

Long-read, whole genome sequencing and metagenomics of Atlantic Rainforest bodies of water

Samir VF Atum^{1,2}, Douglas MM Soares¹, João C Setubal², Renato S Freire¹ and Cassius V Stevani¹

1. Departamento de Química Fundamental, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil
2. Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Background:

- The Atlantic Forest is an important Brazilian biome, where most of the population, industrial centers, and agricultural production is located. As such, it's also Brazil's most degraded biome.
- Despite its importance, there haven't been many metagenomic studies conducted in the biome, and none in its waters.
- Whole Genome Sequencing (WGS) metagenomics provide not just the taxonomic distribution of a microbiome, but also physiological information, metabolic functions, and enables the elucidation of whole genomes (binning).
- Two natural and two artificial bodies of water were analysed.

Methodology:

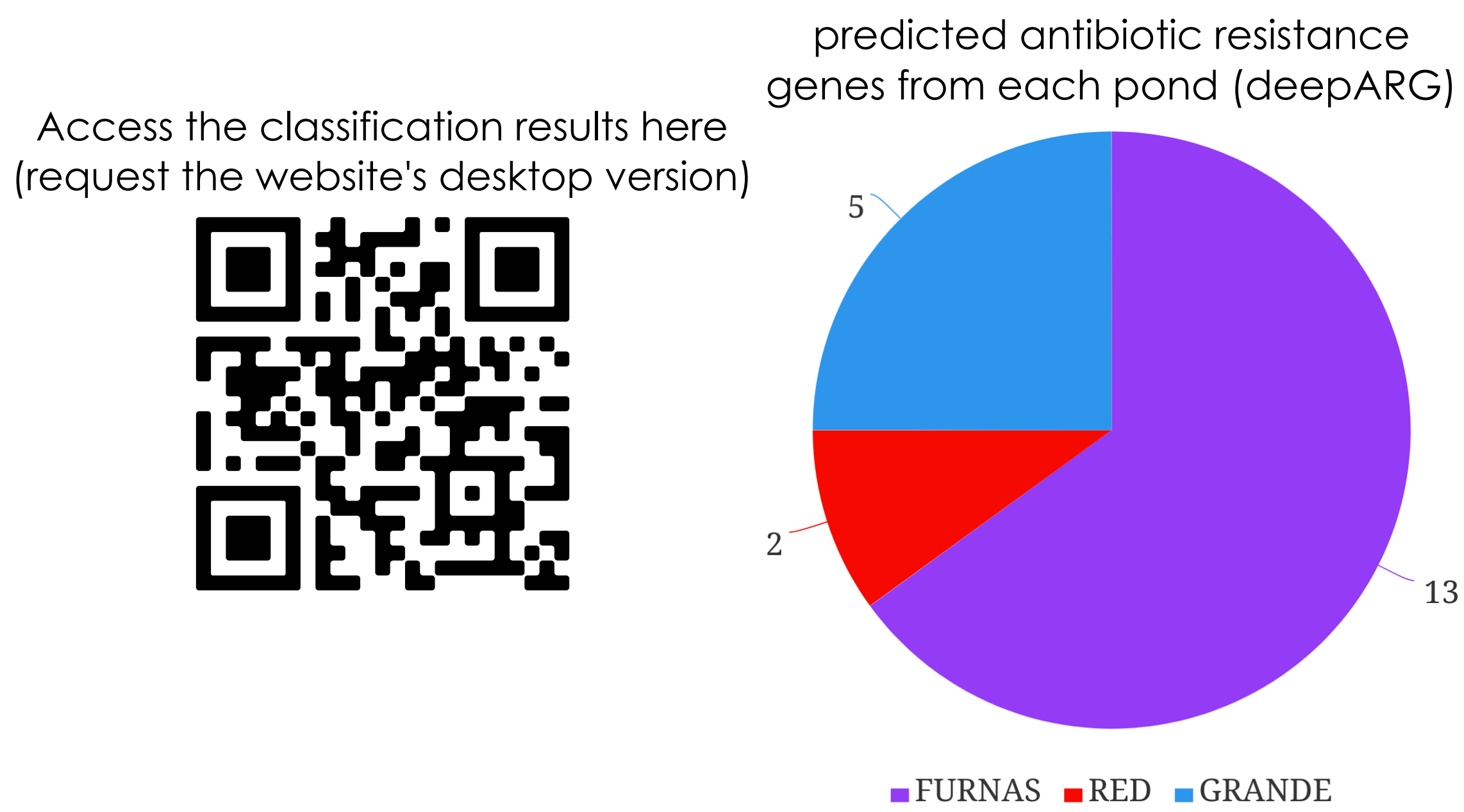
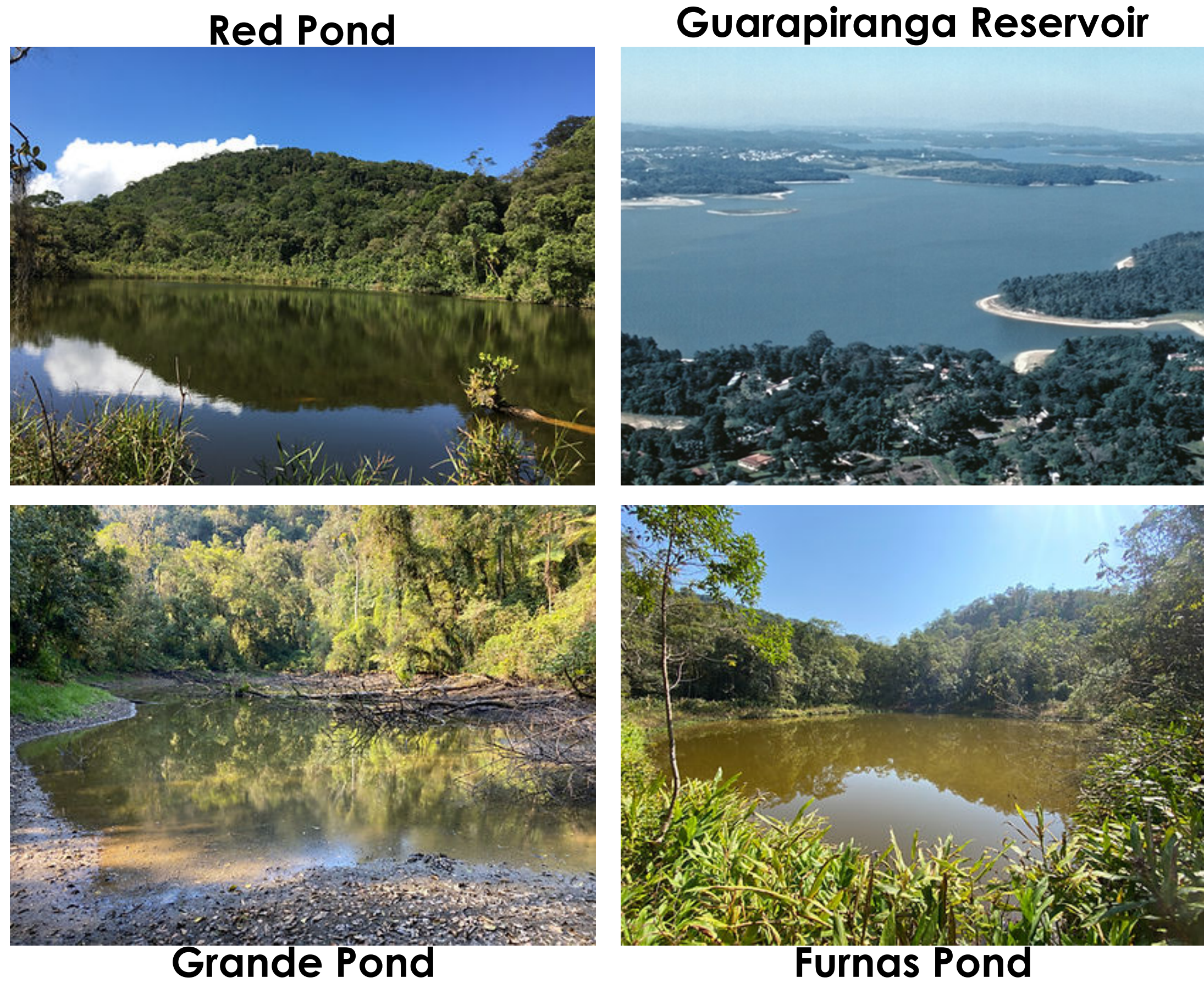
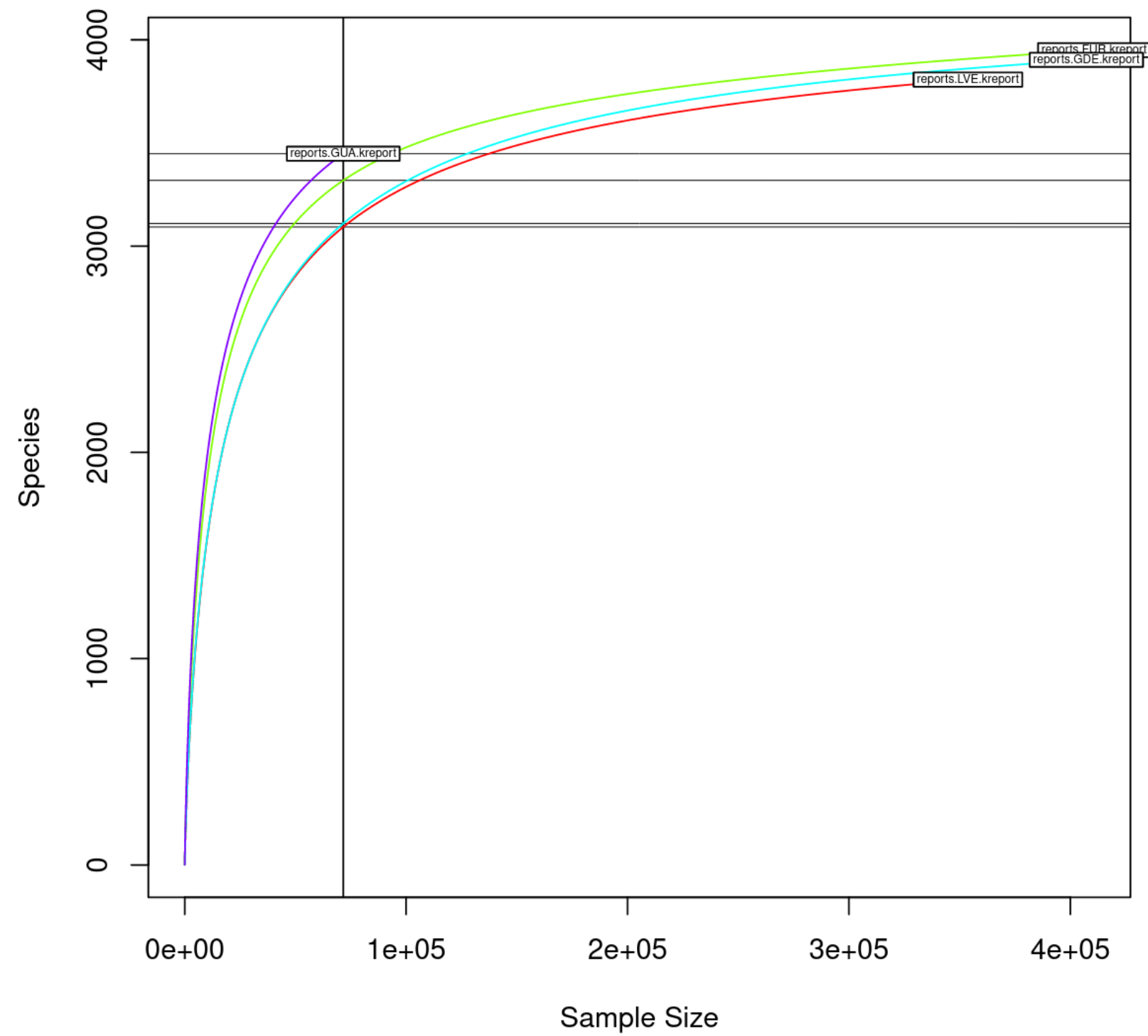


sequencing

- Water was sampled from the surface.
- Vacuum filtration of pooled water samples (1 L) using 0.22 µm filters.
- gDNA extraction using Dneasy PowerWater kit (QIAGEN).
- DNA concentration and purity evaluation by both spectrophotometer and fluorometer.
- Sequencing library preparation from 3 g of gDNA.

Results:

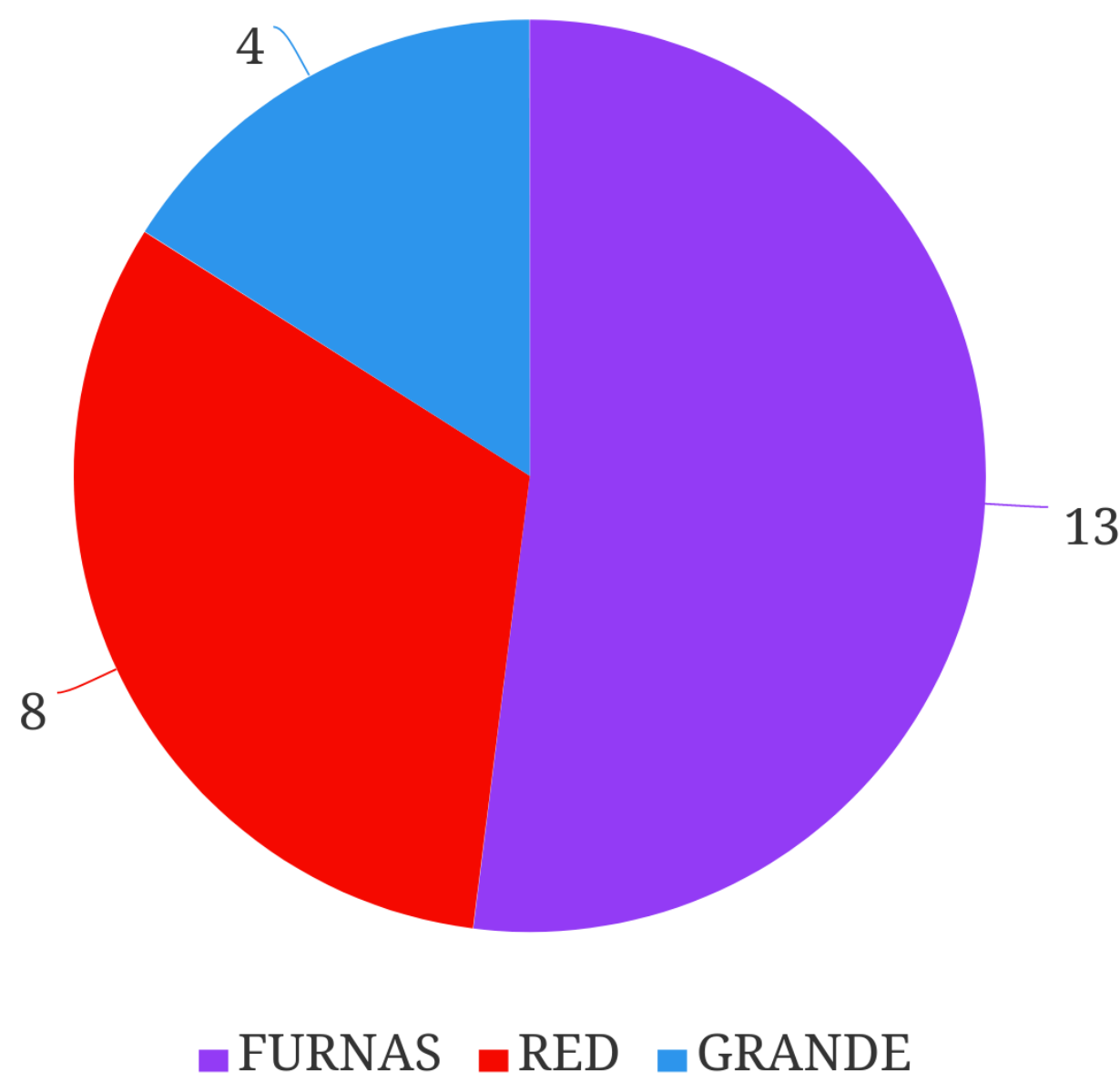
rarefaction graphs (kraken2 and vegan R package)



Most abundant contigs in the assembly and their classification:

Guarapiranga reservoir			RED pond			GRANDE pond			FURNAS pond		
Read count	Contig size (bp)	Classification	Read count	Contig size (bp)	Classification	Read count	Contig size (bp)	Classification	Read count	Contig size (bp)	Classification
1,239	251,358	<i>Aquirufa nivalsilvae</i>	56,131	3,787,723	<i>Leptolyngbya</i>	8,463	2,210,778	<i>Phenylobacterium*</i>	23,486	2,223,420	<i>Niabella</i>
1,212	155,302	<i>Cyanophage S-RIM32</i>	6,096	1,193,138	<i>Fuerstia marisgermanicae</i>	8,254	2,142,414	<i>Rhizobiaceae*</i>	19,267	1,407,122	<i>Candidatus Nanopelagicus</i>
917	102,618	<i>Delftia sp. Cs1-4</i>	5,622	1,035,040	<i>Fuerstia marisgermanicae</i>	7,889	1,287,417	<i>Rhizobiaceae*</i>	16,368	906,744	<i>Lacunisphaera limnophila</i>
784	351,605	<i>Oligotropha carboxidovorans</i>	3,048	713,816	<i>Rickettsia*</i>	6,049	427,251	<i>Synechococcus phage S-IOM18</i>	13,525	748,692	<i>Microcystis</i>
674	126,556	<i>Comamonadaceae</i>	2,791	636,565	<i>Bacteria</i>	4,847	130,931	<i>Thermosynechococcus</i>	8,895	505,884	<i>Heterosigma akashiwo virus</i>

MAGs obtained from each pond



The only Metagenome-Assembled Genome (MAG) classified as a described species was a *Microcystis novacekii* genome, from Red Pond

