

RESEARCH HIGHLIGHT

Repeated phenotypic changes highlight molecular targets of convergent evolution

Maggie R Wagner and Thomas Mitchell-Olds*

Abstract

How predictable is evolution at the molecular level? An example of repeated evolution in rice and Brassica illustrates how selection might preferentially target certain genes and mutations.

Keywords Convergence; parallel evolution; *cis*-regulatory evolution; fruit development; *Brassica*; natural selection; artificial selection; pleiotropy.

Evolution sometimes repeats itself. Similar phenotypes arise not only from genetic heritage but also from the shared need to adapt to certain ecological challenges, a need that is shared even between lineages that are widely separated in time and space. For instance, succulent stem tissue and leaves that are modified into spines have arisen independently in the Old-World Euphorbiaceae and the New-World Cactaceae as morphological adaptations to life in deserts. Other shared adaptations not due to common ancestry include C_4 photosynthesis [1], which has independently evolved more than 45 times, and secondary metabolism [2].

Any occurrence of parallel adaptation is remarkable considering the miniscule probabilities that two divergent genomes will independently respond to a given selection pressure in an identical manner. Thus, its prevalence raises several questions. Are some phenotypes more visible to natural selection than others? Do selectively advantageous mutations target certain parts of genes and pathways? In short, how predictable is adaptive evolution?

The genetic bases of most shared adaptations are unknown, but the complexity of genomes suggests that similar phenotypes could have different molecular origins. Recently, however, Arnaud *et al.* [3] used a fruit morphology phenotype that is shared by two genetically well-characterized taxa to demonstrate that convergent

nucleotide changes can underlie similar phenotypic traits. Their investigation began by looking at variation within the Brassicaceae in the morphology of the replum, a structure dividing the two valves of the fruit. The long seed capsules, or siliques, of wild-type *Arabidopsis* plants boast a wide replum that fully separates the valve margins, whereas *Brassica* fruits have a thin replum that is breached by the valve margins [3], similar to that of the *Arabidopsis* mutant *replumless* (*rpl*). Phylogenetic footprinting followed by transgenic verification revealed that a C>T point mutation in a *cis*-regulatory element of *RPL*, which encodes a transcription factor, explains the evolution of a thin replum in *Brassica* from the ancestral wide replum of *Arabidopsis*. The wide- and thin-replum forms of this *cis*-element (*Shattering element-like* (*Shl*)) are named *C-Shl* and *T-Shl* after the *Arabidopsis* and *Brassica* sequences, respectively.

A reduced abscission zone and the loss of seed shattering in domesticated rice are attributed to a C>A transversion at the same position in the *Shl* region of the rice *RPL* ortholog. Thus, a single point mutation that has a strong phenotypic effect has been fixed in orthologous genes in two distantly related organisms.

Are some genes more evolvable than others?

RPL encodes a transcription factor that inhibits the expression of several other genes: *SHATTERPROOF1* (*SHP1*), *SHP2*, and *INDEHISCENT* (*IND*), which all induce cells to differentiate as valve margin cells [4]. Down-regulation of *RPL* causes the valve margins to span the width of the fruit [3,5]. Theoretically, up-regulation of *SHP1*, *SHP2*, or *IND* should produce a similar phenotype. Why, then, did this feature evolve in both rice and *Brassica* through the perturbation of *RPL* rather than by mutations elsewhere in the pathway?

Several properties of a gene could make it more evolvable. First, physically large genes present bigger targets for mutation. Assuming a uniform mutation rate, a greater number of selectable mutations occur in a long gene than in a short one. Related to mutational target size is the proportion of potential mutations that actually change the function of the gene [6]. For genes that encode enzymes, this depends on the proportion of the gene in

*Correspondence: tmo1@duke.edu
University Program in Genetics and Genomics, Institute for Genome Sciences and Policy Box 90338, Duke University, Durham, NC 27708, USA

which a substitution would alter the biochemical properties of the encoded protein, such as its solubility, substrate affinity, or protein stability. For genes that encode transcription factors, this may manifest as the proportion of the encoded protein that determines its affinity for DNA sequences. DNA-protein interactions that require a close match between binding site and *cis*-element are more sensitive to mutations than those that allow some variation in the protein and target sequences.

All else being equal and assuming that mutations occur randomly throughout the genome, a gene whose functionality is easily changed by one or a few mutations is more likely to be involved in parallel evolution than a gene whose functionality is less likely to be affected [6]. Such genes are also less able to maintain intraspecific variation because they are tightly constrained by purifying selection: within-species variation is more likely to evolve in downstream genes that can tolerate mildly deleterious mutations [7].

Second, a gene's position in its biochemical or regulatory network could also affect its evolvability, because certain types of genes contribute disproportionately to quantitative traits. In a computational model of biosynthetic pathway evolution, substitutions in upstream enzymes had greater fitness effects than biochemically equivalent substitutions in downstream enzymes [8]. Adaptive mutations in a gene of large effect should increase in frequency more rapidly than those in genes that have only a mild influence on the selected phenotype [6]. This also holds for transcription-factor genes, which tend to have larger phenotypic effects when they encode global regulators [7] or nodes in a network [6]. Such transcription factors could include those that integrate inputs from many sensory genes or distribute a signal to downstream genes. It is unclear how well RPL fits this description. Although RPL is known to regulate several genes that are active in fruit development, the full network that involves this transcription factor has not been elucidated [4]. Nevertheless, knockout experiments have shown that RPL strongly affects replum morphology [5], a phenotype of putatively high adaptive value.

Do adaptive mutations target certain parts of genes?

It is remarkable that one functionally important point mutation reached fixation in both *Brassica* and rice. Assuming that RPL was targeted by selection because it is a gene that has a large effect on fruit morphology, why did the adaptive phenotype not evolve first through amino acid substitutions in the RPL protein caused by mutation of the RPL gene itself, which presents a larger mutational target than *Shl*?

Most reports of molecular convergence describe identical amino acid changes in homologous proteins [9],

but regulatory mutations might be more selectable for several reasons. First, gene expression variants are usually semi-dominant, causing a selectable phenotype in heterozygotes. By contrast, adaptive protein variants might act recessively, slowing the response to natural selection. Second, variation in gene expression can create more nuanced phenotypes than variation in amino acid sequence because of the quantitative and dynamic nature of gene expression [10]. A crucial implication of transcriptional flexibility is that evolution arising from *cis*-regulatory variation could reduce the pleiotropic effects of new mutations.

For genes encoding proteins that have only one function, mutations that cause amino-acid substitutions are the predominant mechanism of adaptive evolution [6]. Pleiotropic genes, however, are less amenable to coding changes because an improvement to one function may harm another. Such tradeoffs decrease the selection pressure for an otherwise helpful mutation, making it less likely to reach fixation. For such genes, adaptive evolution may occur more readily through *cis*-regulatory mutations that alter certain parts of the expression profile without perturbing others. For instance, such mutations might affect expression only in specific tissues or environmental conditions. In the case of RPL, which also influences leaf and meristem development [4], T-*Shl* alters the gene's expression during fruit development but does not affect its other activities [3]. Both the introduction of T-*Shl* and knockout of RPL caused the thin replum phenotype, but defects in leaf and meristem morphology were observed only in the RPL knockout. This suggests that *Shl* controls RPL expression only during fruit development [3].

Are related organisms predisposed to converge genetically?

Although parallel evolution operates by a different mechanism than simple inheritance, its likelihood still depends on the evolutionary relationship between the species that share a selective pressure. Because genes typically act within networks of other genes, identical mutations that occur in two highly divergent genomes can result in completely different phenotypes [9]. Therefore, genetic convergence is more likely to occur between closely related organisms that have a high degree of sequence similarity. The precise genetic convergence observed between rice (a monocot) and *Brassica* (a eudicot) is impressive given the ancient divergence between these species [3]. Notably, both RPL and *Shl* are highly conserved, and thus have a heightened probability of parallel mutation and selection.

Furthermore, closely related organisms are more likely to share segregating variants that are inherited from their most recent common ancestor. If an environmental shift suddenly makes one variant more advantageous than

another for both species, the favored allele has a better chance of being fixed and is unlikely to be lost to drift because of its initial intermediate frequency [6]. By contrast, the probability of the same new mutation arising twice independently is low, and the likelihood of both mutations becoming fixed is even lower.

Tying mutation to natural selection

The repeated evolution of fruit anatomy through point mutations in *RPL* is a striking illustration of genetic convergence between relatively divergent organisms. The evolutionary paths of both rice and *Brassica* were, however, probably biased toward changes in a *cis*-regulatory element controlling a gene that has a strong and specific effect on fruit development. Assuming that a reduced replum or abscission zone was selected for in both plants, *RPL* (as opposed to other genes in the fruit development pathway) might have been targeted because of its major influence on the advantageous phenotype. Within *RPL*, the *cis*-regulatory element *Shl* was probably targeted because it allows a beneficial change to fruit morphology without perturbing the gene's pleiotropic effects on stem and leaf development. The story is incomplete, however, without knowing why, or whether, natural selection led to the fixation of T-*Shl* in *Brassica*. The timing of dehiscence is important for wild plants, playing an important part in ensuring the germination of their next generation. Nevertheless, the ecological implications of a thinner replum for *Brassica* are not obvious: it is not even clear that T-*Shl* alters fruit dehiscence. In this case, the phenotypic convergence is more mysterious than the underlying genetic convergence.

Abbreviations

C-*Shl*, wide-replum form of *Shl*; *IND*, *INDEHISCENT*; *rpl*, *replumless*; *Shl*, *Shattering element-like*; *SHP1*, *SHATTERPROOF1*; T-*Shl*, thin-replum form of *Shl*.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by grants from the US National Institutes of Health (NIH) (R01 GM086496) and US National Science Foundation (NSF) (EF-0723447) to TMO.

Published: 23 August 2011

References

1. Sage R: The evolution of C(4) photosynthesis. *New Phytol* 2004, **161**:341-370.
2. Pichersky E, Gang DR: Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci* 2000, **5**:439-445.
3. Arnaud N, Lawrenson T, Østergaard L, Sablowski R: The same regulatory point mutation changed seed-dispersal structures in evolution and domestication. *Curr Biol* 2011, **21**:1215-1219.
4. Østergaard L: Don't "leaf" now. The making of a fruit. *Curr Opin Plant Biol* 2009, **12**:36-41.
5. Roeder A, Ferrándiz C, Yanofsky M: The role of the REPLUMLESS homeodomain protein in patterning the *Arabidopsis* fruit. *Curr Biol* 2003, **13**:1630-1635.
6. Gompel N, Prud'homme B: The causes of repeated genetic evolution. *Dev Biol* 2009, **332**:36-47.
7. Jenkins DJ, Stekel DJ: De Novo evolution of complex, global and hierarchical gene regulatory mechanisms. *J Mol Evol* 2010, **71**:128-140.
8. Wright KM, Rausher MD: The evolution of control and distribution of adaptive mutations in a metabolic pathway. *Genetics* 2009, **184**:483-502.
9. Christin P-A, Weinreich D, Besnard G: Causes and evolutionary significance of genetic convergence. *Trends Genet* 2010, **26**:400-405.
10. Wray GA: The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 2007, **8**:206-216.

doi:10.1186/gb-2011-12-8-124

Cite this article as: Wagner MR, Mitchell-Olds T: Repeated phenotypic changes highlight molecular targets of convergent evolution. *Genome Biology* 2011, **12**:124.