Agriculture and Food

AIMS Agriculture and Food, 2(1): 16-39
DOI: 10.3934/agrfood.2017.1.16
Received 9 November 2016,
Accepted 27 December 2016,
Published 6 January 2017

http://www.aimspress.com/journal/agriculture

Review

Understanding potato with the help of genomics

José Héctor Gálvez[1], Helen H. Tai[2], Noelle A. Barkley[3], Kyle Gardner[2], David Ellis[3], and Martina V. Strömvik[1, *]

[1] Department of Plant Science, McGill University, Montreal, Canada
[2] Fredericton Research and Development Centre, Agriculture and Agri-Food Canada, Fredericton, Canada
[3] International Potato Center, Lima, Peru

* **Correspondence:** Email: martina.stromvik@mcgill.ca.

Abstract: Potato (*Solanum tuberosum* L.) is an important staple crop with a highly heterozygous and complex genome. Despite its cultural and economic significance, potato improvement efforts have been held back by the relative lack of genetic resources available to producers and breeders. The publication of the potato reference genome and advances in high-throughput sequencing technologies have led to the development of a wide range of genomic and transcriptomic resources. An overview these new tools, from the updated versions of the potato reference genome and transcriptome, to more recent gene expression, regulatory motif, re-sequencing and SNP genotyping analyses, paints a picture of modern potato research and how it will change our understanding of potato as well as other tuber producing Solanaceae.

Keywords: *Solanum tuberosum*; *Solanum commensonii*; Solanaceae; genomics; transcriptomics; potato reference genome; gene regulation; regulatory motifs; single nucleotide polymorphisms; genotyping-by-sequencing

1. Introduction

Potato (*Solanum tuberosum* L.) is widely recognized as the most important non-grain staple crop worldwide. The latest FAO statistics indicate that over 380 million tonnes of potatoes were produced in 2014 alone [1], illustrating its international economic and agricultural importance. Potato is a member of the Solanaceae family, which includes other significant agricultural species such as tomato, pepper, and tobacco. The cultivated forms of potato are vegetatively propagated and

are predominantly autotetraploids (2n = 4x = 48). However, ploidy ranges from diploid to hexaploid in cultivated potato [2], (for a review on potato genetic diversity, see Machida-Hirano, 2014 [3]). Potatoes were domesticated in the Andes approximately 10,000 years ago and the landraces have a wide variety of shapes, skin and tuber colors, often not seen in modern varieties [4]. It is fairly common in the Andes that landraces of all ploidy levels are grown in the same field and are also grown near wild relatives facilitating cross hybridization and gene flow [5].

Potatoes are valued for their nutritious properties and their wide eco-geographical range. However, due to their high heterozygosity, complex polysomic inheritances, and narrow genetic base, they are difficult to improve through classical breeding methods. Because they are typically vegetatively propagated, many modern cultivars are only separated by a few meiotic generations [6,7] making the genetic diversity among cultivars really low. They are quite susceptible to many pests and also suffer from acute inbreeding depression.

The scientific and economic importance of potato is not new. While other crops such as maize and wheat have seen great increases in yield as a consequence of genetic improvement in the last few decades, this has not been the case with potato. Instead, evidence suggests that yield increases are mostly due to improved agricultural practices. The majority of cultivated potato still comes from a narrow group of cultivars, including Russet Burbank, which was originally released in 1874 [8,9]. While many more recent cultivars have been released since the late 1800s, these have been bred mostly based on phenotypic selection, not genetic information, and they have been developed with a very particular use in mind, such as processing for the potato chip or the French fry industries [10]. Worldwide demand for potato is increasing; therefore, scientists have begun to study potato genetics with the hope that it can provide breeders with more tools to aid crop improvement in terms of yield and disease resistance.

Until recently, the genomic understanding of this crop was held back by its relatively complex genome. The challenges associated with potato improvement have prompted a number of significant genomic and transcriptomic studies in this species and its close relatives, which will provide tools for breeders and additionally shed light into mechanisms behind important molecular processes. In 2011, the first potato reference genome and transcriptome were published [11,12], and two years later, an update was released substantially improving the scaffolds and pseudomolecules of the initial reference [13]. Recently, the first draft genome of a wild potato species, *Solanum commersonii*, was also released [14], in addition to many other genome sequencing efforts in related species, such as tomato (*Solanum lycopersicum*) [15], chili pepper (*Capsicum annuum*) [16], tobacco (*Nicotiana tabacum*) [17] and the parental genomes of petunia (*Petunia axillaris* and *Petunia integrifolia*) [18] which collectively have also provided valuable information on potato.

The potato reference genome is also a starting point for the exploration of biodiversity between potato cultivars and subspecies. Using genome re-sequencing, it is now possible to assemble separate genomes as a reference for specific varieties. These new assemblies can provide useful information about the structural differences between different potato subspecies (*Solanum tuberosum* subsp. *andigena, S. stenotomum* subsp. *goniocalyx*, *S. stenotomum* subsp. *stenotomum* and 36 *S. tuberosum* subsp. *Tuberosum*; species definition from [2]) from Single Nucleotide Polymorphisms (SNPs) and Copy Number Variation (CNV) to large-scale structural variation. Indeed, recent research is already pointing to significant differences in gene copy number between different potato populations [19]. This review will focus on updates to the potato reference genome since its release in 2011, transcriptome assembly in potato, gene expression, regulatory motif discovery, the draft assembly of a wild potato genome, and modern genotyping approaches.

2. The potato reference genome

Because of the high degree of heterozygosity normally found in *S. tuberosum* a homozygous clone of the plant needed to be created in order to produce a high-quality draft genome. This was achieved through the duplication of a monoploid (1n = 1x = 12) specimen that had been previously derived from a heterozygous clone of the *Phureja* group of cultivated potato. This doubled monoploid, named DM1-3 516 R44 (hereafter referred to as DM) was the only source of sequencing data for the potato draft genome. The genome scaffolds assembled from DM were then used to integrate data from a heterozygous diploid breeding line that was a cross between a *S. tuberosum* "dihaploid" (SUH2293) and a diploid clone (BC1034) generated from two *S. tuberosum* group *Phureja* hybrids [12].

The first potato reference genome was completed in 2011 by the Potato Genome Sequencing Consortium (PGSC) using a whole-genome shotgun (WGS) approach [12]. Previous data had already determined that the potato genome is composed of 12 chromosomes and has a total size of approximately 840 Mb. Contigs were assembled *de novo* using the SOAPdenovo [20] assembly program. The assembly consisted of 727 Mb, 93.3% of which were non-gapped sequences. Analysis of the DM assembly revealed that 62.2% of the assembled genome consisted of repetitive content. To assess the quality of the draft genome, Sanger-derived phase-2 BAC sequences, which amounted to approximately 1 Mb, were aligned to the assembly. No gross assembly errors were detected in the aligned data [21]. A reference transcriptome was produced to annotate the genome and it contained around 21,000 high-confidence transcripts [11].

Two years after the publication of the first reference genome, a new assembly of the DM clone was released with a more accurate arrangement of scaffolds and pseudomolecules [13]. This updated assembly of the potato reference genome (version 4.03) was created by integrating linkage data from a segregating diploid potato population derived from the reference sequence clone (DM). This new dataset was used to revise and improve the genome pseudomolecules (PMs) of the original assembly [13]. The new build contains 951 genome superscaffolds of which 90% (655 Mb) have been assigned to an absolute or relative orientation within the PMs. Also, a small number of superscaffolds (about 3%) have been assigned to a random orientation. The exact chromosome position and absolute orientation of the remaining 279 Mb of superscaffold sequences found in the heterochromatin could not be determined. This means that a total of 93% of the assembled genome, comprising a total of 674 Mb, are contained in the chromosome scale PMs of the 4.03 version of the assembly. A total of around 96% of the predicted genes in potato are found in these PMs [13].

A more recent update of the potato reference genome (version 4.04) was released in 2016 [19]. It was built with additional genomic data obtained from foliar and stem tissue of a potato cloned from the original DM reference. It adds 55.7 Mb of novel sequences in the form of >200 bp contigs, including several new genes, that did not map to the v4.03 reference. These contigs were concatenated into an unanchored pseudomolecule called "chrUn", which was then annotated using a standardized pipeline [19]. However, since this new assembly does not anchor the new data into any chromosome, or incorporate new linkage data in any way, it is only useful as further reference for potato sequences that do not align to any of the established pseudomolecules found in v4.03. A summary of the available reference genomes for potato and its close relatives can be found in Table 1.

Table 1. Summary of the different versions of the potato reference genome as well as the genome of its close relative *S. commersonii,* and tomato (*S. lycopersicum*).

	S. tuberosum reference genome			*S. commersonii* reference genome	*S. lycopersicum* reference genome
	v3	v4.03	v4.04		
Source of Genetic Material	*S. tuberosum* group Phureja DM1-3 516 R44.	Same as v3, with additional linkage data from DMDD† mapping population.	Same as v3 with additional DNA from DM1-3 516 R44 stem and leaf tissue.	*S. commersonii* accession PI 243503.	"Heinz 1706" inbred line (Heinz Corporation, Pittsburgh, PA).
Total Length (Mb)	727	723	779	730	760
Scaffold N_{50} (kb)*	1340	4100	4100	44	16,470
GC content	34.80%	34.80%	N/A	34.50%	35.71%
Predicted Number of Genes	39,031	39,031	N/A	37,662	34,727
Comments	Most recent version available in the NCBI Genome database.	No new novel sequences compared with v3.	Only difference with v4.03 is the addition of 55.7 Mb of novel genes and sequences.	Also available in the NCBI Genome database.	Most recent version available in the NCBI Genome database.
Reference	[12]	[13]	[19]	[14]	[15]

* Minimum size in which 50% of the assembly can be found;

† Mapping population of 180 backcross progeny clones derived from an initial cross (DM × D where DM = DM1-3 516 R44 and D = CIP703825) [13].

3. The potato reference transcriptome and gene expression studies

For many years, the main transcriptomic resources available to potato breeders were public EST libraries containing a total of more than 200,000 tags [21-24]. Additionally, EST libraries from other closely related Solanaceae species (such as tomato, eggplant, pepper, tobacco and petunia) also proved to be relevant for potato because many of the genes were shared across the species and genera in this family [25]. The EST sequence data was used to develop microarrays for analysis of gene expression including cDNA arrays [26,27] and a 44k oligo array using the Agilent platform (Potato Oligo Chip Initiative: POCI) [28]. Recently, another Agilent oligo array (JHI *Solanum tuberosum* 60k array), was developed based on the predicted transcripts of potato reference genome v3.4 (see Table 2) [29]. Collectively, these resources were behind significant discoveries in the gene expression profiles of potato under different conditions such as flowering and tuber development [30-32], biotic [29,33,34] and abiotic stress [35-39].

After the publication of the first potato reference genome, it became possible to design potato transcriptomics studies with the potential to analyze gene expression using RNA-seq. In order to assemble the gene models that make up the potato reference transcriptome, the PGSC collected data from 32 different tissues of the same *Phureja* DM clone used for the sequencing of the reference genome. The tissues were selected to represent all the major plant organs, including flower, fruit, leaf, tuber and roots at different developmental stages and stress conditions [11].

Using RNA-Seq, over 550 million reads were obtained from all tissue samples. Petal tissue yielded the lowest amount of reads with only 5.4 million, while the mature whole fruit library had the greatest number of reads, around 30 million. In terms of high- confidence transcripts, one sample of tuber tissue had the lowest amount (11,394), while the highest number (16,276) was found in plants treated with salt (NaCl). Since libraries with the lowest and highest number of reads produced roughly the same number of high-confidence transcripts, it seems there is no significant bias against transcript detection depending on sequencing depth [11].

A transcript was considered as expressed if its abundance, as calculated using the Cufflinks software package [40], had a fragment per kilobase of exon per million fragments mapped (FPKM) value ≥0.001 and the lower bound of the 95% confidence interval was above zero. Using these criteria, a total of 22,704 unique high-confidence transcripts were identified in all 32 libraries. The *S. tuberosum* reference genome contains a total of 39,031 protein-coding genes. If a single transcript is chosen to represent each gene also found in the genome, around 60% of the genes found in the genome are also included in the reference transcriptome. Out of all these transcripts, only 17% have no known function and a total of 1680 (around 8%) were only found in tissues under some type of biotic or abiotic stress [11].

With the goal of facilitating comparative analyses between potato and tomato, the international Tomato Annotation Group (iTAG) used their annotation pipeline to re-annotate the potato reference genome. This newer potato gene annotation, referred to as iTAG, contains less total genes than the original annotation published by the PGSC (35,004 and 39,031 genes, respectively) [15]. However, when compared to an external standard (TAIR10) [41], 92% of the genes models in the iTAG annotation had a corresponding match, whereas more than 30% of the PGSC genes had no match at all [15]. The iTAG annotation is therefore another valuable resource for potato research, especially in studies that involve comparisons with tomato or other members of the Solanaceae family.

Gene expression studies have proven to be a useful tool for investigating plant molecular response to different environmental stimuli [42]. There has been a recent increase in the application of RNA-Seq to understand potato biology and the genes underlying complex traits. *Phytophthora infestans* defense response, tuberization under the control of photoperiod, drought response, tuber pigmentation, PVY resistance, response to nitrogen fertilizer and an activation-tagged mutant with altered growth habit have all been examined using RNA-Seq to quantify gene expression [43-50]. RNA-Seq has also been used to identify genes that are predictive of cold-induced sweetening in tubers [Neilson et al. 2016, in preparation]. It has also been used to quantify gene expression in a wild potato species, *S. commersonii*, which is resistant to bacterial wilt; this was achieved using the potato reference genome for sequence alignment [51]. Finally, the National Centre for Biotechnology (NCBI) Gene Expression Omnibus (GEO) [52] currently lists more than 1500 *S. tuberosum* samples across 102 series of experiments performed using high throughput sequencing, this includes studies on miRNA and other non-coding RNA, in addition to gene expression.

Expression profiling using RNA-Seq is dependent on accurate alignment of short sequence

reads to the reference genome. The number of aligned reads is the basis for quantification of gene expression. However, it has been noted that some genes are recalcitrant to RNA-Seq analysis [53]. This may be due to small transcript size that is excluded with construction of RNA libraries and/or sequence overlap with other transcripts. Improvements to RNA-Seq methodology and bioinformatic data processing are needed. Nevertheless, RNA-Seq studies produce an abundance of gene expression data that have contributed to understanding a range of potato traits. As RNA-Seq becomes more accessible, new datasets can expand upon this knowledge and enable the discovery of additional information, including mechanisms of gene regulation

Biological interpretation of RNA-Seq and other gene expression analyses require functional annotations of genes. Gene Ontology (GO) is frequently used to look for biological processes, molecular functions, and cellular compartments that are enriched in the dataset [54]. Recent efforts have substantially improved the functional annotation of the potato genome using a structure-based pipeline that integrates the results of several different functional annotation software [55]. Despite this improvement, manual curation is still required to further refine functional annotations, especially in plants since they have several unique pathways and processes that are not found in other animals or microorganisms. GoMapMan, a recently developed resource for manual curation, consolidation and visualization of functional annotation in plants, has already been used for several crops including potato [56].

4. Regulatory motif discovery in potato

The availability of a high-quality potato reference genome and transcriptome have, in turn, enabled the development of techniques that allow an accurate quantification of gene transcripts that will aid in the understanding of the complexities potato genetics. This includes the analysis of the *cis*-regulatory elements that are flanking genes, which are important because many polymorphisms associated with crop domestication are found in these regions [57]. Studies performed in *Arabidopsis thaliana* and maize have shown that flanking regions can contain potential binding sites for elements regulating important phenotypic characteristics such as nitrogen (N) response and assimilation [58,59]. Therefore, a greater understanding of gene regulatory mechanisms in potato will provide important information for breeding and genetic modification.

The identification and characterization of regulatory elements has remained a challenge. Techniques such as ChIP-sequencing can reveal the binding sites of regulatory elements, such as transcription factors, by taking advantage of chromatin immuno-precipitation (ChIP) along with DNA sequencing. It has long been known that transcription factors bind to the DNA molecule at specific sites and this interaction is fundamental for the regulation of transcription. In addition, regulation at the translational level also involves sequence motifs and proteins binding to them (RNA binding proteins). The sequences these molecules bind to are usually short (6–15 bp) and conserved both among genes and species, and they are referred to as DNA or RNA motifs [60]. Gene regulatory regions also contribute to phenotypic variation. Studies in *Arabidopsis thaliana* show higher densities of SNPs in environmental response and signaling genes compared to housekeeping genes [61]. Regulatory motif discovery will be important in understanding the impact of genetic variation in regulatory regions.

Throughout the years there have been numerous experimental studies linking specific DNA and RNA motifs with certain regulatory mechanisms, as well as specific phenotypes in plants and other

organisms. Collectively, these studies offer great value because they can be used to annotate sequences and they can be mined for potential genetic modification targets. There are several curated databases containing experimentally validated DNA and RNA regulatory motifs of which two of the largest are JASPAR [62,63] and PLACE [64], the latter is specifically focused on motifs found in plants.

There are a limited number of studies on potato gene regulation. Recent approaches have leveraged increasing amounts of sequencing data to characterize not just a promoter region, but also individual motifs and their binding regulatory elements to provide a better understanding of how gene regulation is carried out at a molecular scale. An example of the regulatory importance of the 5' flanking region of genes can be found in the promoters for the Class I patatin family of genes, which encodes several isoforms of patatin and is the most abundant family of proteins found in the tuber of potato. Putative *cis*-regulatory motifs were identified in the 5'-flanking regions, using alignments of previously reported sequence data and searches in the PLACE database [65]. Several conserved occurrences of previously validated motifs were identified and they had associations with plant functions such as light and sucrose responsive transcriptional regulation, transcription enhancers, and response to abiotic stress. Additionally, by artificially adding the upstream flanking region of these genes to other transgenic genes, it is possible to replicate similar sucrose-induced transcription in other tissues of the plant that are not the tuber [65]. The promoters for the pathogenesis-related *PR-10a*, *chitinase C*, stolon-specific *Stgan*, *snakin-1*, *granule-bound starch synthase (GBSS1)*, and *chalcone isomerase* have also been characterized in a similar fashion [66-71].

In crops outside the Solanaceae family, there have been studies specifically linking N metabolism with certain regulatory motifs. A good example is the nitrate-responsive *cis*-element (NRE) that was identified in the *Arabidopsis* NiR1 gene by aligning the upstream flanking region of the gene the same region in other plants. The NRE consists of a highly-conserved 43 bp sequence and there is evidence this regulatory element is sufficient and necessary for nitrate responsive transcription [72]. The NRE can be found in the upstream flanking regions of NiR genes in several crops (e.g. maize, wheat, bean, tobacco). This seems to confirm that the mechanism for transcriptional regulation in response to nitrate may be conserved in many higher plant species [58]. A Yeast 1 Hybrid (Y1H) screening revealed a Ninein-Like Protein (NLP) that binds to the NRE and activates the nitrate-responsive transcription, indicating that NLPs have a regulatory function in nitrate response [73].

The identification and experimental validation of NRE, as well as the characterization of its interaction with NLPs are a good example of the potential knowledge that can be gained from research focused on the discovery of regulatory mechanisms in plants. However, the exclusive use of sequence alignment to identify conserved targets limits the discovery of motifs to well-annotated datasets available in many different plant species. More recent approaches include *de novo* motif discovery, which has also been favored due to the low costs compared with the relative difficulty and cost of finding and characterizing motifs using only *in vitro* or *in vivo* techniques.

Modern algorithms designed for the purpose of *de novo* motif discovery have different approaches, each with their own set of advantages and disadvantages. Three software packages that have been used successfully in plants are: Weeder [60], MEME [74] and Seeder [75]. To increase the probability of discovering and predicting regulatory elements, it is common to analyze a single dataset using different software, which compensates for the strengths and weaknesses of each algorithm [76,77]. The aggregated results obtained from these tools can then be used to search curated motif databases. If regulatory motifs with no previous experimental validation are found,

targeted studies can be designed to determine the biological function of these motifs either *in vitro* or *in vivo*.

A recent study conducted using a *de novo* motif discovery approach was able to identify nine putative *cis*-regulatory motifs in the upstream flanking region of nitrogen responsive genes in three potato cultivars [43]. The nine motifs had close matches to experimentally validated regulatory motif entries in both PLACE and JASPAR. These sequences could be targeted in experimental studies analyzing steady-state nitrogen response and regulation in field-grown potato, which is a pressing concern for potato producers because of the dependence of the crop on nitrogen supplementation. However, future research on motifs and regulatory elements must also take into account the diversity of potato cultivars and varieties, which requires a deeper knowledge of the genetic differences between them.

5. Genome re-sequencing and genetic diversity

The assembly of the potato reference genome and transcriptome was possible thanks to the development of the double monoploid derived from the *Phureja* group [12]. However, most cultivated potatoes are polyploid and highly heterozygous and until recently [78] were originally classified into seven species and nine taxa [2] which could mean that the genomes of potato landraces and native cultivars might differ significantly from the potato reference genome. The complexity of the potato genome has made the genetic differences between these populations difficult to discern. For example, the taxonomy of the group *Solanum* sect. *Petota* (wild potatoes), as well as the appropriate classification of different potato varieties have been a point of debate among specialists for several years [3-5,78-80].

Although different approaches have been employed to classify potato germplasm (morphological, molecular, cytometric), taxonomy remains challenging due to varying ploidy levels, sexual and asexual reproduction, the ease of interspecific hybridization, and introgressions from various wild species. One example of a characteristic that caused some confusion in taxonomic classification in the past is that potato germplasm was frequently classified as a particular species based on ploidy level or ploidy level was assumed based on classification in a particular species. However, molecular studies have demonstrated that ploidy is not a good indicator of taxonomic classification because potato species have been found with mixed ploidy levels within the species [81] [Barkley et al., in preparation]. Current research programs on genetic resources are working on sorting potato taxonomy and making modifications as needed.

Since the release of the potato reference genome, significant amounts of data have been collected using high-throughput sequencing and SNP arrays. These new datasets have mostly supported, with some exceptions, the current taxonomic tree of tuber-bearing *Solanaceae* and provided a general overview of the genetic diversity of these species [10,82]; however, these tools are just starting to be used to discover specific differences in the genome of potato varieties. For example, SNP arrays have been shown to reveal complex relationships, such as, inter- and intraspecific diversity of the wild species [82]. Evaluation of wild genotypes across loci can also potentially reveal valuable information on genes that differentiate primitive and cultivated germplasm, as well as, determine key loci involved in domestication or enhanced agronomic performance of modern varieties [82]. Identification of novel alleles and their potential utilization is a key factor to assist breeding programs in developing improved varieties in order to advance this important crop.

Important structural variations between different varieties of potato have also been uncovered. A study using a Fluorescence *in situ* Hybridization (FISH) based approach concluded that CNVs were highly abundant in potatoes. However, the limitations of that technology made it impossible to accurately determine the distribution and prevalence of CNVs throughout the genome [9].

High-throughput sequencing data was recently used to identify CNVs within a panel of 12 potato monoploids containing diverse genetic backgrounds [19]. Using CNVnator [83], a program developed to detect CNVs by comparing sequencing read depth to a reference genome, the prevalence of CNVs in potato varieties was confirmed. Results show that CNVs cover approximately 30% of the genome and more than 11,500 individual genes, making them one of the major components of genetic diversity. Genes found exclusively in potato, including disease resistance genes, as well as genes previously identified as dispensable were more likely to be affected by CNVs than genes that were highly conserved among angiosperms. Finally, several large scale CNVs (with sizes above 100 kb) were detected, mostly affecting the heterochromatic or peri-centromeric regions of chromosomes, especially chromosomes 5 and 7 [19].

While CNVs provide useful information about the genetic diversity of potato, another promising approach is to assemble different reference genomes for each potato variety using re-sequencing data. Research in humans has shown there are a number of complex structural variants that are difficult to discover without new assemblies, especially when the heterozygosity found in diploid genomes is taken into account [84,85]. A recent *de novo* assembly of a diploid wild potato species (*Solanum commersonii*) revealed significant differences in the distribution of SNPs, a lower degree of heterozygosity, fewer zones of repetitive DNA, and novel genes, when compared to the potato reference genome (see below) [14]. This study, along with additional experiments performed in other Solanaceae crops such as tomato [86], highlight the potential benefits to further sequence, assemble and analyze close potato varieties and close relatives.

However, genome assembly in potato and other plants remains a complex problem and usually requires more data and computational resources than assemblies for microorganisms or even the human genome. More recent efforts to assemble a *de novo* genome in non-model plant species such as *indica* rice, carrot, and pineapple [87-89], have had to rely on new strategies that combine different types of sequencing data, including long-read technologies, to produce better results especially when dealing with repetitive sequences and polyploidy. Other plant genomes have been assembled using long-read data exclusively, such as the desiccation-tolerant grass *Oropetium thomaeum* [90]. Finally, advances in sequencing library preparation methods, such as those developed by 10X Genomics and Dovetail Genomics, have made it possible to approximate the information of long-sequence reads while still using short-read sequencers to generate the data [91,92]. These new technologies have started to be used in plant genome assembly efforts including one wild potato species [93-96].

An alternative way to overcome challenges associated with assembling new plant genomes is to take advantage of the reference genomes that are currently available as a way to reduce the need for more data. A study in *Arabidopsis* used this approach to assemble four new genome sequences for divergent strains. By doing a whole genome alignment of the sequencing reads to the reference genome, the initial dataset was divided into smaller groups of well-aligned, contiguous reads that were then used as input for assembly software. Unaligned reads were assembled separately and integrated at a later stage in the scaffolding process [97]. The results show that using a reference-guided approach effectively increases the coverage of the resulting assemblies. However, reference-guided assemblies had comparatively worse statistics that those produced *de novo,*

including a lower N_{50}. The reference-based assemblies have enabled the discovery of previously unknown variants, including several large-scale variations and experiments, such as those involving small RNA (sRNA), produce better results when aligned to a strain-specific genome than to the generic *Arabidopsis* reference genome [97].

If the purpose of a genome re-sequencing study is to identify all the non-redundant DNA sequences in a particular population, a novel approach has been developed that utilizes a metagenome-like assembly strategy. Briefly, the procedure consists of sequencing all the individuals of the population at a very low coverage, and then using this data in addition to a well annotated reference to identify unique sequences that are present in at least two of the individuals. The effectiveness of this technique has been shown in rice (*Oryza sativa*) where 1483 accessions were sequenced enabling the assembly and mapping of most of the known agronomically important genes that were previously absent from the Nipponbare rice reference genome [98]. In future studies where the detection of large-scale structural variants is not important, this approach can reduce the amount of sequencing data required while still enabling the discovery of novel sequences found in a sub-set of individuals in a population.

6. Genome assembly of a wild potato species

Solanum commersonii is a wild potato species that is sexually incompatible with *S. tuberosum* due to different endosperm balance numbers [99]. Breeding efforts have allowed introgression of alleles from this species into cultivated potato by overcoming the reproductive barriers; however, little progress has been made on the release of new varieties originating from *S. commersonii* [14]. This species has also been shown to be genetically distinct from cultivated potato by chloroplast restriction sites and nitrate reductase gene sequences [100]. *S. commersonii* has generated interest because it contains important agronomic traits such as resistance to root knot nematode, soft rot and blackleg, bacterial and *Verticillium* wilt, potato virus X, tobacco etch virus, common scab, late blight, and the ability to acclimate to the cold/freezing conditions [2,101,102]. Genomic efforts in this species may help reveal important genes or the molecular pathways for specific traits which could be further utilized to improve cultivated potato.

In 2015, the first draft of a wild potato genome was released using a whole genome shotgun sequence and assembly approach based on size selected, paired end and mate pair libraries ranging from 400 bp to 10 kb [14]. The genome size was slightly smaller (830 Mb) than the cultivated form which was mainly due to variations in intergenic sequences. After filtering the data, a total of 278,460 contigs with an N_{50} of 6506 bp were assembled. In total, 64,665 scaffolds greater than 1 kb were produced with a mean scaffold length of 13,543 bp. The potato reference genome was utilized to map *S. commersonii* scaffolds and anchor them on each chromosome, producing 12 pseudomolecules [14].

Even though *S. commersonii* is known to be an allogamous species, it had a low rate of heterozygosity compared to the reference genome of the cultivated variety (1.5% versus 53–59%), which could be due to the maintenance of this germplasm *ex situ,* artificially reducing the diversity level. The repeated sequences were also reduced (44.5% versus 55%) in *S. commersonii* compared to cultivated reference genome *S. tuberosum*. Ty3-gypsy type long terminal repeat retrotransposons (LTR-RTs) were the predominant transposable elements (TEs) identified in the genome, but the lower frequency of TEs found relative to cultivated potato and tomato may also contribute to its smaller genome size. An evaluation of the diversity demonstrated that the majority of SNPs had a

distance of < 50 bp to their nearest neighbor. The divergence time between cultivated potato and *S. commersonii* was estimated to be approximately 2.3 million years ago [14].

Transcriptome data was produced from leaf, flower, stolon, and tubers, from which a total of 37,662 genes were predicted [14]. The annotated genes for *S. commersonii* were evaluated and compared to cultivated potato and tomato. Pathogen resistance (R) genes were compared between *S. commersonii*, *S. tuberosum*, and *S. lycopersicum*. The wild potato had fewer putative R genes than the cultivated form, but more than the tomato genome. Factors such as genome size, natural and artificial selection, polyploidization, breeding, and gene family interactions can all contribute to pathogen resistance gene evolution. It is possible that the R genes in these three species vary due to different pathogenic pressures [14,103]. However, it could also be an artifact due to the disparity in the quality of each assembly. Further evidence will be required to reach a conclusion.

Cold response genes were also compared [14], resulting in 5853 predicted protein sequences revealed in *S. commersonii* and 8666 in *S. tuberosum*. These predicted proteins were similar to cold responsive genes found in the annotation of the *Arabidopsis thaliana* genome. The expression profiles of *S. commersonii* were further investigated to identify the genes involved in freezing and cold acclimation response. A total of 855 genes were determined to be differentially expressed in plants acclimated to frost stress and non-acclimated plants. A total of 11 transcription factors were negatively correlated and 25 were positively correlated to acclimated and non-acclimated tolerance. Collectively, these results show how comparing related genomes can aid scientists in revealing differences in gene function and regulatory elements. Generally conserved sequences across distant species are likely constrained implying similar biological function [104].

7. Genomics and genotyping

Whole genome re-sequencing can reveal important differences between cultivated potato varieties and related wild species, especially at a large scale. Traditionally, the cost of resequencing entire populations of samples has been prohibitive, and thus, there has been a need for novel solutions to genotype large collections of potato germplasm. The recent and tremendous reduction in costs associated with high-throughput sequencing have enabled the development of genetic markers with a single nucleotide resolution that can be rapidly assayed on hundreds to thousands of individuals. These molecular markers can be used in applications such as marker-assisted breeding, quantitative trait loci (QTL) determination, genome-wide association analyses (GWAS), as well as, evolutionary and diversity studies [105].

Genotyping arrays have been the most common tool for high-throughput SNP genotyping in the last decade. Arrays have been developed for multiple platforms (including Infinium and Axiom) and offer many advantages over low-throughput gel-based genotyping platforms: a relatively low cost per sample, automation and standardization that makes it easy to analyze and compare the results of many individual samples. However, regardless of platform, array development is costly, time consuming, and requires extensive knowledge of the target genome. Additionally, researchers that use arrays are limited to the genes or sequences that are included in the platform [106].

There have been several SNP arrays developed for potato. Currently, one of the most popular is the Infinium 8303 Potato Array [107] which was developed using SNPs discovered in two previous studies: one that mined markers from potato EST databases [108] and a second that analyzed cDNA sequences from six elite potato germplasm accessions [109]. As its name suggests, this array

contains 8303 SNP markers chosen to provide roughly even distribution across all 12 potato chromosomes. Out of the total number of markers, 536 were previously used genetic markers, 3018 were selected from candidate genes of interest, and 4749 were selected for maximum genome coverage [107]. This platform has proven useful in a number of studies, including genetic mapping of important agricultural traits [110-113], retrospective analysis of potato breeding [10] and taxonomic studies [82].

A second recently developed SNP platform is the SolSTW array. It includes a total of 14,530 SNP markers, the majority of which were selected from a previous sequence based genotyping experiment [114]. The design of this array was focused on expanding the genetic sources of the markers, reducing biases and making it more useful for applications such as marker-assisted breeding. As opposed to the Infinium array that used the transcriptome of only six elite cultivars as the main source for markers, the majority of the markers in the SolSTW array are derived from a broad sequencing study (see below) that included 84 unique individuals and included chloroplastic DNA [105,114].

As an alternative to genotyping arrays, several new sequencing based genotyping methods have emerged, leveraging high-throughput, short read sequencing to genotype hundreds of individuals simultaneously at thousands of genetic loci. The two most common methods are: genotyping-by-sequencing (GBS) [115] and RAD-seq [116]. Both techniques have become popular in the agricultural genomics and ecological genetics communities respectively. In each case, a small subset of the genome is sequenced at low coverage, providing a relatively cheap tool to identify molecular markers. This reduced representation of the genome is constructed by digesting the genome with restriction enzymes (GBS) or digestion in combination with physical shearing (RAD-seq). The reduced representation libraries from many individuals can be DNA barcoded, pooled, and then sequenced in the same experiment, greatly reducing the cost per sample. Post sequencing analyses can be performed using available software packages and tools, including TASSEL-GBS [117,118], UNEAK [119], Stacks [120], Haplotag [121] and GBS-SNP-CROP [122].

While there are many benefits to using GBS or RAD over genotyping arrays, including no requirement of a complete reference genome, no array ascertainment bias, and the ability to identify multiple types of genetic markers, significant challenges remain. The main obstacle is the sparse genotype matrix that is missing genotype calls, produced during the computational step that calls and filters SNPs and indels. This is due to the finite amount of sequencing data produced in one experiment, which is spread across many sequenced individuals, in other words, the tradeoff of sequencing coverage and depth among multiplexed DNA samples. It is not uncommon to see sequence-based genotyping studies tolerate between 20–50% missing genotype data [115]. Despite this hurdle, GBS has been successfully implemented in genetic mapping studies of diploid crops such as maize (producing 200,000 markers) [115], wheat and barley (producing 20,000 and 34,000 SNPs, respectively) [123], and polyploid crops such as alfalfa (11,694 SNPs) [124].

In potato, there has been limited application of GBS for molecular marker development perhaps due to the highly heterozygous, tetraploid genome. In one instance, however, a modified GBS approach has been successfully used in marker discovery as part of a study that genotyped a panel of 83 tetraploid potato varieties chosen to represent the most important commercial cultivars and landraces worldwide [105]. This study also included a monoploid clone related to the variety used to develop the potato reference genome. In total 12.4 Gb of sequence data were produced, which resulted in the identification of 129,156 markers. Out of that total, ~111 k corresponded to SNPs, ~13 k were insertions or deletions, and ~5 k were multi-nucleotide polymorphisms. These markers

were then successfully used in analyses to determine population structure, sequence diversity, chloroplast type and genetic association [105].

The successful use of GBS in tetraploid potato cultivars opens the door to future studies exploring the wider diversity of commercial and non-commercial potato varieties. Similar studies in other Solanaceae, such as tomato, show the potential benefits of using this technique to explore wild species diversity [125]. Additionally, it has been recently reported that GBS can be used to aid in the analysis of diploid potato mapping populations [126]. A summary of the different SNP genotyping tools discussed in this section can be found in Table 2.

Marker assisted selection (MAS) increases the efficiency of breeding [127]. Markers are identified using genetic mapping, which is hampered in potato by complex tetraploid genetics and heterozygosity. To date most MAS studies in potato have relied on low-throughput molecular markers, including amplified fragment length polymorphism (AFLP) and simple sequence repeats (SSRs) that have been associated with traits with relatively simple genetic basis such as disease resistance. For example, several studies have identified loci associated with resistance to late blight [128], potato virus Y [129-132], potato virus X [130,133] and *Verticillium* wilt [134,135]. In contrast, there are markedly fewer studies focusing on polygenic (i.e. quantitative) traits such as tuber quality [136], and tuber starch and yield [137]. Regardless of trait, many of these low-throughput, gel-based, markers in their current form are not suitable for large scale screening of progenies, which would be required for application in a breeding program. One option would be to convert the gel based markers to a more efficient platform, as has been recently done for potato virus Y resistance markers [132]. A second option would be to validate the existing marker-trait associations with the array- or sequence-based genotyping platforms, and identify SNP markers linked to the trait of interest. The latter option would be preferable, as it could be done as a byproduct of generating genome-wide marker information, which could in turn be used in future QTL mapping or genome wide association for novel traits. The successful use of high-throughput genotyping platforms (Table 2), in potato, opens the door to exploring the wider diversity of potato genetic variation, and the practical application of MAS in breeding programs. Ultimately, the genome-wide marker information could be used to go beyond MAS at a few loci, to being able to predict the phenotype solely from marker genotypes at all marker loci using whole genome selection methods [138].

Bulked segregant analysis (BSA) is emerging as a method for genetic mapping that has a particularly good compatibility with genome re-sequencing. BSA is an approach for gene mapping where pooled DNA from individuals is genotyped as a single bulked sample. The method was originally applied in lettuce using individuals from a single biparental cross that segregated for a downy mildew resistance [95], but it can also be used for three-way, four-way and multiparental crosses, including those developed with special designs such as diallel design, North Carolina design (NCD), multiparent advanced generation intercross (MAGIC) and nested association mapping (NAM) [139]. Traits are quantified for all individuals in the population. Most commonly, individuals at the two extremes ends of the trait distribution are identified and their DNA is pooled, however other pooling strategies have also been used. Genome re-sequencing of the two pools plus two parents is a cost-effective way of getting high density genotyping data. Sequence data are mapped to a reference sequence and base distribution across the genome is analyzed. Detection of trait-associated variants in pooled sequence data involves use of statistical analysis to compare observed base distributions in the pools with that predicted by parental base distributions [140,141]. The selection of individuals for pools, genetic architecture of the trait and population size are other

factors affecting the power of BSA [139]. BSA was successfully used in potato to map steroidal glycoalkaloid content in tetraploids [141]. As sequencing costs drop the use of whole genome sequencing for genotyping will become more widespread.

Table 2. Summary of popular array and GBS platforms in potato.

	Gene Expression Arrays		*SNP-Arrays*		*Genotyping-by-Sequencing (GBS)*
	POCI 44k	JHI *Solanum tuberosum* 60k	Infinium 8303	SolSTW	
SNPs Markers	N/A	N/A	8303	17,987	111,212
Expression Markers	42,034	52,998	N/A	N/A	N/A
Additional Markers[*]	N/A	N/A	N/A	N/A	17,944
Total Number of Markers	42,034	52,998	8303	17,987	129,156
Year	2005	2013	2012	2015	2013
Source of Genetic Information	Previous data on differentially expressed transcripts and a custom text mining approach for conserved sequences.	Predicted transcripts from the potato reference genome v3.4.	Transcriptomic data from previous experiments, selected for representation of genes of interest and maximum genome coverage.	A combination of GBS derived markers and previously included markers in the Infinium 8303 array.	A panel of 83 tetraploid potato cultivars selected to represent the global gene pool of commercial potato, mostly covering accessions with high breeding value.
Comments	Using the Agilent 60-mer oligo platform.	Using the Agilent 60-mer oligo platform.	Some markers were mapped to the unanchored superscaffold of the potato reference genome.	Includes a small portion of chloroplast markers.	-
Reference	[28]	[29]	[107]	[114]	[105]

[*] Including insertions, deletions and other multinucleotide polymorphisms.

8. Conclusion

Overall, modern sequencing technologies have fundamentally changed the field of plant genomics. It is now possible to identify large structural variations among closely related species, something that was extremely challenging just few years ago. These new resources provide scientists

and producers with better tools to continue working on the discovery of new genes and regulatory mechanisms. In turn, knowledge generated this way can inform future crop improvement efforts. In the case of tuber bearing Solanaceae, there is already a fair amount of evidence pointing to important genetic differences within these species. A summary of additional genomics resources for potato and related species can be found in Table 3. However, more research is required especially in wild relatives of commercial potato, which could be important sources of genetic diversity but have remained relatively unexplored so far.

Table 3. Summary of genomics resources available for potato and related species.

Name of resource	*Description*	*URL*	*Ref.*
Potato Genomics Resources			
Spud DB: Potato Genomics Resource	Latest versions of the potato reference genome, as well as a genome browser and several other potato genomics resources.	http://solanaceae.plantbiology.msu.edu/index.shtml	[142]
NCBI Genome (Potato)	The reference genome listed for *S. tuberosum* in the NCBI Genome database.	https://www.ncbi.nlm.nih.gov/genome/400	[52]
NCBI GEO (Potato)	Gene expression datasets for *S. tuberosum*	https://www.ncbi.nlm.nih.gov/gds/?term=Solanum+tuberosum	[52]
ArrayExpress (Potato)	Array-based gene expression datasets for *S. tuberosum.*	https://www.ebi.ac.uk/arrayexpress/search.html?query=Solanum+tuberosum	[143]
PoMaMo Database	Database containing potato genomic maps and sequences.	http://www.gabipd.org/projects/Pomamo/#Tools	[144]
The NSF Potato Genome Project	Portal containing several potato genomics resources including SSR and microarrays.	http://potatogenome.berkeley.edu/nsf5/	N/A
Potato Variety Databases			
The Potato Association of America Variety Database	Catalogue of potato varieties in the US.	http://potatoassociation.org/industry/varieties#Breeding	N/A
Canadian Potato Varieties Database	Catalogue of potato varieties in Canada.	http://www.inspection.gc.ca/plants/potatoes/potatovarieties/eng/1299172436155/1299172577580	N/A
European Cultivated Potato Database	Catalogue of European potato varieties.	https://www.europotato.org/menu.php	N/A
AHDB Potato Variety Database	Agriculture & Horticulture Development Board Catalogue of British potato varieties.	http://varieties.ahdb.org.uk/	N/A
Potato Germplasm Banks			
International Potato Center (CIP) Genebank	Worldwide collection of potato and sweet potato varieties and wild relatives.	http://cipotato.org/genebank/	N/A

NRSP-6 - United States Potato Genebank	Collection of germplasm of cultivated potato varieties and wild.	http://www.ars-grin.gov/ars/MidWest/NR6/	N/A
Centre for Genetic Resources, The Netherlands (CGN)	Dutch-German collection of wild and Andean cultivated species.	http://www.wur.nl/en/Expertise-Services/Statutory-research-tasks/Centrefor-Genetic-Resources-the-Netherlands-1/Centre-for-Genetic-Resourcesthe-Netherlands-1/Expertise-areas/Plant-Genetic-Resources/CGN-cropcollections/Potato.htm	N/A
N. I Vavilov Institute of Plant Genetic Resources (VIR)	Wild *Solanum* species, cultivatd species and indigenous Chilean cultivars, breeding varieties, hybrids and dihaploids.	http://vir.nw.ru	N/A
Canadian Potato Genetic Resources	Collection of Canadian and international potato germplasm that is part of Plant Gene Resources of Canada.	http://pgrc3.agr.gc.ca/index_e.html	N/A
Commonwealth Potato Collection	United Kingdom genebank of landrace and wild potatoes.	http://germinate.hutton.ac.uk/germinate_cpc/app/	N/A
Other Solanaceae Resources			
Sol Genomics Network	A variety of genomics resources for several of the most important Solanaceae species.	https://solgenomics.net/	[145]
Solanaceae Coordinated Agricultural Project (SolCAP)	A collection of germplasm, phenotype and genotype data on several Solanaceae species.	http://solcap.msu.edu/index.shtml	[107]
GoMapMan	Open-source for manual gene functional annotations in plants, including potato, tomato and tobacco.	http://www.gomapman.org/	[56]

Acknowledgments

The authors acknowledge funding through a Nouvelles Initiatives (Project International) grant from the Centre SÈVE (Fonds de recherche du Québec-Nature et technologies (FRQ-NT) to M.V.S., N.B., D.E., and H.H.T.; the Natural Sciences and Engineering Research Council of Canada (NSERC) (Grant No. 283303) to M.V.S.; A-base funding from Agriculture and Agri-Food Canada to H.H.T. and K.G.; the Mexican National Council of Science and Technology (Consejo Nacional de Ciencia y Tecnología; CONACYT) (Scholarship No. 381158) to J.H.G.; and the McGill Department of Plant Science Graduate Excellence Fund.

Conflict of Interest

The authors declare no competing interests

References

1. Food and Agriculture Organization. Food and Agricultural commodities production / Commodities by regions. FAOSTAT, 2016. Available from: http://faostat3.fao.org/browse/rankings/commodities_by_regions/E
2. Hawkes JG (1990) *The potato: evolution, biodiversity and genetic resources*. Belhaven Press.
3. Machida-Hirano R (2015) Diversity of potato genetic resources. *Breed Sci* 65: 26-40.
4. Ovchinnikova A, Krylova E, Gavrilenko T, et al. (2011) Taxonomy of cultivated potatoes (Solanum section Petota: Solanaceae). *Bot J Linn Soc* 165: 107-155.
5. Huamán Z, Spooner DM (2002) Reclassification of landrace populations of cultivated potatoes (Solanum sect. Petota). *Am J Bot* 89: 947-965.
6. Gebhardt C, Ballvora A, Walkemeier B, et al. (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Mol Breed* 13: 93-102.
7. Simko I, Haynes KG, Jones RW (2006) Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. *Genetics* 173: 2237-2245.
8. Douches DS, Jastrzebski K, Maas D, et al. (1996) Assessment of potato breeding over the past century. *Crop Sci* 36: 1544-1552.
9. Iovene M, Zhang T, Lou Q, et al. (2013) Copy number variation in potato - An asexually propagated autotetraploid species. *Plant J* 75: 80-89.
10. Hirsch CN, Hirsch CD, Felcher K, et al. (2013) Retrospective view of North American potato (Solanum tuberosum L.) breeding in the 20th and 21st centuries. *G3* 3: 1003-1013.
11. Massa AN, Childs KL, Lin H, et al. (2011) The Transcriptome of the Reference Potato Genome Solanum tuberosum Group Phureja Clone DM1-3 516R44. *PLoS One* 6: 1-8.
12. The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475: 189-195.
13. Sharma SK, Bolser D, de Boer J, et al. (2013) Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3* 3: 2031-2047.
14. Aversano R, Contaldi F, Ercolano MR, et al. (2015) The Solanum commersonii Genome Sequence Provides Insights into Adaptation to Stress Conditions and Genome Evolution of Wild Potato Relatives. *Plant Cell* 27: 954-968.
15. The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485: 635-641.
16. Kim S, Park M, Yeom S-I, et al. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nat Genet* 46: 270-278.
17. Sierro N, Battey JND, Ouadi S, et al. (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun* 5: 3833.
18. Bombarely A, Moser M, Amrad A, et al. (2016) Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. *Nat Plants* 2: 16074.

19. Hardigan MA, Crisovan E, Hamilton JP, et al. (2016) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated Solanum tuberosum. *Plant Cell* 28: 388-405.
20. Luo R, Liu B, Xie Y, et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1: 18.
21. Crookshanks M, Emmersen J, Welinder KG, et al. (2001) The potato tuber transcriptome: analysis of 6077 expressed sequence tags. *FEBS Lett* 506: 123-126.
22. Ronning CM, Stegalkina SS, Ascenzi RA, et al. (2003) Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol* 131: 419-429.
23. Flinn B, Rothwell C, Griffiths R, et al. (2005) Potato expressed sequence tag generation and analysis using standard and unique cDNA libraries. *Plant Mol Biol* 59: 407-433.
24. Rensink W, Hart A, Liu J, et al. (2005) Analyzing the potato abiotic stress transcriptome using expressed sequence tags. *Genome* 48: 598-605.
25. Rensink WA, Lee Y, Liu J, et al. (2005) Comparative analyses of six solanaceous transcriptomes reveal a high degree of sequence conservation and species-specific transcripts. *BMC Genomics* 6: 124.
26. Kloosterman B, Vorst O, Hall RD, et al. (2005) Tuber on a chip: Differential gene expression during potato tuber development. *Plant Biotechnol J* 3: 505-519.
27. Rensink WA, Iobst S, Hart A, et al. (2005) Gene expression profiling of potato responses to cold, heat, and salt stress. *Funct Integr Genomics* 5: 201-207.
28. Kloosterman B, De Koeyer D, Griffiths R, et al. (2008) Genes driving potato tuber initiation and growth: Identification based on transcriptional changes using the POCI array. *Funct Integr Genomics* 8: 329-340.
29. Bengtsson T, Weighill D, Proux-Wéra E, et al. (2014) Proteomics and transcriptomics of the BABA-induced resistance response in potato using a novel functional annotation approach. *BMC Genomics* 15: 315.
30. Bachem C, Van Der Hoeven R, Lucker J, et al. (2000) Functional genomic analysis of potato tuber life-cycle. *Potato Res* 43: 297-312.
31. Campbell M, Segear E, Beers L, et al. (2008) Dormancy in potato tuber meristems: Chemically induced cessation in dormancy matches the natural process based on transcript profiles. *Funct Integr Genomics* 8: 317-328.
32. Navarro C, Abelenda, JA, Cruz-Oró E, et al. (2011) Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature* 478: 119-122.
33. Restrepo S, Myers KL, del Pozo O, et al. (2005) Gene profiling of a compatible interaction between Phytophthora infestans and Solanum tuberosum suggests a role for carbonic anhydrase. *Mol Plant Microb Interact* 18: 913-922.
34. Tai HH, Goyer C, Platt HW, et al. (2013) Decreased defense gene expression in tolerance versus resistance to Verticillium dahliae in potato. *Funct Integr Genomics* 13: 367-378.
35. Schafleitner R, Gutierrez Rosales RO, Gaudin A, et al. (2007) Capturing candidate drought tolerance traits in two native Andean potato clones by transcription profiling of field grown plants under water stress. *Plant Physiol Biochem* 45: 673-690.
36. Ginzberg I, Barel G, Ophir R, et al. (2009) Transcriptomic profiling of heat-stress response in potato periderm. *J Exp Bot* 60: 4411-4421.

37. Evers D, Lefèvre I, Legay S, et al. (2010) Identification of drought-responsive compounds in potato through a combined transcriptomic and targeted metabolite approach. *J Exp Bot* 61: 2327-2343.
38. Hancock RD, Morris WL, Ducreux LJM, et al. (2014) Physiological, biochemical and molecular responses of the potato (Solanum tuberosum L.) plant to moderately elevated temperature. *Plant Cell Environ* 37: 439-450.
39. Hammond JP, Broadley MR, Bowen HC, et al. (2011) Gene expression changes in phosphorus deficient potato (Solanum tuberosum L.) leaves and the potential for diagnostic gene expression markers. *PLoS One* 6: e24606.
40. Trapnell C, Williams BA, Pertea G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28: 511-515.
41. Swarbreck D, Wilks C, Lamesch P, et al. (2008) The Arabidopsis Information Resource (TAIR): Gene structure and function annotation. *Nucleic Acids Res* 36: 1009-1014.
42. Hazen SP, Wu Y, Kreps JA (2003) Gene expression profiling of plant responses to abiotic stress. *Funct Integr Genomics* 3: 105-111.
43. Gálvez JH, Tai HH, Lagüe M, et al. (2016) The nitrogen responsive transcriptome in potato (Solanum tuberosum L.) reveals significant gene regulatory motifs. *Sci Rep* 6: 26090.
44. Cho K, Cho KS, Sohn HB, et al. (2016) Network analysis of the metabolome and transcriptome reveals novel regulation of potato pigmentation. *J Exp Bot* 67: 1519-1533.
45. Liu B, Zhang N, Wen Y, et al. (2015) Transcriptomic changes during tuber dormancy release process revealed by RNA sequencing in potato. *J Biotechnol* 198: 17-30.
46. Goyer A, Hamlin L, Crosslin JM, et al. (2015) RNA-Seq analysis of resistant and susceptible potato varieties during the early stages of potato virus Y infection. *BMC Genomics* 16: 472.
47. Frades I, Abreha KB, Proux-Wéra E, et al. (2015) A novel workflow correlating RNA-seq data to Phythophthora infestans resistance levels in wild Solanum species and potato clones. *Front Plant Sci* 6: 718.
48. Zhang N, Yang J, Wang Z, et al. (2014) Identification of novel and conserved microRNAs related to drought stress in potato by deep sequencing. *PLoS One* 9: e95489.
49. Shan J, Song W, Zhou J, et al. (2013) Transcriptome analysis reveals novel genes potentially involved in photoperiodic tuberization in potato. *Genomics* 102: 388-396.
50. Gao L, Tu ZJ, Millett BP, et al. (2013) Insights into organ-specific pathogen defense responses in plants: RNA-seq analysis of potato tuber-Phytophthora infestans interactions. *BMC Genomics* 14: 340.
51. Zuluaga AP, Solé M, Lu H, et al. (2015) Transcriptome responses to Ralstonia solanacearum infection in the roots of the wild potato Solanum commersonii. *BMC Genomics* 16: 246.
52. Wheeler DL, Barrett T, Benson DA, et al. (2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 35: 5-12.
53. Hirsch CD, Springer NM, Hirsch CN (2015) Genomic Limitations to RNAseq Expression Profiling. *Plant J* 84: 491-503.
54. Ashburner M, Ball CA, Blake JA, et al. (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* 25: 25-29.
55. Amar D, Frades I, Danek A, et al. (2014) Evaluation and integration of functional annotation pipelines for newly sequenced organisms: the potato genome as a test case. *BMC Plant Biol* 14: 1-14.

56. Ramšak Ž, Baebler Š, Rotter A, et al. (2014) GoMapMan: Integration, consolidation and visualization of plant gene annotations within the MapMan ontology. *Nucleic Acids Res* 42: 1167-1175.
57. Swinnen G, Goossens A, Pauwels L (2016) Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. *Trends Plant Sci* 21: 506-515.
58. Konishi M, Yanagisawa S (2011) Roles of the transcriptional regulation mediated by the nitrate-responsive cis-element in higher plants. *Biochem Biophys Res Commun* 411: 708-713.
59. Liseron-Monfils C, Bi Y-M, Downs GS, et al. (2013) Nitrogen transporter and assimilation genes exhibit developmental stage-selective expression in maize (Zea mays L.) associated with distinct cis-acting promoter motifs. *Plant Signal Behav* 8: 1-14.
60. Pavesi G, Zambelli F, Pesole G (2007) WeederH: an algorithm for finding conserved regulatory motifs and regions in homologous sequences. *BMC Bioinformatics* 8: 46.
61. Korkuc P, Schippers JHM, Walther D (2014) Characterization and Identification of cis-Regulatory Elements in Arabidopsis Based on Single-Nucleotide Polymorphism Information. *Plant Physiol* 164: 181-200.
62. Sandelin A, Alkema W, Engström P, et al. (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* 32: D91-D94.
63. Mathelier A, Zhao X, Zhang AW, et al. (2014) JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 42: 142-147.
64. Higo K, Ugawa Y, Iwamoto M, et al. (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27: 297-300.
65. Aminedi R, Das N (2014) Class I patatin genes from potato (Solanum tuberosum L.) cultivars: molecular cloning, sequence comparison, prediction of diverse cis-regulatory motifs, and assessment of the promoter activities under field and in vitro conditions. *Vitr Cell Dev Biol Plant* 50: 673-687.
66. Chen M, Zhu WJ, You X, et al. (2015) Isolation and characterization of a chalcone isomerase gene promoter from potato cultivars. *Genet Mol Res* 14: 18872-18885.
67. Bansal A, Kumari V, Taneja D, et al. (2012) Molecular cloning and characterization of granule-bound starch synthase I (GBSSI) alleles from potato and sequence analysis for detection of cis-regulatory motifs. *Plant Cell Tissue Organ Cult* 109: 247-261.
68. Almasia NI, Narhirñak V, Hopp HE, et al. (2010) Isolation and characterization of the tissue and development-specific potato snakin-1 promoter inducible by temperature and wounding. *Electron J Biotechnol* 13: 1-21.
69. Trindade LM, Horvath B, Bachem C, et al. (2003) Isolation and functional characterization of a stolon specific promoter from potato (Solanum tuberosum L.). *Gene* 303: 77-87.
70. Ancillo G, Hoegen E, Kombrink E (2003) The promoter of the potato chitinase C gene directs expression to epidermal cells. *Planta* 217: 566-576.
71. Despres C, Subramaniam R, Matton DP, et al. (1995) The Activation of the Potato Pr-Loa Gene Requires the Phosphorylation of the Nuclear Factor Pbf-1. *Plant Cell* 7: 589-598.
72. Konishi M, Yanagisawa S (2010) Identification of a nitrate-responsive cis-element in the Arabidopsis NIR1 promoter defines the presence of multiple cis-regulatory elements for nitrogen response. *Plant J* 63: 269-282.
73. Konishi M, Yanagisawa S (2013) Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. *Nat Commun* 4: 1617.

74. Bailey TL, Boden M, Buske FA, et al. (2009) MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res* 37: 202-208.
75. Fauteux F, Blanchette M, Strömvik MV (2008) Seeder: Discriminative seeding DNA motif discovery. *Bioinformatics* 24: 2303-2307.
76. López Y, Patil A, Nakai K (2013) Identification of novel motif patterns to decipher the promoter architecture of co-expressed genes in Arabidopsis thaliana. *BMC Syst Biol* 7: S10.
77. Zolotarov Y, Strömvik M (2015) De Novo Regulatory Motif Discovery Identifies Significant Motifs in Promoters of Five Classes of Plant Dehydrin Genes. *PLoS One* 10: e0129016.
78. Spooner DM, Ghislain M, Simon R, et al. (2014) Systematics, Diversity, Genetics, and Evolution of Wild and Cultivated Potatoes. *Bot Rev* 80: 283-383.
79. Spooner DM (2009). DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *Am J Bot* 96: 1177-1189.
80. Spooner DM, Núñez J, Trujillo G, et al. (2007) Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. *Proc Natl Acad Sci U S A* 104: 19398-19403.
81. Ghislain M, Andrade D, Rodríguez F, et al. (2006) Genetic analysis of the cultivated potato Solanum tuberosum L. Phureja Group using RAPDs and nuclear SSRs. *Theor Appl Genet* 113: 1515-1527.
82. Hardigan MA, Bamberg J, Buell CR, et al. (2015) Taxonomy and Genetic Differentiation among Wild and Cultivated Germplasm of sect. Petota. *Plant Genome* 8: 1-16.
83. Abyzov A, Urban AE, Snyder M, et al. (2011) CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 21: 974-984.
84. Chaisson MJP, Wilson RK, Eichler EE (2015) Genetic variation and the de novo assembly of human genomes. *Nat Rev Genet* 16: 627-640.
85. Pendleton M, Sebra R, Pang AWC, et al. (2015) Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat Methods* 12: 780-786.
86. Aflitos S, Schijlen E, De Jong H, et al. (2014) Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. *Plant J* 80: 136-148.
87. Ming R, VanBuren R, Wai CM, et al. (2015) The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet* 47: 1435-1442.
88. Iorizzo M, Ellison S, Senalik D, et al. (2016) A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat Genet* 48: 657-666.
89. Mahesh HB, Shirke MD, Singh S, et al. (2016) Indica rice genome assembly, annotation and mining of blast disease resistance genes. *BMC Genomics* 17: 242.
90. VanBuren R, Bryant D, Edger PP, et al. (2015) Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. *Nature* 527: 508-511.
91. Eisenstein M (2015) Startups use short-read data to expand long-read sequencing market. *Nat Biotechnol* 33: 433-435.
92. Putnam NH, Connell BO, Stites JC, et al. (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* 26: 342-350.
93. Paajanen PM, Giolai M, Verweij W, et al. S. verrucosum, a Wild Mexican Potato As a Model Species for a Plant Genome Assembly Project. Plant and Animal Genome XXIV Conference, 2016. Available from: https://pag.confex.com/pag/xxiv/webprogram/Paper20356.html

94. Bredeson JV, Lyons JB, Prochnik SE, et al. (2016) Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotechnol* 34: 562-570.
95. Michelmore R, Reyes Chin-Wo S, Kozik A, et al. Improvement of the Genome Assembly of Lettuce (Lactuca sativa) Using Dovetail/in vitro Proximity Ligation. Plant and Animal Genome XXIV Conference, 2016. Available from: https://pag.confex.com/pag/xxiv/webprogram/Paper22314.html
96. Reyes Chin-Wo S, Lavelle D, Truco MJ, et al. Dovetail/in vitro Proximity Ligation Data Facilitates Analysis of an Ancient Whole Genome Triplication Event in Lactuca sativa. Plant and Animal Genome XXIV Conference, 2016. Available from: https://pag.confex.com/pag/xxiv/webprogram/Paper19305.html
97. Schneeberger K, Ossowski S, Ott F, et al. (2011) Reference-guided assembly of four diverse Arabidopsis thaliana genomes. *Proc Natl Acad Sci U S A* 108: 10249-10254.
98. Yao W, Li G, Zhao H, et al. (2015) Exploring the rice dispensable genome using a metagenome-like assembly strategy. *Genome Biol* 16: 187.
99. Johnston SA, den Nijs TPM, Peloquin SJ, et al. (1980) The significance of genic balance to endosperm development in interspecific crosses. *Theor Appl Genet* 57: 5-9.
100. Rodríguez F, Spooner DM (2009) Nitrate Reductase Phylogeny of Potato (Solanum sect. Petota) Genomes with Emphasis on the Origins of the Polyploid Species. *Syst Bot* 34: 207-219.
101. Hanneman RE, Bamberg JB (1986) *Inventory of tuber- bearing Solanum species*. University of Wisconsin Press.
102. Micheletto S, Boland R, Huarte M (2000) Argentinian wild diploid Solanum species as sources of quantitative late blight resistance. *Theor Appl Genet* 101: 902-906.
103. Andolfo G, Jupe F, Witek K, et al. (2014) Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC Plant Biol* 14: 120.
104. Alföldi J, Lindblad-Toh K (2013) Comparative genomics as a tool to understand evolution and disease. *Genome Res* 23: 1063-1068.
105. Uitdewilligen JGAML, Wolters AMA, D'hoop BB, et al. (2013) A Next- Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. *PLoS One* 8: 10-14.
106. De Donato M, Peters SO, Mitchell SE, et al. (2013) Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next- generation sequencing. *PLoS One* 8: e62137.
107. Felcher KJ, Coombs JJ, Massa AN, et al. (2012) Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS One* 7: e36347.
108. Anithakumari AM, Tang J, van Eck HJ, et al. (2010) A pipeline for high throughput detection and mapping of SNPs from EST databases. *Mol Breed* 26: 65-75.
109. Hamilton JP, Hansey CN, Whitty BR, et al. (2011) Single nucleotide polymorphism discovery in elite north american potato germplasm. *BMC Genomics* 12: 302.
110. Massa AN, Manrique-Carpintero NC, Coombs JJ, et al. (2015) Genetic Linkage Mapping of Economically Important Traits in Cultivated Tetraploid Potato (Solanum tuberosum L.). *G3* 5: 2357-2364.

111. Manrique-Carpintero NC, Coombs JJ, Cui Y, et al. (2015) Genetic map and QTL analysis of agronomic traits in a diploid potato population using single nucleotide polymorphism markers. *Crop Sci* 55: 2566-2579.
112. Manrique-Carpintero NC, Coombs JJ, Veilleux RE, et al. (2016) Comparative Analysis of Regions with Distorted Segregation in Three Diploid Populations of Potato. *G3* 6: 2617-2628.
113. Endelman JB, Jansky SH (2016) Genetic mapping with an inbred line-derived F2 population in potato. *Theor Appl Genet* 129: 935-943.
114. Vos PG, Uitdewilligen JGAML, Voorrips RE, et al. (2015) Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. *Theor Appl Genet* 128: 2387-2401.
115. Elshire RJ, Glaubitz JC, Sun Q, et al. (2011) A robust, simple genotyping-by- sequencing (GBS) approach for high diversity species. *PLoS One* 6: e19379.
116. Baird NA, Etter PD, Atwood TS, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3: e3376.
117. Bradbury PJ, Zhang Z, Kroon DE, et al. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635.
118. Glaubitz JC, Casstevens TM, Lu F, et al. (2014) TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS One* 9: e90346
119. Lu F, Lipka AE, Glaubitz J, et al. (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet* 9: e1003215.
120. Catchen JM, Amores A, Hohenlohe P, et al. (2011) Stacks: building and genotyping Loci de novo from short-read sequences. *G3* 1: 171-182.
121. Tinker NA, Bekele WA, Hattori J (2016) Haplotag: Software for Haplotype- Based Genotyping-by-Sequencing Analysis. *G3* 6: 857-863.
122. Melo ATO, Bartaula R, Hale I (2016) GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinformatics* 17: 29.
123. Poland JA, Brown PJ, Sorrells ME, et al. (2012) Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by- Sequencing Approach. *PLoS One* 7: e32253.
124. Rocher S, Jean M, Castonguay Y, et al. (2015) Validation of genotyping-by- sequencing analysis in populations of tetraploid alfalfa by 454 sequencing. *PLoS One* 10: e0131918.
125. Labate JA, Robertson LD, Strickler SR, et al. (2014) Genetic structure of the four wild tomato species in the Solanum peruvianum s.l. species complex. *Genome* 57: 169-180.
126. Endelman J. Genotyping-By-Sequencing of a Diploid Potato F2 Population. Plant and Animal Genome XXIII, 2015. Available from: https://pag.confex.com/pag/xxiii/webprogram/Paper15683.html
127. Barone A (2004) Molecular marker-assisted selection for potato breeding. *Am J Potato Res* 81: 111-117.
128. Tiwari JK, Siddappa S, Singh BP, et al. (2013) Molecular markers for late blight resistance breeding of potato: an update. *Plant Breed* 132: 237-245.
129. Song Y-S, Hepting L, Schweizer G, et al. (2005) Mapping of extreme resistance to PVY (Ry sto) on chromosome XII using anther-culture-derived primary dihaploid potato lines. *Theor Appl Genet* 111: 879-887.

130. Gebhardt C, Bellin D, Henselewski H, et al. (2006) Marker-assisted combination of major genes for pathogen resistance in potato. *Theor Appl Genet* 112: 1458-1464.
131. Fulladolsa AC, Navarro FM, Kota R, et al. (2015) Application of Marker Assisted Selection for Potato Virus Y Resistance in the University of Wisconsin Potato Breeding Program. *Am J Potato Res* 92: 444-450.
132. Nie X, Sutherland D, Dickison V, et al. (2016) Development and Validation of High-Resolution Melting Markers Derived from Ry sto STS Markers for High-Throughput Marker-Assisted Selection of Potato Carrying Ry sto. *Phytopathology* 106: 1366-1375.
133. Ritter E, Debener T, Barone A, et al. (1991) RFLP mapping on potato chromosomes of two genes controlling extreme resistance to potato virus X (PVX). *Mol Gen Genet* 227: 81-85.
134. Simko I, Haynes KG, Ewing EE, et al. (2004) Mapping genes for resistance to Verticillium albo-atrum in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. *Mol Genet Genomics* 271: 522-531.
135. Uribe P, Jansky S, Halterman D (2014) Two CAPS markers predict Verticillium wilt resistance in wild Solanum species. *Mol Breed* 33: 465-476.
136. Li L, Tacke E, Hofferbert H-R, et al. (2013) Validation of candidate gene markers for marker-assisted selection of potato cultivars with improved tuber quality. *Theor Appl Genet* 126: 1039-1052.
137. Schönhals EM, Ortega F, Barandalla L, et al. (2016). Identification and reproducibility of diagnostic DNA markers for tuber starch and yield optimization in a novel association mapping population of potato (Solanum tuberosum L.). *Theor Appl Genet* 129: 767-785.
138. Slater AT, Cogan NOI, Forster JW, et al. (2016) Improving Genetic Gain with Genomic Selection in Autotetraploid Potato. *Plant Genome* 9: 1-15.
139. Zou C, Wang P, Xu Y (2016) Bulked sample analysis in genetics, genomics and crop improvement. *Plant Biotechnol J* 14: 1941-1955.
140. Bansal V (2010) A statistical method for the detection of variants from next-generation resequencing of DNA pools. *Bioinformatics* 26: i318-i324.
141. Kaminski KP, Kørup K, Andersen MN, et al. (2016) Next Generation Sequencing Bulk Segregant Analysis of Potato Support that Differential Flux into the Cholesterol and Stigmasterol Metabolite Pools Is Important for Steroidal Glycoalkaloid Content. *Potato Res* 59: 81-97.
142. Hirsch CD, Hamilton JP, Childs KL, et al. (2014) Spud DB: A Resource for Mining Sequences, Genotypes, and Phenotypes to Accelerate Potato Breeding. *Plant Genome* 7: 1-12.
143. Brazma A, Parkinson H, Sarkans U, et al. (2003) ArrayExpress - A public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 31: 68-71.
144. Meyer S, Nagel A, Gebhardt C (2005) PoMaMo — a comprehensive database for potato genome data. *Nucleic Acids Res* 33: 666-670.
145. Fernandez-Pozo N, Menda N, Edwards JD, et al. (2014) The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. *Nucleic Acids Res* 43: 1-6.

AIMS Press