Government of Saskatchewan
Ministry of Agriculture

ADF Project Final Report Format

1. Project title and ADF file number.

20150123, Comparative Genomics of Apomictic Plants: Advancing Novel Tools for Niche Breeding

2. Name of the Principal Investigator and contact information.

Professor Timothy Francis Sharbel; Global Institute for Food Security (GIFS), University of Saskatchewan

3. Name of the collaborators and contact information.

Philippe Riggault, Gydle (http://www.gydle.com)

4. Abstract/ Summary: *This must include project objectives, results, and conclusions for use in publications and in the Ministry database. Maximum of 300 words in lay language.*

Crop breeding requires many (8 to 10) generations in order to produce a new and improved variety at considerable costs. Apomixis, a naturally occurring form of asexual reproduction in plants that is found in several plant families, is a tool that would potentially significantly reduce both the time and costs associated with developing novel crop varieties. Apomixis technology is a game changer for industry, farmers and consumers, and as such presents a plethora of social and economic benefits, together with many challenges for intellectual property sharing, as farmers will become innovators for niche-adapted varieties that could subsequently be bred apomictically by industry. The introduction of apomixis into agriculture represents a disruptive enabling/transformative technology that would benefit both breeders and users by significantly simplifying the fixation and preservation of genetic heterozygosity and its associated hybrid vigor in crop plants, for example by enabling single generation hybrid seed production without the need for inbred lines. As such, the resultant advances in crop diversification arising through apomictic breeding would have immense positive impacts on food security in both advanced and developing economies.

Here we have completed a gold-standard assembled genome of a sexual *Boechera stricta*, and are in the process of finalizing the assemblies of a further 3 apomictic genomes from the same genus, using the sexual genome as a reference. Importantly, the ongoing statistical and bioinformatics analyses of these assembled genomes will enable us to (1) identify DNA polymorphisms that that are specific to either sexual or apomictic reproduction, and (2) elucidate conserved structural variation in the apomictic genomes which have arisen from the disturbed meiosis which characterizes asexual reproduction

5. Introduction: *Brief project background and rationale.*

The world's population is growing and demand for food is beginning to outpace production, resulting in a significant food security challenge on a global scale. Finite land, water and mineral resources prevent producers from simply expanding into undeveloped areas to address capacity needs. Furthermore, the rate of innovation leading to efficient breeding and increased yields has both slowed and become costlier (Phillips McDougall 2011). Agriculture is thus in need of new technologies that (1) increase the rate of production of novel crop varieties; (2) enable the immediate and faithful propagation of desired crop traits; and (3) provide incentives to farmers and industry to work together as innovators for crop breeding.

These 3 criteria are filled by apomixis, a naturally occurring form of asexual reproduction in plants which presents a key technology for crop variety improvement. Its introduction into crop plants would enable the permanent fixation of any heterozygous genotype, regardless of its genomic heredity and phenotypic complexity, in a single generation. By maintaining heterozygosity in offspring, apomixis is superior to doubled haploid production, another technology to fix certain traits. The impact on breeding programs would be enormous, as the fixation of traits demonstrating heterosis in a single generation would permit the rapid development of superior crop varieties adapted to changing environmental conditions, diverse farming niches and systems, and evolving markets.

Apomixis is a game-changing technology that would significantly benefit the breeding industry, whose

present investment of resources into the generation and maintenance of inbred lines could instead be redirected to the development of a more diverse germplasm. The estimated market improvement value of having apomixis introduced to rice alone would be $2.5 billion per year (*not corrected for 2014 market values*; Spillane *et al.*, 2004). Additionally, it has been estimated that apomixis would greatly benefit crops which are presently propagated vegetatively, for example true seed propagation of potato and cassava have an estimated value of $3.2 billion per year (Spillane *et al.*, 2004). The multiple uses of apomixis in agriculture (Jefferson, 1994) have long been recognized as important steps for sustainability and food security (Toennissen, 2001), although as of yet apomixis has not yet been successfully introduced into crops.

Apomixis is a modification of normal sexual seed production with three developmental steps: the production of meiotically-unreduced egg cells (*apomeiosis*), the embryogenic development of these egg cells without fertilization *(parthenogenesis)*, and the formation of endosperm with (*pseudogamy*) or without (*autonomous aopomixis*) fertilization. Despite being a developmentally complex trait, mapping studies have repeatedly demonstrated that apomixis is inherited in a Mendelian fashion (see Albertini *et al.*, 2010). As with all fields of biology, next-generation sequencing (NGS) technologies were rapidly embraced by the apomixis research community, and deep sequencing provides the possibility of defining the sequence polymorphisms that are associated with apomixis. While the discovery and subsequent genetic transformation of such information into crops is the most direct pathway to the introduction of apomixis into agriculture, current NGS limits this leap because of its short-read length.

Here, a proposal is presented to (1) use a hybrid sequencing approach (short and long read sequencing, in addition to Irys optical genome mapping) to assemble the complex genomes of 3 sexual and 3 apomictic *Boechera*, and (2) to perform a genome-wide isoform analysis (identification and expression levels of mRNA variants arising from the same gene) of microdissected reproductive tissues from the same plants. Together, such a combined analysis of gene isoforms from reproductive tissues, in conjunction with an accurately assembled genome to which isoforms could be mapped, would go hand in hand in spurring a jump in our understanding of what factors control and regulate the switch from sexual to apomictic seed formation. Neither of these project goals has been achievable using current NGS technology, and hence the projects proposed here have been made possible by Third Generation single-molecule sequencing and single molecule genome mapping (Lam *et al.*, 2012).

The case for *Boechera* – encouraging developments

The genus *Boechera* (Böcher's rock cress; formerly *Arabis*; Brassicaceae), is a perennial member of the Brassicaceae that is distributed throughout North America and Greenland (Koch *et al.*, 1999). The genus is monophyletic, has a basic chromosome number x = 7 (Koch *et al.*, 1999), and is characterized by diploid sexual, as well as diploid, aneuploid, and polyploid (mostly 2n=3x=21) apomictic forms. Polyploidy has arisen multiple times in geographically and genetically distinct populations (Sharbel *et al.*, 2005).

Boechera species have also been used for comparative genomic analysis, including partial genome sequencing (Windsor *et al.,* 2006), genetic map construction (Schranz *et al.*, 2007), copy number variation (Aliyu *et al.,* 2013), miRNA analyses (Amiteye *et al.,* 2011, 2013), and transcriptome sequencing (Sharbel *et al.,* 2009; Sharbel *et al.,* 2010), and the entire genomes of *B. stricta* and *B. divaricarpa* are being sequenced (DOE Joint Genome Institute; www.jgi.doe.gov). Importantly, comparative genomics between the model plant, *Arabidopsis thaliana,* and *Boechera,* is facilitated by the close phylogenetic relationship between these taxa and the remarkable amount of genetic resources in the Brassicaceae. Genome sequences are available for *Capsella* spp., *Camelina* spp., and many *Arabidopsis* species. These resources form the foundation to construct ancestral alleles, identify orthologs and conserved nucleotide sites and determine patterns of genome evolution of *Boechera* (e.g. Shranz *et al.,* 2005). One particularly useful example of translational genomics comes from the laboratory of Stephen Wright (Haudry *et al.,* 2013), where conserved coding and non-coding sequences were documented across the Brassicaceae.

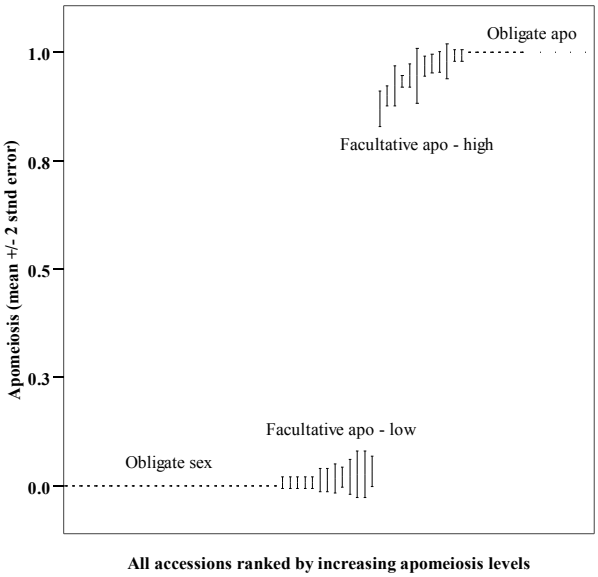
Apomixis and *Boechera*

The genus *Boechera* is an ideal model system to compare sexual and apomictic reproduction and associated life history traits, as it is characterized by naturally occurring diploid sexual and *diploid apomictic* forms (Aliyu *et al.,* 2010). *Thus differences between apomictic and sexual individuals can be compared without the confounding effects of polyploidy or hybridization*.

Quantitative variation for expression of apomixis components

The three components of the apomixis phenotype: apomeiosis, parthenogenesis, and pseudogamy, can be measured using the flow cytometric seed screen (FCSS; Matzk *et al.,* 2000), based on embryo and endosperm ploidy from single seeds. A diploid (2C) embryo and a triploid (3C) endosperm characterize "normal" sexual seed. Alternatively, apomictic seed production will be characterized by deviations from this ratio since the non-reduced (apomeiotically-derived) polar nuclei of the mature gametophyte lead to the formation of endosperm whose ploidy level is determined by autonomous development (4C), or by fertilization with a reduced (C) or unreduced (2C) pollen grain, leading to 5C and 6C endosperm respectively.

We developed a high-throughput flow cytometric seed screen (FCSS) to measure embryo:endosperm ploidy in over 22000 single seeds of 71 genotypes (3 replicates per genotype) of diploid and triploid *Boechera* (**Fig. 1**; Aliyu *et al.*, 2010). Four interrelated features were identified: (1) genotype-specific variation for most traits associated with apomictic seed formation, (2) three levels of apomeiosis penetrance (low, high, obligate) among diploids (**Fig. 1**), (3) genotype-specific correlations between apomeiosis and parthenogenesis/pseudogamy, and (4) most diploid and triploid apomicts are characterized by both meiotically unreduced egg and pollen production (Aliyu *et al.*, 2010). These data are consistent with epistatic influences on the penetrance of individual apomixis components in different genetic backgrounds. Importantly, these data have enabled us to choose highly expressive (obligate) apomictic and sexual genotypes for comparative expression profiling of (1) ovules and (2) anthers (next section).

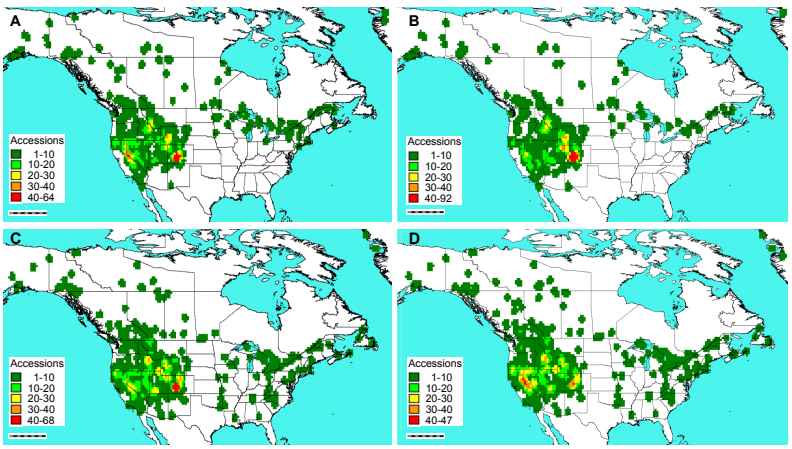


◂**Figure 1.** Ranked mean apomeiosis frequency (± 2 standard error) for 62 diploid and 9 triploid *Boechera* accessions, with three replicates per accession (Taken from Aliyu *et al*., 2010).

A separate candidate gene each for meiotically-unreduced egg and pollen production

Using our high throughput seed screening method, in conjunction with custom-made high-density microarrays for tissue-targeted expression analyses, *we have identified a single factor each for meiotically*

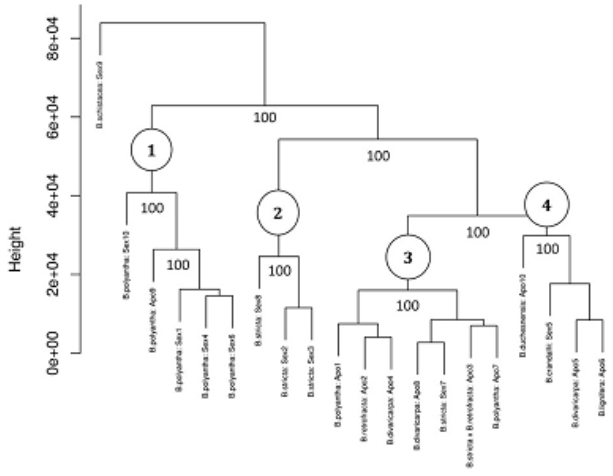
unreduced egg (APOLLO; Corral et al., 2013) and meiotically unreduced pollen (UPGRADE-2; Mau et al., 2013) production in apomictic Boechera. As we have shown that >95% of seeds produced by diploid apomictic *Boechera* are characterized by a diploid embryo (100% maternally-derived) and a hexaploid endosperm (4 maternal genomes and 2 paternal genomes), we hypothesize that both factors are required in order to stabilize apomixis through the combined effects of producing a meiotically-unreduced ovule with balanced (2 maternal : 1 paternal) endosperm formation. In support of this, our continuing work on these candidates has demonstrated that both factors are highly correlated in highly-expressive apomictic *Boechera* from genetically (i.e. different taxa) and geographically (i.e. 1000s of kms) divergent genotypes (**Fig. 2**; Mau *et al.,* 2015).



◂**Figure 2.** Geographic distribution of *Boechera* accessions with and without apomixis marker genes. The PCR-based screen shows similar distributional ranges of *Boechera* accessions with APOLLO (A) and with UPGRADE2 (B) compared to accessions lacking APOLLO or UPGRADE2 (C, D). Bar scale = 1000 kilometers.

Copy number variation demonstrates independent origins of apomictic lineages

Asexual reproduction is expected to lead to mutation accumulation due to the decreased responses to selection that are correlated with an absence of meiosis and sex. To measure this, we have recently compared mutation accumulation in the form of copy number variation (CNV) in the transcribed genomic regions of 10 sexual and 10 apomictic *Boechera* genotypes using a double-validated analysis of comparative genomic hybridization on a high-density (>700 K) custom microarray (Aliyu *et al.*, 2013). Genome-wide patterns of CNV revealed four divergent lineages, three of which contain *both sexual and apomictic* genotypes (**Fig. 3**).



◂**Figure 3.** Hierachical clustering of 10 sexual and 10 apomictic *Boechera* genotypes based upon > 300 000 CNV probe states (i.e. enriched versus depleted). Lineages 1, 3 and 4 (circled numbers) are characterized by both sexual and apomictic genotypes (from Aliyu *et al.,* 2013).

Hence each lineage reflects an independent origin (i.e., expression) of apomixis from a different sexual genetic background. Importantly, multiple origins of the apomixis phenotype are consistent with introgression of our candidate apomixis factors (APOLLO and UPGRADE-2; Corral *et al.*, 2013 and Mau *et al.*, 2013) into different sexual backgrounds via hybridization (Mau *et al.*, 2015).

Multiple origins of the apomictic phenotype occurs via an infectious spread of the APOLLO genomic region

As apomixis is 98% associated with the APOLLO polymorphism, the genotype of this locus offers a high-throughput analytical method to screen for apomixis in *Boechera*. Using this approach, we genotyped and predicted sexuality vs. apomixis in 1502 (an experimental *n* unachievable through traditional flow cytomtetric methods) accessions collected across the phylogenetic and geographic distribution of *Boechera*. APOLLO, UPGRADE2 or both together (**Fig. 2**) are found not only across all major *Boechera* chloroplast haplotypes (from Kiefer *et al.*, 2009), but also across the entire geographic distribution of the genus (Mau *et al.*, 2015).

Importantly, an analysis of 8 *Boechera* species, each of which was characterized by sufficient sexual and apomictic members for statistical comparison, clearly demonstrates (1) sexual species-specific niche patterns, (2) independent origins of apomictic phenotypes sharing the same niche as their sexual backgrounds in 6 of 8 species, and (3) niche "drift" of the apomictic phenotype away from its sexual genetic background (e.g. *B. retrofracta* and *B. stricta*; **Fig. 4**).

Taken together, our data demonstrate (1) sexual speciation and niche differentiation occurs, followed by (2) introgression of the genomic region encompassing APOLLO into different sexual backgrounds.

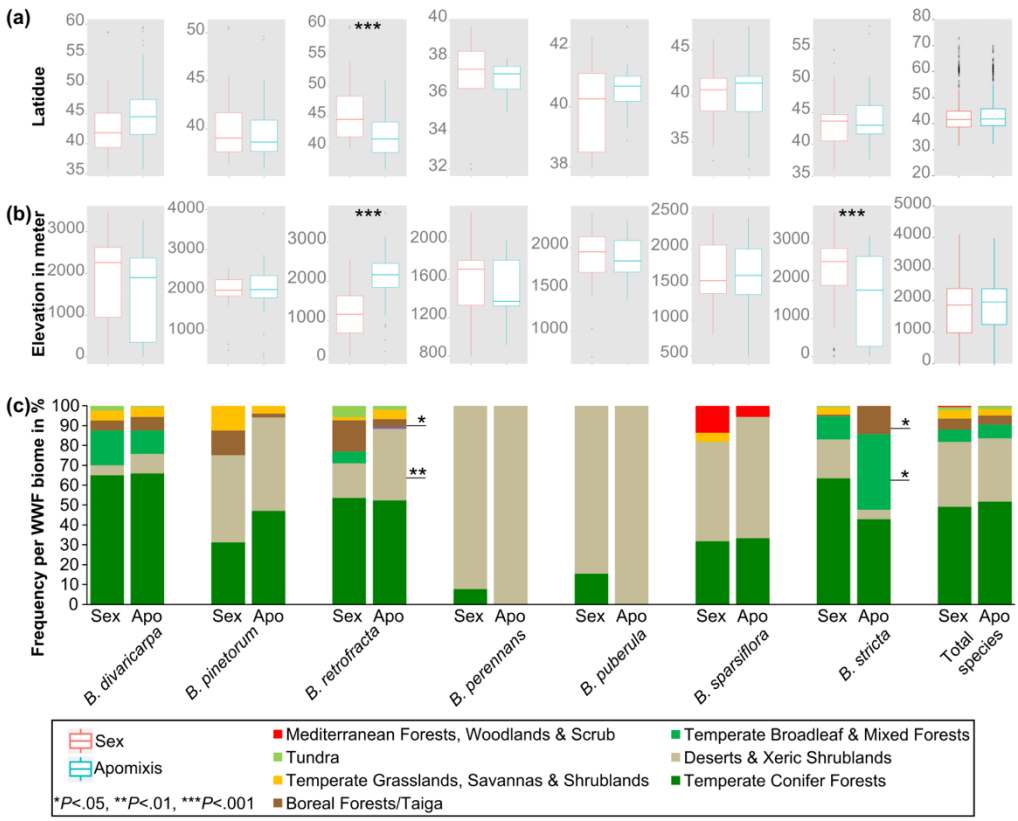


Figure 4. Species-specific variation of niche occupation and niche partitioning between sexual and apomictic *Boechera* in a subset

of species where statistical comparisons could be made (Mau *et al.*, 2015). In some species, apomixis is constrained to a subset of climates (*e.g. B. retrofracta*) while in others apomixis is found across the entire ecological niche (e.g. *B. divaricarpa*). Asterisks denote significant differences between distribution of sexual and apomictic *Boechera* based on two-tailed Fisher's exact tests ($\alpha = 0.05$; *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$).

The importance of this project

Delivering a high-quality genome sequence, including positional accuracy through the non-recombinant (i.e. hemizygous) genomic regions characteristic of asexual species, represents a task of high merit for any single apomictic species. Here we propose the challenging task of delivering a gold standard sexual genome and 6 assembled "silver standard" apomictic genomes in the genus *Boechera*.

Genome sequences, in addition to the fully sequenced isoform spectra of ovules from sexual and apomictic members of *Boechera*, are relatively low risk deliverables that will greatly advance apomixis technology in their own right. The ability to subsequently perform comparative analyses using different levels of biological replication provides a profound source of information that will have far-reaching benefits to apomixis research.

6. Methodology: *Include approaches, experimental design, methodology, materials, sites, etc.*

Hybrid genome assembly and mapping: highly accurate positional genome sequence data

Relative to their sexual relatives, apomicts typically possess complex and large genomes, characterized by an abundance of repeats, transposable elements and aneu- and polyploidy. One challenge with respect specifically to apomixis research is related to the fact that apomixis factors have been repeatedly localized to genomic regions that are characterized by suppressed or absent meiotic recombination, whereby homologous chromosomal regions diverge (in structure and DNA sequence) from one another (e.g. hemizygous regions; Roche *et al.*, 2001). These regions are typically "genomic deserts", enriched for repetitive sequences, heterochromatin and transposable elements, and as such are notoriously difficult to assemble using NGS short-read sequences and standard techniques. Dispersed amongst this complex sequence background are apomixis factors and their regulators (e.g. Ozias-Akins and van Dijk, 2007). As these genomic regions are under decreased natural selection pressure with respect to normal meiotic chromosome pairing, their evolution likely reflects that of non-recombining regions in heteromorphic sex chromosomes (e.g. Y-chromosome; Bachtrog and Charlesworth, 2002). The chromatin modifications associated with chromosomal divergence in hemizygous regions further attest to the potential for these regions to generate novel splice variants (Luco and Misteli, 2011).

To address these challenges during the assembly and analysis of the genomes under investigation, we will use complementary state-of the art NGS and mapping technologies and innovative bioinformatics software tools designed to integrate data from these technologies into accurate genome representations.

For genomic DNA sequencing, we will use complementary technologies combining short and long DNA insert sizes and read lengths. All samples will undergo short-read Illumina DNA sequencing, with short-insert libraries (400bp-800bp – PCR free) sequenced at 60x coverage for both 2x125bp and 2x250bp read lengths, and for long-insert libraries (a.k.a mate-pair, 1kb-20kb insert size) sequenced at 40x coverage for both 2x125bp and 2x250bp read lengths. In addition, 5x coverage of finished-quality long-reads (3kb-20kb) will be added in the form of Illumina synthetic long reads (a.k.a Moleculo). The Gydle bioinformatics software tools will integrate these data into hybrid genome assemblies that will enable us to (1) assemble the genome into long high-quality contigs (N50 ≥ 100 KB; *sensu* Au *et al.*, 2014), (2) assess the genomic abundance of each genome segment, thus resolving repetitive elements and distinguishing allelic variants from separate genomic copies, and (3) call high-quality single nucleotide polymorphisms (*sensu* Pellino *et al.*, 2013).

We will complement the genome sequencing and assembly process with single-molecule optical mapping technology (Lam *et al.*, 2012; Hastie *et al.*, 2013 e.g. Irys genome maps, BioNano Genomics), which provide genome maps with positional information of particular sequence motifs that can be anchored to the

sequence assembly, thus scaffolding the genome assembly over very large regions to entire chromosomes and providing the ordering of genetic factors. Specifically, the mapping of repeating elements will greatly advance assembly and analysis of apomixis-specific non-recombining (e.g. hemizygous; Roche *et al.*, 2001) regions. These technologies and software tools will be employed in a sequential fashion to accurately map both recombining and non-recombining regions in apomicts, whose genomes were previously too complex and/or large to be economically feasible for a sequencing survey.

The genome assemblies generated through this multifaceted approach (**Fig. 5**) will provide scaffolds for deliverables 2 and 3, and will enable the comparison of genomic structural variation (e.g. copy number variation; Aliyu *et al.*, 2013) between sexual and apomictic samples.

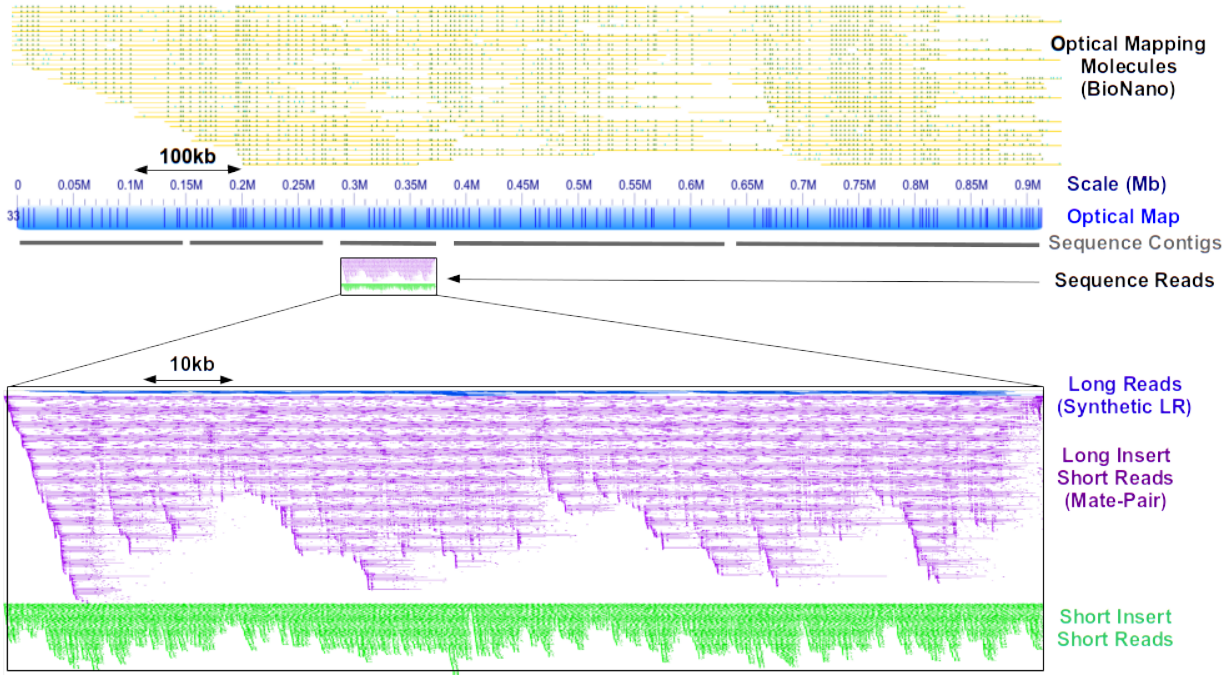


Figure 5. Strategy of sequencing, assembly and optical mapping proposed for each sexual and apomictic pair per species.

7. Research accomplishments: *(Describe progress towards meeting objectives. Please use revised objectives if Ministry-approved revisions have been made to original objectives.)*

Objectives	Progress
Activity 1. High quality sexual and apomictic genome assemblies	**Complete** - A single diploid sexual genotype of *Boechera stricta* (accession ES672) was grown, and 10 high molecular weight and quality DNA samples were collected from leaf tissue. The 10 samples were sent to NRGene (3 Golda Meir St. Ness Ziona 74036 Israel) for whole genome sequencing. From the same genotype, 2 RNA samples from each of the following: Leaf, steam, Flower. Samples were delivered to the Genome sequencing facility at the NRC. Illumina and Kapa library preparation for MiSeq sequencing (2x300) was performed.

Activity 1.1. Library construction, test run (MiSeq), initial quality control	**Complete** - We had initially wanted to outsource sequencing to a Canadian provider (e.g. Genome Quebec), but earlier in 2016 NRGene visited the UofS to enquire about possible future collaborations using their proprietary assembly pipeline, which has been used to generate high quality assemblies of larger, more complex genomes (e.g. lentil, wheat, barley). We have thus outsourced a single sexual Boechera genotype to them for high quality genome assembly, and this will be used as a backbone to which genome sequences from additional sexual and apomictic genotypes will be assembled by Gydle.
Activity 1.2. Illumina sequencing (first individual): short-read, mate-pair data (2x250bp) and SLR data.	**Complete** - Data were generated at the NRC sequencing facility, and data delivered to Gydle.
Activity 1.3. Assembly of Genome Survey Sequence	**Complete** - Initial data from NRGene have been assembled into the 7 chromosomes of a sexual *Boechera*.
Activity 1.4. Optical mapping	**Complete** - BioNano optical data (Irys chip, 735,000 molecules) generated for first (reference) individual. Assembly of optical data produced 155 optical maps covering 200Mb of DNA, providing a high-quality optical backbone for use in Activity 1.5.
Activity 1.5. Integrated genome assembly (first individual).	**Complete** - Integration of genome survey sequence with optical data including iterative read mapping, local sequence reassembly, gap filling and assembly of complete pseudomolecules representing the organelles and 7 nuclear chromosomes of a sexual *Boechera*. Genome in finishing stage (95% covered), annotation produced together with activity 2.3.
Activity 1.6. Illumina sequencing (as in 1.2) for all remaining individuals.	**Complete** - Samples grown, DNA extracted and sent for standard Illumina Sequencing (10X linked-reads sequencing).
Activity 1.7. Integrated genome assembly (as in 1.6) for all remaining individuals.	Bioinformatics in progress
Activity 2.1. Tissue preparation from microdissected ovules	**Complete** - Reproductive tissues microdissected, RNA extracted and sent to Genome Quebec for RNAseq.
Activity 2.2. RNA-Seq sequencing	**Complete** - Completed RNA-Seq for generation of Gene-space assembly: Illumina libraries representing all major tissues of reference individual sequenced on Illumina MiSeq (2x300b) generated 20M paired reads.
Activity 2.3. Gene-space assembly	**Complete** - Creation of a catalogue of dual transcriptome/genome sequences for *Boechera* genes, by combining *de-novo* assembly of RNA-Seq reads and mapping reads and transcript assemblies the genome sequence, resolving complete gene sequences with intron/exon structures. Current assembly explains 95% of RNA-Seq reads. Complete annotation of genome (in finishing stage) shared with activity 1.5.
Activity 2.4. Analysis of isoform spectra and validation experiments	Bioinformatics in progress
Activity 3.1. Alternative splicing spectra	Bioinformatics in progress

Activity3.2. SNP variation in alternative splice variants	Bioinformatics in progress
Activity3.3. Comparisons to alternative splicing patterns in transformed *Boechera*, and ongoing experiments in Arabidopsis, Maize and Brassica	Bioinformatics in progress
add additional lines as required	

8. Discussion: *Provide discussion necessary to the full understanding of the results. Where applicable, results should be discussed in the context of existing knowledge and relevant literature. Detail any major concerns or project setbacks.*

Agricultural improvement relies upon novel varieties and crop traits whose complex underlying genetic networks and interactions cannot be rapidly and efficiently stabilized by state of the art plant breeding methods. F1 hybrid seed production, the backbone of modern agriculture, is a limiting step to innovation considering the time and resources required to develop and maintain inbred lines, and to test and select the best parental combinations. Apomictic plants produce normal seeds containing embryos that grow into plants that are genetic clones of themselves, regardless of their homo- or heterozygous state on a genome-wide scale. Apomixis technology would provide an enormous advance to crop breeding by significantly shortening the production time of novel hybrid varieties, and by enabling the exploitation of traits previously too complex to harness using standard breeding approaches.

A comparative genome-sequencing (short and long read sequencing), single-molecule mapping and isoform analysis project has been completed in the sexual and apomictic members of the genus *Boechera*, a wild relative of *Brassica*. As apomictic genomes are by definition non-recombining, efforts to assemble short read sequences from apomicts are notoriously difficult due to mutation accumulation (*sensu* Y chromosomes in mammals). Furthermore, as apomixis factors are typically mapped in hemizygous (i.e. having no homologue) chromosomal regions, the unprecedented dual use of long-read sequencing, in conjunction with single-molecule mapping, will enable such difficult genomic regions to be assembled.

The genus *Boechera* has a basic chromosome number x = 7 (Koch *et al.*, 1999), and is characterized by diploid sexual, as well as diploid, aneuploid, and polyploid (mostly 2n=3x=21) apomictic forms (Böcher 1951). Accordingly, our polished sexual *Boechera* genome assembly is characterized by 7 nuclear chromosomes and the complete mitochondrial and chloroplast genomes (Table 1).

Table 1. Assembly statistics from polished sexual *Boechera stricta* genome.

Molecule	Size
Chloroplast	155,276 bp
Mitochondria	271,608 bp in two sub genomes (161,796 and 109,812 bp)
Chr1	27.6 Mbp
Chr2	28.3 Mbp
Chr3	32.2 Mbp
Chr4	27.9 Mbp
Chr5	30.5 Mbp
Chr6	27.4 Mbp
Chr7	27.3 Mbp

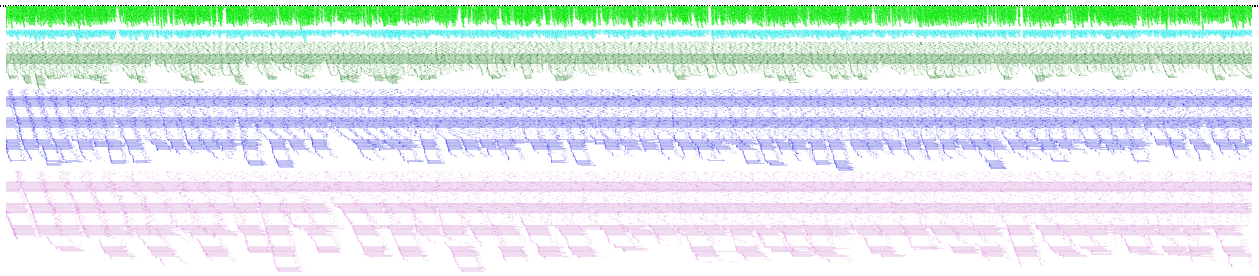


Figure 6. *Boechera*_Assembly_Reads using Gydle software: showing a 500kb segment (chr4 around 10Mb-10.5Mb) of the finished-quality (no gap, no N) assembled sequence, with reads from the 5 libraries aligned to it: light green: 470bp fragment size; light blue: 800bp fragment size; dark green: 2-4kb fragment size (mate-pair); dark blue: 5-7kb fragment size (mate-pair); dark blue: 8-12kb fragment size (mate-pair).

Using the Gydle software, in conjunction with our use of 5 different DNA sequence library sizes, we were able to create a gold-standard genome assembly for the sexual *B. stricta*, having no gaps nor unresolved nucleotide positions (Fig. 6). The different read data were assembled into super contigs, pseudomolecules having a size greater than 100 kb.

Figure 7. Boechera_BioNano_Contig.png using Gydle software: showing that same region from Fig. 2 aligned to a BioNano map, perfectly validating that sequence. Vertical lines indicate aligned restriction sites (imaged by BioNano), colors code for the distances between sites.

The super contigs which were assembled using the 5 different DNA insert libraries were ordered relative to one another using optical mapping on the BioNano system (Fig. 7). Using this approach, we were able to assemble all pseudomolecules into the 7 chromosomes and organellar genomes expected from *B. stricta* (Table 1; Fig. 8).

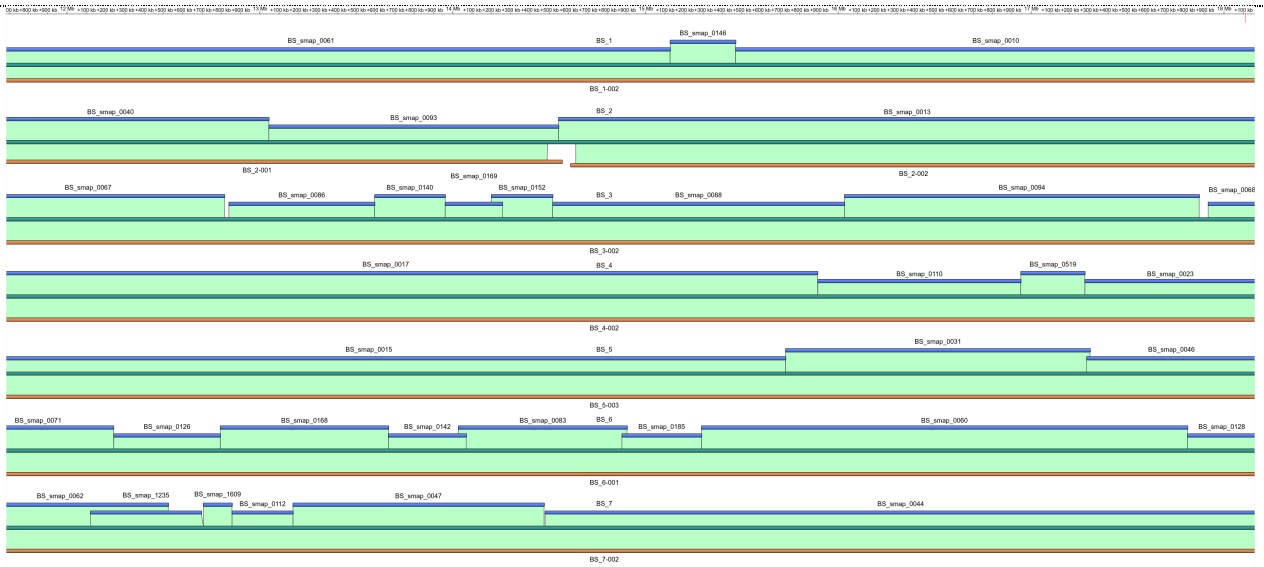


Figure 8. *Boechera* Assembly Overview using Gydle software: showing a ~6Mbp region in the 7 chromosomes (stacked one on top of another) with for each chromosome a ruler in the middle, BioNano maps on top and Contigs on the bottom.

Bioinformatics and statistical analyses in progress

Now that all data have been collected, we are in the process of performing high-powered bioinformatics and statistical analyses of the remaining apomictic genomes, in addition to the sequenced and mapped mRNA isoform spectra isolated from ovules from each individual from each individual. These data will be included in a revised version of this final report in the coming months, as the analyses are completed.

References

Albertini et al. (2010) Apomixis in the era of biotechnology. In: Pua E-C, Davey M R (Eds.): Plant Developmental Biology - Biotechnological Perspectives: Vol. 1. Springer Verlag, Berlin Heidelberg.

Aliyu, O.M., M.E. Schranz and T.F. Sharbel (2010) Quantitative variation for apomixis components in the genus *Boechera*. Am. J. Botany 97(10):1719-1731.

Aliyu et al. (2013) Copy number variation in transcriptionally active regions of sexual and apomictic *Boechera* demonstrates independently-derived apomictic lineages. Plant Cell 25:3808-3823.

Amiteye, S., J.M. Corral, H. Vogel, M. Kuhlmann, M.F. Mette, T.F. Sharbel (2013) Novel microRNAs and microsatellite-like small RNAs in sexual and apomictic *Boechera* species. MicroRNA 2: 46-63.

Amiteye S, Corral J M, Vogel H, Sharbel T F. (2011). Analysis of conserved microRNAs in floral tissues of sexual and apomictic *Boechera* species. BMC Genomics 12 (2011) 500.

Au et al. 2013. Characterization of the human ESC transcriptome by hybrid sequencing. PNAS E4821-E4830.

Ayele M, Dolezel J, Van Duren M, Brunner H, Zapata-Arias FJ (1996) Flow cytometric analysis of the nuclear genome of the Ethiopian cereal Tef [*Eragrostis tef* (Zucc.) Troter]. Genetica 98: 211–215

Bachtrog D, Charlesworth D (2002) Reduced adaptation of a non-recombining neo-Y chromosome. Nature, 416: 323-326.

Beintema, N., Stads, G, Fuglie, K., and Heisey, P. (2012) *ASTI Global Assessment of Agricultural R&D Spending: Developing Countries Accelerate Investment*. Washington: IFPRI.

Buckley YM, Briese DT, Rees M (2003) Demography and management of the invasive plant species *Hypericum perforatum* L. using multi-level mixed-effects models for characterizing growth, survival and fecundity in a long-term data set. Journal Applied Ecology, 40: 481-493.

Carman, J.G. (1997) Asynchronous expression of duplicate genes in angiosperms may cause apomixis,

bispory, tetraspory, and polyembryony. Biol. J. Linn. Soc. 61: 51–94.
Chaisson et al. (2014) Resolving the complexity of the human genome using single-molecule sequencing. Nature doi:10.1038/nature13907
Corral et al. 2013. A conserved apomixis-specific polymorphism is correlated with exclusive exonuclease expression in premeiotic ovules of apomictic *Boechera* species. Plant Physiol. 163:1660-1672.
De Klerk, E., 't Hoen, PA. (2015) Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. Trends Genet. 31(3): 128-139.
Graff. G. D., D. Zilberman and A. B. Bennett. (2009) The contraction of agbiotech product quality innovation. *Nature Biotechnology* 27: 8: 702-704.
Haggui, F., P. Phillips and R. Gray (2006) Opposition to genetically modified wheat and global food security. Chapter 15 in R. Evenson and V. Santaniello (eds), *International Trade and Policies for GM Products. Wallingford:* CABI.
Hajheidari M, Koncz C and Eick D (2013) Emerging roles for RNA polymerase II CTD in *Arabidopsis*. Trends Plant Science 18(11): 633-643.
Hastie et al. (2013) Rapid genome mapping in nanochannel arrays for highly complete and accurate *de novo* sequence assembly of the complex *Aegilops tauschii* Genome. PLoS ONE 8(2): e55864
Haudry A, et al. (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat Genet. 45: 891–898.
Hojsgaard D.H., Klatt S., Baier R., Carman JG., Hörandl E. (2014) Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics. Critical Reviews of Plant Sciences 33: 414-427.
Jefferson RA. (1994) Apomixis: a social revolution for agriculture? Biotechnol. Dev. Monitor 19: 14-16.
Lam et al. (2012) Genome mapping on Nanochannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology 30(8): 771
Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. (2011) Epigenetics in alternative pre-mRNA splicing. Cell. 144(1):16-26. doi:10.1016/j.cell.2010.11.056.
Matzk, F. , A. Meister , and I. Schubert (2000) An efficient screen for reproductive pathways using mature seeds of monocots and dicots. Plant Journal 21: 97–108.
Mau et al. (2013) The conserved chimeric transcript UPGRADE-2 is associated with unreduced pollen formation and is exclusively found in apomictic *Boechera*. Plant Physiol. 163:1640-1659.
Mau et al. (2015) Hybrid apomicts trapped in the ecological niches of their sexual ancestors. PNAS (in press).
Pellino et al. (2013) Asexual genome evolution in the apomictic *Ranunculus auricomus* complex: examining the effects of hybridization and mutation accumulation. Mol. Ecol. 22:5908-5921
Perri. (2001) 'Governing by technique: judgement and prospects for the governance of and with technology', in OECD, *Governance in the 21St Century*, Paris: OECD, accessed on August 20, 2005 at : http://www.oecd.org/dataoecd/15/0/17394484.pdf, pp. 67-120.
Poverene MM, Voigt PW (1997) Isozyme variation and germplasm relationships in the *Eragrostis curvula* Complex. Biochem Syst Ecol 25:21-32.
Roche et al. (2001) Gametophytic apomixis, polyploidy, and supernumerary chromatin. Sex. Plant Reprod. 13:343-349.
Schranz, M.E., A.J. Windsor, B.H. Song, A. Lawton-Rauh, and T. Mitchell-Olds (2007) Comparative genetic mapping in *Boechera stricta*, a close relative of *Arabidopsis*. Plant Physiol. 144: 286-298.
Sharbel, T. F. T. Mitchell-Olds, C. Dobes, L. Kantama and H. de Jong (2005) Biogeographic distribution of polyploidy and B chromosomes in the apomictic *Boechera holboellii* complex. Cytogenetic and Genome Research 109: 283-292.
Sharbel, T. F. M.L. Voigt, J.M. Corral, G. Galla, J. Kumlehn, C. Klukas, F. Schreiber, H. Vogel and B. Rotter (2010) Apomictic and sexual ovules of *Boechera* display heterochronic global gene expression

patterns. Plant Cell 10.1105/tpc.109.072223.
Sharbel, T. F. M.L. Voigt, J.M. Corral, T. Thiel, A. Varshney, J. Kumlehn, H. Vogel and B. Rotter (2009) Molecular signatures of apomictic and sexual ovules. Plant Journal 58(5): 870-882.
Spillane et al. (2004) Apomixis technology development—virgin births in farmers' fields? Nature Biotechnology, 22:687 – 691.
Toennissen GH (2001) Feeding the world in the 21st century: plant breeding, biotechnology and the potential role of apomixis. In Y Savidan, JG Carman, T Dresselhaus, eds, The Flowering of Apomixis: From Mechanisms to Genetic Engineering. CIMMYT, IRD, European Commission DG VI (FAIR), Mexico D.F., Mexico, pp 1-7.
Treutlein et al. (2014) Cartography of Neurexin Alternative Splicing Mapped by Single-Molecule Long-Read mRNA Sequencing. *PNAS* 111(13): E1291–99. doi:10.1073/pnas.1403244111.
Windsor, A.J., M.E. Schranz, N. Formanova, S. Gebauer-Jung, J.G. Bishop, D. Schnabelrauch, J. Kroymann and T. Mitchell-Olds (2006) Partial shotgun sequencing of the *Boechera stricta* genome reveals extensive microsynteny and promoter conservation with *Arabidopsis*. *Plant Physiol* 140:1169-1182.

9. Conclusions and Recommendations: *Highlight significant conclusions based on the previous sections, with emphasis on the project objectives specified above. Provide recommendations for the application and adoption of the project.*

Apomixis technology is a game changer for industry, farmers and consumers, and as such presents a plethora of social and economic benefits, together with many challenges for intellectual property sharing, as farmers will become innovators for niche-adapted varieties that could subsequently be bred apomictically by industry. The introduction of apomixis into agriculture represents a disruptive enabling/transformative technology that would benefit both breeders and users by significantly simplifying the fixation and preservation of genetic heterozygosity and its associated hybrid vigor in crop plants, for example by enabling single generation hybrid seed production without the need for inbred lines. This new technique will open the opportunity for more customization of target traits in local varieties. In this sense, apomixis could potentially change the distribution and scale of investment in new varieties, disrupt existing commercial supply chains and lead to greater uptake and use of advanced agronomic and nutritional traits.

Here we have completed a gold-standard assembled genome of a sexual *Boechera stricta*, and are in the process of finalizing the assemblies of a further 3 apomictic genomes from the same genus, using the sexual genome as a reference. Importantly, the ongoing statistical and bioinformatics analyses of these assembled genomes will enable us to (1) identify DNA polymorphisms that that are specific to either sexual or apomictic reproduction, and (2) elucidate conserved structural variation in the apomictic genomes which have arisen from the disturbed meiosis which characterizes asexual reproduction. Together with the mRNA isoforms analyses (underway) from the microdissected ovules of sexual and apomictic individuals of *Boechera*, we will be able to integrate and interpret a number of ongoing projects in the Sharbel lab which compare the proteomics, degradomes (i.e. post transcriptional products of mRNA degeneration) and genome-wide expression analyes of sexual and apomictic reproductive tissues.

The importance of the data generated here cannot be overstated. Considering the number of ongoing projects in the Sharbel lab which center around apomixis in the model *Boechera* system, this project has succeeded in placing the apomixis research done at GIFS in a world-leading position, with multiple planned high-impact publications to arise from the work.

10. Success stories/ practical implications for producers or industry: *Identify new innovations and /or technologies developed through this project; and elaborate on how they might impact the producers /industry.*

In the coming months, the completed data generated here will be integrated with the various projects underway in the Sharbel lab to accelerate our progress in transferring apomixis to crops. With respect to the impacts on industry, apomixis would have significant influences on a number of levels.

The world market for seeds is currently (2014) US$45 billion. This value has almost doubled since 2000 primarily because of the use of molecular biology and biotechnology in adding value for farmers through traits like herbicide tolerance and insect- and viral-resistance (see Table 2). More than 75% of the value of seeds is 'proprietary seed', which includes F1 hybrids or traits protected by patent, which command a premium or a technology access fee. Often it is a combination of both. The days of small, local seed companies have been superseded by the emergence of seed industry giants such as Pioneer Hi Bred (Du Pont) and DeKalb (Monsanto). In fact, about 50% of the total seed market worldwide is in the hands of 4 companies. However, innovation and value-creation in these markets still comes from smaller companies (and universities) who provide the discovery science and intellectual property needed to develop a trait to commercial reality. The disparity between proprietary and non-proprietary seed is dramatic. In the US, two-thirds of the total seed market is accounted for by corn and soybeans, while wheat (non-proprietary) accounts for only 5% of the market (~US400M per annum).

Year	2000	2005	2010
World (US$ B)	24.4	34	42
US (US$ B)	5.05	7.08	8.7
US % of seed market	20.7	20.8	20.7

Table 2: Development of global and US seed markets since 2000

It is clear therefore that the business model that drives value in the seed sales requires a proprietary position, with either repeat business (as with F1 Hybrids) or a technology access fee paid *in lieu* of a licence fee for crops.

The introduction of a new technology such as **apomixis** presents an interesting challenge to the current global seed business for two reasons.

Firstly, the F1 hybrid seed business could be radically disrupted by the introduction of apomixis. If an F1 hybrid, were induced to reproduce by apomixis after generating germplasm with substantial heterosis, the heterotic yield gain would be maintained essentially through clonal seed propagation. This would mean that farmers could save seed and still enjoy the benefits of F1 hybrids. In developed countries where patents are properly enforced, value could still be returned to the seed company, but it would have to be via mechanisms such as technology access fees. This would not inhibit further plant breeding to produce newer and better varieties, but these might be introduced less frequently depending on need for new traits such as drought tolerance, which will become more prevalent due to climate change. In some developing countries, where less heed is paid to patents or plant variety protection, the use of apomixis to "fix' heterosis in asexually propagated germplasm would result in higher yield potential for farmers. Whether that yield potential is realised would depend on other inputs like fertilizer, but the potential yield increases might make purchasing of fertilizer a more economic proposition.

The second disruptive consequence of apomixis being used in major crop species that are currently not 'proprietary' such as wheat would be the ability to deliver hybrid vigour to traditionally inbred crops due to their cleistogamous flower structure. These include large acreage crops such as wheat, soybeans, barley, lentils and other pulses. In all of these cases, it has been shown that crosses between diverse germplasm in each of these species will result in significant hybrid vigour. However, normally, the cost of seed production

with these cleistogamous species is prohibitive as manual emasculation and hand pollination is excessive due to labour costs. However, if it is possible to create, by manual cross pollination, an F1 hybrid that could then be maintained through asexual reproduction (apomixis), then the result would offer much higher yield potential than conventional self-pollinated inbreds which are typical of the above-mentioned crops. A profitable business model would still require a technology access fee or a well-enforced royalty scheme, but existing precedents suggest that this could be developed. If this were applied to a large acreage crop such as wheat, the value of just the North American wheat seed market could jump from $400M to over $2 billion if a technology access fee were levied.

In addition to these disruptions to markets and business models, apomixis would also result in dramatically reduced costs for any kind of hybrid production. This is because the desired F1 hybrid would only be made once rather than every year. Hybrid production, although routine either through emasculation or by cytoplasmic male sterility, is still a significant element to cost-of-goods in a hybrid seed business as is quality control/assurance. With these costs greatly reduced, the overall profitability of a seed business would be improved. Similarly, distribution costs would be substantially reduced. As most of the 'inventory' would now be on farm, the overall costs of maintaining inventory would also be reduced.

The changes in cost structure that would be introduced by the exploitation of apomixis, could have one additional consequence: enhanced competition. Currently, as described above, the world seed business is dominated by four major players: DuPont, Monsanto, Syngenta and Limagrain. Collectively, they dominate 49% of the global seed market. However, with changes in the economics of hybrid production, the smaller players might have greater incentives to develop apomictic hybrids due to the reduced importance of economies of scale, particularly in seed production and distribution. The better margins and reduction in capital costs (working- and infrastructure) may be tempting for the next tier of seed company such as Land o' Lakes, KWS, Bayer and Dow Agro to develop their hybrid businesses.

The benefits of apomixis for Canola, wheat and maize

Using **Canola** as a specific example, apomixis could *fundamentally* impact the conventional system of hybrid breeding and production in three ways:

(a) Canola F1 seed production currently involves seed fields composed of males and male sterile female lines. The female parents must be bred to contain a male sterility trait, a process that requires additional breeding effort and perhaps additional time. The hybrid seed is produced by planting rows of fertile male plants in between male sterile female rows. Bees must be brought to the field to pollinate the male sterile female plants. This adds cost and risk to production costs. Apomixis would enable seed production to be made directly from F1 hybrids, thereby eliminating these costs and extra steps.

(b) The introduction of apomictic F1 seed production would enable the full seed volume to be attained in 1 to 2 years, thereby enabling more rapid genetic/yield gains at the farm level, and providing more dependable supplies to market (i.e. food security).

(c) From a breeding perspective, apomixis would provide the greatest and most immediate benefit considering that F1 hybrids would be available in their first generation for testing. Presently, test crosses are made from F2 or F3 segregating lines, and variation in the resultant hybrids is used to select lines to advance for further testing in the F6 or F7 generation (i.e. relatively genetically fixed). As these lines are nonetheless segregating, the hybrid genetics between the tested lines and the ultimately fixed hybrids that a farmer would grow are not the same. This breeding scheme also means that potentially good hybrids are discarded while poor ones are advanced. Apomixis would enable hybrids to be tested in the first generation, thus providing more rapid and accurate data to breeders for making advancement decisions.

Using **wheat** as a specific example, apomixis could *fundamentally* impact the conventional system of hybrid breeding and production in two ways:

(a) Wheat is currently grown primarily as a self-pollinated crop. This means it is true breeding and its seeds will produce a plant identical to the mother plant. Therefore apomixis is not needed to 'fix' the genotype. It is already fixed because of its inbred nature. However, where apomixis could be used to improve wheat is by making it possible to affordably make hybrid wheat varieties. To make hybrid wheat currently, the female flowers must be hand pollinated through removal of the anthers (male sex organ) and then bringing pollen from a different male plant to this emasculated female flower. There are chemical means now to induce sterility. Then rows of female plants would be grown in fields interspersed with rows of male plants to cross pollinate the females and produce hybrids seed. But to date, almost all wheat grown in the world remains self-pollinated varieties. With apomixis, it would be possible to produce hybrid wheat seed rapidly and cheaply. Once a hybrid plant is made, by using apomixis, it would continue to self-pollinate but now produce hybrid seed instead of inbred seed.

(b) The introduction of apomictic F1 seed production would enable the widescale production of inexpensive hybrid wheat seed and could leverage hybrid vigor, something not currently possible in self-pollinated crops today. This should enable greater yields and genetic/yield gains at the farm level, and providing more dependable supplies to market (i.e. food security). The same would be true for any other self-pollinated crop such as soybean.

Using **maize** as a specific example, apomixis could *fundamentally* impact the conventional system of hybrid breeding and production in 4 ways:

(a) Maize F1 seed production currently involves seed fields composed of males and detasseled female inbred lines, both of which have lower yield and are vulnerable to crop failure due to susceptibility to environmental stress compared to F1 hybrids. Detasseling furthermore contributes significantly to production costs. Apomixis would enable seed production to be made directly from F1 hybrids, thereby eliminating these costs and unpredictability.

(b) Some significantly improved crosses never make it to market, the result of asynchrony in maturity/flowering times between the parental lines (i.e. nicking). The introduction of apomixis would mean that genetic fixation of a cross would only have to be successful one time, followed by apomictic seed increase of that cross. Hence more potential hybrids would be available to market.

(c) The typical "life cycle" of commercial maize hybrids is presently 3 to 5 years, 2 to 3 years of which are used to multiply and inbreed seed stock. The introduction of apomictic F1 seed production would enable the full seed volume to be attained in 1 to 2 years, thereby enabling more rapid genetic/yield gains at the farm level, and providing more dependable supplies to market (i.e. food security).

(d) From a breeding perspective, apomixis would provide the greatest and most immediate benefit considering that F1 hybrids would be available in their first generation for testing. Presently, test-crosses are made from F2 or F3 segregating lines, and variation in the resultant hybrids is used to select lines to advance for further testing in the F6 or F7 generation (i.e. relatively genetically fixed). As these lines are nonetheless segregating, the hybrid genetics between the tested lines and the ultimately fixed hybrids that a farmer would grow are not the same. This breeding scheme also means that potentially good hybrids are discarded

while poor ones are advanced. Apomixis would enable hybrids to be tested in the first generation, thus providing more rapid and accurate data to breeders for making advancement decisions.

11. Patents/ IP generated/ commercialized products: *List any products developed from this research.*

None so far

12. List technology transfer activities: *Include presentations to conferences, producer groups or articles published in science journals or other magazines.*

Multiple instances expected in 2018

13. List any industry contributions or support received.

The work here was performed in the GIFS labs at the University of Saskatchewan, and the infrastructure and staff of the Sharbel lab provided support to successfully complete the work.

14. Is there a need to conduct follow up research? *Detail any further research, development and/or communication needs arising from this project.*

The data generated here, in combination with other ongoing apomixis projects funded by GIFS and third parties, will lead to future projects of validation and functional genetics as we move more closely to crop plants. For the moment these goals are undefined as we complete the statistical analyes of the data generated here.

15. Acknowledgements. *Include actions taken to acknowledge support by the Ministry of Agriculture and the Canada-Saskatchewan Growing Forward 2 bilateral agreement.*

None so far, although ADF will clearly be acknowledged on any publications, IP and presentations which describe this work.

16. Appendices: *Include any additional materials supporting the previous sections, e.g. detailed data tables, maps, graphs, specifications, literature cited*

None so far.