

Adaption in plant genomes: mutational target size and genome size

February 11, 2017

How plant adapted to different environment has been a long standing questions interested in by many evolutionary biologists. “adaptationist programme” has been dominated the evolutionary thought in the field from early to middle 19th century. Afterwards, Gould and Lewontin published their famous critique for the adaptationist programme, particular pointing out the role of genetic drift and discussed about different scenarios between selection and adaptation. Now we know most of mutations are neutral in the genomes and only the proportional of mutations are adaptive for the organism in response to the environments. It is important to understand what is the underlying evolutionary selection for the adaptive trait (eg. hard sweep, soft sweep and polygenic trait). Does adaption always occur in the certain place of the genome? Does mutational target size diff across different species?

1. Where are functional mutations in maize?

1.1 Different types of adaptation in maize

- In general, hard sweep play minor role in maize and eg. hard sweep (protein change) - *tga1*.
- soft sweep including: standing genetic variation, regulatory change (enhancers expression) - *tb1*; multiple mutations (modify expression) - *gt1*.
- polygenic adaptation - Small effect polygenic variants are responsible for most of the standing variation for domestication-related traits in maize.

1.2 Lots of functional mutations are outside of coding sequence

- 70% of GWAS hits are outside of annotated genes, with the peak at 1-1.5kb away from genes, which indicates the regulation role of cis-regulatory elements.
- MNase HS regions (less than 1% of the maize genome) explain a remarkably large amount (about 40%) of heritable phenotypic variance in diverse complex traits.
- certain TE nearby genes can activate maize gene expression upon abiotic stress.

1.3 Lots of stuff showing selection includes regulatory sequence, and some regions showing selection have no genes in them

- selection scan the teosinte, landrace and improved maize lines, 6% domestication regions and 11% improvement regions identified by cross-population composite likelihood ratio test contained no annotated sequence.
- 1,100 genes show transcriptome expression change and rewiring, and significantly enriched previously identified genes with target of selection during maize domestication and improvement.
- In teosinte, candidate SNPs associated with environmental variables enrichment in nongenic regions.

2. What selection looks like in other organisms?

2.1 Enrichment of selection on nonsynonymous sites

- Current study suggested that whereas *Drosophila*, mice, and bacteria have undergone extensive adaptive evolution.
- In *Arabidopsis*, non-synonymous SNPs enriched in the environmental adaption.

- In *Drosophila simulans*, the adaption is primary on the non-synonymous substitutions, and estimate approximately 13% of amino acid substitutions were adaptive.
- In *Capsella grandiflora*, the selection is strong in coding regions, and weak on noncoding regions.

2.2 No enrichment for nonsyn in humans (large genome)

- Human genome show little or no evidence of adaptive evolution in protein-coding sequences.
- Classic selective sweeps were rare in recent human evolution.
- Soft sweeps predominant form of adaptation in humans.
- Adaptation was frequent in human evolution and provide support for the hypothesis of King and Wilson that adaptive divergence is primarily driven by regulatory changes.

2.3 for most plants

One hypothesis is that the pattern of difference in adaptive evolution is related to effective population size. The estimated adaptive rate of evolution is low in 11 plant species surveyed across monocots and eudicots, and effective population size does not seem to explain the pattern observed.

3. Selection on noncoding sites

- approximately 90,000 conserved noncoding sequences (CNSs) that show evidence of transcriptional and post-transcriptional regulation provide the evidence of selection on regulatory region.
- fast turn over rate of regulatory elements, balance selection might be able to explain the enhancer/functional elements fast turn over rate and polymorphism within functional elements.
- delete megabase conserved non-coding sequence between human and mice and result is indistinguishable.
- 15,363 orthologous conserved non-coding sequence identified conserved across pan-grass including rice, foxtail millet, sorghum, brachypodium, and maize.

4. Thoughts on soft sweeps

- Bring up the concept of mutational target size.
- parallel adaptation depends on mutational target size, if you have a really narrow mutational target size, magic parallel adaptation (eg. seed shattering across grass, starch metabolism pathway).
- comparing highland adaption in maize between mesoamerican and south American, have very low level of parallel, maybe

because maize has a large mutation target space or high land adaption?

5. Mutational target size

How to decide mutation target size? (DNase I hypersensitive Sites?), large genome has large mutational target size, so the selection primary on non-coding region, small region has small mutational target size, so the selection primary on coding region of the genome. Maize has normal genome size compared to other angiosperms, so the selection/adaption pattern observed in maize might be occur in many other plants. Maize, not Arabidopsis or other small genome plants, is the model to study plant adaption.