

Supplementary material

Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding

Henrik Krehenwinkel^{1 2}, Madeline Wolf¹, Jun Ying Lim¹, Andrew J. Rominger¹, Warren B. Simison², Rosemary G. Gillespie¹

¹Department of Environmental Sciences, Policy and Management

University of California Berkeley

Mulford Hall, Berkeley, California, USA

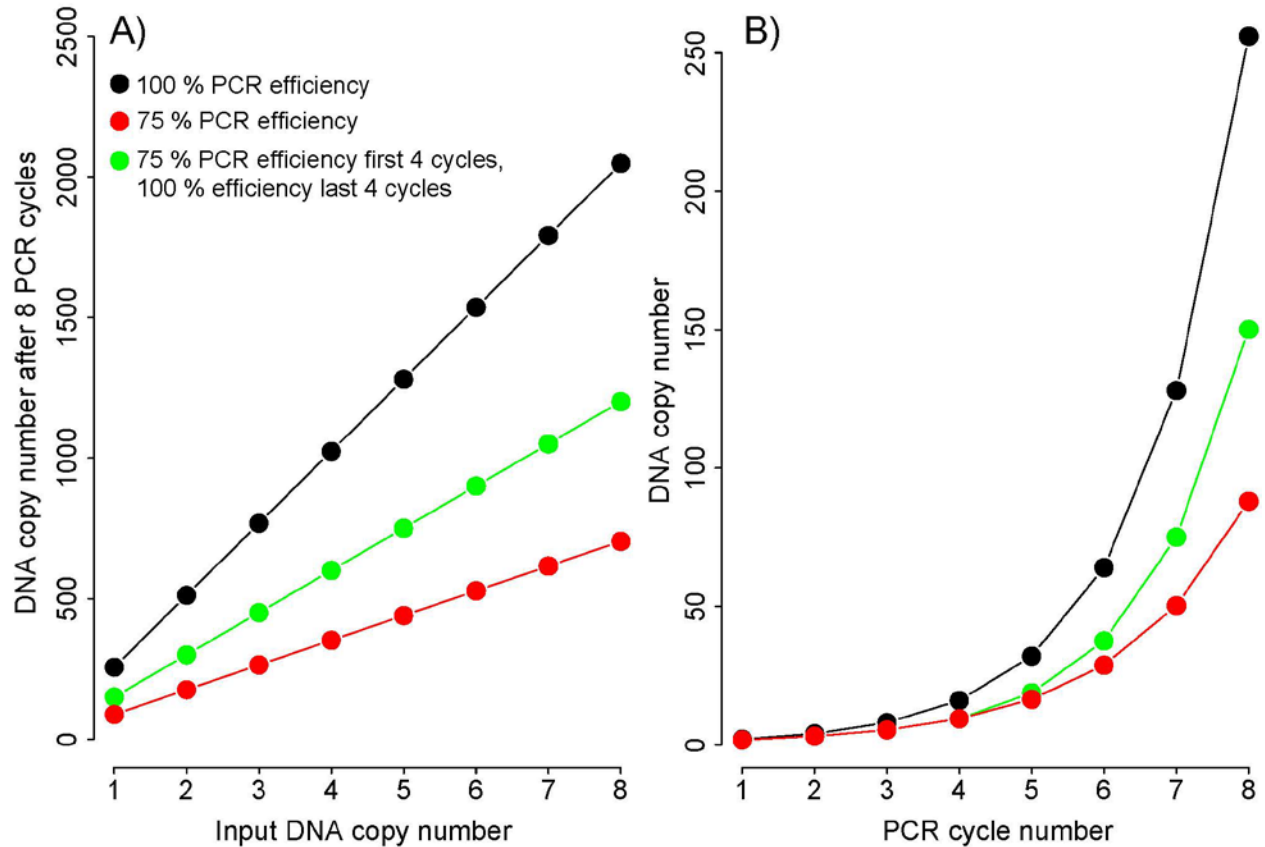
²Center for Comparative Genomics

California Academy of Sciences

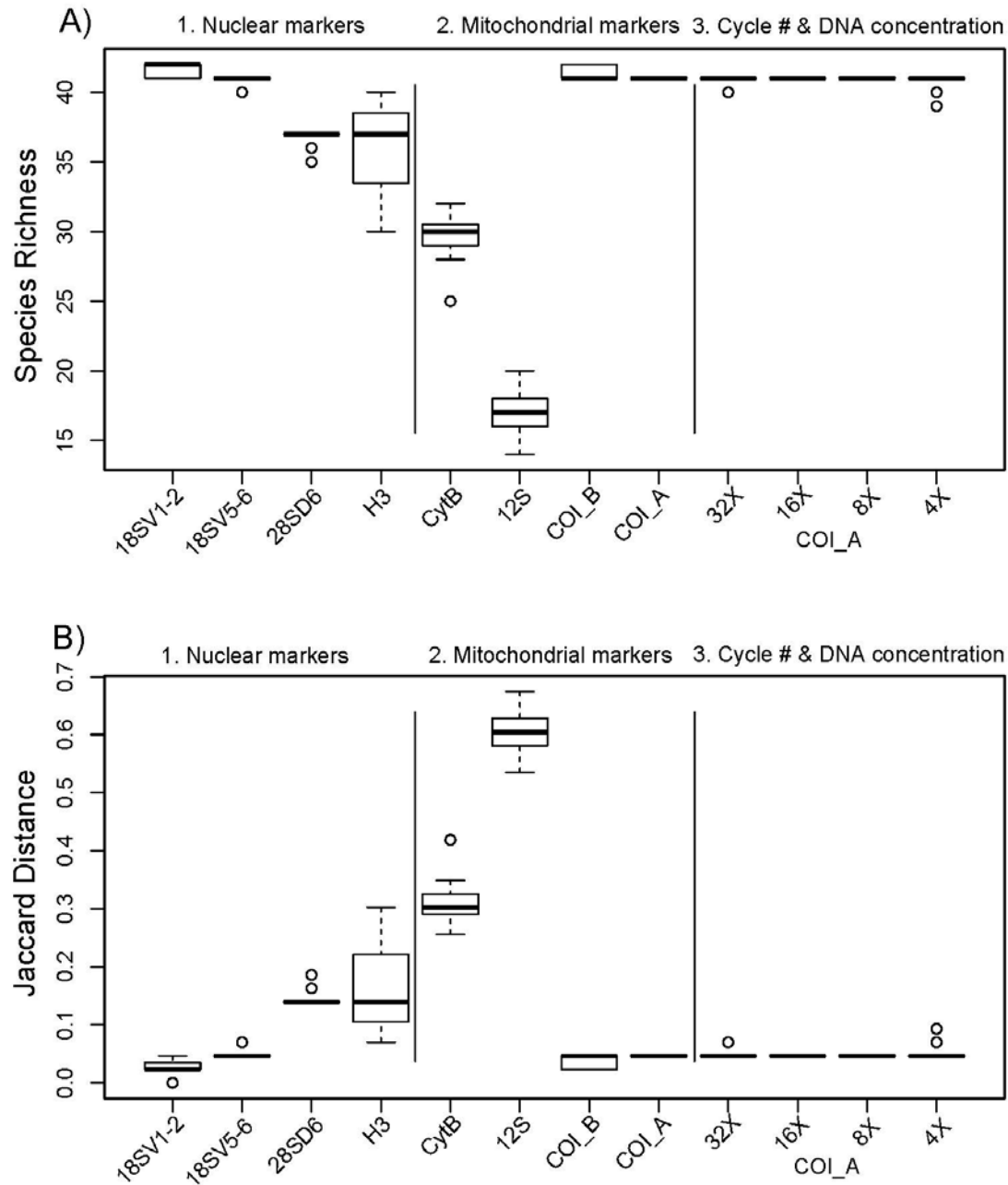
Music Concourse Drive, San Francisco, California, USA

Krehenwinkel@berkeley.edu

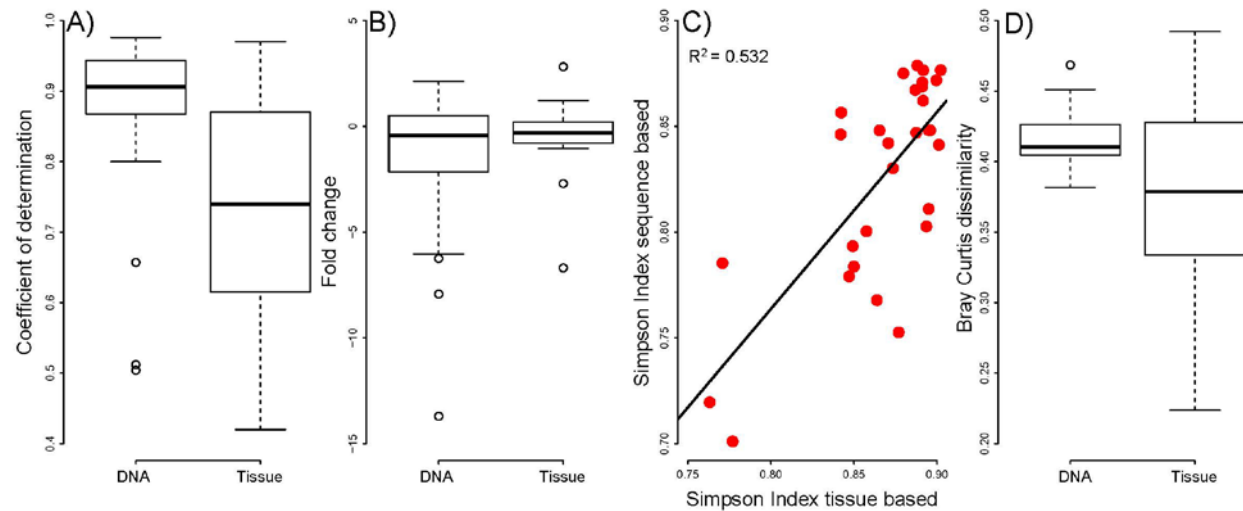
+1-510-646-3409



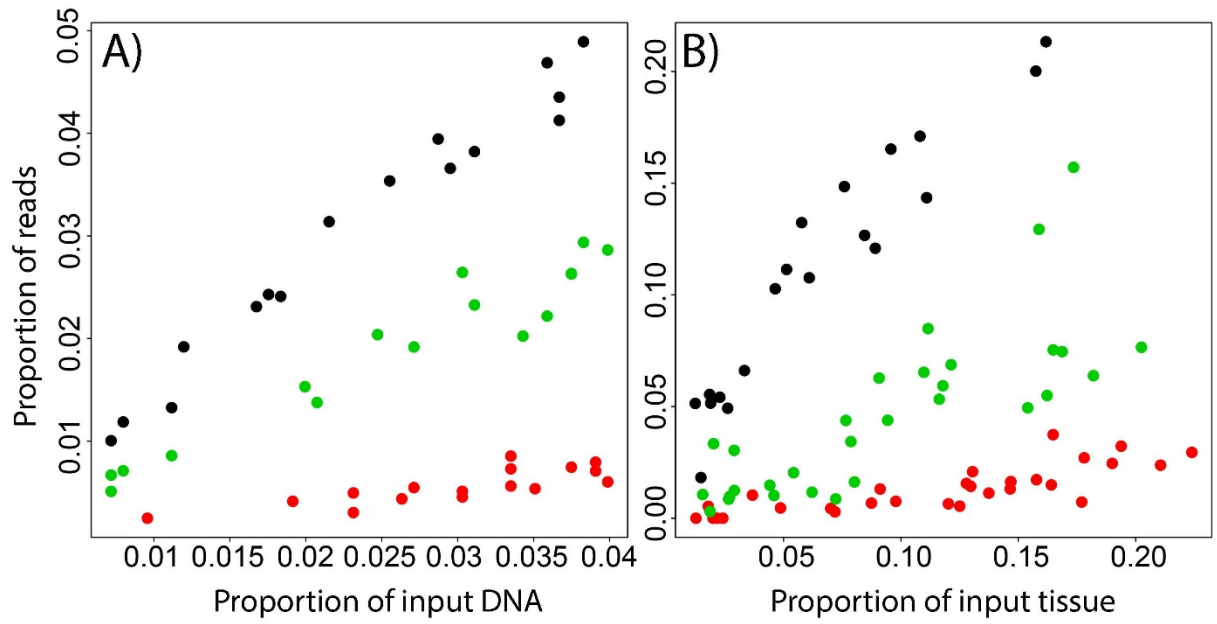
Supplementary Figure 1 A) Hypothetical association of input DNA copy number and recovered DNA copy number after 8 PCR cycles for two taxa with a difference in PCR efficiency of 25 % (red and black circles). By using a nested PCR on conserved priming sites after 4 cycles, the recovered copy number is corrected (green circles), leading to less biased abundance estimates. **B)** Association of PCR cycle number and DNA copy number for the same taxa. By removing PCR bias after four cycles, the difference in copy number between the two taxa is reduced.



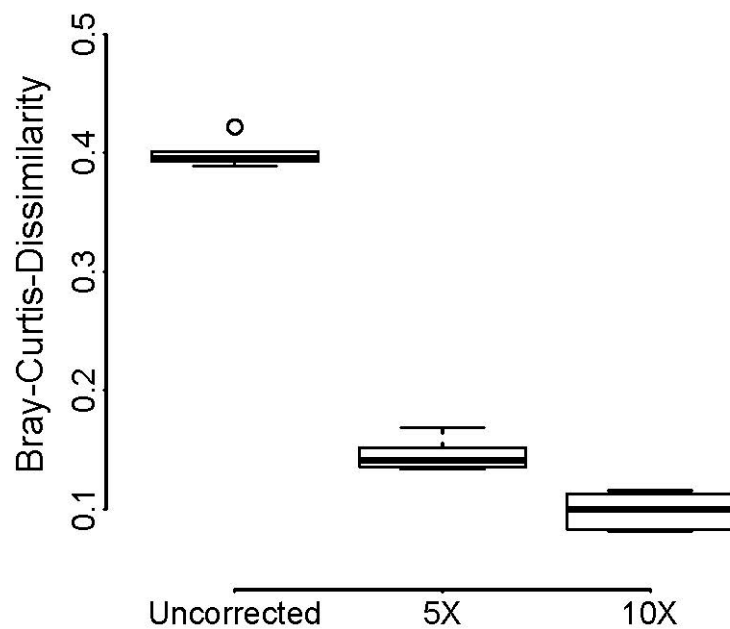
Supplementary Figure 2 A) Species richness for all mock communities of 43 taxa, based on DNA sequencing for **1.** nuclear and **2.** mitochondrial markers, and **3.** for mitochondrial COI at varying PCR cycle numbers and increased DNA template amount during PCR. **B)** Beta diversity (Jaccard distances) between actual morphospecies based mock communities and sequence based analyses. The boxplots present the same experimental conditions as described above.



Supplementary Figure 3 A) Coefficients of determination (R^2) of the linear association between actual abundance and read abundance for DNA based and tissue based mock communities. **B)** Fold change between actual abundance and read abundance for the same communities. **C)** Association of alpha diversity (Simpson indexes) based on actual abundances and sequence based abundance estimates for tissue mock communities. **D)** Beta diversity (Bray Curtis dissimilarity) between actual specimen based and sequence based communities and based on DNA or tissue mock communities.



Supplementary Figure 4 A) Association of the proportion of input DNA and recovered reads for three arthropod species (red: *Porcellio scaber*, green: *Proteroiulus fuscus*, black: *Tomocerus minor*) and **B)** for the same three species based on tissue pools.



Supplementary Figure 5 Bray Curtis dissimilarity between specimen based and sequencing based community samples for uncorrected read data, and using five or ten mock communities to derive corrections factors and correct read abundances.

Supplementary Table 1 Mean and standard deviation of fold change of input DNA and read abundance, as well as alpha diversity (Simpson index) and beta diversity (Bray Curtis dissimilarity) for PCR amplicon based and metagenomic samples. Fold change was calculated in relation to a hypothetical 1:1 association. The table shows all eight targeted loci. Amplicons were prepared with 32 PCR cycles following six cycles of indexing PCR. Metagenomic samples were prepared from untreated genomic DNA with only six cycles of indexing PCR.

Primer pair	FC _{amplicon}	FC _{gDNA}	α _{amplicon}	α _{gDNA}	B _{amplicon}	B _{gDNA}
ArF1/Fol-degen-rev	-1.073 ± 2.384	-0.454 ± 2.11	0.945 ± 0.008	0.937 ± 0.002	0.389 ± 0.019	0.347 ± 0.008
mlCOlintF/Fol-degen-rev	-1.336 ± 3.047		0.947 ± 0.007		0.417 ± 0.021	
CB3/CB4	-11.330 ± 22.497	0.312 ± 2.069	0.815 ± 0.050	0.910 ± 0.004	0.702 ± 0.028	0.420 ± 0.009
12sai/12sbi	-0.603 ± 6.180	0.691 ± 1.462	0.756 ± 0.047	0.928 ± 0.002	0.778 ± 0.023	0.369 ± 0.012
SSU_FO4/SSU_R22	-1.792 ± 4.965	-0.770 ± 2.491	0.931 ± 0.013	0.940 ± 0.001	0.435 ± 0.019	0.403 ± 0.005
18s_2F/18s_4R	-1.894 ± 6.694	-0.583 ± 2.637	0.933 ± 0.009	0.934 ± 0.002	0.412 ± 0.017	0.405 ± 0.007
28s_3F/28s_4R	-1.198 ± 3.857	-0.644 ± 2.221	0.922 ± 0.011	0.930 ± 0.002	0.470 ± 0.018	0.418 ± 0.005
H3aF/H3aR	-4.124 ± 9.968	0.680 ± 1.715	0.925 ± 0.006	0.932 ± 0.002	0.528 ± 0.031	0.389 ± 0.013