Computational and Molecular Investigation of Helitrons in the Maize Genome

Research Team: Elizabeth Barrows, John Higgins, Sarah Smith
Advisors: Dr. Lal, Dr. Mili
Introduction and Outline of Presentation

- Background of Helitrons
- Computational Approach to Identifying Helitrons
- Experimental Approach to Identifying Helitrons
- Further Analysis to Take Place

- Picture of Everyone in Group!!!!!
Central Dogma

http://www.nature.com/nrg/journal/v3/n9/full/nrg890.html
Transposable Elements

- “Jumping Genes”—can move genetic material around within a genome
- Discovered in maize by Barbara McClintock in 1941
- Importance was not initially recognized, but she was awarded the Nobel Prize in Physiology or Medicine in 1983

http://www.osti.gov/accomplishments/mcclintock.html
Importance of Transposable Elements

- TEs represent a large portion of the genome of many species; approximately 40% of the human genome, 70% of the maize genome.
- Since TEs can move genetic material around, they have the potential to cause mutations.
- This can be a driving force for evolution of novel genes or the regulation of genes.

http://universe-review.ca/F11-monocell.htm#DNA
Helitrons are a Superfamily of TEs

- Discovered using computational methods searching model organisms’ genomes
- Make up about 2% of the maize genome
- Also present in plant species such as rice and morning glories, and animals such as the brown bat, C. elegans, and platy fish
Structure of Helitrons

- Helitrons capture genes or gene fragments
- Have conserved 5’ and 3’ sequences
- Insert between A and T
- Palindrome sequence at the 3’ end that forms a hairpin loop structure

5’ ATC

100bp - 25kb

CTRR T

3’
An Introduction to Computational Tools for Genomic Sequence Analysis

National Center for Biotechnology Information Tools

• BLAST Suite
  • Basic Local Alignment Search Tool
    • Aligns input sequences of multiple types against existing databases of sequences submitted by researchers.

• BLAST-N
  • Used to align nucleotide sequences (genomic DNA, mRNA, etc.) against sequence database.

• BLAST-X
  • Used to align nucleotide sequences (mRNA, cDNA, etc.) against protein sequences in multiple frames.

• ORF Finder
  • Translates nucleotide sequences into amino acid sequences in six reading frames and looks for start and stop codons.
Example Helitron Annotation BLAST-N

Input Sequence

Matching Alignments

Scoring

Query 435  GCCCGCGCTGGGCAGCGACTGG  ...  ACTTTCCATCATTATTTATTTTCTT  1555

Sbjct 1  GCCCGCGCTGGGCAGCGACTGG  ...  ACTTTCCATCATTATTTATTTTCTT  1030

>gb|BT062945.1|  Zea mays full-length cDNA clone ZM_BFc0010I18 mRNA, complete cds
Length=1030

Score = 1366 bits  Expect = 5e-97
Identities =1030/1030 (100%), Gaps = 0/1030 (0%)
Strand=Plus/Plus
Continuing Helitron Annotation BLAST-X

>gb|ACN27642.1| unknown [Zea mays]
Length=152

Score = 330 bits (846), Expect = 1e-88
Identities = 152/152 (100%), Positives = 152/152 (100%), Gaps = 0/152 (0%)
Frame = -3

Query 728  MDTKPWVSTYLTKLICLMQEHL...FIKPHHSPLGSIFRAQSCKIEK  273
            MDTKPWVSTYLTKLICLMQEHL...FIKPHHSPLGSIFRAQSCKIEK
Sbjct  1    MDTKPWVSTYLTKLICLMQEHL...FIKPHHSPLGSIFRAQSCKIEK  152
Continuing Helitron Annotation ORF-Finder

Possible Open Reading Frames (ORF)

Selected ORF and Start/Stop Codons

Length: 152 aa
Maize Specific Computational Tools

Combining it all with GeneSeqer

• Combines BLASTN–type alignment with open reading frame detection in all six frames.

• Predicts the splicing of EST introns/exons against genomic DNA using heuristics specific to maize genomic characteristics.

![Diagram showing ESTs, consensus ESTs, and open reading frames.](image-url)
GeneSeqer

• A much more efficient and accurate way of predicting the structure and characteristics of genes within a Helitron sequence.

• Allows for easy identification of possible protein products from Helitron sequence.

BLAST Suite

• Allows users to get a broader view of the context of the Helitron in the genome.

• Can provide clues into possible mechanisms and outcomes behind Helitron insertions.
Identifying Helitron Context Example

HTGS and GSS

• Collection of lengthy nucleotide sequences that were assembled in an attempt to sequence the entire genome of an organism.

5’

A T

Cut Helitron Out
Leaving Flanking Sequence

3’

AT

Connect Flanking Sequences and BLAST-N against HTGS and GSS
Identifying Helitron Context

BLAST-N Results

• Hits register in different strains in which some contain the insertion and others do not. Provides supporting evidence of a Helitron insertion.
Previous Approaches to Computational Helitron Discovery

HelitronFinder

• Chunguang Du, Hugo K. Dooner et al. Collaboration between Montclair and Rutgers University. Written in Perl programming language.

• Use terminus detection coupled with palindrome sequence identification algorithm to identify possible Helitron candidates.

Limitations

• Heuristic for terminus detection less stringent than experimental findings.

• Algorithm for palindrome detection searches in a vary specific region.

• Searches for Helitrons with lengths far beyond what is considered to be probable size.
Computational Approaches for Helitron Detection

<table>
<thead>
<tr>
<th>TENTATIVE ANNOTATION</th>
<th>5’ TERMINUS</th>
<th>3’ TERMINUS</th>
<th>FIRST REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>sh2-7527*</td>
<td>TCTCTACTCTAT</td>
<td>CGCTCCGTAGC- AACGCACGGGACATTACCTAG</td>
<td>Lal et al., 2003</td>
</tr>
<tr>
<td>ba1-Ref*</td>
<td>TCTCTACTCTTA</td>
<td>CGTTCCGTAGC- AACGCACGGGATACTACCTAG</td>
<td>Gallavotti et al., 2004</td>
</tr>
<tr>
<td>Rp1873**</td>
<td>TCTCTACTCTAT</td>
<td>GTCGCCGTCTG- AACGCACGGGACACTACCTAG</td>
<td>Gupta et al., 2005</td>
</tr>
<tr>
<td>ZeinBSSS53**</td>
<td>TCTCTACTCTCTAT</td>
<td>ACTTTCTGGCTAAACGCACGGGACATTACCTAG</td>
<td>Gupta et al., 2005</td>
</tr>
<tr>
<td>P450B73**</td>
<td>TCTCTACTCTAT</td>
<td>ACCTCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lal et al., unpublished</td>
</tr>
<tr>
<td>He1A-1*</td>
<td>TCTCTACTCTTA</td>
<td>--TCTGCGTCGCAACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>He1A-2**</td>
<td>TCTCTACTCTCTAT</td>
<td>--TCTGCGTCGCAACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>GHIJKLM9002*</td>
<td>TCTCTACTCTCTAT</td>
<td>GAACACGTAGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>NOPQ9002*</td>
<td>TCTCTACTCTCTAT</td>
<td>GTCGCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>NOPQB73_14578**</td>
<td>TCTCTACTCTCTAT</td>
<td>GTCGCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>NOPQB73_9002**</td>
<td>TCTCTACTCTCTAT</td>
<td>GTCGCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>Mo17NOPQ_14577**</td>
<td>TCTCTACTCTCTAT</td>
<td>GTCGCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>RST9002*</td>
<td>TCTCTACTCTCTAT</td>
<td>GTCGCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>U9002*</td>
<td>TCTCTACTCTCTAT</td>
<td>GTCGCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>HI9002*</td>
<td>TCTCTACTCTCTAT</td>
<td>GTCGCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>He18*</td>
<td>TCTCTACTCTCTAT</td>
<td>GTCGCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
<tr>
<td>TUVw9002*</td>
<td>TCTCTACTCTCTAT</td>
<td>GTCGCCGTGGC- AACGCACGGGACACTACCTAG</td>
<td>Lai et al., 2005</td>
</tr>
</tbody>
</table>

*discovered initially by +/- variation
**discovered by similarity of either terminal or internal sequence to other Helitrons

3’ terminal hairpins are shown in red
Developing a Heuristic for Computational Detection

100bp - 25kb
HelRaizer

Helitron Detection and Classification Algorithm

• Written in Python programming language using modules from the BioPython developers suite

Part 1 Helitron Detection

Input Genomic Sequence → Detect Conserved Helitron Termini → Finished Genome?

Yes → ID Helitrons Based on Termini Order and Proximity

No → Store Termini Type, Location, and Sequence ID → Store Helitron Sequence and 1000bp Flanking Sequence
Part 2 Determining Helitrons of Interest

Scoring Considerations

• Length of Alignment
• Number of Identities
• Number of Nucleotide Gaps
• Error Scoring (Statistical Probability)
## HelitronFinder vs. HelRaizer

<table>
<thead>
<tr>
<th>Program Name</th>
<th>Total Helitrons Found</th>
<th>Average Helitron Length (bp)</th>
<th>Median Helitron Length (bp)</th>
<th>Maximum Helitron Length (bp)</th>
<th>Minimum Helitron Length (bp)</th>
<th>Total Nucleotide Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helitron Finder</td>
<td>2,791</td>
<td>11,900</td>
<td>7,342</td>
<td>50,000</td>
<td>126</td>
<td>33.4Mb 1.39%</td>
</tr>
<tr>
<td>HelRaizer</td>
<td>2,573</td>
<td>10,252</td>
<td>6,757</td>
<td>49,975</td>
<td>168</td>
<td>26.4 Mb 1.09%</td>
</tr>
</tbody>
</table>
• Sample HelRaizer detected Helitrons have been shown to be absent in paralogous sequences.

• HelRaizer has detected Helitrons identical to HelitronFinder results, as well as Helitrons that have been discovered through experimental means.

• Further validation of HelRaizer results will come from future genomic analysis through experimental testing and procedures.
Genomic DNA Extraction

- Allows extraction of DNA from plant sample
- Plant sample pulverized using liquid nitrogen
- RNase to lyze cells
- Buffer to precipitate detergent, proteins, polysaccharides- filtered through column
- Manipulate environment using buffers, causing DNA to bind to another column, then release binding and go into solution
Gel Electrophoresis

- Gel produced from agarose and TAE buffer
- Ethidium bromide stain
- Application of electric current

- Add dye to DNA samples
- DNA migrates from negative to positive electrode
- Migration of DNA fragments is relative to size
PCR (Polymerase Chain Reaction)

- Amplifies a few copies of a DNA sequence into millions of the same sequence
- Components:
  - Primer
  - Taq enzyme
  - Deoxyribonucleoside triphosphates
  - Magnesium sulfate
  - Buffer
  - DNA template

http://www.obgynacademy.com/basicsciences/fetology/genetics/
Primer Design

- Primers are used in PCR; needed so that nucleotides can be added to a growing DNA chain
- Characteristics of a good primer
  - About 20 base pairs long
  - Equal percentages of A and T, G and C
  - Not self-hybridizing
  - Specific to the sequence to be amplified (not repeated throughout the genome)
Stages of PCR Reaction

- Initiation
- Denaturation
- Annealing
- Extension and Elongation
- Final Elongation
- Final Hold

Premise of Experimental Approach

- Putative helitrons with possible full-length genes located in the maize genome using computational approaches
- To prove the sequences are helitrons, must show plus-minus polymorphism
- Positive control- B-73 strain of maize
- Negative control- water control

http://www.biomedcentral.com/1471-2164/9/467/figure/F1?highres=y
Experimental Design for Plus-Minus Polymorphism

- **First procedure:** design a primer to excise a region overlapping flanking sequence and exon of helitron (positive result denotes helitron presence)
- **Second procedure:** design a primer to excise a region overlapping paralocus sequence of the helitron (positive result denotes no helitron)
  - Helitrons are too large to be picked up in entirety by PCR
Genomic DNA Extraction and PCR

- Practiced laboratory techniques for first weeks, extracted genomic DNA from ten inbred maize lines
  - Teosinte leaves showed no band on genomic DNA extraction
- 6/18: Ran PCR on twelve strains of genomic DNA using both sets of primers- 5’ primer showed no bands, 3’ primer showed several bands in several inbred strains
Primer Testing with PCR

• 6/19: Ran PCR using both sets of primers on B-73 to optimize thermal cycle- 5’ primer showed one band at expected length, 3’ primer showed multiple bands

• 6/22: Attempted to run PCR across ten inbred lines using 5’ primer- no bands for any strains, no amplification
• 6/23: Ran PCR with B-73 and HP-301 samples with our designed 5’ primer and a standard primer- both standard primers showed bands, 5’ primer showed band only for B-73
  – Plus-minus polymorphism?
Further Analysis

- Confirm plus-minus polymorphism of Helitron of interest
- Continue computational analysis of other interesting Helitrons
- Continue individual annotation of Helitrons of interest
- Investigate Helitrons of interest for polymorphism through paralocus primer design

http://www.nature.com/nature/journal/v443/n7111/full/443521a.html
Final Goals

• Contribute findings in order to assist in the understanding of the mechanism of gene capture by Helitrons

• Generate, analyze, and publish data to further research in the field of transposons
Questions?