Characterization of a potential pseudogene in *Leucaena trichandra*

Ed Barge

INTRODUCTION

Pseudogenes are regions of DNA that are related to functional genes but are themselves noncoding. However, some still play important functional roles (Vanin 1985).

Processes which can lead to the creation of pseudogenes:

1) Whole-genome or tandem duplication: Changes can occur during or after duplication (mutations, insertions, deletions, frame shifts) which result in loss of gene function at the transcription or translation level (Lynch & Conery 2000).

2) Retrotransposition: Reverse transcription of mRNA back into the genome can lead to the formation of pseudogenes (Jurka 2004).

OBJECTIVES

1) Characterize potential pseudogene in *Leucaena trichandra*.

2) Identify process leading to its formation (duplication or retrotransposition).

RESULTS

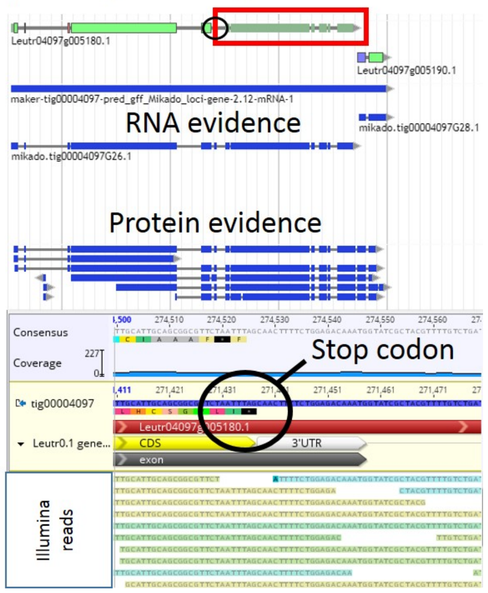
Potential pseudogene:
Leutr04097g005180

Protein (BLAST2GO):
Auxin transport BIG

Function (uniprot):
Required for auxin efflux and polar auxin transport

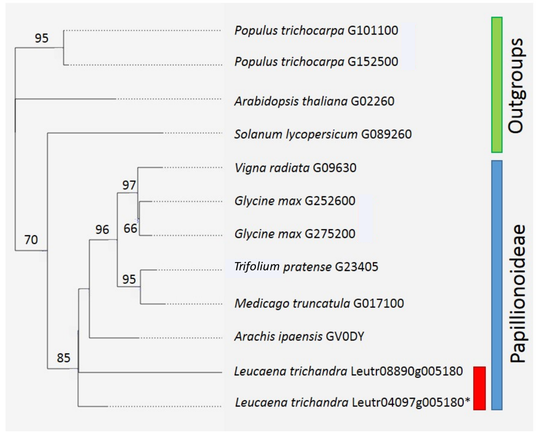
Evidence: RNA and protein evidence support presence of gene. However, there is a premature stop codon (black circle) at the beginning of the putatively untranslated portion of the gene (red rectangle).

Gene 5180 containing potential psuedogenized region



RESULTS CONT'D

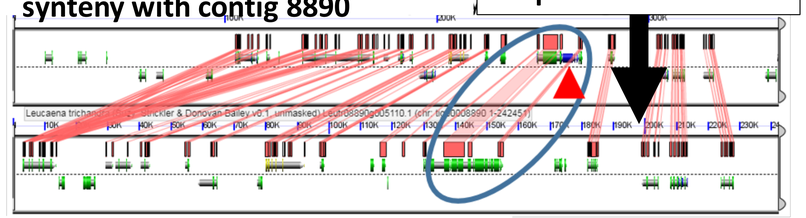
It appears the gene of interested has been duplicated independently several times based on the phylogeny generated in PLAZA showing the position of the putative pseudogene and its in-paralog (red bar) in relation to other legumes and 3 outgroup taxa.



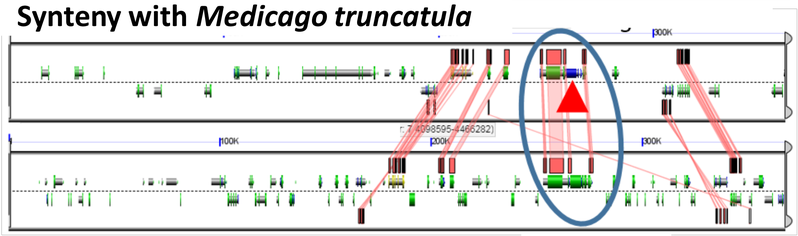
Based on high synteny (below) and clear intron-exon structure, recent whole-genome duplication was likely the process that led to the formation of the pseudogene (red triangle).

Recent genome duplication

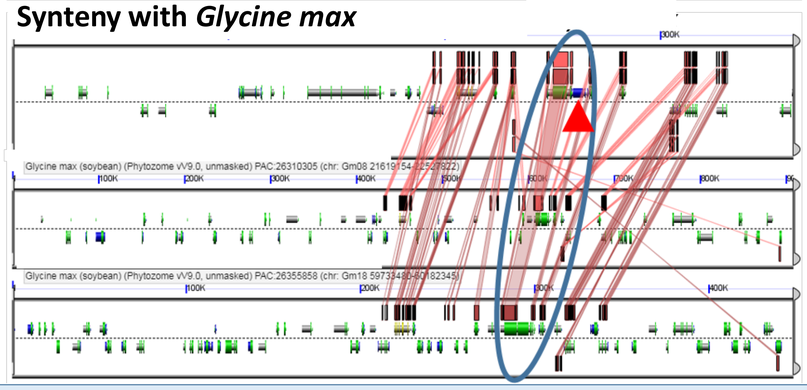
***Leucaena trichandra* contig 4097 synteny with contig 8890**



Synteny with *Medicago truncatula*



Synteny with *Glycine max*



MATERIALS AND METHODS

The potential pseudogene of interest was identified by finding a gene with a premature stop codon in the *Leucaena trichandra* contig 4097 PacBio assembly using Apollo (Lewis et al. 2002). Support for the premature stop codon was assessed by viewing Illumina reads mapped to the PacBio assembly in geneious (Kearse et al. 2012) to determine whether the stop codon may be due to sequencing error in the PacBio assembly.

Similar genes in *Glycine max*, *Medicago truncatula*, and elsewhere in the *L. trichandra* genome were identified using blat (Kent 2002). To explore the evolutionary history of the putative pseudogene, a custom phylogenetic tree was created with PLAZA (Van Bel 2017) using the putative pseudogene and its *Leucaena trichandra* in-paralog, as well as putatively orthologous sequences in the available legumes and 3 outgroup species.

Synteny between the putative pseudogene and genes in *Glycine max*, *Medicago truncatula*, and other regions of the *L. trichandra* genome was explored using SynFind and GEvo within the CoGe framework (Lyons & Freeling 2008) .

The process leading to the creation of the putative pseudogene was inferred by examining sequence similarity between the pseudogene and its in-paralog as well as through visually inspecting the phylogeny and synteny.

DISCUSSION

Leutr04097g005180.1 may be a pseudogene or in the process of becoming a psuedogene. However, there could still be an upstream error (not detected) causing a frameshift leading to a false-positive.

Clear intron-exon structure and synteny results suggest recent whole-genome duplication as the most-likely underlying process which led to the formation of the pseudogene as opposed to tandem duplication or retrotransposition.

It is unclear whether the 5' end of the gene still codes for a functional protein.

References

Jurka J. 2004. Current Opinion in Genetics & Development 14(6):603-608.
Kearse M, et al. 2012. Bioinformatics 28(12):1647-1649.
Kent WJ. 2002. Genome Research 12(4):656-664.
Lewis SE, et al. 2002. Genome Biology 3(12):research0082-1.
Lynch M, Conery JS. 2000. Science 290:1151-1155.
Lyons E, Freeling M. 2008. The Plant Journal 53(4):661-673.
Van Bel M, et al. 2017. Nucleic Acids Research 46(D1):D1190-D1196.
Vanin EF. 1985. Annual Review of Genetics 19:253-272.