Identification of Woodland Strawberry Gene Coexpression Networks

Figure 1. The diploid woodland strawberry, *F. vesca*.

What we think of as a strawberry is botanically not a berry or even a fruit, but rather multiple fruits (achenes that contain the seeds) on the outside of a swollen receptacle. This technicality aside, strawberries are both economically important and a useful system in which to study seed-fruit communication. While cultivated strawberries have a complex octoploid genome, one of their likely progenitors, the woodland strawberry (*Fragaria vesca*; Fig. 1), is a rapidly growing model system for the Rosaceae family due to its short generation time and capacity to be transformed. A draft of the woodland strawberry diploid genome sequence was released in 2011 (Shulaev et al., 2011), and the recent publication of a high-quality genome based on PacBio sequencing has added almost 1,500 genes to the annotation (Edger et al., 2018). Genetic and epigenetic resources have also been developed for this species (Xu et al., 2016; Hilmarsson et al., 2017).

Given the unusual botanical nature of the strawberry fruit, the question arises of how developmental pathways have been co-opted or altered to result in a fleshy receptacle covered in achenes. A key approach in the study of strawberry fruit development has been the characterization of transcriptomes from various tissues and developmental stages to identify gene expression patterns or gene regulatory networks driving the transition between developmental stages. As an example of the power of this approach, a previous analysis of RNA-seq data from woodland strawberry has suggested that the endosperm and seed coat are key in producing the phytohormones auxin and gibberellin that coordinate the development of achenes and receptacle (Kang et al., 2013).

In this issue of *Plant Physiology*, the group of Zhongchi Liu at the University of Maryland identifies gene coexpression networks through integration and analysis of data from their three previous RNA-seq studies (Kang et al., 2013; Hollender et al., 2014; Hawkins et al., 2017). They used weighted gene coexpression network analysis to determine which genes have similar changes in expression across developmental stages and tissues, with the expectation that coexpressed genes may have a functional relationship. The authors used a number of variations on the weighted gene coexpression network analysis method to produce gene clusters of differing stringency. As all variations of the analysis are available through a Web interface (www.fv.rosaceaefruits.org), researchers can choose to query either many small gene clusters of high robustness or fewer large clusters of relatively low stringency. The association of gene clusters with specific tissues and developmental stages can be useful to confirm the results of previous studies. To illustrate this point, Shahan and colleagues show that core transcription factor regulators of meristem activity appear in a gene coexpression cluster that is strongly correlated with expression in the receptacle, thus confirming their previous finding that the receptacle has gained meristematic function (Hollender et al., 2014).

By studying gene ontology terms associated with genes in each cluster, new functions may be connected to specific developmental stages or tissues. Shahan and colleagues demonstrate the utility of this approach by identifying a high prevalence of iron transport and sequestration gene ontology terms associated with the genes in cluster 2. Since the genes in this cluster displayed high expression in the cortex, pith, and ghost (combined endosperm and seed coat) tissues, the authors hypothesized that the receptacle and ghost transport iron. Indeed, visualization of the concentration of free iron in the vasculature tissues between receptacle and seeds confirmed this hypothesis. Hence, the receptacle and ghost may actively transport iron to the embryo, where it acts as a cofactor for crucial enzymes that are necessary postfertilization.

[1]Author for contact: lisa.m.smith@sheffield.ac.uk.
[2]Senior author: lisa.m.smith@sheffield.ac.uk.
www.plantphysiol.org/cgi/doi/10.1104/pp.18.00880

Examination of the expression networks of duplicated genes can also be used to determine divergent or parallel functions. In Arabidopsis (*Arabidopsis thaliana*), the transcription factor LFY interacts with the F-box protein UFO to activate *AP3* gene expression, with the B class AP3 transcription factor then specifying petal and stamen identity (Lee et al., 1997; Chae et al., 2008). In woodland strawberry, there are three copies of *UFO*. *FveUFO1* appears to maintain its interaction with *FveLFY*, as is evidenced through correlation of their expression patterns. *FveUFO2* and *FveUFO3* are expressed later in flower development; therefore, the cognate proteins have likely diverged in function from FveUFO1. Exploring the *FveLFY-FveUFO1* coexpression data further, Shahan and colleagues suggest that FveLFY and FveUFO1 may not regulate class B genes, such as *FveAP3* in woodland strawberry, but may instead regulate class A and class E genes, including *FveAP1*, *FveAP2*, *FveSEP1*, and *FveSEP4*. The authors isolated a putative mutant in *FveUFO1*, which has a phenotype resembling those of selected mutants in homeotic genes belonging to classes A, B, and C. Chromatin immunoprecipitations using LFY antibodies will be needed to confirm which classes of transcription factors are directly regulated by FveLFY and therefore determine the FveLFY-FveUFO1 gene regulatory network.

Future development of the coexpression visualization platform by Shahan et al. (2018) will reveal the resilience of this method in comparing RNA-seq datasets from different laboratories and other woodland strawberry varieties. For example, integration of transcriptome and metabolome data from white versus red varieties, as presented by Härtl et al. (2017), would increase our knowledge about the robustness of coexpression networks and the variability in gene coexpression between different growth environments. Additionally, the parallel presentation of gene coexpression networks from other model fruits, such as tomato, would assist in identification of commonalities of gene functionality across broad evolutionary scales. Finally, the integration of gene coexpression networks with extant eFP browsers (e.g. Hawkins et al., 2017) would allow for streamlining of gene analysis. Taken together, the coexpression networks presented in this issue by Shahan and colleagues represent a significant step forward in the analysis of gene functionality in the fruits of woodland strawberry.

Lisa M. Smith[1,2]
Department of Animal and Plant Sciences and Plant Production and Protection (P3) Centre, University of Sheffield, Sheffield S10 2TN, United Kingdom
ORCID ID: 0000-0003-2364-8187

LITERATURE CITED

Chae E, Tan QK, Hill TA, Irish VF (2008) An Arabidopsis F-box protein acts as a transcriptional co-factor to regulate floral development. Development **135**: 1235–1245

Edger PP, VanBuren R, Colle M, Poorten TJ, Wai CM, Niederhuth CE, Alger EI, Ou S, Acharya CB, Wang J, (2018) Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (Fragaria vesca) with chromosome-scale contiguity. Gigascience **7**: 1–7

Härtl K, Denton A, Franz-Oberdorf K, Hoffmann T, Spornraft M, Usadel B, Schwab W (2017) Early metabolic and transcriptional variations in fruit of natural white-fruited Fragaria vesca genotypes. Sci Rep **7**: 45113

Hawkins C, Caruana J, Li J, Zawora C, Darwish O, Wu J, Alkharouf N, Liu Z (2017) An eFP browser for visualizing strawberry fruit and flower transcriptomes. Hortic Res **4**: 17029

Hilmarsson HS, Hytönen T, Isobe S, Göransson M, Toivainen T, Hallsson JH (2017) Population genetic analysis of a global collection of Fragaria vesca using microsatellite markers. PLoS One **12**: e0183384

Hollender CA, Kang C, Darwish O, Geretz A, Matthews BF, Slovin J, Alkharouf N, Liu Z (2014) Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks. Plant Physiol **165**: 1062–1075

Kang C, Darwish O, Geretz A, Shahan R, Alkharouf N, Liu Z (2013) Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry Fragaria vesca. Plant Cell **25**: 1960–1978

Lee I, Wolfe DS, Nilsson O, Weigel D (1997) A LEAFY co-regulator encoded by UNUSUAL FLORAL ORGANS. Curr Biol **7**: 95–104

Shahan R, Zawora C, Wight H, Sittmann J, Wang W, Mount SM, Liu Z (2018) Consensus coexpression network analysis identifies key regulators of flower and fruit development in wild strawberry. Plant Physiol **178**: 202–216

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, (2011) The genome of woodland strawberry (Fragaria vesca). Nat Genet **43**: 109–116

Xu J, Tanino KK, Robinson SJ (2016) Stable epigenetic variants selected from an induced hypomethylated *Fragaria vesca* population. Front Plant Sci **7**: 1768