SUPPLEMENTARY MATERIALS 1

**Gene Enrichment Method**

The gene enrichment filter power calculation for *E. guineensis* and *E. oleifera* as described in Bedell et al., (2005) is:

pmf

g

\_\_\_\_\_\_\_

pμf

g

pmf is the probability that a methylation filtered (MF) read sample gene coding sequence

g

pμf is the probability that an unfiltered (UF) read sample gene coding sequence

g

These probabilities are estimated by looking at the proportions of MF and UF reads that hit genes

^

pmf = Nmf

g

g

\_\_\_\_\_\_\_\_\_\_\_

Nmf

^

pμf = Nμf

g

g

\_\_\_\_\_\_\_\_\_\_\_

Nμf

It was observed that out of Nmf methylation filtered reads, Nmf of them hit a gene and for Nμf unfiltered reads, Nμf of them hit genes. The 95% confidence intervals for these binomial proportions are calculated using the standard approximate formula, *viz*.:

g

g

^

^

^

pμf ± 1.96 pμf (1 - pμf )

g

g

g



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nμf

For the palm data analysis, we used the mixed UF and MF assembly to sample in depth the filtered and unfiltered reads.

Nmf = number of filtered reads in contigs (excluding singletons)

Nμf = number of unfiltered reads in contigs (excluding singletons)

Nmf = number of filtered reads in (repeat-masked) contigs with a hit to a gene

g

Nμf = number of filtered reads in (repeat-masked) contigs with a hit to a gene

g