Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Considerations when choosing a genetic model organism for metabolomics studies

Laura K Reed [1], Charles F Baer[2] and Arthur S Edison[3]

Model organisms are important in many areas of chemical biology. In metabolomics, model organisms can provide excellent samples for methods development as well as the foundation of comparative phylometabolomics, which will become possible as metabolomics applications expand. Comparative studies of conserved and unique metabolic pathways will help in the annotation of metabolites as well as provide important new targets of investigation in biology and biomedicine. However, most chemical biologists are not familiar with genetics, which needs to be considered when choosing a model organism. In this review we summarize the strengths and weaknesses of several genetic systems, including natural isolates, recombinant inbred lines, and genetic mutations. We also discuss methods to detect targets of selection on the metabolome.

Addresses
[1] Department of Biological Sciences, University of Alabama, 300 Hackberry Lane, Tuscaloosa, AL 35487, USA
[2] Department of Biology, University of Florida Genetics Institute, University of Florida, 621 Bartram Hall, 876 Newell Dr., Gainesville, FL 32611-8525, USA
[3] Departments of Genetics and Biochemistry, Institute of Bioinformatics and Complex Carbohydrate Center, University of Georgia, 315 Riverbend Rd, Athens, GA 30602, USA

Corresponding author: Edison, Arthur S (aedison@uga.edu)

Current Opinion in Chemical Biology 2017, **36**:7–14

This review comes from a themed issue on **Omics**

Edited by **Georg Pohnert** and **Frank C Schroeder**

http://dx.doi.org/10.1016/j.cbpa.2016.12.005

Introduction

Metabolomics is an essential component of systems biology and is at the forefront of precision medicine [1]. The process by which genetic information (DNA) is transmitted into a measureable phenotype (e.g. disease) works through a cascade of physiological levels, each of which affects the ultimate phenotype (Figure 1). Like all physiological levels, the metabolome is highly influenced by genetic, environmental, and genotype-by-environment interactions [2,3••,4••,5]; however, the proximity of the metabolome to the ultimate phenotype makes it an especially good measure of an organism's physiology [1]. Metabolomics is still in its early stages and has many unsolved challenges, such as chemically identifying metabolites [6]. The chemical diversity of the metabolome — which includes both endogenous and environmental sources such as food, microbes, and toxins — is vast [7], and the concentration range of metabolites spans about 12 orders of magnitude. Presently, no single technology is capable of a complete identification and quantification of the metabolome.

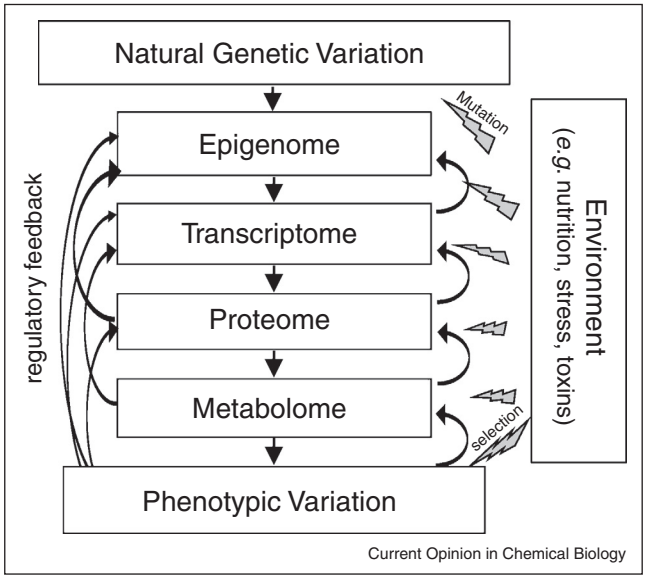
Model organisms were important for completing the human genome. They were used for developing new DNA sequencing technologies, data analysis techniques, and overall project organization [8]. Model organisms also have an important role in metabolomics in technique developments (e.g. [9–11]). And just as the annotation of genomes has been accelerated by comparative phylogenomics [12•], similar advances will undoubtedly come through model organisms and 'phylometabolomics' [13]. An interesting recent example of the power of phylometabolomics was the finding that human subjects suffering from chronic fatigue syndrome have metabolomic profiles similar to the dauer phenotype in *Caenorhabditis elegans* [14••]. Also, by taking an evolutionary approach, we will identify metabolites most critical to survival by determining which metabolites are experiencing selection.

Many choices are possible in metabolomics studies with model organisms, including the organism itself and the genetic system. The choice must be driven by the scientific question. Here we provide an overview of the strengths and limitations of different types of organisms and genetic systems, especially when the goal is to associate variation in the metabolome with specific factors within the genome (e.g. genes). We focus on potential metabolomics applications in model organisms that we suspect will soon have broad utility. Our areas of expertise are flies (*Drosophila melanogaster*) and worms (*C. elegans*), so these receive most of the emphasis, but we try to highlight important aspects of other systems.

The purpose(s) of model organisms

Model organisms serve two purposes in biology. First, they are employed as a proxy for a specific organism when it is impractical to address a question in that organism. In the case of biomedical research, the organism of interest is usually humans. For genetic questions, humans are difficult to study because, firstly, one cannot study a specific

Figure 1



Sources of interesting biological variation that contribute to observable phenotypes. The components of a system in a population of organisms include: genetic variation (hard coded genetic information in DNA), the epigenome (controls on how and when the genetic information is expressed), the transcriptome (all of the RNA gene transcripts), the proteome (the complete set of proteins translated and subsequently modified from the RNA transcripts), the metabolome (all the small molecule chemical compounds from both endogenous and exogenous sources). To understand complex traits, including disease, ultimately we have to understand each 'omics' level, how it evolves, and the mechanisms by which it can be perturbed. Phenotypic variation derives not only from genetic variation but also from the environment that can introduce both predictable and random perturbation (lightning bolts) of the physiological system. Metabolic homeostasis is achieved through interactions between different physiological or 'omics' levels within an organism, including regulatory feedback (curved arrows) and the fitness effects on the evolutionary genetics of the species. Each physiological level is likely to have a distinct reaction to environmental perturbation and thus needs to be understood as part of a larger system to predict the phenotypic consequence.

gene while controlling the rest of the organism's genetic background and secondly, it is extremely difficult to unambiguously separate the effects of genes from effects of the environment. Interactions between genes within a genome (epistasis) and genotype-by-environment interactions are unavoidable, thus carefully constructed experimental systems must be used to control for these non-linear effects. In model systems, replicates of specific genotypes can by systematically tested across different environments to characterize how certain genes interact with the environment through the metabolome (Figure 2).

Second, model organisms serve as a proxy for living organisms in general since a large number of scientists working on the same organism will make more progress toward understanding the fundamental properties of biology than will the same number of scientists working on a diversity of organisms. For that reason, many biologists have converged on a small number of organisms to serve as models — especially mouse and rat (Mammal), Xenopus and Zebrafish (Vertebrate), *D. melanogaster* and *C. elegans* (animal), *Arabidopsis thaliana* (plant), *Saccharomyces cerevisiae* (Eukaryote), *Escherichia coli* (Eubacteria, Prokaryote). However, there are also many questions in biology for which a model organism may not be appropriate, such as exploring the metabolic mechanisms of ecological specialization, identifying novel natural products, or evolution of development; such questions require the use of specific organisms (e.g. [15–20]).

Choosing a model organism and genetic system

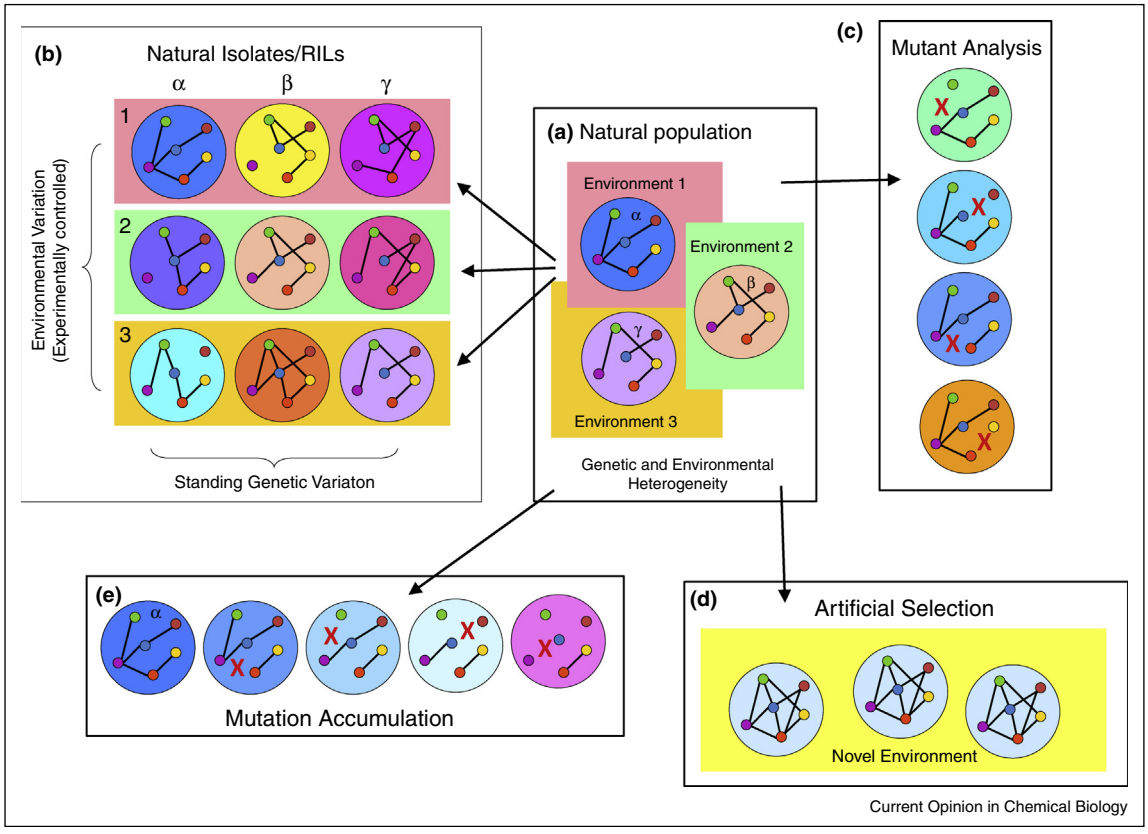
A tradeoff in the choice of a model organism is whether it is critical to be able to attribute the phenotype in question to an individual or a population. If the former, systems in which metabolomics of individual organisms is not yet experimentally tractable (e.g. yeast, *C. elegans*) are not appropriate. The choice of organism should also reflect the level of the research question. For example, if the metabolic pathway of interest is evolutionarily conserved across all eukaryotes then yeast might be the most expedient model. However, if you need to know how metabolic processes are coordinated between two distinct vertebrate tissues then zebrafish or mice would be appropriate.

There are several genetic approaches that can be used to associate metabolites with genes, and they each have specific advantages and disadvantages (Table 1). We know the metabolome can vary greatly among genotypes within a species as well as across environments [2,3••,4••,5], and using genetically characterized models can assist in identifying the differences in the metabolome due to genetic and environmental effects (Figure 2). Many model organisms can be genetically engineered to test specific mechanistic questions, and panels of lines bearing mutations distributed across the model's genome are easily obtained (e.g. [21–25]). Other resources can capture natural genetic variation to map genetic effects on the metabolome. It is our belief that systematic comparisons between genes and metabolites for many organisms will not only enhance our understanding of evolution and biology but will also be a very powerful aid in the identification of unknown metabolites.

Natural isolates

Medical genetics is essentially a study of natural human isolates: we are outbred and our environments are not controlled. Although most work with model organisms is done on a few standard stocks and the mutants derived from them (especially when the candidate genetic pathway is known *a priori*), it is often useful to characterize the genetic variation in natural populations. Wild isolates must acclimate to the lab for several generations to

Figure 2



The genetic basis for variation in the metabolome can be identified by several strategies. The metabolome is symbolized here by small colored nodes for metabolites with edges between the nodes indicating a gene function that links two metabolites (e.g. an enzyme). The color of the larger circles containing the metabolome represents the phenotype of the organism. **(a)** A natural population exhibits inter-individual variation in the structure of their metabolomes and resulting phenotypes due to a combination genotype (α, β, or γ) and environmental (1, 2, or 3) effects. **(b)** The relative contribution of genotype and environment to metabolome variation can be identified by systematically testing the distinct genotypes (α, β, and γ) across each of the three environments. This can be a powerful way to map the genes contributing to metabolome variation in natural isolates and recombinant inbred lines (RILs). **(c)** Systematic mutation of the genes (red X) contributing to the metabolome can identify which are most critical to maintaining the organism's phenotype when compared to the wildtype (α) genetic control (e.g. orange vs. blue phenotype). **(d)** Artificial selection exposes a genetically variable population to a selective force such as a novel environment that selects for a phenotype and underlying metabolome controlled by particular genes. By analyzing how the metabolome network adapts one can identify especially important components (e.g. new connection between the blue and orange metabolites). **(e)** Mutation accumulation (MA) experiments allow a wildtype progenitor genotype (α) to progressively acquire mutations that are largely deleterious. The comparison between the wildtype and MA lines allows for the identification of especially critical genetic controls and the overall robustness of the metabolome (e.g. blue vs. pink phenotypes).

remove residual transgenerational environmental effects carried from the field. A large panel of wild isolates will capture a much broader spectrum of genetic variation than exists in standard lab stocks, which suffer from small population size and adaption to the lab environment. Most model organisms are widespread in nature and can be easily collected, making it feasible to obtain large panels of wild isolates to map with high resolution, sometimes even finding the causal genetic variant. Also, each genetic isolate can be systematically tested across different environments to estimate the relative contributions of genetics and the environment to metabolome variation (Figure 2b). Many model organisms have extensive collections of natural isolates [26–30,31•,32]. In Drosophila, the *Drosophila* Genetic Reference Panel (205 lines) represents genetic diversity within one wild population [27,30], while the Global Diversity Lines include 84 isolates from five continents [33]. Global diversity is also represented in the RegMap panel in Arabidopsis (1307 isolates), in *C. elegans* (200 lines), in *E. coli* (202 strains) and yeast (36 *S. cerevisiae* and 35 *S. paradoxus* lines) [26,28,31•,32,34].

One key consideration is the mating system. Upon collection if individual genomes are to be preserved as a unit for genetic mapping, they must be maintained as isofemale lines by selfing or sib-mating. Isofemale lines quickly become homozygous. Naturally selfing organisms like

Table 1

Strengths and weaknesses of some genetic systems

Type	Strengths	Weaknesses
Natural isolates (inbred)	1. Using what evolution provides in natural populations 2. Heterozygosity can be restored with genetic crosses 3. Better able to localized causal genetic variants.	1. Critical genes are harder to study 2. Need large numbers of isolates to find causal variants 3. Homozygosity leads to removal of natural deleterious recessive alleles
Recombinant inbred lines (RILs)	1. Powerful for genetic mapping to the genomic neighborhood of causal variants 2. Heterozygosity can be restored with genetic crosses	1. Lacks representative genetic variation 2. Isolating specific causal genetic variant difficult, homozygosity represents an artificial genetic architecture 3. Need large numbers of replicates
Genetic mutations ('nearly isogenic lines', NILs)	1. Many available 2. Can make more with gene editing tools 3. Can target specific genes	1. Often different genetic backgrounds 2. Labor intensive to create
Mutation accumulation lines (MA)	1. Provide a nearly unbiased sample of the genetic variation introduced by mutation	1. Slow in multicellular organisms 2. Labor intensive 3. Difficult to accumulate on a heterozygous background or assay when heterozygous except in Drosophila.

yeast, *C. elegans* and *Arabidopsis* are highly homozygous in nature, and inbred laboratory populations will reflect the natural condition. On the other hand, flies, mice and zebrafish are naturally outbred and thus heterozygous at many loci in the wild, but with inbreeding in the lab deleterious mutations will be underrepresented because recessive mutations will be exposed to purifying selection. That distinction will be important to evolutionary biologists, but perhaps less so to biomedical researchers whose primary interest is likely to be in associating genes with phenotypes. If heterozygous effects are of interest, inbred lines can be crossed to a standard stock or each other and the phenotype measured in F_1s.

In contrast to humans, where there are now studies encompassing thousands of individuals with the advent of precision medicine initiatives [35–37], studies of natural genetic variation in the metabolome in model organisms are in their infancy. Some studies in flies have included a few tens of wild isolates, targeted toward capturing the variation among genotypes in the response to environmental factors. For example, Reed and colleagues quantified substantial genetic and dietary effects across 187 metabolites in a panel of 20 isofemale lines, and associated that variation with transcriptomic and disease phenotype variation [2,3••]. Similar small-scale studies were conducted by Promislow and coworkers [38•].

Recombinant inbred lines

Recombinant inbred lines (RILs) provide an especially powerful resource for genetic mapping [39]. RILs are made by crossing two or more distinct genotypes, then inbreeding independent lines derived from the F_1s (or more advanced generations) by selfing or sib-mating until they are nearly completely homozygous. Each RIL genome consists of a unique combination of parental alleles. RILs have a significant advantage over F_2 (or more advanced) mapping panels because each RIL is essentially a clone, so variation can be cleanly partitioned into genetic and non-genetic components. As usual, power to detect associations between genes and phenotype will depend on the sample size. RILs are especially useful when the specific genetic pathway responsible for the phenotype is unknown and one wants to find candidate genes or genomic regions. Panels of many hundreds or even thousands of RILs are available for most model organisms [29,40–44] although as yet there have been only a few applications to metabolomics.

One good example of this approach was from yeast, where Breunig *et al.* examined 74 metabolites in about 100 related genetic strains and found that a gene for cell signaling (*ira2*) also had a large influence on central carbon metabolism [45••]. The power of a large panel of RILs to identify the architecture of quantitative traits was demonstrated by Kruglyak and collegues when they characterized the additive and epistatic genetic effects on growth rate across 1008 RILs in 46 chemically distinct environments [43].

Genetic mutations

Genetic model organisms all have extensive collections of mutations, perhaps their most well-known attribute. The primary advantage of genetic mutations is the ability to examine directly the function of a specific gene or pathway by comparing a mutant with a reference strain (Figure 2c). Care must be taken to control for genetic background in which the mutations lie. Mutant strains are available for a small fee from major stock centers

(e.g. http://modelorganisms.nih.gov). A nice example of this approach applied to the metabolome in *C. elegans* demonstrated the role of peroxisomal β-oxidation in the biosynthesis of ascarosides [46•,47]. The Schroeder laboratory has developed an effective method called DANS (Differential Analysis by 2D NMR Spectroscopy) to compare NMR data to uncover metabolite differences between two genetic strains [48••].

Mutations can be made through chemical mutagenesis or, more recently, through gene editing technologies such as CRISPR/Cas9 [49,50] or RNAi [51,52]. These newer technologies have made it possible to manipulate the 'genetic characteristics' of an organism directly, thus making it possible to test modification in the same genomic location across many different genetic backgrounds. A number of studies have shown that a single genetic manipulation can produce distinct phenotypes on distinct genetic backgrounds [53–56]. In recent work, Kim and coworkers investigated 100 mutant Arabidopsis plants for changes in over 1300 metabolites. They found remarkable stability of the metabolome, with only about 10 metabolites significantly changing with most mutations. They conclude that the metabolome is relatively insensitive to individual genetic changes [57••]. In contrast, in the study cited above using RILs in yeast [45••], some unexpected genes disrupted several metabolites, which they may not have found if they had focused only on specific single gene mutations.

Detecting targets of selection in the metabolome

A particularly powerful genetic tool in model organisms is artificial selection [4••], due to their short generation times. The population can be subjected to a selective force, such as a novel nutrient or growth temperature then tracked over many generations of adaptation (Figure 2d). It is possible to elucidate the underlying architecture of the physiological system [58–60] by analyzing the genetic modifications [61] and changes in the metabolome that are selected. Such an approach can also isolate the most critical genomic locations matching the metabolome to an environment. An example of such a study in Drosophila selected flies for their chill-coma recovery times and found that flies with short recovery times had metabolic networks that were more robust to cold-induced perturbations [4••].

There is also much to be learned from identifying features of the metabolome that appear to be significant targets of natural selection. One powerful approach is an organism of interest that is composed of populations with local adaptation to distinct ecological niches, partitioning the genetic variance in a wild sample into within-population and among-population components. If the among-population variance for the quantitative traits (QST) is significantly different from that estimated from genetic loci (FST), it suggests that the trait has evolved adaptively [62]. With a larger number of traits (e.g. metabolites) there is additional power to detect features of functional importance falling in the tails of distribution of QST values [63].

The second way to infer long-term evolution by natural selection is to compare the genetic variance introduced by mutation (the mutational variance, VM) with the genetic variance observed in natural populations (VG). VM can be estimated by means of a 'mutation accumulation' (MA) experiment, in which populations are allowed to evolve in the near-absence of natural selection through many generations of selfing or inbreeding (Figure 2e) [64]. The ratio of VM/VG provides a measure of the strength of selection acting on a trait [65]. If estimates of VM and VG are available for many traits, comparisons between broad classes of traits can reveal underlying generalities about their relative functional importance [66]; comparisons within a class of traits (e.g. metabolite pools) can identify outliers that may be under atypically strong or weak selection [67,68]. Comparison of trait means of MA populations with the ancestral mean provides a powerful clue of the direction of natural selection, because a directional change in MA indicates that natural selection in the ancestor has purged mutations that would have changed the trait in that direction. Interestingly, several metabolite traits in *C. elegans* exhibit among the largest directional changes under MA of any traits yet measured [69•].

Community resources needed

Model organism communities need to coordinate metabolomics efforts to address questions that are difficult to study in humans such as variation in environmental exposures and the role of sub-metabolomes (e.g. tissues (e.g. [70–72]), developmental stages [73,74,75•,76], and sub-cellular compartments). The metabolic inventory for Drosophila tissues completed by Chintapalli et al. [77] is a good model of how such efforts might be undertaken. In addition, a systematic characterization of the metabolomes of model organisms carrying mutations in key physiological pathways (e.g. initiated in yeast [78]) would be a broadly useful tool to all researchers.

Many potentially important metabolites have yet to be chemically defined, and unless we are tracking what little information we do know about these compounds across studies, we are us at risk of overlooking fundamental biological mechanisms. However, as metabolomic databases grow [79–82], the power to link phenotype to metabolome variation both within and between species will accelerate, because we will be able to prioritize unknowns for chemical elucidation. The data that need to be captured take two basic forms, firstly, chemical information about given compounds (e.g. spectra, structure, chemical identification) and

secondly, quantitative information the compound's concentration and phenotype across experimental conditions. Also, while there is extensive conservation in primary metabolic pathways across model organisms and humans there are also organism-specific idiosyncrasies not captured by pathway databases like KEGG and HMDB [83,84] that can be described using informatics tools like MetaCyc [85]. An organized effort across several model systems to relate metabolites to genetic pathways will provide phylometabolomic maps that impact basic biology and biomedicine.

Acknowledgements

LKR was supported by NIH R01GM098856. CFB was supported by NIH R01GM107227. ASE was supported by the Southeast Center for Integrated Metabolomics (SECIM) NIH/NIDDK, 1U24DK097209-01A1 and the Georgia Research Alliance. We thank Nick Batora for helpful feedback on the manuscript.

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. Beger RD, Dunn W, Schmidt MA, Gross SS, Kirwan JA, Cascante M, Brennan L, Wishart DS, Oresic M, Hankemeier T *et al.*: **Metabolomics enables precision medicine: ''a white paper, community perspective''**. *Metabolomics* 2016, **12**:149.

2. Williams S, Dew-Budd K, Davis K, Anderson J, Bishop R, Freeman K, Davis D, Bray K, Perkins L, Hubickey J *et al.*: **Metabolomic and gene expression profiles exhibit modular genetic and dietary structure linking metabolic syndrome phenotypes in Drosophila**. *G3 (Bethesda)* 2015, **5**:2817-2829.

3. •• Reed LK, Lee K, Zhang Z, Rashid L, Poe A, Hsieh B, Deighton N, Glassbrook N, Bodmer R, Gibson G: **Systems genomics of metabolic phenotypes in wild-type *Drosophila melanogaster***. *Genetics* 2014, **197**:781-793.

In one of the first studies to simultaneously quantify the natural genetic and environmental (dietary) contribution to variation in the metabolome, this study found that one major axes of variation in the metabolome in larvae (PC2) was predictive of the heart disease risk in adult flies. In addition, they found that the relative effect of diet on the metabolome was substantially larger than the diet effect on the transcriptome.

4. •• Williams C, Watanabe M, Guarracino M, Ferraro M, Edison A, Morgan T, Boroujerdi A, Hahn D: **Cold adaptation shapes the robustness of metabolic networks in *Drosophila melanogaster***. *Evolution* 2014, **68**:3505-3523.

Drosophila populations were selected for increased and decreased chill-coma recovery times, an important physiological trait in the wild. They found that flies with reduced chill-coma time exhibited a metabolic network that was more robust to perturbation by cold. The metabolites affected by selection for cold tolerance included those involved in membrane lipid synthesis and amino acid metabolism, thus identifying pathways that may be targets of selection in the wild.

5. Stupp GS, Clendinen CS, Ajredini R, Szewc MA, Garrett T, Menger RF, Yost RA, Beecher C, Edison AS: **Isotopic ratio outlier analysis global metabolomics of *Caenorhabditis elegans***. *Anal Chem* 2013, **85**:11858-11865.

6. Creek DJ, Dunn WB, Fiehn O, Griffin JL, Hall RD, Lei ZT, Mistrik R, Neumann S, Schymanski EL, Sumner LW *et al.*: **Metabolite identification: are you sure? And how do your peers gauge your confidence?**. *Metabolomics* 2014, **10**:350-353.

7. Markley JL, Brüschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, Wishart DS: **The future of NMR-based metabolomics**. *Curr Opin Biotechnol* 2017, **43**:34-40.

8. Wilson RK: **How the worm was won. The *C. elegans* genome sequencing project**. *Trends Genet* 1999, **15**:51-58.

9. Clendinen CS, Stupp GS, Wang B, Garrett TJ, Edison AS: **C-13 metabolomics: NMR and IROA for unknown identification**. *Curr Metab* 2016, **4**:116-120.

10. Edison AS, Clendinen CS, Ajredini R, Beecher C, Ponce FV, Stupp GS: **Metabolomics and natural-products strategies to study chemical ecology in nematodes**. *Integr Comp Biol* 2015, **55**:478-485.

11. Tennessen JM, Barry WE, Cox J, Thummel CS: **Methods for studying metabolism in Drosophila**. *Methods* 2014, **68**:105-115.

12. • Spatafora JW, Bushley KE: **Phylogenomics and evolution of secondary metabolism in plant-associated fungi**. *Curr Opin Plant Biol* 2015, **26**:37-44.

Phylogenomics was used to compare endophytic fungal species to discover gene clusters with potential for more secondary metabolites than previously understood.

13. Edison AS, Hall RD, Junot C, Karp PD, Kurland IJ, Mistrik R, Reed LK, Saito K, Salek RM, Steinbeck C *et al.*: **The time is right to focus on model organism metabolomes**. *Metabolites* 2016, **6**:8.

14. •• Naviaux RK, Naviaux JC, Li K, Bright AT, Alaynick WA, Wang L, Baxter A, Nathan N, Anderson W, Gordon E: **Metabolic features of chronic fatigue syndrome**. *Proc Natl Acad Sci U S A* 2016, **113**:E5472-E5480.

The investigators studied a relatively small cohort of people with chronic fatigue syndrome and controls. They targeted 612 metabolites and found 20 pathways that were different between CFS and control. Surprisingly, many of these pathways correspond to known dauer pathways from *C. elegans*.

15. Choe A, von Reuss SH, Kogan D, Gasser RB, Platzer EG, CC SF, Sternberg PW: **Ascaroside signaling is widely conserved among nematodes**. *Curr Biol* 2012, **22**:772-780.

16. Choe A, Chuman T, von Reuss SH, Dossey AT, Yim JJ, Ajredini R, Kolawa AA, Kaplan F, Alborn HT, Teal PEA *et al.*: **Sex-specific mating pheromones in the nematode *Panagrellus redivivus***. *Proc Natl Acad Sci USA* 2012, **109**:20949-20954.

17. Bose N, Ogawa A, von Reuss SH, Yim JJ, Ragsdale EJ, Sommer RJ, Schroeder FC: **Complex small-molecule architectures regulate phenotypic plasticity in a nematode**. *Angew Chem Int Ed Engl* 2012, **51**:12438-12443.

18. Scharf ME: **Omic research in termites: an overview and a roadmap**. *Front Genet* 2015, **6**:76.

19. Chandrasekaran S, Rittschof CC, Djukovic D, Gu H, Raftery D, Price ND, Robinson GE: **Aggression is associated with aerobic glycolysis in the honey bee brain(1)**. *Genes Brain Behav* 2015, **14**:158-166.

20. Hou Y, Wang XL, Saha TT, Roy S, Zhao B, Raikhel AS, Zou Z: **Temporal coordination of carbohydrate metabolism during mosquito reproduction**. *PLoS Genet* 2015, **11**:e1005309.

21. Mohr SE, Hu Y, Kim K, Housden BE, Perrimon N: **Resources for functional genomics studies in *Drosophila melanogaster***. *Genetics* 2014, **197**:1-18.

22. Brown SD, Moore MW: **The international mouse phenotyping consortium: past and future perspectives on mouse phenotyping**. *Mamm Genome* 2012, **23**:632-640.

23. Wang Z, Sherwood DR: **Dissection of genetic pathways in *C. elegans***. *Methods Cell Biol* 2011, **106**:113-157.

24. Steensels J, Snoek T, Meersman E, Picca Nicolino M, Voordeckers K, Verstrepen KJ: **Improving industrial yeast strains: exploiting natural and artificial diversity**. *FEMS Microbiol Rev* 2014, **38**:947-995.

25. Hilton IB, Gersbach CA: **Enabling functional genomics with genome engineering**. *Genome Res* 2015, **25**:1442-1455.

26. Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, Felix MA, Kruglyak L: **Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity**. *Nat Genet* 2012, **44**:285-290.

27. Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM *et al.*: **The**

***Drosophila melanogaster* genetic reference panel**. *Nature* 2012, **482**:173-178.

28. Seren U, Vilhjalmsson 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.

29. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W *et al.*: **The collaborative cross, a community resource for the genetic analysis of complex traits**. *Nat Genet* 2004, **36**:1133-1137.

30. Huang W, Massouras A, Inoue Y, Peiffer J, Ramia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF *et al.*: **Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines**. *Genome Res* 2014, **24**:1193-1208.

31. • Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V *et al.*: **Population genomics of domestic and wild yeasts**. *Nature* 2009, **458**:337-341.

32. The 1001 Genomes Consortium: **1135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana***. *Cell* 2016, **166**:481-491.

33. Grenier JK, Arguello JR, Moreira MC, Gottipati S, Mohammed J, Hackett SR, Boughton R, Greenberg AJ, Clark AG: **Global diversity lines — a five-continent reference panel of sequenced *Drosophila melanogaster* strains**. *G3 (Bethesda)* 2015, **5**:593-603.

34. Souza V, Rocha M, Valera A, Eguiarte LE: **Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents**. *Appl Environ Microbiol* 1999, **65**:3373-3385.

35. Dunn WB, Lin W, Broadhurst D, Begley P, Brown M, Zelena E, Vaughan AA, Halsall A, Harding N, Knowles JD *et al.*: **Molecular phenotyping of a UK population: defining the human serum metabolome**. *Metabolomics* 2015, **11**:9-26.

36. Karaman I, Ferreira DL, Boulange CL, Kaluarachchi MR, Herrington D, Dona AC, Castagne R, Moayyeri A, Lehne B, Loh M *et al.*: **A workflow for integrated processing of multi-cohort untargeted ^{1}H NMR metabolomics data in large scale metabolic epidemiology**. *J Proteome Res* 2016.

37. Rhee EP, Ho JE, Chen MH, Shen D, Cheng S, Larson MG, Ghorbani A, Shi X, Helenius IT, O'Donnell CJ *et al.*: **A genome-wide association study of the human metabolome in a community-based cohort**. *Cell Metab* 2013, **18**:130-143.

38. • Hariharan R, Hoffman JM, Thomas AS, Soltow QA, Jones DP, Promislow DE: **Invariance and plasticity in the *Drosophila melanogaster* metabolomic network in response to temperature**. *BMC Syst Biol* 2014, **8**:139.
A panel of fifteen inbred natural isolates of *Drosophila melanogaster* were surveyed for changes in metabolite levels across seven time points as the adults aged in both males and females. The researchers identified multiple metabolic pathways that varied significantly across age, genotype, and sex.

39. Crow JF: **Haldane, Bailey, Taylor and recombinant-inbred lines**. *Genetics* 2007, **176**:729-732.

40. Andersen EC, Shimko TC, Crissman JR, Ghosh R, Bloom JS, Seidel HS, Gerke JP, Kruglyak L: **A powerful new quantitative genetics platform, combining *Caenorhabditis elegans* high-throughput fitness assays with a large collection of recombinant strains**. *G3 (Bethesda)* 2015, **5**:911-920.

41. King EG, Macdonald SJ, Long AD: **Properties and power of the Drosophila synthetic population resource for the routine dissection of complex traits**. *Genetics* 2012, **191**:935-949.

42. Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, Maciver F, Mueller M, Hummel O, Monti J, Zidek V *et al.*: **Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease**. *Nat Genet* 2005, **37**:243-253.

43. Bloom JS, Ehrenreich IM, Loo WT, Lite TL, Kruglyak L: **Finding the sources of missing heritability in a yeast cross**. *Nature* 2013, **494**:234-237.

44. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R: **A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana***. *PLoS Genet* 2009, **5**:e1005309.

45. •• Breunig J, Hackett S, Rabinowitz J, Kruglyak L: **Genetic basis of metabolome variation in yeast**. *PLoS Genet* 2014, **10**:e1004142.
100 segregants from a cross of *S. cerevisiae* were analyzed by LC–MS metabolomics. Although only 74 metabolites were characterized, the authors were able to associate 34 to genetic loci. A surprising finding was that ira2 had a big influence on central carbon metabolism. Previously there was no known association between ira2 and metabolism.

46. • Butcher RA, Ragains JR, Li W, Ruvkun G, Clardy J, Mak HY: **Biosynthesis of the *Caenorhabditis elegans* dauer pheromone**. *Proc Natl Acad Sci U S A* 2009, **106**:1875-1879.
Mutants defective in dauer pheromone production were examined for their ascaroside pheromone composition. The daf-22 mutant was known for years, and this study showed that long-chain ascarosides build up and are not processed to shorter, active pheromones. This led to the conclusion that daf-22 is involved in peroxisomal β-oxidation.

47. von Reuss SH, Bose N, Srinivasan J, Yim JJ, Judkins JC, Sternberg PW, Schroeder FC: **Comparative metabolomics reveals biogenesis of ascarosides, a modular library of small-molecule signals in *C. elegans***. *J Am Chem Soc* 2012, **134**:1817-1824.

48. •• Pungaliya C, Srinivasan J, Fox BW, Malik RU, Ludewig AH, Sternberg PW, CC SF: **A shortcut to identifying small molecule signals that regulate behavior and development in *Caenorhabditis elegans***. *Proc Natl Acad Sci U S A* 2009, **106**:7708-7713.
Differential Analysis of 2D NMR Spectroscopy (DANS) is a general method developed in the Schroeder laboratory. The approach is to compare high-resolution dqf-COSY NMR data from two different samples to see the chemical difference between them. The primary application is to examine mutants compared with reference strains to determine small molecules that differ between the two genotypes.

49. Jiang W, Marraffini LA: **CRISPR-Cas: new tools for genetic manipulations from bacterial immunity systems**. *Annu Rev Microbiol* 2015, **69**:209-228.

50. Sternberg SH, Doudna JA: **Expanding the biologist's toolkit with CRISPR-Cas9**. *Mol Cell* 2015, **58**:568-574.

51. Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J: **Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference**. *Nature* 2000, **408**:325-330.

52. Taylor J, Woodcock S: **A perspective on the future of high-throughput RNAi screening: will CRISPR cut out the competition or can RNAi help guide the way?** *J Biomol Screen* 2015, **20**:1040-1051.

53. Paaby AB, White AG, Riccardi DD, Gunsalus KC, Piano F, Rockman MV: **Wild worm embryogenesis harbors ubiquitous polygenic modifier variation**. *Elife* 2015:4.

54. Chandler CH, Chari S, Tack D, Dworkin I: **Causes and consequences of genetic background effects illuminated by integrative genomic analysis**. *Genetics* 2014, **196**:1321-1336.

55. Chari S, Dworkin I: **The conditional nature of genetic interactions: the consequences of wild-type backgrounds on mutational interactions in a genome-wide modifier screen**. *PLoS Genet* 2013, **9**:e1003661.

56. Taylor MB, Phan J, Lee JT, McCadden M, Ehrenreich IM: **Diverse genetic architectures lead to the same cryptic phenotype in a yeast cross**. *Nat Commun* 2016, **7**:11669.

57. •• Kim T, Dreher K, Nilo-Poyanco R, Lee I, Fiehn O, Lange BM, Nikolau BJ, Sumner L, Welti R, Wurtele ES *et al.*: **Patterns of metabolite changes identified from large-scale gene perturbations in Arabidopsis using a genome-scale metabolic network**. *Plant Physiol* 2015, **167**:1685-1698.
136 mutant arabidopsis plants covering 11 different functional categories were investigated for metabolite changes. The authors identified and quantified 1348 metabolites and found only that in about 80% of the mutants examined, fewer than 10 metabolites changed. Over 50 metabolites were changed for 2 different mutants. This stability was far greater than anticipated, suggesting that metabolic networks are stable in the presense of most genetic mutations.

58. Samal A, Wagner A, Martin OC: **Environmental versatility promotes modularity in genome-scale metabolic networks**. *BMC Syst Biol* 2011, **5**:135.

59. Winterbach W, Van Mieghem P, Reinders M, Wang H, de Ridder D: **Topology of molecular interaction networks**. *BMC Syst Biol* 2013, **7**:90.

60. Conrad TM, Lewis NE, Palsson BO: **Microbial laboratory evolution in the era of genome-scale science**. *Mol Syst Biol* 2011, **7**:509.

61. Schlotterer C, Kofler R, Versace E, Tobler R, Franssen SU: **Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation**. *Heredity (Edinb)* 2015, **114**:431-440.

62. Spitze K: **Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation**. *Genetics* 1993, **135**:367-374.

63. Whitlock MC: **Evolutionary inference from QST**. *Mol Ecol* 2008, **17**:1885-1896.

64. Halligan DL, Keightley PD: **Spontaneous mutation accumulation studies in evolutionary genetics**. *Ann Rev Ecol Evol Syst* 2009, **40**:151-172.

65. Barton NH: **Pleiotropic models of quantitative variation**. *Genetics* 1990, **124**:773-782.

66. Houle D, Morikawa B, Lynch M: **Comparing mutational variabilities**. *Genetics* 1996, **143**:1467-1483.

67. Farhadifar R, Ponciano JM, Andersen EC, Needleman DJ, Baer CF: **Mutation is a sufficient and robust predictor of genetic variation for mitotic spindle traits in *Caenorhabditis elegans***. *Genetics* 2016, **203**:1859-1870.

68. Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK: **The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans***. *Nat Genet* 2005, **37**:544-548.

69. • Davies SK, Leroi AM, Burt A, Bundy JG, Baer CF: **The mutational structure of metabolism in *Caenorhabditis elegans***. *Evolution* 2016.

The first assessment of the cumulative effects of spontaneous mutations on the metabolome in any organism. Averaged over 29 metabolites, the vulnerability to mutation as measured by the increase in the genetic variance and the absolute change in the mean was typical of other types of traits. However, several metabolites were among the most mutationally vulnerable of any traits quantified in that way, in any organism.

70. Bozek K, Wei Y, Yan Z, Liu X, Xiong J, Sugimoto M, Tomita M, Paabo S, Pieszek R, Sherwood CC *et al.*: **Exceptional evolutionary divergence of human muscle and brain metabolomes parallels human cognitive and physical uniqueness**. *PLoS Biol* 2014, **12**:e1001871.

71. Paine MR, Kim J, Bennett RV, Parry RM, Gaul DA, Wang MD, Matzuk MM, Fernandez FM: **Whole reproductive system non-negative matrix factorization mass spectrometry imaging of an early-stage ovarian cancer mouse model**. *PLOS ONE* 2016, **11**:e0154837.

72. Menger RF, Clendinen CS, Searcy LA, Edison AS, Yost RA: **MALDI mass spectrometric imaging of *Caenorhabditis elegans***. *Curr Metab* 2015, **3**:130-138.

73. Tennessen JM, Bertagnolli NM, Evans J, Sieber MH, Cox J, Thummel CS: **Coordinated metabolic transitions during *Drosophila* embryogenesis and the onset of aerobic glycolysis**. *G3 (Bethesda)* 2014, **4**:839-850.

74. Kaplan F, Srinivasan J, Mahanti P, Ajredini R, Durak O, Nimalendran R, Sternberg PW, Teal PE, CC. SF, Edison AS *et al.*: **Ascaroside expression in *Caenorhabditis elegans* is strongly dependent on diet and developmental stage**. *PLoS ONE* 2011, **6**:e17804.

75. • Srinivasan J, Kaplan F, Ajredini R, Zachariah C, Alborn HT, Teal PE, Malik RU, Edison AS, Sternberg PW, CC SF: **A blend of small molecules regulates both mating and development in *Caenorhabditis elegans***. *Nature* 2008, **454**:1115-1118.

The authors collected the exometabolomes of *C. elegans* as a function of developmental stage. The material was bioassayed for male-specific attraction, and the authors found that only at L4 stages and older were males attracted to hermaphroditic cultures. This allowed for the purification and identification of the first nematode mating pheromone.

76. An PN, Yamaguchi M, Bamba T, Fukusaki E: **Metabolome analysis of *Drosophila melanogaster* during embryogenesis**. *PLOS ONE* 2014, **9**:e99519.

77. Chintapalli VR, Al Bratty M, Korzekwa D, Watson DG, Dow JA: **Mapping an atlas of tissue-specific *Drosophila melanogaster* metabolomes by high resolution mass spectrometry**. *PLoS ONE* 2013, **8**:e78066.

78. Cooper SJ, Finney GL, Brown SL, Nelson SK, Hesselberth J, MacCoss MJ, Fields S: **High-throughput profiling of amino acids in strains of the *Saccharomyces cerevisiae* deletion collection**. *Genome Res* 2010, **20**:1288-1296.

79. Sud M, Fahy E, Cotter D, Azam K, Vadivelu I, Burant C, Edison A, Fiehn O, Higashi R, Nair KS *et al.*: **Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools**. *Nucleic Acids Res* 2016, **44**:D463-D470.

80. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S *et al.*: **HMDB: a knowledgebase for the human metabolome**. *Nucleic Acids Res* 2009, **37**:D603-D610.

81. Haug K, Salek RM, Conesa P, Hastings J, de Matos P, Rijnbeek M, Mahendraker T, Williams M, Neumann S, Rocca-Serra P *et al.*: **MetaboLights — an open-access general-purpose repository for metabolomics studies and associated meta-data**. *Nucleic Acids Res* 2013, **41**:D781-D786.

82. Johnson SR, Lange BM: **Open-access metabolomics databases for natural product research: present capabilities and future potential**. *Front Bioeng Biotechnol* 2015, **3**:22.

83. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: **Data, information, knowledge and principle: back to metabolism in KEGG**. *Nucleic Acids Res* 2014, **42**:D199-D205.

84. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E *et al.*: **HMDB 3.0 — the human metabolome database in 2013**. *Nucleic Acids Res* 2013, **41**:D801-D807.

85. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A *et al.*: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases**. *Nucleic Acids Res* 2014, **42**:D459-D471.