American Journal of Botany

American Journal of Botany 99(2): 312–319. 2012.

# TARGETED SEQUENCE CAPTURE AS A POWERFUL TOOL FOR EVOLUTIONARY ANALYSIS[1]

CORRINNE E. GROVER[2,4], ARMEL SALMON[3], AND JONATHAN F. WENDEL[2]

[2]Department of Ecology, Evolution, & Organismal Biology, Iowa State University, Ames, Iowa 50011 USA; and [3]Evolution des Genomes et Speciation (Equipe MOB); UMR CNRS 6553 Ecobio—CAREN, Universite de Rennes 1; Bat. 14A, Campus de Beaulieu; 35 042 Rennes Cedex (France)

Next-generation sequencing technologies (NGS) have revolutionized biological research by significantly increasing data generation while simultaneously decreasing the time to data output. For many ecologists and evolutionary biologists, the research opportunities afforded by NGS are substantial; even for taxa lacking genomic resources, large-scale genome-level questions can now be addressed, opening up many new avenues of research. While rapid and massive sequencing afforded by NGS increases the scope and scale of many research objectives, whole genome sequencing is often unwarranted and unnecessarily complex for specific research questions. Recently developed targeted sequence enrichment, coupled with NGS, represents a beneficial strategy for enhancing data generation to answer questions in ecology and evolutionary biology. This marriage of technologies offers researchers a simple method to isolate and analyze a few to hundreds, or even thousands, of genes or genomic regions from few to many samples in a relatively efficient and effective manner. These strategies can be applied to questions at both the infra- and interspecific levels, including those involving parentage, gene flow, divergence, phylogenetics, reticulate evolution, and many more. Here we provide a brief overview of targeted sequence enrichment, and emphasize the power of this technology to increase our ability to address a wide range of questions of interest to ecologists and evolutionary biologists, particularly for those working with taxa for which few genomic resources are available.

**Key words:** next-generation sequencing; phylogenetics; population genetics; sequence capture.

Next-generation sequencing (NGS) technologies have rapidly become transformative in many areas of biology. The ability to generate massive amounts of sequence data quickly and cheaply has been considered as revolutionary as the introduction of PCR, "with one's imagination being the primary limitation to its use" (Metzker, 2010, p. 31). Recent reviews and articles have focused not only on the types of NGS, but also outlined some of their potential uses in arenas as diverse as genome sequencing (Weigel and Mott, 2009; Argout et al., 2011), evolutionary genomics (Gilad et al., 2009), genomic ecology (Hudson, 2008; Ekblom and Galindo, 2011), gene expression (Wang et al., 2009; Grabherr et al., 2011; Ozsolak and Milos, 2011), epigenetics (Laird, 2010; Schmitz and Zhang, 2011), functional genomics (Morozova and Marra, 2008), population-level analyses (Davey and Blaxter, 2010; Pool et al., 2010; Bansal et al., 2011), crop breeding (Varshney et al., 2009), drug discovery/development (Woollard et al., in press), cancer or disease genetics (Schweiger et al., 2011; Shen et al., 2011), and ancient DNA analyses (Wolinsky, 2010). NGS has greatly increased our ability to attain the previously unattainable, and this newfound capability has been transformative for ecological and evolutionary research, particularly for nonmodel systems that lack extensive genomic resources (e.g., genome sequences, EST libraries, BAC libraries). Freed from the previous constraints imposed by absence of genomic resource, ecologists and evolutionary biologists interested in even the most obscure taxa can begin to dream up powerful new experiments to enhance their research objectives.

[1]Manuscript received 13 July 2011; revision accepted 28 October 2011.

[4]Author for correspondence (e-mail: corrinne@iastate.edu)

doi:10.3732/ajb.1100323

This is not to suggest that all biological questions require genome-scale data; in many cases, experimental objectives are readily pursued by (1) targeting specific genes or genomic regions rather than by surveying the entire genome or specific transcripts of interest and (2) improving sequencing depth of the regions targeted, especially for complex genomes. But in many situations, both infraspecific-level questions (e.g., phylogeography, gene flow, parentage) and interspecies-level questions (e.g., divergence, phylogenetics, reticulate evolution) become more effectively addressed using the relatively inexpensive, high-throughput sequencing afforded by NGS, provided the sequencing is targeted to genes or regions of interest. Recent developments have led to several strategies that enable targeted enrichment of genomic fractions specifically for NGS, each of which have their own benefits and limitations (Baird et al., 2008; Summerer, 2009; Mamanova et al., 2010; Elshire et al., 2011). Here we provide a synopsis of how targeted genomic enrichment coupled with NGS can create a powerful toolset for a wide range of questions of interest to ecologists and evolutionary biologists and provide examples of the types of biological questions targeted sequence enrichment is well suited to address.

## TARGETED SEQUENCE ENRICHMENT: CONCEPT AND TECHNOLOGIES

Targeted sequence enrichment refers to the suite of technologies designed to isolate a specific genomic fraction (e.g., genes, molecular markers, larger genomic regions) for subsequent NGS, ultimately resulting in an enriched pool of target sequences such that there is overall reduction in the genomic sequencing space, and hence greater sequence coverage for each targeted region. Plant genomes tend to be large, containing many highly

repetitive sequences, and thus many of the sequences returned by randomly sheared and prepared whole-genome NGS libraries are derived from this interesting, but often nonpertinent genomic fraction. The reduction in sequencing space offered by target enrichment strategies confers three important advantages over randomly sequencing the entire genome. First, by reducing the sequencing space required per sample, multiplexing of samples becomes feasible, reducing the overall cost of sequencing per sample and thereby permitting more samples to be sequenced for the same cost. Second, by targeting only the fraction of the genome that is either necessary (e.g., where a specific number of genes is required for adequate genomic coverage) or informative (e.g., genes of a specific pathway) for the biological question being addressed, the complexity of the analysis is significantly reduced. Third, given the redundancy of plant genomes, which typically include myriad gene duplications, both recent and ancient (Van de Peer et al., 2009; Jiao et al., 2011), and the fact that all plants have experienced multiple rounds of polyploidy (Jiao et al., 2011), the depth of sequencing afforded by targeted NGS increases the possibility of identifying both the precise region (orthologs) of interest and its paralogs in population and infraspecific genomics assays.

Several technologies exist for target enrichment, which can be classified by the mode of enrichment: (1) hybridization-based sequence capture, (2) PCR-based amplification, and (3) molecular inversion probe-based amplification. The merits of each of these have been discussed in depth (Mamanova et al., 2010) so need not be repeated here. Of the currently available technologies, sequence capture has the advantage of being extraordinarily quick and simple, while remaining relatively inexpensive. While PCR amplification, multiplexed or uniplexed and pooled prior to sequencing, is perhaps a more familiar and comfortable method of target enrichment, it is time consuming and laborious, as well as more expensive on a per-sequence-generated basis. Molecular inversion probes (MIPS) circumvent some of the labor and expense associated with PCR selection; however, MIPs still require experimental optimization, and it may not be possible to capture all targeted regions under uniform conditions (Porreca et al., 2007; Krishnakumar et al., 2008; Mamanova et al., 2010).

Sequence capture is a technology that permits the isolation of targeted loci from the background of the entire genome. The scale of the capture can range from several targeted loci to over a million target regions (Agilent, 2011; Mycroarray, 2011; NimbleGen, 2011), making it adaptable for both small-scale and large-scale projects. Many protocols exist for hybridization-based sequence capture. Some of these have been developed in such a way that the entire protocol can be accomplished in a single standard research laboratory, from probe design through sequence capture (Albert et al., 2007; Hodges et al., 2007; Okou et al., 2007; Porreca et al., 2007; Gnirke et al., 2009; Ng et al., 2009; Turner et al., 2009; Zheng et al., 2009). While many technological alternatives are available for sequence capture, the two primary approaches involve hybridization of samples either to microarrays or to solution-based, pooled RNA-baits, both of which are complementary to the targeted genes. In both approaches, all targeted sequences are captured in one hybridization reaction, and, since samples can be multiplexed using appropriate barcoding of NGS adapters, the potential exists to capture thousands of sequences simultaneously. All of these approaches have the advantage of being inexpensive once established, yet may be more expensive at the outset than, say, low coverage whole-genome shotgun sequencing, with respect to costs associated with necessary start-up materials and with troubleshooting. Sequence capture solutions are also available through several companies, who provide probe design and synthesis services accompanied by a streamlined protocol suitable for laboratory use (Agilent, 2011; Mycroarray, 2011; NimbleGen, 2011). These have the added advantage of minimizing start-up trial and error, and, consequently, both cost and time. As with any technique, key elements must be taken into consideration when designing sequence capture experiments, including (1) the number of targets and the mean depth of coverage needed for each target, (2) the probe specificity, (3) the biological system studied, (4) the expected enrichment efficiency, and (5) the NGS technology used. As mentioned, sequence capture is fully scalable, making it useful for experiments requiring few genes or genomic regions, as well as those involving entire exomes. There are several obvious tradeoffs required in experimental design, between, for example, the total number of targets and depth of coverage for any given amount of NGS sequencing; an increase in the number of targets will require either more sequencing space or will result in decreased sequencing coverage on a per target basis. The mean coverage depth required will vary by target, depending upon the specificity of the probe or probes used for that target, the number of closely related sequences (orthologs and paralogs) in the genome, and level of heterozygosity; both probe characteristics (e.g., conserved domains vs. unique regions) and the biological system being studied (e.g., ploidy or genome size) will affect the number of different copies represented in the enriched pool, each of which will occupy part of the sequencing space. These factors, combined with the assumed enrichment efficiency (either calculated from prior captures or assumed from the literature) will determine the amount of sequencing required and will influence the degree of multiplexing possible for the experiment. For example, if the required mean coverage depth per 1 kb target is 20×, the total capture space required for that target is 20 kb. Given an assumed capture efficiency of 60% (Hodges et al., 2007; Gnirke et al., 2009; Hedges et al., 2011), the total sequencing effort required for that target alone is ca. 33 kb, if it is single copy. If paralogs or variant alleles are expected, this should be adjusted for each additional sequence type.

The target probe design itself is similar to the design of microarray probes, for which there is much literature and software, and as with microarrays, most users will turn to an established company for probe design and bait synthesis. For those working on organisms that currently possess few genomic resources, previously generated sequence data can be considered as targets for sequence capture. For those organisms for which no appropriate data currently exists, data can be quickly and easily generated through a variety of NGS strategies aimed at marker development, including, and probably most usefully, RNA-seq. By sequencing a single Illumina lane from one or more species, a wealth of data can be generated for any organism from which RNA can be obtained. For those whose taxon of choice cannot supply adequate amounts of RNA (e.g., rare samples, herbarium samples), anonymous sequencing methods targeting mostly euchromatic regions (e.g., genotyping by sequencing, GBS; see below) can provide sequence from which baits can be designed. A final consideration in preparing for sequence capture is the type of NGS technology applied to the experiment. The newest Roche-454 platform (GS FLX Titanium XL+) produces 700 megabases of sequence per plate (Roche-454, 2011), whereas the newest Illumina platform (HiSEq. 2000) is capable of producing significantly more at 300 gigabases per flow cell

(Illumina, 2011); however, the read length produced by the GS FLX is significantly longer than the HiSEq. 2000 (1000 bp vs. 100 bp). This creates a trade-off between sequence coverage per target (or number of targets multiplexed) and the ease and accuracy of assembly; that is, while the amount of data produced by the Illumina HiSEq. 2000 is significantly greater, it may prove challenging, or in some cases nearly impossible, to accurately assemble extremely similar sequences (e.g., close homoeologs or paralogs) from such short reads.

## APPLICATIONS

***Ecological and evolutionary questions within species—*** Ecological and evolutionary questions that rely on evaluating population structure and diversity within species often have been constrained by the extent and quality of the data that can be applied to any given question. The ability to obtain extensive nuclear sequence from large numbers of loci and organisms has been referred to as the "Holy Grail in molecular ecology and evolution" (Avise, 2010, p. 665). Sequence capture has the unique ability to allow ecologists and evolutionary biologists to quickly and easily evaluate sequence polymorphisms and divergence for hundreds, or even thousands, of genes, gene families, and genomic regions, and apply those results to inferences of parentage, gene flow, population divergence, phylogeography, diversity, domestication and improvement, and many other questions. This massive increase in attainable data can be gathered for any species, model or nonmodel, offering greatly enhanced precision and, possibly insight, into biological questions.

One of the most widespread marker systems used today, for a variety of applications, are microsatellite loci, which have a long history of productive use in ecology and evolutionary biology, but which also require time and some degree of genomic resources to develop for each species. In addition, locus number often can be limiting, and polymorphism levels need to remain below the threshold at which homoplasy becomes problematic. Amplified (or restriction) fragment length polymorphism (AFLP or RFLP) and sequence-specific amplified polymorphism (SSAP) analyses have higher-throughput and have the potential to capture a greater amount of polymorphism, but these tools suffer from problems inherent in translating gel-based phenotypes into accurate genotypes. Sequence capture obviates these problems while simultaneously providing access to thousands of nucleotide level polymorphisms (Barbazuk et al., 2007; Gilad et al., 2009; Shen et al., 2010). In addition, the nature of sequence capture is such that previously successful markers derived from lower-throughput technologies can be included in the target enrichment design, while simultaneously allowing new markers to be incorporated. This shifts the scale of population level analyses from gene to genome and the range of applicability from model to all species.

The application of large-scale sequence capture assays at the population level promises to enhance plant molecular population genetic (and now, population genomic) analyses of diversity, phylogeography, gene flow, and any number of studies designed to explore the genetic underpinnings of phenotypes in natural settings and the forces that shape those phenotypes. Because sequence capture generates nucleotide data sets as opposed to gel-based phenotypes (as with microsatellites and AFLP analyses, for example), a large body of population genetic theory is applicable, so statistical methods for approaching many questions will already be mature and readily applied. As an example, pioneering studies on plant domestication have demonstrated that genes underlying traits that have experienced strong selection comprise a relatively small fraction of genes in the genome and that these genes are often inferred from their exceptionally low levels of genetic diversity (Wang et al., 1999; Tenaillon et al., 2001, 2004). More generally, genes under directional selection, such as those involved in domestication, are mainly detected through three methods that rely on measures of nucleotide diversity (Wright and Gaut, 2005; Burger et al., 2008; Siol et al., 2010): (1) the overall amount of nucleotide diversity (positive selection leading to a reduction in nucleotide diversity), (2) the frequency distribution of diversity, and (3) the patterns of linkage disequilibrium (LD). The use of targeted sequencing methods enhances all three of these important parameters that control the power of inference provided by these methods: the number of individuals sampled, which is increased with higher-order multiplexing, the number of genes evaluated, and the improved power of estimating and locating higher LD regions because of higher saturation of genomic coverage. Hence, the broader and deeper sampling of nucleotide diversity afforded by sequence capture increases the power of these and related experiments.

It is worth noting that other methods are currently being adopted for high throughput marker discovery using NGS technologies, e.g., restriction-site-associated DNA (RAD) sequencing (Baird et al., 2008) and genotyping-by-sequencing (GBS; Elshire et al., 2011), both of which generate markers associated with genomic restriction sites in an attempt to secure similar genomic fractions from different species or accessions to screen for polymorphic markers. Both of these techniques have some of the advantages of sequence capture, i.e., utilization of the low cost, high output nature of NGS and the richness that sequencing data provides; however, neither has the ability to significantly reduce the sequencing space that sequence capture does. Recent research utilizing RAD sequencing or GBS initiated with over 1 million potential markers, which then were reduced to only those that could be mapped to a single genomic location and are polymorphic (Baird et al., 2008; Elshire et al., 2011). While mapping is not strictly necessary and is not possible for taxa lacking robust maps or gene sequences, inferences of orthology among the sequenced regions are most confidently made when using a reference genome. This same logic holds for sequence capture, but the focused targeting of genes simplifies the process of orthology inference.

Although currently there are few examples that have used sequence capture, recent research in maize, wheat, and pine have demonstrated the potential of the technology. One of the first applications of sequence capture to a large, complex plant genome was the resequencing of an ca. 2.3 Mb chromosomal interval from maize, while in the same capture targeting 43 specific genes (Fu et al., 2010). This experiment was designed primarily to evaluate the utility of sequence capture in facilitating resequencing of complex genomes; however, during the study, more than 2500 SNPs were identified between the two maize lines used. Briefly, NGS libraries from the maize lines B73 and Mo17 were applied to a sequence capture array derived solely from maize line B73. Target enrichment was successful for 80–98% of the targeted bases, facilitating the high level of SNP identification reported. Furthermore, the authors note that paralogous sequences were also present in the capture, an added benefit to those interested in gene family characterization.

A recent paper in wheat also demonstrated the effectiveness of sequence capture on evaluating diversity, this time between

wild and domesticated accessions (Saintenac et al., 2011). Because the wheat genome is even larger and more complex than that of maize (10 Gb for tetraploid wheat and 16 Gb for hexaploid wheat), resequencing of the wheat genome to evaluate diversity is simply not practical; thus, the authors employed sequence capture to resequence a targeted 3.5 Mb region of exonic sequence from both wild and domesticated wheat genomes. From the targeted region, the rate of target recovery ranged from 79–88% in cultivated and wild wheat, respectively, and ~63% of targets were sufficiently covered in both wheat lines to allow comparison between the two (Saintenac et al., 2011). From the data generated, the authors identified more than 4000 SNPs distinguishing the wild and domesticated wheat lines; over 14 000 SNPs distinguishing the component genomes of wheat and shared between the lines; and 129 small-scale indels. In addition, the authors were able to detect both copy number variation and presence/absence variation. It is worth noting that the baits used to facilitate enrichment in this experiment were designed from only one homoeologous copy of each gene represented, yet the experiment captured the divergent homoeologs, which attests to the ability of sequence capture to successfully enrich for a given set of targets, even across divergent genomes.

As a final example, two association studies employing sequence capture using cross-species baits are currently underway in pine, with initial results promising success in both SNP discovery and genetic trait association (Kirst et al., 2011). These studies hold promise for improving breeding design in these long-lived and economically important perennials.

***Ecological and evolutionary questions among species and higher taxa—***Much of our understanding in evolution (e.g., the evolution of a character, trait, behavior, genomic property, species relationships) relies on our ability to make meaningful comparisons among species. This, in turn, is dependent on our ability to generate data that are appropriate to the context of the question. For many reasons, including complexities in both genomic and organismal-level histories, attempts to efficiently recover orthologous loci with which to conduct interspecific comparisons can be challenging. Early attempts at phylogenetics and interspecific comparisons on the molecular level sought out genomic components that were, by their nature, less complicated to obtain and sequence (e.g., ITS, nuclear rDNA, chloroplast loci), but as a consequence, often less informative or sometimes even misleading (Álvarez and Wendel, 2003; Small et al., 2004). The nuclear genome is much preferred for resolving relationships, although it is not without complication (Doyle, 1992; Cronn et al., 2002; Small et al., 2004; Hughes et al., 2006; Philippe et al., 2011).

Perhaps the most prohibitive complication in using nuclear markers, which has impeded their adoption on a broad scale, is the ability to generate useful data. Because nuclear DNA can evolve rapidly, conserved primers for low-copy nuclear markers can be difficult to obtain due to the need for substantial existing sequence (e.g., EST sequence data) or wide comparisons utilizing degenerate primers. The potential for any given nuclear locus to become unusable can be high, either due to the inability to amplify it from all species or the level of sequence divergence relative to the taxonomic question being addressed, leading some to address the question of the best method for obtaining phylogenetically useful nuclear markers (Xu et al., 2004; Schlüter et al., 2005; Whittall et al., 2006; Álvarez et al., 2008; Steele et al., 2008). The advent of sequence capture has largely surmounted the hurdles encountered in attempting to generate large amounts of orthologous sequence data. Using baits that target tens to thousands of genes, sequence capture can enrich next-generation sequencing libraries for not only the targeted loci, but also for any existing close paralogs that could complicate subsequent analyses, if unwittingly excluded from the sample. Since the capture baits used to enrich the sequencing library may be used across taxa (at low levels of divergence), a priori sequence information is required for only one taxon; this can readily be obtained from existing sequencing information or generated through analysis of cDNA libraries (Denoeud et al., 2008; Nagalakshmi et al., 2008; Gibbons et al., 2009; Hittinger et al., 2010). Importantly, because the use of capture baits results in capture targets that are larger than the baits themselves (due to extension of captured DNA fragments beyond both ends of the bait sequence), intron information is not required a priori. Instead, in plants, where most introns are generally about the same size as exons, the intronic sequence for each gene will be recovered even if the bait set is designed from coding sequence alone. This key and important advantage of sequence capture technology, i.e., the ability to recover target sequence without any a priori knowledge, would have been unfathomable to those working with nonmodel systems (i.e., most of us) until recently.

Even when employing nuclear-derived markers, resolving questions regarding phylogenetic relationships is complex at best, as they often are complicated by current and historical hybridization and introgression, cryptic polyploidy, unknown paralogy, and incomplete lineage sorting. The best methods for resolving relationships, in light of these complications, is a topic of great interest (Sang and Zhong, 2000; Barton, 2001; Holland et al., 2004; Linder and Rieseberg, 2004; Degnan and Salter, 2005; Nakhleh et al., 2005; Huson and Bryant, 2006; Maddison and Knowles, 2006; Sanderson and McMahon, 2007; Holland et al., 2008; Brito and Edwards, 2009; Degnan and Rosenberg, 2009; Liu et al., 2009; Meng and Kubatko, 2009; Joly et al., 2009; Burleigh et al., 2011; Griffin et al., 2011), and from these discussions, several aspects of sampling have been highlighted as important. Because sequence capture is designed to facilitate sequencing of large numbers of loci from many samples, issues of sampling become lessened, presumably leading to improved power for phylogenetic inference. Multiple loci typically are required for phylogenetic reconstruction because any given locus could have a history of complex evolution that leads to contradictory gene and "species" trees. Likewise, taxonomic sampling can have a significant effect on the inference of relationships due to phenomena such as long-branch attraction and unexpectedly high infrataxon variation, for example. The traditional method of PCR followed by Sanger sequencing quickly becomes laborious as taxa are added, especially when multiple loci are used or when cloning is required, whereas the increase in labor for sequence capture is minimal; each additional taxon merely adds one additional sequencing library and one additional hybridization. In addition to these sampling considerations, unintentional sampling of paralogs as orthologs for one or more taxa can significantly affect phylogenetic inferences; here again, sequence capture can help minimize the problem by easily acquiring both orthologs and close paralogs to those loci targeted, in the process facilitating their discrimination.

It bears mention that the efficiency of sequence capture at higher levels of divergence (e.g., among genera or families) has yet to be evaluated in plants; however, cross-species sequence capture has successfully been applied in mammals (Burbano et al., 2010;

George et al., 2011), where the bulk of the sequence capture literature currently resides. One might expect that both capture efficiency and coverage should decrease as divergence increases, in a fashion similar to that observed in cross-species microarrays (Bar-Or et al., 2007; Lu et al., 2009; Nazar et al., 2010); however, given the large number of loci that can be included in probe design, it is likely that the dropout rate of useful probes will be nonlinear, related to the degree of divergence and hence variable among genes, and that because of this a subset of probes will prove useful at the level of the genus and perhaps above, for particularly slowly evolving genes. It is highly likely that for most species, levels of genic divergence among individuals and populations are low enough that sequence capture will be effective for most targeted genes. Exceptions include cases of rapid gene loss, although this would be discovered by absence of target recovery.

***Evolutionary (epi)genomics*—**Plant species that become widely studied are not random; instead, economically (e.g., crop species) and ecologically (e.g., weedy or invasive species) important species clearly attract preferential research attention. As such, the identification and characterization of the underlying genetic components of specific traits of economic importance often is the research focus. Many methods exist for identifying genes controlling traits of interest, including nucleotide diversity scans that detect bottlenecks in genetic diversity concomitant with natural selection or domestication (Doebley et al., 2006; Gross and Olsen, 2010), which, as mentioned, can now be facilitated through the coupling of sequence capture and NGS. Other methods, including positional cloning and comparative differential expression analyses, may identify genes that differ between taxa that exhibit the phenotypic differences under study. Of course, individual genes rarely act in isolation, being instead components of integrated pathways and networks. Thus, to understand all but the simplest of phenotypes, entire pathways become focal points of interest. Sequence capture and NGS, coupled with pathway information from related model species, are well-suited to facilitate comparative pathway and network analysis. By taking sequence information derived from genomic resources such as EST collections and/or genomic sequence data, and comparing it to pathway information derived from the Plant Metabolic Network database (Plant Metabolic Network and TAIR, 2011), sequence capture baits can be designed for most known pathway members, which can then be used on multiple related species, ultimately allowing for comparative pathway analysis. As an example, the flowering time network in *Gossypium* is of agronomic interest because domesticated cottons are day-length neutral, whereas the wild forms are photoperiod sensitive. The transformation to day-length neutrality could involve any number of genes from the flowering time network, which have not been described for *Gossypium*. Using the *Arabidopsis* Metabolic Network (Plant Metabolic Network and TAIR, 2011), cotton homologs to those described in the *Arabidopsis* flowering time network were isolated from cotton-specific EST libraries, and the resulting ESTs were used to design sequence baits. Use of these baits may prove useful in describing changes to the flowering time network in day-length sensitive and day-length neutral cottons (C. Grover et al., unpublished data), at least at the genomic level. Just as the evolution of a gene is better understood in the context of the pathways in which it is involved (Lu and Rausher, 2003; Rausher et al., 2008), the evolution of a trait needs to incorporate understanding of genic change at both the DNA and gene expression levels. Next-generation sequencing and RNA-seq have made data acquisition for comparative transcriptomics routine, and the coming years will likely see continued growth in the application of comparative transcriptomics to questions in ecology and evolution.

One of the exciting frontiers in ecology and evolution is the interaction between ecological and evolutionary properties and heritable epigenetic marks (Richards and Wendel, 2011; Scoville et al., 2011). It has become clear that gene expression is at least partially controlled by heritable epigenetic marks, which accordingly become traits of interest themselves, traits that may also be targeted through modified protocols of targeted sequence enrichment. In most eukaryotes, the interplay of epigenetic marks, such as DNA methylation, covalent histone modifications, and small RNAs, regulate chromatin conformation and genome expression (Henderson and Jacobsen, 2007; Zhang, 2008). Among these epigenetic marks, DNA methylation is the most commonly studied, historically detected using several complementary approaches, such as bisulfite sequencing, methylation-sensitive restriction enzyme analysis, methyl-cytosine affinity or immunoprecipitation, and microarray technologies (Zhang, 2008). Similarly, histone modifications have also been evaluated through chromatin immunoprecipitation followed by microarray analysis. Coupling epigenetic analysis with NGS has been of great interest, with several strategies already outlined (Cokus et al., 2008; Down et al., 2008; Park, 2008; Zhu, 2008; Laird, 2010; Satterlee et al., 2010; Varley and Mitra, 2010), capture-based sequencing among them (Hodges et al., 2009; Nautiyal et al., 2010). As capture-based target enrichment is similar in principle to microarray technologies, any comparative analyses previously conducted using microarrays should be adaptable to a sequence capture and NGS protocol, which will quickly produce high-resolution data for epigenetic analysis.

## CONCLUSIONS AND PROSPECTS

The goals of evolutionary biology are to understand the history and diversity of life and the forces that shape this diversity. The advent of sequence capture has opened up avenues of research that were, in the past, prohibitive for nonmodel taxa, and that will increase the resolving power of many kinds of studies. We now have the power to analyze, in detail and with high resolution, the population structure and dynamics of virtually any species, allowing increased insight into questions concerning parentage, gene flow, population divergence, phylogeography, diversity, domestication and improvement, the detection of selection, and many other arenas. Likewise, the coupling of sequence capture and NGS will allow in-depth and detailed assessments of phylogenetic relatedness, taxon or hybrid identification, instances of historical introgression, polyploid parentage, and so much more. Finally, the vastly enhanced power of sequence capture technologies, relative to the more conventional "gene at a time" approaches, promises to yield new insights into many aspects of plant evolutionary and ecological genomics, as well as controlling epigenetic marks. For those fortunate few working on model organisms whose genome sequences have been available for years, the methodological advance of NGS and sequence capture has represented a significant leap forward in technical capacity; for all others, however, who study organisms with more limited genomic resources, this technology is revolutionary.

## LITERATURE CITED


AGILENT. 2011. SureSelect Target Enrichment System Kits website http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Product&SubPageType=ProductDetail&PageID=1659.

ALBERT, T. J., M. N. MOLLA, D. M. MUZNY, L. NAZARETH, D. WHEELER, X. SONG, T. A. RICHMOND, ET AL. 2007. Direct selection of human genomic loci by microarray hybridization. *Nature Methods* 4: 903–905.

ÁLVAREZ, I., A. COSTA, AND G. FELINER. 2008. Selecting single-copy nuclear genes for plant phylogenetics: A preliminary analysis for the Senecioneae (Asteraceae). *Journal of Molecular Evolution* 66: 276–291.

ÁLVAREZ, I., AND J. F. WENDEL. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.

ARGOUT, X., J. SALSE, J.-M. AURY, M. J. GUILTINAN, G. DROC, J. GOUZY, M. ALLEGRE, ET AL. 2011. The genome of *Theobroma cacao*. *Nature Genetics* 43: 101–108.

AVISE, J. 2010. Perspective: Conservation genetics enters the genomics era. *Conservation Genetics* 11: 665–669.

BAIRD, N. A., P. D. ETTER, T. S. ATWOOD, M. C. CURREY, A. L. SHIVER, Z. A. LEWIS, E. U. SELKER, ET AL. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: e3376.

BANSAL, V., R. TEWHEY, E. M. LEPROUST, AND N. J. SCHORK. 2011. Efficient and cost effective population resequencing by pooling and in-solution hybridization. *PLoS ONE* 6: e18353.

BAR-OR, C., H. CZOSNEK, AND H. KOLTAI. 2007. Cross-species microarray hybridizations: A developing tool for studying species diversity. *Trends in Genetics* 23: 200–207.

BARBAZUK, W. B., S. J. EMRICH, H. D. CHEN, L. LI, AND P. S. SCHNABLE. 2007. SNP discovery via 454 transcriptome sequencing. *Plant Journal* 51: 910–918.

BARTON, N. H. 2001. The role of hybridization in evolution. *Molecular Ecology* 10: 551–568.

BRITO, P., AND S. EDWARDS. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* 135: 439–455.

BURBANO, H. A., E. HODGES, R. E. GREEN, A. W. BRIGGS, J. KRAUSE, M. MEYER, J. M. GOOD, ET AL. 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* 328: 723–725.

BURGER, J. C., M. A. CHAPMAN, AND J. M. BURKE. 2008. Molecular insights into the evolution of crop plants. *American Journal of Botany* 95: 113–122.

BURLEIGH, J. G., M. S. BANSAL, O. EULENSTEIN, S. HARTMANN, A. WEHE, AND T. J. VISION. 2011. Genome-scale phylogenetics: Inferring the plant tree of life from 18,896 gene trees. *Systematic Biology* 60: 117–125.

COKUS, S. J., S. FENG, X. ZHANG, Z. CHEN, B. MERRIMAN, C. D. HAUDENSCHILD, S. PRADHAN, ET AL. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452: 215–219.

CRONN, R., M. CEDRONI, T. HASELKORN, C. GROVER, AND J. F. WENDEL. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics* 104: 482–489.

DAVEY, J. W., AND M. L. BLAXTER. 2010. RADSeq: Next-generation population genetics. *Briefings in Functional Genomics* 9: 416–423.

DEGNAN, J. H., AND N. A. ROSENBERG. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.

DEGNAN, J. H., AND L. A. SALTER. 2005. Gene tree distributions under the coalescent process. *Evolution* 59: 24–37.

DENOEUD, F., J. M. AURY, C. DA SILVA, B. NOEL, O. ROGIER, M. DELLEDONNE, M. MORGANTE, ET AL. 2008. Annotating genomes with massive-scale RNA sequencing. *Genome Biology* 9: R175.

DOEBLEY, J. F., B. S. GAUT, AND B. D. SMITH. 2006. The molecular genetics of crop domestication. *Cell* 127: 1309–1321.

DOWN, T. A., V. K. RAKYAN, D. J. TURNER, P. FLICEK, H. LI, E. KULESHA, S. GRAF, ET AL. 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology* 26: 779–785.

DOYLE, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144–163.

EKBLOM, R., AND J. GALINDO. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107: 1–15.

ELSHIRE, R. J., J. C. GLAUBITZ, Q. SUN, J. A. POLAND, K. KAWAMOTO, E. S. BUCKLER, AND S. E. MITCHELL. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379.

FU, Y., N. M. SPRINGER, D. J. GERHARDT, K. YING, C.-T. YEH, W. WU, R. SWANSON-WAGNER, ET AL. 2010. Repeat subtraction-mediated sequence capture from a complex genome. *Plant Journal* 62: 898–909.

GEORGE, R. D., G. MCVICKER, R. DIEDERICH, S. B. NG, A. P. MACKENZIE, W. J. SWANSON, J. SHENDURE, ET AL. 2011. Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. *Genome Research* 21: 1686–1694.

GIBBONS, J. G., E. M. JANSON, C. T. HITTINGER, M. JOHNSTON, P. ABBOT, AND A. ROKAS. 2009. Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Molecular Biology and Evolution* 26: 2731–2744.

GILAD, Y., J. K. PRITCHARD, AND K. THORNTON. 2009. Characterizing natural variation using next-generation sequencing technologies. *Trends in Genetics* 25: 463–471.

GNIRKE, A., A. MELNIKOV, J. MAGUIRE, P. ROGOV, E. M. LEPROUST, W. BROCKMAN, T. FENNELL, ET AL. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182–189.

GRABHERR, M. G., B. J. HAAS, M. YASSOUR, J. Z. LEVIN, D. A. THOMPSON, I. AMIT, X. ADICONIS, ET AL. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644-652. doi:10.1038/nbt.1883

GRIFFIN, P., C. ROBIN, AND A. HOFFMANN. 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biology* 9: 19.

GROSS, B. L., AND K. M. OLSEN. 2010. Genetic perspectives on crop domestication. *Trends in Plant Science* 15: 529–537.

HEDGES, D. J., T. GUETTOUCHE, S. YANG, G. BADEMCI, A. DIAZ, A. ANDERSEN, W. F. HULME, ET AL. 2011. Comparison of three targeted enrichment strategies on the SOLiD sequencing platform. *PLoS ONE* 6: e18595.

HENDERSON, I. R., AND S. E. JACOBSEN. 2007. Epigenetic inheritance in plants. *Nature* 447: 418–424.

HITTINGER, C. T., M. JOHNSTON, J. T. TOSSBERG, AND A. ROKAS. 2010. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proceedings of the National Academy of Sciences, USA* 107: 1476–1481.

HODGES, E., A. SMITH, J. KENDALL, Z. XUAN, K. RAVI, M. ROOKS, M. ZHANG, ET AL. 2009. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Research* 19: 1593–1605.

HODGES, E., Z. XUAN, V. BALIJA, M. KRAMER, M. N. MOLLA, S. W. SMITH, C. M. MIDDLE, ET AL. 2007. Genome-wide in situ exon capture for selective resequencing. *Nature Genetics* 39: 1522–1527.

HOLLAND, B., S. BENTHIN, P. LOCKHART, V. MOULTON, AND K. HUBER. 2008. Using supernetworks to distinguish hybridization from lineage-sorting. *BMC Evolutionary Biology* 8: 202.

HOLLAND, B. R., K. T. HUBER, V. MOULTON, AND P. J. LOCKHART. 2004. Using consensus networks to visualize contradictory evidence for species phylogeny. *Molecular Biology and Evolution* 21: 1459–1461.

HUDSON, M. E. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* 8: 3–17.

HUGHES, C. E., R. J. EASTWOOD, AND C. DONOVAN BAILEY. 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philosophical Transactions of the Royal Society B. Biological Sciences* 361: 211–225.

HUSON, D. H., AND D. BRYANT. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

ILLUMINA. 2011. HiSeq 2000. Website http://www.illumina.com/systems/hiseq_2000.ilmn.

JIAO, Y., N. J. WICKETT, S. AYYAMPALAYAM, A. S. CHANDERBALI, L. LANDHERR, P. E. RALPH, L. P. TOMSHO, ET AL. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

JOLY, S., P. A. MCLENACHAN, AND P. J. LOCKHART. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *American Naturalist* 174: E54–E70.

KIRST, M., M. RESENDE, P. MUNOZ, AND L. NEVES. 2011. Capturing and genotyping the genome-wide genetic diversity of trees for association mapping and genomic selection. *BMC Proceedings* 5: I7.

KRISHNAKUMAR, S., J. ZHENG, J. WILHELMY, M. FAHAM, M. MINDRINOS, AND R. DAVIS. 2008. A comprehensive assay for targeted multiplex amplification of human DNA sequences. *Proceedings of the National Academy of Sciences, USA* 105: 9296–9301.

LAIRD, P. W. 2010. Principles and challenges of genomewide DNA methylation analysis. *Nature Reviews Genetics* 11: 191–203.

LINDER, C. R., AND L. H. RIESEBERG. 2004. Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* 91: 1700–1708.

LIU, L., L. YU, L. KUBATKO, D. K. PEARL, AND S. V. EDWARDS. 2009. Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution* 53: 320–328.

LU, Y., P. HUGGINS, AND Z. BAR-JOSEPH. 2009. Cross species analysis of microarray expression data. *Bioinformatics* 25: 1476–1483.

LU, Y., AND M. D. RAUSHER. 2003. Evolutionary rate variation in anthocyanin pathway genes. *Molecular Biology and Evolution* 20: 1844–1853.

MADDISON, W. P., AND L. L. KNOWLES. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30.

MAMANOVA, L., A. J. COFFEY, C. E. SCOTT, I. KOZAREWA, E. H. TURNER, A. KUMAR, E. HOWARD, ET AL. 2010. Target-enrichment strategies for next-generation sequencing. *Nature Methods* 7: 111–118.

MENG, C., AND L. S. KUBATKO. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theoretical Population Biology* 75: 35–45.

METZKER, M. L. 2010. Sequencing technologies—The next generation. *Nature Reviews Genetics* 11: 31–46.

MOROZOVA, O., AND M. A. MARRA. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255–264.

MYCROARRAY. 2011. MYselect: Custom bait libraries for targeted sequencing. Website http://www.mycroarray.com/myselect/index.html.

NAGALAKSHMI, U., Z. WANG, K. WAERN, C. SHOU, D. RAHA, M. GERSTEIN, AND M. SNYDER. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344–1349.

NAKHLEH, L., T. WARNOW, C. R. LINDER, AND K. S. JOHN. 2005. Reconstructing reticulate evolution in species—Theory and practice. *Journal of Computational Biology* 12: 796–811.

NAUTIYAL, S., V. E. CARLTON, Y. LU, J. S. IRELAND, D. FLAUCHER, M. MOORHEAD, J. W. GRAY, ET AL. 2010. High-throughput method for analyzing methylation of CpGs in targeted genomic regions. *Proceedings of the National Academy of Sciences, USA* 107: 12587–12592.

NAZAR, R., P. CHEN, D. DEAN, AND J. ROBB. 2010. DNA chip analysis in diverse organisms with unsequenced genomes. *Molecular Biotechnology* 44: 8–13.

NG, S. B., E. H. TURNER, P. D. ROBERTSON, S. D. FLYGARE, A. W. BIGHAM, C. LEE, T. SHAFFER, ET AL. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461: 272–276.

NIMBLEGEN. 2011. Sequence capture. Website http://www.nimblegen.com/seqcap/.

OKOU, D. T., K. M. STEINBERG, C. MIDDLE, D. J. CUTLER, T. J. ALBERT, AND M. E. ZWICK. 2007. Microarray-based genomic selection for high-throughput resequencing. *Nature Methods* 4: 907–909.

OZSOLAK, F., AND P. M. MILOS. 2011. RNA sequencing: Advances, challenges and opportunities. *Nature Reviews Genetics* 12: 87–98.

PARK, P. J. 2008. Epigenetics meets next-generation sequencing. *Epigenetics; Official Journal of the DNA Methylation Society* 3: 318–321.

PHILIPPE, H., H. BRINKMANN, D. V. LAVROV, D. T. J. LITTLEWOOD, M. MANUEL, G. WÖRHEIDE, AND D. BAURAIN. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology* 9: e1000602.

PLANT METABOLIC NETWORK. 2011. Plant metabolic pathway databases. Website www.plantcyc.org.

PLANT METABOLIC NETWORK, AND TAIR. 2011. AraCyc. Website http://www.arabidopsis.org/biocyc/index.jsp.

POOL, J. E., I. HELLMANN, J. D. JENSEN, AND R. NIELSEN. 2010. Population genetic inference from genomic sequence variation. *Genome Research* 20: 291–300.

PORRECA, G. J., K. ZHANG, J. B. LI, B. XIE, D. AUSTIN, S. L. VASSALLO, E. M. LEPROUST, ET AL. 2007. Multiplex amplification of large sets of human exons. *Nature Methods* 4: 931–936.

RAUSHER, M., Y. LU, AND K. MEYER. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *Journal of Molecular Evolution* 67: 137–144.

RICHARDS, C. L., AND J. F. WENDEL. 2011. The hairy problem of epigenetics in evolution. *New Phytologist* 191: 7–9.

ROCHE-454. 2011. GS FLX+ System. Website http://454.com/products/gs-flx-system/index.asp.

SAINTENAC, C., D. JIANG, AND E. AKHUNOV. 2011. Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biology* 12: R88.

SANDERSON, M., AND M. MCMAHON. 2007. Inferring angiosperm phylogeny from EST data with widespread gene duplication. *BMC Evolutionary Biology* 7: S3.

SANG, T., AND Y. ZHONG. 2000. Testing hybridization hypotheses based on incongruent gene trees. *Systematic Biology* 49: 422–434.

SATTERLEE, J. S., D. SCHUBELER, AND H. H. NG. 2010. Tackling the epigenome: Challenges and opportunities for collaboration. *Nature Biotechnology* 28: 1039–1044.

SCHLÜTER, P. M., T. F. STUESSY, AND H. F. PAULUS. 2005. Making the first step: Practical considerations for the isolation of low-copy nuclear sequence markers. *Taxon* 54: 766–770.

SCHMITZ, R. J., AND X. ZHANG. 2011. High-throughput approaches for plant epigenomic studies. *Current Opinion in Plant Biology* 14: 130–136.

SCHWEIGER, M., M. KERICK, B. TIMMERMANN, AND M. ISAU. 2011. The power of NGS technologies to delineate the genome organization in cancer: From mutations to structural variations and epigenetic alterations. *Cancer and Metastasis Reviews* 30: 199–210.

SCOVILLE, A. G., L. L. BARNETT, S. BODBYL-ROELS, J. K. KELLY, AND L. C. HILEMAN. 2011. Differential regulation of a MYB transcription factor is correlated with transgenerational epigenetic inheritance of trichome density in *Mimulus guttatus*. *New Phytologist* 191: 251–263.

SHEN, P., W. WANG, S. KRISHNAKUMAR, C. PALM, A. K. CHI, G. M. ENNS, R. W. DAVIS, ET AL. 2011. High-quality DNA sequence capture of 524 disease candidate genes. *Proceedings of the National Academy of Sciences, USA* 108: 6549–6554.

SHEN, Y., Z. WAN, C. COARFA, R. DRABEK, L. CHEN, E. A. OSTROWSKI, Y. LIU, ET AL. 2010. A SNP discovery method to assess variant allele probability from next-generation resequencing data. *Genome Research* 20: 273–280.

SIOL, M., S. I. WRIGHT, AND S. C. BARRETT. 2010. The population genomics of plant adaptation. *New Phytologist* 188: 313–332.

SMALL, R. L., R. C. CRONN, AND J. F. WENDEL. 2004. L. A. S. Johnson review no. 2. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17: 145–170.

STEELE, P. R., M. GUISINGER-BELLIAN, C. R. LINDER, AND R. K. JANSEN. 2008. Phylogenetic utility of 141 low-copy nuclear regions in taxa at different taxonomic levels in two distantly related families of rosids. *Molecular Phylogenetics and Evolution* 48: 1013–1026.

SUMMERER, D. 2009. Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* 94: 363–368.

TENAILLON, M. I., M. C. SAWKINS, A. D. LONG, R. L. GAUT, J. F. DOEBLEY, AND B. S. GAUT. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Sciences, USA* 98: 9161–9166.

TENAILLON, M. I., J. U'REN, O. TENAILLON, AND B. S. GAUT. 2004. Selection versus demography: A multilocus investigation of the domestication process in maize. *Molecular Biology and Evolution* 21: 1214–1225.

TURNER, E. H., C. LEE, S. B. NG, D. A. NICKERSON, AND J. SHENDURE. 2009. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nature Methods* 6: 315–316.

VAN DE PEER, Y., S. MAERE, AND A. MEYER. 2009. The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics* 10: 725–732.

VARLEY, K. E., AND R. D. MITRA. 2010. Bisulfite patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. *Genome Research* 20: 1279–1287.

VARSHNEY, R. K., S. N. NAYAK, G. D. MAY, AND S. A. JACKSON. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27: 522–530.

WANG, R.-L., A. STEC, J. HEY, L. LUKENS, AND J. DOEBLEY. 1999. The limits of selection during maize domestication. *Nature* 398: 236–239.

WANG, Z., M. GERSTEIN, AND M. SNYDER. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57–63.

WEIGEL, D., AND R. MOTT. 2009. The 1001 Genomes project for *Arabidopsis thaliana*. *Genome Biology* 10: 107.

WHITTALL, J. B., A. MEDINA-MARINO, E. A. ZIMMER, AND S. A. HODGES. 2006. Generating single-copy nuclear gene data for a recent adaptive radiation. *Molecular Phylogenetics and Evolution* 39: 124–134.

WOLINSKY, H. 2010. History in a single hair. *EMBO Reports* 11: 427–430.

WOOLLARD, P. M., N. A. L. MEHTA, J. J. VAMATHEVAN, S. VAN HORN, B. K. BONDE, AND D. J. DOW. In press. The application of next-generation sequencing technologies to drug discovery and development. *Drug Discovery Today Corrected Proof.*

WRIGHT, S. I., AND B. S. GAUT. 2005. Molecular population genetics and the search for adaptive evolution in plants. *Molecular Biology and Evolution* 22: 506–519.

XU, W., W. J. BRIGGS, J. PADOLINA, R. E. TIMME, W. LIU, C. R. LINDER, AND D. P. MIRANKER. 2004. Using MoBIoS' scalable genome join to find conserved primer pair candidates between two genomes. *Bioinformatics* 20: i355–i362.

ZHANG, X. 2008. The epigenetic landscape of plants. *Science* 320: 489–492.

ZHENG, J., M. MOORHEAD, L. WENG, F. SIDDIQUI, V. E. H. CARLTON, J. S. IRELAND, L. LEE, ET AL. 2009. High-throughput, high-accuracy array-based resequencing. *Proceedings of the National Academy of Sciences, USA* 106: 6712–6717.

ZHU, J. K. 2008. Epigenome sequencing comes of age. *Cell* 133: 395–397.