Meta-kaks: a tool to measure natural selection strength on a metagenome

Vito A. Cantu, Cynthia B. Silveira, Robert A. Edwards

San Diego State University

CSRC — an interdisciplinary gateway to the future

COMPUTATIONAL SCIENCE RESEARCH CENTER, SAN DIEGO STATE UNIVERSITY

What is the Ka/Ks ratio?

The Ka/Ks ratio is a measure of the strength of natural selection over a protein coding gene. It is calculated by dividing the number of non-synonymous mutations (those that result in an amino-acid change) per non-synonymous site (Ka) over the number of synonymous mutation per synonymous site (Ks). As such Ka/Ks is already normalized by the length of the protein and its amino-acid composition, so Ka/Ks ratios obtained from different genes are directly comparable.

AA	Val	Met	Arg	Thr	Total
Codon	G T T	A T G	A A G	A C C	
non-syn sites	1 1 0	1 1 1	1 1 ⅔	1 1 0	9⅔
syn sites	0 0 1	0 0 0	0 0 ⅓	0 0 1	2⅓

Mutated to	Val	**Leu**	Arg	Thr	Total
Codon	G T **A**	**C** T G	A A **A**	A C C	
non-syn substitutions	0 0 0	1 0 0	0 0 0	0 0 0	1
syn substitutions	0 0 1	0 0 0	0 0 1	0 0 0	2

$$K_a = \frac{non-synsubstitution}{non-synsites} = \frac{1}{9⅔} \simeq 0.857 \qquad K_s = \frac{synsubstitution}{synsites} = \frac{2}{2⅓} \simeq 0.103 \qquad \frac{K_a}{K_s} = \frac{0.103}{0.857} = 0.12$$

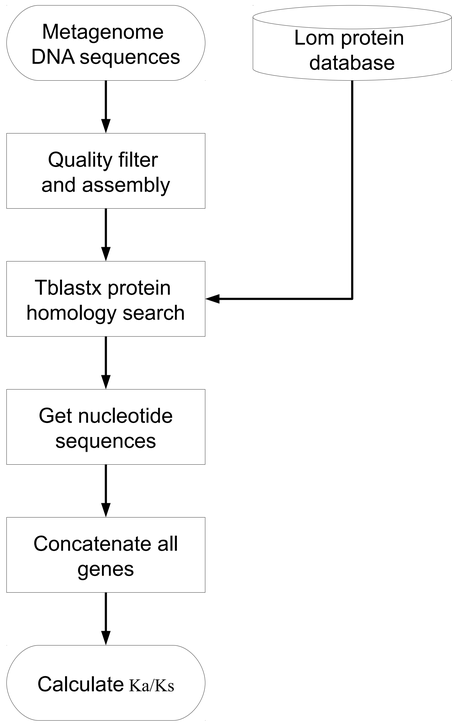
If Ka/Ks > 1 the gene is said to be under positive selection; new variants are advantageous and genetic diversity grows in the population. If Ka/Ks < 1 the gene is under negative selection and new variants are eliminated from the population. If Ka/Ks=0 the gene is under neutral selection. Two genes in the same genome can be under different kinds and strengths of selection.

Ka/Ks in a metagenome

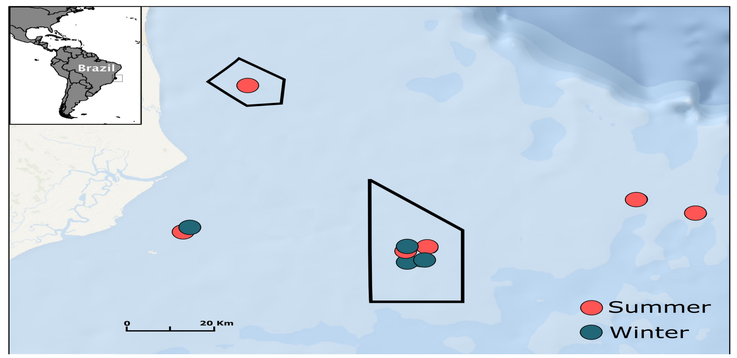
Normally, Ka/Ks is calculated from the sequences of two or more known genes. In this work we developed a tool to calculate the strength of selection acting on a protein family in a particular environment.

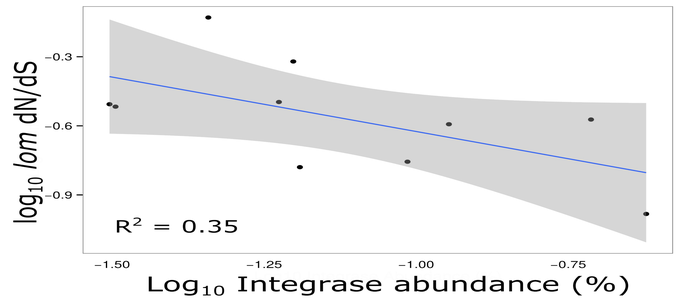
Our tool use Spades to assemble meta-genomes, the SEED database to get protein families, tblastx to do homology based searches, custom scripts to construct axt files from tblastx output and kaks-calculator to get the Ka/Ks ratios from those axt files. The metagenomes were provided by Forest Rohwer's lab. All code can be found at: https://github.com/Adrian-Cantu/meta-kaks .

Methods



Results





We investigated the viral assemblages in the Abrolhos Bank, the largest reef complex in the South Atlantic. One of the most abundant genes was lom, lambda outer membrane protein, which is known to confer pathogenicity to Enterobacteria via Lysogenic conversion. This well-known mechanism in human pathogens has recently been proposed in the coral pathogen Vibrio coralliitycus, where it could be the cause of disease outbreaks observed in reefs worldwide.

Conclusions

Our results show that we can detect genes that are under strong selective pressures in metagenomes. That may indicate how the environment acts on the genome.

References

Zhang Zhang et al., "KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging," Genomics, Proteomics & Bioinformatics 4, no. 4 (2006): 259–63, doi:10.1016/S1672-0229(07)60007-2.

Anton Bankevich et al., "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing," Journal of Computational Biology 19, no. 5 (May 2012): 455–77, doi:10.1089/cmb.2012.0021.

This research was supported by
NSF Grants (MCB-1330800, and DUE-132809)

CONACYT Consejo Nacional de Ciencia y Tecnología