

ECOLOGICAL GENOMICS OF MODEL EUKARYOTES¹

John K. McKay^{1,2} and John R. Stinchcombe³

¹Department of Bioagricultural Sciences and Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO 80523

²E-mail: J.McKay@ColoState.EDU

³Department of Ecology and Evolutionary Biology and Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON M5S3B2, Canada

Received July 22, 2008

Accepted July 22, 2008

“It is of basic importance to understand the hereditary mechanisms that function in the perpetuation of the existing natural races and species and in the evolution of new ones, but investigations in this field are difficult and have largely been neglected. The commonly used laboratory organisms are inadequate to resolve differences that characterize natural biological entities found in the wild”

—Clausen and Hiesey, 1958

Fifty years ago, the publication of “The experimental nature of species (part 4)” mapped out the field of ecological genetics. Clausen and Hiesey (1958) outlined three major areas in which data were needed: (1) genetic analysis of the traits involved in local adaptation, (2) understanding acclimation, or the plasticity of traits and gene expression across ecologically relevant conditions, and (3) understanding the pathways that underlie such acclimation and adaptation. Much has changed in the past 50 years since their seminal work. First and foremost, the work of Clausen, Keck, and Hiesey (1940, 1948) has inspired a tremendous amount of effort in the fields of ecological genetics and evolutionary ecology to study ecologically important traits with statistical methods that describe population-level patterns of genetic and phenotypic variance. Such quantitative genetic approaches can provide many basic answers, including predictions of the evolution of traits,

even in the absence of any knowledge about the loci underlying them (Falconer and MacKay 1996). An undisputed finding of this research is that many phenotypes of ecological significance are quantitative in nature; variation in these traits results from several loci as well as the environment. For example, individual size, tolerance to abiotic stress, and fecundity are all traits that generally exhibit a continuous distribution. Such traits have been studied using experimental breeding designs and quantitative genetic analysis, largely borrowed from agricultural genetics, which together provide estimates of the amount of genetic variance in these traits. These experimentally determined estimates of genetic variance for a particular population and environment can be combined with estimates of natural or artificial selection to predict the short-term evolutionary response. Together these data revealed surprising amounts of variation in ecological traits and revealed which traits were subject to strong selection.

The revolutionary advances in genomics, molecular biology, and biotechnology have now provided the prospects of moving beyond quantitative genetics, and actually identifying the actual loci underlying ecological variation. This prospect raises a natural question: Are there additional insights that can be gained from knowing the genes underlying traits, as opposed to simply knowing that some unknown number of polymorphic loci must exist? The answer to this question is a resounding yes. Several longstanding questions in evolutionary and ecological genetics can only be addressed by identifying and examining the history

¹Symposium at the 2007 Annual Meetings of the Society for the Study of Evolution, Christchurch, New Zealand

of the actual loci underlying variation in ecologically important traits. For instance, abundant genetic variation has been described over the last 50 years (both quantitative genetic and sequence variation) but we have historically been unable to distinguish between alternatives causing this variation. Is this variation due to mutation–selection balance, where variation is due to new mutations that have yet to be eliminated by selection or balancing selection, where variation is maintained over long periods of time by fluctuating selection pressure? Yet, these competing models make alternative predictions about the frequencies of alleles with effects on ecologically important traits and fitness. Although we have relatively little data with which to evaluate these models, examining sequence data for functional traits allows for testing predictions these predictions.

With regard to the overall field of ecological genetics, there are a host of issues related to the genetics of adaptation for which we currently have more questions than answers (Orr 2005): How many genes underlie adaptation, what are their effect sizes, and are they consistent with existing theoretical models? For example, Orr (1998) predicts an exponential distribution, where the effect size decreases as genotypes approach an optimal phenotype. A related question is whether the response to selection is limited by beneficial new mutations, or standing genetic variation within populations? Are the loci that have diverged between populations and species also polymorphic within populations? How often are adaptations due to dominant or recessive mutations? What genes underlie continuous variation: regulatory or protein coding? Are genetic correlations between traits due to linkage disequilibrium or pleiotropy? We have surprisingly little data with which to evaluate these issues. Indeed, the number of cases in which these questions have actually been answered pales in comparison to our ability to enumerate the questions, and the excitement over the prospects of addressing them.

Identifying the loci underlying variation in ecologically important traits allows investigators to study biological phenomena at multiple levels of organization. Evolution is clearly both an ecological and a genetic process: ecological forces produce selection, whereas alleles change in frequency between generations. Although few would argue with this definition, consider how rarely we understand the evolutionary process on multiple levels of biological organization from DNA sequence variation to the complex, multispecies ecological communities that are producing selection. For instance, how many systems or traits do we understand (1) the genes that influence a trait, (2) the genetic variances, covariances, and heritabilities of that trait in natural populations, (3) the strength of natural selection on that trait in the wild, (4) the agents of selection, and finally (5) how any of the above vary depending on the community context, habitat, or demographic or evolutionary history of the populations? Although for a variety of traits and systems we have detailed evidence on

one or a few of these questions, we have a dearth of examples to point to where the ecology, evolution, and genetics of adaptations are well understood at multiple levels of biological organization.

Until recently, evolutionary and ecological geneticists have simply lacked the tools for rigorously addressing these questions. Advances in molecular biology, model systems, and computational approaches are now allowing the first glimpse of the molecular basis of polygenic variation, and with it, the possibility of addressing longstanding theoretical and empirical issues that have been out of reach.

The Promise and Limitations of Model Systems

As the second sentence in the Clausen and Hiesey quote suggests, in the recent past, model systems offered experimental tractability such as short generation times, ease of crossing and mutant screening, and simple conditions for growth. Today, the great advantages of model systems are fully sequenced and annotated genomes. With current sequencing and expression profiling technology, the gap is closing between model and nonmodel species. At present, however, it is only in a handful of model species, where we have the tools to dissect the genetic basis of within species variation in complex traits. Such tools include genome resequencing of multiple individuals to identify which loci and regions of the genome vary within the species, readily available mutants or gene silencing, genetic transformation, and a large research community to provide experimental evidence for gene function. Although such tools could in principle be created for any organism, the massive effort required makes it worthwhile to ask what evolutionary questions we can and cannot answer with the ideal genomic tools. In other words, given the best-case scenario of genomic resources, how much (genetic) progress can be made?

The current limitation of model systems is that they are often removed from nature, are often human commensals, and include a limited set of life histories. For example, *Arabidopsis*, *Drosophila*, and *Lycopersicon* accessions that are frequently studied (see below) are often obtained from a stock center. Although originally collected in nature, often little information is available for the habitat or ecology of the accessions collected. In many cases the researchers have never been to the collection sites. This lack of basic knowledge about the ecology, such as when the local population is growing and reproducing, can limit the ability to interpret phenotypic and molecular variation. Although we are all tempted to correlate variation among genotypes in our favorite traits or loci with aspects of the native climate, this can be a fishing expedition if we do not know how and when the organism is interacting with habitat variables. This situation appears worst for *Caenorhabditis elegans*, where the lack of basic knowledge on the ecology of the

organism limits our ability to understand which traits might be under selection in nature.

The Symposium

The symposium in Christchurch, NZ, had the goal of exploring the progress and prospects for ecological genomics, and assessment of how close the field has come to answering the longstanding and challenging questions noted above. The symposium featured a range of talks on both established model systems (*Drosophila*, *Arabidopsis thaliana*, *Homo sapiens*, *C. elegans*) as well as emerging model systems (*Lycopersicon*, *Mimulus*, *Daphnia*). These systems range from ones in which the genomic resources far outstrip the ecological and natural history knowledge of the species to ones in which a detailed ecological understanding was developed first, and for which genomic resources are now being developed. A consistent theme throughout the symposium was the need to achieve greater balance between the ecological and genomic knowledge available for all of these systems—this will entail more ambitious, painstaking, and clever ecological experiments with established model systems, and continued investment in genome sequencing, transformation technology, and gene-expression platforms for emerging model systems.

At the symposium, Brian Eads (Indiana University) described the recent progress made in developing genomic and expression tools for *Daphnia pulex*, an organism with a long and rich history in population ecology and limnology. Work in this system has the potential to unravel how changes in gene expression affect the switch between parthenogenetic reproduction and sexual reproduction (e.g., Eads et al. 2007). Theodore Morgan (Kansas State University) described work started in collaboration with Trudy McKay and colleagues, and continued in his laboratory, on the genetic basis of thermotolerance in *Drosophila melanogaster* (Morgan and Mackay 2006), a complex trait likely to be important to many insects, and possibly under novel selection due to global climate change (Balanyá et al. 2006). Corbin Jones (University of North Carolina) provided an update on his laboratory's work on unraveling the genetic basis of specialization in *Drosophila sechellia* on the fruits of *Morinda citrifolia*, which is typically toxic to most *Drosophila* species. By combining hybridization of antenna tissue from *D. sechellia* to the Affymetrix *Drosophila* genome version 2.0 microarrays and phylogenetic analysis of gene expression patterns in the closely related *D. simulans* and *D. melanogaster*, it is possible to narrow in on genes whose expression were upregulated during the evolution of host specialization (Kopp et al. 2008). Matthew Rockman (New York University) described research that could be described with the converse approach to that used by Jones: rather than starting with ecologically important traits, Rockman advocated using “reverse ecology.” Using the power of genetic mapping (in his case, in *C. elegans*),

this approach finds the genes underlying arbitrary traits and then lets molecular population genetic approaches reveal the selective consequences of variation in those traits. Although this approach is unbiased by human conceptions of which traits are ecologically consequential, the next (and quite formidable) challenge is work outwards from the genes and their evolution to the ecological mechanisms of selection. For some traits, such as copulatory plugging (Palopoli et al. 2008), such inferences are straightforward, but in other cases, such as embryonic lethality maintained by balancing selection (Seidel et al. 2008), the ecological context remains deeply mysterious.

The remaining five speakers in the symposium, whose papers are included in the special feature, used a mixture of established and emerging model systems. Both McKay et al. (2008) and Samis et al. (2008) used *A. thaliana* to ask questions about the genetic basis of ecologically important traits in plants. McKay et al. (2008) make an important contribution by developing a new set of recombinant inbred lines for mapping ecologically important traits. These lines were developed for explicitly ecological reasons, with the parents originating in locations with dramatically different water availabilities. As such, loci identified with this cross pertaining to water relations are strong candidates to be loci that have differentiated adaptively in response to climate-driven selection. Interesting, McKay et al. (2008) find QTL in the same region as flowering time QTL, echoing past suggestions that flowering time in *A. thaliana* may have evolved as a correlated response to selection on water relations traits (McKay et al. 2003; Stinchcombe et al. 2004). Also working with *A. thaliana*, Samis et al. (2008) use a complementary approach: rather than studying the genetic differences between a focal pair of accessions using crosses, they take an association-mapping approach that uses many natural accessions. Taking advantage of candidate gene information about photoperiod responses in *A. thaliana*, they test for latitudinal clines in photoperiod sensitivity and the contribution of *PHYC*. Interestingly, they find strong evidence for longitudinal clines, suggesting that the critical photoperiod that accurately predicts the onset of winter temperatures varies across Europe.

Li et al. (2008) take up Rockman's call for a “reverse ecology,” and apply it to three organisms for which experimental ecology is quite difficult (humans, yeast, and fruit flies) with the goal of identifying loci that underlie adaptive evolutionary change. Li et al. (2008) compare three commonly applied methods for detecting natural selection from sequence data, the McDonald-Kreitman test, the mkprf method, and likelihood methods for determining if $dn/ds > 1$. They use three population genomics datasets (*H. sapiens*, *S. cerevisiae*, and *D. simulans*). In addition to providing explicit data on the statistical power of commonly used methods for detecting selection from polymorphism and divergence data, Li et al. (2008) detail the subtle differences between

the types of data used by these three tests, and the potentially dramatic sensitivity of the results to the starting statistical assumptions.

Fishman et al. (2008) and Moyle (2008) provide examples of the prospects of ecological genomics in emerging model systems. For example, Fishman et al. (2008) investigated the genetic architecture of conspecific pollen precedence in F_2 and hybrid backcross mapping populations between *Mimulus guttatus* and *M. nasutus*. Conspecific pollen precedence, or the competition between pollen genotypes that leads to a transmission advantage for one species or genotype in mixed pollinations, is an important contributor to the evolution of reproductive isolation in plants. Fishman et al. (2008) take advantage of the newly developed genomic resources for *Mimulus* (nearly isogenic lines, EST libraries, physically and genetically mapped markers, a whole genome sequence) to move one step closer to identifying the genetic basis of gametophytic divergence in plants.

Moyle (2008) details the tremendous opportunities available for tapping into *Lycopersicon* (wild tomatoes) for studies in ecological and evolutionary genomics. The close evolutionary relationships between wild *Lycopersicon* species and cultivated tomatoes suggest that moving genomic, biochemical, and molecular tools into wild species will not be an insurmountable challenge. Importantly, wild *Lycopersicon* species span strong environmental gradients and have extensive historical occurrence records (Rick 1963). Moyle (2008) makes a convincing case that the incomplete reproductive barriers between species make *Lycopersicon* an ideal group to study speciation in plants. Other promising areas include plant–herbivore and plant–pathogen interactions, as well as adaptation to abiotic climatic variation. As Moyle (2008) points out, such studies would be greatly facilitated by international collaborations and projects that supported in situ ecological and genetic research.

Future Directions

Although the studies included in the symposium and the special feature are promising forays into answering the challenge described by Clausen and Hiesey 50 years ago, no one would argue that we have sufficient data to address the longstanding questions that can only be resolved once the loci and alleles for ecologically important traits are in hand. How, then, do we get from where we are now to meeting those lofty goals?

A clear future direction emphasized in the symposium and in the papers is the need to complement genomic and molecular approaches with “traditional” evolutionary ecology. For many model systems, this will entail more fieldwork in the tradition of Clausen, Keck and Hiesey: ecological experimentation, and in situ measurements of natural selection, across time, space, and habitats. These experiments are logistically challenging and difficult

to execute, but the potential ecological context that they provide is invaluable. A corollary to this view is the need to expand genomic resources and tools beyond the limited life histories, ecologies, and taxonomic units currently at our disposal. As the cost of sequencing and expression studies continues to fall, the net effect of these limitations should be reduced. Finally, these problems need to be attacked by interdisciplinary teams. In the past 50 years, acclimation and plasticity have been studied by molecular biologists and physiologists in the absence of an evolutionary context. Those in the field of ecological genomics have studied adaptation and selection in different environments, but lacked molecular and physiological approaches to understand how the organisms are responding plastically. Merging of these approaches can answer a final question, as organisms are moving to new environments do the plastic responses of traits or transcripts predict the evolutionary responses?

LITERATURE CITED

- Balanyá, J., J. M. Oller, R. B. Huey, G. W. Gilchrist, and L. Serra. 2006. Global genetic change tracks global climate warming in *Drosophila subobscura*. *Science* 313: 1773.
- Clausen, J., and W. M. Hiesey. 1958. Experimental studies on the nature of species. IV. Genetic structure of ecological races. Carnegie Institution of Washington Publication 615. Washington, DC.
- Clausen, J., D. D. Keck, and W. M. Hiesey. 1940. Experimental studies on the nature of species. I. Effect of varied environments on North American plants. Carnegie Institute of Washington Publication 520. Washington, DC.
- Clausen, J., D. D. Keck, and W. M. Hiesey. 1948. Experimental studies on the nature of species. III. Environmental responses of climatic races of *Achillea*. Carnegie Institute of Washington Publication 581. Washington, DC.
- Eads, B., J. Colbourne, E. Bohuski, and J. Andrews. 2007. Profiling sex-biased gene expression during parthenogenetic reproduction in *Daphnia pulex*. *BMC Genom.* 8:464.
- Falconer, D. S., and T. F. C. MacKay. 1996. Introduction to quantitative genetics. 4th ed. Longman, Essex.
- Fishman, L., J. Aagaard, and J. C. Tuthill. 2008. Towards the evolutionary genomics of gametophytic divergence: patterns of transmission ratio distortion in monkeyflower (*Mimulus*) hybrids reveal a complex genetic basis for conspecific pollen precedence. *Evolution*, in press.
- Kopp, A., O. Barmina, A. M. Hamilton, L. Higgins, L. M. McIntyre, and C. D. Jones. 2008. Evolution of gene expression in *Drosophila* olfactory system. *Mol. Biol. Evol.* 25:1081–1092.
- Li, Y. F., J. C. Costello, A. K. Holloway, and Matthew W. Hahn. 2008. “Reverse ecology” and the power of population genomics. *Evolution*, in press.
- McKay, J. K., J. H. Richards, and T. Mitchell-Olds. 2003. Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Mol. Ecol.* 12:1137–1151.
- McKay, J. K., J. H. Richards, K. S. Nemali, S. Sen, T. Mitchell-Olds, S. Boles, E. A. Stahl, T. Wayne, and T. E. Juenger. 2008. Genetics of drought adaptation in *Arabidopsis thaliana* II. QTL analysis of a new mapping population, Kas-1 x Tsu-1. *Evolution*, in press.
- Morgan, T. J. and Mackay, T. F. C. 2006. Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity* 96:232–242.
- Moyle, L. C. 2008. Ecological and evolutionary genomics in the wild tomatoes (*Solanum* sect. *Lycopersicon*). *Evolution*, in press.

- Orr, H. A. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949.
- Orr, H. A. 2005. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* 6:119–127.
- Palopoli, M. F., M. V. Rockman, A. TinMaung, C. Ramsay, S. Curwen, A. Aduna, J. Laurita, and L. Kruglyak. 2008. Molecular basis of the copulatory plug polymorphism in *Caenorhabditis elegans*. *Nature* 454:1019–1022.
- Rick, C. M. 1963. Barriers to interbreeding in *Lycopersicon peruvianum*. *Evolution* 17:216–232.
- Samis, K. E., K. D. Heath, and J. R. Stinchcombe. 2008. Longitudinal clines in flowering time and PHYTOCHROME C in *Arabidopsis thaliana*. *Evolution*, in press.
- Seidel, H. S., M. V. Rockman, and L. Kruglyak. 2008. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* 319:589–594.
- Stinchcombe, J. R., C. Weinig, M. Ungerer, K. M. Olsen, C. Mays, S. S. Halldorsdottir, M. D. Purugganan, and J. Schmitt. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* is modulated by the flowering time gene *FRIGIDA*. *Proc. Natl. Acad. Sci. USA* 101:4712–4717.