Task Group	Task	Description	Number papers reporting task	Number papers not reporting software	Total number of software tools	Total number of software functions	Number of papers performing manually
Read preparation	***quality control***	*Generating a report of sequence quality information from a sample or set of samples - no modification is done to data*	**19**	**0**	**4**	**4**	**0**
	adapter trimming	*Trimming of sequencing adapters*	**9**	**1**	**6**	**6**	**0**
	demultiplexing	*Separation of sequences from a mixed pool into separate pools based on the occurence of a unique set of bases (index or tag)*	**55**	**17**	**16**	**19**	**0**
	pair merging	*The assembly of mate pair reads into a single contig*	**63**	**1**	**10**	**18**	**0**
	quality trimming	*The removal of bases from either or both ends of sequences in a pool based on quality scores*	**20**	**1**	**8**	**10**	**0**
	mate pairing	*The identification and syncronisation of mate pair reads between two samples, often involving arranging reads in identical orders and/or removal of reads without a mate pair*	**3**	**0**	**3**	**3**	**0**
	primer trimming	*Trimming of PCR primers*	**66**	**8**	**15**	**17**	**0**
	reverse complementation	*Reverse complementing the sequences in a pool*	**7**	**3**	**2**	**2**	**0**
	sequence conversion	*Converting sequences from fastq to fasta*	**3**	**0**	**2**	**3**	**0**
	length trimming	*The removal of bases from either or both ends of sequences in a pool, either the removal of a fixed number of bases or the removal of a variable number of bases to reduce sequences to a standard length*	**10**	**3**	**6**	**7**	**0**
	pair concatenation	*Concatenating mate pair reads into a single contig (where reads don't overlap)*	**8**	**4**	**4**	**4**	**0**
	assembly	*The assembly of reads into contigs, applied when more than one pair of overlapping fragments have been metabarcoded*	**6**	**0**	**4**	**4**	**0**
	degapping	*Removal of gaps from sequences*	**1**	**0**	**1**	**1**	**0**
Sequence processing	***dereplication***	*The removal of duplicate reads to retain only unique sequences in a pool; often the total number of copies of a sequence is recorded in the header of the retained sequence*	**58**	**10**	**11**	**19**	**0**
	size sorting	*The sorting of a fasta file according to a size annotation in the header*	**10**	**2**	**3**	**4**	**0**
Filtering	***quality filtering***	*Removal and/or trimming of sequences from a pool based on quality information. Also often converts from fastq to fasta.*	**81**	**11**	**20**	**27**	**0**
	similarity filtering	*Removal of sequences based on similarity to an alignment, either based on sequence identity or alignment position*	**9**	**1**	**4**	**4**	**0**

	length filtering	*The removal of sequences from a pool that are less than, more than, or fall within or outside of a specified length threshold or thresholds*	**54**	**21**	**17**	**23**	**0**
	preclustering	*Reduction of sequence variation in a dataset prior to further processing - a form of denoising*	**12**	**1**	**3**	**6**	**0**
	denoising	*The removal of reads containing putative PCR or sequencing errors based on statistical assessment*	**18**	**1**	**8**	**8**	**0**
	normalisation	*A process by which the number of sequences for each of a set of samples is reduced where necessary such that the output set of samples all have the same number of sequences while maintaining the relative frequencies of OTUs*	**2**	**0**	**1**	**1**	**1**
	chimera filtering	*The filtering of putative chimeric assemblies from a pool of mate paired reads*	**63**	**4**	**6**	**16**	**1**
	translation filtering	*Removal of sequences from a set of sequence based on their translation, usually removing sequences with inframe stop codons or frameshifts due to erroneous indels or substitutions caused by sequencing errors*	**22**	**3**	**11**	**12**	**0**
	frequency filtering	*Removal of sequences based on their frequency in a pool*	**51**	**37**	**11**	**15**	**1**
	taxonomy filtering	*Removal of sequences based on an assigned taxonomy or a taxonomic classification*	**9**	**5**	**1**	**1**	**1**
	mistag filtering	*Removal of sequences based on putative tagging errors*	**3**	**1**	**1**	**1**	**0**
Data generation	***OTU delimitation***	*The grouping of a set of sequences into OTUs by some method*	**84**	**5**	**12**	**22**	**0**
	OTU mapping	*The mapping of sequences to OTUs to provide read counts for each OTU*	**30**	**3**	**7**	**11**	**0**
	uncurated taxonomic assignment	*The assignment (identification or classification) of taxonomy to OTUs using a global uncorated reference database (e.g. GenBank, BOLD)*	**55**	**2**	**11**	**13**	**0**
	reference taxonomic assignment	*The assignment (identification or classification) of taxonomy to OTUs using a purpose-built and/or specially curated reference set of sequences*	**60**	**9**	**18**	**23**	**1**

Table 1: *Table of all bioinformatic tasks performed across the core papers set. Tasks are grouped into four groups by broad purposes, and a detailed definition of each task is given along with summary statistics of the implementation of each task across the 111 papers. For a list of the software used for each task, Table S1 is an expanded version of this table.*