Thank you Charles Darwin: A description of candidate wheat genes that may control wheat rust, derived from an evolutionarily based approach

Goal: To identify possible gene candidates for control of wheat stem rust Ug99.

Approach: Assess the genome of *Triticum monococcum* for positively selected genes that may have recently evolved in response to selection pressure from Ug99.
Our approach depends upon pairwise comparisons of transcriptomes on a gene-by-gene basis using the signature of positive selection as a screen.

Positive selection is here narrowly defined as changes in protein sequence that have occurred as a response to selective pressures, rather than simply as a result of stochastic fixation of random amino acid replacements. Positively selected changes give rise to novel protein conformations and thus new functions.
Previous Work

- Our approach uses pairwise comparisons in a system where there appears to have been a recent selective shift that occurred as a response to strong selective pressure

- Cassava: candidate genes for resistance to cassava brown streak disease (CBSD) using wild relatives of cassava

- Soybean: we have validated genes for control of soybean cyst nematodes (SCN)

- Candidates were validated in transgenic assays that demonstrate we can induce SCN resistance in previously susceptible commercial soybean cultivars. We are negotiating with a major agricultural company to license this genes for application in their germplasm

- Candidate yield genes being field tested by our rice seed partner
Positive selection can be detected in protein-coding genes by pairwise comparisons of the ratios of nonsynonymous nucleotide substitutions per nonsynonymous site \([K_a]\) to synonymous substitutions per synonymous site \([K_s]\)
Selected and Conserved Genes

• **Positively selected genes:** $\frac{Ka}{Ks} > 1$
  non-synonymous nucleotide changes (Ka) exceed synonymous (Ks) nucleotide changes.

• **Conserved genes:** $\frac{Ka}{Ks} < 1$
  synonymous nucleotide changes (Ks) exceed non-synonymous (Ka) nucleotide changes.
Molecular Evolution Analysis

EG’s Adapted Traits Platform is a unique comparative bioinformatics approach incorporating a molecular evolution algorithm (Ka/Ks) to find positively selected genes.

**Positively selected genes: Ka/Ks > 1**
non-synonymous nucleotide changes per non-synonymous (Ka) exceed synonymous (Ks) nucleotide changes per synonymous site.

**Conserved genes: Ka/Ks < 1**
synonymous nucleotide changes per synonymous site (Ks) exceed non-synonymous (Ka) nucleotide changes per non-synonymous site.

Align orthologs

Ka/Ks analysis

Ka/Ks > 1 (positively selected)

Candidate genes

Ka/Ks ≤ 1, archive 99.9999 % of comparisons
Results

• Analysis on-going: 3 positively selected genes have been confirmed so far

• Our first candidate appears to mediate drought/salinity tolerance

• Bowman-Birk type trypsin inhibitor (Serpin 1)

• Metallothionein type-2 protein (MT2)
Bowman-Birk type trypsin inhibitors

- Serine protease inhibitors that block trypsin
- Present in archaea, bacteria, fungi, animals, plants
- Aka Serpin (Serine peptidase inhibitor)
- Serpins found in cereal grains are ‘double-headed’ with 2 BBI domains containing reactive loops/binding sites connected by a linker region
- Expressed in wounded plant tissues and seeds
- Mediate protease regulation, PCD, and plant pathogen defense
Serpin 1

● Overall, Serpin 1 shows conservation of the ‘double-headed’ structure seen in cereal serpins, nonetheless, this protein shows the signature of strong positive selection

● Full length coding sequence Ka/Ks= 0.35

● Positive selection in Serpin 1 is localized in two regions that code for binding sites

● Ka/Ks in regions of positively selection = 1.7 - 2.1

● Conserved regions Ka/Ks = 0.0 – 0.34

● Leader Sequence Ka/Ks = 0.89
Serpin 1 Sequence Alignment

monococcum

MGMRKCI  LPSIMMLMALQASAAL  IFVADVGAILLPSQGKVADQAMAAA  AKRPWKKCDQAV

56

turgidum

MGMRKCI  VPSILLMLALQ-AALTVAVDVGDIILLPSQGQ-DEAAMAA  KKRWPWKCDQAV

60

monococcum

CTRSIEPPICTCMDQVFECPSCTKACGSPMDPSRRICQDQYVGDGPICRPEWCCDSPTC

116

turgidum

CTRSIEPPIICRMDQVFECPSCTKSCGPSMADEPSRLVCQDQYVLPGPICRPEWCCDSPTC

120

monococcum

TKSNPPTCRCGDEVDKCAPTCKTCLPSRPRPSRVRCLDSYFGAFPACTPSEAVAAGGN

175

turgidum

TKSNPPTCRCGDEVDKCAPTCKTCLPARSHPSRVRCLDSYFGAFPACTPSEAVAAGGN

179

Legend:

- Weakly Conserved
- Moderately to Strongly Conserved
- Positively Selected
- Protease Binding Site/Reactive Loop
- Leader Sequence
- BBI Domain
Metallothioneins

- Superfamily of short proteins, rich in cysteines
- In many organisms, these proteins are metal-binders that mediate metal uptake, transport and storage; this is less clear in plants
- MTs may protect against metal toxicity; MTs may also function to protect against oxidative stress. One recent study suggests that MTs in plants might be involved in plant defense against some fungal species, at least in citrus.
- Metallothioneins can be upregulated by biotic or abiotic stress
Candidate MT2

- Plant metallothioneins: conserved primary protein structure (Cys-Xaa-Cys): two cysteine-rich terminal domains, with more variation in the central linker domain

- No evidence of localization of positive selection in our candidate

- Full length coding sequence Ka/Ks= 2.0

- Short protein: 79 aa

- Located on wheat chromosome 3A

- Sequence variation within *monococcum* is 2% while sequence variation between *monococcum* and *turgidum* is 5%

We are looking for partners

• EG’s goal in this project was to identify potential gene candidates for wheat stem rust resistance

• We are looking for partners who have an interest in working with us to validate these candidates
Thank you

Tepper Koga
Geoff Hirsch
Kari Koga
David Johnson
Ginny Orndorff
Steve Warnecke
Charles Darwin

We thank the Bill and Melinda Gates Foundation for funding this work.