



# Jonah Ventures NGS Report Files Explained

## Summary data (read-data file)

The % of base pairs in the sequence that match to the species in the library.

The detected sequence.

Sample identifiers  
Numbers after the decimal point represent lab replicates of the same sample.

Exact Sequence Variant unique identifier. Use this to cross reference with the ESV file

Consensus Taxonomy  
Taxonomic ranking from the library matched to each sequence.

Number of species matching the sequence at a given level. Reference the ESV file to see which species matched.

ESV	Family	Genus	Species	%match	# Species	Sequence	S10001.1	S10002.1	S10003.1	S10003.2
ESV_000080	Fagaceae	Quercus	NA	100	6	AAAAAG...	3066	1340	5780	3462
ESV_000818	Orobanchaceae	Orthocarpus	Orthocarpus luteus	100	1	AAAAAG...	0	2582	249	0
ESV_000006	Poaceae	NA	NA	100	77	GAAAAG...	584	101	0	1039
ESV_000050	Poaceae	NA	NA	100	12	GAAAAG...	1328	0	0	0
ESV_000308	Polygonaceae	Polygonum	NA	98	2	AAAAAG...	902	0	0	278
ESV_000027	Polygonaceae	Eriogonum	Eriogonum umbellatum	98	1	ATAAAG...	341	124	282	381
ESV_000520	Polygonaceae	Fallopia	Fallopia multiflora	96	1	AAAAAG...	126	0	0	994

"NA" at a taxonomic ranks means multiple species present in the sample differ in that taxonomic rank

The number in each cell is the absolute number of times a given sequence was read by the sequencer.

## ESV detail (esv-data file)

Exact Sequence Variant

Index identification number. Each sample has an individual index number.

Species Taxonomy

The % of base pairs in the sequence that match to the species. 100% indicates all base pairs matched

ESV	Accession	Phylum	Order	Family	Genus	Species	%match
ESV_000818	EF103883	Streptophyta	Lamiales	Orobanchaceae	Orthocarpus	Orthocarpus luteus	100
ESV_000080	AB043749	Streptophyta	Fagales	Fagaceae	Quercus	Quercus serrata	100
ESV_000080	AB043744	Streptophyta	Fagales	Fagaceae	Quercus	Quercus mongolica	100
ESV_000080	KX667936	Streptophyta	Fagales	Fagaceae	Quercus	Quercus robur	100
ESV_000080	KF718308	Streptophyta	Fagales	Fagaceae	Quercus	Quercus pubescens	100
ESV_000080	MG709405	Streptophyta	Fagales	Fagaceae	Quercus	Quercus stellata	100
ESV_000080	AF327594	Streptophyta	Fagales	Fagaceae	Quercus	Quercus petraea	100

The sequence represented by an ESV is generated from a denoising algorithm. ESVs are generated by examining the most abundant sequences and then removing sequences that have random differences in them as opposed to consistent differences. Those with consistent differences, even just one base difference, become their own ESV instead of being lumped together.

100% indicates all base pairs matched  
99% is roughly 1-2 mismatches per sequence  
95% is about 7 base pair mismatches.