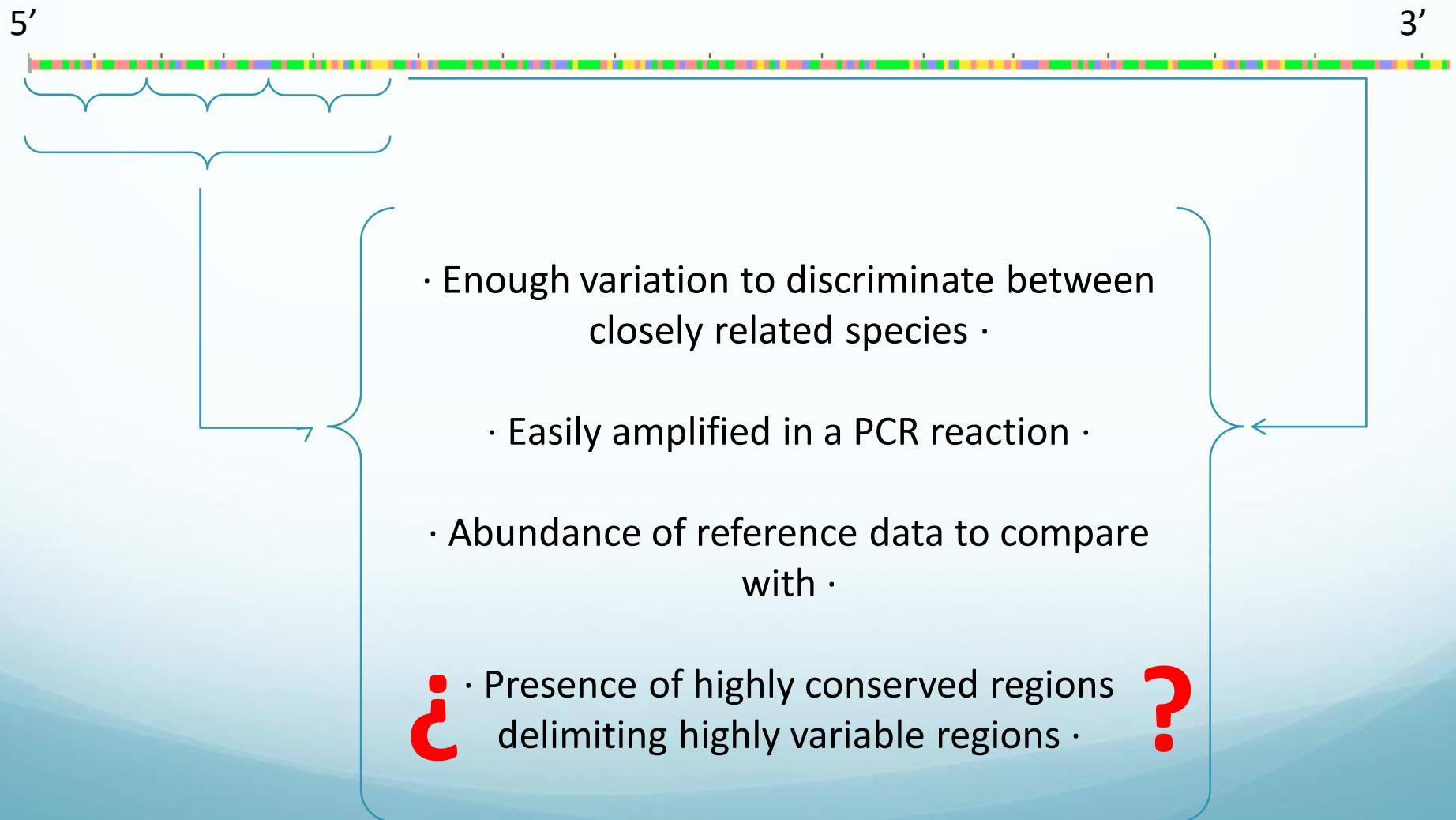


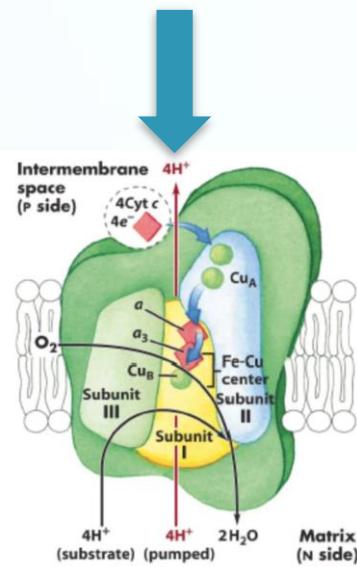
Problems with amplicon-based approaches (and how to overcome them).

“Barcodeing ist kein ponyhof”

Is COI metabarcoding really a ponyfarm?



Is COI metabarcoding really a ponyfarm?



Protein-coding gene:

- Function
 - Structure
 - Location
- Selective pressure upon mutation in certain regions (in the $\alpha\beta$ sequence).

Redundancy in the genetic code.

→ The third position in the codon is irrelevant for the translation in many cases.

	Second Letter				
	T	C	A	G	
T	TTT TTC TTA TTG } Phe	TCT TCC TCA TCG } Ser	TAT TAC TAA TAG } Tyr Stop Stop	TGT TGC TGA TGG } Cys Stop Trp	TCA AGG } T GAG } C
C	CTT CTC CTA CTG } Leu	CCT CCC CCA CCG } Pro	CAT CAC CAA CAG } His Gln	CGT CGC CGA CGG } Arg	TCA AGG } T GAG } C
A	ATT ATC ATA ATG } Ile Met	ACT ACC ACA ACG } Thr	AAT AAC AAA AAG } Asn Lys	AGT AGC AGA AGG } Ser Arg	TCA AGG } T GAG } C
G	GTT GTC GTA GTG } Val	GCT GCC GCA GCG } Ala	GAT GAC GAA GAG } Asp Glu	GGT GGC GGA GGG } Gly	TCA AGG } T GAG } C

Even in the $\alpha\beta$ -conserved regions, mutations in the 3rd position will be much less constrained.

HYPERVARIABILITY

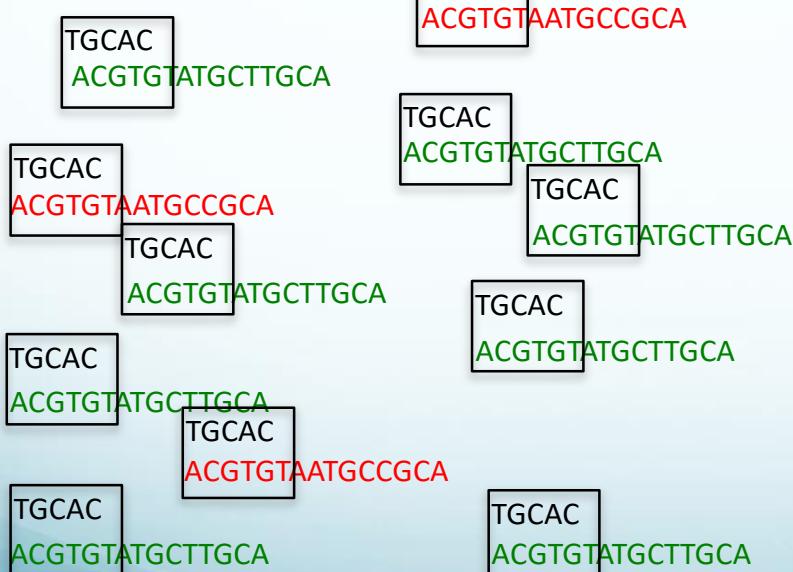
Is COI metabarcoding really a ponyfarm?

What does **HYPERVARIABILITY** imply?

- 1) Amplification of non-target groups (e.g spider's diet).

Primer: TGCAC; Spdier: Fly

Sample



Amplification

ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACGTGTAAATGCCGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACGTGTAAATGCCGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACGTGTAAATGCCGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA
ACCTGTATGCTTGCA

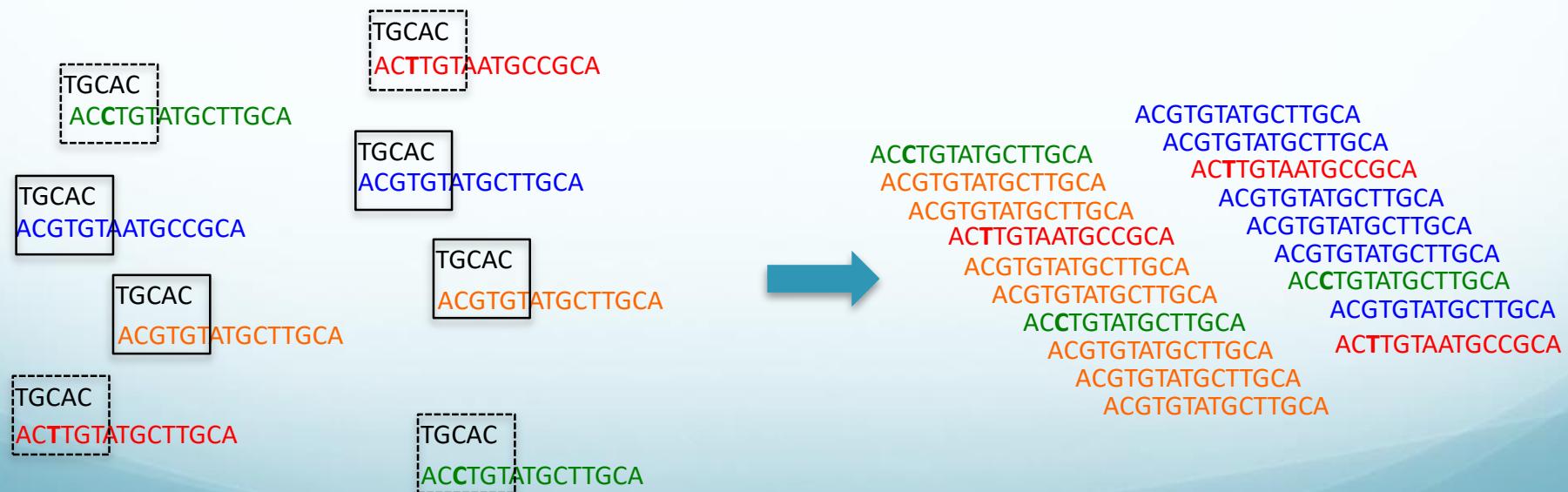
Sequencing and analysis

Is COI metabarcoding really a ponyfarm?

What does **HYPERVARIABILITY** imply?

- 1) Amplification of non-target groups (e.g spider's diet).
 - 2) Amplification bias within the target group

Primer: TGCAC; Diptera; Hymenoptera; Coleoptera; Lepidoptera



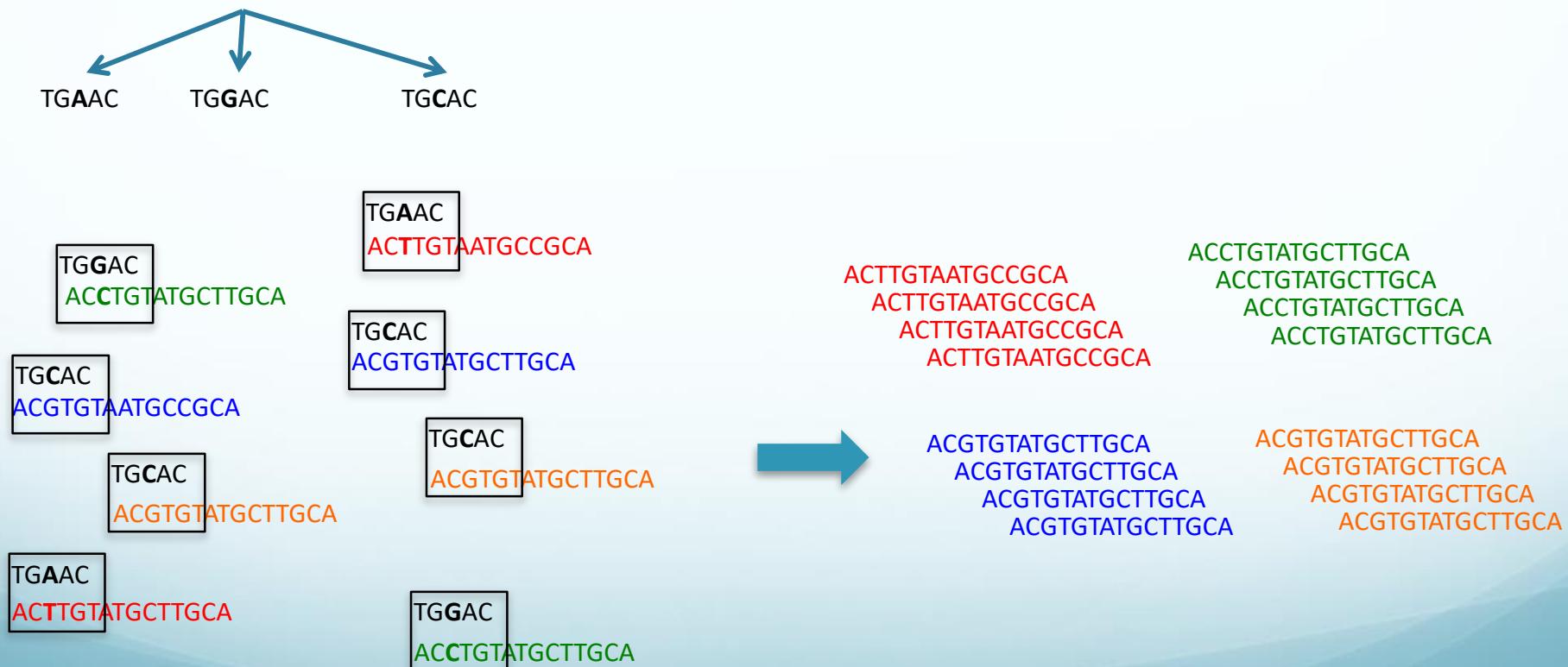


Won't somebody
please think of the
children?!

Is there any solution?

Solution 1: DEGENERATE PRIMERS

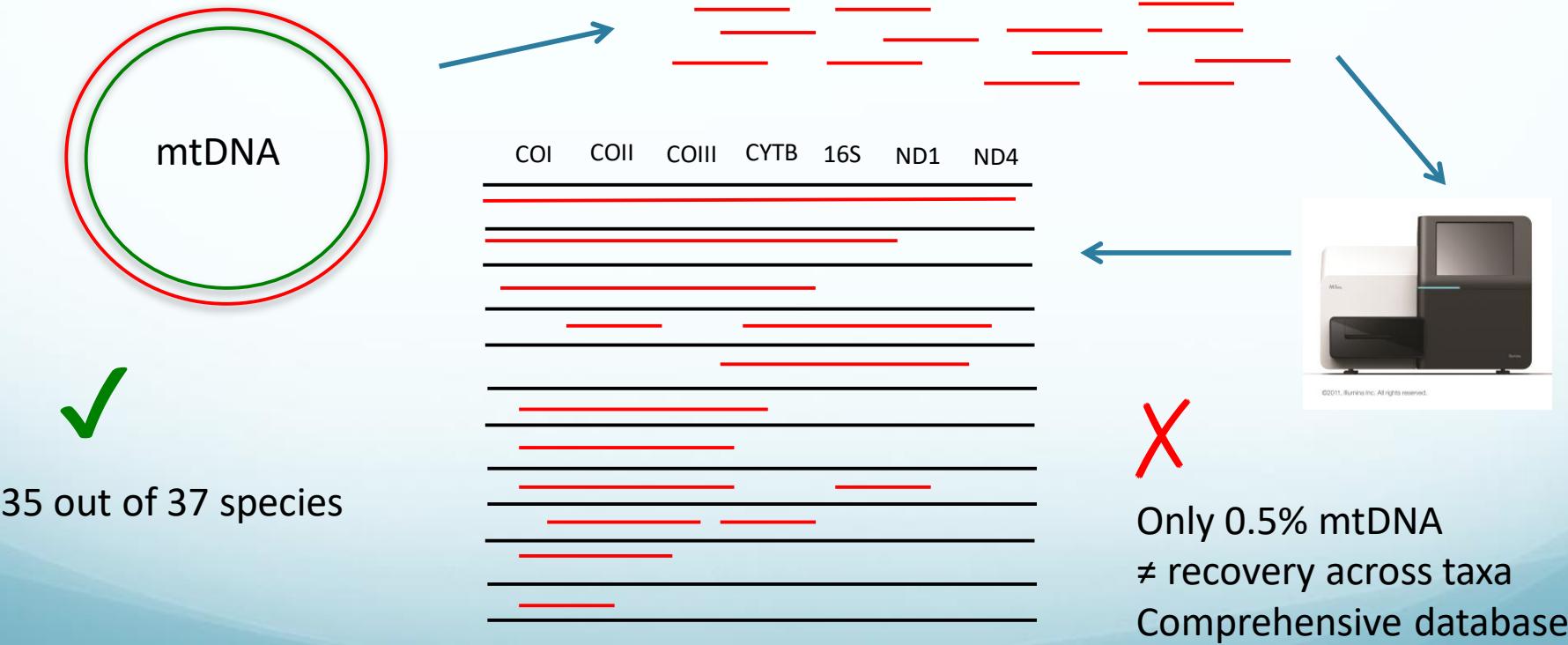
Primer: TGVAC; **Diptera**; **Hymenoptera**; **Coleoptera**; **Lepidoptera**



Is there any solution?

Solution 2: PCR-FREE PROTOCOLS

2.1 Shotgun sequencing

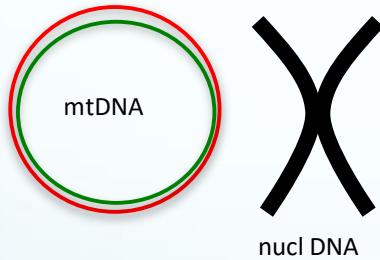


Is there any solution?

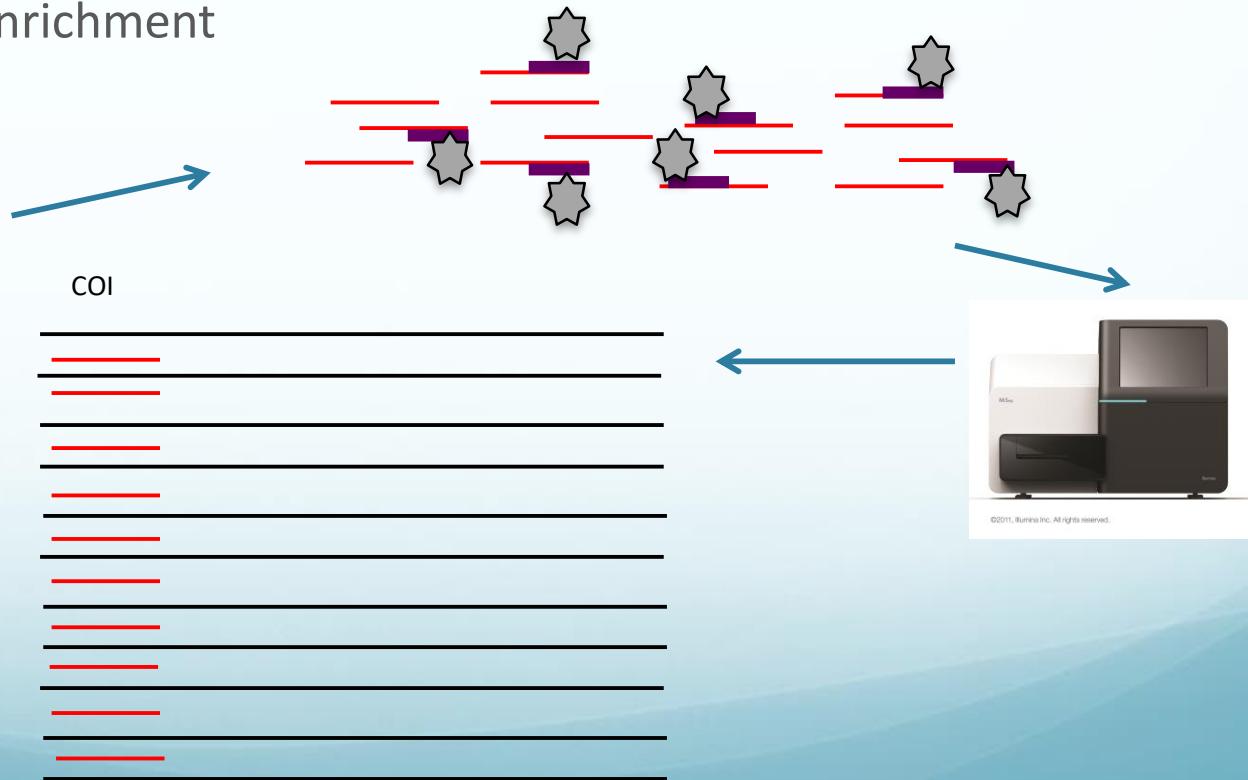
Solution 2: PCR-FREE PROTOCOLS

2.1 Shotgun sequencing

2.2 Gene enrichment



No bias
Good species recovery



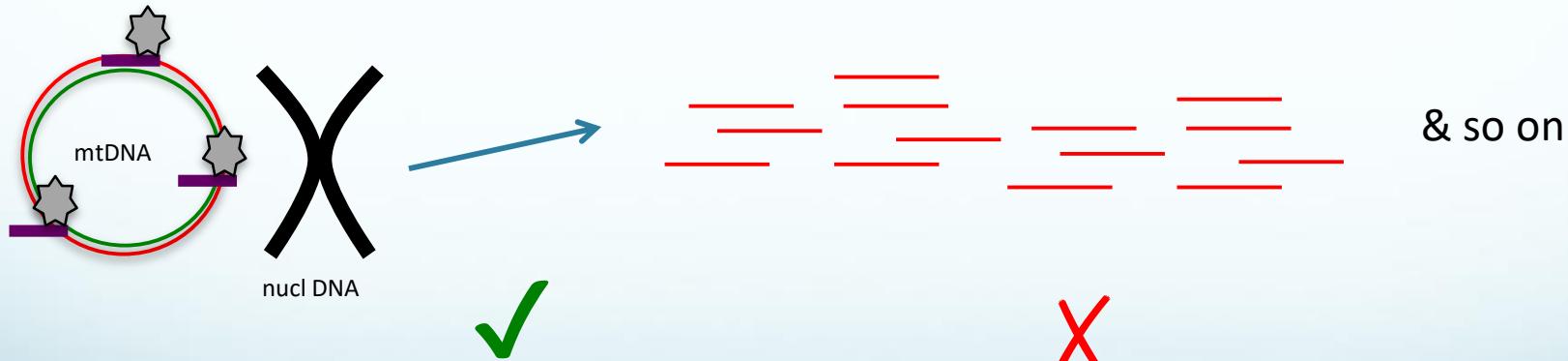
Is there any solution?

Solution 2: PCR-FREE PROTOCOLS

2.1 Shotgun sequencing

2.2 Gene enrichment

2.3 Mitochondrial enrichment



Huge coverage of the mt.
genome (still 5%)
Good species recovery

Expensive
Comprehensive database

Is there any solution?

Solution 3: USE OTHER MARKERS

- Broad taxonomic scope: **18S** (soil, meiofauna).
- Vertebrates: **12S**.
- Insects, some orders: **CytB**.
- Insects, all orders: **16S**. As good resolution as COI from 200bp, consistent amplification through 23 orders.

Building local databases helps a lot

That's all Folks!