

# INPUT / OUTPUT

		FASTA file	Foreach split file	Less stringent	<b>Suggested</b>	More stringent
1.	No	File with > 1 MID *	Yes Split by multiplex identifier (MID)		2 mismatches	No mismatch
	No		Yes Trim tag from 5' (and 3') end	3+ mismatches	2 mismatches	No mismatch
	No		Yes Trim poly-A/T tails **	At least 5 bp long	At least 20 bp long	At least 50 bp long
2. TRIM ENDS	mRNA / cDNA data					
	Filter too short and/or too long reads			60 bp minimum	Mode ±2 SD	Mode ±1 SD
	Filter reads with ambiguous base N			≤ 5% (5 out of 100)	≤ 1% (1 out of 100)	No Ns
	Filter low complexity reads **			Entropy ≥ 40	Entropy ≥ 50	Entropy ≥ 70
	Filter read duplicates			5' only	5' and rev. compl.	5', 3' and rev. compl.
3. FILTER SEQUENCES	Remove sequence contaminants **			Cov ≥ 80, Ident ≥ 80	Cov ≥ 90, Ident ≥ 90	Cov ≥ 95, Ident ≥ 94
	FASTA file			Less stringent	<b>Suggested</b>	More stringent

\* MID = Multiplex identifier as used for 454/Roche sequencing of multiple samples

\*\* Optional

## PROGRAM

<span style="display:inline-block; width:15px; height:15px; background-color:#80CBC4;"></span> Perl script	<span style="display:inline-block; width:15px; height:15px; background-color:#00BCD4;"></span> PRINSEQ
<span style="display:inline-block; width:15px; height:15px; background-color:#C8E6C9;"></span> TagCleaner	<span style="display:inline-block; width:15px; height:15px; background-color:#FF9800;"></span> DeconSeq