

## GRIIDC Genetic Data Guidance Table

Genetic Data Type	Existing National Data Archive						Files to submit to GRIIDC					
	Raw reads NCBI SRA**	NCBI GenBank	NCBI-GEO	Tree BASE	NCBI Popset	MG-Rast	Assembled contigs***	Assembled scaffolds***	Annotated assembly of FASTA files***	Raw sequence data	File with sample info*	Other files
genome assemblies and annotations	x						x	x	x	x	x	
transcriptome assemblies and annotations	x						x		x		x	
gene expression	x		x				x	x	x		x	
methylation	x		x				x	x	x		x	
SNP	x										x	matrix of SNPs
microsatellite											x	matrix of microsat scores
comet assay											x	images showing results
microarray											x	
quantitative PCR											x	data/curve for calculating concentration
metagenomics	x					x					x	
sanger sequences		x									x	
terminal restriction fragments (RFLP)											x	matrix of how fragments were scored
aligned sequences (evolution/ population genetics)				x	x						x	nexus or aligned fasta
microbial ribosomal sequence data	x	x									x	

\* All data should have sample information – generally this is a table with sample IDs and accession numbers AND either 1) if samples are collected in the field: latitude, longitude, depth, date collected, and other environmental data collected; OR 2) if data are derived from laboratory experiments exposing organisms to different treatments in the lab include treatment data.

\*\* SRA data should have a reference to the home page of the data in the Supplemental Information section of Dataset Submission.

\*\*\* PI may not have always annotated the assembly or created scaffold but if they do they should submit these to GRIIDC.

Note: NCBI has a Transcriptome Shotgun Assembly (TSA) Database; this is not user-friendly for all organisms. PIs should deposit assemblies either with GRIIDC or TSA, or they may do both.