Molecular-genetic Extensions of Vavilov's Predictions

Kevin M. Folta[1]

Horticultural Sciences Department, University of Florida, Gainesville, FL 32606

Additional index words. evolution, genetic traits, parallel selection, photoperiod, shattering

***Abstract*. Vavilov's *Law of Homologous Series* indicates that heritable variation for a given trait will occur in different species based on parallel selection. The conclusion comes from Vavilov's study of extensive collections and careful attention to phenotypic variation across taxa. The same examination of variation can be applied to traits using the power of genetic and genomic resolution, because parallel traits would be expected to be governed by the same genetic loci, and perhaps even common mutations. In this review, these concepts are applied to two central traits—the control of "shattering" of kernels in cereals and in the control of photoperiodic flowering. One of the strengths of the law is that it can make predictions about traits and perhaps the genes or genomic regions that control them. With respect to genetic variation, the occurrence and physical location of genes associated with kernel retention may be predicted. Many grains share mutations, such as the *Sh 1* gene, which were selected in parallel. Selection of the *Sh1* gene led to higher yields due to better kernel retention. While the genes affected are often the same, the types of mutations are not, implying convergent selection. Flowering time is governed by multiple loci, so variation may be attributed only to a few candidates, yet because of the number of regulators the predictive power of the law is lower. The modern application of the *Law of Homologous Series* is that it allows basic researchers or plant breeders to make predictions about the genes controlling key traits, although the genetic basis of variation is likely not conserved.**

Paleontologists have identified many gaps in the fossil record. One gap in particular noted a conspicuous absence of a species bridging late Devonian fishes and the first amphibious tetrapods. Based on careful study of form and function, paleontologists hypothesized a water-dwelling creature that would have specialized limbs for exploiting the water–land interface, probably once flopping around near shore, in ponds, or in drying lake beds. They also could guess at when and where it would have lived—probably 350–400 million years ago. Evidence of the hypothetical beast might be found in the fossil strata from the corresponding period.

In 2004, a team led by Dr. Neil Shubin searched the Devonian shales of Northern Canada for this particular predicted specimen. Only two years later the group published its findings in the journal *Nature* (Shubin et al., 2006). As predicted, they had found the organism and named it *Tiktaalik roseae*. The species bore features of fish and amphibians, particularly predicted features consistent with exploitation of land. This is the kind of evolutionary prediction that N.I. Vavilov made for plant species in the early 1900s. Throughout the study of evolution and speciation, scientists have carefully analyzed organismal form, function, and diversity to make predictions about the unknowns and instances of convergent evolution. New information refines our deficiencies. These fresh areas of inquiry present a renewed foundation for more elaborate or precise hypotheses that seek to fill these scientific vacancies.

Received for publication 18 Feb. 2015. Accepted for publication 20 Mar. 2015.

This paper was presented as part of a Workshop entitled "Nikolai Ivanovich Vavilov: Plant Genetic Resources Pioneer and Hero" on July 28, 2014 at the ASHS Annual Conference in Orlando, FL.

[1]To whom reprint requests should be addressed; e-mail kfolta@ufl.edu.

Nikolai Vavilov was an expert at analyzing information and making predictions about the unknowns, as inferred from analysis of plant traits in extensive collections. In a time coincident with first discussions of genes and their connection to traits, Vavilov posited that similar traits would be found in neighboring taxa, and that similar geographical or environmental influences may shape similar form and function (Vavilov, 1922). This was the basis of the *Law of Homologous Series* that states closely-related organisms would show parallel variation in shared traits. It was also noted that common characters in similar species would be identified in different geographical regions.

REVISITING VAVILOV'S LAW OF HOMOLOGOUS SERIES

Today it is possible to revisit Vavilov's concepts, not just with an eye on plant morphology and behavior, but also through use of molecular tools that allow scientists to define the genes associated with biological traits. From these analyses, we find cases where principles of homologous series apply to the molecular underpinnings. In other situations the whole-plant similarity arises from different genes that contribute to the ultimate phenotype. When a number of important plant traits are analyzed to the gene level, we find frequent evidence of parallel variability.

Seed shattering. One of the best examples of common genetic factors contributing to a homologous series is kernel retention on seed heads in grains (Li et al., 2006). Early grain production was limited by the loss of kernels from the panicle. As grains dried in the field, were harvested or moved, kernel loss led to reductions in yield. Identification of genotypes that resisted seed dispersal from shattering seed heads was an important step in grain cultivation. Not only would seed retention allow greater harvests, it would also facilitate more permissive transportation and would decrease radiation of species with invasive tendencies (Doust, 2013). Today seed retention remains an important trait selected in breeding programs (Konishi et al., 2006). On the other hand, a tendency to disperse seed is favored under strong natural selection, so nonshattering populations are almost certainly remnants of domestication.

Vavilov's *Law of Homologous Series* would have predicted that seed retention would be a common trait and governed by common mechanisms, even across grains. The same sentiment was noted in Paterson et al. (1995) who predicted that "A conserved mechanism would underlie seed retention in all grasses." These initial predictions were made from comparison of emerging genetic maps, and would find substantial support in the genomics era. While grass genomes vary widely in size based on repetitive DNA, there is significant retention of colinearity between orthologous genomes (Gaut, 2002), particularly in the regions germane to agronomic traits. Today the literature shows that during the domestication of wild grasses, certain key traits would in fact be selected in parallel in ranging regions of the globe—and the same genes were affected in many of these cases.

Sorghum has a genetically simple basis for controlling seed dispersal. In this species, the shattering phenotype is controlled by a single locus, called *Sh1*. The trait was uncovered by examining a cross between a wild sorghum (*Sorghum virgatum*) and a domesticated, nonshattering line called Tx340. Segregation in the F2 showed that a single locus was controlling the shattering phenotype. Fine mapping and cloning of the causal gene show that *Sh1* encodes a *YABBY-type* transcription factor (Lin et al., 2012). Examination of shattering and nonshattering genotypes showed that a haplotype consistent with *S. virgatum* was represented across all shattering genotypes. Three other alleles were found among the nonshattering genotypes. Association analysis comparing many shattering and nonshattering genotypes also showed that the same transcription factor was

affected in these accessions. All domesticated sorghums show a mutation in *Sh1*, although they were three separate mutation events (Lin et al., 2012). These alleles are separated in distinct regions of the African continent. This distribution suggests that three separate groups identified nonshattering sorghum and fixed the causal alleles in the subsequent breeding populations.

While *Sh1* was identified as a strong regulator of inflorescence disarticulation in sorghum, the question was whether the same gene might affect the trait in other species. Later it was shown to contribute to the same phenomenon in some rice accessions, maize, and likely in foxtail millet. The absolute contribution of this gene appears to be prevalent, but is not the sole basis for domestication. *Sh1* has only a minor effect in many rice genotypes. In rice, shattering is primarily controlled by other genes that create a condition where mutation of the *Sh1* gene is most readily observable, such as the morphological changes seen by Ishii et al. (2013). Here a change in the inflorescence architecture is shown to affect seed retention, and may represent an alternative mechanism derived through domestication. Still, the prevalence of one major locus controlling this key trait across diverse species demonstrates its role in domestication.

Sh1 does play a role in rice seed retention, depending on the genotype. Two rice genotypes, Nipponbare (nonshattering) and Kasalath (shattering) were compared using QTL analysis (Konishi et al., 2006). The results showed that the major contributor to shattering was a locus on chromosome 1. Fine mapping revealed that the defect again was in the *Sh1* gene, in this case a single point mutation in the same YABBY-type transcription factor. Across grains, there is a strong conservation of syntenic blocks associated with shattering and the *Sh1* locus, aligning well across maize, rice, and foxtail millet (Fig. 1).

Different disruptions in *Sh1* happen within and between species. In rice, *Sh1* mutations lead to changes in the abscission layer, and were epistatic to other loci controlling shattering (Onishi et al., 2007). Mapping of the mutation found a ≈4 kb insertion even into the third intron of the gene that disrupted its function. Analysis across many rice varieties suggests that the *Sh1* gene was under strong artificial selection (He et al., 2011). In maize, there are *Sh1*-containing loci on chromosomes 1 and 5. The *sh1* mutations are present in a large number of maize inbreds, and only in a few teosinte inbreds (Lin et al., 2012). A sample of the mutation diversity is shown in Figure 2.

The conservation of genes affected in a similar strongly selected domestication trait would likely be expected when considering genes associated with a common trait in a homologous series. However, this assumption is far too simplistic, as revealed by analysis that identifies other genes contributing to these domestication traits. For instance, even in sorghum, additional mutations have been identified in nonshattering lines that are independent from the *Sh1* locus. Association mapping has identified another transcription factor, in this case a WRKY-type that appears to play a synergistic role with *Sh1* (Tang et al., 2013). It is hypothesized that the two may play roles in lignin deposition as a mechanism toward seed retention.

These findings are not surprising, as multiple loci would be expected to contribute differentially depending upon other structural and physiological attributes of the species. Thus, it seems as though a Vavilovian homologous series does not extend as well to the gene as it does to the trait itself. While a locus like *Sh1* shows dominant effects across grains, the existence of other influential genes is certainly noted.

Flowering control. Another set of traits worth examining in the context of homologous series are the traits associated with flowering control. Precise control of the transition to flowering has profound consequences ecologically and commercially, so it is subject to intense natural and artificial selection. In nature, the ability of plants of the same species to flower in synchrony can have great advantage, as it increases the likelihood of outcrossing. Examination of flowering-related genes shows that many flowering-related traits are governed by a common set of genes, consistent with the concept of a common trait in a homologous series being governed by a common genetic mechanism.

Flowering-related traits range from sensitivity to photoperiodic stimulus to a complete lack of tie to light duration cues. Continuous flowering, or remontancy, is a desirable trait in many plant species, especially as it relates to commercial production of various crops. Continuous flowering has been of interest in strawberry (*Fragaria* spp.) and roses (*Rosa* spp.), and both crops have representatives that exhibit this trait. In strawberry the genetic control is conferred by SEASONAL FLOWERING LOCUS (SFL; Brown and Wareing, 1965) and in rose it is encoded by RECURRENT BLOOMING (RB; Semeniuk, 1971). As rose and strawberry are phylogenetic neighbors, it may be assumed that a common mechanism controls this trait.

Work by Iwata et al. (2012) showed that this was in fact the case. In these two species, continuous flowering is due to mutation in a gene called *TERMINAL FLOWER 1* (*TFL1*), which is encoded by the *SFL* and *RB* loci in their respective species. In roses, the gene is disrupted by a retrotransposon in the second intron, whereas in strawberry the same trait is conferred by a 2 bp deletion causing a frameshift in the gene. TFL1 is a floral repressor that when repressed by mutation, allows for remontant flowering habits (Hanano and Goto, 2011).

TFL1 also plays a central role in other crop species. In grapevine, a TFL ortholog controls the proliferation of meristems in the formation of flower clusters, as deduced from a somatic mutant with extensive flower meristem production that was shown to possess a transposon-mediated insertion in the TFL1A gene (Fernandez et al., 2010). Analysis of sequence polymorphism across grape accessions showed three distinct variations in the TFL1A, likely representing three genetic pools of domestication (Fernandez et al., 2014).

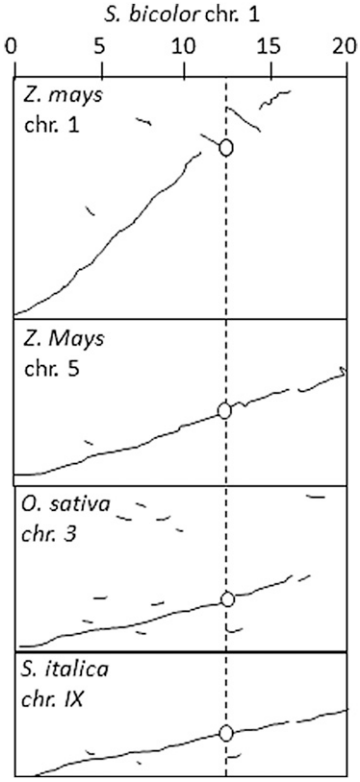

Fig. 1. A diversity of mutations in the *Sh1* gene confers the nonshattering trait across species. *S. bicolor* features a 4 bp deletion of a proximal promoter regulatory element. The *Z. mays* B73 *Sh1-1* allele features a 19.3 kb insertion into the first intron, as well as an 83 bp insertion into an exon. The *Z. mays Sh1-5.1* allele contains a ≈23 kb insertion that represents translocation of another chromosome sequence, containing two exons from a non-*Sh1* gene (grey boxes). *O. sativa Sh1* function is disrupted by a ≈4 kb insertion into an intron of the native gene. The *O. sativa qSh1* allele is caused by a single base transversion, apparently 12 kb upstream of translational start.

Study of the genes controlling flowering time variation likely narrows down gene candidates to a half-dozen potential genetic regulators that continually emerge in the literature as controlling flowering behaviors, consistent with predictions from homologous series. QTL analysis on sorghum shows that the CONSTANS protein (CO; *heading date 1*) confers long-day behaviors crucial to maximizing biomass accumulation (Yang et al., 2014). Polymorphism in the *Co* gene also appears to be central to speciation between two species of red oak (Lind-Riehl et al., 2014). The *Co* and *Co-like* genes have been analyzed in soybean and suggest that variation in flowering behaviors is linked to polymorphism in these genes (Wu et al.,

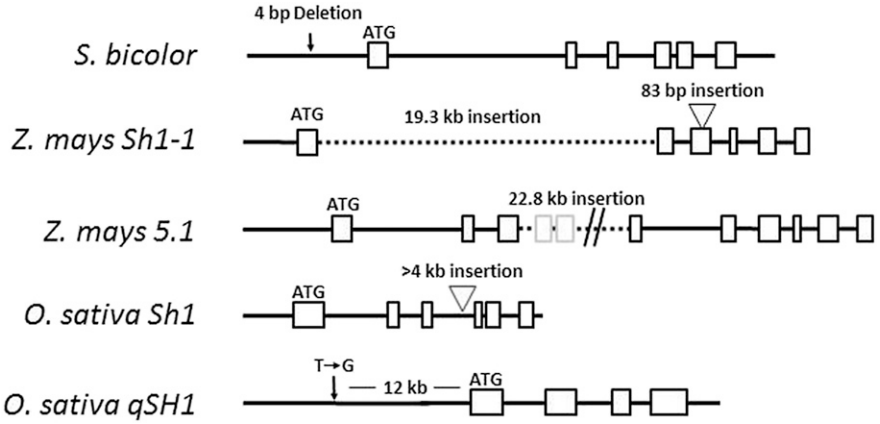

Fig. 2. Maintenance of synteny within the *Sh1* genomic region. The *Sh1*-containing region of *S. bicolor* (horizontal axis) is compared with the corresponding genome regions from *Z. mays* chromosome 1 and 5, *O. sativa* chromosome 3 and foxtail millet (*S. italica*) chromosome IX. White circles represent the *Sh1* locus and the lines represent corresponding genomic regions. Redrawn from data in Lin et al. (2012).

2014). Polymorphism in photoreceptors such as phytochromes and cryptochromes have been described as the genetic basis for flowering behaviors in a number of species, including natural variation in Arabidopsis (Aukerman et al., 1997; El-Din El-Assal et al., 2001). Analysis of flowering time shows that there are several dominant genes that account for variation observed in nature.

SUMMARY

This short review extends the concepts of Vavilov's *Law of Homologous Series* to the level of the gene. There are several interpretations that can be taken from this concept. First, because many traits are controlled by multiple loci, the concepts of homologous series apply best at the gene level to traits were they are controlled by a single major dominant locus. The SH1 shattering trait is a good example, especially in species like sorghum. On the other hand, when a trait is governed by at least a dozen strong loci, prediction of the applicable genes is less precise, as in the vegetative-floral transition.

The predictive power of identifying the molecular-genetic basis underlying traits in homologous series falls apart even more when the variation is examined at the level of DNA. Common mutations are not typically found across species, suggesting that common traits are obtained by convergent selection around these few critical loci. This observation was especially apparent in the examination of *Sh1* in grains, where maize, rice, sorghum, and millet (Lin et al., 2012), where disruption of the central locus controlling shattering behaviors varies from simple frameshifts to disruptions with retrotransposons.

In plant breeding and in basic science there still is an interest to understand the genetic basis underlying key traits. While the ability to predict the molecular basis of change is impractical, there is significant informational inertia to allow good predictions of which genes are best to study first. The concept is important to apply to breeding, as variation in a specific gene tied to a key trait can serve as the basis for molecular markers to hasten selection. In contemporary genomic analysis, an understanding of the genes associated with key traits allows researchers to focus efforts down to a number of candidate regulators that may be studied within massive data sets. These approaches are based on the concept that independently selected traits will be based on common regulators, and are consistent with Vavilov's *Law of Homologous Series*.

Literature Cited

Aukerman, M.J., M. Hirschfeld, L. Wester, M. Weaver, T. Clack, R.M. Amasino, and R.A. Sharrock. 1997. A deletion in the PHYD gene of the Arabidopsis Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. Plant Cell 9:1317–1326.

Brown, T. and P.F. Wareing. 1965. Genetical control of everbearing habit and 3 other characters in varieties of *Fragaria vesca*. Euphytica 14:97–112.

Doust, A.N. 2013. New ways to gather grains. Dev. Cell 25:114–115.

El-Din El-Assal, S., C. Alonso-Blanco, A.J. Peeters, V. Raz, and M. Koornneef. 2001. A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. Nat. Genet. 29:435–440.

Fernandez, L., L. Le Cunff, J. Tello, T. Lacombe, J. M. Boursiquot, A. Fournier-Level, G. Bravo, S. Lalet, L. Torregrosa, P. This, and J.M. Martinez-Zapater. 2014. Haplotype diversity of VvTFL1A gene and association with cluster traits in grapevine (*V. vinifera*). BMC Plant Biol. 14:209.

Fernandez, L., L. Torregrosa, V. Segura, A. Bouquet, and J.M. Martinez-Zapater. 2010. Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. Plant J. 61:545–557.

Gaut, B.S. 2002. Evolutionary dynamics of grass genomes. New Phytol. 154:15–28.

Hanano, S. and K. Goto. 2011. Arabidopsis TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression. Plant Cell 23:3172–3184.

He, Z., W. Zhai, H. Wen, T. Tang, Y. Wang, X. Lu, A.J. Greenberg, R.R. Hudson, C.I. Wu, and S. Shi. 2011. Two evolutionary histories in the genome of rice: The roles of domestication genes. PLoS Genet. 7:e1002100.

Ishii, T., K. Numaguchi, K. Miura, K. Yoshida, P. T. Thanh, T.M. Htun, M. Yamasaki, N. Komeda, T. Matsumoto, R. Terauchi, R. Ishikawa, and M. Ashikari. 2013. OsLG1 regulates a closed panicle trait in domesticated rice. Nat. Genet. 45:462–465.

Iwata, H., A. Gaston, A. Remay, T. Thouroude, J. Jeauffre, K. Kawamura, L.H.-S. Oyant, T. Araki, B. Denoyes, and F. Foucher. 2012. The TFL1 homologue KSN is a regulator of continuous flowering in rose and strawberry. Plant J. 69:116–125.

Konishi, S., T. Izawa, S.Y. Lin, K. Ebana, Y. Fukuta, T. Sasaki, and M. Yano. 2006. An SNP caused loss of seed shattering during rice domestication. Science 312:1392–1396.

Li, C., A. Zhou, and T. Sang. 2006. Rice domestication by reducing shattering. Science 311:1936–1939.

Lin, Z., X. Li, L.M. Shannon, C.-T. Yeh, M.L. Wang, G. Bai, Z. Peng, J. Li, H.N. Trick, T.E. Clemente, J. Doebley, P.S. Schnable, M.R. Tuinstra, T.T. Tesso, F. White, and J. Yu. 2012. Parallel domestication of the Shattering1 genes in cereals. Nat. Genet. 44:720–724.

Lind-Riehl, J.F., A.R. Sullivan, and O. Gailing. 2014. Evidence for selection on a CONSTANS-like gene between two red oak species. Ann. Bot. 113:967–975.

Onishi, K., K. Takagi, M. Kontani, T. Tanaka, and Y. Sano. 2007. Different patterns of genealogical relationships found in the two major QTLs causing reduction of seed shattering during rice domestication. Genome 50:757–766.

Paterson, A.H., Y.-R. Lin, Z. Li, K.F. Schertz, J.F. Doebley, S.R.M. Pinson, S.-C. Liu, J.W. Stansel, and J.E. Irvine. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269: 1714–1718.

Semeniuk, P. 1971. Inheritance of recurrent blooming in *Rosa wichuraiana*. J. Hered. 62:203–204.

Shubin, N.H., E.B. Daeschler, and F.A. Jenkins. 2006. The pectoral fin of *Tiktaalik roseae* and the origin of the tetrapod limb. Nature 440:764–771.

Tang, H., H.E. Cuevas, S. Das, U.U. Sezen, C. Zhou, H. Guo, V.H. Goff, Z. Ge, T.E. Clemente, and A.H. Paterson. 2013. Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. Proc. Natl. Acad. Sci. USA 110:15824–15829.

Vavilov, N.I. 1922. The law of homologous series in variation. J. Genet. 12:47–89.

Wu, F., B.W. Price, W. Haider, G. Seufferheld, R. Nelson, and Y. Hanzawa. 2014. Functional and evolutionary characterization of the CONSTANS gene family in short-day photoperiodic flowering in soybean. PLoS One 9:e85754.

Yang, S., B.D. Weers, D.T. Morishige, and J.E. Mullet. 2014. CONSTANS is a photoperiod regulated activator of flowering in sorghum. BMC Plant Biol. 14:148.