DNA Barcodes Provide a Quick Preview of Mitochondrial Genome Composition

Xiang Jia Min and Donal A. Hickey
Department of Biology
Concordia University
Montreal, Quebec

The primary function of DNA barcodes is species-level identification.

The power of DNA barcoding lies in the fact that a very short fragment of genomic sequence provides a useful sample of the whole genome sequence

A potential secondary function of DNA barcodes is that they might also provide a useful sampling of differences in genome composition.

In this study, we have analyzed the correlation between the nucleotide content of the short DNA barcode sequences and the genomes from which they are derived.

Our results show that such short sequences can yield important, and surprisingly accurate, information about the composition of the entire genome.

In other words, the DNA barcodes can provide a quick preview of the whole genome composition.

Questions?

- **Can the nucleotide composition of a DNA barcode reflect the overall nucleotide composition (i.e., GC content) of the parent genome?**

- **Can the nucleotide skews of a DNA barcode reflect the overall nucleotide skews of the parent genome?**

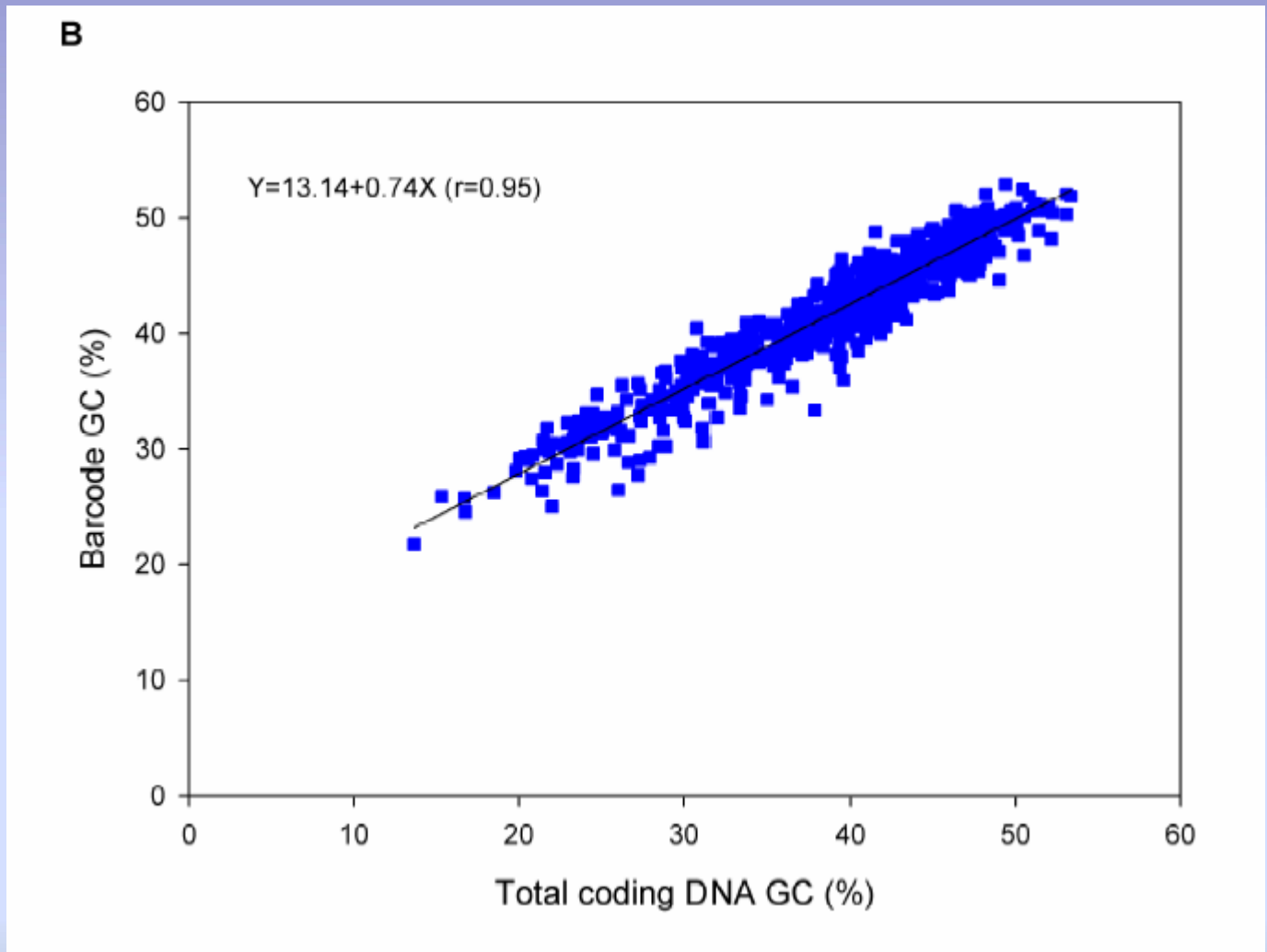
Data and Methods

- **A total of 849 complete mitochondrial genomes (mtDNA) in metazoa were collected.**
- **Standard barcodes of ~648 nt from cytochrome c oxidase subunit 1 (COI) were retrieved from each genome.**
- **Nucleotide frequencies of mtDNA, total coding DNA, and DNA barcode sequences were analyzed.**
- **The GC- and AT- skews were measured according to the following formulae:**

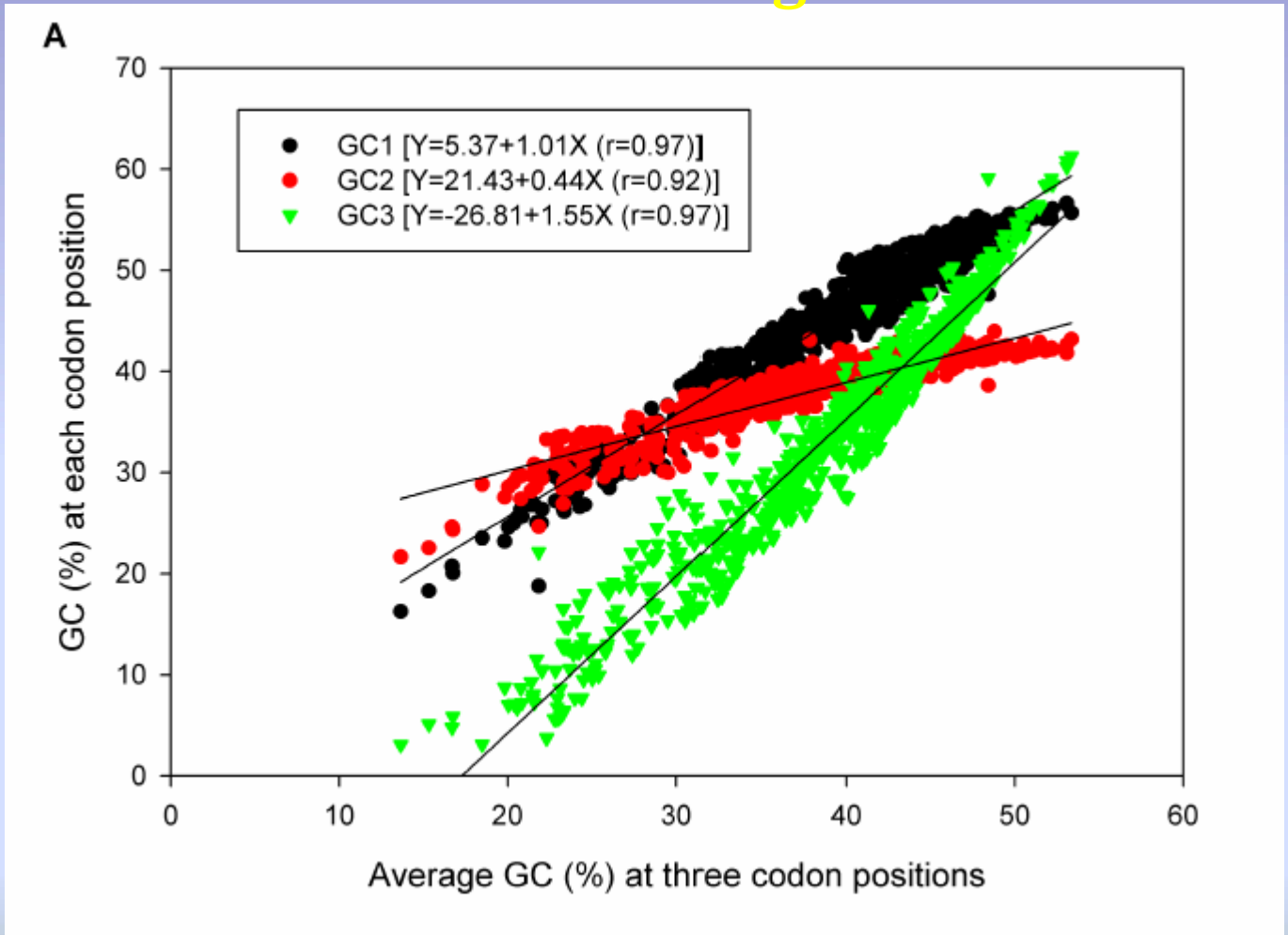
$$\text{GC-skew} = (G-C)/(G+C)$$

$$\text{AT-skew} = (A-T)/(A+T)$$

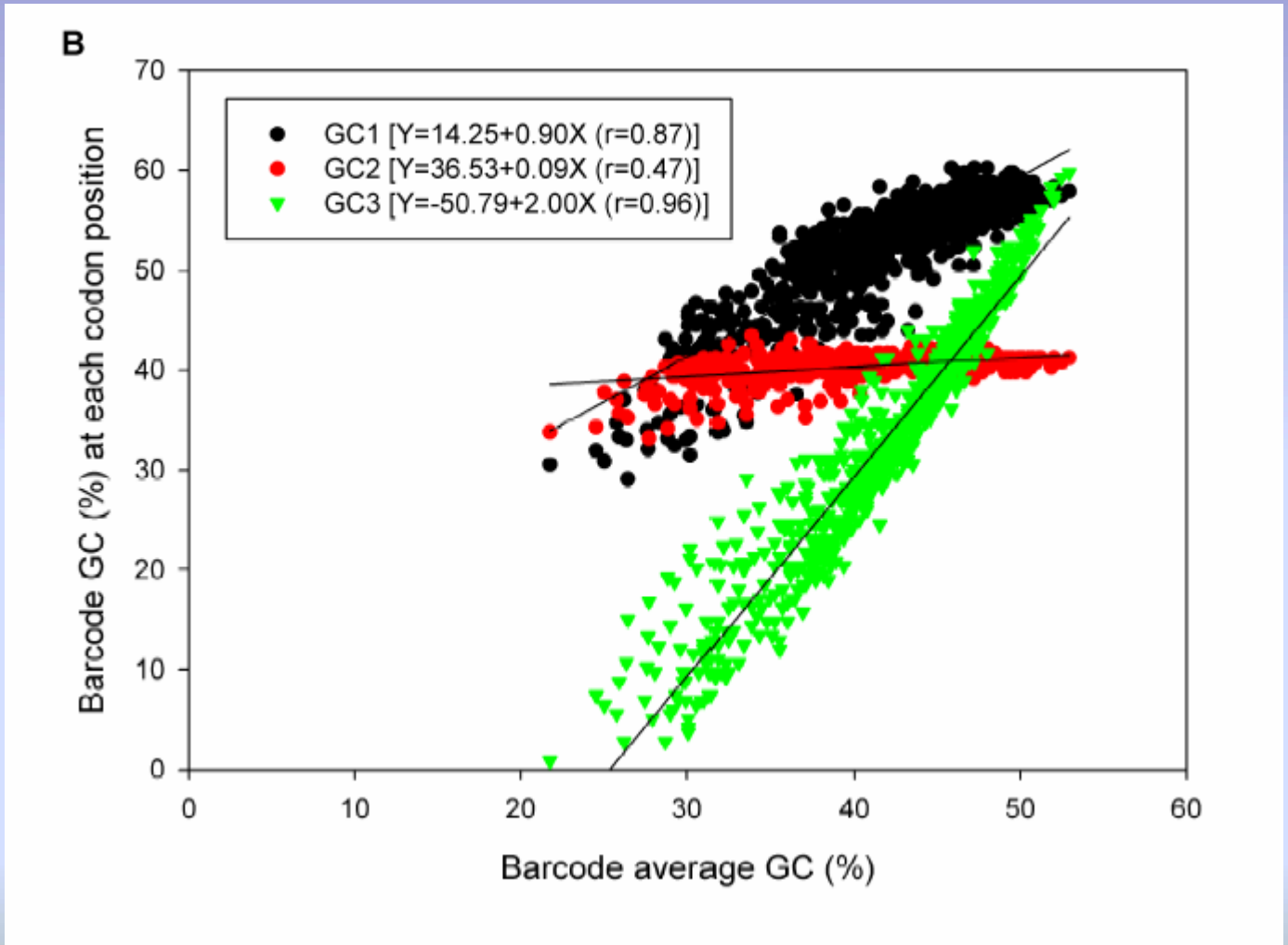
Correlation between nucleotide content of DNA barcodes and the nucleotide content of the total mitochondrial coding sequences.



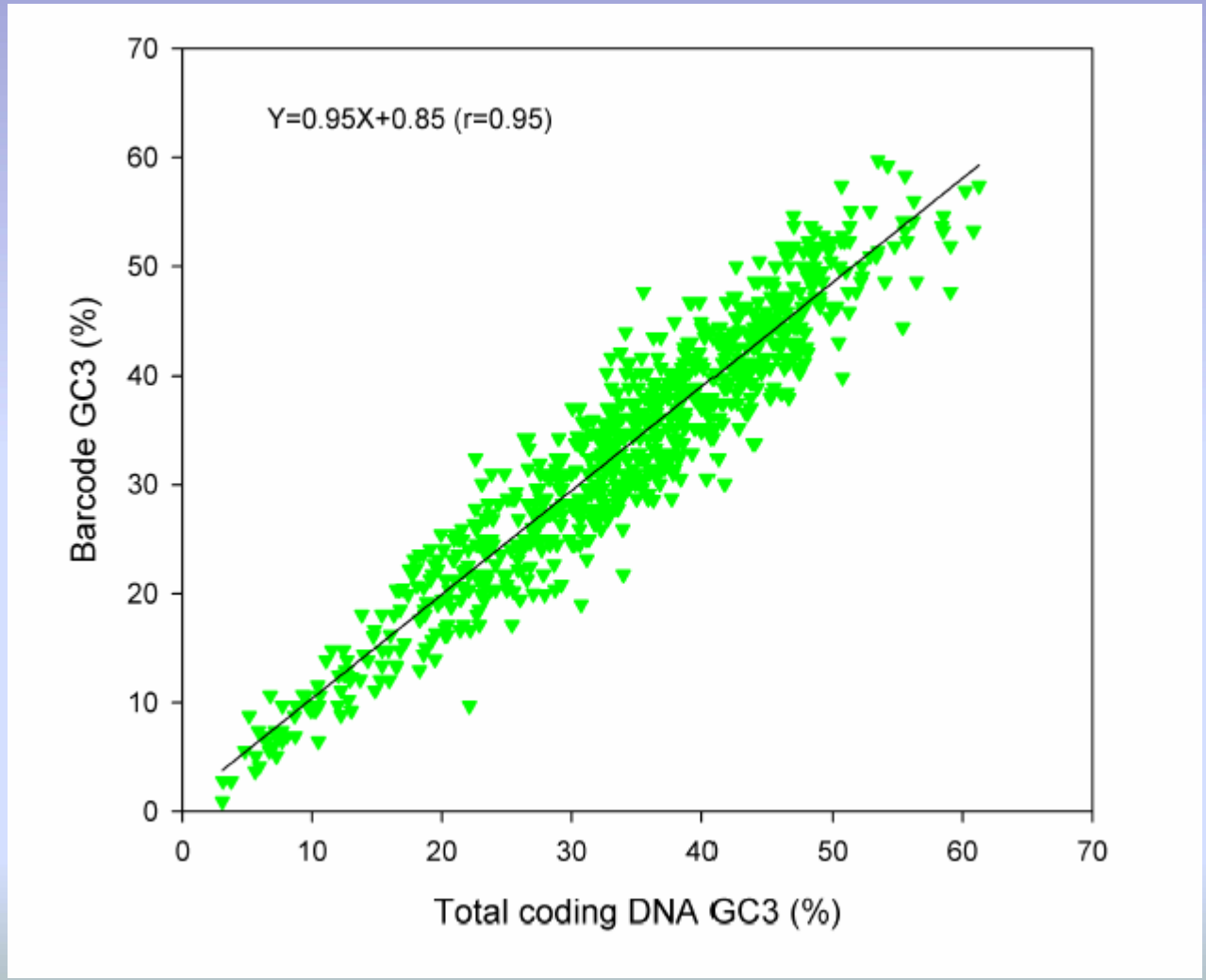
Correlation between GC content at each codon position and the average GC content for all protein coding sequences in the mitochondrial genome.



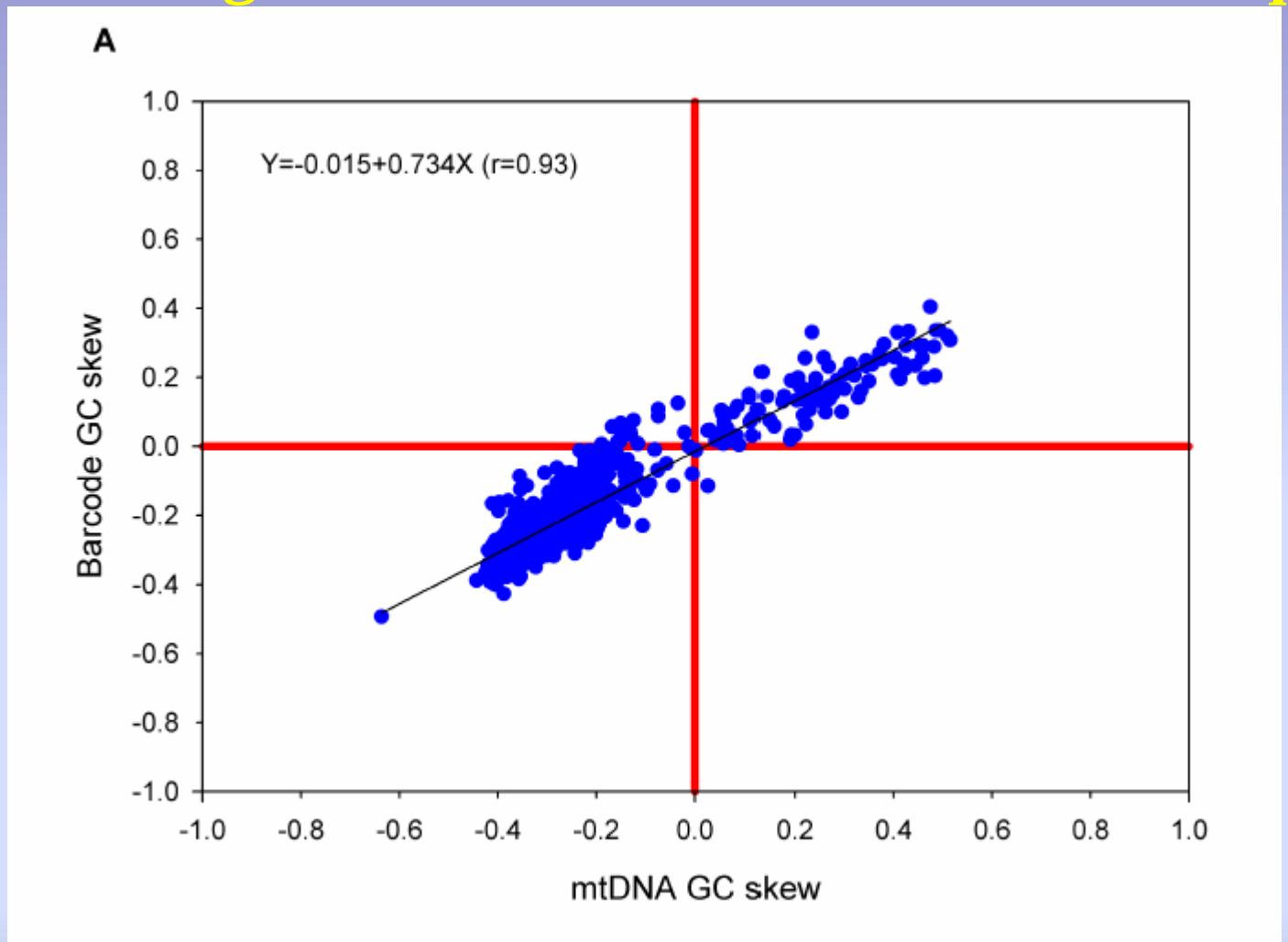
Correlation between GC content at each codon position and the average GC content for DNA barcode sequences



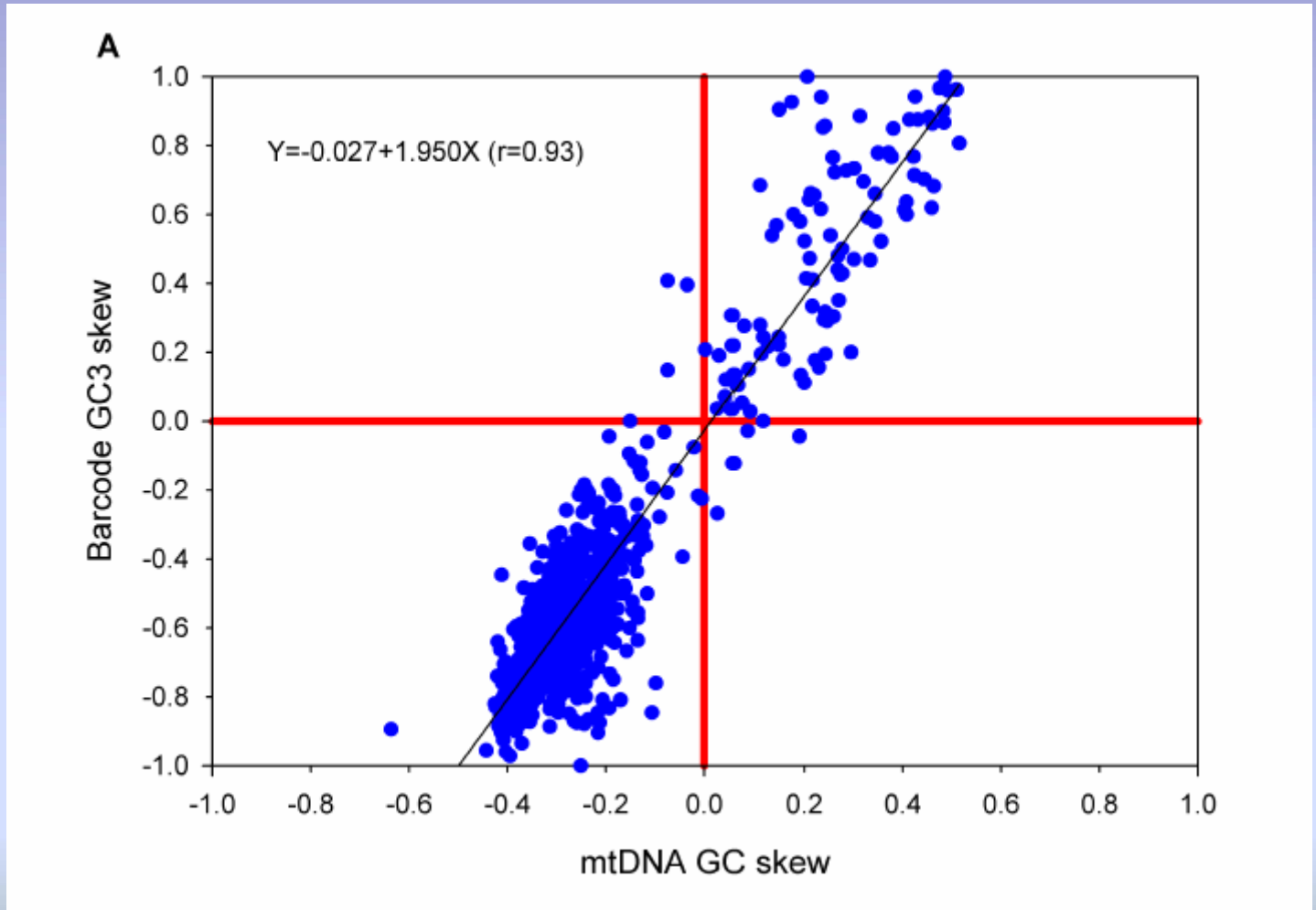
Correlation between nucleotide content of DNA barcodes and the nucleotide content of the total mitochondrial coding sequences (3rd codon position only).



Correlation between nucleotide skews in the entire mitochondrial genome and in the DNA barcode sequences.



Correlation between nucleotide skews in the entire mitochondrial genome and in the DNA barcode sequences (3rd codon position only).



Conclusions

1. The nucleotide composition of DNA barcode sequences is highly correlated with nucleotide content of the whole mitochondrial genomes.

2. The variation of nucleotide composition at each codon position in barcodes also reflects the variation of nucleotide composition of total coding DNA sequences of the entire genomes.

3. The nucleotide composition skews (GC- and AT-skews) of DNA barcodes are highly correlated with nucleotide skews of the entire mitochondrial genomes.

Therefore DNA barcodes can not only enable us to assign unknown specimens to known species, but it can also help us to get a quick snapshot of genome composition.

The information on nucleotide composition and nucleotide skews can, in turn, provide practical information for designing primers.

Future work: could this approach be used for nuclear genomes and nuclear DNA barcodes?

Acknowledgments

This research was supported through funding to the Canadian Barcode of Life Network from Genome Canada through the Ontario Genomics Institute, NSERC and other sponsors listed at www.BOLNET.ca.