

Unique region identification in genomes using a k-mer approach

<https://github.com/bcgsc/unikseq>

Warren RL, Allison MJ, Lopez ML, Acharya-Patel N, Coombe L, Yang CL, Helbing CC, Birol I

NON-DESTRUCTIVE PRECISION GENOMICS FOR ENVIRONMENTAL IMPACT TRACKING IN A GLOBAL CLIMATE CHANGE ERA



ENVIRONMENTAL DNA (eDNA)

GENETIC MATERIAL EXTRACTED FROM ENVIRONMENTAL SAMPLES



RAPID
ACCURATE
COST-EFFECTIVE
NON-INVASIVE
ACTIONABLE

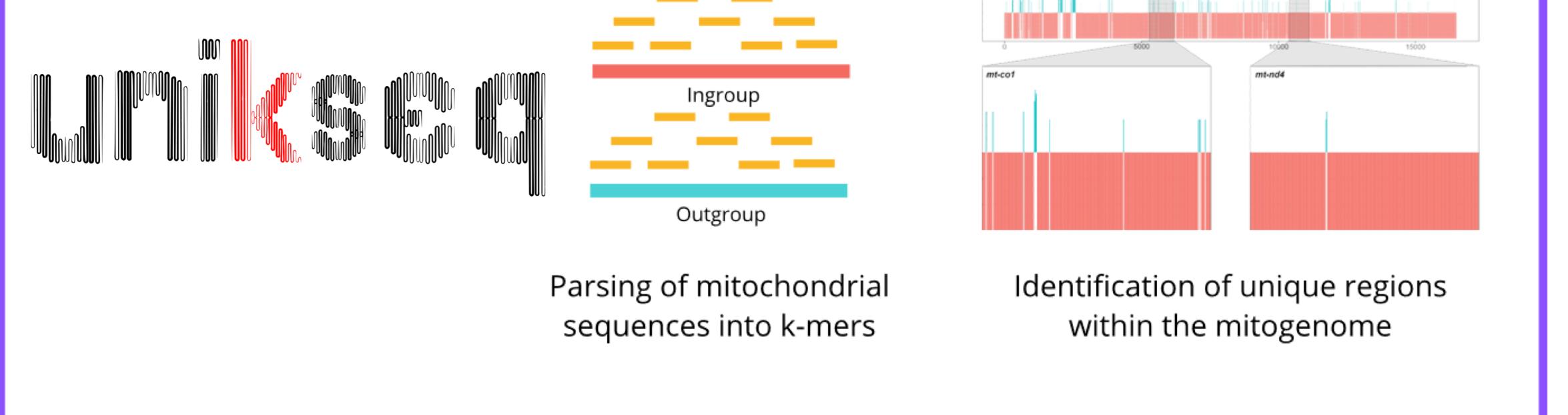
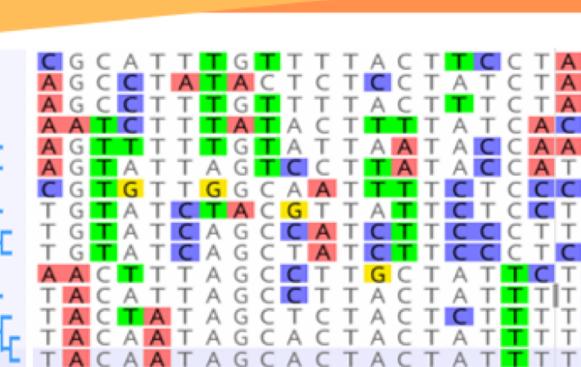
Relevant species identification



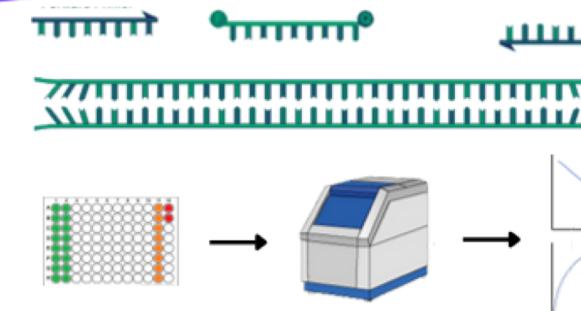
Mitogenome acquisition from public repositories or direct sequencing



Multiple mitogenome sequence alignment and phylogenetic analysis



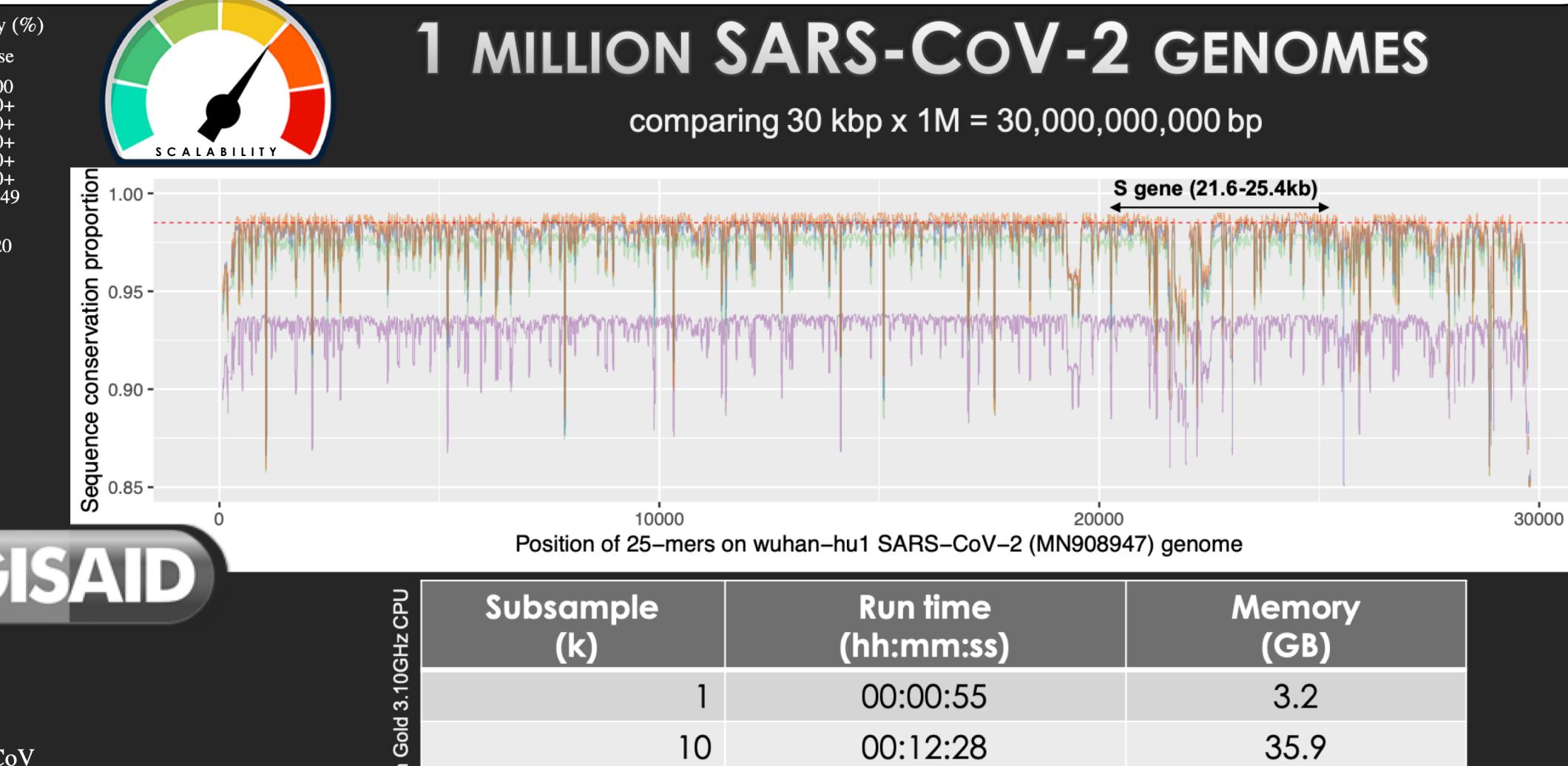
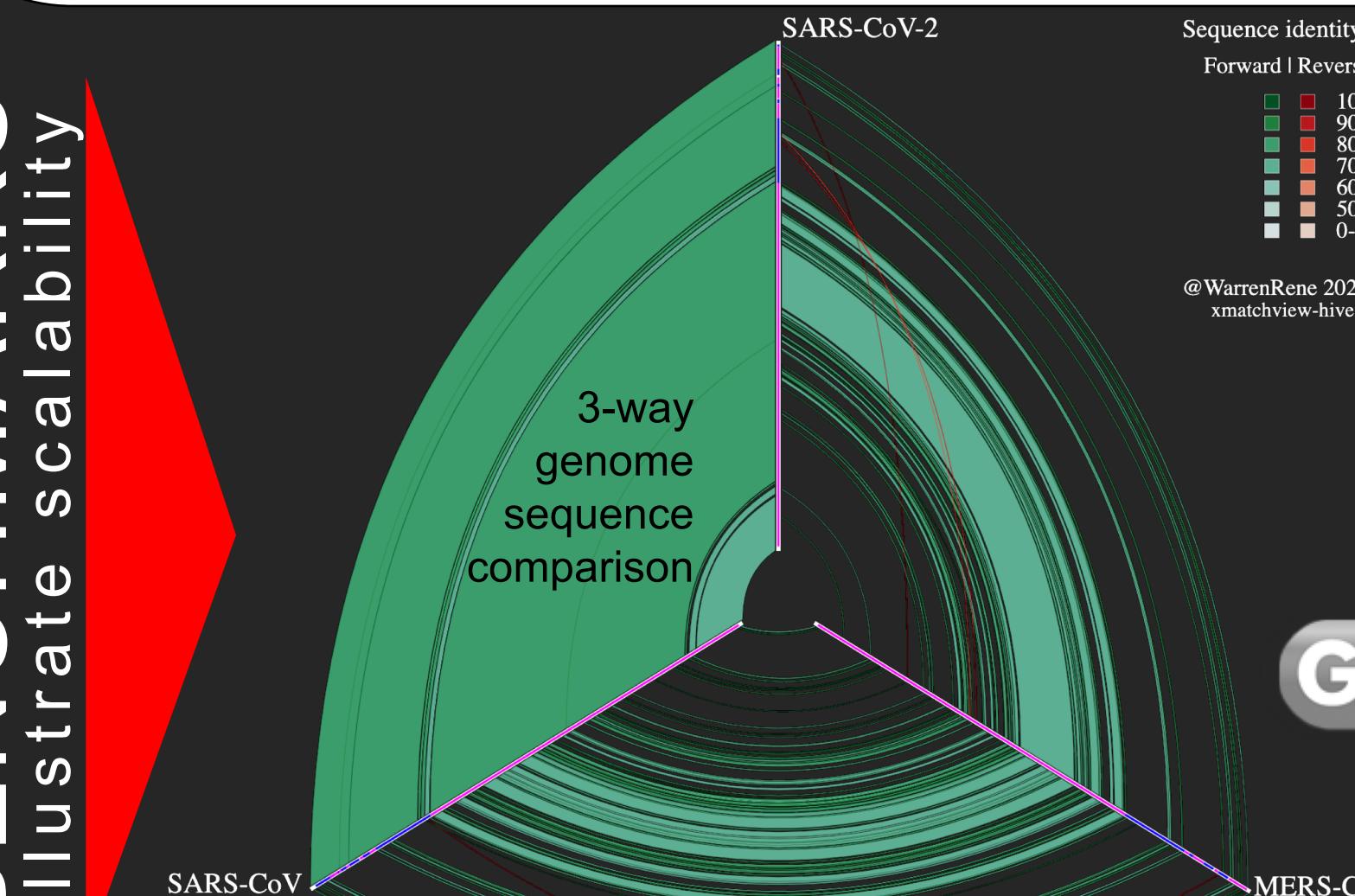
eDNA assay design and *in vitro* analysis



eDNA assay *in situ* (field) validation



BENCHMARKS illustrate scalability



Subsample (k)	Run time (hh:mm:ss)	Memory (GB)
1	00:00:55	3.2
10	00:12:28	35.9
100	02:02:11	323.5
1000	11:27:02	1238.5

FUNDING



Canadian Institutes of Health Research



Instituts de recherche en santé du Canada



Genome Québec



Genome Canada

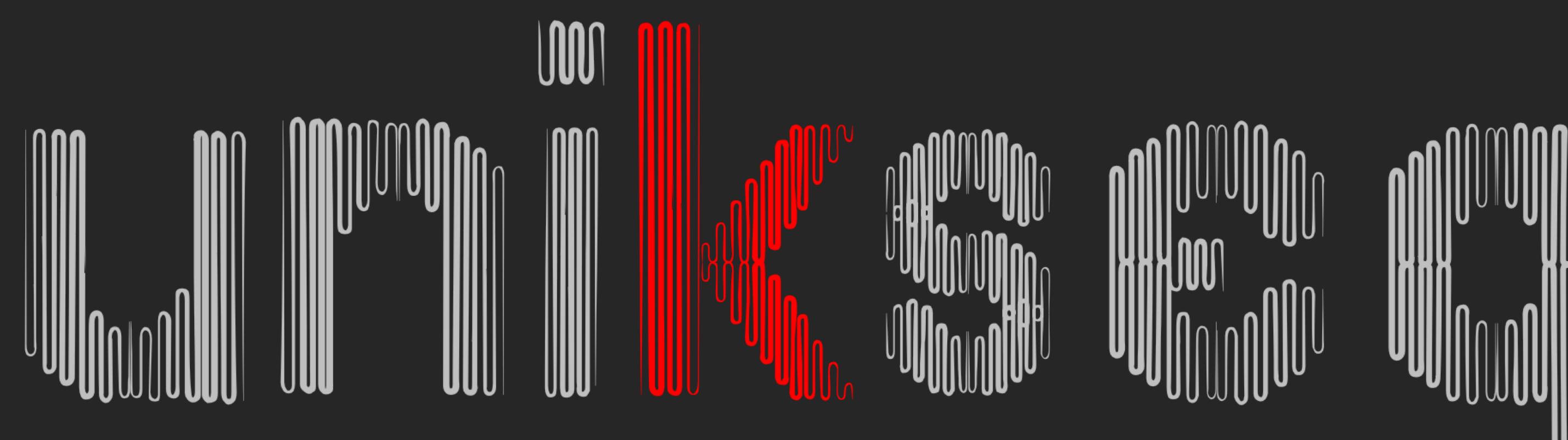


JOHN JAMBOR KNOWLEDGE FUND

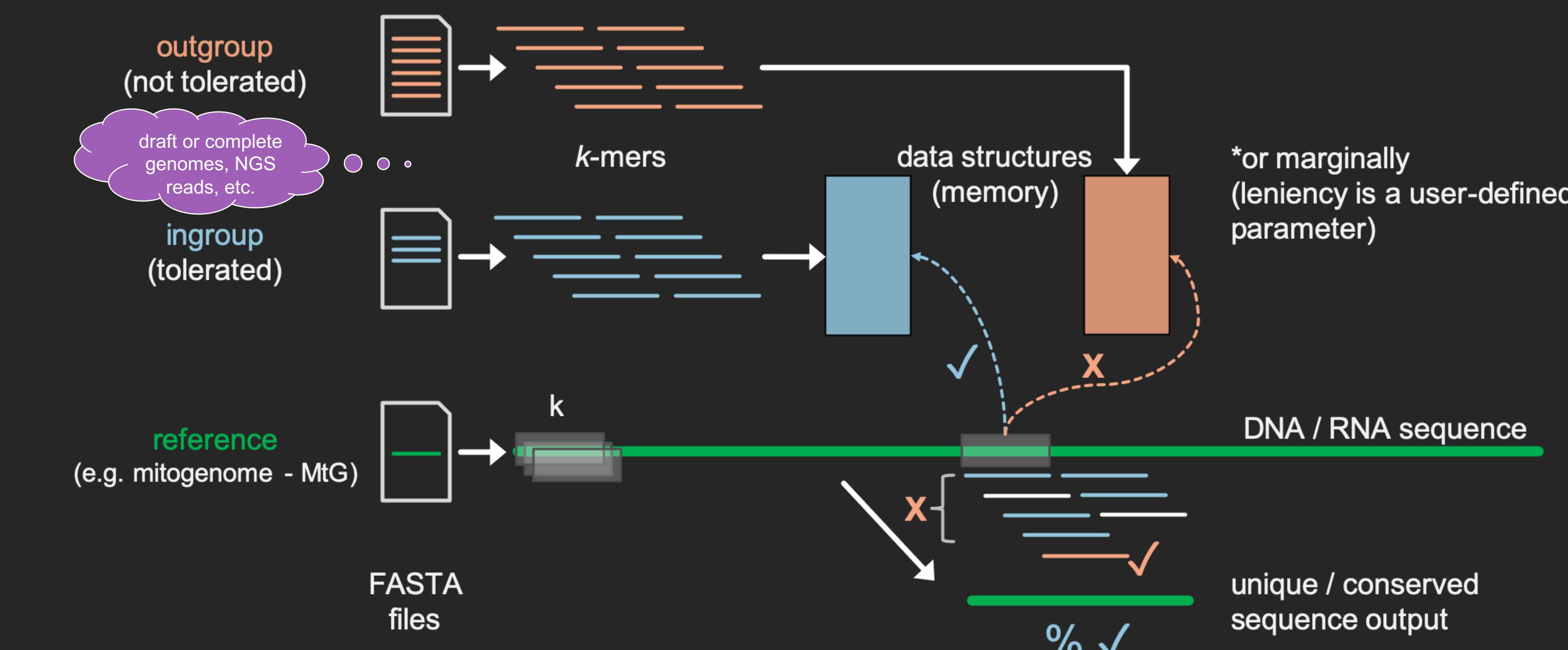
<http://warrenrl.github.io>



<http://birol-lab.ca>



UNIQUE REGION IDENTIFICATION IN A REFERENCE, TOLERATED IN INGROUP, NOT* IN OUTGROUP



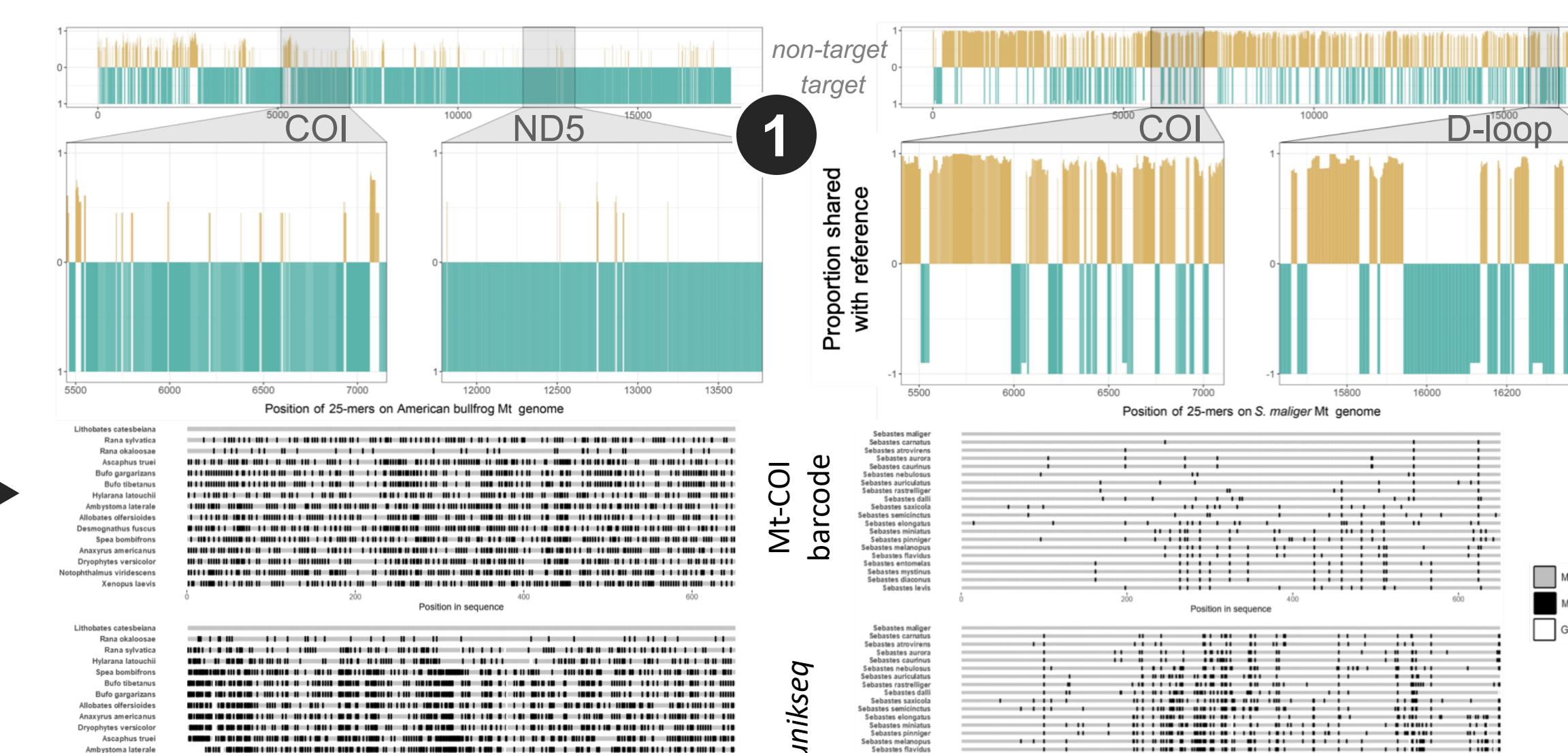
Features

- IDENTIFY MISSING GENOMIC REGIONS
- INSPECT SEQUENCE CONSERVATION
- EXTRACT UNIQUE SEQUENCES

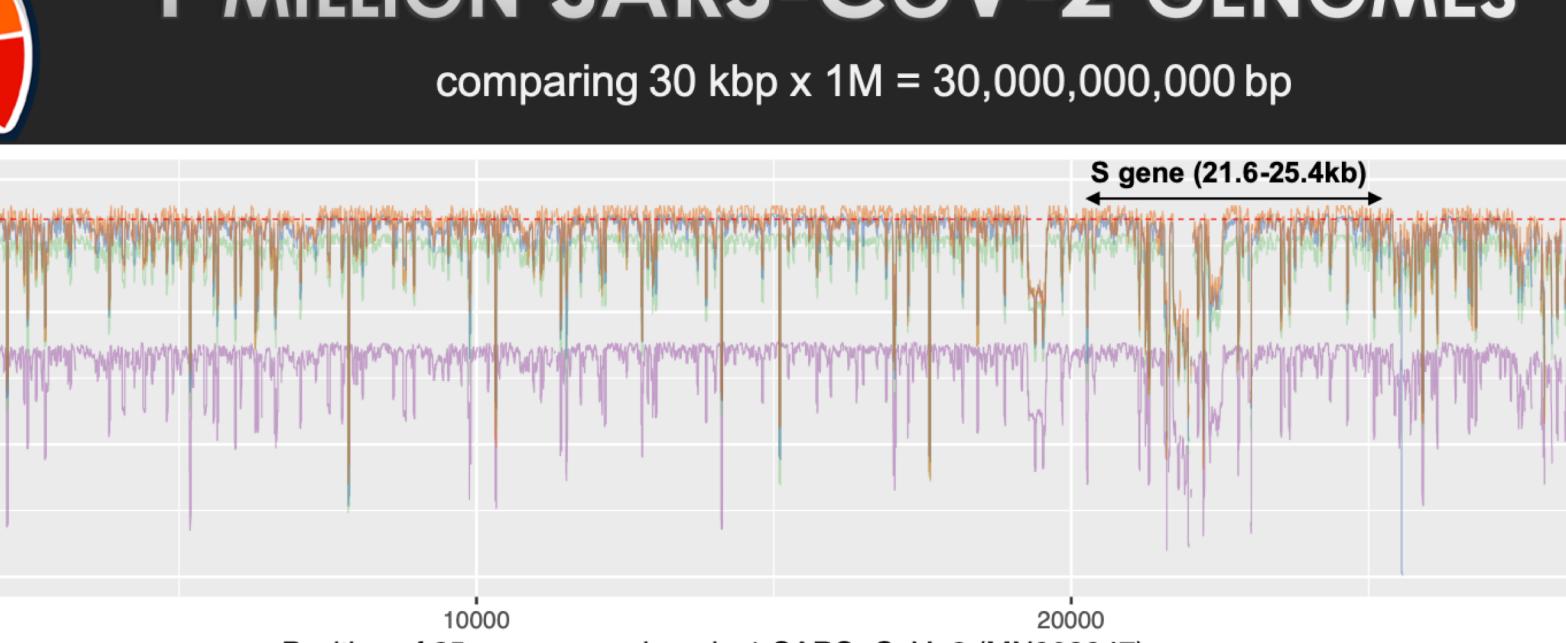
eDNA ASSAY DESIGN

Unlocks multiple MtG regions beyond Mt-COI barcode

- 1 Bullfrog strong assay design potential
- 2 Rockfish closely related taxa limits MtG qPCR design



1 MILLION SARS-CoV-2 GENOMES



Subsample	Run time	Memory
1	00:00:55	3.2
10	00:12:28	35.9
100	02:02:11	323.5
1000	11:27:02	1238.5

