Next-generation sequencing for understanding and accelerating crop domestication

Robert J. Henry

Advance Access publication date 23 October 2011

Abstract

Next generation Sequencing (NGS) provides a powerful tool for discovery of domestication genes in crop plants and their wild relatives. The accelerated domestication of new plant species as crops may be facilitated by this knowledge. Re-sequencing of domesticated genotypes can identify regions of low diversity associated with domestication. Species-specific data can be obtained from related wild species by whole-genome shot-gun sequencing. This sequence data can be used to design species specific polymerase chain reaction (PCR) primers. Sequencing of the products of PCR amplification of target genes can be used to explore genetic variation in large numbers of genes and gene families. Novel allelic variation in close or distant relatives can be characterized by NGS. Examples of recent applications of NGS to capture of genetic diversity for crop improvement include rice, sugarcane and Eucalypts. Populations of large numbers of individuals can be screened rapidly. NGS supports the rapid domestication of new plant species and the efficient identification and capture of novel genetic variation from related species.

Keywords: domestication; crops; wild relatives; plant improvement

INTRODUCTION

Crop improvement is facilitated by harnessing the gene pool of the species and related species to find genotypes and recombine genes to deliver superior plant performance in agriculture, food, energy and biomaterial production [1]. Expansion of the available gene pool to incorporate genes from increasingly distant relatives can be assisted by an improved knowledge of the events associated with domestication. Next-generation sequencing (NGS) allows whole-genome analysis to determine the genetic basis of phenotypic differences [2]. This facilitates the combination of traits essential in a domesticated genotype with genes for desirable traits in wild genetic resources. NGS provides very large amounts of DNA sequence data but often only as short sequence reads. The amount of data and the length of the reads are increasing as the technology continues to be developed. NGS is beginning to be applied to crop systems [3].

NGS provides the opportunity to explore genetic diversity in plants and their wild relatives on a much larger scale than was possible with earlier technologies [3, 4] and allows even the most complex plant genomes to be tackled [5]. This review will examine the contribution of next-generation sequencing to understanding domestication and the use of this knowledge in capturing genes from diverse sources for plant improvement and in the domestication of new crop species. Examples of whole-genome sequencing [6] (shot-gun sequencing of genomic DNA), transcriptome sequencing [7] (sequencing of cDNA from different tissues and developmental stages) and amplicon sequencing [8] (sequencing of DNA amplified using primers for regions or genes of interest and DNA bulked from all the individuals from the population of interest) will be discussed. Amplicon sequencing may involve the sequencing of a small number of genetic loci for a large number of genotypes, a large number of genetic
Importance of understanding traits important to humans

Understanding domestication of plants as crops requires that we identify those traits that are important to humans and have been altered during domestication. These include traits such as colour and taste associated with desirability of the product and traits such as loss of shattering and control of flowering time [9] that make the plants more suitable for growing in an agricultural context. Traits selected in domesticated plants may often reduce the fitness or adaptation of the plant in the wild [10]. Some of these traits, such as seed size in cereals, are common to different domestication events while others may be very specific to the species. NGS allows efficient discovery of the genes that have been selected during domestication [11] and may contribute to a better understanding of the processes of domestication. Moving useful traits from wild plants into domesticated varieties requires that the key genes that need to be preserved in the domesticated variety are known so that the introgression can be managed to avoid their loss in recombination with the wild relative.

Examples of use of NGS in the capture of novel variation for established crops

NGS can be applied to the discovery of novel useful variation (Table 1). Analysis of the primary gene pool and of more distantly related wild relatives has potential to identify genes and alleles that can be used to improve the performance of major crop species [12]. NGS allows rapid expansion of genomic analysis to investigation of non-model species [13]. Discovery of variation in plant genomes has been limiting the application of sequence-based selection in plant improvement [14] until the development of NGS. Examples of the use of NGS in discovery of this useful variation with the aim of domesticating new genes or species will be described to illustrate different approaches in diverse applications to food (rice), energy (sugarcane) and forest (Eucalypt) species.

Rice

Rice is a key global food crop. Rice was the first crop to have a reference genome sequence [15, 16] and is an excellent model for use of NGS in re-sequencing to discover variation in both domesticated genotypes and in the wider gene pool of related Oryza species. Understanding the domestication of rice remains challenging despite significant efforts probably because of a complex origin involving some gene movement between the major japonica and indica groups [17, 18].

Whole-genome re-sequencing of rice genotypes has allowed analysis of domestication genes identified in areas of low diversity in the genome as a result of human selection [19]. This approach has shown that both japonica and indica rice share common areas of low diversity possibly due to introgression from one population to the other [20] following selection. Many areas of low diversity contain only a single domestication gene. Genes for shattering upright growth habit and white grain pericarp are all associated with these areas of low diversity in domesticated rice [19]. These studies may provide a useful understanding of the domestication of rice and the relationships between modern rice populations.

NGS of wild rice populations has recently been used to identify the distinctness of Asian and Australian wild rice populations [21]. This analysis made use of whole-chloroplast genome sequences identified in shot-gun sequencing of total plant DNA samples [22]. Oryza rufipogon was domesticated in Asia. The Australian populations of O. rufipogon seem to be more closely related to O. meridionalis than to Asian populations of O. rufipogon. However, all these populations are important genetic resources for rice improvement. Along with clarifying phylogenetic relationships, NGS of the nuclear genome of wild rice relatives will identify genetic diversity of potential value for rice improvement.

Starch and food properties of rice

NGS provides new tools for evaluation of the grain quality [23]. The cooking properties and texture of rice are traits selected by humans during domestication. The properties of the starch largely determine these characteristics of rice. The biochemical steps in the biosynthesis of starch have been elucidated by conventional biochemistry. The genetic basis of differences in rice starch properties and the changes introduced during domestication have been revealed by analysis of allelic diversity in the genes of starch metabolism and association of this variation with differences in starch properties or phenotype. NGS has been applied to understanding the contribution of variation in genes of starch biosynthesis to variations.
in grain quality in the germplasm within a breeding program [8]. The genes of starch metabolism were amplified from a breeding germplasm population and sequenced as a single sample. This allowed detection of all of the polymorphisms in the population. Mutations likely to alter the sequence and associated function of the expressed enzymes were identified. Genotyping of the population allowed these variants to be associated with trait variation. This approach provides an analysis of the full set of alleles at all candidate genes in the population and allows association to identify the alleles that need to be selected to achieve any required phenotype. This approach has potential for routine application to define breeding targets and options in a given gene pool.

**Sugarcane**

Sugarcane is probably the most important energy crop globally. Modern sugarcane is a hybrid between two species of *Saccharum*. *Saccharum officinarum* was domesticated as a sweet (high sucrose) type from wild *S. robustum*. Hybridization of this high-sugar domesticated species with wild *S. spontaneum* has generated the high-performing high-sugar modern sugarcane genotypes [24].

**Sugar and energy potential of sugarcane**

Sugarcane is a major sugar crop and is being developed as a major energy crop with the potential to convert not only the sugar but also the fibre in the plant to fuel and other biomaterials. Sugar content controlled by genes of sucrose metabolism and transport are key domestication traits in this crop. Sucrose phosphate synthase (SPS) is the enzyme responsible for sucrose formation and has been studied in some detail [25]. Specific alleles of key sucrose metabolism genes are expressed in different tissues and may explain the sugar phenotype. The highly polyploid nature of sugarcane has made analysis of allelic diversity of multifamily genes like SPS very difficult. NGS offers for the first time technology that has the power to resolve the complexity of this system. NGS of amplified target genes has been used to identify SNP for mapping of gene of interest in sugarcane breeding [26]. This approach will replace earlier molecular marker technologies that have been used to explore relationships between sugarcane and the pre-domestication progenitors and related species [27].

Sorghum is the closed species with a reference genome sequence [28]. This sequence provides a useful resource for sugarcane genome analysis.

The sugarcane genome is the subject of a coordinated sequencing effort incorporating NGS approaches [29]. A genome sequence will provide a platform for rapid advances in understanding how the component genomes contribute to the performance of this complex hybrid. This analysis may provide clues to new paths for domestication of new sugarcane types based upon combinations of the progenitor genomes.

**Eucalypts**

Eucalypts are among the most widely planted forest species in the world especially in tropical and sub-tropical areas. A reference genome sequence is available for *Eucalyptus grandis*. The Eucalypts include >700 taxa adapted to a wide range of environments. Many species have not been domesticated significantly and plantings are often from wild material. Re-sequencing using NGS is an attractive option for exploring variation in this important group. Eucalypts are a major global option for a woody energy crop [30].

**Wood properties and bioenergy traits of Eucalypts**

The properties of wood are major determinants of value and as such very important traits for selection in the breeding of forest species. Association genetic has been widely applied to the identification of genes controlling key wood quality traits. The genes of cell wall biosynthesis (polysaccharide and lignin metabolism) are the primary candidate genes for wood properties. NGS of gene of cell wall metabolism and association with wood traits has been used to identify genes in Eucalypts influencing wood quality [31]. Eucalypts are only at the very early stages of domestication with many forest plantings still based upon selected seed collected from wild plants. Eucalypts have become the most widely planted hardwood forest species especially in tropical environments. Understanding of the genome and the diversity of the genome in wild populations will be essential to the domestication of high-performing trees with a composition that suits solid wood, pulp and energy end uses. Short rotation cropping systems may be preferred requiring the selection of genotypes suitable for this production system.

**Use of NGS for domestication of new species**

Satisfying the increasing demand for food and energy is challenging because of the finite nature of arable
land and fossil fuels and complicated by the risks of climate change. The food and energy security requirements of a growing world population with increasing affluence place the continued development of existing crop species and may also be supported by the domestication of new crops [1]. This is an important option especially for energy crops since most species of domesticated plants were not selected originally for their energy crop potential. Food crops suitable for production in a changing climate may also require domestication of new species. The domestication of new species requires the selection of plants with characteristics that suit production in a domesticated cropping environment. The domestication of the major cereal crops was associated with the selection of non-shattering and larger seeded genotypes. Shattering or shedding of mature seeds aids dispersal of wild grass seeds but is not desirable in a crop that should have seeds that are retained on the plant until they can be harvested. Larger seeds contribute to greater yields and ease of processing as food. Genes controlling these traits have been identified in model crops such as rice. Targeted selection of mutants or variants at these genetic loci would allow rapid domestication of wild grasses as new crops.

NGS has been used to identify homologues of domestication genes in a wild relative of rice [32]. Genes for traits such as heading date, seed shattering, grain number, pericarp colour, seed width, plant stature and starch properties were analysed in *Microlaena stipoides* a distant relative of rice in the same sub-family (Ehrhartioideae). This wild relative is a smaller seeded grass adapted to temperate climates and used as a pasture species but had enough homology to rice at the sequence level to allow successful identification of homologues of these domestication genes in the whole-genome sequence data. Primers designed for amplification of these genes using the whole-genome sequence facilitated the capture of 92% of the sequence of the genes in the wild relative [32].

### NGS-based plant identification and IP protection for new crops

Identification of plant genotypes can be important in plant production and processing to food or biomaterials. Discriminating identification tools are also important for the protection of intellectual property associated with plant genotypes.

NGS offers a new level of protection that will allow distinction of key genes determining functional traits. Recently, NGS was demonstrated as a cost-effective method for plant identification with a useful strategy being the analysis of the chloroplast genome sequence from whole-genome shot-gun sequencing of plant DNA [22]. Nuclear sequences selected in domestication would allow protection of newly developed crop species with NGS providing a method for effectively policing the protection of the germplasm.

Transcriptome sequencing of barley genotypes has been used to identify large numbers of SNP that can distinguish the genotype and define genetic variation within the cultivar. Even in-breeding crops such as barley show genetic diversity with the variety because of the breeding strategies used to generate the foundation population for the variety. NGS has the power to identify even minor variants with the variety. Variety identification tools need to focus on SNP that are common to all individuals of the variety. NGS facilitates managing this diversity and any changes in crop performance over time due to genetic drift. NGS technology has potential for expansion to cover wild populations and to support introgression of specific genome regions with desirable traits. Drought tolerance in barley has been found to be associated with specific chromosomal regions introgressed for wild barley (*Hordeum spontaneum*) [33]. NGS should allow characterization of the part of the genome that has been introduced into drought tolerant domesticated genotypes developed using wild barley germplasm.

### NGS for development of novel hybrid crops

The performance of hybrid crops is a key contributor to sustainable food production with the heterosis of wide crosses delivering greater productivity.

---

**Table I: Examples of NGS for discovery in gene and species domestication**

<table>
<thead>
<tr>
<th>Species</th>
<th>Traits</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice</td>
<td>Domestication genes</td>
<td>Whole genome</td>
<td>1, 9</td>
</tr>
<tr>
<td></td>
<td>Agronomic traits</td>
<td>Whole genome</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Starch related end-use quality</td>
<td>Amplicon</td>
<td>8</td>
</tr>
<tr>
<td>Maize</td>
<td>Root development</td>
<td>Transcriptome</td>
<td>7</td>
</tr>
<tr>
<td>Wild grasses</td>
<td>Domestication genes</td>
<td>Whole genome</td>
<td>32</td>
</tr>
<tr>
<td>Eucalypts</td>
<td>Wood properties</td>
<td>Amplicon</td>
<td>31</td>
</tr>
<tr>
<td>Sugarcane</td>
<td>Sugar content</td>
<td>Amplicon</td>
<td>26</td>
</tr>
</tbody>
</table>
NGS has been applied to the analysis and prediction of heterosis in well-studied systems such as maize [34]. Patterns of gene expression have been evaluated in hybrids [35] using NGS. Rice miRNA and small RNAs have been shown to be potential contributors to heterosis in hybrid rice [36]. NGS could be extended to species in which the technology has not been well developed. Whole-genome sequencing and transcriptome sequencing can be used to compare the genomes and transcriptomes of parental lines and direct the crossing of individuals with greater potential for superior hybrid performance. Transcriptome sequencing in triticale has been used to identify the genes from rye that are expressed in the hybrid [37]. This technology has the potential to rapidly advance the development of new crop species as productive new hybrid cropping options.

**Future prospects using NGS**

Plant breeders using NGS will have more opportunity than ever before to deliberately introduce diversity into their varieties [38] while retaining their objectives for plant performance and product quality. Application of NGS to genome analysis in vegetatively propagated species such as banana may identify genes that may be introduced into the domesticated gene pool by transformation [39]. Greatly accelerated domestication of wild germplasm can be anticipated [40]. Introduction of new traits or alleles will be facilitated while simultaneously retaining major domestication genes [41]. These developments will be critical in adapting crops and agriculture to a changing climate [42, 43]. Molecular selection tools will be more easily discovered using NGS [44] and applied to diverse wild resources [45] using advanced screening methods that allow automation and high through put [46] in gene discovery and transfer into commercial genotypes.

### Key Points

- Capture of novel genes from wild species will be made easier by the molecular events associated with crop domestication.
- Re-sequencing of domesticated species can identify low-diversity regions resulting from selection during domestication.
- Low-coverage short-gun NGS of whole genomes of wild relatives can be used to identify gene-specific sequences to aid the cloning of homologues of key domestication genes from wild relatives.
- NGS of amplicons of large numbers of candidate genes from wild and domesticated plant populations can define diversity of target genes in wild populations and lead to the discovery of key genes for important traits by association analysis.

### References


