UC Davis

UC Davis Previously Published Works

Title

Synthetic biology of metabolism: using natural variation to reverse engineer systems.

Permalink

<https://escholarship.org/uc/item/5j18p9g5>

Author

Kliebenstein, Daniel J

Publication Date

2014-06-01

DOI

10.1016/j.pbi.2014.03.008

Peer reviewed

Synthetic biology of metabolism: using natural variation to reverse engineer systems

Daniel J Kliebenstein^{1,2}

A goal of metabolic engineering is to take a plant and introduce new or modify existing pathways in a directed and predictable fashion. However, existing data does not provide the necessary level of information to allow for predictive models to be generated. One avenue to reverse engineer the necessary information is to study the genetic control of natural variation in plant primary and secondary metabolism. These studies are showing that any engineering model will have to incorporate information about 1000s of genes in both the nuclear and organellar genome to optimize the function of the introduced pathway. Further, these genes may interact in an unpredictable fashion complicating any engineering approach as it moves from the one or two gene manipulation to higher order stacking efforts. Finally, metabolic engineering may be influenced by a previously unrecognized potential for a plant to measure the metabolites within it. In combination, these observations from natural variation provide a beginning to help improve current efforts at metabolic engineering.

Addresses

¹ Department of Plant Sciences, University of California at Davis, Davis, CA 95616, USA

² DynaMo Center of Excellence, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark

Corresponding author: Kliebenstein, Daniel J (kliebenstein@ucdavis.edu)

Current Opinion in Plant Biology 2014, **19**:20–26

This review comes from a themed issue on **Physiology and metabolism**

Edited by **Sarah O'Connor** and **Tom Brutnell**

1369-5266/\$ – see front matter, © 2014 Elsevier Ltd. All rights reserved.

<http://dx.doi.org/10.1016/j.pbi.2014.03.008>

Introduction

A significant goal of synthetic biology or biological engineering is to take an existing organism and alter it in a directed and predictive fashion to introduce a new phenotype that had not previously existed in that organism. Some of these efforts start with introducing a new metabolic pathway into an organism to provide a better source of a commercially important chemical or to make new chemicals entirely [1–3]. Other efforts are focused at taking agronomically important traits such as defensive chemicals, nitrogen fixation or perennial growth from one species and transferring these traits into other species that

do not contain them [4^{••},5,6]. However these efforts while exciting frequently run into great difficulty and for this review article, I will focus on one the potential of using information from natural variation studies to facilitate these efforts for metabolic engineering.

To predictively engineer a new metabolic pathway into a plant with optimal efficiency requires a base level of knowledge that likely does not exist to any full level. To generate full predictive ability we would need much deeper understanding of the following questions. First, how many genes within the plant need to be manipulated to facilitate the integration of the new pathway into the organisms metabolic and regulatory network. How do these genes interact, are they solely additive or is there evidence of more complex epistasis that complicates predictive capacity? Finally, are these genes solely in the nuclear genome or is there causal variation also in the organellar genomes? Most efforts to answer these questions focus largely on forward or single-gene reverse genetics approaches which are highly time consuming and difficult to scale up to the necessary level.

An alternative approach to understanding a system is to use the concept of reverse engineering [7,8]. The goal of reverse engineering is to take a functioning system, say an existing metabolic pathway, and disassemble it to understand how it functions. The need however is to attempt this in a complex highly parallelized manner to decompose the system as much as possible. This requires a method to introduce a large number of perturbations into the system, mix these perturbations and measure the outputs. There are efforts to create massive numbers of mutations in individual lines but these are currently limited to single cell lineages without the creation of different mixes [9]. An alternative approach is to use the polymorphisms/perturbations in metabolism that occur within the diverse natural accessions of any organism. Then recombination and independent assortment provides the necessary mixture of these perturbations in combinations that can be studied using quantitative genetics approaches. Thus, it is possible to use populations containing natural diversity to provide some reverse engineering information for metabolic systems engineering.

This review article will discuss how natural variation studies can inform our attempts to engineer plant metabolic systems with a focus on the influence of variation in the genetic and metabolomic background. This will include a survey of natural variation research showing the

influence of genetic background on transgenic engineering attempts. It will then proceed through reverse engineering observations about the number of genes that influence the accumulation of endogenous metabolic pathways, the interaction of these links and where these links are located in the genome.

The role of genetic background in synthetic biology/engineering

There are several major lines of evidence that the genetic background plays a determinant role in controlling the success of any synthetic or forward engineering approach in an organism. First in almost every species tested, introduction of a specific transgene into different natural varieties led to significantly different results in each genotype [10–12]. This was even when using multiple independent events per background as with mice or *Drosophila* [10,11]. In mice, the background determined the epigenetic stability of expression from the transgene [10]. In White clover, a specific transgenic event was outcrossed into a polymorphic population and the background was shown to have quantitative effects leading to a range of function from the transgene [12]. Thus, the genetic background can impact specific transgenes via a multitude of mechanisms. Further, the white clover results shows that while different backgrounds introduce a difficulty in generating predictive function from a transgene, it may be possible to use breeding approaches to optimize a transgenes function. Unfortunately, most studies showing the effect of genetic background upon transgene function are likely not published due to their complexity. This absence of published data is a significant hindrance to developing an understanding of how the background interacts with an introduced transgene.

In addition to the background influencing the function of a specific transgene, there is evidence that transgenes can have differential effects on the organism depending upon the background. Within *Drosophila*, different backgrounds had different extensions in life-span when having the Cu/Zn superoxide dismutase introduction [11]. Similarly, the overexpression of a secondary metabolite specific transcription factor in *Arabidopsis* had dramatically different effects depending upon the genetic background. When *MYB28* was introduced into the Col-0 background, it was possible to obtain up to 10× increases in the content of the endogenous glucosinolates without showing significant impact on growth or development of the plant [13,14]. In contrast, the introduction of a similar *MYB28* construct into the Landsberg *erecta* (*Ler*) background led to maximally a 3× induction of glucosinolates and always concomitant with a dramatic stunting of the plants growth and development [15]. This stunting was associated with significant pleiotropic impacts on primary metabolism [15]. This difference in genotypes response to the transgenes was occurring even though

Ler and Col-0 start off with similar glucosinolate content and the Col-0 transgenic lines accumulated higher glucosinolate [13–17]. This argues that the stunting is not a simple response to the cost of producing more glucosinolate. If this was the case, then the Col-0 lines would have been stunted. While there are numerous polymorphisms controlling glucosinolates and their regulation between *Ler* and Col-0, it remains to be tested if any of these influence this differential growth response to the introduction of *MYB28* [16,18]. This variation does highlight the significant potential impact of genetic background on the ability of engineering specific pathways in plants.

Another source of information on the ability of genetic background to affect synthetic biology is the large body of literature on introducing synthetic pathways into new organisms. Over the past decade numerous biochemical pathways have been moved into new organisms that had previously not contained this pathway [1,19–21]. A standard part of the protocol for optimizing the expression of the resulting metabolite is to conduct a round of mutagenesis with selection for higher accumulating lines in both bacteria and fungi [1,19–21]. The goal of this mutagenesis step is to randomly engineer the genetic background and use selection to identify the lines with better backgrounds. The ability of background modification to increase synthetic pathway efficiency is so effective as to have driven the development of new approaches to speed up this process and link it more directly to the pathway [22,23]. To date, this approach has been left to random mutation and selection with minimal identification of the background changes to test if there are methods to predict *a priori* how the background should be manipulated.

In addition to the synthetic introduction of new pathways, the selective pressures of *de novo* pathway generation during evolution also support the role of background changes in increasing pathway efficiency [24]. The generation of new pathways within species by evolution and selection requires the manipulation of genes to create new enzymes as well as new developmental structures to house these pathways via a variety of mechanisms [24–26]. An analysis of genetic variation within *Arabidopsis* showed that there was a high level of genetic diversity within the glucosinolate secondary metabolite pathway which arose within the Brassicales [27,28]. This secondary metabolite pathway diversity was also linked to elevated diversity in the genes of the sulfur associated pathways [27]. The glucosinolates are a heavy draw on the sulfur pathways. Thus, this linkage suggests that the introduction of the glucosinolate pathway imparted a selective pressure to also alter the background sulfur assimilation and biosynthetic pathways to optimize the secondary metabolite pathway. Thus, natural evolution appears to also manipulate the genetic background when new pathways are evolved.

Genome wide association mapping to reverse engineer the number of gene connections

A key aspect of introducing a synthetic pathway into a new organism is how many background genes need to be coordinated to properly integrate the new pathway into the organism's regulatory and metabolic architecture. Natural variation has been used to characterize and clone a number of secondary metabolite pathways within plants [16,29^{*},30–32]. More recently the study of natural variation is providing insight into how many connections occur between metabolic pathways and physiological or regulatory processes. These observations are coming from the use of genome wide association (GWA) mapping which utilizes the standing variation present within a natural population. Measuring phenotypes in this same population then allows the researcher to find putative genes where genetic variation correlates with phenotypic variation [33–36].

GWA studies have recently begun to be applied to investigating the genetic variation underlying phenotypic variation in both primary and secondary metabolism [37^{*},38,39,40^{*}]. An GWA study of genetic variation in the glucosinolate secondary metabolite pathway present in *Arabidopsis thaliana* showed that it was possible to identify causal genes with >70% accuracy using a combination of GWA and transcriptomics [37^{*},39]. More importantly for systems engineering, this provided a baseline estimate that >1000 genes likely have the ability to influence glucosinolate content at levels that were ecologically and likely economically relevant [37^{*},39]. These 1000+ genes were clustered into networks that centered on known (ABA signaling, blue light signaling, herbivory resistance, etc.) and unknown physiological processes. Similar results were found for primary metabolic networks suggesting that this level of connectedness between a metabolic pathway and the plants physiological network is true across most pathways [38]. Thus, while the number of genes to consider is daunting, a solution may be to consider the physiological processes linked to a metabolic pathway as the basis of engineering rather than treating the genes as independent units.

Quantitative trait locus mapping to reverse engineer the shape of metabolic connections

With metabolic pathways showing a high level of connectivity to numerous physiological and regulatory processes, this raises the potential for interactions in genes across these pathways and processes [41]. However, little is known about the style of these interactions, are they largely additive suggesting that they can be readily modeled or predicted. Or alternatively do they show some form of epistasis that complicates modeling and prediction. Comparing epistasis versus additive interactions is largely impossible in GWA studies do to the non-random structure of the populations and instead requires more randomly structured population like recombinant inbred

lines [38,42,43]. Analysis of secondary metabolite pathways using RILs has shown a high level of epistatic interactions amongst causal genes [29^{*},44,45]. This includes genes typically thought of as peripheral to the pathway like those involved in the core circadian clock [46,47].

For primary metabolism, the interaction of genes is less clear. A number of metabolomics studies have been done on RILs [29^{*},32,48,49,50^{**},51]. Some studies find high levels of epistasis while others find mainly additive interactions. One complication to this analysis is that the studies finding mainly additivity typically involve fewer lines than were used in the studies that found epistasis. This difference in power between the studies is more critical because the epistasis typically found for primary metabolism is not pairwise interactions of loci but instead higher-order interactions involving three or more loci leading to large allele specific effects [50^{**},51]. The observation of multi-locus epistasis in metabolism is supported by other observations but more work on larger biparental populations needs to be conducted to test if these gene interactions are additive or higher-order epistasis [39,52,53].

The ultimate goal of testing if the 1000+ causal genes is to help guide efforts at computational modeling to improve our power to predictively engineer plant metabolism. Quantitative trait locus mapping with human insight has enabled the characterization and prediction of numerous pathways. And this has lead initiatives to create an automated algorithm which takes the metabolite and genetic data to predict metabolic pathways [45,54–58]. However, a direct comparison of computational predictions to empirical pathway evidence showed the none of the current algorithms are sufficient and new approaches need to be developed to enable this automated pathway discovery [59].

What is the genomic background that needs manipulating?

When considering what parts of the genetic background need to be engineered to optimize the expression of a pathway the focus largely centers upon the genes located within the nuclear genome of a eukaryote. This is partly for technical reasons in that it is harder to manipulate genes within the mitochondria or plastid. However there is also an issue of ascertainment as most mechanistic studies within model systems do not focus on mechanisms encoded within the organellar genomes. This is in contrast to the ecological and evolutionary importance that genetic variation within the organellar genetic background can play [60]. Recent analysis of metabolomic variation within a population segregating for both nuclear and organellar genetic variation showed that the genetic variation within the organelle genome in *A. thaliana* can cause metabolic variation of a magnitude at least as large

if not typically larger than any individual nuclear locus [50^{••}]. This was true for both primary and secondary metabolism [61]. Interestingly, for primary metabolism a large component of the total variation in their accumulation was due to epistatic interactions between genes in the nucleus and organelle, cytonuclear epistasis [50^{••}]. The mechanistic basis of these interactions remains to be determined. These observations do suggest that when engineering new metabolic pathways into plants it is equally important to consider the role of variation in both the nuclear and organellar genetic backgrounds. This cytonuclear epistasis also suggests that placing pathways into the organelle does not insulate them from variation in the nuclear background.

What about the engineered metabolite?

When engineering a new metabolite into a plant or modifying a metabolite already existent in the plant, the effort typically assumes, consciously or subconsciously, that the metabolite itself is relatively inert beyond the cost of its production and storage. In some instances, the potential for toxicity is considered [5,62–64]. However, vary rarely is the potential for the organism to directly perceive the engineered metabolite and cause regulatory disruption of metabolic network considered. This is in spite of the significant and broad evidence of chemicals to be sensed and cause both transcriptional and post-transcriptional regulatory changes. Studies of quantitative trait loci within *A. thaliana* that influence the circadian clock has recently lead to the observation that genetic polymorphisms within the glucosinolate secondary metabolite pathway can alter the oscillatory behavior of the circadian clock [46]. The study showed that this link was not a pleiotropic result of altered sulfur assimilation affecting the clock but instead a potential direct effect of one of the glucosinolate metabolites. Other studies have also come across evidence that glucosinolates might have regulatory properties within *Arabidopsis* and other plants [47,65,66]. Further, the rapid explosion of new regulatory metabolites for diverse processes within plants suggests that we should consider the likely potential that any chemical introduced into a plant may be sensed and cause regulatory changes within the plant itself [67–69,70^{••}]. Thus, part of any manipulation of the genetic background may have to involve removing the plants perception of the introduced compound or its precursors.

Conclusions and future perspectives

Future efforts at engineering of plant metabolism via synthetic biology will have to begin moving beyond current models of metabolic regulation and toward more integrative models involving vastly more links than had been previously considered (Figure 1). This will likely go a long ways toward developing more predictive engineering efforts. This will however require new approaches to simultaneously modify dozens to 100s of background

Figure 1

Reverse engineering insights from natural variation. The abbreviations are as follows: JA – jasmonic acid signaling, SA – salicylic acid signaling, ABA – abscisic acid, EnzA – Enzyme A, TF – transcription factor. The Blue hexagon to the right shows a precursor metabolite while the red shows the product metabolite produced via the enzymatic activity of Enzyme A. (a) Shown is a simplistic model pathway with a single enzyme converting a blue to a red metabolite. Single linear inputs with two transcription factors and enzymatic feedback by the red metabolic product are shown as in solid lines to represent a generic base model of metabolic regulation. In this example, the model is for a wound or herbivory responsive enzyme. (b) Extensions of this base model suggested by reverse engineering insights obtained from natural variation include the role of the organellar genome, the extension of the number of regulatory inputs and the potential for metabolic to transcription connections. New potential regulatory connections suggested by natural variation for this herbivory responsive enzyme are presented in solid lines. The generic base regulatory connections from the model in A are in dashed lines. TFC is shown in dotted fill to indicate that it has natural variation for the presence or absence of the TF. This presence or absence variation would lead to differential ABA inputs into the regulation of Enzyme A across the genotypes.

genes in both the nucleus and organelles within multicellular plants. A broad survey to identify how many metabolites can be sensed for transcriptional or other regulatory responses in a plant is a key component to develop any predictive engineering system. Additionally, this will require new approaches to measure and develop a theoretical framework for how polymorphisms in this number of genes can function together to generate a diverse array of output metabolite phenotypes. Current models rarely progress beyond 10 genes while empirical assessment of gene interactions rarely go above five genes. The effort of developing cohesive models of how a full multi-cellular eukaryotic system integrates to coordinate a single phenotype is an important avenue forward that will greatly illuminate both the study of natural variation as well as to reverse engineer a more predictive synthetic biology framework.

Acknowledgements

I apologize to all the authors whose work relates to the specific topics discussed in this review but whom I was unable to cite because of space limitations. This work was supported by the US National Science Foundation (DBI Award 0820580 and MCB Award 1330337) and by the Danish National Research Foundation (DNRF99).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J *et al.*: **Production of the antimalarial drug precursor artemisinic acid in engineered yeast.** *Nature* 2006, **440**:940-943.
2. Glenn WS, Nims E, O'Connor SE: **Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor.** *J Am Chem Soc* 2011, **133**:19346-19349.
3. Rungtaphan W, Qu X, O'Connor SE: **Integrating carbon-halogen bond formation into medicinal plant metabolism.** *Nature* 2010, **468**:461-464.
4. Moldrup ME, Geu-Flores F, de Vos M, Olsen CE, Sun J, Jander G, Halkier BA: **Engineering of benzylglucosinolate in tobacco provides proof-of-concept for dead-end trap crops genetically modified to attract *Plutella xylostella* (diamondback moth).** *Plant Biotechnol J* 2012, **10**:435-442.
- One of the first instances of engineering a multi-step biosynthetic pathway into a new organism and showing its potential agronomic application to insect control.
5. Pfalz M, Mikkelsen MD, Bednarek P, Olsen CE, Halkier BA, Kroymann J: **Metabolic engineering in *Nicotiana benthamiana* reveals key enzyme functions in *Arabidopsis* indole glucosinolate modification.** *Plant Cell* 2011, **23**:716-729.
6. Vance CP: **Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources.** *Plant Physiol* 2001, **127**:390-397.
7. Chikofsky EJ, Cross JH: **Reverse engineering and design recovery—a taxonomy.** *IEEE Softw* 1990, **7**:13-17.
8. Eilam E: *Reversing: Secrets of Reverse Engineering.* Wiley Publishing; 2005.
9. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: **RNA-guided human genome engineering via Cas9.** *Science* 2013, **339**:823-826.
10. Allen ND, Norris ML, Surani MA: **Epigenetic control of transgene expression and imprinting by genotype-specific modifiers.** *Cell* 1990, **61**:853-861.
11. Orr WC, Mockett RJ, Benes JJ, Sohal RS: **Effects of overexpression of copper-zinc and manganese superoxide dismutases, catalase, and thioredoxin reductase genes on longevity in *Drosophila melanogaster*.** *J Biol Chem* 2003, **278**:26418-26422.
12. Scott A, Woodfield D, White DWR: **Allelic composition and genetic background effects on transgene expression and inheritance in white clover.** *Mol Breed* 1998, **4**:479-490.
13. Sønderby IE, Burow M, Rowe HC, Kliebenstein DJ, Halkier BA: **A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in *Arabidopsis*.** *Plant Physiol* 2010, **153**:348-363.
14. Sønderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ: **A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates.** *PLoS ONE* 2007, **2**:e1322.
15. Malitsky S, Blum E, Less H, Venger I, Elbaz M, Morin S, Eshed Y, Aharoni A: **The transcript and metabolite networks affected by the two clades of *Arabidopsis* glucosinolate biosynthesis regulators.** *Plant Physiol* 2008, **148**:2021-2049.
16. Kliebenstein DJ, Gershenzon J, Mitchell-Olds T: **Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds.** *Genetics* 2001, **159**:359-370.
17. Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T: **Genetic control of natural variation in *Arabidopsis thaliana* glucosinolate accumulation.** *Plant Physiol* 2001, **126**:811-825.
18. Kliebenstein DJ, Figuth A, Mitchell-Olds T: **Genetic architecture of plastic methyl jasmonate responses in *Arabidopsis thaliana*.** *Genetics* 2002, **161**:1685-1696.
19. Alonso-Gutierrez J, Chan R, Batth TS, Adams PD, Keasling JD, Petzold CJ, Lee TS: **Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production.** *Metab Eng* 2013, **19**:33-41.
20. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD: **Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids.** *Nat Biotechnol* 2003, **21**:796-802.
21. Steen EJ, Chan R, Prasad N, Myers S, Petzold CJ, Redding A, Ouellet M, Keasling JD: **Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol.** *Microb Cell Fact* 2008, **7**.
22. Chou HH, Keasling JD: **Programming adaptive control to evolve increased metabolite production.** *Nat Commun* 2013, **4**.
23. Woo HM, Murray GW, Batth TS, Prasad N, Adams PD, Keasling JD, Petzold CJ, Lee TS: **Application of targeted proteomics and biological parts assembly in *E. coli* to optimize the biosynthesis of an anti-malarial drug precursor, amorpha-4,11-diene.** *Chem Eng Sci* 2013, **103**:21-28.
24. Osbourn AE, Kliebenstein DJ: **Making new molecules—evolution of pathways for novel metabolites in plants.** *Curr Opin Plant Biol* 2012, **15**:415-423.
25. Kliebenstein DJ: **Making new molecules—evolution of structures for novel metabolites in plants.** *Curr Opin Plant Biol* 2013, **16**:112-117.
26. Kliebenstein DJ: **New synthesis-regulatory evolution, the veiled world of chemical diversification.** *J Chem Ecol* 2013, **39**:349.
27. Kliebenstein DJ: **A role for gene duplication and natural variation of gene expression in the evolution of metabolism.** *PLoS ONE* 2008, **3**:e1838.
28. Kliebenstein DJ, West MAL, Van Leeuwen H, Kyunga K, Doerge RW, Michelmore RW, Clair StDA: **Genomic survey of gene expression diversity in *Arabidopsis thaliana*.** *Genetics* 2006, **172**:1179-1189.
29. Gong L, Chen W, Gao Y, Liu X, Zhang H, Xu C, Yu S, Zhang Q, Luo J: **Genetic analysis of the metabolome exemplified using a rice population.** *Proc Natl Acad Sci USA* 2013, **110**:20320-20325.
- A paper showing that the use of epistasis and metabolomics combined can allow for the inference of metabolic pathways in agronomic model systems.
30. Kliebenstein D, Lambrix V, Reichelt M, Gershenzon J, Mitchell-Olds T: **Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in *Arabidopsis thaliana*.** *Plant Cell* 2001, **13**:681-693.
31. Kliebenstein DJ: **A quantitative genetics and ecological model system: understanding the aliphatic glucosinolate biosynthetic network via QTLs.** *Phytochem Rev* 2009, **8**:243-254.
32. Keurentjes JJB, Fu JY, de Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M: **The genetics of plant metabolism.** *Nat Genet* 2006, **38**:842-849.
33. Nemri A, Atwell S, Tarone AM, Huang YS, Zhao K, Studholme DJ, Nordborg M, Jones JDG: **Genome-wide survey of *Arabidopsis* natural variation in downy mildew resistance using combined association and linkage mapping.** *Proc Natl Acad Sci U S A* 2010, **107**:10302-10307.

34. Atwell S, Huang Y, Vilhjalmsón BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone A, Hu TT *et al.*: **Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* in-bred lines.** *Nature* 2010, **465**:627–631.
 35. Seren U, Vilhjalmsón BJ, Horton MW, Meng D, Forai P, Huang YS, Long Q, Segura V, Nordborg M: **GWAPP: a web application for genome-wide association mapping in *Arabidopsis*.** *Plant Cell* 2012, **24**:4793–4805.
 36. Nordborg M, Weigel D: **Next-generation genetics in plants.** *Nature* 2008, **456**:720–723.
 37. Chan EK, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ:
 - **Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*.** *PLoS Biol* 2011, **9**:e1001125.

A paper showing how to identify causal genes from GWA data with a high success rate. The results implicate the role of 100s to 1000s of genes in controlling metabolite accumulation of the glucosinolate defense compounds.
 38. Chan EKF, Rowe HC, Kliebenstein DJ: **Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping.** *Genetics* 2010, **185**:991–1007.
 39. Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ: **The complex genetic architecture of the metabolome.** *PLoS Genet* 2010, **6**:e1001198.
 40. Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE:
 - **Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize.** *Proc Natl Acad Sci U S A* 2012, **109**:8872–8877.

A paper on GWA in maize metabolomics that provides a counter image to that found in *Arabidopsis* with large effect loci controlling central metabolites. This may potentially arise from the influence of domestication.
 41. Segre D, DeLuna A, Church GM, Kishony R: **Modular epistasis in yeast metabolism.** *Nat Genet* 2005, **37**:77–83.
 42. Platt A, Vilhjalmsón BJ, Nordborg M: **Conditions under which genome-wide association studies will be positively misleading.** *Genetics* 2010, **186**:1045–1052.
 43. Mackay TFC: **The genetic architecture of quantitative traits.** *Annu Rev Genet* 2001, **35**:303–339.
 44. Byrne PF, McMullen MD, Wiseman BR, Snook ME, Musket TA, Theuri JM, Widstrom NW, Coe EH: **Maize silk maysin concentration and corn earworm antibiosis: QTLs and genetic mechanisms.** *Crop Sci* 1998, **38**:461–471.
 45. Giamoustaris A, Mithen R: **Genetics of aliphatic glucosinolates. IV. Side-chain modification in *Brassica oleracea*.** *Theor Appl Genet* 1996, **93**:1006–1010.
 46. Kerwin RE, Jiménez-Gómez JM, Fulop D, Harmer SL, Maloof JN, Kliebenstein DJ: **Network quantitative trait loci mapping of circadian clock outputs identifies metabolic pathway-to-clock linkages in *Arabidopsis*.** *Plant Cell* 2011, **23**:471–485.
 47. Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ: **Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways.** *PLoS Genet* 2007, **3**:1687–1701.
 48. Sulpice R, Pyl ET, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC *et al.*: **Starch as a major integrator in the regulation of plant growth.** *Proc Natl Acad Sci U S A* 2009, **106**:10348–10353.
 49. Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Torjek O, Selbig J, Altmann T *et al.*: **Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations.** *Plant J* 2008, **53**:960–972.
 50. Joseph B, Corwin JA, Li B, Atwell S, Kliebenstein DJ: **Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome.** *eLife* 2013, **2**:e00776.
- The authors show that the genetic variation controlling the accumulation of primary metabolites within *Arabidopsis* is conditioned genetic variation present within the organellar genomes, either mitochondrial or plastidic.
51. Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ: **Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome.** *Plant Cell* 2008, **20**:1199–1216.
 52. Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR: **Mode of inheritance of primary metabolic traits in tomato.** *Plant Cell* 2008, **20**:509–523.
 53. Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J *et al.*: **Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement.** *Nat Biotechnol* 2006, **24**:447–454.
 54. Keurentjes JJB, Fu JY, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC: **Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci.** *Proc Natl Acad Sci U S A* 2007, **104**:1708–1713.
 55. Keurentjes JJB, Molenaar J, Zwaan BJ: **Predictive modelling of complex agronomic and biological systems.** *Plant Cell Environ* 2013, **36**:1700–1710.
 56. Fu J, Swertz MA, Keurentjes JJB, Jansen RC: **MetaNetwork: a computational protocol for the genetic study of metabolic networks.** *Nat Protoc* 2007, **2**:685–694.
 57. Mithen R, Campos H: **Genetic variation of aliphatic glucosinolates in *Arabidopsis thaliana* and prospects for map based gene cloning.** *Entomol Exp Appl* 1996, **80**:202–205.
 58. Mithen R, Clarke J, Lister C, Dean C: **Genetics of aliphatic glucosinolates. III. Side-chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*.** *Heredity* 1995, **74**:210–215.
 59. Hageman RB, Kliebenstein DJ, Churchill GA: **What can causal networks tell us about metabolic pathways?** *PLoS Comput Biol* 2012, **8**.
 60. Greiner S, Bock R: **Tuning a menage a trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants.** *Bioessays* 2013, **35**:354–365.
 61. Joseph B, Corwin JA, Zuest T, Li B, Iravani M, Schaepman-Strub G, Turnbull LA, Kliebenstein DJ: **Hierarchical nuclear and cytoplasmic genetic architectures for plant growth and defense within *Arabidopsis*.** *Plant Cell* 2013, **25**:1929–1945.
 62. Morant AV, Jorgensen K, Jorgensen B, Dam W, Olsen CE, Moller BL, Bak S: **Lessons learned from metabolic engineering of cyanogenic glucosides.** *Metabolomics* 2007, **3**:383–398.
 63. Kristensen C, Morant M, Olsen CE, Ekstrom CT, Galbraith DW, Moller BL, Bak S: **Metabolic engineering of dhurrin in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome.** *Proc Natl Acad Sci U S A* 2005, **102**:1779–1784.
 64. Geu-Flores F, Nielsen MT, Nafisi M, Moldrup ME, Olsen CE, Motawia MS, Halkier BA: **Glucosinolate engineering identifies gamma-glutamyl peptidase.** *Nat Chem Biol* 2009, **5**:575–577.
 65. Clay NK, Adio AM, Denoux C, Jander G, Ausubel FM: **Glucosinolate metabolites required for an *Arabidopsis* innate immune response.** *Science* 2009, **323**:95–101.
 66. Hasegawa T, Yamada K, Kosemura S, Yamamura S, Hasegawa K: **Phototropic stimulation induces the conversion of glucosinolate to phototropism-regulating substances of radish hypocotyls.** *Phytochemistry* 2000, **54**:275–279.
 67. Xiao Y, Savchenko T, Baidoo EEK, Chehab WE, Hayden DM, Tolstikov V, Corwin JA, Kliebenstein DJ, Keasling JD, Dehesh K: **Retrograde signaling by the plastidial metabolite MECPP**

- regulates expression of nuclear stress-response genes. *Cell* 2012, **149**:1525-1535.
68. Liu P-P, von Dahl CC, Park S-W, Klessig DF: **Interconnection between methyl salicylate and lipid-based long-distance signaling during the development of systemic acquired resistance in Arabidopsis and Tobacco.** *Plant Physiol* 2011, **155**:1762-1768.
 69. Chanda B, Xia Y, Mandal MK, Yu K, Sekine K-T, Gao Q-m, Selote D, Hu Y, Stromberg A, Navarre D *et al.*: **Glycerol-3-phosphate is a critical mobile inducer of systemic immunity in plants.** *Nat Genet* 2011, **43**:421-427.
 70. Finkemeier I, Koenig A-C, Heard W, Nunes-Nesi A, Phuong Anh P, Leister D, Fernie AR, Sweetlove LJ: **Transcriptomic analysis of the role of carboxylic acids in metabolite signaling in Arabidopsis leaves.** *Plant Physiol* 2013, **162**:239-253.
- The authors begin to provide evidence that metabolites in primary metabolism may provide information to the cell to alter transcriptional regulation.