Alfalfa Breeding Benefits from Genomics of *Medicago truncatula*

Bernadette Julier · Isabelle Meusnier

Summary: International programs aim at developing knowledge and tools in the model species *Medicago truncatula*. Genetic resources, DNA sequences, markers, genetic and physical maps are now publicly available. These efforts contribute to improve breeding schemes of crop species such as alfalfa. However, transfer of information from *M. truncatula* to alfalfa is not straightforward. The article reviews the gain given by the model species to better analyze genetic determinism of breeding traits in alfalfa. It also shows that investments in alfalfa genomics (DNA sequences, SNP development) are needed to benefit from the model species.

Key words: breeding, candidate gene, marker, *Medicago sativa*, *Medicago truncatula*, model species

Introduction

Alfalfa (*Medicago sativa* L.) is a major forage species over the world. Its genetics is complex (allozyme and autotetraploidy), it is a perennial plant and cultivars are synthetic varieties that contain a high genetic diversity. Its wild relative *Medicago truncatula* was chosen by the scientific community as a model for legume species (Barker et al. 1990). This plant, close to many important crops (alfalfa, clovers, pea, soybean, bean, etc.) was selected given its relative small genome size (450Mb/1C, twice less than alfalfa), its simple genetics (diploidy and autogamy) that simplify genetic analysis, its easy growing (short generation time, prolific seed production) and its ability to undergo genetic transformation. The international design strategy was to focus research on the model plant and after all to transfer the knowledge to more complex species such as alfalfa. Many tools are being developed in the model species, such as genomic sequences, EST, markers, genetic and physical maps and expression data (www.medicago.org). It is expected that genomic resources developed in the model species will accelerate genetic analysis in related crop species. However, transfer of information from a model to a cultivated species needs investigation. The aim of this article is to review which data or tools issued from the model species would be useful in a breeding program of alfalfa. After a description of the genetic analyses that were feasible in alfalfa before genomics on *M. truncatula*, the knowledge and tools that are being developed on *M. truncatula* genomics is described. The last part shows that these tools are useful to enhance breeding program strategies.

Genetic Analyses in Alfalfa before *M. truncatula* Genomics

Alfalfa cultivars are synthetic populations coming from the cross pollination of several parents during three or four generations. This structure is a way to valorise heterosis and minimize inbreeding depression (Gallais 1990). Each parent being an individual plant or a family, the diversity among plants in a variety is high. Traditional breeding is based on different steps: the identification of breeding criteria related to agronomic traits, the evaluation of genetic diversity for these criteria, a recurrent selection program for the breeding criteria and finally the selection of parental plants or families to create a new variety. Breeding traits are basically forage yield, forage quality, disease and pest resistances, lodging resistance and seed yield. They are mostly quantitative traits governed by several genes. In order to improve breeding schemes, quantitative genetic studies were conducted to analyse genetic diversity, trait heritability and genetic correlations. Although, the genetic progress of these breeding programs is low in alfalfa, due to tetraploidy and synthetic structure of the varieties, significant
Before the development of genomic tools in model species, molecular biology was rarely used to analyse genetic diversity or genetic determination of traits. For example, as the number of markers was insufficient, it was almost impossible to build a complete genetic map (Yu & Pauls 1993, Brouwer et al. 2000). However, few markers were available and used in combination with phenotypic traits, particularly to analyse genetic diversity. With phenotypic traits related to agronomic value, it was shown that variations among and within populations were high (Julier et al. 1995, Julier et al. 1996, Crochemore et al. 1998, Julier et al. 2000, Bolaños-Aguilar et al. 2001). Isoenzymatic, RAPD, AFLP and RFLP markers also showed a large among individual variation (Pupilli et al. 1996, Crochemore et al. 1996, Jenczewski et al. 1999, Muller et al. 2001, Maureira et al. 2004). However, these markers suffered from several limitations: isoenzymes and RFLP markers exhibited a low level of polymorphism, and RAPD and AFLP were dominant and anonymous. At this time, the availability of SSR markers in alfalfa that combine high polymorphism, codominance and locus-specificity was restricted to about a dozen (Diwan et al. 1997, Diwan et al. 2000).

The history of domestication and the relationship between wild and cultivated populations were studied using mitochondrial and nuclear polymorphisms. A demographical bottleneck probably occurred during domestication, even if a low level of divergence was detected between the wild and the domesticated populations (Muller et al. 2001, Muller et al. 2003, Muller et al. 2006). In the 2000's, the available high throughput and reproducible markers, i.e. AFLP, were used in a Bulk Segregant Analysis (BSA) to identify markers linked to a disease resistance. Using two populations specially designed to differ only for their resistance to downy mildew (Peronospora trifoliorum), Obert et al. (2000) identified two AFLP markers that were linked to the resistance. BSA is still an efficient method to test a lot of markers with a reasonable input of funds and time, but it is difficult to sequence an AFLP marker, to get genomic sequence in its region and to convert it into a codominant marker.

Progress in M. truncatula Genomics and Genetics

From 1990's onwards, international programs were launched on the model legume M. truncatula. Extensive reviews are available (Young & Udvardi 2009, Rispail et al. 2010). Genetic resources of M. truncatula were collected, population structure was analysed (Ronfort et al. 2006) and their genetic diversity for many traits was described, including first nitrogen-fixing symbiosis (Barker et al. 1990), and more recently other traits that may have an agronomic interest in related species (Duc 2004, Moussart et al. 2006, Julier et al. 2007). Sequencing efforts were first devoted to ESTs (Bell et al. 2001) and lately to gene-rich regions (Young et al. 2005). In parallel, F3 (Thoquet et al. 2002, Choi et al. 2004a, Ané et al. 2008) and RILs (Julier et al. 2007, Pierre et al. 2008) mapping populations were developed.

Knowledge in genome sequence of M. truncatula allows to develop low cost and easy-to-use markers, mainly Simple Sequence Repeats (SSRs). Sequence data (EST or BAC) was screened for SSR motifs and primers surrounding these motifs were designed and tested. Large sets of EST-SSR markers or SSR from BAC sequences are now available (www.medicago.org).

The constant sequencing technique innovation with higher throughput offers the possibility of more ambitious projects. The next generation sequencing technologies (454, Solexa or Solid) (Rothberg & Leamon 2008, Stanger 2009, Rounsley & Last 2010) gave the opportunity to acquire genomic and/or transcriptomic (eDNA, (Wheat 2010)) sequences of a set of lines for a reasonable cost. Such data are being obtained on M. truncatula by sequencing 30 lines (http://www.medicagoohapmap.org/) (Branca et al. 2010). These data sets will reveal sequence polymorphisms that will be used to better understand genetic evolution of the species and will be exploited to identify single-nucleotide polymorphism (SNP) markers. Such markers are supposed to cover the entire genome with a low cost (Ganal et al. 2009).

Mapping populations are used to analyse genetic architecture of traits of interest. Having identified genetic variation for the target trait, a QTL analysis indicates the zones of the genome that could be involved in trait variation in a population. A more accurate QTL position is obtained by a fine-mapping strategy developed on a large population and the bio-analysis of the genomic sequence of the confidence interval of the QTL gives a list of candidate genes. Within this list, some genes can be selected by making hypotheses on their role on the trait. Such strategy was applied for flowering date. QTL analysis in three M. truncatula mapping populations revealed a major QTL on chromosome 7 (Pierre et
al. 2008). A pseudo-F2 population of 1640 individual was derived from the self-pollination of a F6 individual that was heterozygous at the QTL. A confidence interval of 2.4 cM was calculated. The bio-analysis of the BACs comprised in this region revealed 577 genes. Among them, six genes were described in other species to be involved in the flowering pathway (Pierre et al. 2010). Finally, genomic and cDNA sequences of these genes were obtained and their expressions were compared between the two parents. Two genes (CONSTANS-like and FT) showed sequence polymorphism but only CONSTANS-like was differentially expressed. Similarly for the resistance to anthracnose (Colletotrichum trifolii), three genes having motifs specific to resistance genes were identified after a QTL analysis followed by a fine mapping step in M. truncatula (Yang et al. 2007, Yang et al. 2008). A susceptible alfalfa clone was transformed by each of the three genes but only one gene, called RCT1, induced resistance to anthracnose. The genes identified in M. truncatula (CONSTANS-like, FT, RCT1 or others) could possibly explain variation in alfalfa, but it has to be proved.

**Marker Development**

Sequence data (EST or BAC) was exploited to acquire low cost and easy-to-use SSR markers. Indeed, many of the SSRs developed on M. truncatula from EST were transferred to alfalfa (Julier et al. 2003, Sledge et al. 2005), due to synteny. The transfer success of SSR from the model plant to cultivated plant depends of the marker location in the genome. In coding regions, the primers designed on M. truncatula often gave an amplification product when tested on alfalfa (Julier et al. 2003). Conversely, SSR developed in non-coding regions were hardly usable.

**Sequence Polymorphism**

Sequence polymorphism is not expected to be the same across legume species because point mutations are different. Thus SNPs identified in M. truncatula are not transferrable in alfalfa. The development of SNPs on alfalfa requires sequencing efforts on this specific species. As on M. truncatula and other species, the next generation sequencing technologies give the opportunity to acquire genomic and/or transcriptomic sequences to reveal polymorphisms in alfalfa. These high throughput sequencers are especially useful for heterozygote species such as alfalfa because all the alleles can be sequenced at once without cloning. Even if a reference sequence of alfalfa is not available, the interpretation of short DNA reads obtained by next generation sequencers will gain from the knowledge of M. truncatula sequence, the model species genome would be used as a reference. In fact, sequencing of alfalfa is closer to a re-sequencing effort than to a de novo sequencing (Young & Udvardi 2009). Specific bioinformatics tools will be needed to handle the high level of polymorphism due to heterozygosity and tetraploidy.

**Analysis of Alfalfa Genetic Resources**

The genetic diversity of alfalfa was also analysed with codominant SSR markers that have a higher level of polymorphism than isoenzymes or RFLP. With SSR markers developed from M. truncatula, their number is sufficient to cover the alfalfa genome. However, SSR markers in alfalfa suffer from two limitations: (1) it is often difficult to read their allelic dosage, so they are not truly "codominant" markers: an allele that is present can be in one to four doses, (2) they contain a large number of null alleles, probably due to a lack of cross amplification or sequence reads obtained by next generation sequencers will gain from the knowledge of M. truncatula sequence, the model species genome would be used as a reference. In fact, sequencing of alfalfa is closer to a re-sequencing effort than to a de novo sequencing (Young & Udvardi 2009). Specific bioinformatics tools will be needed to handle the high level of polymorphism due to heterozygosity and tetraploidy.

polymorphisms in alfalfa itself. This reduces the information provided by the markers. SSR markers confirmed the large within-population diversity and the high heterozygosity of the individuals (Flajoulot et al. 2005, Herrmann et al. 2008, Bagavathiannan et al. 2010). They proved that alfalfa varieties or landraces are at panmictic equilibrium, as expected from their mode of production with several cross-pollination generations. Population structure was not described more precisely with SSRs than with AFLPs; populations were highly similar on average. But when sp. *falcata* wild populations were studied, they were easily differentiated from the sp. *sativa* cultivated populations (Jenezewski et al. 1999, Muller et al. 2001, Maureira et al. 2004).

The consequences of this information on alfalfa breeding are clear: breeding for agronomic traits using the current methods does not reduce the neutral genetic diversity. The distinctness of populations or varieties is difficult except if a large range of variation including both sp. *falcata* and *sativa* is concerned. However, it would be interesting to revise this evaluation by an analysis of genetic resources with non-neutral markers. These markers related to agronomic traits would reveal another view of the genetic diversity present in this species and perhaps give other prospects to use genetic resources in breeding programs.

Development of Genetic Maps and QTL Detection

The SSR markers were used for genetic mapping in alfalfa. As they amplify several alleles per locus, they help significantly to find the four homologous chromosomes that could be detected only by repulsion linkage with dominant markers (Hackett et al. 1998). SSRs contributed deeply to build dense genetic maps of cultivated alfalfa (Julier et al. 2003, Sledge et al. 2005). Genetic mapping confirmed the autotetraploidy of the genome, as supposed by cytogenetic studies (Armstrong 1954). Mapping populations were either F1 or backcross as the high inbreeding depression restricts the analysis of F2 populations and prevents to build RIL populations. QTL detection was carried out for some agronomic traits: disease resistance (Mackie et al. 2007), morphogenesis (Julier et al. 2002), forage yield (Robins et al. 2007), water use efficiency (Julier et al. 2008). In most cases, many QTLs, each with a limited effect, were detected. For disease resistance, a lower number of QTLs was found. Usually, for a trait, each parent carried both positive and negative alleles that explained the presence of transgressive individuals in the progeny. These QTLs could be used to select individuals carrying both valuable phenotypic traits and positive alleles at the QTLs in the mapping population and to create an improved population. However, this strategy was never reported to be used on alfalfa.

The QTL position is usually imprecise due to the tetraploidy of alfalfa. To get the same precision than on a diploid species it would require much larger mapping populations. It is a typical case in which a cultivated plant would take advantage of the information available on the model plant. As alfalfa and *M. truncatula* genomes are syntenic, the position of a QTL detected on alfalfa is the same on *M. truncatula*. Using the sequence database of *M. truncatula* and its annotation, it is possible to list all the genes described on *M. truncatula* in this region. These genes could be responsible for the trait variation in the mapping population, especially if they are involved in the pathway described in other species to contribute to the trait.

Association Mapping

Association genetics is currently often used in plant species to establish a link between a gene or a genomic region and a trait. Compared to mapping populations based on only two parents, association genetics is based on a wide genetic diversity. In allogamous species such as alfalfa, the linkage disequilibrium (LD: the non-random occurrence of alleles at different loci) is expected to be short (Flint-Garcia et al. 2003). Experimental results confirmed this expectation (Auzanneau et al. 2007). In such a situation, an association study relies on a candidate gene approach in which the polymorphism of the gene is linked to the trait polymorphism. This analysis provides a high accuracy on the location of the causal mutation compared to a genome scan analysis that is carried out when the LD is large.

LD can be estimated with markers spread at known distances along chromosomes. For example, in a study of Constans-like gene that was partly sequenced in alfalfa, LD decayed shortly and after 1200 bp, no more linkage was observed (Herrmann et al. 2010). This result is in agreement with the theory and means that only a candidate gene approach is valuable in this species. Moreover, for long genes, it is important to analyse the whole sequence of the gene or several short parts evenly spaced along the gene.

In the case of alfalfa, the use of anonymous markers such as SSRs or AFLPs for an association
study would hardly give positive results. Indeed the chance that the markers are located in the proper gene is extremely low. It would require to genotype all the individuals of the studied population with many markers. In the next years, the development of SNPs covering the alfalfa genome and the ability to genotype large populations will allow such association study. However, if populations having a longer LD decay would be developed, especially by polycrossing inbred parents, a genome-scan approach could be implemented to link a genomic region to a trait.

In this situation when LD decay is short, it is valuable to have an association study based on a candidate gene approach. Genes taken in the literature or identified in *M. truncatula* (CONSTANS, FT, RCT1 or others) could possibly explain trait variation in alfalfa. In the case of the CONSTANS-like gene on chromosome 7, two gene portions were sequenced in an alfalfa population of 400 individuals. Compared to a neutral gene, CONSTANS-like gene showed less SNPs (1 SNP every 125 pb compared to 1 SNP every 30 pb for the neutral gene), indicating that CONSTANS-like gene is probably under selection. All individuals were phenotyped for flowering date and stem height, the SNPs explained up to 4% of the variation (Herrmann et al. 2010). It proved that a gene identified in *M. truncatula* from a QTL analysis is involved in trait variation in alfalfa.

The next generation sequencers can be used to test if a candidate gene, identified in *M. truncatula*, is involved in a trait variation in alfalfa. Currently, these apparatus are mainly employed to sequence or re-sequence organisms as described above. They can also be useful to sequence a large number of sequences coming from a single gene amplified in numerous genotypes. Bulk segregant analysis, in which the individuals of one population are grouped into two bulks according to their phenotype, is a method to establish interesting plant material. With next generation sequencer, all alleles of the candidate gene can be sequenced in all individuals (total number of alleles = number of populations x 2 bulks x number of individuals per bulk x 4 alleles per individual). Once polymorphism is evidenced, it has to be related to the trait: the frequency of mutations must be different in the two contrasted bulks. Such analysis is handled by bioinformatics tools. The following step is to develop markers in order to genotype individual plants and to assess the link between the gene polymorphism and the phenotype. The reliable markers will be exploited in marker-assisted selection.

**Conclusions**

The development of genetic and genomic tools was delayed on alfalfa because of its biology (allogamous species) and genetics (autotetraploidy). However, efforts in sequencing and analysis of the model species *M. truncatula* give new perspectives for alfalfa breeding. Locus-specific markers such as SSRs are now available and offer the possibility to analyse genetic diversity, build genetic maps and detect QTLs. All these results can be exploited into breeding programs. Research studies in *M. truncatula* are progressively identifying genes involved in traits of agronomic interest. These genes are candidates to explain trait variation in the cultivated species and can be tested in association genetics. In addition, considering *M. truncatula* genome sequence as a reference, next generation sequencers give the ability to sequence parts or even all the alfalfa genome and to develop large set of SNP markers. All this knowledge and tools developed on alfalfa will assist breeding programs.
References


References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References

References
accessions to infection by three major pathogens of pea (Pisum sativum) and alfalfa (Medicago sativa). Eur. J. Plant Pathol. 117: 57-69
Doprinos genomike *Medicago truncatula* oplemenjivanju lucerke

Bernadet Žilije · Izabel Mesnije

Nacionalni institut za poljoprivredna istraživanja, IJ 4, Multidisciplinarna istraživačka jedinica za travnjaštvo i krmno bilje, Luzinjan, Francuska

**Izvod:** Međunarodni programi usmereni su na sticanje znanja i razvoj genomičkih tehnika na model vrsti *Medicago truncatula*. Genetički resursi, sekvence DNK, marker, genetičke i fizičke mape sada su javno dostupne. Ovi napori doprinose usavršavanju metoda oplemenjivanja gajenih biljaka poput lucerke. Međutim, prenos informacija sa *M. truncatula* na lucerku nije jednostavan. Ovaj rad daje pregled doprinosa ove model vrste za bolju genetičku analizu poželjnih svojstava lucerke, a takođe prikazuje i da su ulaganja u genomiku lucerke (sekvence DNK, razvoj SNP) neophodne za dobijanje koristi od model vrste.

**Ključne reči:** geni kadidati, marker, Medicago sativa, Medicago truncatula, model vrsta, oplemenjivanje