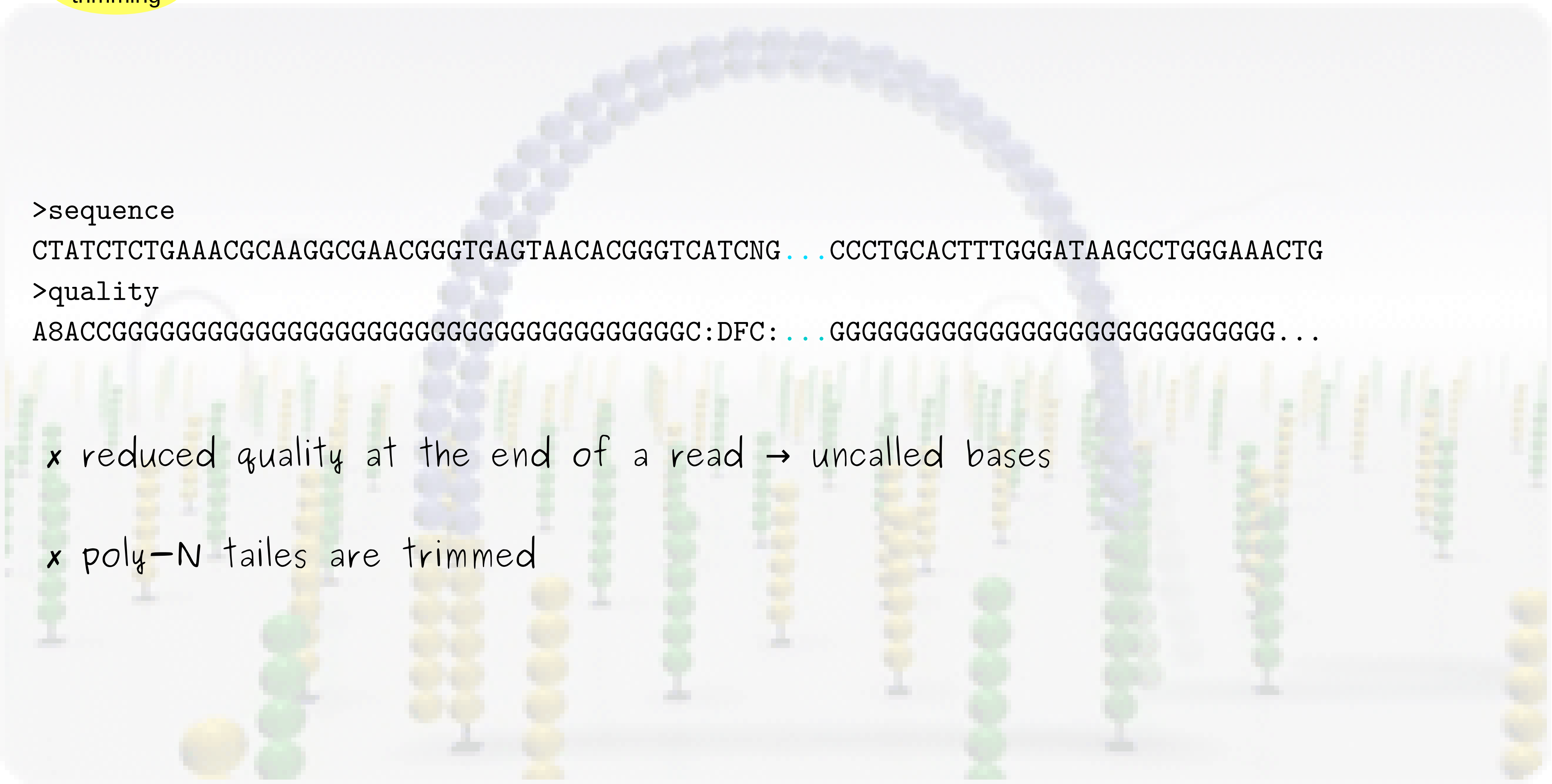
✗ reduced quality at the end of a read → uncalled bases

poly-N tail

The pipeline

single-read mode

poly-N trimming



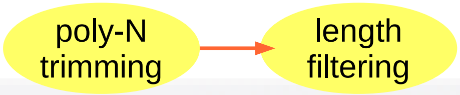
```
>sequence
CTATCTCTGAAACGCAAGGCGAACGGGTGAGTAACACGGGTCATCNG...CCCTGCACTTTGGGATAAGCCTGGGAAACTG
>quality
A8ACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC:DFC:...GGGGGGGGGGGGGGGGGGGGGGGGGGGG...
```

✗ reduced quality at the end of a read → uncalled bases

✗ poly-N tailes are trimmed

The pipeline

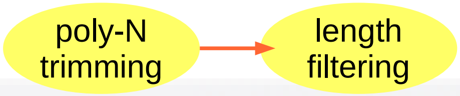
single-read mode

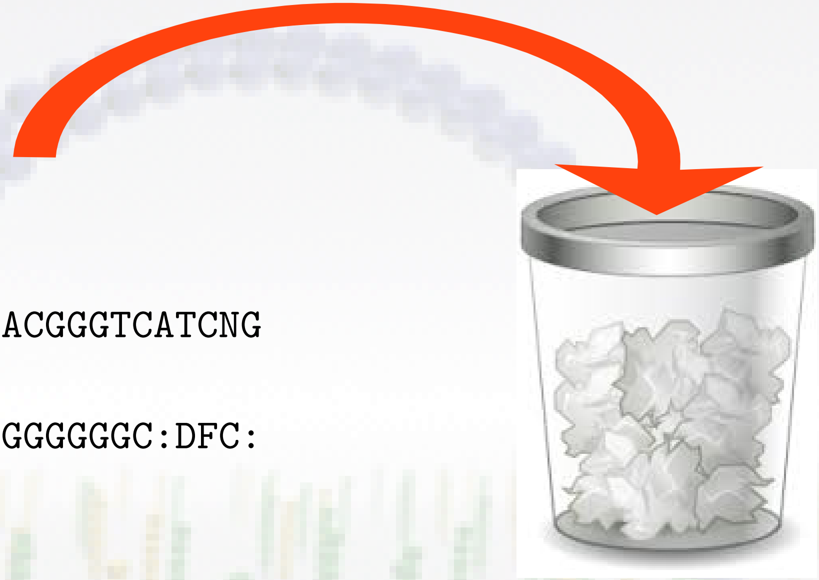


```
>sequence
CTATCTCTGAAACGCAAGGCGAACGGGTGAGTAACACGGGTCATCNG
>quality
A8ACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC:DFC:
```

The pipeline

single-read mode



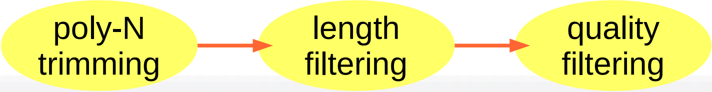


```
>sequence
CTATCTCTGAAACGCAAGGCGAACGGGTGAGTAACACGGGTCATCNG
>quality
A8ACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC:DFC:
```

✗ sequences below a certain length cutoff are discarded

The pipeline

single-read mode

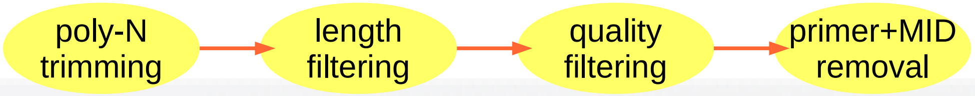


```
>sequence
CTATCTCTGAAACGCAAGGCGAACGGGTGAGTAACACGGGTCATCNG...CCCTGCACTTTGGGATAAGCCTGGGAAACTG
>quality
A8ACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC:DFC:...GGGGGGGGGGGGGGGGGGGGGGGGGGG...
```

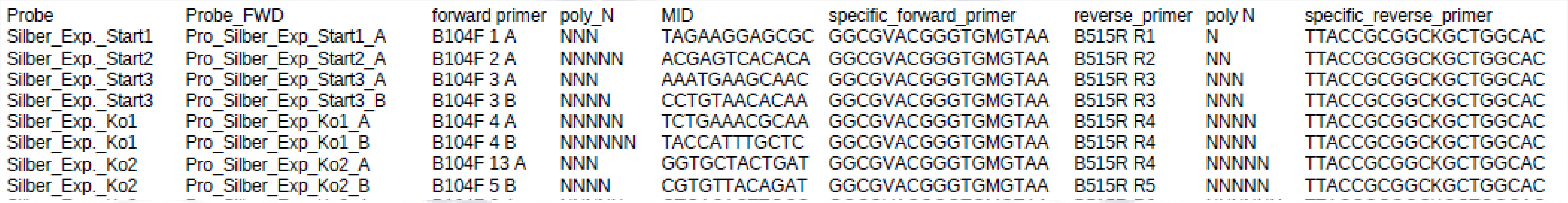
- ✗ calculates for each sequence the mean Phread score (mPs)
- ✗ determines the lowest Phread score for a base in the sequences (lPs)
- ✗ if mPs < a given treshold **OR** lPs < a given treshold → sequence is discarded
- ✗ sequences are saved as fasta, quality values are no longer required

The pipeline

single-read mode



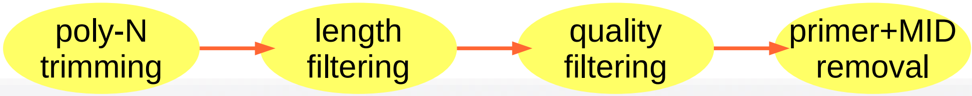
poly-N trimming → length filtering → quality filtering → primer+MID removal

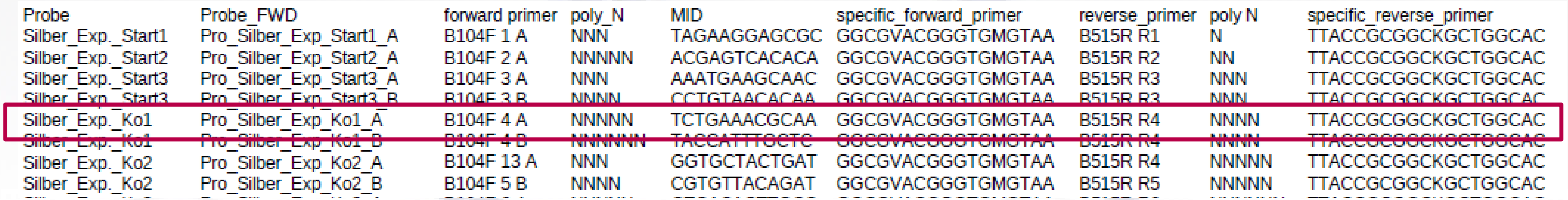


Probe	Probe_FWD	forward primer	poly_N	MID	specific_forward_primer	reverse_primer	poly N	specific_reverse_primer
Silber_Exp._Start1	Pro_Silber_Exp_Start1_A	B104F 1 A	NNN	TAGAAGGAGCGC	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start2	Pro_Silber_Exp_Start2_A	B104F 2 A	NNNNN	ACGAGTCACACA	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_A	B104F 3 A	NNN	AAATGAAGCAAC	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_B	B104F 3 B	NNNN	CCTGTAACACAA	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_A	B104F 4 A	NNNNN	TCTGAAACGCAA	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_B	B104F 4 B	NNNNNN	TACCATTTGCTC	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_A	B104F 13 A	NNN	GGTGCTACTGAT	GGCGVACGGGTGMGTAA	B515R R4	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_B	B104F 5 B	NNNN	CGTGTTACAGAT	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC

The pipeline

single-read mode


poly-N trimming → length filtering → quality filtering → primer+MID removal

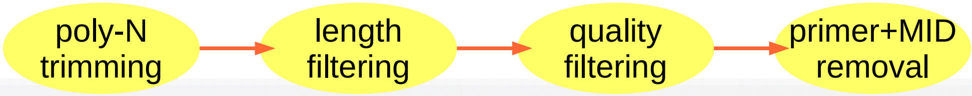


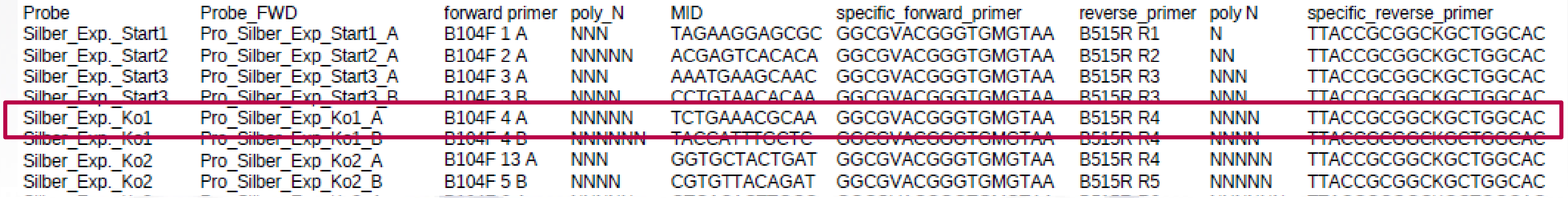
Probe	Probe_FWD	forward primer	poly_N	MID	specific_forward_primer	reverse_primer	poly N	specific_reverse_primer
Silber_Exp._Start1	Pro_Silber_Exp_Start1_A	B104F 1 A	NNN	TAGAAGGAGCGC	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start2	Pro_Silber_Exp_Start2_A	B104F 2 A	NNNNN	ACGAGTCACACA	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_A	B104F 3 A	NNN	AAATGAAGCAAC	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_B	B104F 3 B	NNNN	CCTGTAACACAA	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_A	B104F 4 A	NNNNN	TCTGAAACGCAA	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_B	B104F 4 B	NNNNNN	TACCATTTGCTC	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_A	B104F 13 A	NNN	GGTGCTACTGAT	GGCGVACGGGTGMGTAA	B515R R4	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_B	B104F 5 B	NNNN	CGTGTTACAGAT	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC

- poly-N + MID + primer: NNNNNTCTGAAACGCAAGGCGVACGGGTGMGTAA → 34 nt
- looks for exact match in reads

The pipeline

single-read mode





Probe	Probe_FWD	forward primer	poly_N	MID	specific_forward_primer	reverse_primer	poly N	specific_reverse_primer
Silber_Exp._Start1	Pro_Silber_Exp_Start1_A	B104F 1 A	NNN	TAGAAGGAGCGC	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start2	Pro_Silber_Exp_Start2_A	B104F 2 A	NNNNN	ACGAGTCACACA	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_A	B104F 3 A	NNN	AAATGAAGCAAC	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_B	B104F 3 B	NNNN	CCTGTAACACAA	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_A	B104F 4 A	NNNNN	TCTGAAACGCAA	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_B	B104F 4 B	NNNNNN	TACCATTTGCTC	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_A	B104F 13 A	NNN	GGTGCTACTGAT	GGCGVACGGGTGMGTAA	B515R R4	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_B	B104F 5 B	NNNN	CGTGTTACAGAT	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC

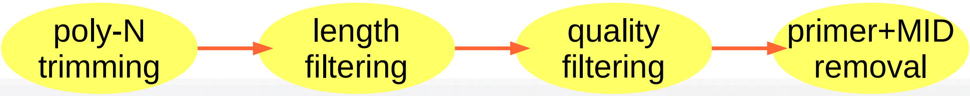
- poly-N + MID + primer: NNNNNTCTGAAACGCAAGGCGVACGGGTGMGTAA → 34 nt
- looks for exact match in reads

```
width seq
```

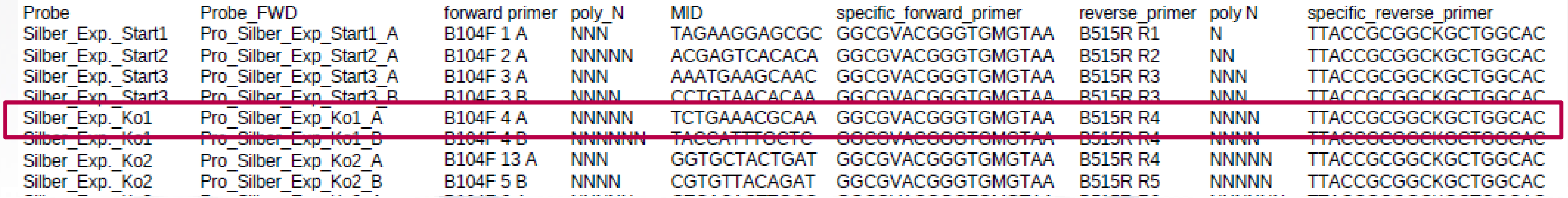
$$300\ \underbrace{\texttt{AGTGA}}_{\text{poly-N}}\underbrace{\texttt{TCTGAAACGCAA}}_{\text{MID}}\underbrace{\texttt{GGCGGACGGGTGAGTAA}}_{\text{primer}}\texttt{TACATCGGAACGTACCTTATCGTGGGGGATAACGCAGCGAAAGCTG...}$$

The pipeline

single-read mode



poly-N trimming → length filtering → quality filtering → primer+MID removal



Probe	Probe_FWD	forward primer	poly_N	MID	specific_forward_primer	reverse_primer	poly N	specific_reverse_primer
Silber_Exp._Start1	Pro_Silber_Exp_Start1_A	B104F 1 A	NNN	TAGAAGGAGCGC	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start2	Pro_Silber_Exp_Start2_A	B104F 2 A	NNNNN	ACGAGTCACACA	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_A	B104F 3 A	NNN	AAATGAAGCAAC	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_B	B104F 3 B	NNNN	CCTGTAACACAA	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_A	B104F 4 A	NNNNN	TCTGAAACGCAA	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_B	B104F 4 B	NNNNNN	TACCATTTGCTC	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_A	B104F 13 A	NNN	GGTGCTACTGAT	GGCGVACGGGTGMGTAA	B515R R4	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_B	B104F 5 B	NNNN	CGTGTTACAGAT	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC

- ✗ poly-N + MID + primer: NNNNNTCTGAAACGCAAGGCGVACGGGTGMGTAA → 34 nt
- ✗ looks for exact match in reads

```
width seq
300 AGTGATCTGAAACGCAAGGCGGACGGGTGAGTAATACATCGGAACGTACCTTATCGTGGGGGATAACGCAGCGAAAGCTG...
```

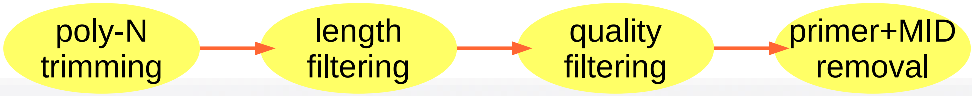
poly-N MID primer

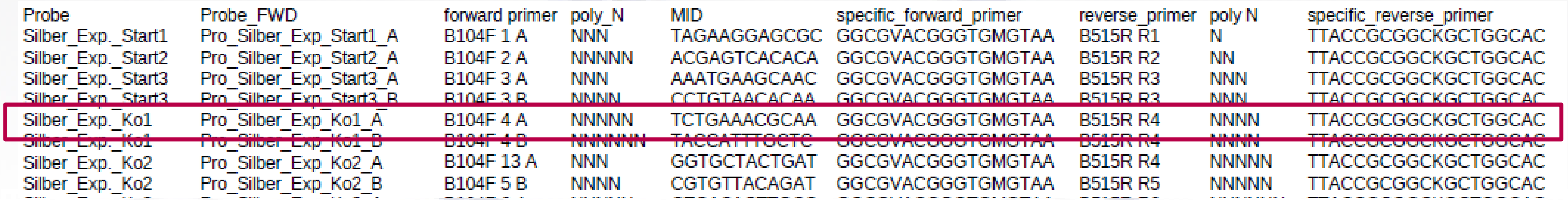
```
width seq
300 GGCCTTCTGAAACGCAAGGCGCACGGGTGAGTAACGCGTAAGAATCTAACTTCAGGACGGGGACAACAGTTGGAAACGAC...
```

poly-N MID primer

The pipeline

single-read mode





Probe	Probe_FWD	forward primer	poly_N	MID	specific_forward_primer	reverse_primer	poly N	specific_reverse_primer
Silber_Exp._Start1	Pro_Silber_Exp_Start1_A	B104F 1 A	NNN	TAGAAGGAGCGC	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start2	Pro_Silber_Exp_Start2_A	B104F 2 A	NNNNN	ACGAGTCACACA	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_A	B104F 3 A	NNN	AAATGAAGCAAC	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_B	B104F 3 B	NNNN	CCTGTAACACAA	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_A	B104F 4 A	NNNNN	TCTGAAACGCAA	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_B	B104F 4 B	NNNNNN	TACCATTTGCTC	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_A	B104F 13 A	NNN	GGTGCTACTGAT	GGCGVACGGGTGMGTAA	B515R R4	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_B	B104F 5 B	NNNN	CGTGTTACAGAT	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC

- poly-N + MID + primer: NNNNNTCTGAAACGCAAGGCGVACGGGTGMGTAA → 34 nt
- looks for exact match in reads

```
width seq
```

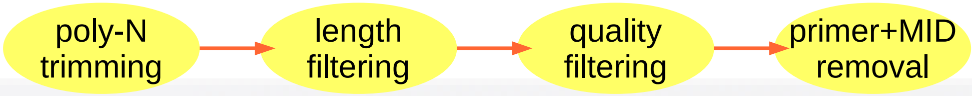
$$300\ \underbrace{\texttt{AGTGA}}_{\text{poly-N}}\underbrace{\texttt{TCTGAAACGCAA}}_{\text{MID}}\underbrace{\texttt{GGCGGACGGGTGAGTAA}}_{\text{primer}}\texttt{TACATCGGAACGTACCTTATCGTGGGGATAACGCAGCGAAAGCTG...}$$

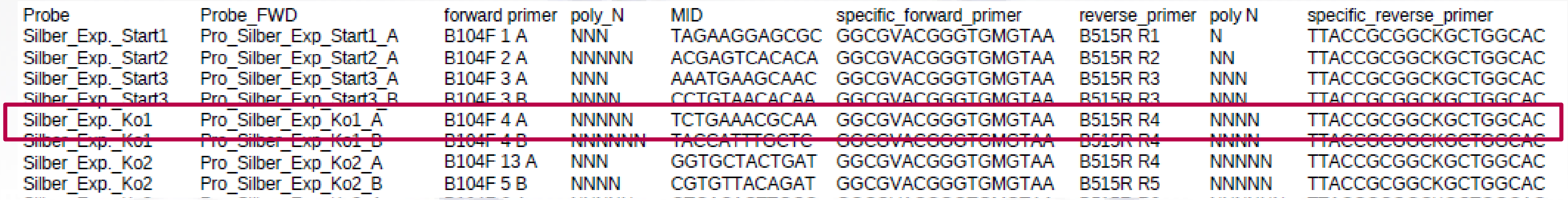
```
width seq
```

$$300\ \underbrace{\texttt{GGCCT}}_{\text{poly-N}}\underbrace{\texttt{TCTGAAACGCAA}}_{\text{MID}}\underbrace{\texttt{GGCGCACGGGTGAGTAA}}_{\text{primer}}\texttt{CGCGTAAGAATCTAACTTCAGGACGGGGACAACAGTTGGAAACGAC...}$$

The pipeline

single-read mode





Probe	Probe_FWD	forward primer	poly_N	MID	specific_forward_primer	reverse_primer	poly N	specific_reverse_primer
Silber_Exp._Start1	Pro_Silber_Exp_Start1_A	B104F 1 A	NNN	TAGAAGGAGCGC	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start2	Pro_Silber_Exp_Start2_A	B104F 2 A	NNNNN	ACGAGTCACACA	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_A	B104F 3 A	NNN	AAATGAAGCAAC	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_B	B104F 3 B	NNNN	CCTGTAACACAA	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_A	B104F 4 A	NNNNN	TCTGAAACGCAA	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_B	B104F 4 B	NNNNNN	TACCATTTGCTC	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_A	B104F 13 A	NNN	GGTGCTACTGAT	GGCGVACGGGTGMGTAA	B515R R4	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_B	B104F 5 B	NNNN	CGTGTTACAGAT	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC

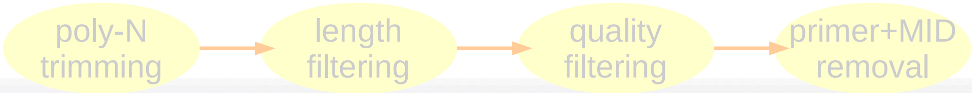
- poly-N + MID + primer: NNNNNTCTGAAACGCAAGGCGVACGGGTGMGTAA → 34 nt
- looks for exact match in reads

```
width seq
266   TACATCGGAACGTACCTTATCGTGGGGATAACGCAGCGAAAGCTG...
```

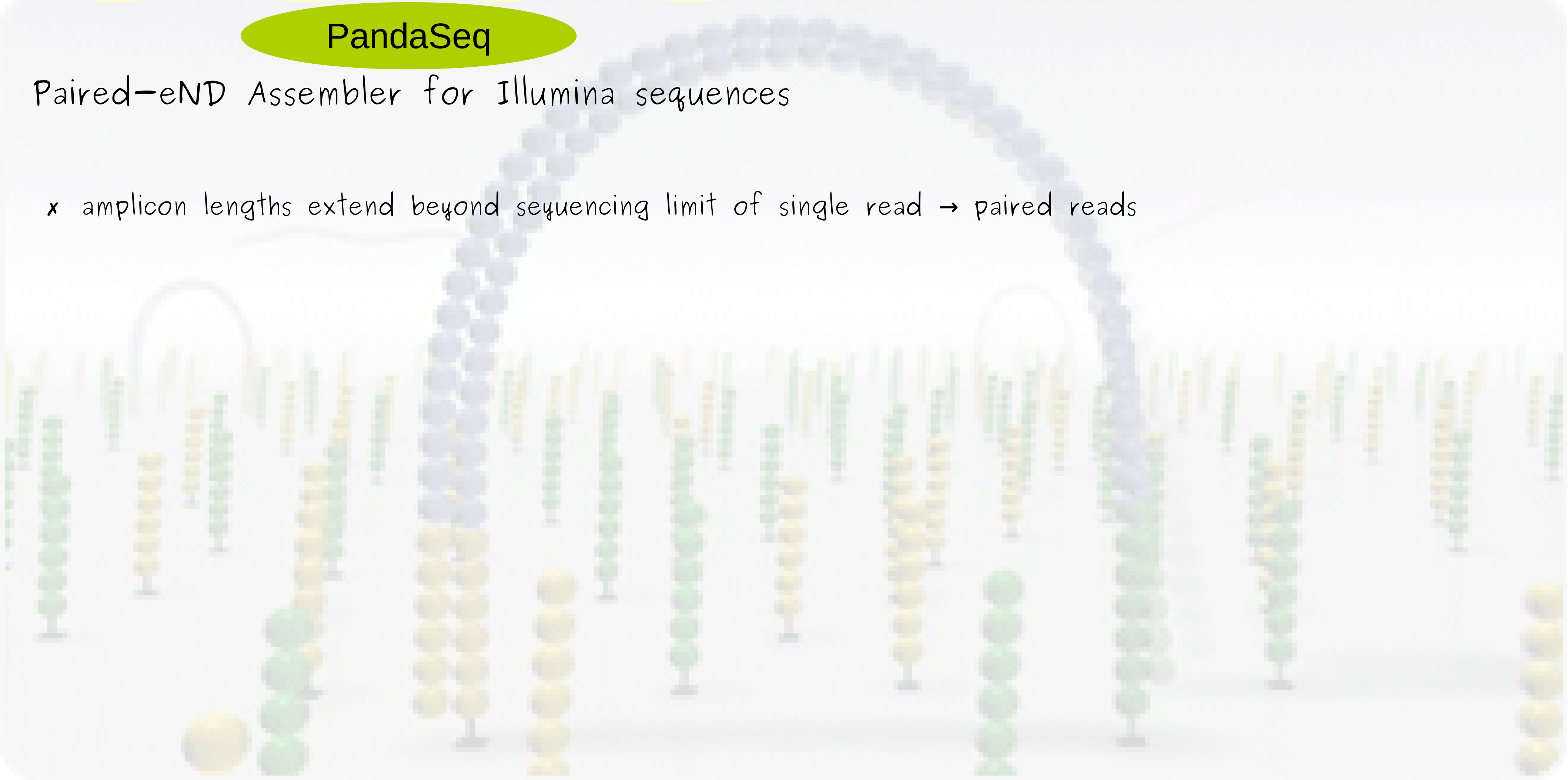
```
width seq
266   CGCGTAAGAATCTAACTTCAGGACGGGGACAACAGTTGGAAACGAC...
```

The pipeline

paired-end mode


poly-N trimming → length filtering → quality filtering → primer+MID removal

PandaSeq

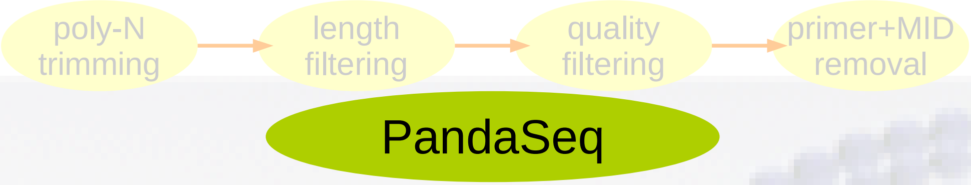


Paired-eND Assembler for Illumina sequences

✗ amplicon lengths extend beyond seyuencing limit of single read → paired reads

The pipeline

paired-end mode



Paired-eND Assembler for Illumina sequences

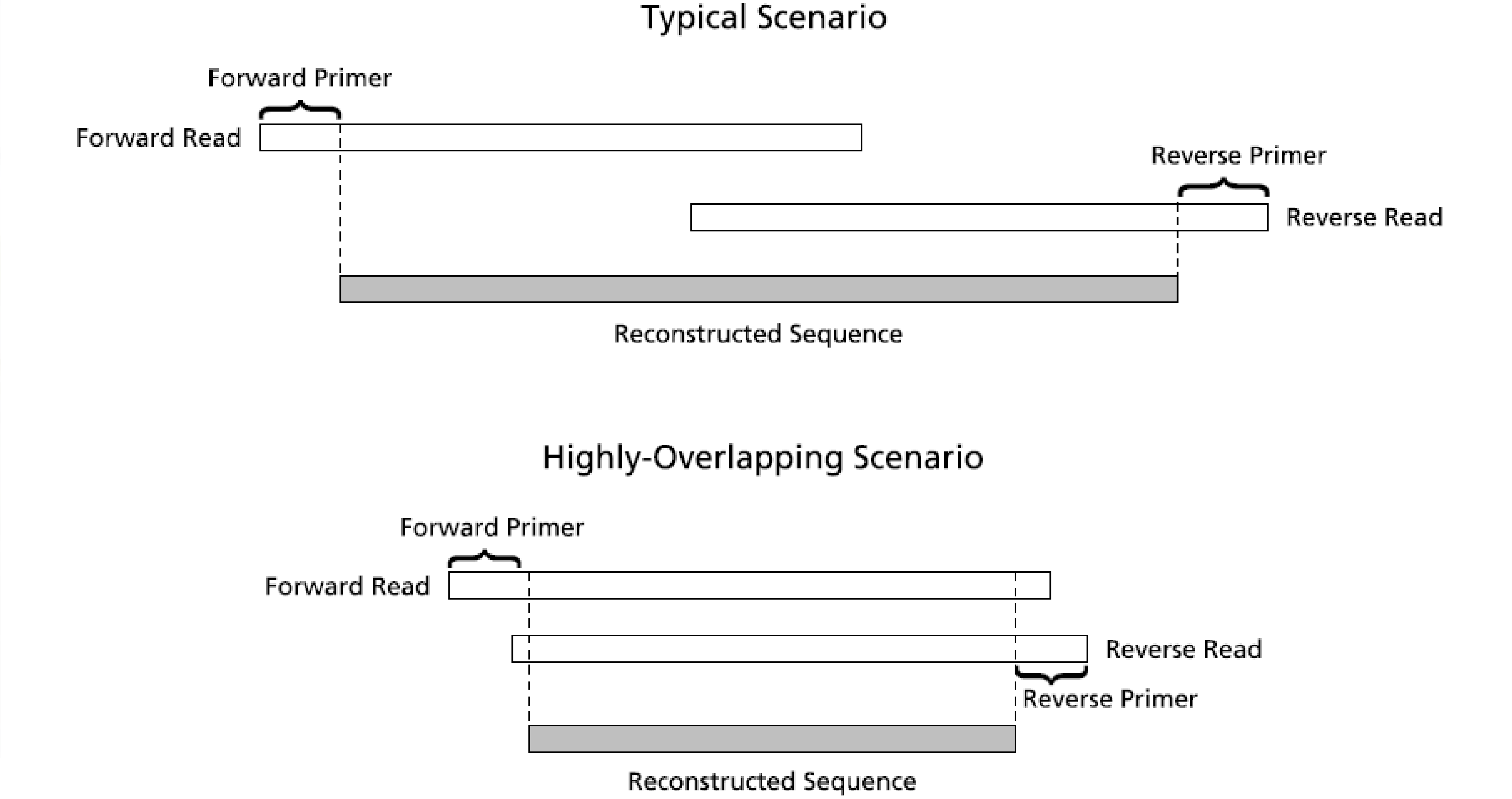
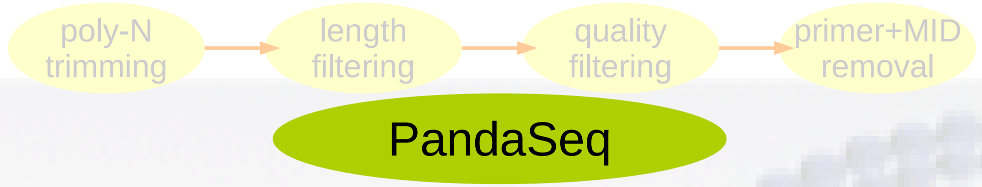


Figure 1 Schematic of paired-end assembly. Typical scenario: forward and reverse reads are overlapped and the primer regions are removed to reconstruct the sequences. Highly overlapping scenario: for short templates, the overlapping region may include the primer regions.

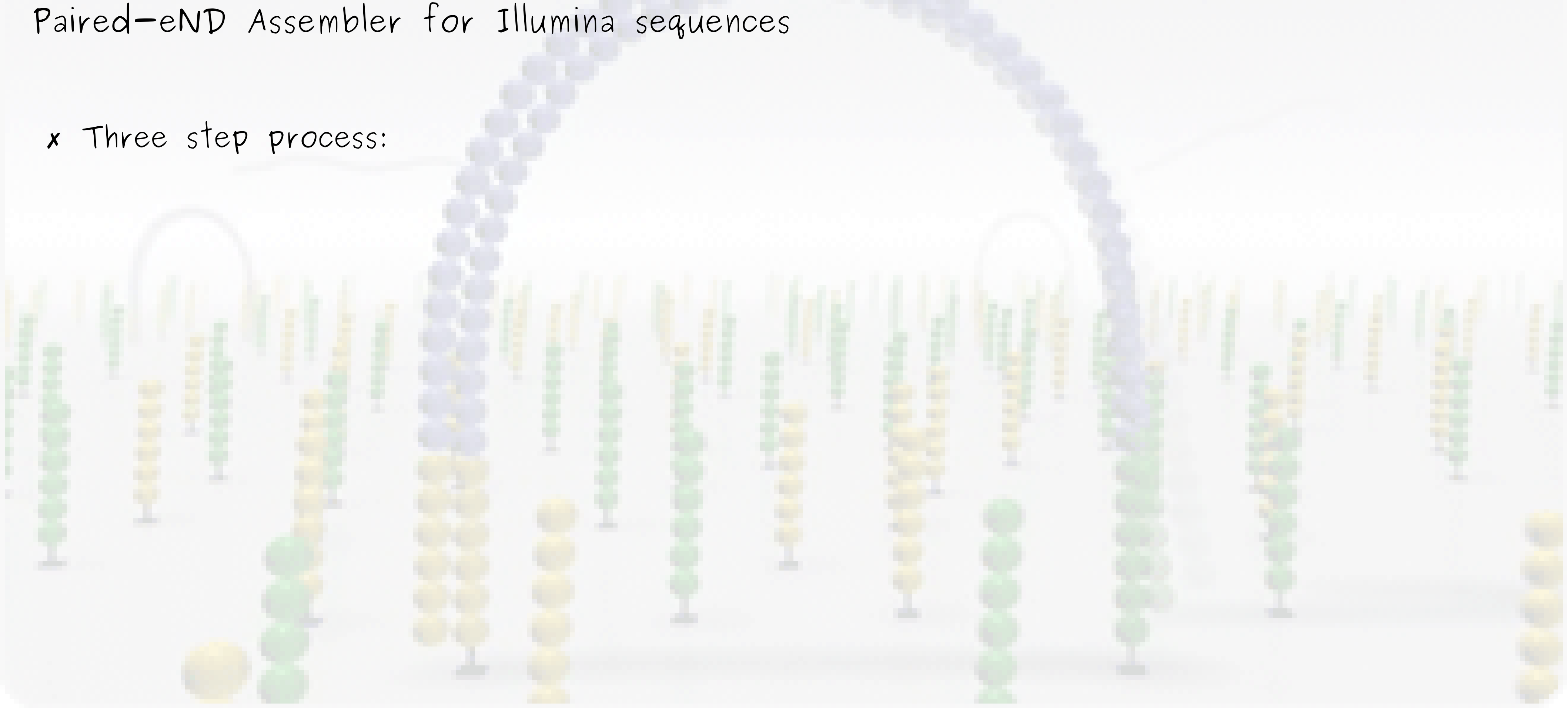
Masella et al. BMC Bioinformatics 2012, 13:31

The pipeline

paired-end mode



Paired-eND Assembler for Illumina sequences

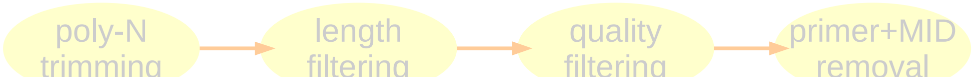


- ✗ Three step process:

Masella et al. BMC Bioinformatics 2012, 13:31

The pipeline

paired-end mode



PandaSeq

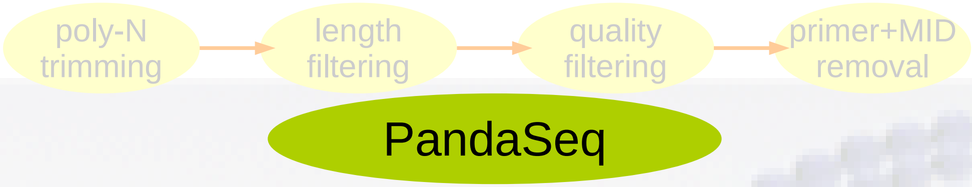
Paired-eND Assembler for Illumina sequences

- Three step process:
  - locates sequencing primers

Masella et al. BMC Bioinformatics 2012, 13:31

The pipeline

paired-end mode



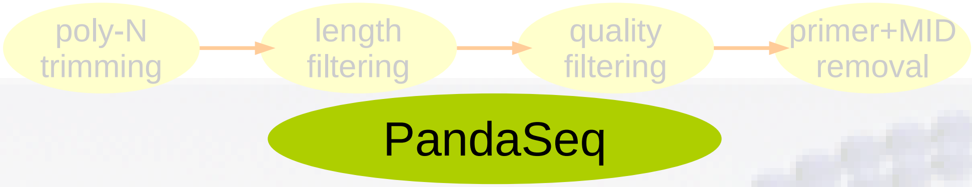
Paired-eND Assembler for Illumina sequences

✗ Three step process:

- ✗ locates sequencing primers
- ✗ identifies optimal overlap
  - Uses the Phred values to estimate the probabilities that
    - a) the true bases match, given the sequenced bases mismatch
    - b) the true bases match, given the sequenced bases match
    - c) the true bases match, given that one of the bases is uncalled

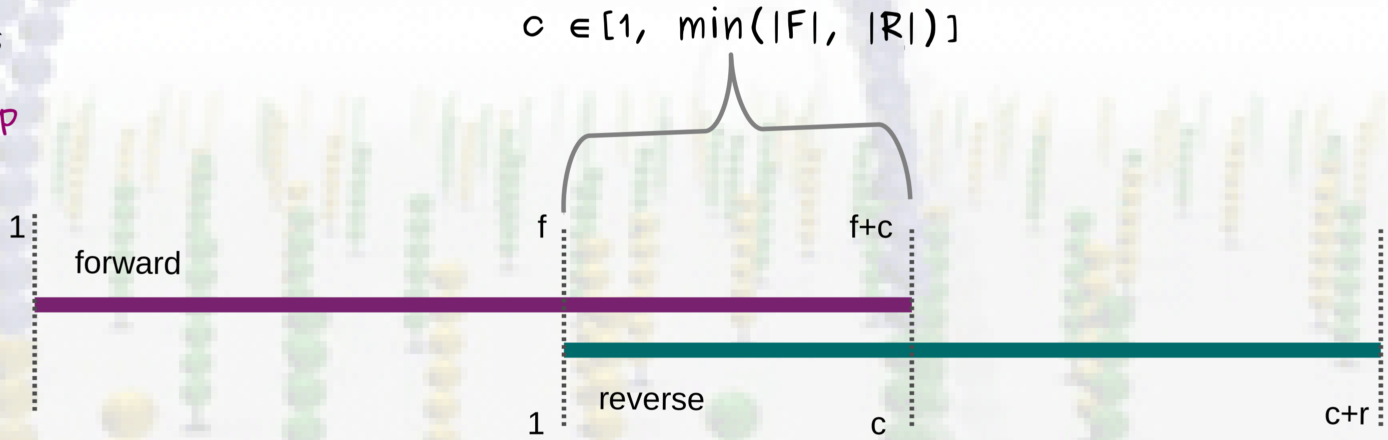
The pipeline

paired-end mode



Paired-eND Assembler for Illumina sequences

- Three step process:
  - locates sequencing primers
  - identifies optimal overlap



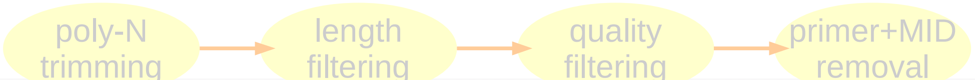
c, the range of overlap is choosen to maximize:

$$\Pr[F,R|c] = \prod_{i=1\ldots f} \Pr[F_i] \cdot \prod_{i=1\ldots c} \Pr[\hat{F}_{i+f} = \hat{R}_i] \cdot \prod_{i=1\ldots r} \Pr[R_{i+c}]$$

Masella et al. BMC Bioinformatics 2012, 13:31

The pipeline

paired-end mode



PandaSeq

Paired-eND Assembler for Illumina sequences

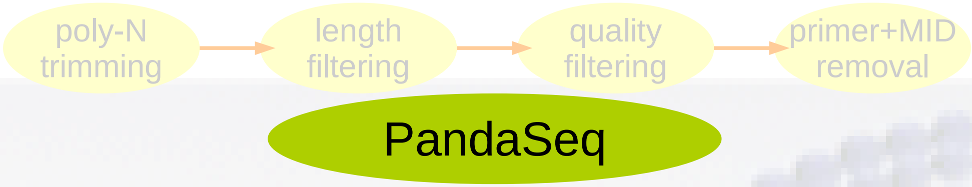
- Three step process:
  - locates sequencing primers
  - identifies optimal overlap
  - reconstructs complete sequence

- unpaired regions are copied
- overlapping regions:
  - quality score is corrected
  - if bases don't match, base with higher quality score is choosen
- calculates an overall quality score
- primer are removed

Masella et al. BMC Bioinformatics 2012, 13:31

The pipeline

paired-end mode

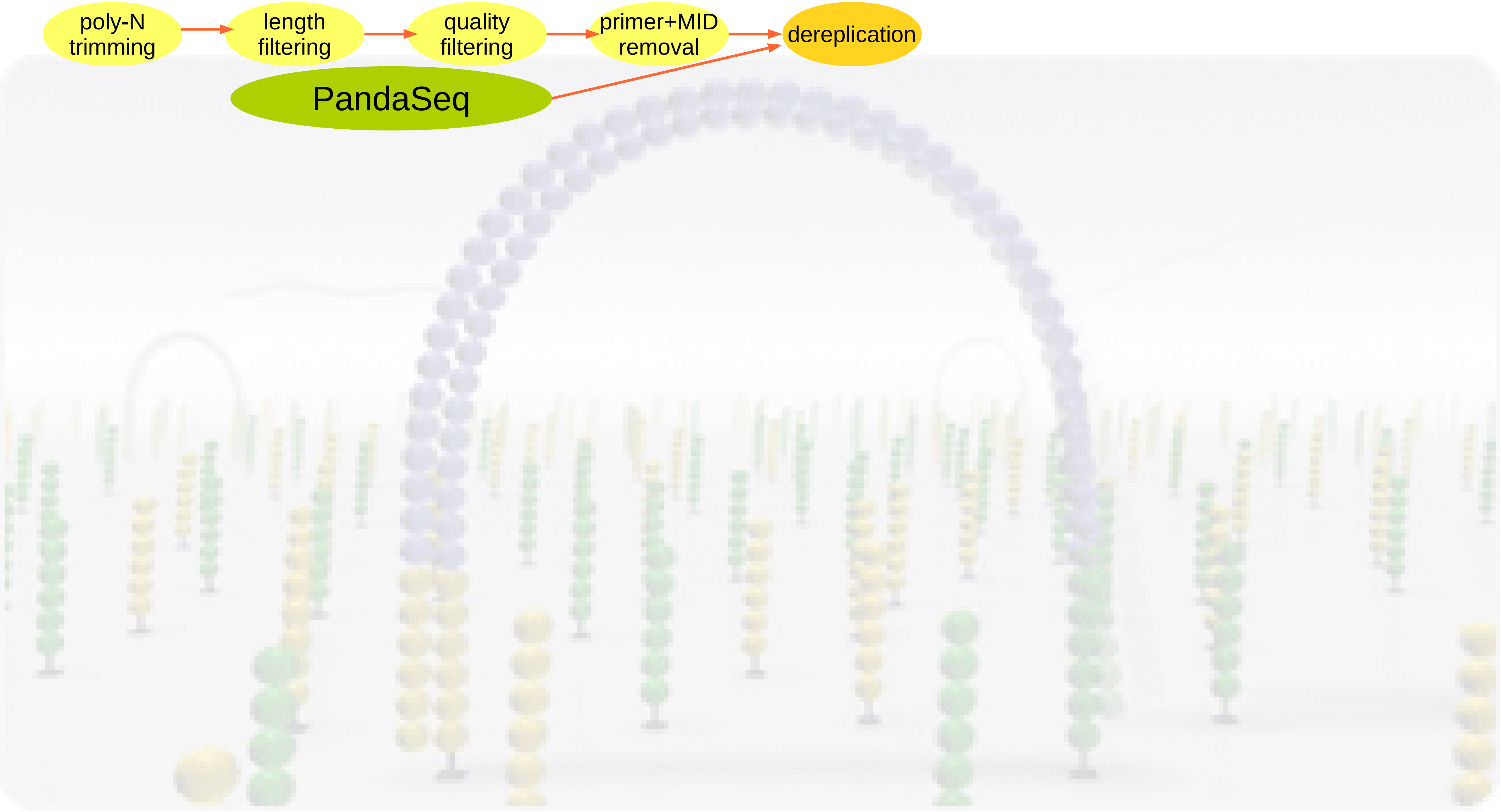


Paired-eND Assembler for Illumina sequences

- ✗ Three (Four) step process:
  - ✗ locates sequencing primers
  - ✗ identifies optimal overlap
  - ✗ reconstructs complete sequence
  - ✗ rejects sequences based on user specified parameters
    - low quality score
    - length of assembled sequence
    - length of overlap
    - presence of uncalled bases

Masella et al. BMC Bioinformatics 2012, 13:31

The pipeline



poly-N trimming

length filtering

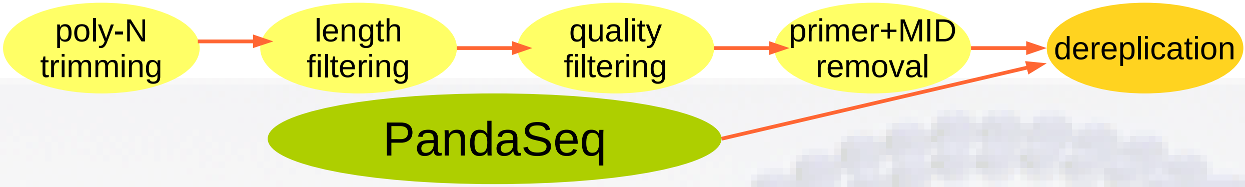
quality filtering

primer+MID removal

dereplication

PandaSeq

The pipeline

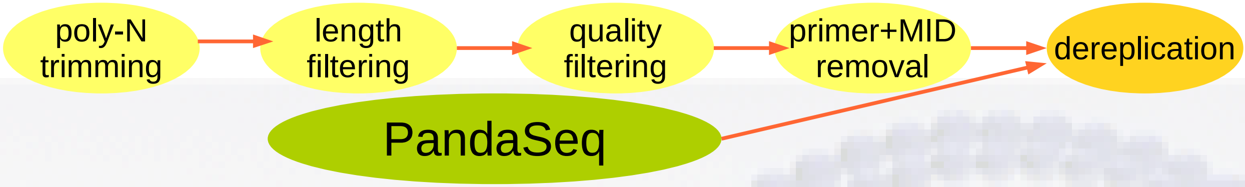


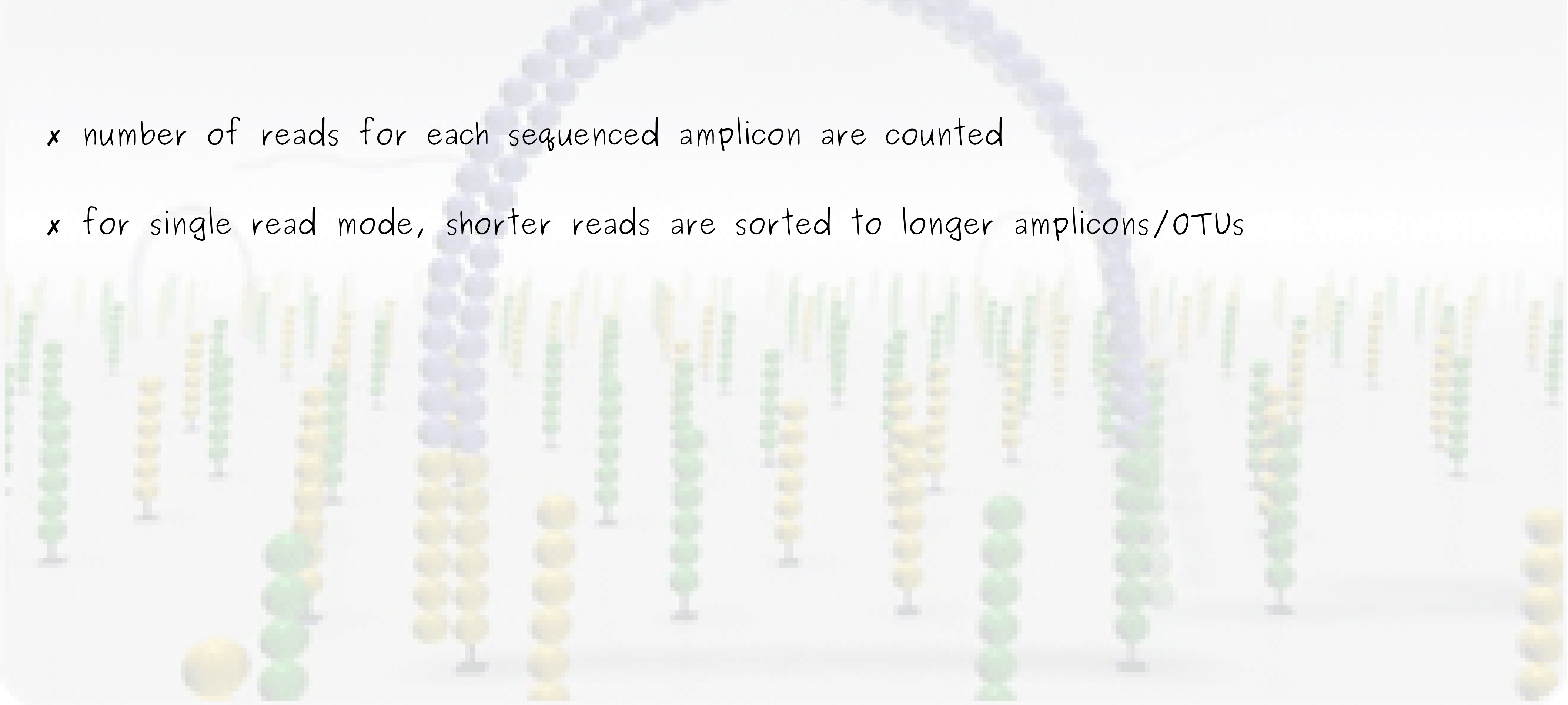
poly-N trimming → length filtering → quality filtering → primer+MID removal → dereplication

PandaSeq → dereplication

- ✗ number of reads for each sequenced amplicon are counted

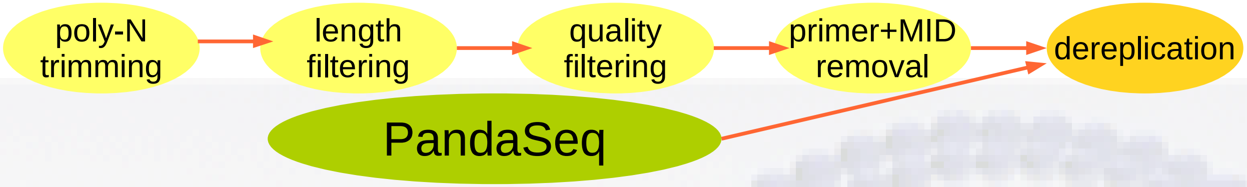
The pipeline





- ✗ number of reads for each sequenced amplicon are counted
- ✗ for single read mode, shorter reads are sorted to longer amplicons/OTUs

The pipeline

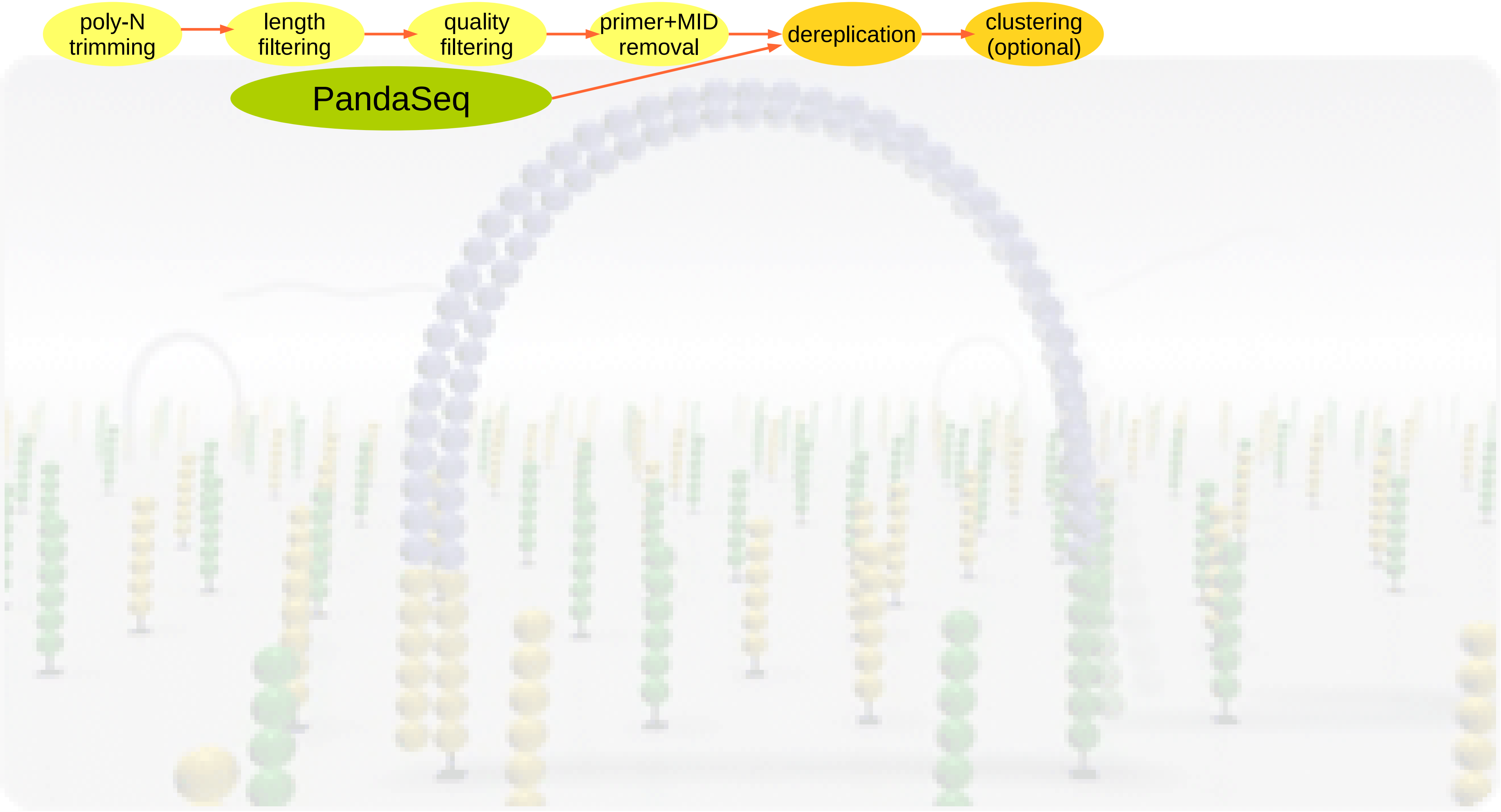


✗ number of reads for each sequenced amplicon are counted

✗ for single read mode, shorter reads are sorted to longer amplicons/OTUs

```
>Pro_Silber_Exp_Ko1_A_1;size=16066;
CGCGTAAGAACTTACCTTTTGGTGTGGGATAACAGCTGGAAACGGCTGCTAATACCGCATAGTGCTGAGAAGCTAAAAGTGA
AAACTGCCAAGAGAGAGGCTTGCGTCTGATTAGCTAGTTGGTGGAGGTAAAGGCTCCCCAAGGCGACGATCAGTAGCTGGT
CTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGAGGAATTTTCCACAAT
GGGCGAAAGCCTGATGGAGCAATACCGCGTGGAGGAAGACGGATCGTGGTCTGTAAACTCCTTTTCTTAGAGAAGACAACC
GACGGTATCTAAGGAATAAGCACCGGCTAACTCC
>Pro_Silber_Exp_Ko1_A_2;size=3883;
CGCGTAAGAATCTAACTTCAGGACGGGGACAACAGTTGGAAACGACTGCTAATACCCGATGTGCCGCAAGGTGAAACCTAAT
TGGCCTGGAGAAGAGCTTGCGTCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCAGTAGCTGGTCTGA
GAGGATGAGCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTCCGCAATGGG
CGAAAGCCTGACGGAGCAACGCCGCGTGAGGGAGGAAGGTCTTTGGATTGTAAACCTCTTTTCTCAAGGAAGAAGTTCTGA
CGGTACTTGAGGAATCAGCCTCGGCTAACTCC
>Pro_Silber_Exp_Ko1_A_3;size=3072;
CACGTATGCAACCTACCTTACATTGGGGGATAGCCTTTCGAAAGGGAGATTAATACCGCATAAGACAGTAGCTGGGCATCCAG
CAGCTGTTAAAGATTTATCGATGTAAGATGGGCATGCGTCCAATTAGTTAGTTGGCGAGGTAATGGCTCACCAAGACTTTGATT
GGTAGGGGAACTGAGAGGTCAATCCCCCACACTGGCACTGAGATACGGGCCAGACTCCTACGGGAGGCAGCAGTAGGGAA
TATTGGGCAATGGACGCAAGTCTGACCCAGCCATGCCGCGTGCAGGATGAAGGCGTTATGCGTTGTAAACTGCTTTTATACA
GGAAGAAACGACTCTTGCGAGAGGCATTGACGGTACTGTATGAATAAGCACCGGCTAACTCC
```

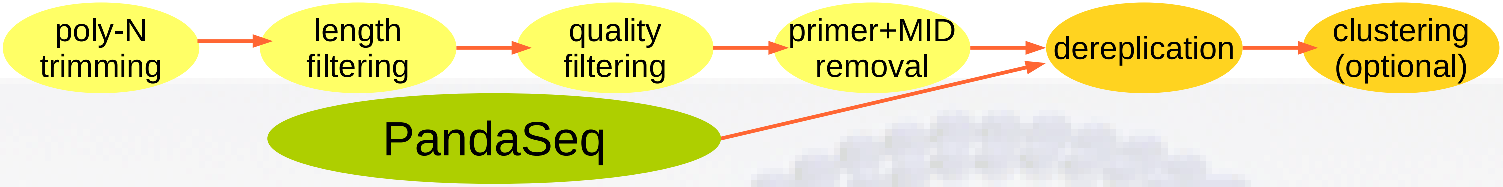
The pipeline

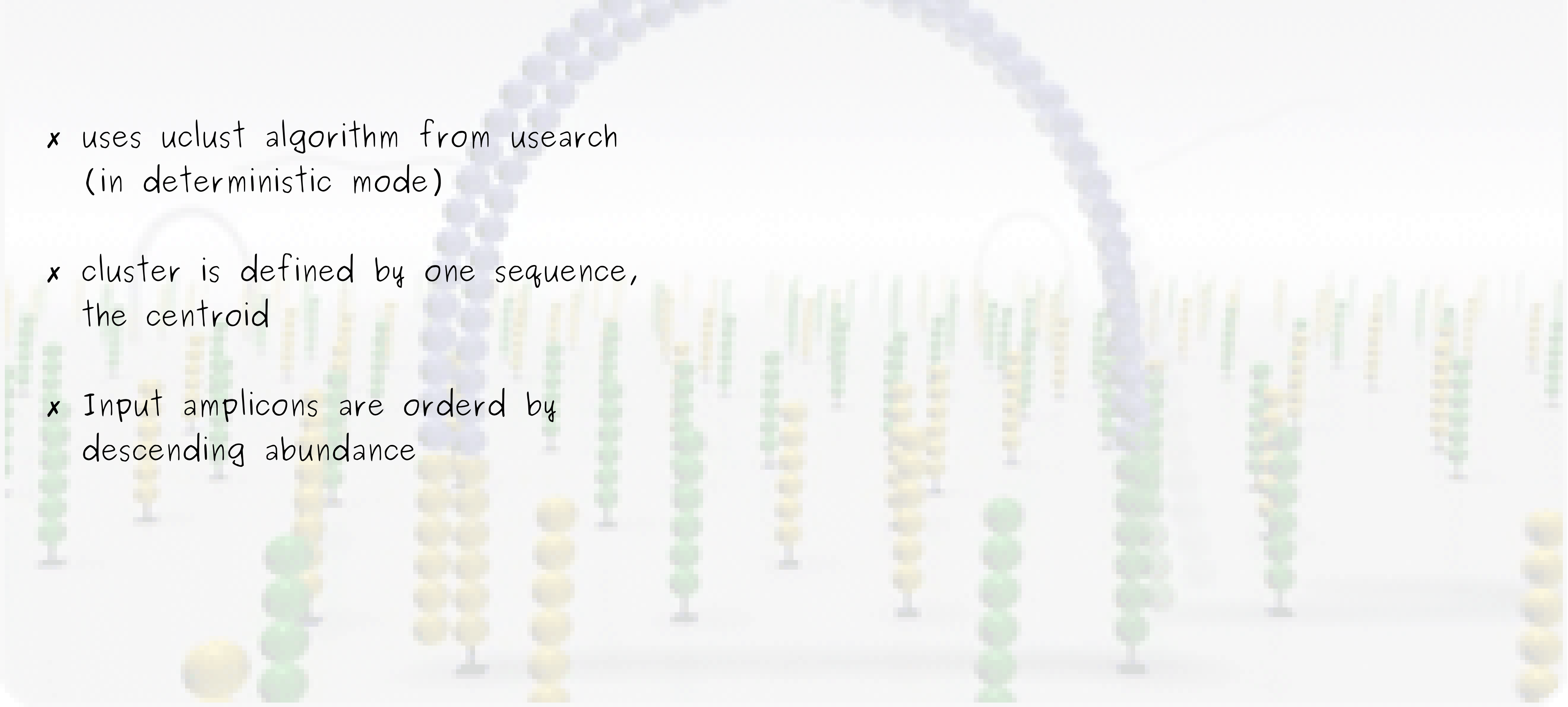


poly-N trimming → length filtering → quality filtering → primer+MID removal → dereplication → clustering (optional)

PandaSeq → dereplication

The pipeline





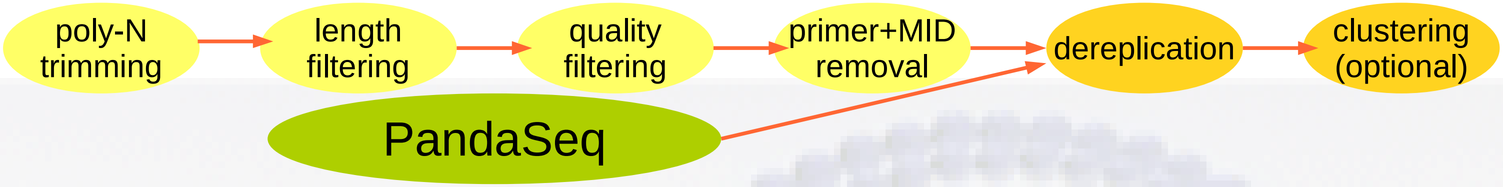
✗ uses uclust algorithm from usearch (in deterministic mode)

✗ cluster is defined by one sequence, the centroid

✗ Input amplicons are orderd by descending abundance

http://drive5.com/usearch/manual6/uclust_algo.html

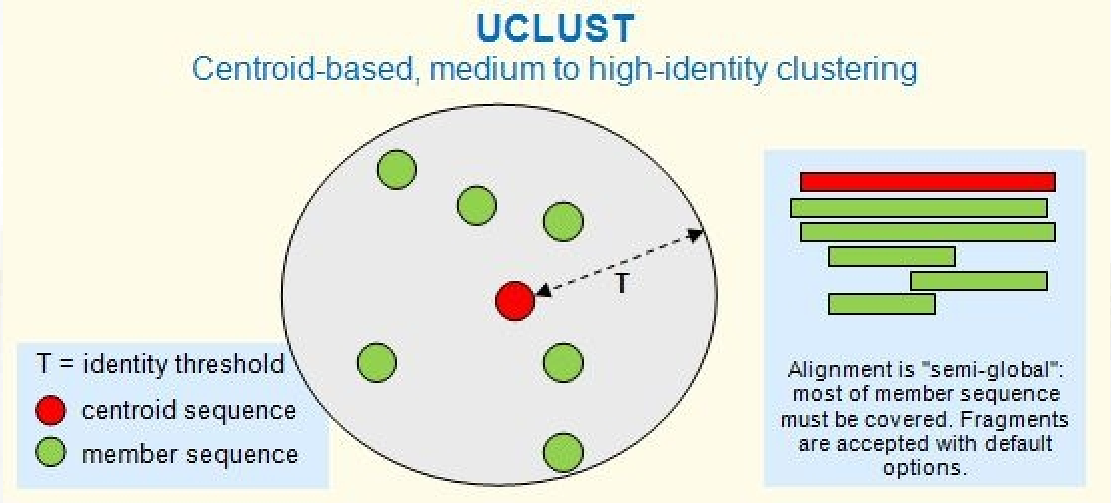
The pipeline



- uses uclust algorithm from usearch (in deterministic mode)

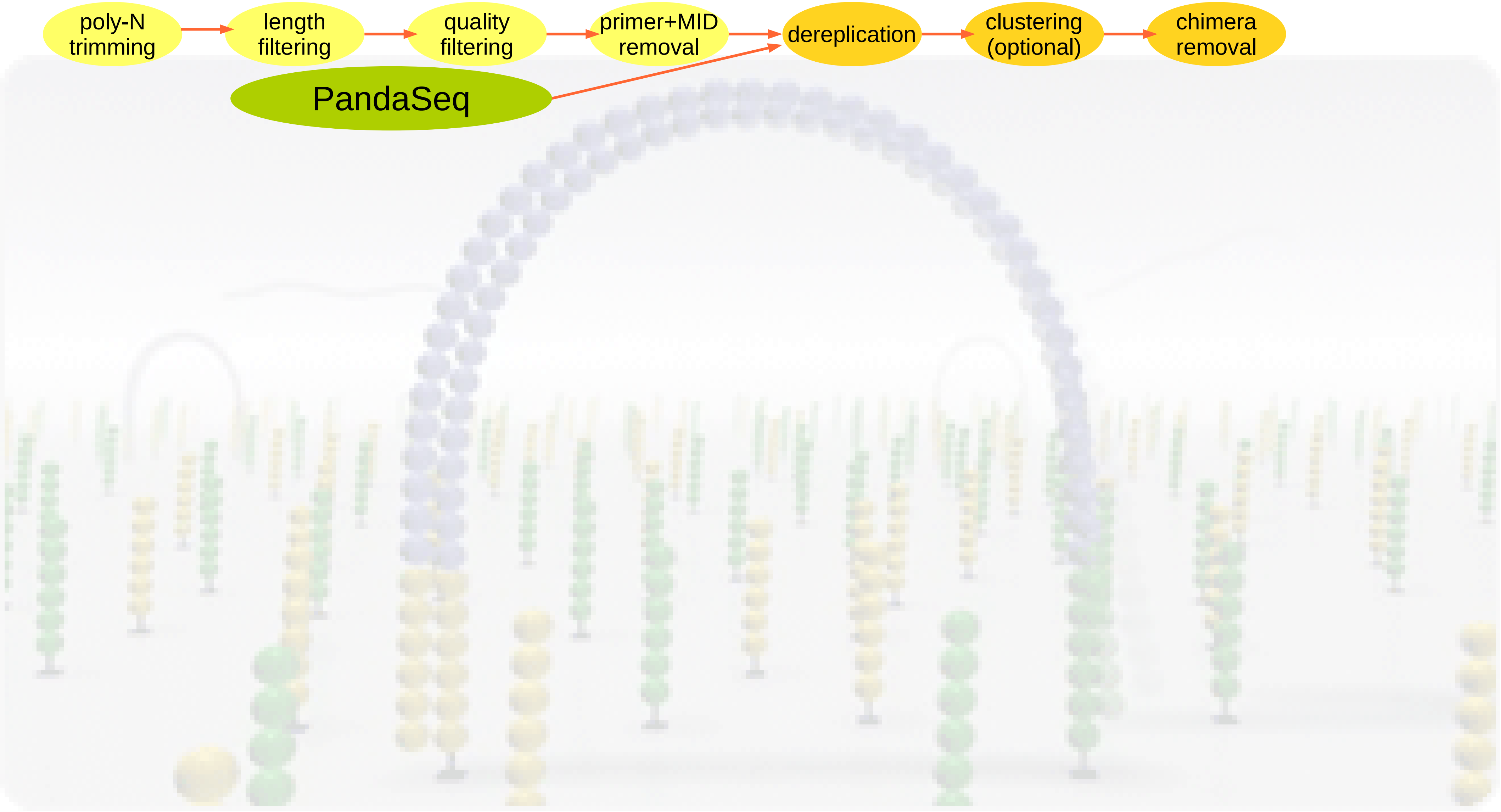
- cluster is defined by one sequence, the centroid

- Input amplicons are orderd by descending abundance


UCLUST
Centroid-based, medium to high-identity clustering
T
T = identity threshold
● centroid sequence
● member sequence
Alignment is "semi-global": most of member sequence must be covered. Fragments are accepted with default options.

http://drive5.com/usearch/manual6/uclust_algo.html

The pipeline



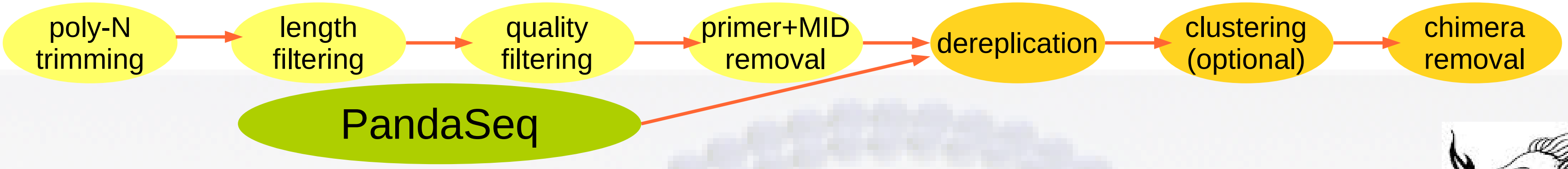
poly-N trimming → length filtering → quality filtering → primer+MID removal → dereplication → clustering (optional) → chimera removal

PandaSeq → dereplication

http://drive5.com/usearch/manual/uchime_algo.html

Bioinformatics. Aug 15, 2011; 27(16): 2194–2200.

The pipeline

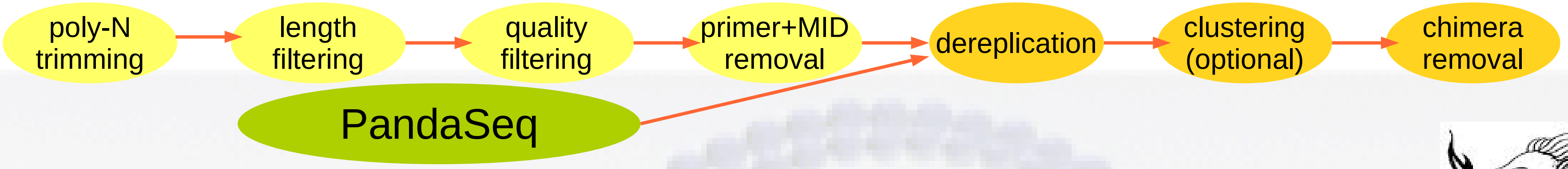


✗ chimera: sequences that stem from 2+ original sequences

http://drive5.com/usearch/manual/uchime_algo.html

Bioinformatics. Aug 15, 2011; 27(16): 2194–2200.

The pipeline



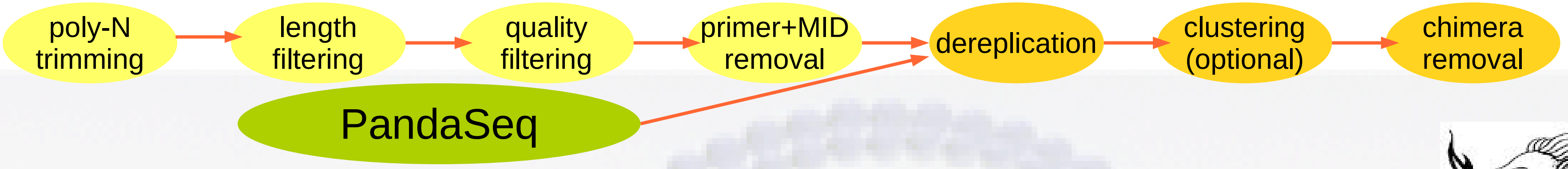
- chimera: sequences that stem from 2+ original sequences

- UCHIME – reference mode
  – denovo mode

http://drive5.com/usearch/manual/uchime_algo.html

Bioinformatics. Aug 15, 2011; 27(16): 2194–2200.

The pipeline

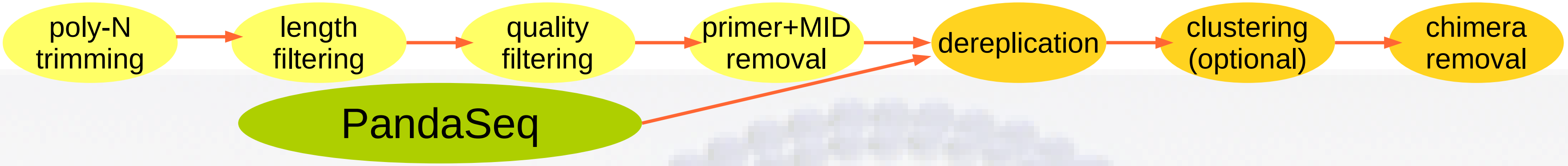


- chimera: sequences that stem from 2+ original sequences

- UCHIME
  - reference mode
  - denovo mode

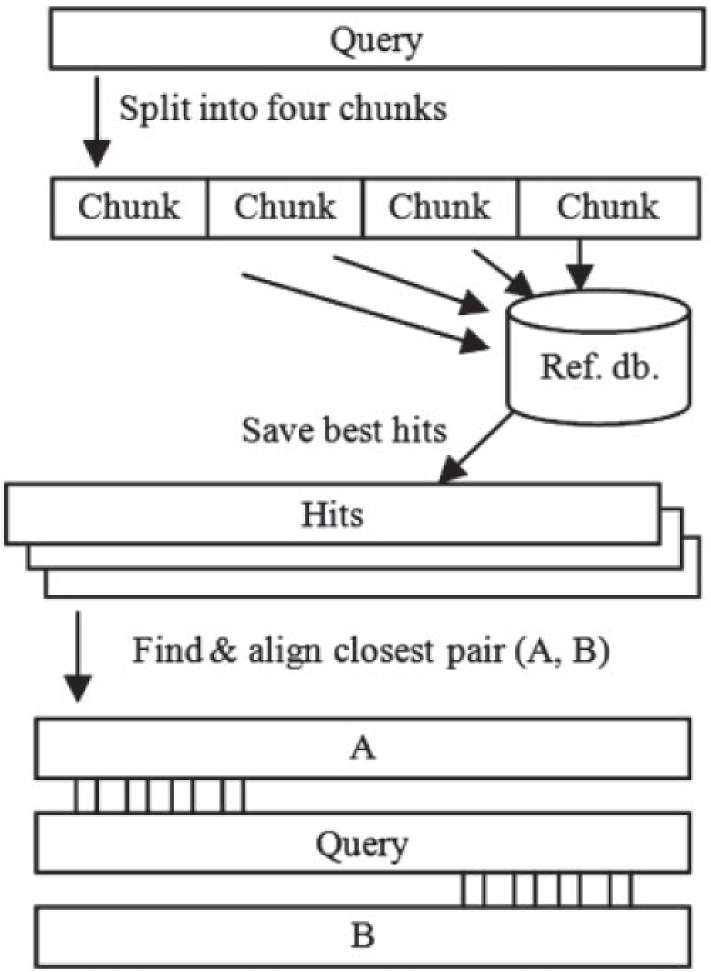
Bioinformatics. Aug 15, 2011; 27(16): 2194–2200.

The pipeline



✗ chimera: sequences that stem from 2+ original sequences

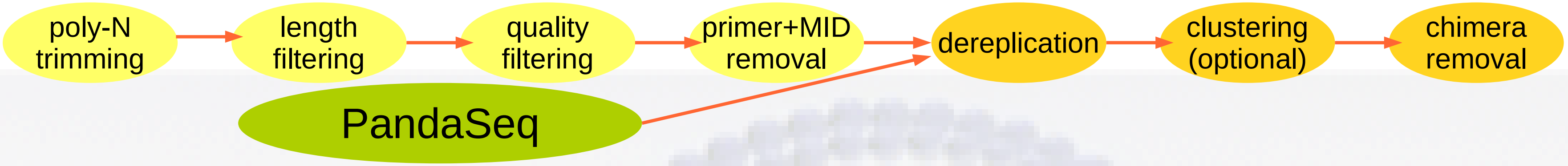
✗ UCHIME – reference mode
– denovo mode



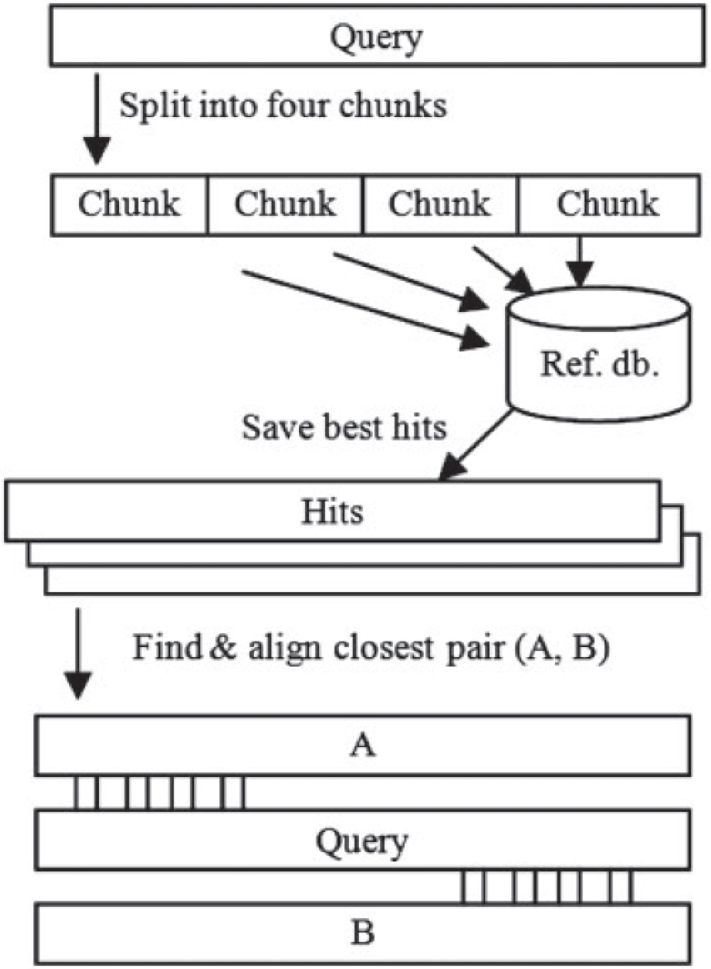
http://drive5.com/usearch/manual/uchime_algo.html

Bioinformatics. Aug 15, 2011; 27(16): 2194–2200.

The pipeline



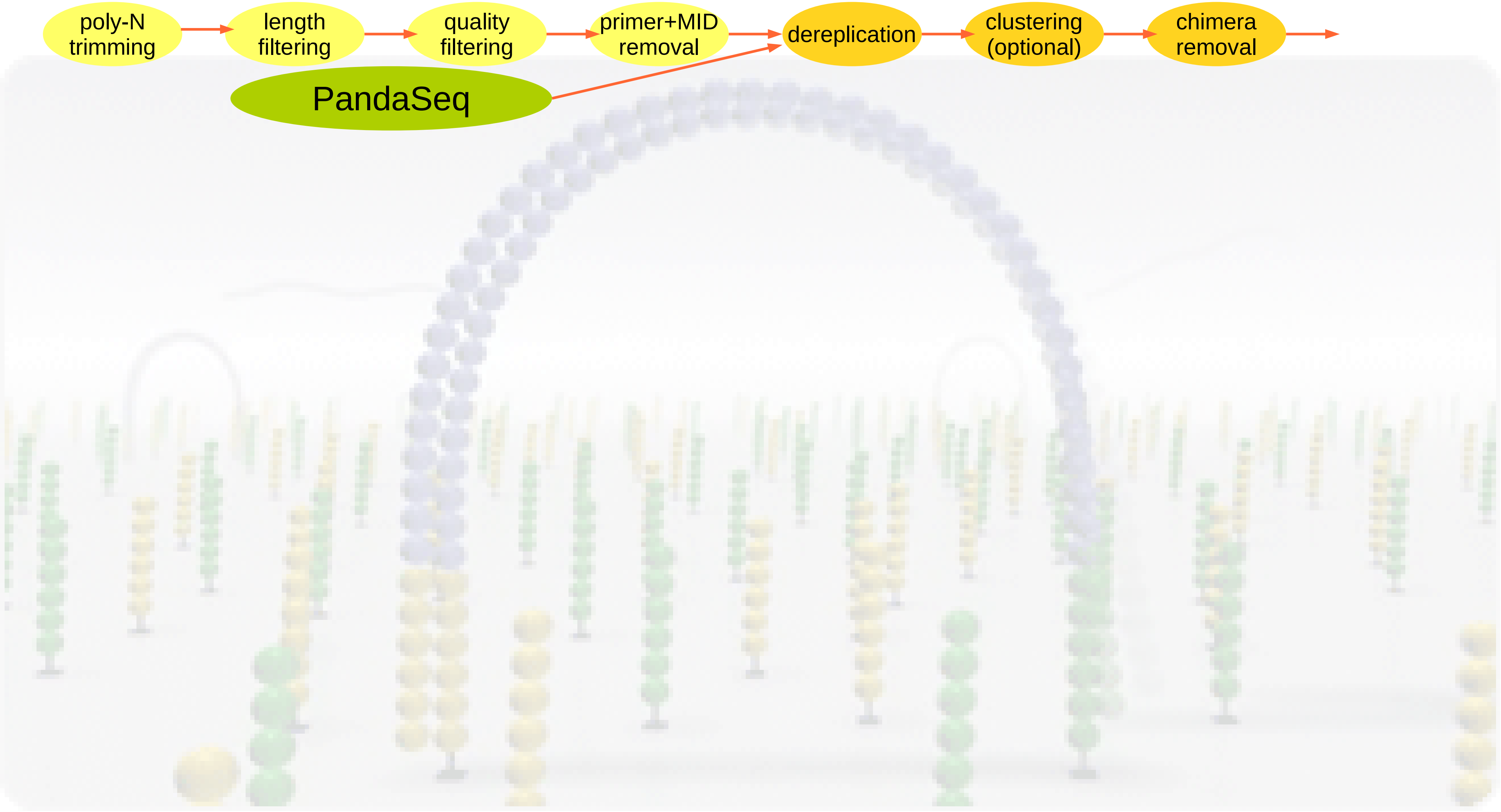
- ✗ chimera: sequences that stem from 2+ original sequences
- ✗ UCHIME – reference mode
  – denovo mode
- ✗ algorithm:
  - query is devided into 4 chunks
  - each chunk is used to search a refernce database
  - 2 best candidate parents are identified, at least n times more abundant then query
  - three–way multiple alignment is constructed
  - calculates a score h for the alignment
  - if h is above a user specified treshold → query is classified as a chimera
  - any sequence classified as non chimeric is added to the reference DB



http://drive5.com/usearch/manual/uchime_algo.html

Bioinformatics. Aug 15, 2011; 27(16): 2194–2200.

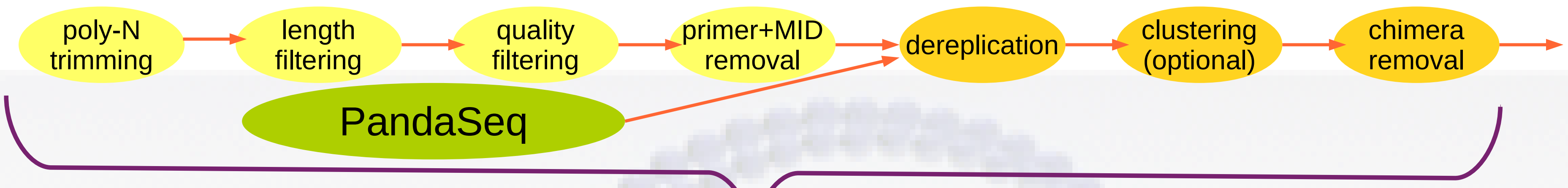
The pipeline



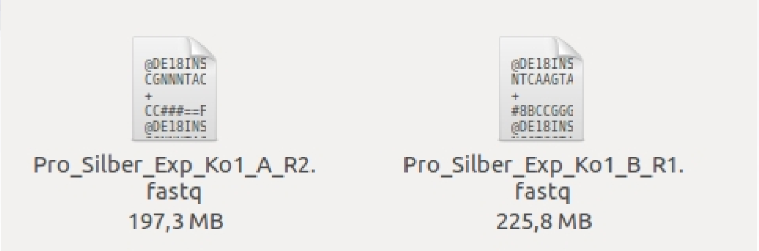
poly-N trimming → length filtering → quality filtering → primer+MID removal → dereplication → clustering (optional) → chimera removal →

PandaSeq → dereplication

The pipeline



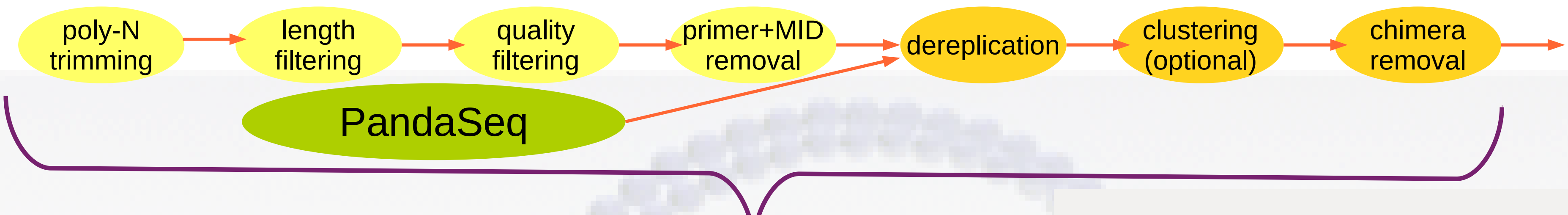
for each sample seperately



Pro_Silber_Exp_Ko1_A_R2.fastq
197,3 MB

Pro_Silber_Exp_Ko1_B_R1.fastq
225,8 MB

The pipeline



poly-N trimming → length filtering → quality filtering → primer+MID removal → dereplication → clustering (optional) → chimera removal →

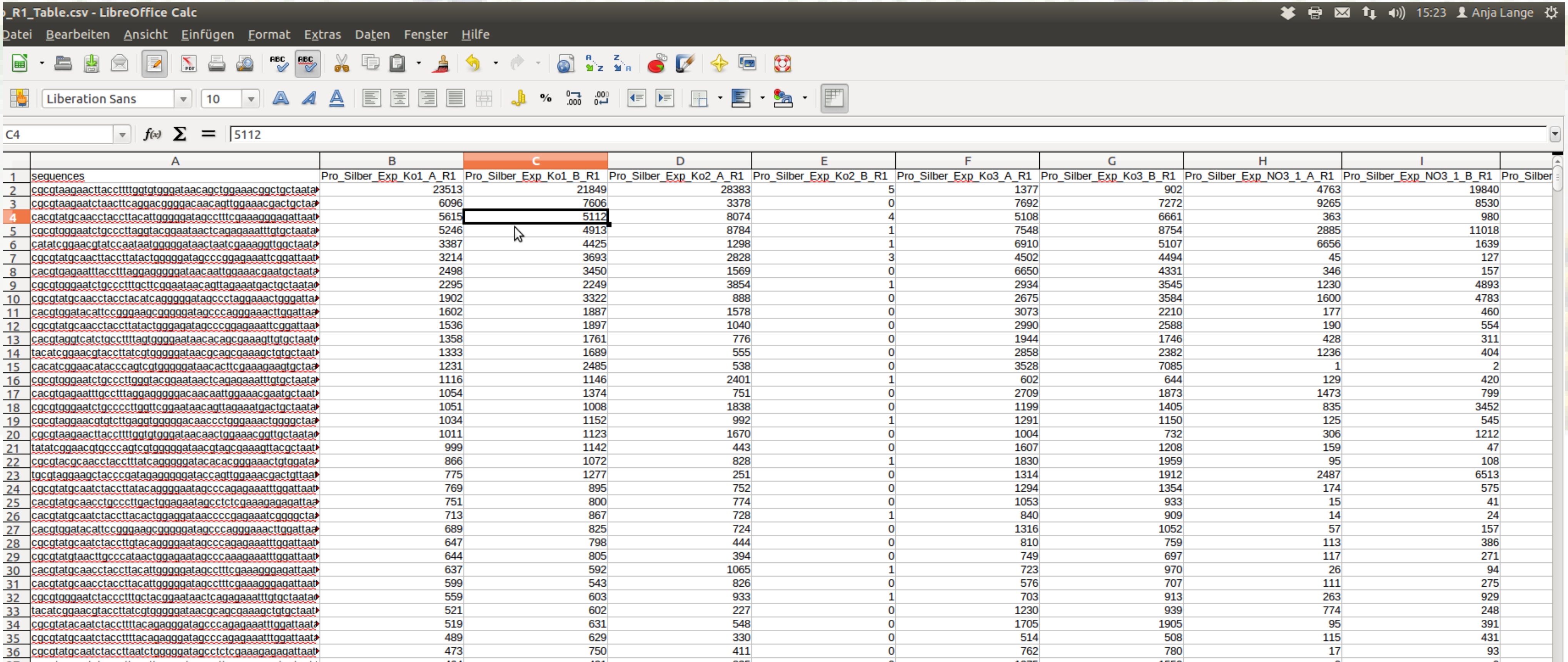
PandaSeq → dereplication

for each sample seperately

Pro_Silber_Exp_Ko1_A_R2.fastq
197,3 MB

Pro_Silber_Exp_Ko1_B_R1.fastq
225,8 MB

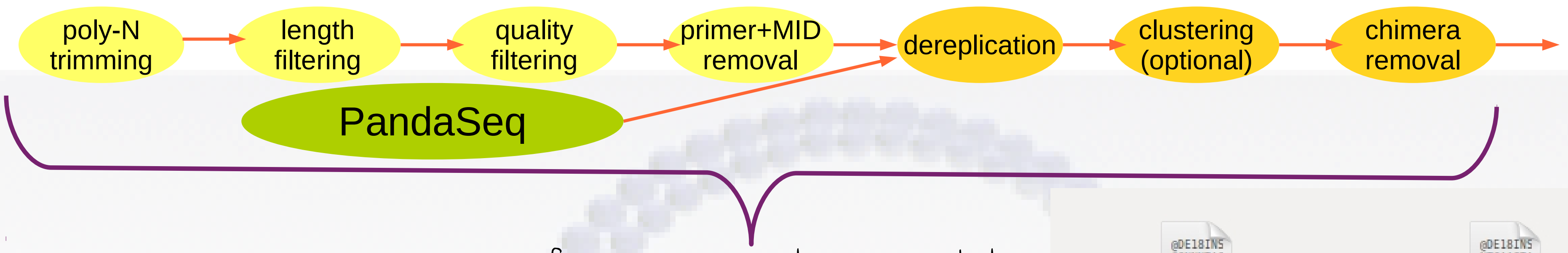
one table with all samples is generated, sequences used as key values



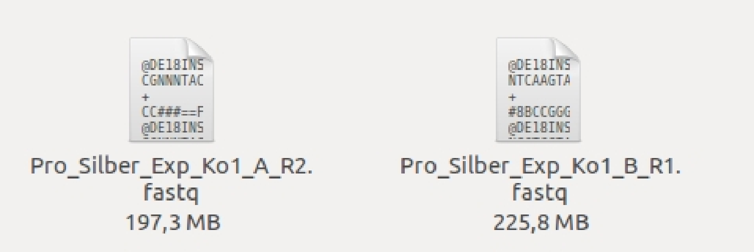
C4 = 5112

	A	B	C	D	E	F	G	H	I	
1	sequences	Pro_Silber_Exp_Ko1_A_R1	Pro_Silber_Exp_Ko1_B_R1	Pro_Silber_Exp_Ko2_A_R1	Pro_Silber_Exp_Ko2_B_R1	Pro_Silber_Exp_Ko3_A_R1	Pro_Silber_Exp_Ko3_B_R1	Pro_Silber_Exp_NO3_1_A_R1	Pro_Silber_Exp_NO3_1_B_R1	Pro_Silber
2	cgcgtaagaacttaccttttggtgtgggataacagctggaaacggctgctaata	23513	21849	28383	5	1377	902	4763	19840	
3	cgcgtaagaatctaacttcaggacggggacaacagttggaaacgactgctaa	6096	7606	3378	0	7692	7272	9265	8530	
4	cacgtatgcaacctaccttacattgggggatagcctttcgaaagggagattaat	5615	5112	8074	4	5108	6661	363	980	
5	cgcgtgggaatctgcccttaggtacggaataactcagagaaatttgtgctaata	5246	4913	8784	1	7548	8754	2885	11018	
6	catatcggaacgtatccaataatgggggataactaatcgaaaggttggctaata	3387	4425	1298	1	6910	5107	6656	1639	
7	cgcgtatgcaacttaccttatactgggggatagcccggagaaattcggattaat	3214	3693	2828	3	4502	4494	45	127	
8	cacgtgagaatttacctttaggagggggataacaattggaaacgaatgctaata	2498	3450	1569	0	6650	4331	346	157	
9	cgcgtgggaatctgccctttgcttcggaataacagttagaaatgactgctaata	2295	2249	3854	1	2934	3545	1230	4893	
10	cgcgtatgcaacctacctacatcaggggatagccctaggaaactgggatta	1902	3322	888	0	2675	3584	1600	4783	
11	cacgtggatacattccgggaagcgggggatagcccagggaaacttggattaa	1602	1887	1578	0	3073	2210	177	460	
12	cgcgtatgcaacctaccttatactgggagatagcccggagaaattcggattaa	1536	1897	1040	0	2990	2588	190	554	
13	cacgtaggtcatctgccttttagtggggaataacacagcgaaagttgtgctaat	1358	1761	776	0	1944	1746	428	311	
14	tacatcggaacgtaccttatcgtgggggataacgcagcgaaagctgtgctaat	1333	1689	555	0	2858	2382	1236	404	
15	cacatcggaacatacccagtcgtgggggataacacttcgaaagaagtgctaa	1231	2485	538	0	3528	7085	1	2	
16	cgcgtgggaatctgcccttgggtacggaataactcagagaaatttgtgctaata	1116	1146	2401	1	602	644	129	420	
17	cacgtgagaatttgcctttaggagggggacaacaattggaaacgaatgctaat	1054	1374	751	0	2709	1873	1473	799	
18	cgcgtgggaatctgccccttggttcggaataacagttagaaatgactgctaata	1051	1008	1838	0	1199	1405	835	3452	
19	cgcgtaggaacgtgtcttgaggtggggacaaccctgggaaactggggctaa	1034	1152	992	1	1291	1150	125	545	
20	cgcgtaagaacttaccttttggtgtgggataacaactggaaacggttgctaata	1011	1123	1670	0	1004	732	306	1212	
21	tatatcggaacgtgcccagtcgtgggggataacgtagcgaaagttacgctaat	999	1142	443	0	1607	1208	159	47	
22	cgcgtacgcaacctacctttatcaggggatacacacgggaaactgtggata	866	1072	828	1	1830	1959	95	108	
23	tgcgtaggaagctacccgatagagggggataccagttggaaacgactgttaa	775	1277	251	0	1314	1912	2487	6513	
24	cgcgtatgcaatctaccttatacaggggaatagcccagagaaatttggattaat	769	895	752	0	1294	1354	174	575	
25	cacgtatgcaacctgcccttgactggagaatagcctctcgaaagagagattaa	751	800	774	0	1053	933	15	41	
26	cacgtatgcaatctaccttacactggaggataaccccgagaaatcggggcta	713	867	728	1	840	909	14	24	
27	cacgtggatacattccgggaagcgggggatagcccagggaaacttggattaa	689	825	724	0	1316	1052	57	157	
28	cgcgtatgcaatctaccttgtacaggggaatagcccagagaaatttggattaat	647	798	444	0	810	759	113	386	
29	cgcgtatgtaacttgcccataactggagaatagcccaaagaaatttggattaat	644	805	394	0	749	697	117	271	
30	cacgtatgcaacctaccttacattgggggatagcctttcgaaagggagattaat	637	592	1065	1	723	970	26	94	
31	cacgtatgcaacctaccttacattgggggatagcctttcgaaagggagattaat	599	543	826	0	576	707	111	275	
32	cgcgtgggaatctaccctttgctacggaataactcagagaaatttgtgctaata	559	603	933	1	703	913	263	929	
33	tacatcggaacgtaccttatcgtgggggataacgcagcgaaagctgtgctaat	521	602	227	0	1230	939	774	248	
34	cgcgtatacaatctaccttttacagagggatagcccagagaaatttggattaata	519	631	548	0	1705	1905	95	391	
35	cgcgtatgcaatctaccttttacagagggatagcccagagaaatttggattaata	489	629	330	0	514	508	115	431	
36	cgcgtatgcaatctaccttaatctgggggatagcctctcgaaagagagattaat	473	750	411	0	762	780	17	93	

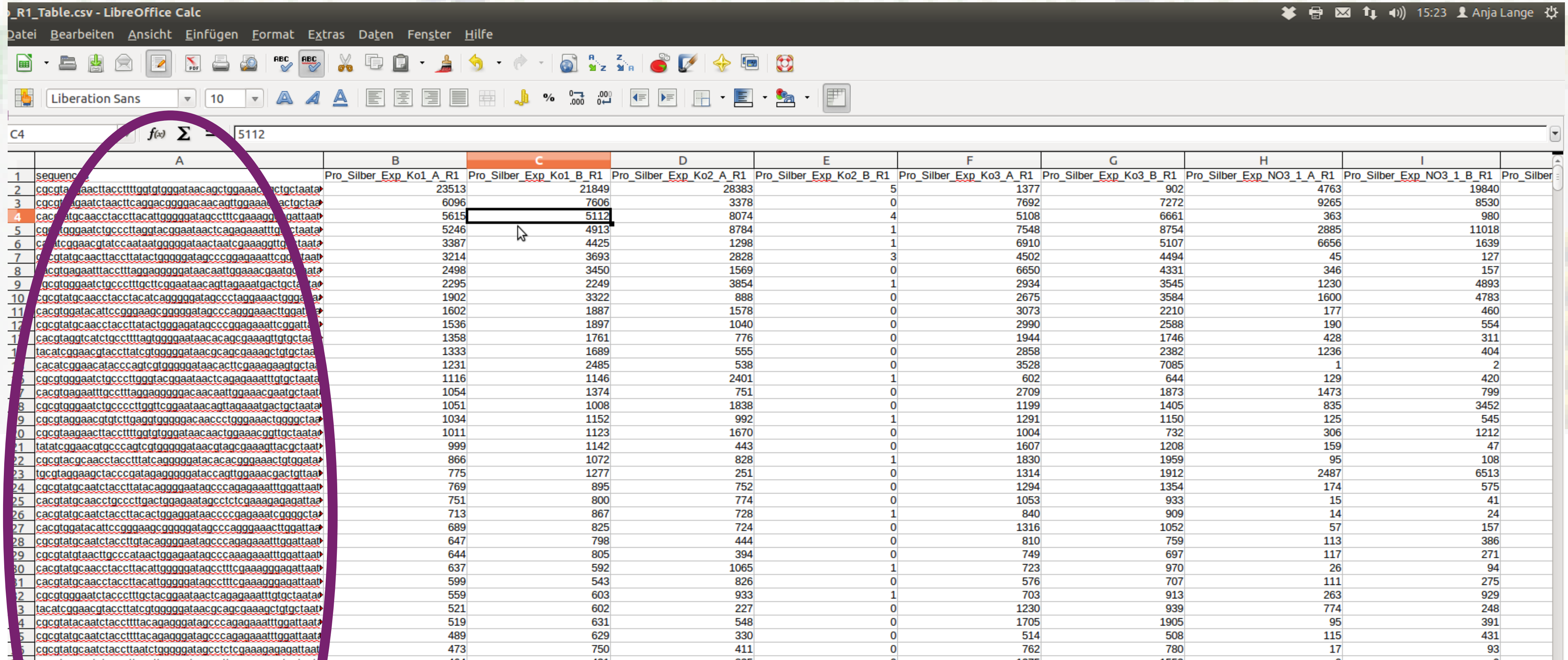
The pipeline



for each sample seperately

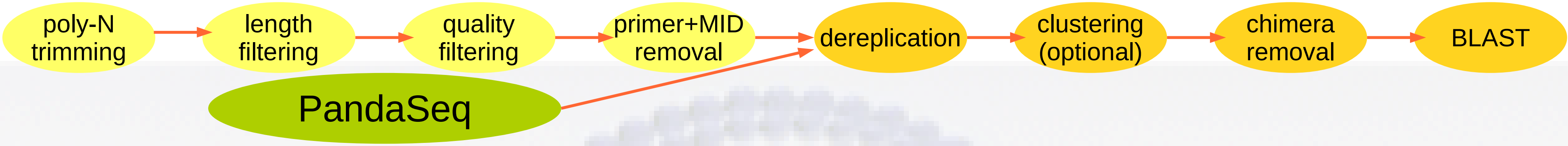


one table with all samples is generated, sequences used as key values



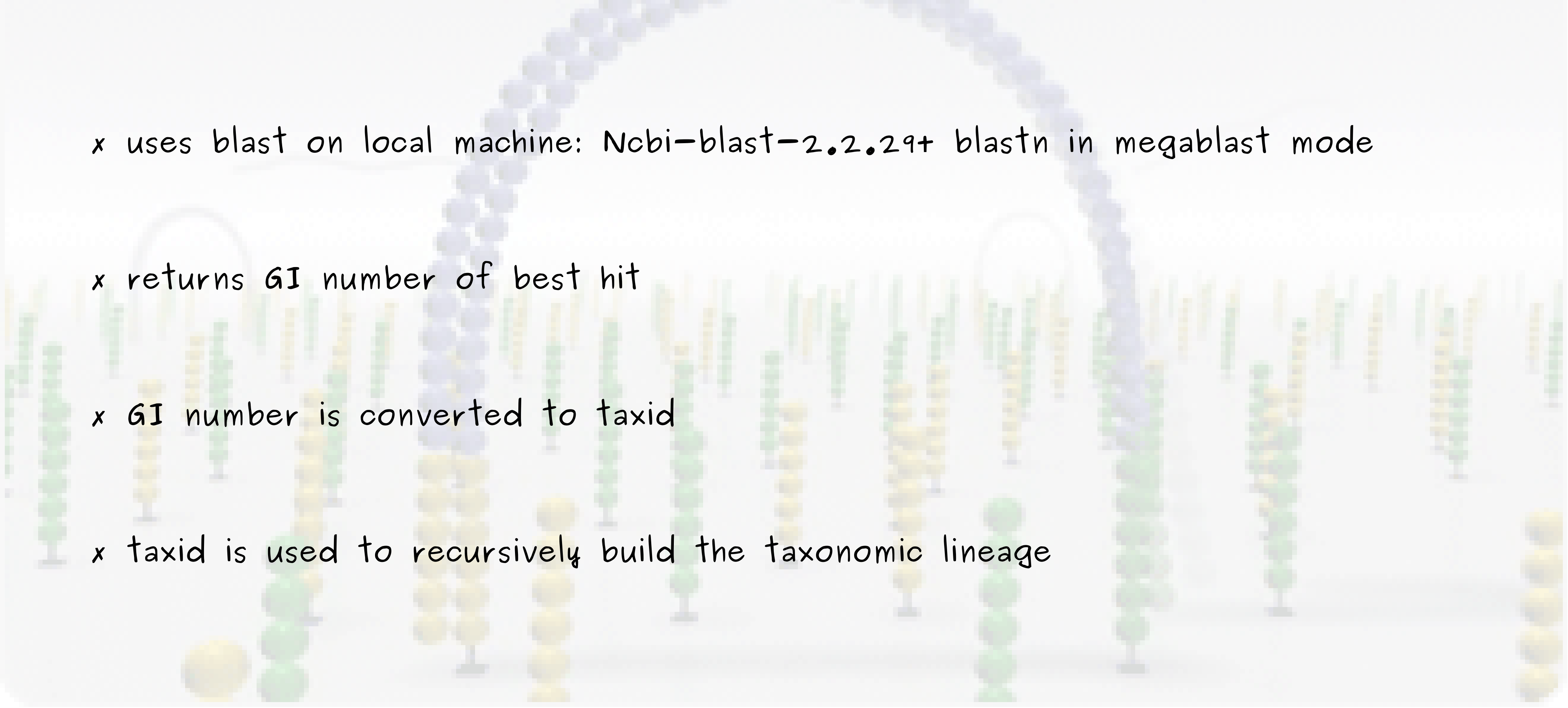
	A	B	C	D	E	F	G	H	I	
1	sequenc[illegible]	Pro_Silber_Exp_Ko1_A_R1	Pro_Silber_Exp_Ko1_B_R1	Pro_Silber_Exp_Ko2_A_R1	Pro_Silber_Exp_Ko2_B_R1	Pro_Silber_Exp_Ko3_A_R1	Pro_Silber_Exp_Ko3_B_R1	Pro_Silber_Exp_NO3_1_A_R1	Pro_Silber_Exp_NO3_1_B_R1	Pro_Silber
2	cgcgta[illegible]gaacttaccttttggtgtgggataacagctggaaac[illegible]gctgctaata	23513	21849	28383	5	1377	902	4763	19840	
3	cgcgt[illegible]agaatctaacttcaggacggggacaacagttggaaa[illegible]actgctaa	6096	7606	3378	0	7692	7272	9265	8530	
4	cac[illegible]atgcaacctaccttacattgggggatagcctttcgaaagg[illegible]gattaat	5615	5112	8074	4	5108	6661	363	980	
5	cg[illegible]tgggaatctgcccttaggtacggaataactcagagaaatttg[illegible]ctaata	5246	4913	8784	1	7548	8754	2885	11018	
6	ca[illegible]atcggaacgtatccaataatgggggataactaatcgaaaggttg[illegible]taata	3387	4425	1298	1	6910	5107	6656	1639	
7	c[illegible]cgtatgcaacttaccttatactgggggatagcccggagaaattcgg[illegible]taat	3214	3693	2828	3	4502	4494	45	127	
8	[illegible]acgtgagaatttacctttaggaggggataacaattggaaacgaatgc[illegible]ata	2498	3450	1569	0	6650	4331	346	157	
9	[illegible]gcgtgggaatctgccctttgcttcggaataacagttagaaatgactgcta[illegible]ta	2295	2249	3854	1	2934	3545	1230	4893	
10	cgcgtatgcaacctacctacatcaggggggatagccctaggaaactggga[illegible]a	1902	3322	888	0	2675	3584	1600	4783	
11	cacgtggatacattccgggaagcgggggatagcccagggaaacttggat[illegible]	1602	1887	1578	0	3073	2210	177	460	
12	cgcgtatgcaacctaccttatactgggagatagcccggagaaattcggatta	1536	1897	1040	0	2990	2588	190	554	
13	cacgtaggtcatctgccttttagtggggaataacacagcgaaagttgtgctaa	1358	1761	776	0	1944	1746	428	311	
14	tacatcggaacgtaccttatcgtgggggataacgcagcgaaagctgtgctaa	1333	1689	555	0	2858	2382	1236	404	
15	cacatcggaacatacccagtcgtgggggataacacttcgaaagaagtgctaa	1231	2485	538	0	3528	7085	1	2	
16	cgcgtgggaatctgcccttgggtacggaataactcagagaaatttgtgctaata	1116	1146	2401	1	602	644	129	420	
17	cacgtgagaatttgcctttaggaggggacaacaattggaaacgaatgctaat	1054	1374	751	0	2709	1873	1473	799	
18	cgcgtgggaatctgccccttggttcggaataacagttagaaatgactgctaata	1051	1008	1838	0	1199	1405	835	3452	
19	cgcgtaggaacgtgtcttgaggtgggggacaaccctgggaaactggggctaa	1034	1152	992	1	1291	1150	125	545	
20	cgcgtaagaacttaccttttggtgtgggataacaactggaaacggttgctaata	1011	1123	1670	0	1004	732	306	1212	
21	tatatcggaacgtgcccagtcgtgggggataacgtagcgaaagttacgctaat	999	1142	443	0	1607	1208	159	47	
22	cgcgtacgcaacctacctttatcaggggatacacacgggaaactgtggata	866	1072	828	1	1830	1959	95	108	
23	tgcgtaggaagctacccgatagagggggataccagttggaaacgactgttaat	775	1277	251	0	1314	1912	2487	6513	
24	cgcgtatgcaatctaccttatacaggggaatagcccagagaaatttggattaat	769	895	752	0	1294	1354	174	575	
25	cacgtatgcaacctgcccttgactggagaatagcctctcgaaagagagattaa	751	800	774	0	1053	933	15	41	
26	cacgtatgcaatctaccttacactggaggataaccccgagaaatcggggcta	713	867	728	1	840	909	14	24	
27	cacgtggatacattccgggaagcgggggatagcccagggaaacttggattaa	689	825	724	0	1316	1052	57	157	
28	cgcgtatgcaatctaccttgtacaggggaatagcccagagaaatttggattaat	647	798	444	0	810	759	113	386	
29	cgcgtatgtaacttgcccataactggagaatagcccaaagaaatttggattaat	644	805	394	0	749	697	117	271	
30	cacgtatgcaacctaccttacattgggggatagcctttcgaaagggagattaat	637	592	1065	1	723	970	26	94	
31	cacgtatgcaacctaccttacattgggggatagcctttcgaaagggagattaat	599	543	826	0	576	707	111	275	
32	cgcgtgggaatctaccctttgctacggaataactcagagaaatttgtgctaata	559	603	933	1	703	913	263	929	
33	tacatcggaacgtaccttatcgtgggggataacgcagcgaaagctgtgctaat	521	602	227	0	1230	939	774	248	
34	cgcgtatacaatctaccttttacagagggatagcccagagaaatttggattaata	519	631	548	0	1705	1905	95	391	
35	cgcgtatgcaatctaccttttacagagggatagcccagagaaatttggattaata	489	629	330	0	514	508	115	431	
36	cgcgtatgcaatctaccttaatctgggggatagcctctcgaaagagagattaat	473	750	411	0	762	780	17	93	

The pipeline



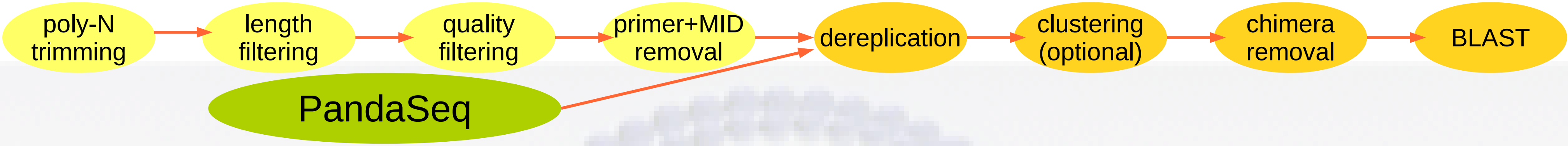
poly-N trimming → length filtering → quality filtering → primer+MID removal → dereplication → clustering (optional) → chimera removal → BLAST

PandaSeq → dereplication



- uses blast on local machine: Ncbi-blast-2.2.29+ blastn in megablast mode
- returns GI number of best hit
- GI number is converted to taxid
- taxid is used to recursively build the taxonomic lineage

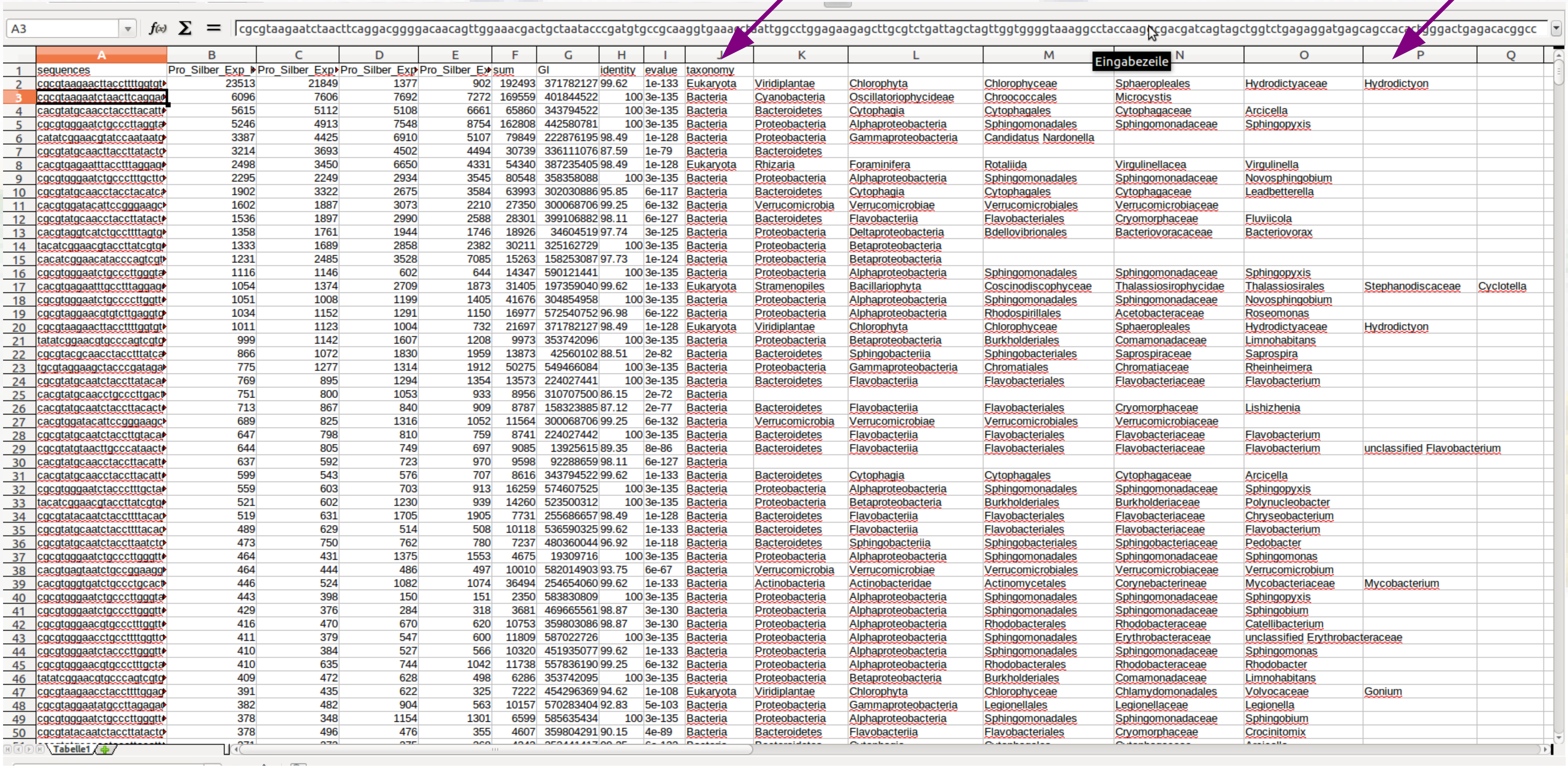
The pipeline



the final output

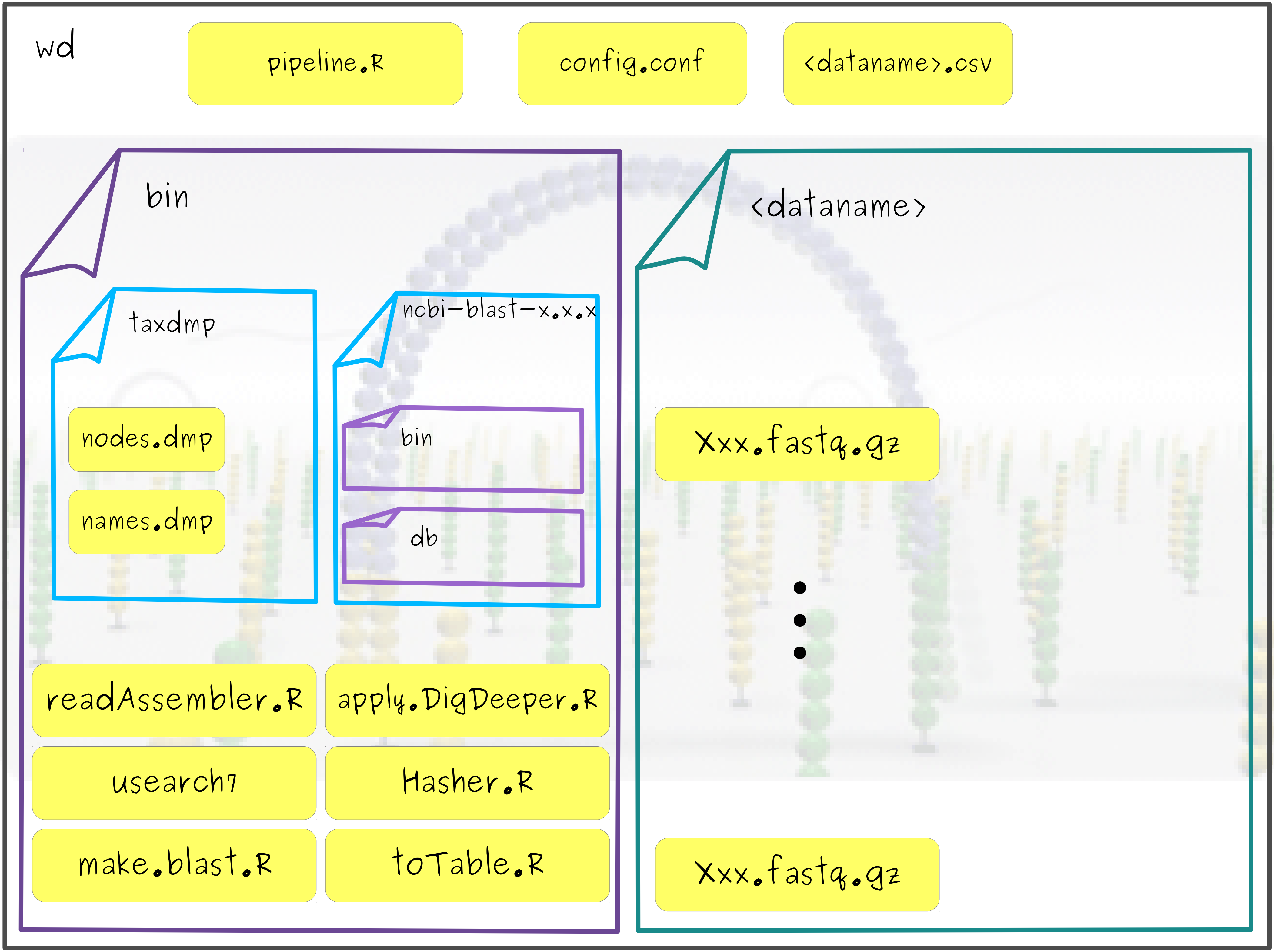
superkingdom

superclass



cgcgtaagaatctaacttcaggacggggacaacagttggaaacgactgctaatacccgatgtgccgcaaggtgaaactaattggcctggagaagagcttgcgtctgattagctagttggtggggtaaaggcctaccaaggcgacgatcagtagctggtctgagaggatgagcagccacactgggactgagacacggcc

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	sequences	Pro_Silber_Exp_	Pro_Silber_Exp	Pro_Silber_Exp	Pro_Silber_Ex	sum	GI	identity	evalue	taxonomy							
2	cgcgtaagaacttaccttttggtgt	23513	21849	1377	902	192493	371782127	99.62	1e-133	Eukaryota	Viridiplantae	Chlorophyta	Chlorophyceae	Sphaeropleales	Hydrodictyaceae	Hydrodictyon	
3	cgcgtaagaatctaacttcaggac	6096	7606	7692	7272	169559	401844522	100	3e-135	Bacteria	Cyanobacteria	Oscillatoriophycideae	Chroococcales	Microcystis			
4	cacgtatgcaacctaccttacatt	5615	5112	5108	6661	65860	343794522	100	3e-135	Bacteria	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	Arcicella		
5	cgcgtgggaatctgcccttaggta	5246	4913	7548	8754	162808	442580781	100	3e-135	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingopyxis		
6	catatcggaacgtatccaataatg	3387	4425	6910	5107	79849	222876195	98.49	1e-128	Bacteria	Proteobacteria	Gammaproteobacteria	Candidatus Nardonella				
7	cgcgtatgcaacttaccttatactc	3214	3693	4502	4494	30739	336111076	87.59	1e-79	Bacteria	Bacteroidetes						
8	cacgtgagaatttacctttaggag	2498	3450	6650	4331	54340	387235405	98.49	1e-128	Eukaryota	Rhizaria	Foraminifera	Rotaliida	Virgulinellacea	Virgulinella		
9	cgcgtgggaatctgccctttgcttc	2295	2249	2934	3545	80548	358358088	100	3e-135	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Novosphingobium		
10	cgcgtatgcaacctacctacatca	1902	3322	2675	3584	63993	302030886	95.85	6e-117	Bacteria	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	Leadbetterella		
11	cacgtggatacattccgggaagc	1602	1887	3073	2210	27350	300068706	99.25	6e-132	Bacteria	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae			
12	cgcgtatgcaacctaccttatact	1536	1897	2990	2588	28301	399106882	98.11	6e-127	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	Fluviicola		
13	cacgtaggtcatctgccttttagtg	1358	1761	1944	1746	18926	34604519	97.74	3e-125	Bacteria	Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae	Bacteriovorax		
14	tacatcggaacgtaccttatcgtg	1333	1689	2858	2382	30211	325162729	100	3e-135	Bacteria	Proteobacteria	Betaproteobacteria					
15	cacatcggaacatacccagtcgt	1231	2485	3528	7085	15263	158253087	97.73	1e-124	Bacteria	Proteobacteria	Betaproteobacteria					
16	cgcgtgggaatctgcccttgggta	1116	1146	602	644	14347	590121441	100	3e-135	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingopyxis		
17	cacgtgagaatttgcctttaggag	1054	1374	2709	1873	31405	197359040	99.62	1e-133	Eukaryota	Stramenopiles	Bacillariophyta	Coscinodiscophyceae	Thalassiosirophycidae	Thalassiosirales	Stephanodiscaceae	Cyclotella
18	cgcgtgggaatctgccccttggtt	1051	1008	1199	1405	41676	304854958	100	3e-135	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Novosphingobium		
19	cgcgtaggaacgtgtcttgaggtg	1034	1152	1291	1150	16977	572540752	96.98	6e-122	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Roseomonas		
20	cgcgtaagaacttaccttttggtgt	1011	1123	1004	732	21697	371782127	98.49	1e-128	Eukaryota	Viridiplantae	Chlorophyta	Chlorophyceae	Sphaeropleales	Hydrodictyaceae	Hydrodictyon	
21	tatatcggaacgtgcccagtcgtg	999	1142	1607	1208	9973	353742096	100	3e-135	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Limnohabitans		
22	cgcgtacgcaacctacctttatca	866	1072	1830	1959	13873	42560102	88.51	2e-82	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Saprospiraceae	Saprospira		
23	tgcgtaggaagctacccgataga	775	1277	1314	1912	50275	549466084	100	3e-135	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Chromatiaceae	Rheinheimera		
24	cgcgtatgcaatctaccttataca	769	895	1294	1354	13573	224027441	100	3e-135	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium		
25	cacgtatgcaacctgcccttgact	751	800	1053	933	8956	310707500	86.15	2e-72	Bacteria							
26	cacgtatgcaatctaccttacact	713	867	840	909	8787	158323885	87.12	2e-77	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	Lishizhenia		
27	cacgtggatacattccgggaagc	689	825	1316	1052	11564	300068706	99.25	6e-132	Bacteria	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae			
28	cgcgtatgcaatctaccttgtaca	647	798	810	759	8741	224027442	100	3e-135	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium		
29	cgcgtatgtaacttgcccataact	644	805	749	697	9085	13925615	89.35	8e-86	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	unclassified Flavobacterium	
30	cacgtatgcaacctaccttacatt	637	592	723	970	9598	92288659	98.11	6e-127	Bacteria							
31	cacgtatgcaacctaccttacatt	599	543	576	707	8616	343794522	99.62	1e-133	Bacteria	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	Arcicella		
32	cgcgtgggaatctaccctttgcta	559	603	703	913	16259	574607525	100	3e-135	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingopyxis		
33	tacatcggaacgtaccttatcgtg	521	602	1230	939	14260	523500312	100	3e-135	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Polynucleobacter		
34	cgcgtatacaatctaccttttacac	519	631	1705	1905	7731	255686657	98.49	1e-128	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Chryseobacterium		
35	cgcgtatgcaatctaccttttacac	489	629	514	508	10118	536590325	99.62	1e-133	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium		
36	cgcgtatgcaatctaccttaatctc	473	750	762	780	7237	480360044	96.92	1e-118	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter		
37	cgcgtgggaatctgcccttgggtt	464	431	1375	1553	4675	19309716	100	3e-135	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas		
38	cacgtgagtaatctgccggaagg	464	444	486	497	10010	582014903	93.75	6e-67	Bacteria	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	Verrucomicrobium		
39	cacgtgggtgatctgccctgcact	446	524	1082	1074	36494	254654060	99.62	1e-133	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae	Mycobacteriaceae	Mycobacterium	
40	cgcgtgggaatctgcccttgggta	443	398	150	151	2350	583830809	100	3e-135	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingopyxis		
41	cgcgtgggaatctgcccttgggtt	429	376	284	318	3681	469665561	98.87	3e-130	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium		
42	cgcgtgggaacgtgccctttggtt	416	470	670	620	10753	359803086	98.87	3e-130	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Catellibacterium		
43	cgcgtgggaacctgccttttggttc	411	379	547	600	11809	587022726	100	3e-135	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Erythrobacteraceae	unclassified Erythrobacteraceae		
44	cgcgtgggaatctacccttgggtt	410	384	527	566	10320	451935077	99.62	1e-133	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas		
45	cgcgtgggaacgtgccctttgcta	410	635	744	1042	11738	557836190	99.25	6e-132	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacter		
46	tatatcggaacgtgcccagtcgtg	409	472	628	498	6286	353742095	100	3e-135	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Limnohabitans		
47	cgcgtaagaacctaccttttggag	391	435	622	325	7222	454296369	94.62	1e-108	Eukaryota	Viridiplantae	Chlorophyta	Chlorophyceae	Chlamydomonadales	Volvocaceae	Gonium	
48	cgcgtaggaatatgccttagagac	382	482	904	563	10157	570283404	92.83	5e-103	Bacteria	Proteobacteria	Gammaproteobacteria	Legionellales	Legionellaceae	Legionella		
49	cgcgtgggaatctgcccttgggtt	378	348	1154	1301	6599	585635434	100	3e-135	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium		
50	cgcgtatacaatctaccttatactc	378	496	476	355	4607	359804291	90.15	4e-89	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	Crocinitomix		



wd

pipeline.R

config.conf

<dataname>.csv

bin

taxdmp

nodes.dmp

names.dmp

ncbi-blast-x.x.x

bin

db

readAssembler.R

apply.DigDeeper.R

usearch7

Hasher.R

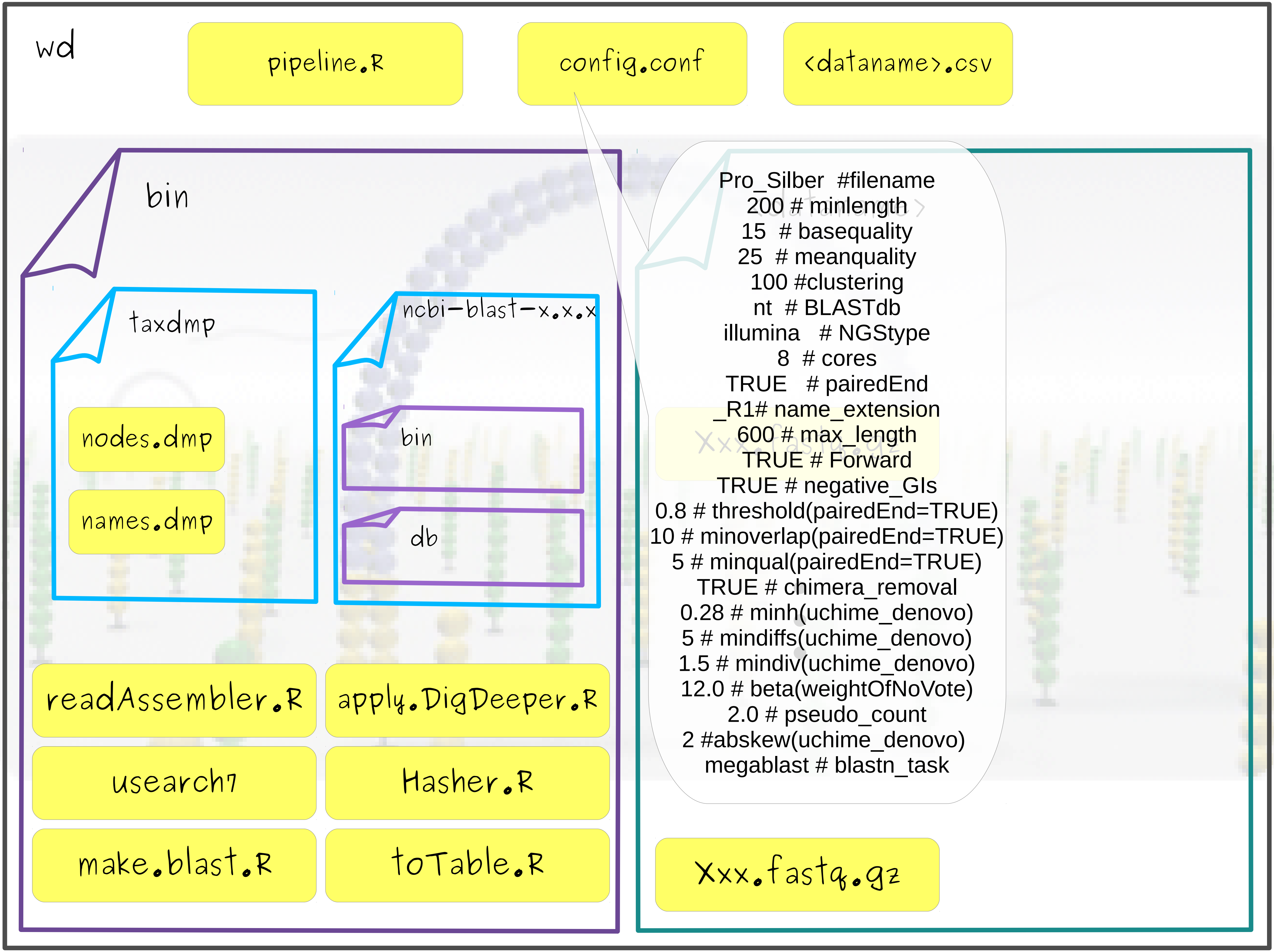
make.blast.R

toTable.R

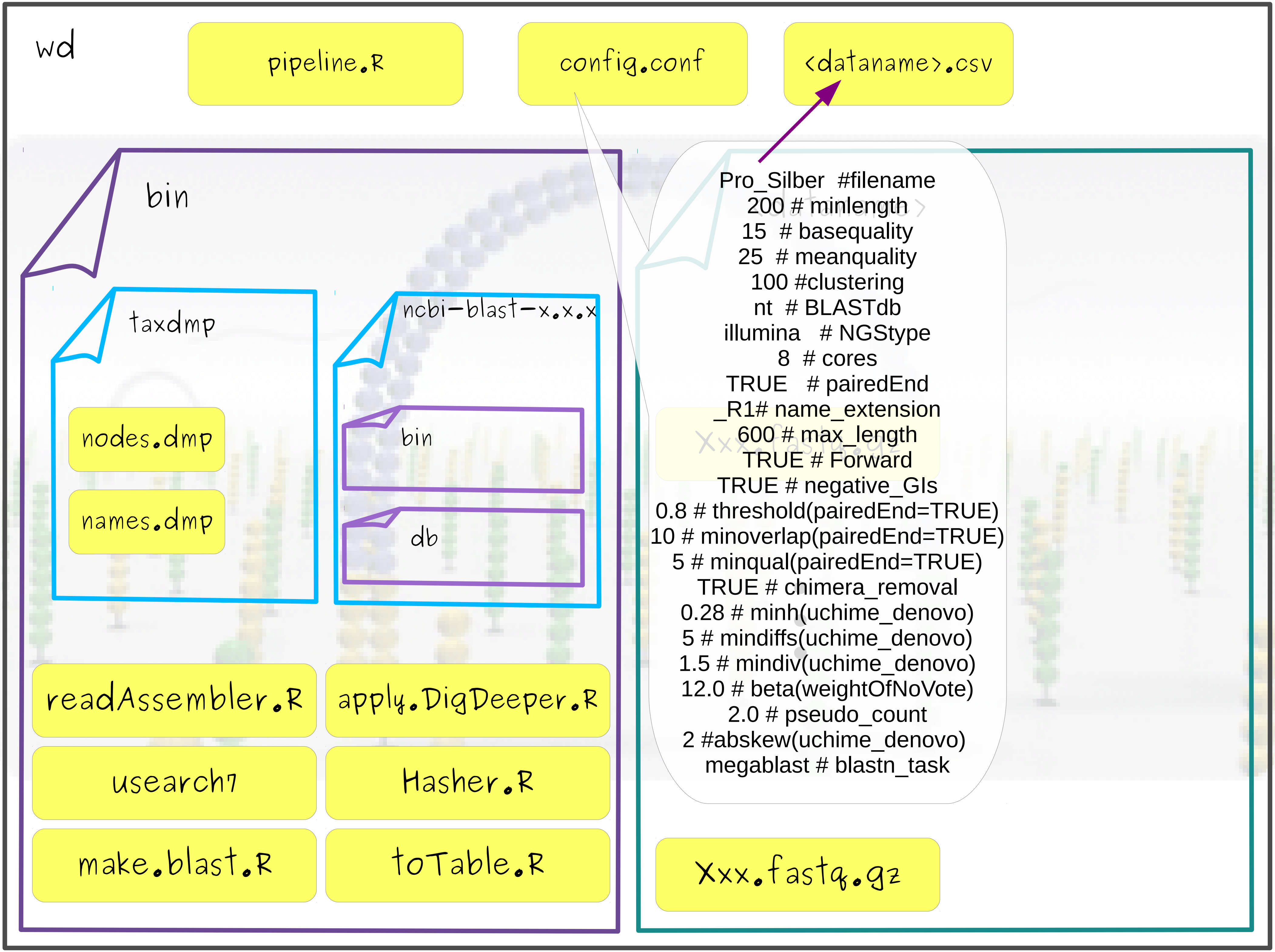
<dataname>

Xxx.fastq.gz

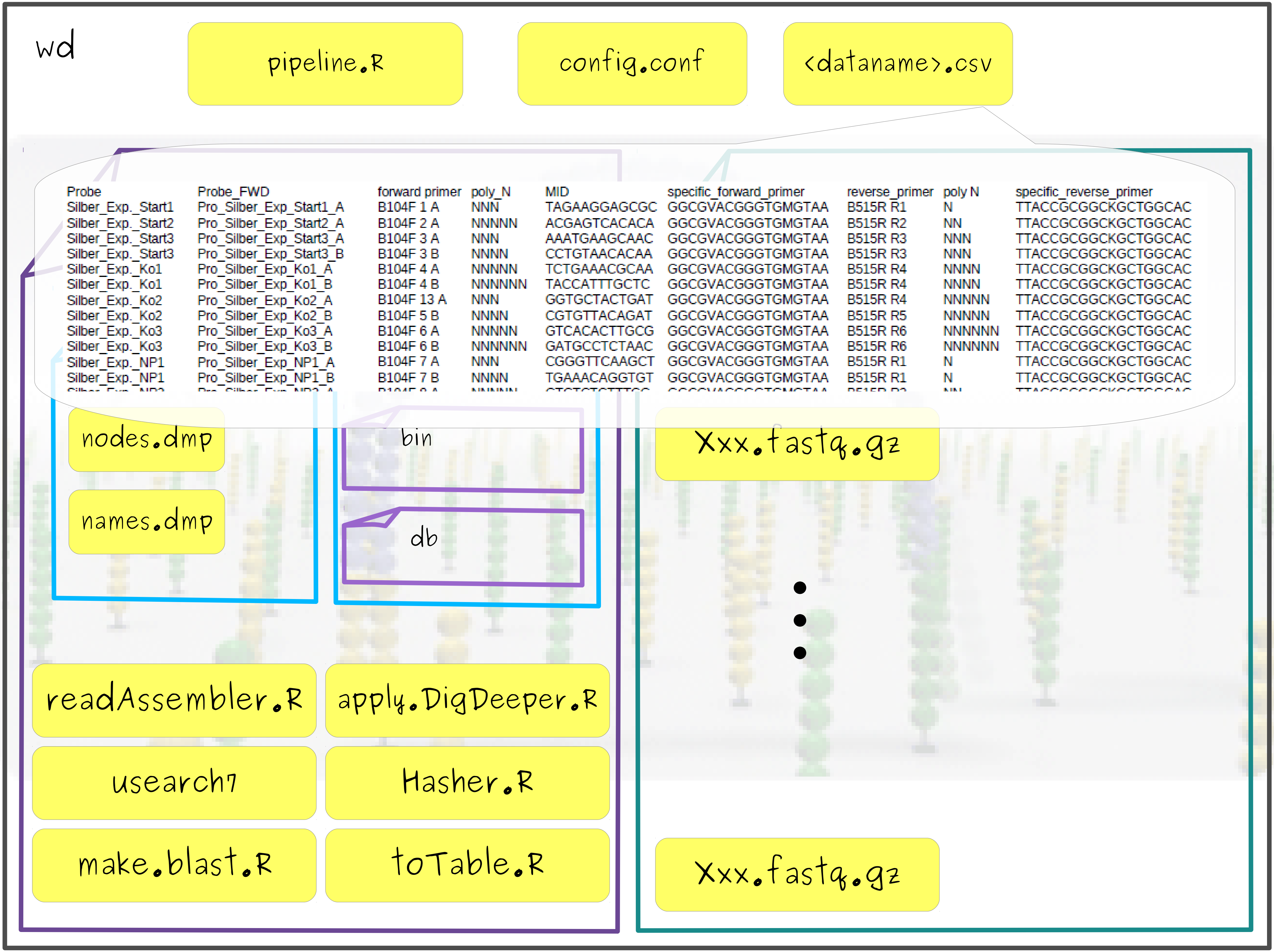
Xxx.fastq.gz



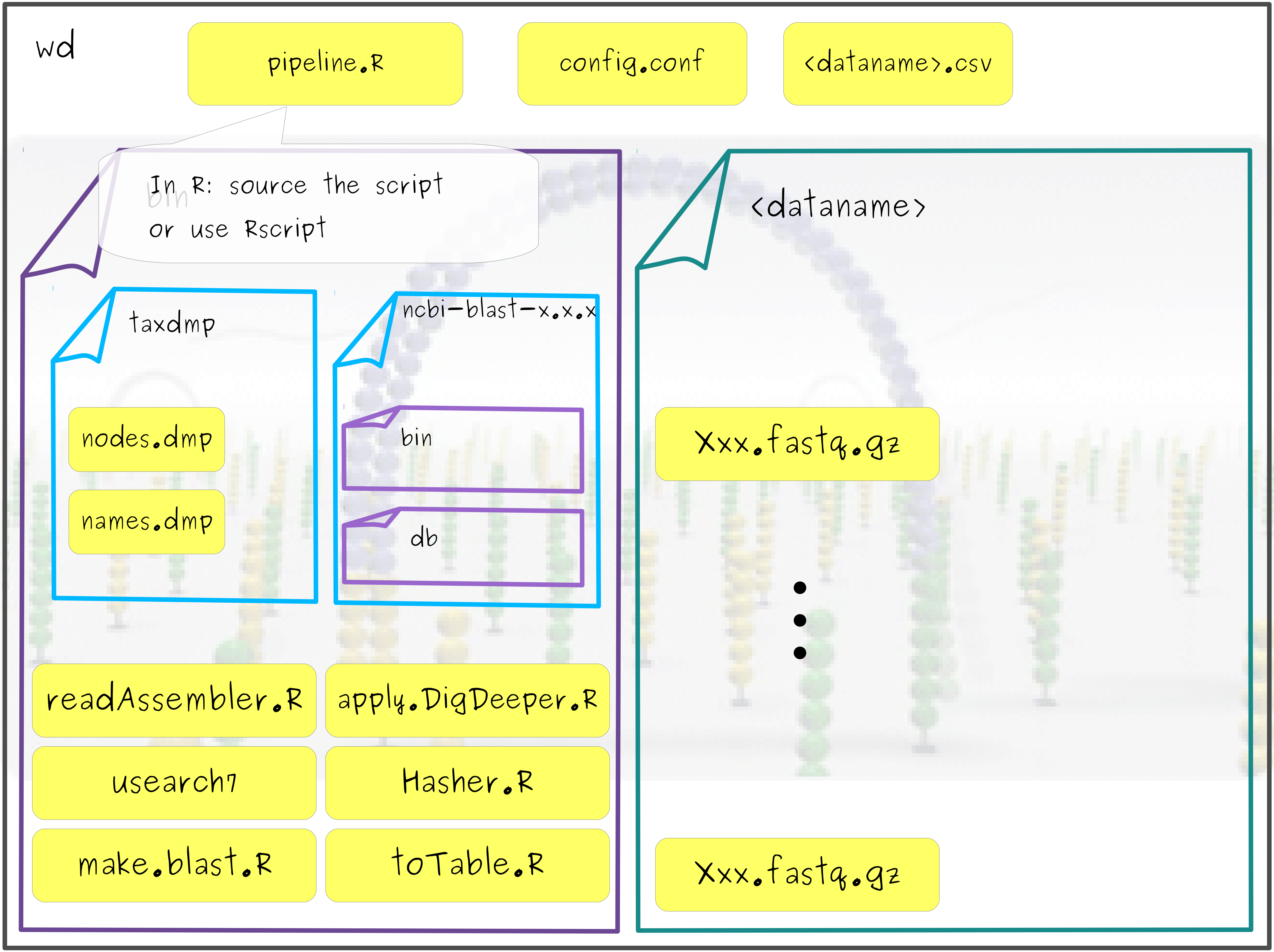
```
Pro_Silber  #filename
200 # minlength
15  # basequality
25  # meanquality
100 #clustering
nt  # BLASTdb
illumina   # NGStype
8  # cores
TRUE   # pairedEnd
_R1# name_extension
600 # max_length
TRUE # Forward
TRUE # negative_GIs
0.8 # threshold(pairedEnd=TRUE)
10 # minoverlap(pairedEnd=TRUE)
5 # minqual(pairedEnd=TRUE)
TRUE # chimera_removal
0.28 # minh(uchime_denovo)
5 # mindiffs(uchime_denovo)
1.5 # mindiv(uchime_denovo)
12.0 # beta(weightOfNoVote)
2.0 # pseudo_count
2 #abskew(uchime_denovo)
megablast # blastn_task
```



```
Pro_Silber  #filename
200 # minlength
15  # basequality
25  # meanquality
100 #clustering
nt  # BLASTdb
illumina   # NGStype
8  # cores
TRUE   # pairedEnd
_R1# name_extension
600 # max_length
TRUE # Forward
TRUE # negative_GIs
0.8 # threshold(pairedEnd=TRUE)
10 # minoverlap(pairedEnd=TRUE)
5 # minqual(pairedEnd=TRUE)
TRUE # chimera_removal
0.28 # minh(uchime_denovo)
5 # mindiffs(uchime_denovo)
1.5 # mindiv(uchime_denovo)
12.0 # beta(weightOfNoVote)
2.0 # pseudo_count
2 #abskew(uchime_denovo)
megablast # blastn_task
```



Probe	Probe_FWD	forward primer	poly_N	MID	specific_forward_primer	reverse_primer	poly N	specific_reverse_primer
Silber_Exp._Start1	Pro_Silber_Exp_Start1_A	B104F 1 A	NNN	TAGAAGGAGCGC	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start2	Pro_Silber_Exp_Start2_A	B104F 2 A	NNNNN	ACGAGTCACACA	GGCGVACGGGTGMGTAA	B515R R2	NN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_A	B104F 3 A	NNN	AAATGAAGCAAC	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Start3	Pro_Silber_Exp_Start3_B	B104F 3 B	NNNN	CCTGTAACACAA	GGCGVACGGGTGMGTAA	B515R R3	NNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_A	B104F 4 A	NNNNN	TCTGAAACGCAA	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko1	Pro_Silber_Exp_Ko1_B	B104F 4 B	NNNNNN	TACCATTTGCTC	GGCGVACGGGTGMGTAA	B515R R4	NNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_A	B104F 13 A	NNN	GGTGCTACTGAT	GGCGVACGGGTGMGTAA	B515R R4	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko2	Pro_Silber_Exp_Ko2_B	B104F 5 B	NNNN	CGTGTTACAGAT	GGCGVACGGGTGMGTAA	B515R R5	NNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko3	Pro_Silber_Exp_Ko3_A	B104F 6 A	NNNNN	GTCACACTTGCG	GGCGVACGGGTGMGTAA	B515R R6	NNNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._Ko3	Pro_Silber_Exp_Ko3_B	B104F 6 B	NNNNNN	GATGCCTCTAAC	GGCGVACGGGTGMGTAA	B515R R6	NNNNNN	TTACCGCGGCKGCTGGCAC
Silber_Exp._NP1	Pro_Silber_Exp_NP1_A	B104F 7 A	NNN	CGGGTTCAAGCT	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC
Silber_Exp._NP1	Pro_Silber_Exp_NP1_B	B104F 7 B	NNNN	TGAAACAGGTGT	GGCGVACGGGTGMGTAA	B515R R1	N	TTACCGCGGCKGCTGGCAC



wd

pipeline.R

config.conf

<dataname>.csv

In R: source the script
or use Rscript

bin

taxdmp

nodes.dmp

names.dmp

ncbi-blast-x.x.x

bin

db

readAssembler.R

apply.DigDeeper.R

usearch7

Hasher.R

make.blast.R

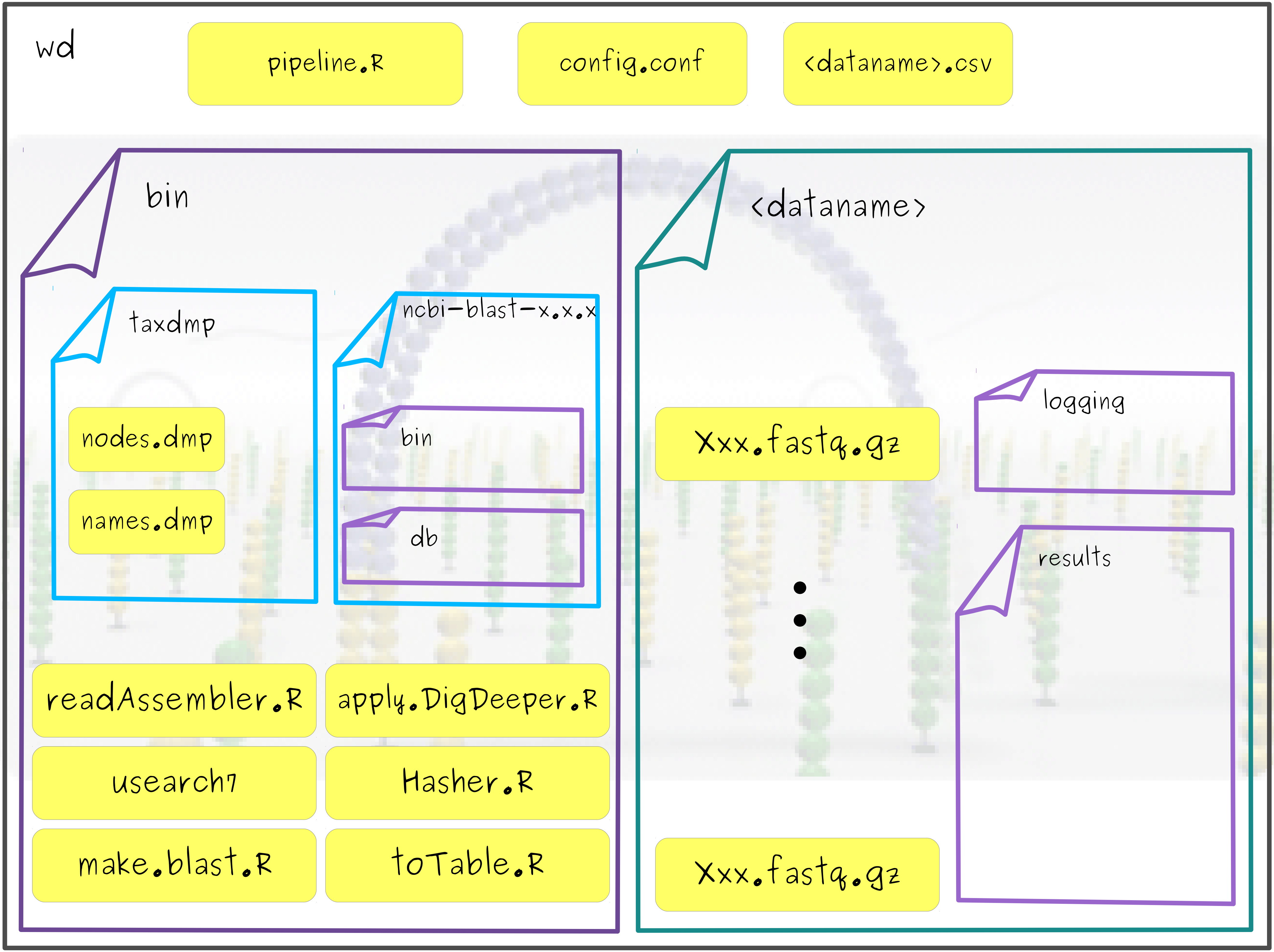
toTable.R

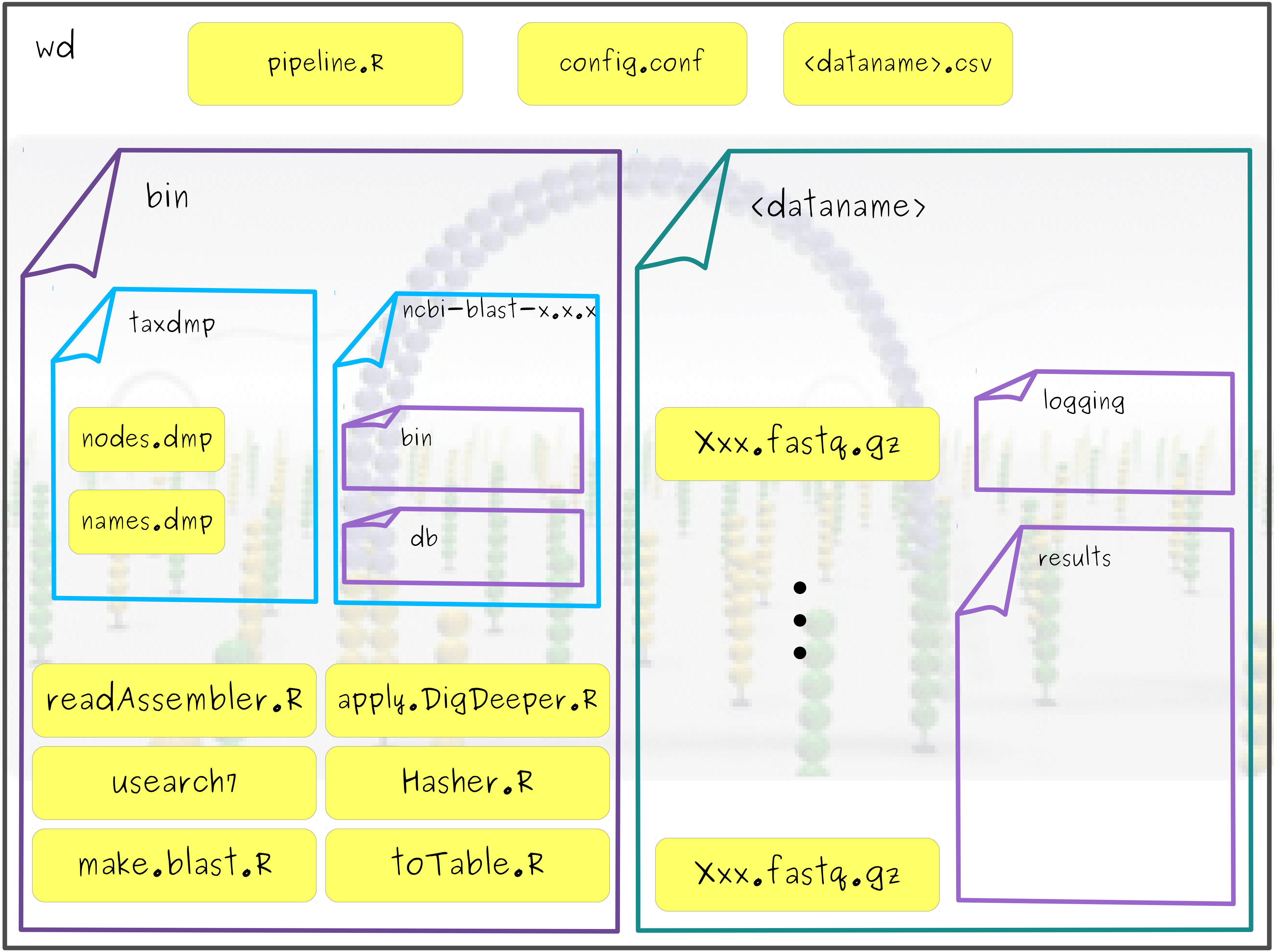
<dataname>

Xxx.fastq.gz

⋮

Xxx.fastq.gz


wd
pipeline.R
config.conf
<dataname>.csv
bin
taxdmp
nodes.dmp
names.dmp
ncbi-blast-x.x.x
bin
db
readAssembler.R
apply.DigDeeper.R
usearch7
Hasher.R
make.blast.R
toTable.R
<dataname>
Xxx.fastq.gz
logging
results
Xxx.fastq.gz



wd

pipeline.R

config.conf

<dataname>.csv

bin

taxdmp

nodes.dmp

names.dmp

ncbi-blast-x.x.x

bin

db

readAssembler.R

apply.DigDeeper.R

usearch7

Hasher.R

make.blast.R

toTable.R

<dataname>

Xxx.fastq.gz

•
•
•

Xxx.fastq.gz

logging

results