

In-silico Assessment of five chloroplast intergenic regions in *Poaceae* family for DNA barcoding

Mohamed Awad¹, Mohamed Helmy^{2,3}, Kareem A. Mosa^{1,4} and Fawzy A. El-feky¹.

1. Department of Biotechnology, Faculty of Agriculture, Al-Azhar University, Cairo, Egypt

2. The Donnelly Centre for Cellular and Biomedical Research, University of Toronto, Toronto, ON, Canada

3. Head of Bioinformatics, BenchSci.com, Toronto, ON, Canada

3. Department of Applied Biology, College of Science, University of Sharjah, United Arab Emirates

Introduction:

- DNA Barcode, in its simplest sense, is one or more short gene sequences taken from standardized region of the genome used for a sustainable species identification through reference sequences from DNA sequence libraries and databases.
- The Consortium for the Barcode of Life (CBOL) was keen to form a plant working-group to propose standard DNA Barcode loci. The plant working group recommended using a combination of two plastid loci, *matK* + *rbcl*, as the standard plant DNA barcodes.
- In spite of the initial CBOL agreement to rely on *matK* and *rbcl* as standard DNA barcodes, a problem appeared in their ability to discriminate between some plant species.
- After identified the DNA barcode as a key method in the UN strategic plan for biodiversity in Cancun 2016, the scientific community become in urgent need to identify and evaluate a standard DNA barcode loci in plants that could be utilized for a wide range of plant species.
- In our recent report, through a chloroplast Genome-wide *in-silico* analysis for *Triticum* genus we proposed combination of (*trnfM*^{CAU} – *trnT*^{GGU}) with either (*trnD-psbM*), (*petN-trnC*), (*matK-rps16*) or (*rbcl-psaI*) as an effective combination to be used as DNA barcodes for the *Triticum* species (Awad et al., 2017).
- Here we aim to assess the five intergenic regions proposed in our previous study as DNA Barcodes for plants of the family *Poaceae*.

Results:

- We investigated five intergenic regions (*trnfM-trnT*, *trnD-psbM*, *petN-trnC*, *matK-rps16* and *rbcl-psaI*) in the available 147 chloroplast genome in *poaceae* family.
- Our designed primers can extract sequences ranged between:
 - 400 and 1000 bp for *petN-trnC* primer.
 - 800 and 2145 bp for *rbcl-psaI* primer.
 - 1230 and 1515 bp for *matK-rps16* primer.
 - 575 and 1250 bp for *trnD-psbM* Primer.
 - 950 and 2650 bp for *trnfM-trnT* primer.
- Our results showed outstanding species discrimination power for the five regions, constituting 94.55, 93.20, 89.10, 87.07 and 86.40 % for *trnfM-trnT*, *rbcl-psaI*, *matK-rps16*, *petN-trnC* and *trnD-psbM*, respectively.
- Ten combination between the intragenic regions showed a very high capacity for discrimination ranged between 93.20 % in *matK-rps16/trnD-psbM* and 98% in *trnfM-trnT/petN-trnC*.
- From 147 species and sub species a combination of three regions (*trnfM-trnT* + *petN-trnC* + *rbcl-psaI*) can discriminate all species except *Phyllostachys nigra* var. *henonis* and *Phyllostachys sulphurea*

Discrimination ability using single intragenic region locus

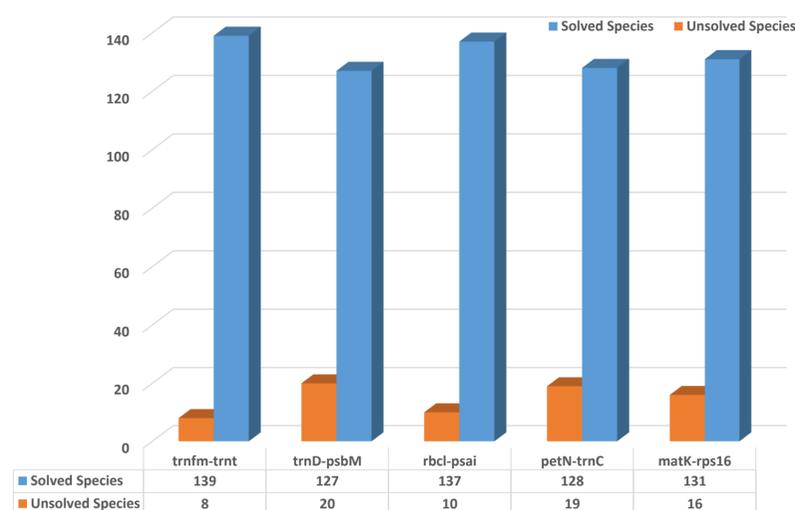


Figure1: Species discrimination effectiveness using single intragenic regions. Orange: number of un-discriminated species (Unsolved species). Blue: number of discriminated species (solved species). Intragenic regions *trnfM-trnT* and *rbcl-psaI* are the best regions for species discrimination.

Discrimination ability using combination of intragenic region locus

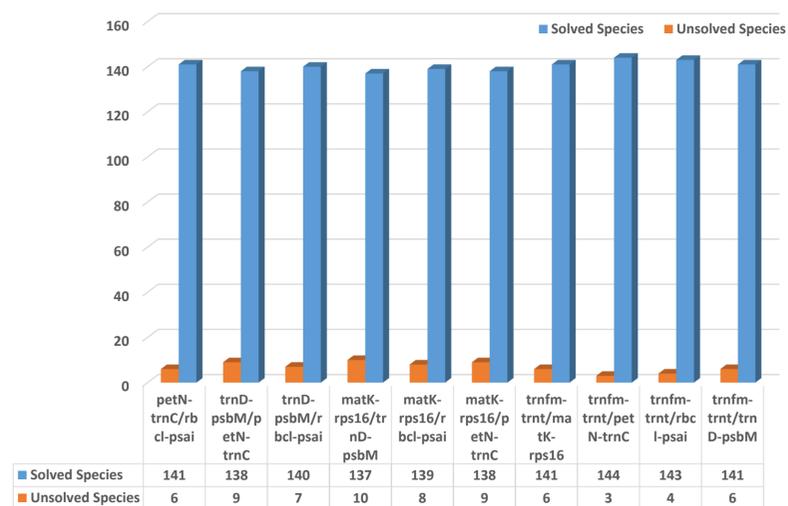


Figure2: Species discrimination effectiveness using combination of Intragenic regions. All intragenic can improve the discrimination ability when making combination between them. Orange: number of un-discriminated species (Unsolved species). Blue: number of discriminated species (solved species). *trnfM-trnT/petN-trnC* and *trnfM-trnT/rbcl-psaI* are the best combinations.

Work Flow

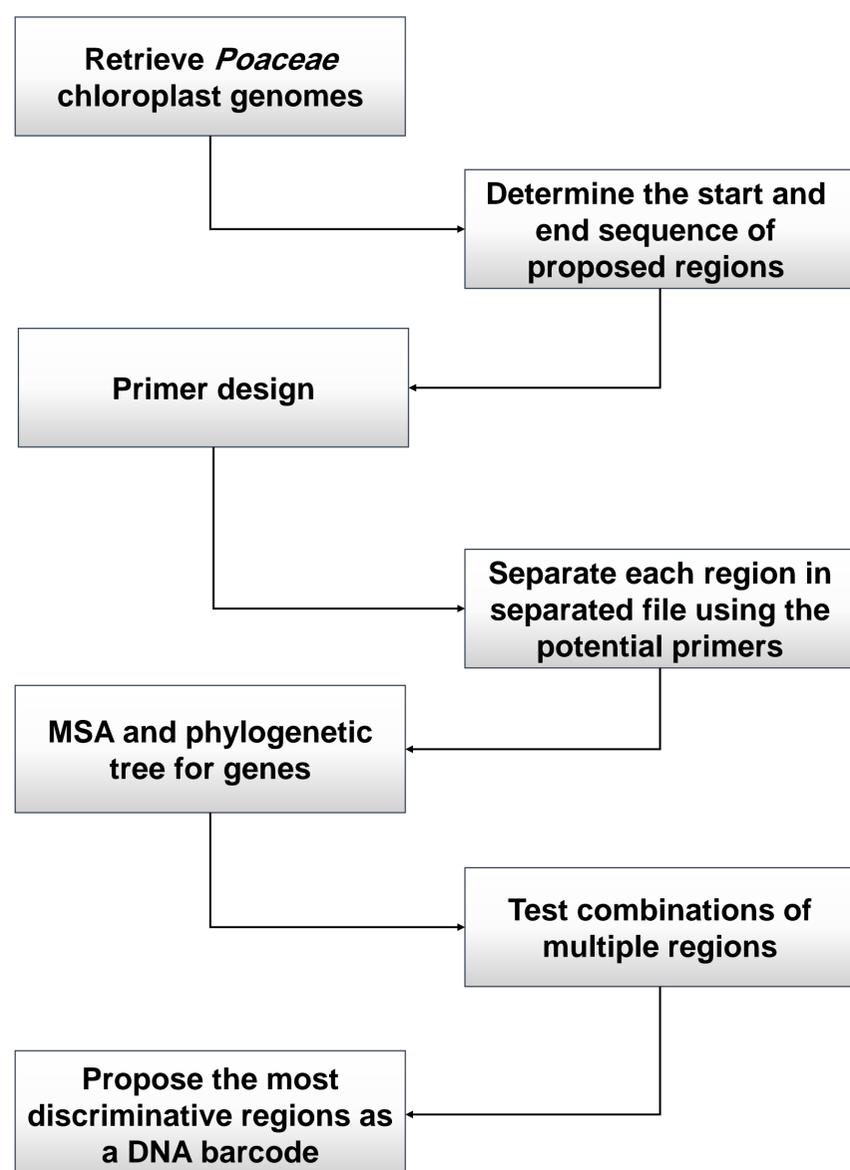


Table 1. List of designed primers

Primers	Frward	Rrverse
<i>matK-rps16</i>	AWTTTTKYSAWTTDYAYT	CTCGRAGAGARGGRAGAGAT
<i>petN-trnC</i>	GCGAGACTTAYTATATCCAT	TGCAAATCYTTTAYCCCCAG
<i>rbcl-psaI</i>	GGTAATGAAATTATC	TAAAGAAATAAGGAA
<i>trnD-psbM</i>	GCACCGCCCTGTCAAGGCGGAA	GTTTTTACATAAATAATAAG
<i>trnfM-trnT</i>	TTCATTCGGCTCCTTTATG	TACCATGGCATTACTCTRCC

References:

- Mohamed Awad, Ragab M. Fahmy, Kareem A. Mosa, Mohamed Helmy, Fawzy A. El-Feky, Identification of Effective DNA Barcodes for Triticum Plants through Chloroplast Genome-wide Analysis (2017), Computational Biology and Chemistry.

Conclusion:

- We recommend the intragenic combination of (*trnfM-trnT/petN-trnC*) and (*trnfM-trnT/rbcl-psaI*) as an effective combination to be used as DNA barcodes for the family *Poaceae*.
- Further investigations are required to assess the effectiveness of these barcodes at the lab and the effectiveness of same DNA barcodes in other plant families.

Acknowledgment:

MA would like to acknowledge the University of Johannesburg for the "Travel Award for African students", which enabled him to attend this conference.

