Trends in Genetics

CellPress

Opinion

# The Power of Natural Variation for Model Organism Biology

Audrey P. Gasch,[1,*] Bret A. Payseur,[1,*] and John E. Pool[1,*]

**Genetic background effects have long been recognized and, in some cases studied, but they are often viewed as a nuisance by molecular biologists. We suggest that genetic variation currently represents a critical frontier for molecular studies. Human genetics has seen a surge of interest in genetic variation and its contributions to disease, but insights into disease mechanisms are difficult since information about gene function is lacking. By contrast, model organism genetics has excelled at revealing molecular mechanisms of cellular processes, but often de-emphasizes genetic variation and its functional consequences. We argue that model organism biology would benefit from incorporating natural variation, both to capture how well laboratory lines exemplify the species they represent and to inform on molecular processes and their variability. Such a synthesis would also greatly expand the relevance of model systems for studies of complex trait variation, including disease.**

## Trends

Laboratory lines of model organisms have provided a wealth of information about biological functions. Many of these lines are derived from species with significant genetic and functional variation in nature.

The longstanding focus on particular genetic backgrounds has enabled the elucidation of core biological functions, but has limited our understanding of the levels and impacts of natural variation found in wild populations.

A broader synthesis incorporating natural variation into model organism research would illuminate biological mechanisms, reveal the genetic basis for phenotypic variation, and solidify the importance of model organisms for understanding both basic and personalized genetics.

## Introduction

Model organisms provide powerful experimental platforms to elucidate the basic biological functions of cells. With the implicit understanding that fundamental mechanisms and processes are likely conserved across species, experimentally facile organisms emerged as choice tools for biological exploration. The intense focus on representative models, including bacteria, budding yeast, fruit flies, round worms, zebrafish, mustard weed, and house mice, has provided mechanistic insights across the kingdoms of life. These insights include, for example, a basic understanding of cellular processes (such as genome replication, transcription, and translation), organismal functions (including reproduction, development, and cell–cell communication), and environmental interactions (spanning immunity, stress responses, and circadian relations).

The initial power of model organism genetics emerged in part from the early acceptance of common, often inbred, lines of laboratory models, such that the knowledge generated in one lab could be readily incorporated by other researchers studying the same lines. Focusing on a limited set of lines minimized the impact of genetic background, so that mechanisms underlying fundamental processes could be worked out. However, while the focus on the same laboratory lines fostered a deep understanding of those models, it has limited the scope of biological knowledge not only to a limited set of model species, but also in reality to only a few individual genotypes.

On first principles, individuals within a species are expected to vary in their genetic properties; indeed, this variation is often the subject of genetic and evolutionary investigation. However, the potential for functional differences across genetic backgrounds is routinely ignored in molecular studies. Here, we define functional genetic differences as DNA variants that confer molecular or organismal phenotypic differences among individuals, regardless of the effects on evolutionary fitness. Understanding the phenotypic consequences of naturally occurring genetic variation remains a major challenge in biology. This is acutely clear in the case of personalized medicine,

[1]Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706, USA

*Correspondence: agasch@wisc.edu (A.P. Gasch), bret.payseur@wisc.edu (B.A. Payseur), and jpool@wisc.edu (J.E. Pool).

where the hope is that an individual's genotype will inform on disease predisposition, prognosis, and treatment. Although genetic variation has been documented extensively in humans, the ability to make predictions about phenotypic variation remains limited.

It is in this light that the experimental tools of model organisms offer tremendous potential to amplify our knowledge of how genetic variation modulates molecular processes and phenotypic outcomes. In this perspective, we argue that natural variation is a critical parameter that should be incorporated into models of molecular and cellular biology. Variation in gene expression, protein function, molecular interactions, and network organization can be thought of as an orthogonal dimension of biology that could be incorporated into molecular and cellular models (Figure 1). We believe that expanding model organism biology to incorporate a view of natural variation will cement the importance of model organisms for understanding both basic and applied genetics, including the causes of complex trait variation.

## The Pluses and Perils of Laboratory Lines

The amount of information that has been collected from laboratory lines is remarkable. Decades of classical breeding, forward- and reverse-genetic screens, and molecular dissection have produced detailed descriptions of cellular and organismal function. Many model organisms were among the first in their clades to be sequenced at the genome scale, including *Saccharomyces cerevisiae* [1], *Escherichia coli* [2], *Caenorhabditis elegans* [3], *Drosophila melanogaster* [4],

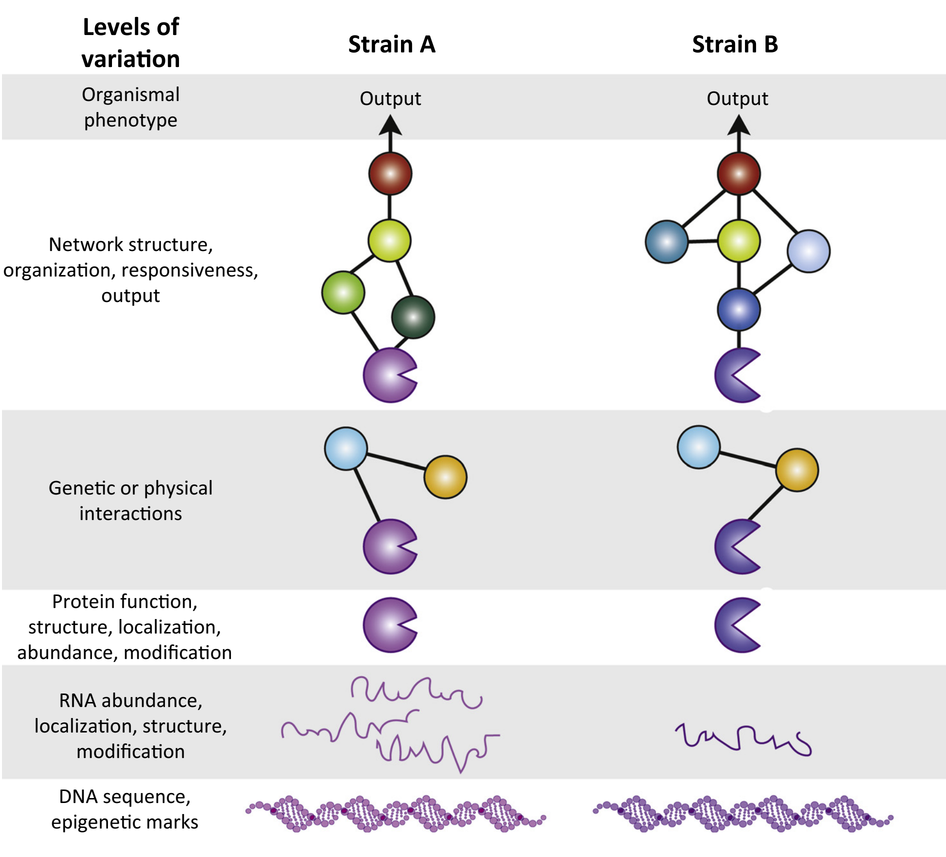


Figure 1. Natural Variation at Multiple Levels. The diagram depicts natural variation emerging for a given gene and its mRNA and/or protein product in one strain (left column) and for the orthologous gene and products in a second genetic background (right column). Circles and lines represent proteins and the interactions between them, respectively.

*Arabidopsis thaliana* [5], and *Mus musculus* [6]. Subsequent technological innovations, including high-throughput sequencing, methods to quantify and locate proteins and metabolites, and robotic and molecular tools to rapidly generate and phenotype mutants, have enabled unprecedented genetic characterization of a few strains of model organisms. While the technology exists to generate these data in a wider range of strains and species, it is really only in the context of the extensive background knowledge for model genetic lines that such high-throughput data can be integrated in a unifying framework. Each discovery is interpreted in light of considerable existing knowledge for that strain, often substantially accelerating understanding. This is especially important for systems-biology models that integrate diverse large-scale data sets, preferably generated from the same genetic background. Hence, there remains an inherent advantage in generating more knowledge on the same laboratory lines.

However, the strategy of focusing on particular lines has important drawbacks that should be recognized. It is becoming increasingly realized that laboratory lines of model organisms are often outliers of their species. These individuals have been selected, deliberately or inadvertently, for their ease of manipulation. Several recent reviews present the unusual features of model genetic backgrounds [7–12] and so we highlight only a few cases here. For example, common lab strains of *S. cerevisiae* are unusual in that they exist as stable haploids, lack the natural flocculence common to many wild isolates, harbor a variety of auxotrophies, and are easy to manipulate genetically [9]. Laboratory stocks of *D. melanogaster*, including the sequenced reference strain, are less active than wild flies, often have impaired senses [13,14], and are mostly descended from North American populations that have a complex, genetically admixed history [15,16]. Commonly used classical inbred strains of house mice, which are in fact hybrids of multiple subspecies [17], grow faster, reproduce earlier, generate larger litters, gain weight more easily, and are less aggressive than wild mice [18–20]. More fundamentally, specific aspects of molecular biology can be unusual in laboratory lines, perhaps obscuring natural processes from view. For example, laboratory lines of *D. melanogaster* lack transposable P elements as well as the regulators that suppress their transposition [21]; had it not been for studies crossing natural and laboratory fly lines, the discovery and subsequent exploitation of P elements may not have occurred.

The phenotypic differences between research lines and wild individuals emerge for a variety of reasons. Some features are the result of artificial selection for adaptation to laboratory environments; others reflect the historical bottleneck of choosing a few strains to serve as the common references for the species. Many of the accumulated differences in commonly used lab strains may in fact be deleterious in nature (even if they are beneficial to life in the lab). For example, mice from one widely used laboratory strain go blind within the first few months of life, a characteristic that would almost certainly be detrimental in their natural habitat [22]. Despite such unusual phenotypes, lab lines are often taken as sole representatives of their species and even of their clades. We argue that for most biological processes, the extent to which mechanistic details vary within species is only beginning to emerge.

## How Much Variation Is There?

DNA sequencing has revealed tremendous genetic variation within and among natural populations of model organism species. On average, a random pair of homologous, nuclear, noncoding sequences differs by 0.3–0.8% in *S. cerevisiae* [23–25], 0.5–1% in *D. melanogaster* [26], 0.1–0.4% in *C. elegans* [27,28], and 0.1–0.2% in *M. musculus* [29]. A large number of variants have the potential to affect phenotypes of interest. For example, surprising numbers of nonsense mutations are segregating in natural populations of both model organisms [30–32] and humans. The average human genome harbors predicted loss-of-function variants in 250–300 genes, according to the 1000 Genomes Project [33]. Yet, the functional consequences of genetic variation in model organisms (and in other species) remain mostly uncharacterized.

The common assumption is that 'important' features of the organism will be conserved within and even between species and, indeed, conservation can be a signature of evolutionary constraint (but can also emerge from low mutation rates). However, a lack of conservation does not necessarily imply an absence of importance to the organism. In fact, a low-conservation signature can be produced by rapid evolution in one or more lineages, specifically because of genetic changes that are adaptively significant [34–37]. Adaptive variation within model species (e.g., in populations from contrasting environments) could be particularly informative in studies dissecting molecular variation, since this type of variation is less likely to emerge clearly from mutagenesis in the laboratory. The lack of conservation can also arise if one genetic change is compensated by other mutations through epistasis. This phenomenon is clearly seen in the evolution of transcription factor-binding sites across closely related yeast species, where the disappearance of a binding element at one promoter position is compensated by the appearance of a novel instance of that element elsewhere in the regulatory region [38].

Another example of the complex interplay between molecular mechanism and constraint is the wholesale rewiring of transcriptional regulatory networks over long evolutionary time frames in fungi. The coregulation of ribosomal protein (RP) genes is highly conserved across kingdoms, and the acute response of RP genes to environmental insults persists over a billion years of fungal evolution [39]. Yet, the regulation of RP genes is perhaps the clearest case where the upstream regulatory mechanisms are completely different across fungal species. Concerted evolution in the network required changes in both upstream regulatory connections and downstream transcriptional elements controlling each of the more than 100 genes [40–43]. In this example, regulatory mechanisms have diverged substantially across species even as tight coregulation and environmental responsiveness of the genes has been highly conserved, and it is likely the coregulated response, rather than the mechanism producing it, that is important for fitness. While it might be expected that long evolutionary timeframes are required to enable wholesale network rewiring, evidence for within-species regulatory rewiring exists: regulation of yeast cell adhesion requires the filamentous MAP kinase pathway in one strain but an entirely different upstream regulatory process in another [44]. These examples illustrate the potential for mechanistic variation in producing the same cellular output.

Although the mechanistic underpinnings in many cases remain murky, high-throughput experiments are now quantifying the prevalence of genetic background effects on functional variation within a species [45]. Comparing gene-knockout libraries from two strains of *S. cerevisiae* revealed that nearly 5% of genes scored essential in one strain are dispensable for survival in the other [46]. How this disparity is tolerated across strains is largely unknown, but could be informative from a mechanistic perspective. In *C. elegans*, the reduced expression of approximately 20% of interrogated genes produced a phenotype that varied considerably across two different lines [47]. Genetic interactions can also vary at a high rate. One study screened two *D. melanogaster* lines for genetic modifiers of a hypomorphic allele of *scalloped*, a transcription factor involved in wing development: nearly 74% of all modifiers produced different effects in the two lines. While most of these differences related to the severity of the phenotype, several modifiers produced opposite effects in the two strain backgrounds. Dissecting the genetic basis of such phenotypic differences has been a longstanding interest of many biologists; we contend that leveraging such differences specifically to understand cellular mechanisms presents an exciting opportunity.

## Incorporating Natural Variation as an Orthogonal Variable in Cellular Networks

Recently, the genetic tools of laboratory lines have been exploited to understand the consequences of variation found in nature. We submit that the reverse can also be true: natural variation could significantly illuminate our understanding of molecular mechanisms. Natural phenotypic variation can be conceptualized as an orthogonal dimension of biology. Just as an individual's DNA sequence can be mapped onto a reference genome, we argue that natural variation in gene expression, protein function, and molecular interactions can be mapped onto cellular networks elucidated in reference lines of model organisms (Figure 1). Adding this orthogonal level of information could reveal profound insights into how cells function, by pointing to individual genes, proteins, or other elements that vary functionally, by revealing mechanistic plasticity across genetic backgrounds, and by suggesting the features of the cellular network (or its output) that are under constraint.

Several case studies highlight the potential for genetic variation to elucidate underlying molecular or physiological mechanisms, particularly those relevant to human health. One recent study incorporated genetic variation from the *Drosophila* Genetic Reference Panel (DGRP) [48,49] into a *D. melanogaster* model of traumatic brain injury (TBI) [50]. The DGRP comprises >200 inbred lines of wild fly isolates, selected to represent common polymorphisms in a natural population, and has served as a powerful community resource for mapping complex traits [48,49]. Mapping TBI sensitivity implicated genes linked to intestinal barrier function; this link, coupled with subsequent molecular studies in the laboratory strain, revealed that a low-sugar diet following TBI protects against death in multiple different genetic backgrounds. This example highlights the ability of natural variation to reveal physiological connections that determine and affect phenotypes, including disease. Another example is provided by a series of experiments exploring differences in pathogen susceptibility in two different commonly used strains of *C. elegans*. Genetic mapping of differential pathogen responses, using recombinant inbred lines emerging from a cross of the two parental strains, implicated a neuronal peptide receptor and, through subsequent mapping, an upstream ubiquitin ligase that regulates a behavioral response to pathogen recognition [51,52]. Combining this information with molecular, structural, and physiological experiments in the lab strain not only implicated the genetic basis for the variable phenotype, but also revealed a new molecular model for pathogen avoidance and contributed to a broader understanding of mechanosensory neuronal circuits [53].

There are those who would argue that genetic and phenotypic variation segregating in nature is less informative for molecular studies than deep laboratory mutagenesis in a single purebred line. We recognize that laboratory mutagenesis is highly effective at dissecting gene and protein function in a single genetic background. However, we argue that the outcome and interpretation of such experiments could be significantly different if done in a different genetic background, as highlighted by several examples above. At issue then is the question being asked: if the goal is to elucidate molecular mechanisms in a single individual, saturating laboratory mutagenesis may indeed be the best approach. However, if the aim is to dissect a molecular mechanism to serve as a model for other organisms, especially for human disease, which is likely to be complex, then the additional step of at least considering natural variation may significantly expand the scope.

## Capturing Natural Functional Variation in Model Organisms

Realizing the power of natural variation for model organism biology will require changes in perspective and in practice. Ultimately, different labs may choose to incorporate natural variation to different extents, from simply scanning allele frequencies of mechanistically implicated genes to performing detailed experiments in multiple backgrounds. While it is impractical to expect a deep understanding of all genetic backgrounds found in nature, we envisage several realistic avenues through which natural variation can be incorporated into molecular studies.

### Understand the Phenotypic Distribution of Natural Strains and Determine where Laboratory Lines Fall Along that Spectrum

Surveying the distribution of a particular phenotype is now often feasible, thanks to medium- and high-throughput techniques and broader availability of natural lines to test. Knowledge of the phenotypic distribution would inform our perspective on the potential diversity of underlying mechanisms and may help prioritize follow-up experiments. For example, comparing phenotypic differences with sequence variation in genes of interest could suggest alternate genetic backgrounds whose subsequent study may be particularly informative from a mechanistic perspective. Knowledge of the phenotypic distribution will also clarify whether lab-strain phenotypes are representative of segregating variation or are unusual compared with other individuals in the species. Even when laboratory lines lie at the outer ranges of the phenotypic distribution, understanding the mechanistic underpinnings in those strains can still be useful; in fact, studying unusual strains in the context of the broader population may be a powerful model for human disease, in which specific genetic backgrounds are particularly susceptible to disease.

### Expand Detailed Molecular Studies to a Broader Set of Diverse Genetic Backgrounds

To expand our understanding of mechanistic possibilities, model organism communities should consider establishing diverse sets of representative lines, chosen to capture the genetic and phenotypic variation of the species. We advocate a two-step framework: (i) a broad survey of phenotypic diversity, as outlined above; and (ii) detailed molecular dissection, when informative, on strategically chosen genetic backgrounds that could be particularly illuminating from a mechanistic perspective. The first step may be more realistically facilitated by a community-wide approach to establish strain panels, which would enable targeted investigation of specific lines by individual labs, to validate detailed molecular studies done in laboratory strains. Additional reference lines should be chosen to maximize variation, such as geographically diverse isolates including representatives from the ancestral range of the species, and adaptively unique natural populations, such as those from extreme environments or distinct substrates. A promising example of a community-wide effort is the *Arabidopsis* 1001 Genomes Project, in which ecologically diverse *A. thaliana* lines sampled from across the ancestral species range are being sequenced to catalyze the functional dissection of genetic variation [54]. This and other mapping resources for model organism communities [28,48,49,54–57] not only enable mapping of phenotypic differences in parental lines, but can also unmask cryptic variation that provides a powerful complement to laboratory mutagenesis (e.g., [58,59]).

### Represent the Potential for Natural Variation when Presenting Mechanistic Studies

Finally, it is important to recognize that variation exists, even when it is not the main point of a study. Results should be described in terms of the strain or genetic background used and not as the sole representation of the species (unless broader information in other lines exists). This is especially important for cross-species comparisons, which frequently consider only one strain, in many cases a laboratory line, as a single representative of each species. We also encourage model organism databases to organize our knowledge of molecular, genomic, and phenotypic variation; a nice example is the Mouse Phenome Database, which captures various phenotypes of diverse mouse lines [60].

## Concluding Remarks

Major advances in biology have often come from combining established perspectives. A synthesis incorporating natural variation into molecular and cellular biological studies could accelerate our understanding of life, leveraging the depth achieved by studying laboratory lines of model organisms and the breadth afforded by incorporating natural variation (see Outstanding Questions). This synthesis would also reinforce the relevance of model organism genetics for understanding disease biology and its interaction with genetic variation.

**Acknowledgments**

We thank our colleagues David Wassarman and Barry Ganetzky and two reviewers for insightful comments. The authors were supported by grants from the NIH (R01GM083989 to A.P.G, R01GM100426 to B.A.P., and R01GM111797 to J.E.P.) and NSF (DEB1353737 to B.A.P.).

### Outstanding Questions

To what degree are molecular interactions conserved across individuals, populations, and species? How does variation at the molecular level impact phenotypic variation at the organismal level?

What are the molecular underpinnings of complex trait variation, and how can that information in turn provide a better mechanistic understanding of biological functions?

In what ways can an expanded understanding of natural variation in model organisms influence our understanding or study of the functional impacts of natural variation influencing human disease?

## References


1. Goffeau, A. *et al.* (1996) Life with 6000 genes. *Science* 274, 563–567
2. Blattner, F.R. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462
3. *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018
4. Adams, M.D. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195
5. Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
6. Mouse Genome Sequencing Consortium *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562
7. Alfred, J. and Baldwin, I.T. (2015) New opportunities at the wild frontier. *Elife* 4, 06956
8. Frezal, L. and Felix, M.A. (2015) *C. elegans* outside the Petri dish. *Elife* 4, 05849
9. Liti, G. (2015) The fascinating and secret wild life of the budding yeast *S. cerevisiae*. *Elife* 4, 05835
10. Markow, T.A. (2015) The secret lives of *Drosophila* flies. *Elife* 4, 06793
11. Phifer-Rixey, M. and Nachman, M.W. (2015) Insights into mammalian biology from the wild house mouse *Mus musculus*. *Elife* 4, 05959
12. Sterken, M.G. *et al.* (2015) The laboratory domestication of *Caenorhabditis elegans*. *Trends Genet.* 31, 224–231
13. Kain, J.S. *et al.* (2012) Phototactic personality in fruit flies and its suppression by serotonin and white. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19834–19839
14. Senthilan, P.R. *et al.* (2012) *Drosophila* auditory organ genes and genetic hearing defects. *Cell* 150, 1042–1054
15. Kao, J.Y. *et al.* (2015) Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from the south-eastern United States and Caribbean Islands. *Mol. Ecol.* 24, 1499–1509
16. Pool, J.E. (2015) The mosaic ancestry of the *Drosophila* genetic reference panel and the *D. melanogaster* reference genome reveals a network of epistatic fitness interactions. *Mol. Biol. Evol.* 32, 3236–3251
17. Yang, H. *et al.* (2011) Subspecific origin and haplotype diversity in the laboratory mouse. *Nat. Genet.* 43, 648–655
18. Bronson, F.H. (1984) Energy allocation and reproductive development in wild and domestic house mice. *Biol. Reprod.* 31, 83–88
19. Miller, R.A. *et al.* (2000) Mouse (*Mus musculus*) stocks derived from tropical islands: new models for genetic analysis of life-history traits. *J. Zool.* 250, 95–104
20. Harper, J.M. (2008) Wild-derived mouse stocks: an underappreciated tool for aging research. *Age* 30, 135–145
21. Engels, W.R. (1983) The P family of transposable elements in *Drosophila*. *Annu. Rev. Genet.* 17, 315–344
22. Chang, B. *et al.* (2002) Retinal degeneration mutants in the mouse. *Vision Res.* 42, 517–525
23. Liti, G. *et al.* (2009) Population genomics of domestic and wild yeasts. *Nature* 458, 337–341
24. Schacherer, J. *et al.* (2009) Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458, 342–345
25. Wang, Q.M. *et al.* (2012) Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol. Ecol.* 21, 5404–5417
26. Lack, J.B. *et al.* (2015) The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* 199, 1229–1241
27. Cutter, A.D. *et al.* (2009) Evolution of the *Caenorhabditis elegans* genome. *Mol. Biol. Evol.* 26, 1199–1234
28. Andersen, E.C. *et al.* (2012) Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat. Genet.* 44, 285–290
29. Geraldes, A. *et al.* (2011) Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Mol. Ecol.* 20, 4722–4736
30. Jelier, R. *et al.* (2011) Predicting phenotypic variation in yeast from individual genome sequences. *Nat. Genet.* 43, 1270–1274
31. Wohlbach, D.J. *et al.* (2014) Comparative genomics of *Saccharomyces cerevisiae* natural isolates for bioenergy production. *Genome Biol. Evol.* 6, 2557–2566
32. Wang, F. *et al.* (2015) FlyVar: a database for genetic variation in Drosophila melanogaster. *Database* 2015, bav079
33. Genomes Project *et al.* (2010) A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073
34. Bishop, J.G. *et al.* (2000) Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5322–5327
35. Bergelson, J. *et al.* (2001) Evolutionary dynamics of plant R-genes. *Science* 292, 2281–2285
36. Swanson, W.J. and Vacquier, V.D. (2002) The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* 3, 137–144
37. Obbard, D.J. *et al.* (2006) Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr. Biol.* 16, 580–585
38. Doniger, S.W. and Fay, J.C. (2007) Frequent gain and loss of functional transcription factor binding sites. *PLoS Comput. Biol.* 3, e99
39. Gasch, A.P. (2007) Comparative genomics of the environmental stress response in ascomycete fungi. *Yeast* 24, 961–976
40. Gasch, A.P. *et al.* (2004) Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biol.* 2, e398
41. Tanay, A. *et al.* (2005) Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7203–7208
42. Hogues, H. *et al.* (2008) Transcription factor substitution during the evolution of fungal ribosome regulation. *Mol. Cell.* 29, 552–562
43. Lavoie, H. *et al.* (2010) Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol.* 8, e1000329
44. Chin, B.L. *et al.* (2012) Genetic variation in *Saccharomyces cerevisiae*: circuit diversification in a signal transduction network. *Genetics* 192, 1523–1532
45. Chandler, C.H. *et al.* (2013) Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends Genet.* 29, 358–366
46. Dowell, R.D. *et al.* (2010) Genotype to phenotype: a complex problem. *Science* 328, 469
47. Vu, V. *et al.* (2015) Natural variation in gene expression modulates the severity of mutant phenotypes. *Cell* 162, 391–402
48. Mackay, T.F. *et al.* (2012) The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482, 173–178
49. Huang, W. *et al.* (2014) Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res.* 24, 1193–2108
50. Katzenberger, R.J. *et al.* (2015) Death following traumatic brain injury in *Drosophila* is associated with intestinal barrier dysfunction. *Elife* 4, 04790
51. Reddy, K.C. *et al.* (2009) A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in *C. elegans*. *Science* 323, 382–384
52. Chang, H.C. *et al.* (2011) Natural polymorphisms in *C. elegans* HECW-1 E3 ligase affect pathogen avoidance behaviour. *Nature* 480, 525–529
53. Schafer, W.R. (2015) Mechanosensory molecules and circuits in *C. elegans*. *Pflugers Arch.* 467, 39–48
54. Cao, J. *et al.* (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* 43, 956–963

55. Churchill, G.A. *et al.* (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* 36, 1133–1137

56. King, E.G. *et al.* (2012) Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. *Genome Res.* 22, 1558–1566

57. Long, A.D. *et al.* (2014) Dissecting complex traits using the Drosophila Synthetic Population Resource. *Trends Genet.* 30, 488–495

58. Campbell, M. and Ganetzky, B. (2013) Identification of Mob2, a novel regulator of larval neuromuscular junction morphology, in natural populations of *Drosophila melanogaster*. *Genetics* 195, 915–926

59. Paaby, A.B. *et al.* (2015) Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *Elife* 4, 09178

60. Grubb, S.C. *et al.* (2014) Mouse phenome database. *Nucleic Acids Res.* 42, D825–D834