Potato Genomics Three Ways: Quantification of Endoreduplication in Tubers, a Romp Through the Transposon Terrain, and Elucidation of Flower Color Regulation

Francis Parker Effingham Laimbeer

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Horticulture

Richard E. Veilleux, Chair

Aureliano Bombarely Gomez

James G. Tokuhisa

M.A. Saghai Maroof

6.8.18

Blacksburg, VA

Keywords: *Solanum tuberosum*, endoreduplication, flow cytometry, ploidy, transposons, MITEs, anthocyanins

ABSTRACT

Investigations of potato (*Solanum tuberosum)* have been hampered by its complicated genetics and high genetic load. This dissertation applies genome reduction techniques to investigate a broad swath of genomic and physiological phenomena. It begins with the presentation and evaluation of a protocol to characterize endoreduplication within potato tubers, demonstrating substantial variation between tissue types and among wild species which may facilitate research into the genesis and growth of these starchy underground stems. Next, we transitioned to explore the distribution and consequences of a specific class of transposable element, Miniature Inverted Transposable Elements (MITEs), showing that they comprise approximately 5% of the potato genome, occur more frequently in genes with stress-related functions, and may be associated with changes, especially decreases, in gene expression. We then combined homology and sparsity based approaches to predict recent MITE activity, identifying five families as especially active. Finally, we expose the gene underlying the potato flower color locus, a homolog of *AN2*, while showing the effects it exerts on the flavonoid biosynthesis and fruit ripening pathways. This region was shown to be particularly dynamic, replete with MITEs and structural variants which we hypothesize to be the ultimate cause of differences in *AN2* expression within the germplasm we examined. While the separate topics of this dissertation are quite disparate, each addresses an important topic in potato genetics, the in-depth study of which is only possible through the utilization of genomic reduction approaches to acquire homozygous genotypes for study and currently available genomic resources.

ABSTRACT (Public)

Despite their humble appearance and routine consumption, potatoes have a complex genetic structure and a life cycle capable of both sexual reproduction through flowers, fruit and seed, and asexual reproduction through the tubers which also comprise the edible product. From an agronomic perspective, one of the most important qualities of a potato tuber is size, a feature influenced by genetics and environment. Cell-to-cell variation for the amount of DNA per cell, one component that influences tuber size, is known to occur, yet our ability to measure DNA content in starchy tuber cells has been obscured by debris generated through routine preparation techniques. We present and evaluate a new method for measuring the DNA content of potato tuber cells, which provides reliable results across a range of different potato varieties and species. 'Jumping genes' also known as transposons, first reported in maize but now known to occur in most advanced plant and animal species, have been found to comprise ~5% of the recently sequenced potato genome. We show that a particular class of transposons is more likely to occur adjacent or actually in certain types of genes, such as those which confer resistance to disease, where they may have meaningful effects on how those genes operate. We then proceed to predict the current activity of the various families of these jumping genes to understand how they continue to alter the genetic landscape of potato. Finally we identify a particular gene which dictates flower color in potato (purple vs. white). We demonstrate that several transposons occur in some forms of the flower color gene. Originally we hypothesized that transposons were associated with the turning off of the purple flower color form; however, on closer examination, we could express the white flower form in transgenic plants that were originally white-flowered and convert them to have purple flowers, demonstrating that even the white flower form was functional. While the separate topics of this dissertation are quite disparate, each addresses an important topic in potato genetics, the in-depth study of which is only possible through the availability of the special strains of potatoes with reduced chromosome number and the publication of the potato genome.

Acknowledgements

First and foremost, I would like to thank my mentor, advisor, and friend, Dr. Richard Veilleux for his kindness, patience, and constant encouragement. Richard is a seemingly unending fount of positivity which has made working with, and learning from, him as rewarding an experience as I have ever had. I am truly grateful for the opportunity to be his 41st and final graduate student and cannot emphasize enough the positive impact his mentorship has had on me over the last five years.

I would like to express my love and gratitude to my parents and sister: Rick, Alice, and Margot. They have been consistent sources of encouragement, love, and stability even in the face of sacrificed family time and my own self-doubt. I have benefitted immensely from all their hard work; allowing me to pursue my education with a blanket of love and security which too few are afforded.

Thanks to my wonderful collaborators and former lab mates who have been an absolute pleasure to work with and taught me so much, especially Sarah Holt, Hua Xiao, Nan Lu, Pris Sears, Kendal Upham, Norma Manrique-Carpintero, Michael Hardigan, John Hamilton, Gina Pham, Joe Coombs, Dave Douches, Aureliano Bombarely, and Robin Buell. I also owe a debt of gratitude to my graduate student peers, Colin Davis, John Herlihy, and especially Stephen Rigoulot for their support and distractions they provided, both good and bad.

Thanks to everyone in the Horticulture department and Translational Plant Sciences program for their constant support, collaborative atmosphere, and intellectual stimulation.

I also want to thank my college and highschool friends, Andy Osheroff, Will Hubbard, Chuck Vogel, Harwood Hoskins, Andrew Kohler, Big Guy, Sean Page, John and James Minshall, and Jared Balavender, who have been an inexhaustible resource of potato jokes over the past five years.

Last, and certainly not least, I thank my wonderful girlfriend Margeaux Malone for her constant support and putting up with my longwinded explanations about why potato genetics are fascinating.

I dedicate this dissertation to my parents, Richard Howard Laimbeer and Alice Cutting Laimbeer. Thank you and I love you more!

Table of Contents

Table of Figures

Table of Tables

Introduction: Ploidy and its relevance to improvement and understanding of potato (*Solanum tuberosum*)

Potato, the ubiquitous tuber crop whose association with the mundane stands in contrast to a rich history and complex genetic profile, was domesticated in the South American Andes circa 8000 BC (Brush et al. 1981). Subsequently, its cultivation spread throughout South America, accruing mutations and introgressions adaptive for growth outside of the Peruvian highlands. It was first introduced to Europe in the late 16th century by the Spanish conquistadors who had recently conquered Peru and promptly spread throughout Europe in the next few decades, mostly as a curiosity rather than a serious food source. However, frequent crop failures and famines in 18th century Europe opened the door for potato's acceptance as a reliable food source (Salaman and Burton 1985; Hawkes 1992). Today, potatoes are a worldwide staple crop, considered the most leading vegetable crop by the USDA (https://www.ers.usda.gov/topics/crops/vegetables-pulses/potatoes/), grown in over 100 countries, and are especially important in developing nations (https://cipotato.org/crops/potato/potato-facts-and-figures/). Such widespread cultivation is likely due to the potato's unique combination of nutrition, versatility, and growth efficiency. Potato ranks first among major crops in both land use and water use efficiency in terms of calories per input unit (Renault and Wallender 2000). Potato also boasts a high nutritional value, with a single potato providing half an adult's recommended daily allowance of vitamin C as well as considerable amounts of B vitamins, balanced protein, and micronutrients (Kaldy 1972; Kolasa 1993). While U.S. corn yields have rose by roughly 500% since the 1940s as a result of genetic improvement, a shift toward hybrid breeding, and modern cultivation practices, potato yields have increased by over 600% over the same timeframe (Tracy et al. 2004). However,

whereas crop improvement has played a substantial role in the gains of corn, almost all of the realized improvement in potato yields is due to management practices, with genetic gains being rather trivial (Douches et al. 1996).The lack of meaningful genetic gains in potato is perhaps best underscored by the lifespan of leading 'modern' cultivars. The widest grown variety in the UK, cv. Maris Piper, was released in 1966 (Bradshaw and Ramsay 2005) while the leading US cultivar cv. Russet Burbank was not even deliberately bred but rather a spontaneous mutant of cv. Burbank and was first release in 1902 under the name Netted Gem (Bethke et al. 2014). The relatively modest genetic improvements afforded to potato over the past few decades may be in part attributed to barriers presented by its genetic and reproductive circumstances.

As most commercially cultivated potato is autotetraploid (2n=4x=48) and asexually reproduced, it has a number of characteristics which make it particularly difficult to improve through traditional breeding efforts. An extremely heterozygous background combined with a high genetic load results in severe inbreeding depression while tetravalent pairing drastically increase the complexity of segregation and gives rise to particularly esoteric challenges to strategic breeding such as double reduction. While tetraploidy has often stymied breeder's attempts to improve potato through conventional breeding techniques, the basic phenomenon of "ploidy", both on a cellular and organismal scale, may be investigated and manipulated to enhance researcher's understanding of potato physiology and genomics. The practice of haploid extraction, where diploid progeny are obtained from tetraploids or monoploid offspring from diploids (either through anther culture or prickle pollination), may greatly reduce the genomic complexity thereby allowing for isolation of genetic variables for study or breeding (Karp et al. 1984; Uijtewaal et al. 1987; Kotch et al. 1992; Sopory and Munshi 1996; Hardigan et al. 2016b). It was just such a manipulation of ploidy that allowed for the sequencing of the potato genome

itself; an anther culture derived doubled monoploid, DM, was selected after heterozygosity of the initially selected diploid germplasm precluded genome assembly (PGSC 2011).

While a single organism or species is typically thought of a given ploidy, be it diploid, tetraploid, hexaploid or even pentaploid, this captures just a fraction of the inherent complexity. In reality, individual organisms may vary in ploidy from cell to cell; each somatic cell containing a multiple of the base ploidy of the organism. This phenomenon occurs through a process known as endoreduplication, where DNA replication occurs in the absence of cell division. The resulting increase in nuclear DNA content is purported to accommodate a concomitant increase in total cellular volume (known as the karyoplasmic theory) which may be required for certain structural or nutritive functions (Wilson 1925a; Chevalier et al. 2014). Common examples of endoreduplicated tissues include leaf trichomes, which require large cell sizes to prevent herbivory and secrete protective compounds, as well as seed endosperm, wherein a greater cell volume allows for more nutrients to be stored and provided to the developing embryo (Grafi and Larkins 1995; Schweizer et al. 1995; Szymanski and Marks 1998; Walker et al. 2000). Furthermore, genetic studies have demonstrated that manipulation of endoreduplication levels can alter organ size, potentially providing a direct means of increasing yields (Cheniclet et al. 2005; Bourdon et al. 2010; Chevalier et al. 2014).

Recently, there has been a renewed interest in pivoting potato breeding efforts to a traditional hybrid breeding scheme through the use of diploid germplasm and true potato seed (Tiwari et al. 2017; Jansky et al. 2016). If this vision can be realized, plant breeders would have the tools to exert a great deal more control over their germplasm, rapidly increasing breeding efficiency and allowing the scientific gains of basic research to be more readily translated into crop improvement. It is with that hope that this research, which spans potato physiology,

genetics, and genomics, was undertaken. Specifically the goals of this study were as follows: To develop and implement a protocol for evaluation of endoreduplication in potato tubers (Chapters 1 & 2); to combine ploidy manipulation and computational biology approaches to evaluate the potato transposon landscape and its influence on gene expression (Chapter 3); and finally to examine the regulation of anthocyanin production in potato corollas as an example of locus instability and a model for anthocyanin production in the organism as a whole (Chapter 4). While the chapters together are an eclectic collection, each leverages our understanding of ploidy, as either a potential mechanism to increase tuber size or a means to reduce complexity, with the ultimate goal of providing a meaningful contribution to the understanding of possibilities for potato improvement.

Works Cited

Bethke PC, Nassar AM, Kubow S, Leclerc YN, Li X-Q, Haroon M, Molen T, Bamberg J, Martin M, Donnelly DJ (2014) History and origin of Russet Burbank (Netted Gem) a sport of Burbank. Amer J Potato Res 91 (6):594-609

Bourdon M, Frangne N, Mathieu-Rivet E, Nafati M, Cheniclet C, Renaudin J-P, Chevalier C (2010) Endoreduplication and growth of fleshy fruits. In: Lüttge U, Beyschlag W, Büdel B, Francis D (eds) Progress in Botany 71. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 101-132. doi:10.1007/978-3-642-02167-1_4

Bradshaw JE, Ramsay G (2005) Utilisation of the Commonwealth Potato Collection in potato breeding. Euphytica 146 (1-2):9-19

Brush SB, Carney HJ, Huam, xe, n Z, xf, simo (1981) Dynamics of andean potato agriculture. Econ Bot 35 (1):70-88

Cheniclet C, Rong WY, Causse M, Frangne N, Bolling L, Carde J-P, Renaudin J-P (2005) Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. Plant Physiol 139 (4):1984-1994. doi:10.1104/pp.105.068767

Chevalier C, Bourdon M, Pirrello J, Cheniclet C, Gévaudant F, Frangne N (2014) Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. J Exp Bot 65 (10):2731-2746. doi:10.1093/jxb/ert366

Douches D, Maas D, Jastrzebski K, Chase R (1996) Assessment of potato breeding progress in the USA over the last century. Crop Sci 36 (6):1544-1552

Grafi G, Larkins BA (1995) Endoreduplication in maize endosperm: involvement of M phase--promoting factor inhibition and induction of S phase--related kinases. Science 269 (5228):1262-1264. doi:10.1126/science.269.5228.1262

Hardigan MA, Crisovan E, Hamilton JP, Kim J, Laimbeer P, Leisner CP, Manrique-Carpintero NC, Newton L, Pham GM, Vaillancourt B (2016) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated Solanum tuberosum. Plant Cell 28 (2):388-405

Hawkes JG (1992) History of the potato. In: The potato crop. Springer, pp 1-12

Jansky SH, Charkowski AO, Douches DS, Gusmini G, Richael C, Bethke PC, Spooner DM, Novy RG, De Jong H, De Jong WS (2016) Reinventing potato as a diploid inbred line–based crop. Crop Sci 56 (4):1412-1422

Kaldy M (1972) Protein yield of various crops as related to protein value. Econ Bot 26 (2):142-144

Karp A, Risiott R, Jones MGK, Bright SWJ (1984) Chromosome doubling in monohaploid and dihaploid potatoes by regeneration from cultured leaf explants. Plant Cell Tiss Org Cult 3 (4):363-373. doi:10.1007/bf00043089

Kolasa KM (1993) The potato and human nutrition. Am Potato J 70 (5):375-384

Kotch GP, Ortiz R, Peloquin S (1992) Genetic analysis by use of potato haploid populations. Genome 35 (1):103-108

PGSC PGSC (2011) Genome sequence and analysis of the tuber crop potato. Nature 475 (7355):189-195

Renault D, Wallender W (2000) Nutritional water productivity and diets. Agric Water Manage 45 (3):275-296

Salaman RN, Burton WG (1985) The history and social influence of the potato. Cambridge University Press,

Schweizer L, Yerk-Davis G, Phillips R, Srienc F, Jones R (1995) Dynamics of maize endosperm development and DNA endoreduplication. Proc Natl Acad Sci 92 (15):7070-7074

Sopory SK, Munshi M (1996) Anther culture. In: In vitro haploid production in higher plants. Springer, pp 145-176

Szymanski DB, Marks MD (1998) *GLABROUS1* overexpression and *TRIPTYCHON* alter the cell cycle and trichome cell fate in Arabidopsis. Plant Cell 10 (12):2047-2062

Tiwari JK, Luthra SK, Kumar V, Bhardwaj V, Singh R, Sridhar J, Zinta R, Kumar S (2017) Genomics in True Potato Seed (TPS) Technology: Engineering Cloning Through Seeds. In: The Potato Genome. Springer, pp 297-305

Tracy WF, Goldman IL, Tiefenthaler AE, Schaber MA (2004) Trends in Productivity of US Crops and Long-term Selection. Plant Breeding Reviews: Long-term Selection: Crops, Animals, and Bacteria, Volume 24, Part 2:89-108

Uijtewaal BA, Huigen DJ, Hermsen JG (1987) Production of potato monohaploids (2n=x=12) through prickle pollination. Theor Appl Genet 73 (5):751-758. doi:10.1007/bf00260786

Walker JD, Oppenheimer DG, Concienne J, Larkin JC (2000) *SIAMESE*, a gene controlling the endoreduplication cell cycle in *Arabidopsis thaliana* trichomes. Development 127 (18):3931-3940

Wilson E (1925) The karyoplasmic ratio. The Macmillan Company, New York

Chapter 1: Protoplast isolation prior to flow cytometry reveals clear patterns of endoreduplication in potato tubers, related species, and some starchy root crops

F. Parker E. Laimbeer[1], Sarah H. Holt[1], Melissa Makris[2], Michael Alan Hardigan[3], C. Robin Buell[3], and Richard E. Veilleux[1]

[1] Department of Horticulture, Virginia Tech, Blacksburg, VA 24061, USA

Parker Laimbeer: plaimbeer@gmail.com (corresponding author)

Sarah Holt: sarahhudson06@gmail.com

Richard Veilleux: potato@vt.edu

[2] Department of Biomedical Sciences and Pathobiology, Center for Molecular Medicine and Infectious Diseases, Virginia-Maryland Regional College of Veterinary Medicine, Virginia Tech, Blacksburg, VA 24061, USA

Mellissa Makris: mmakris@vt.edu

[3] Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

Michael A Hardigan: hardiga3@msu.edu

C. Robin Buell: buell@msu.edu

Author's Contributions

FPEL conducted the research, analyzed the data and drafted the manuscript. FPEL, SHH, and REV developed the protocol. MM performed the flow cytometry. MAH and CRB suggested inclusion of and provided material for the diversity panel. All authors read and approved the final manuscript.

Abstract

Background: Endoreduplication, the process of DNA replication in the absence of cell division, is associated with specialized cellular function and increased cell size. Genes controlling endoreduplication in tomato fruit have been shown to affect mature fruit size. An efficient method of estimating endoreduplication is required to study its role in plant organ development. Flow cytometry is often utilized to evaluate endoreduplication, yet some tissues and species, among them the tubers of *Solanum tuberosum*, remain intractable to routine tissue preparation for flow cytometry. We aimed to develop a method through the use of protoplast extraction preceding flow cytometry, specifically for the assessment of endoreduplication in potato tubers.

Results: We present a method for appraising endoreduplication in potato (*Solanum tuberosum*) tuber tissues. We evaluated this method and observed consistent differences between pith and cortex of tubers and between different cultivars, but no apparent relationship with whole tuber size. Furthermore, we were able to observe distinct patterns of endoreduplication in 16 of 20 wild potato relatives, with mean endoreduplication index (EI) ranging from 0.94 to 2.62 endocycles per cell. The protocol was also applied to a panel of starchy root crop species and, while only two of five yielded reliable flow histograms, the two (sweet potato and turnip) exhibited substantially lower EIs than wild and cultivated potato accessions.

Conclusions: The protocol reported herein has proven effective on tubers of a variety of potato cultivars and related species, as well as storage roots of other starchy crops. This method provides an important tool for the study of potato morphology and development while revealing natural variation for endoreduplication which may have agricultural relevance.

Keywords: Solanaceae, endopolyploidization, *Solanum tuberosum*, karyoplasmic ratio

Background

Endoreduplication is the replication of the nuclear genome without subsequent cytokinesis, resulting in cells that have greater DNA content than somatic cells remaining in the mitotic cycle. The exact purpose of this alternate cell cycle process remains enigmatic but it is thought to play a role in differentiation and specialization of cells and is common in plants, especially within storage and nutritive tissues (Grafi and Larkins 1995; Larkins et al. 2001; Walker et al. 2000). One hypothesis, predicated on the positive correlation between cell size and nuclear DNA content, is the ‘karyoplasmic ratio’; that an increase in nuclear size and DNA content is required to maintain homeostasis with the large cell sizes necessitated by certain storage, secretory, or structural functions (Wilson 1925b; Barow 2006). It may be tempting to reduce the role of endoreduplication in cell size determination to one of direct causality; however, the real picture is considerably more complex. For instance, Chevalier et al. demonstrated that manipulation of cell cycle regulatory genes may alter the endoreduplication index (EI), the mean number of replication cycles per cell, and concomitantly the size of the tomato fruit itself (2014). Conversely, experiments in *Arabidopsis* have shown that ectopic overexpression of EI promoting

genes often yields plants with severe dwarfism, perhaps due to perturbations in the normal organizational structure of the developing plant (Larson-Rabin et al. 2009). Additional research is required to understand the nuanced influences endoreduplication has on plant organ size if it is to be a potential target for selective breeding or other approaches in crop improvement.

As Chevalier et al. [5] demonstrated profound effects caused by alterations of EI on tomato fruit, we aimed to investigate the extent of endoreduplication in another nutrient sink tissue, potato tubers. Typically, endoreduplication can be assessed by flow cytometry through the use of crude tissue preparations or fixations and DNA-binding fluorophores (Hare and Johnston 2011; Doležel et al. 2007b). Fluorescent intensity is directly correlated with DNA content, meaning that for each endocycle a given cell has undergone, its fluorescent intensity will double relative to a 2C somatic cell. The relative abundance of peaks corresponding to different nuclear DNA contents (C-values) can be observed and an average number of endocycles per cell, the endoreduplication index, can be calculated (Bourdon et al. 2011). Unfortunately, we found the protocol that we routinely employ to establish ploidy and endoreduplication in potato leaves, which uses a crude preparation in a modified Galbraith's buffer, yielded poor and unreliable results when applied to tuber preparations, likely due to the abundance of starch-storing amyloplasts and/or modified cell wall composition (Galbraith et al. 1983; Owen HR 1988). To the authors' knowledge there are only two research articles that describe investigation of tuber endoreduplication levels, both of which paired acetic acid fixation to pectinase digestion (Chen and Setter 2003; Chen and Setter 2012); upon attempting this approach, we occasionally obtained reasonable histograms but found the protocol to be unreliable. Previously, protoplast generation prior to flow cytometry has been used to reduce debris and increase clarity of histograms for other plant tissues (Doležel et al. 2007a; Ulrich and Ulrich 1991; Ochatt 2008).

After digestion of the cell wall and extracellular matrix, free cell suspensions can easily be obtained, potentially allowing for more effective release of nuclei and greater binding of the fluorophore to the DNA while removing the effects of complicating debris. Isolation of potato tuber protoplasts has also been reported (Jones et al. 1989; Doke and Tomiyama 1980a; Davis and Currier 1986); however no protocol combining the use of tuber protoplasts with flow cytometry for the evaluation of endoreduplication has been described. The goal of this research was to develop such a protocol, assess the breadth of its application, and employ it to investigate how endoreduplication varies among tuber tissues, tuber sizes, and accessions of potato species.

Methods

Plant Material

Field grown tubers from the Montcalm Research Farm, Michigan, of *Solanum tuberosum* Group Tuberosum cv. Superior were kindly provided by David Douches (Michigan State University) and used within 8 weeks of harvest to optimize the protocol and then compare the EI between tuber cortex and tuber pith in samples from small (10-20 g), medium (50-60 g), or large (>100 g) tubers. The tissue and tuber size experiment was performed using a randomized complete block design with three replications, blocked by day. A single longitudinal core was taken of each tuber and the innermost 1 cm of tissue was used as the pith sample while the outermost, excluding the epidermis, 5 mm sections from each side of the core were combined for the cortex sample. The protocol was evaluated on a wide range of tuber-bearing *Solanum* spp. comprising a diversity panel, using greenhouse tubers harvested from 3-month old plants grown from February-May in East Lansing, MI under 23˚C days and an increasing 11-14 h ambient photoperiod, then stored at 4˚C/95% RH for 8-9 weeks prior to sampling. Due to variation in

storage response, some tubers were in poor condition (dehydrated, soft, or signs of fungal rot) prior to sampling; however that does not appear to have entirely precluded success with the protocol. To accommodate the variation in tuber size and shape, we used an entire longitudinal core for sampling, with some exceptions. Small tubers (< 1.5 cm length) were prepared using a scalpel to recreate a longitudinal core. For the largest tubers (> 6 cm length), the longitudinal core was halved in the center, as to avoid overtaxing the digestion solution, and the distal half was used. To evaluate the protocol on a range of starchy vegetables, we purchased plant material for a panel of other starchy root and tuber crops from a local grocery store: Beets (*Beta vulgaris* subsp. *vulgaris*), carrots (*Daucus carota* subsp. *sativus*), turnip (*Brassica rapa* subsp. *rapa*), cassava (*Manihot esculenta*), sweet potato (*Ipomoea batatas*), and potatoes of different varieties and pigments (*Solanum tuberosum;* purple, red, yellow, russet). The carrots and beets were marked as organic; otherwise their history and origin are unknown. Here, no blocking was performed and all samples were processed on the same day. A 1 cm central fraction of a longitudinal core of the pith was used for all samples with the exception of cassava (*Manihot esculenta*) where cortex was taken because the pith is tough and inedible. For the ImageStream analysis and protoplast images, cv. Superior was used for both the leaf and tuber pith samples. Leaves were taken from *in vitro* plants while the tubers were treated as above.

Tuber Flow Cytometry Protocol

A visual diagram of the protocol is displayed in **Supplementary Figure 1.1**. The protoplast isolation steps are similar to previously published potato tuber protoplast protocols with some modifications (Davis and Currier 1986; Doke and Tomiyama 1980a). Washed tubers were surface sterilized by submersion in 75% ethanol for 5 min and then allowed to dry under a laminar flow hood. A longitudinal core was taken from the apical to the basal end using a heat

sterilized 7 mm cork borer. Depending on the experiment and thus tissue being sampled, a specific 1 cm section of the core was taken and sliced into approx. 3 mm^3 pieces and placed into a 50 ml conical centrifuge tube containing 15 ml filter-sterilized (0.45 μm) plasmolysis incubation solution (PIS) (0.55 M mannitol, 2 mM $CaCl_2$, 1 mM KH_2PO_4, 1 mM $MgCl_2$, 50 mM Tris buffer, pH adjusted to 7.5 with HCl) and incubated overnight at 4°C. Following plasmolysis, the incubation solution was aspirated off with a sterile serological pipette and 10 ml of filter-sterilized (0.45 μm, aPES or other low-protein-binding) enzyme solution [0.71 M mannitol, 3 mM $CaCl_2$, 1 mM KH_2PO_4, 1 mM $MgCl_2$, 4% "Onozuka" R-10 cellulase (Yakult Pharmaceutical Co. Ltd., Tokyo, Japan), 0.8% macerozyme R-10 (Yakult Pharmaceutical Co. Ltd., Tokyo, Japan), 1% hemicellulase (Sigma-Aldrich, St.Louis, MO), 10 mM MES buffer, pH adjusted to 5.8 with NaOH] was added. The samples were then incubated overnight (18-20 h) at 29°C at 180 rpm horizontal shaking. At this point, some deterioration of the tissue was usually visible but often the pieces remained partially intact. After digestion, aseptic technique was no longer necessary. Next, the samples were gently shaken, then allowed to rest for 5 min to allow the protoplasts to settle. The enzyme solution was removed using a serological pipet and replaced with 15 ml modified plasmolysis incubation solution (PW) (0.71 M mannitol, otherwise identical) and the tubes inverted to wash the protoplasts (**Figure 1.1**). After another 5 min rest, the wash solution was then aspirated off, using a pipette to remove as much residual liquid as possible without disturbing the pellet. Next, 1.5 ml of ice-cold modified Galbraith's buffer (13.6 mM sodium citrate-trisodium, 8 mM MOPS, 18 mM $MgCl_2$, 0.4% v/v Triton X-100) were added to each sample prior to 2-5 sec vortexing to break up the aggregated protoplasts. Samples were immediately placed on ice and passed through a 106 μm mesh filter into a 2 ml microcentrifuge tube. Then, RNase A dissolved in modified Galbraith's buffer was added to each sample to a

final concentration of 0.16 mg/ml prior to a brief inversion and a 30 min incubation at RT. Finally, each sample received propidium iodide (PI), also dissolved in modified Galbraith's solution, to 0.04 mg/ml, the tube was then inverted to mix, and incubated on ice for at least 15 min. Samples were allowed to sit on ice for no longer than 2 h prior to flow cytometry.

Flow Cytometry

The stained nuclei samples were analyzed with a BD FACSCalibur (BD Biosciences, San Jose, CA) flow cytometer. Propidium iodide fluorescence was measured with 488 nm laser and a 585/42 bandpass filter. Data were analyzed with FlowJo VX software (Treestar, Inc, Ashland, OR). Nuclei were gated using a FSC-H vs log-scale PI-H plot to remove debris. At least 2000 gated events were collected per sample. Median fluorescent intensity and frequency were calculated for nuclei populations and reported in arbitrary units (AU). Endoreduplication index (EI) represents the mean number of endocycles per cell and was calculated using the following equation: EI = [*4C*] + 2[*8C*] + 3[*16C*] + 4[*32C*] + 5[*64C*] where [*4C*] is the percentage of 4C nuclei, [*8C*] is the percentage of 8C nuclei and so on.

For the ImageStream, nuclei samples of tubers and *in vitro* leaves of potato cv. Superior were obtained through the method described above. Samples were analyzed, imaged, and sized with an Amnis ImageStream Mark II (Amnis Corp, Seattle, WA) imaging flow cytometer. PI fluorescence was measured with 488 nm laser and a 610/30 bandpass filter. Data were analyzed with IDEAS software (Amnis Corp, Seattle, WA). Nuclei were gated using an area vs. fluorescent intensity plot to visualize populations and remove debris.

Results

Tuber flow cytometry protocol

The development of a reliable and straightforward means to investigate the levels of endoreduplication in potato tubers and other starchy organs was the major goal of this project. The success of the protocol can be evaluated by examining the clarity and separation of peaks in the flow cytometry histograms (**Figure 1.1**). Over the course of developing and validating the protocol we noted a few critical steps which greatly influenced the quality of results. Foremost, as much of the PW solution as possible should be removed in the step following enzyme digestion and preceding the addition of the modified Galbraith's buffer. As little as 100 µl remaining high osmotic pressure PW solution is enough to severely degrade the nuclei rendering the sample useless. It is not clear if this is due to the osmotic pressure itself or activity of residual enzyme but, during troubleshooting, this step was found to be the most frequent cause of failed samples. Thus, while a serological pipette is appropriate for aspirating during other steps of this protocol, greater care must be taken to ensure complete removal of liquid prior to release of nuclei via the Galbraith's buffer. The other critical step is the vortexing or vigorous shaking of the sample after the addition of the modified Galbraith's buffer. Here, much of the tissue will remain aggregated despite the enzyme treatment and it is necessary to break up these aggregates to ensure complete exposure of the protoplasts to the buffer as required for nuclei release. Since different genotypes may vary in degree of digestion, the duration of the vortex or shake will vary by sample; however care must be taken not to 'over-vortex' as damage to the nuclei may be incurred. For instance, samples of cv. Superior required only 2-3 seconds of vortexing while at least 5 seconds were necessary for PI 234011 samples. Another note of some significance is the use of aseptic technique to ensure samples are not contaminated prior to incubation and

digestion. While care was taken to avoid such contamination, it was nevertheless observed in a few samples. From our experience, bacterial contamination did not render a sample entirely useless but rather 'messier' with less distinction between peaks of different C-values which was presumably caused by bacterial damage to protoplasts and the nuclei they contain. Hence, it is possible that aseptic technique may not be necessary for certain gross applications requiring maximum throughput but remains best practice for obtaining the highest quality results. Lastly, all samples should be run on the flow cytometer within 2 h of the addition of the modified Galbraith's buffer as after this window we saw considerable decline in sample quality as well as peak shifting, indicating deterioration of nuclei. It is possible this could be avoided through the use of fixatives however this has not been tested.

Tissues of cv. Superior

To test the consistency of results obtained with the protocol, as well as examine how EI differs across sizes and tissues of potato tubers, we evaluated the cortex and tuber tissues of various tuber sizes of cv. Superior. We observed a significant difference in endoreduplication levels between the cortex and pith tissue across tuber sizes ($p = 0.002$); however there was no significant difference due to tuber size itself ($p = 0.89$). Pith of cv. Superior had an average EI of 1.73 endocycles per cell while cortex tissue yielded an average of 1.31. Next, we applied the protocol to a diverse panel of semi-domesticated and wild *Solanum* species to test if the protocol could accommodate samples with potentially different cell wall composition and variable amounts of cellular and extracellular debris.

Solanum spp. diversity panel

We found the protocol was successful on at least one sample for the majority (16/20 tested) of the accessions in the diversity panel indicating its application, while broad, is not universal. Of note is that these genotypes varied greatly in tuber quality after storage and one of the unresponsive genotypes, PI 458355, was of particularly poor quality at the time of sampling (**Figure 1.3 B**). Of the other three unresponsive lines, PI 473385 and PI 546023 appeared to have limited or no digestion after the enzyme incubation while PI 558050 seem to oxidize rapidly, despite being promptly immersed in the plasmolysis solution. We observed significant differences among the responsive genotypes of the diversity panel (**Figure 1.3 A**). For instance, PI 498359, a diploid *S. kurtzianum* accession, demonstrated the lowest mean EI of the panel at 0.94 while the greatest was PI 265863, a diploid *S. candolleanum* line, with a mean EI of 2.62 (**Figure 1.3 C,D**). It should be noted that the diversity panel exhibited greater sample-to-sample variation than either the cv. Superior tissue samples or other cultivar tubers and potential reasons are considered in the discussion.

Cultivars and root crop species

Next, we elected to apply the protocol on a variety of other root crops to further evaluate its effectiveness on other starchy plant organs wherein endoreduplication levels may be germane to development. We included four store-bought potato varieties (purple, red, russet, yellow) with unknown histories of growing conditions, harvest, storage and exact cultivar identities to investigate if significant EI differences exist among cultivated potato varieties. We observed slight but significant differences between some of the potato varieties. For instance, yellow potato averaged 1.98 endocycles per cell while the red variety displayed a mean 1.56 per cell. Of the other starchy root crops, only turnip and sweet potato yielded distinct peaks while cassava,

beet, and carrot did not. Due to variation in nuclear DNA content we expected to observe some shifting of the 2C peak compared to potato. As expected, we observed a minor decrease in fluorescent intensity from potato to sweet potato and an even further decrease in turnip. This is in accordance with previous literature where potato has been found to contain 3.32-3.86 pg/2C nucleus whereas sweet potato and turnip contain 3.31 pg and 1.06 pg, respectively (Arumuganathan and Earle 1991). Nevertheless we observed moderate levels of endoreduplication in turnip, although much lower than what was observed in potato and has been previously reported in fruits of many crop species. Sweet potato demonstrated low levels of endoreduplication with an average of only 0.2 endocycles per cell (**Figure 1.4**). A summary of all evaluated species and accessions is presented in **Table 1. 1** and all raw data for successful samples are available in a **Supplemenary Table 1. 1**.

ImageStream flow cytometry

To verify that the protocol identified and classified nuclei based of their size and fluorescence intensity we employed an ImageStream flow cytometer to capture images of potato cv. Superior nuclei and other particles which were released during the sample preparation. Hence, the ImageStream allowed us to produce the same results as the FACSCalibur flow cytometer, albeit at a slower pace, with the notable advantage of being able to capture an image of each event. With the ImageStream, we were able to verify the relationship between nucleus size and fluorescent intensity as evidenced by the separate populations observed (**Figure 1.5 A,B**). Furthermore, a representative nucleus, one of the events depicted in panels A and B, from each C-value population is displayed to illustrate the increase in size that accompanies each endocycle (**Figure 1.5 C- J**). Taken together, this demonstrates the stepwise increase in DNA content in an

endopolyploid nucleus is accompanied by an increase in total nucleus size and greater intensity of fluorescence which is then detectable through the protoplast-flow cytometry protocol.

Discussion

Tuber flow cytometry protocol

Here, we describe a novel protocol for the evaluation of endoreduplication levels of potato tubers; a phenomenon which has previously received limited investigation despite its relevance to development and potential as a target for crop improvement. We demonstrate that the protocol is effective on an assortment of potato relatives as well as some tuberous root crops. The protocol requires two overnight incubations as well as the process of sterilizing and coring the tuber samples which may be time consuming. Therefore, while it is demonstrably appropriate for small to moderate sized experiments, the throughput may limit its application in ventures requiring high-throughput analysis such as fine mapping or assessment of large quantities of breeding stock. Some protoplast isolation methods, such as that presented by Doke and Tomiyama, utilize a shorter enzymatic incubation period which we found to garner protoplast quantities insufficient for flow cytometry (Doke and Tomiyama 1980a). Apart from the time investment, another limitation of the approach is the high cost of reagents, particularly enzymes. To mitigate this, researchers may consider using a smaller volume of enzyme solution and/or tissue per preparation. We have observed satisfactory results with as much as 3 cm of core (7 mm diameter) in 10 ml of enzyme solution, indicating some tolerance for adjustment in the ratio of enzyme to tissue. Alternatively, it may be possible to reuse enzyme through filtration and subsequent pH adjustment as reported in Saxena and King (1985) where similar yields were observed reusing enzyme solution as many as two times; however, this was not attempted here.

Finally, researchers aiming to make additional improvements to the protocol may consider altering the type and ratios of the enzymes utilized for protoplast isolation. For instance, the protocol presented herein employs Onozuka R-10 cellulase which may contain impurities, including nucleases, thus potentially damaging nuclei (Bhojwani and Dantu 2013; Bengochea and Dodds 1986). More purified cellulase preparations are available (e.g., Onozuka RS), albeit at increased cost, which may expand the protocol to recalcitrant accessions or improve the quality of data obtained. While the protocol does have its shortcomings, we believe it represents a substantial step forward for the assessment of endoreduplication in root and tuber crops, as evidenced by the differences we observed between tissues, accessions, and species.

Tissues of cv. Superior

The results of our experiment on the chipping cultivar, Superior, indicate that, at maturity, there is a difference between the EI of cortical and pith tissues but no apparent relationship to overall tuber size. It has been previously noted from microscopy work that, especially in the latest stages of development, perimedullary parenchyma, in the form of cell division and enlargement, is the predominant contributor to increases in total tuber size, while pith and cortex tissues are comparatively static (Xu et al. 1998). Additionally, prior research has demonstrated that across the developmental spectrum, pith cells are consistently larger than cortical cells and may therefore be expected to display greater levels of endoreduplication (Peterson et al. 1985). Hence, our results are consistent with prior literature as even the smallest tubers we examined (10-20 g) were well beyond the initial stage where cortical and pith division and enlargement are most abundant (< 8 mm). Further research must be performed to determine if the increase in cell number and size in the perimedullary tissue during later development is accompanied by increases in EI. Additionally it would be interesting to determine the degree to which maximum

attainable tuber size of various cultivars is dictated by increases in cell size versus cell proliferation.

Solanum spp. diversity panel

The evaluation of the *Solanum* species panel establishes that the protocol's application extends beyond cultivated potato and thus might be appropriate for diversity and domestication studies. The differences observed in these wild and semi-domesticated lines demonstrate substantial natural variation that may have been selected for in the domestication of potato. Furthermore, considering that some of these tubers had stored poorly or displayed defects (dehydration, fungal growth, hollow heart) yet the majority were still responsive shows the protocol may be robust to other tuber quality defects such as tuber end rot. Three of the accessions responded poorly despite minimal visual decline in storage. For two of these, PI 473385 and PI 546023, these samples remained fully intact following the enzyme treatment and vortexing, indicating minimal digestion. It seems plausible that an alteration of the digestion solution, or addition of other enzymes commonly used in protoplast generation, may yield greater decomposition and thereby allow these genotypes to be evaluated as well. As previously mentioned, the diversity panel yielded greater variation between samples of the same genotype than either the cv. Superior tissue or the root crop panel experiments. We believe this was caused by a few separate factors: sampling, size, and quality. As these were wild and semi-domesticated genotypes, large, well-storing tubers were not consistently available for all accessions and the diminutive size of some of the tubers precluded sampling of specific tissues. Therefore, we elected to sample all tissues via a longitudinal core. This allowed us to maintain a consistent approach for tubers of all sizes but may have introduced greater variation if there were different relative abundances of tissues between tubers in a given genotype. As mentioned previously, the tubers exhibited various

defects which may or may not have been present in all tubers of a genotype, perhaps contributing to sample variation as well.

Cultivars and root crop species

The collection of grocery store tubers or roots of other starchy crops indicated the broad applicability of the protocol for potato cultivars with little regard of growing conditions or storage but limited suitability to other starchy crops. While sample culture conditions for this experiment are unknown and environmental factors such as stress are known to influence EI (Scholes and Paige); the significant differences observed between the pith of the various tuber accessions likely indicate that variation for EI present within cultivated germplasm allows the possibility of selection. However, whether selection for greater endoreduplication would actually yield plants that produce larger tubers or higher yields remains to be seen. What is apparent is that all of our potato tuber samples were highly endoreduplicated and a reliable protocol permits the study of genetic control of the process. While sweet potato and turnip reacted well, cassava, beet, and carrot yielded no usable results. This indicates that other, untested, starchy crops may or may not be responsive. The stark contrast between EI values in potato pith, where the lowest variety (red potato) was 1.56, and the other crops which ranged from 0.22 (sweet potato) to 0.87 (turnip) may be related in part to their differences in organogenesis; While potato tubers originate from stolons, a modified subterranean stem, the other organs are modified tuberous roots. While tuberous roots such as sweet potato may gradually swell over the lifecycle of the plant, bulking of potato tubers is a relatively rapid process, perhaps requiring a more acute manner of starch deposition and cellular expansion made possible by dramatic increase in nuclear DNA content. For instance, signs of sweet potato storage root initiation can be observed in as little as 13 days after transplant despite a growth period ranging from 90 to 150 days with

an average as high as 140 days in some areas, not including growth that occurs prior to transplanting (Woolfe 1992; Nedunchezhiyan et al. 2012; Villordon et al. 2009). Conversely, potato tuber initiation is highly dependent on photoperiod and occurs later in plant development compared with sweet potato (Kloosterman et al. 2013). For instance, cv. Russet Burbank may take around 9 weeks after planting to initiate tuberization when grown in North America despite being a late maturity variety with a time requirement similar to that of sweet potato (Ojala et al. 1990). Evidence for a negative correlation between endoreduplication levels and duration of organ development has been presented in fleshy fruits and a similar association may be uncovered given more investigation of tubers and tuberous roots (Bourdon et al. 2010).

Conclusions

Here, we have demonstrated a novel procedure for evaluating the endoreduplication levels of tubers of potato, its close relatives, and some starchy root crops. The procedure, while time consuming, is reliable, relatively straightforward, and somewhat accommodating of common tuber defects. Nevertheless, it is possible further alterations may be made to increase throughput, minimize costs or further expand its application to other crop species. In establishing the efficacy of the protocol, we observed significant differences in endoreduplication between tuber tissues, cultivars, wild accession and species while there was no apparent difference between mature tubers of different sizes of the same variety. This protocol, in addition to the results we report, may serve as an early glimpse of the variation of endoreduplication in starchy root crops which may have significant agronomic relevance.

Abbreviations

EI: Endoreduplication Index

PI: Propidium iodide or Plant Introduction, depending on context

PIS: Plasmolysis incubation solution

PW: Plasmolysis wash solution (modified incubation solution)

AU: Arbitrary units

C-value: (DNA content value)

Declarations

Ethics approval and consent to participate

Not applicable.

Authors' contributions

FPEL conducted the research, analyzed the data and drafted the manuscript. FPEL, SHH, and REV developed the protocol. MM performed the flow cytometry. MAH and CRB suggested inclusion of and provided material for the diversity panel. All authors contributed to the concepts and various drafts of the manuscript.

Acknowledgements

The authors would like to thank David Douches and Joseph Coombs of Michigan State University for supplying the cv. Superior tubers.

Funding

This research was supported in part by National Science Foundation Award number 1237969, "Unraveling the Heterozygosity, Allelic Composition, and Copy Number Variation of Potato" to CRB and RV and USDA Special Grant 2014-34141-22266 (University of Maine) to RV.

Competing interests

The authors declare they have no competing interests.

Consent for publication

All the authors have approved the manuscript and have made all required statements and declarations.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional file which is available at:

https://plantmethods.biomedcentral.com/articles/10.1186/s13007-017-0177-3

Additional files

Additional_file_1.pdf provides a simplified overview of the entire tuber protoplast flow cytometry protocol. The detailed protocol is described within the text.

The raw data of the tissue, diversity panel, and root crop experiments are available as separate sheets in Additional_file_2.xls. Additional_file_2.xls contains the relative abundances, in percentage, of each c-value peak as well as the EI value calculated from those percentages. **Open Access**

Figures

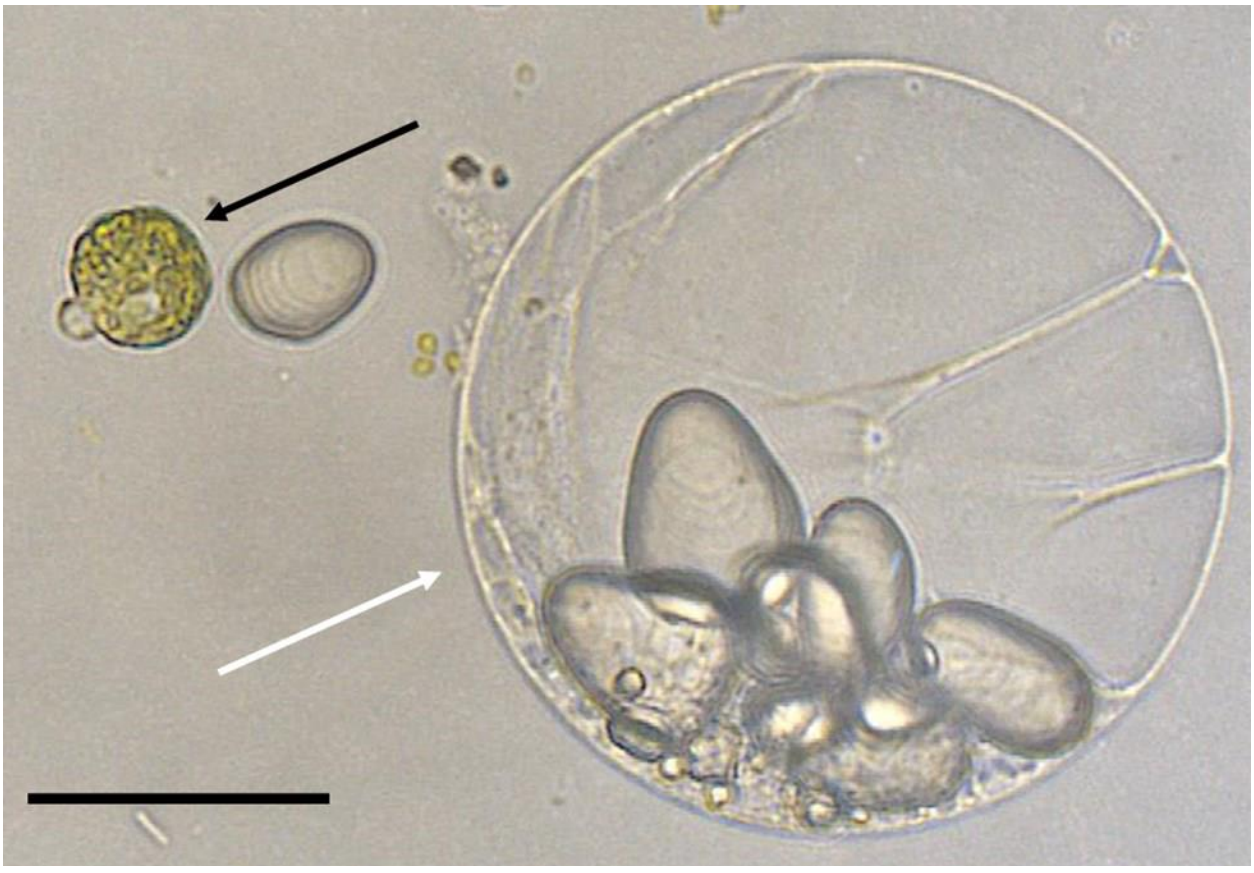


Figure 1.1: Image of leaf (black arrow) and tuber (white arrow) protoplasts prior to the addition of modified Galbraith's flow cytometry buffer. A starch granule lies between the two protoplasts and many more are visible within the tuber protoplast. Scale bar = 50 µm

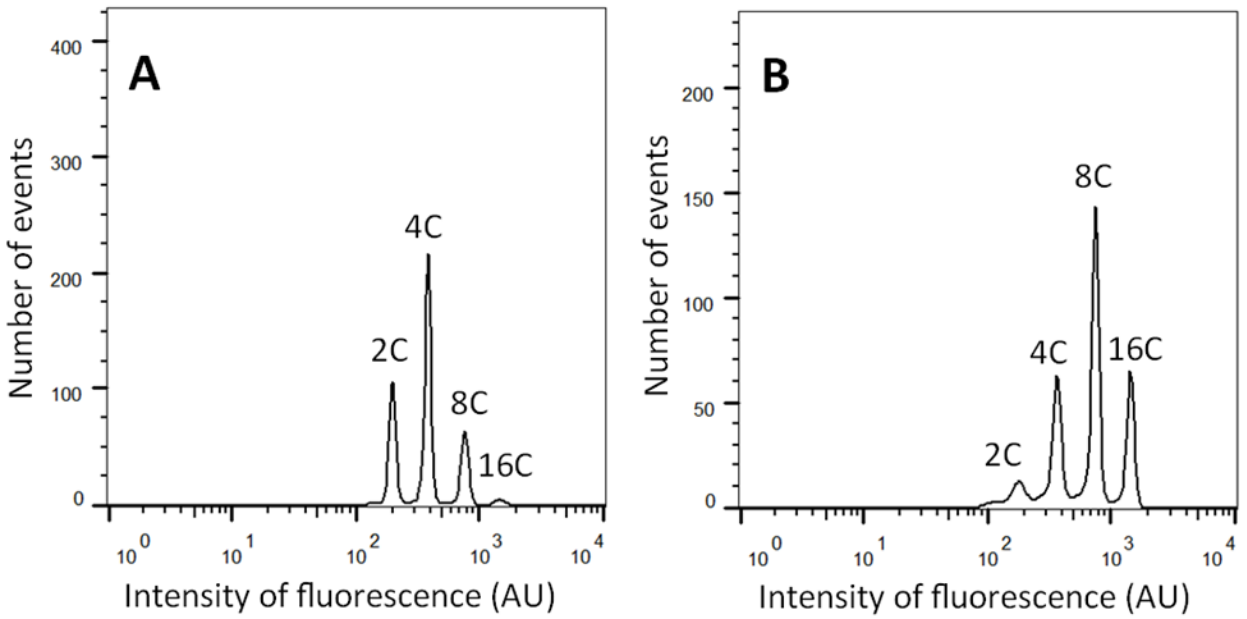


Figure 1.2: Example of flow cytometry histogram obtained from protoplast preparation of *Solanum tuberosum* cv. Superior tuber cortex (A) and pith (B) showing relative abundance of nuclei with differing DNA contents. Note the separation between peaks allowing for reliable frequency estimations.

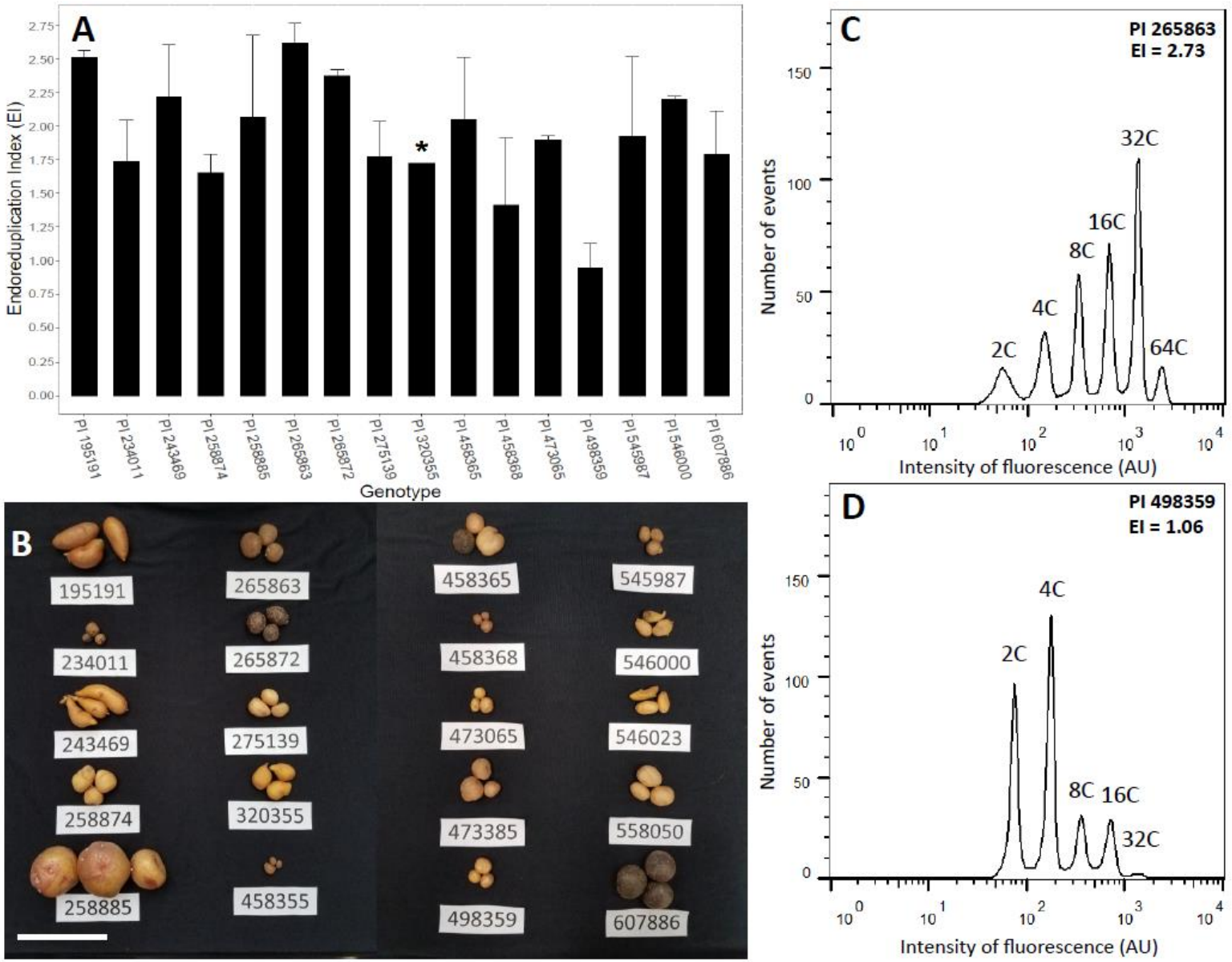


Figure 1.3: The results of the tuber endoreduplication protocol applied to the diversity panel. Panel A shows mean EI for each accession using a longitudinal core. Asterisk indicates genotype with only one successful rep. Panel B displays the tubers at the time of sampling, including those genotypes for which successful results were not obtained. Note variation in tuber size, shape, and rigidity. Scale bar = 10 cm. Panels C and D show representative flow cytometry histograms and EI values of the genotypes with the highest (C, PI 265863) and lowest (D, PI498359) mean endoreduplication values.

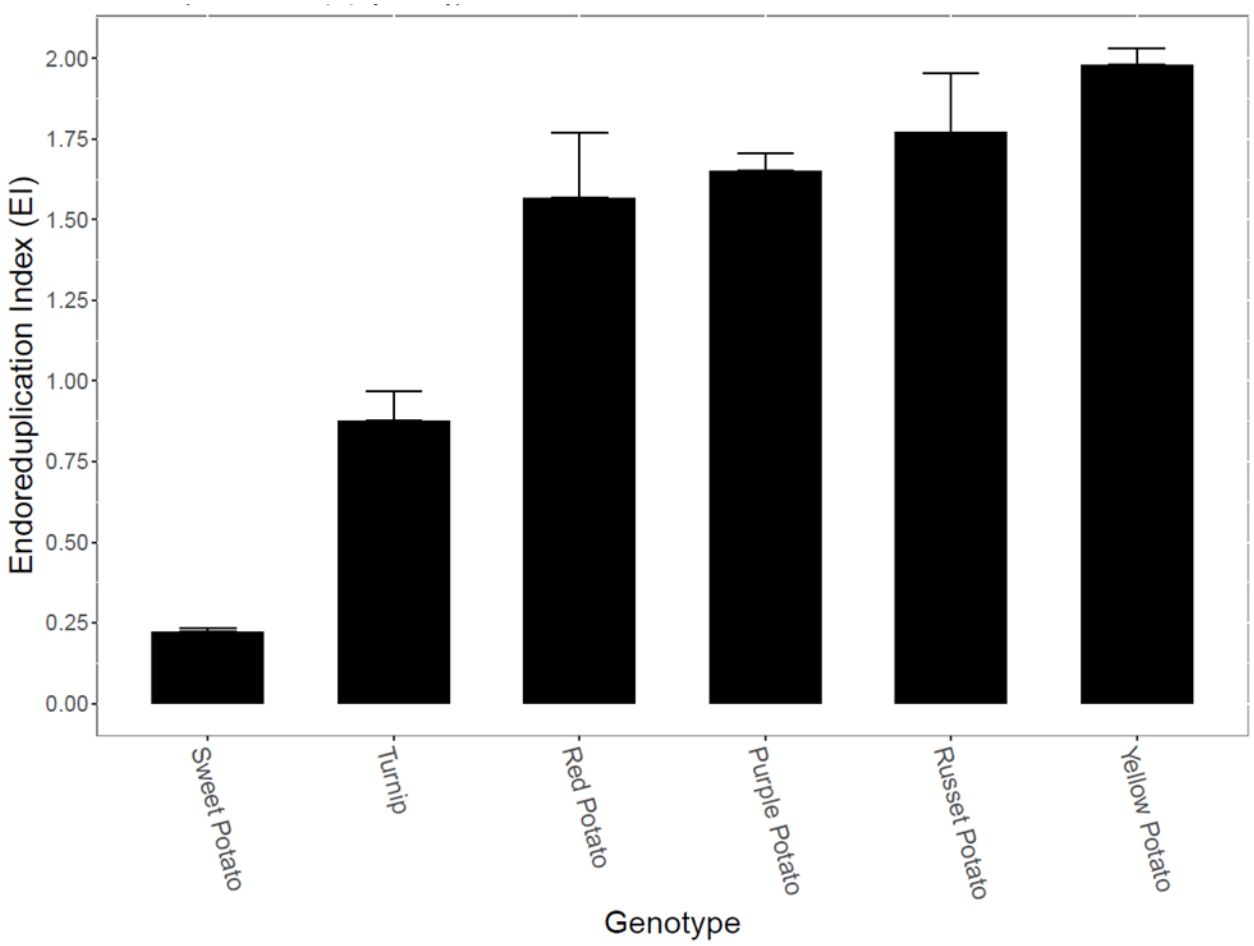


Figure 1.4: Mean EI values obtained with the protocol for the tuber and tuberous root crop panel. Unresponsive species are not shown.

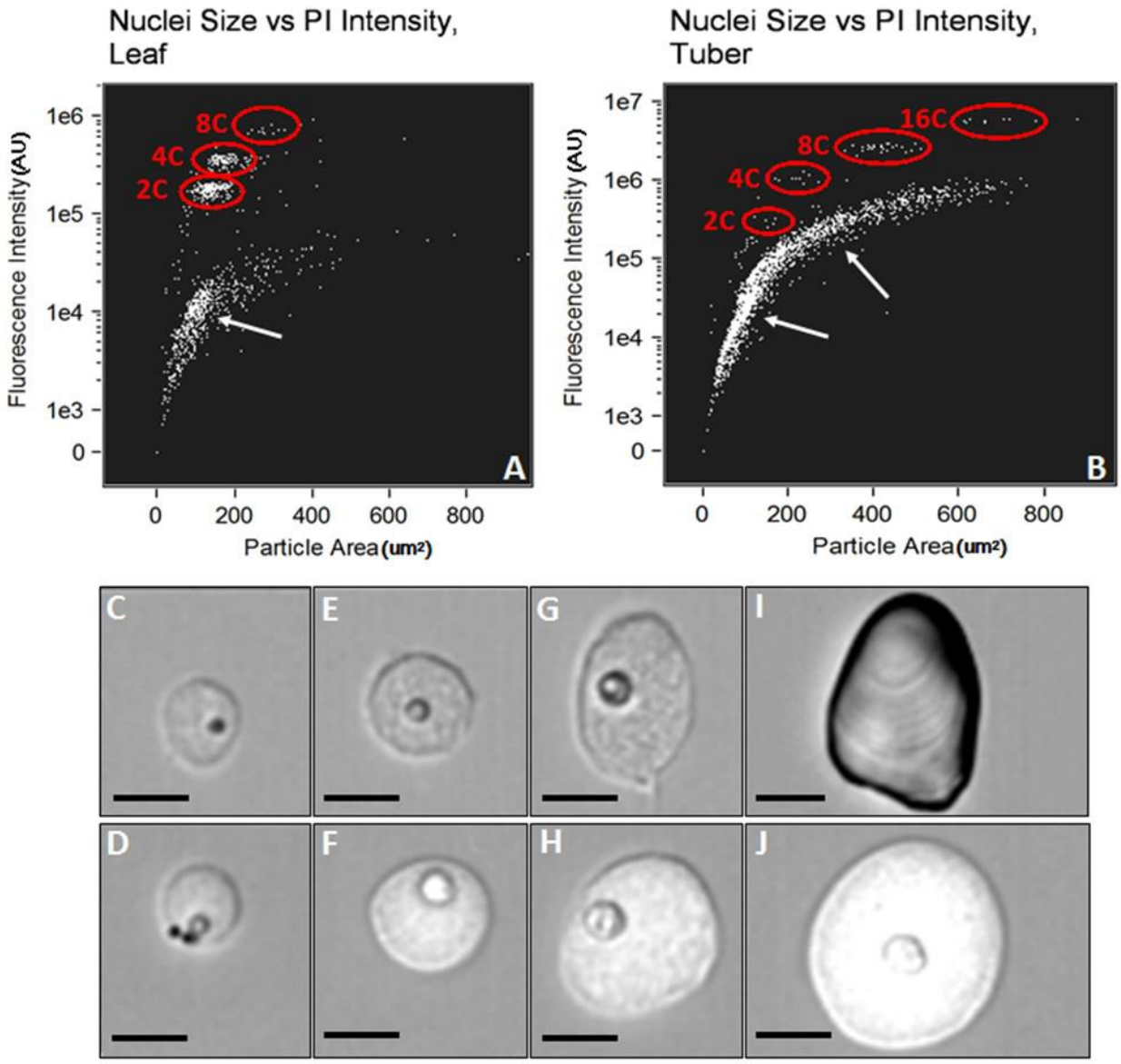
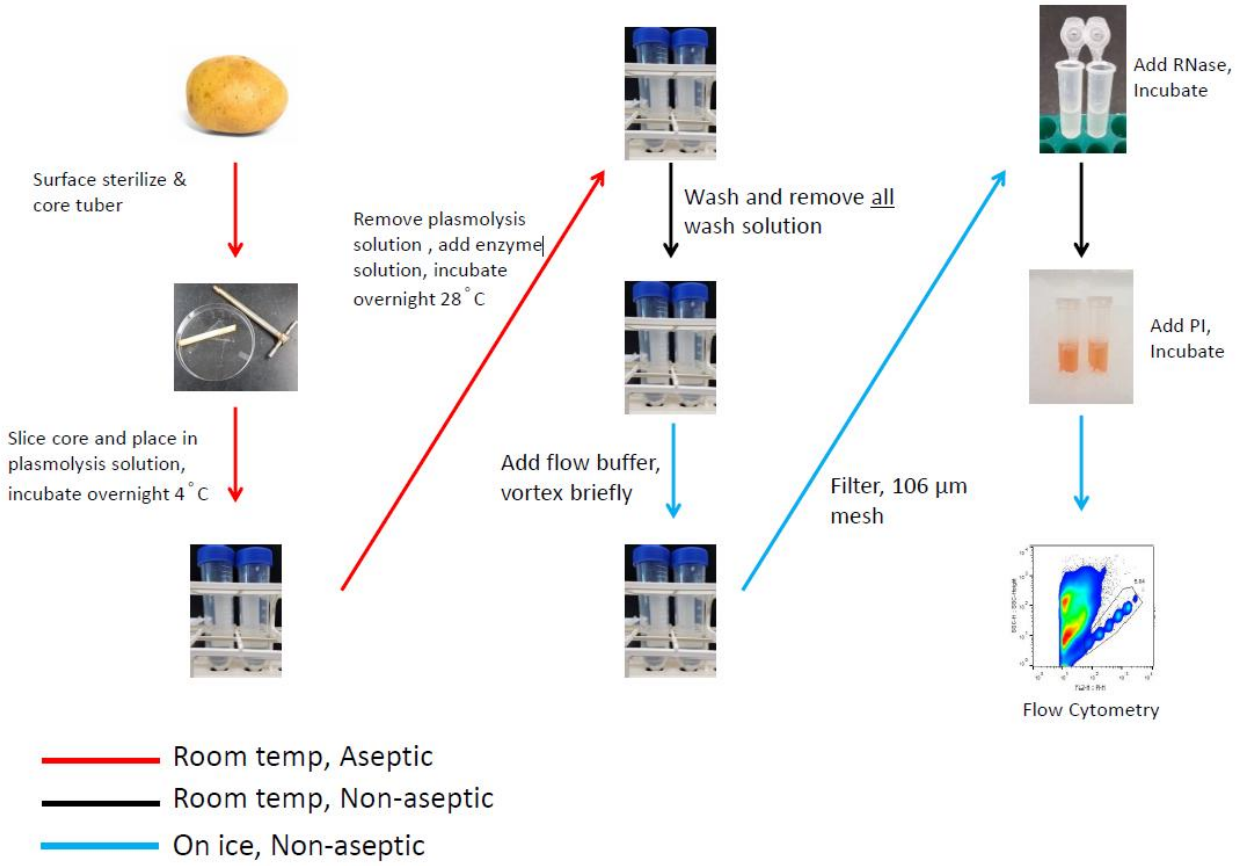


Figure 1.5: Relationship between fluorescence intensity and particle area in leaf (A) and tuber pith (B) nuclei obtained from protoplast preparations. Populations of different C-values are circled. Other points represent debris (arrows) C-J: Bright-field images obtained during ImageStream flow cytometry of leaf (C, E, G) and tuber (D, F, H, J) nuclei obtained from protoplast preparations with increasing C-values which are representative of the populations circled in panels A and B. Panel I depicts a large starch granule from the "debris" in B. Scale bar = 10 µm. Note: scatterplots are meant to be illustrative and were not used to calculate relative abundances of nuclei due to lower flow rate of ImageStream cytometer.



Supplementary Figure 1.1: The general process of the protoplast extraction and flow cytometry protocol.

Table 1.1: Potato accessions and other species used for flow cytometric analysis of starchy tissues. PI: Plant introduction; Response: + acceptable flow cytometry histograms obtained; - no meaningful histograms obtained.

Species	PI Number	Ploidy	Response
S. tuberosum	195191	2*x*	+
S. tuberosum	234011	2*x*	+
S. tuberosum	243469	2*x*	+
S. tuberosum	258874	4*x*	+
S. tuberosum	258885	4*x*	+
S. candolleanum	265863	2*x*	+
S. medians	265872	2*x*	+
S. chacoense	275139	2*x*	+
S. tuberosum	320355	2*x*	+
S. microdontum	458355	2*x*	-
S. berthaultii	458365	2*x*	+
S. okadae	458368	2*x*	+
S. brevicaule	473065	2*x*	+
S. brevicaule	473385	2*x*	-
S. kurtzianum	498359	2*x*	+
S. brevicaule	545987	2*x*	+
S. boliviense	546000	2*x*	+
S. tuberosum	546023	4*x*	-
S. commersonii	558050	2*x*	-
S. tuberosum	607886	4*x*	+
S. tuberosum cv. Superior	-	4*x*	+
S. tuberosum (russet)	-	4*x*	+
S. tuberosum (red)	-	4*x*	+
S. tuberosum (yellow)	-	4*x*	+
S. tuberosum (purple)	-	4*x*	+
Brassica rapa subsp. *rapa (*turnip)	-	2*x*	+
*Manihot esculenta (*cassava)	-	2*x*	-
Beta vulgaris subsp. *vulgaris* (beet)	-	2*x*	-
Daucus carota subsp. *sativus* (carrot)	-	2*x*	-
*Ipomoea batatas (*sweet potato)	-	6*x*	+
Total	**-**	**-**	**24/30**

Works Cited

1. Grafi G, Larkins BA. Endoreduplication in maize endosperm: Involvement of M phase--promoting factor inhibition and induction ofS phase--related kinases. Science. 1995;269:1262-4.
2. Larkins BA, Dilkes BP, Dante RA, Coelho CM, Woo YM, Liu Y. Investigating the hows and whys of DNA endoreduplication. J Exp Bot. 2001;52:183-92.
3. Walker JD, Oppenheimer DG, Concienne J, Larkin JC. *SIAMESE*, a gene controlling the endoreduplication cell cycle in arabidopsis thaliana trichomes. Development. 2000;127:3931-40.
4. Wilson E. The karyoplasmic ratio. 3rd ed. New York: The Macmillan Company; 1925.
5. Barow M. Endopolyploidy in seed plants. Bioessays. 2006;28:271-81.
6. Chevalier C, Bourdon M, Pirrello J, Cheniclet C, Gévaudant F, Frangne N. Endoreduplication and fruit growth in tomato: Evidence in favour of the karyoplasmic ratio theory. J Exp Bot. 2014;65:2731-46.
7. Larson-Rabin Z, Li Z, Masson PH, Day CD. *FZR2/CCS52A1* expression is a determinant of endoreduplication and cell expansion in Arabidopsis. Plant Physiol. 2009;149:874-84.
8. Hare EE, Johnston JS. Genome size determination using flow cytometry of propidium iodide-stained nuclei. In Molecular methods for evolutionary genetics. Orgogozo V, Rockman MV, editors. Totowa, NJ: Humana Press. 2011. p.3-12.
9. Doležel J, Greilhuber J, Suda J. Flow cytometry with plants: An overview. In Flow cytometry with plant cells. editors. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. 2007. p.41-65.
10. Bourdon M, Coriton O, Pirrello J, Cheniclet C, Brown SC, Poujol C, Chevalier C, Renaudin JP, Frangne N. *In planta* quantification of endoreduplication using fluorescent *in situ* hybridization (FISH). Plant J. 2011;66:1089-99.
11. Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science. 1983;220:1049-51.
12. Owen HR VR, Levy D, Ochs DL Environmental, genotypic, and ploidy effects on endopolyploidization within a genotype of *Solanum phureja* and its derivatives. Genome. 1988:506-10.
13. Chen C-T, Setter TL. Response of potato tuber cell division and growth to shade and elevated CO2. Ann Bot. 2003;91:373-81.
14. Chen C-T, Setter TL. Response of potato dry matter assimilation and partitioning to elevated CO2 at various stages of tuber initiation and growth. Environ Exp Bot. 2012;80:27.

15. Doležel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry. Nat Protoc. 2007;2:2233-44.

16. Ulrich I, Ulrich W. High-resolution flow-cytometry of nuclear-DNA in higher-plants. Protoplasma. 1991;165:212-5.

17. Ochatt SJ. Flow cytometry in plant breeding. Cytometry A. 2008;73:581-98.

18. Jones H, Karp A, Jones MG. Isolation, culture, and regeneration of plants from potato protoplasts. Plant Cell Rep. 1989;8:307-11.

19. Doke N, Tomiyama K. Effect of hyphal wall components from P*hytophthora infestans* on protoplasts of potato tuber tissues. Physiol Plant Pathol. 1980;16:169-76.

20. Davis DA, Currier WW. The effect of the phytoalexin elicitors, arachidonic and eicosapentaenoic acids, and other unsaturated fatty acids on potato tuber protoplasts. Physiol Mol Plant Pathol. 1986;28:431-41.

21. Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. Plant Mol Biol Rep. 1991;9:208-18.

22. Saxena PK, King J. Reuse of enzymes for isolation of protoplasts. Plant Cell Rep. 1985;4:319-20.

23. Bhojwani SS, Dantu PK. Tissue and cell culture. In Plant tissue culture: An introductory text. Springer. 2013. p.39-50.

24. Bengochea T, Dodds JH. Plant protoplasts : A biotechnological tool for plant improvement. London : Chapman and Hall; 1986.

25. Xu X, Vreugdenhil D, Lammeren AAMv. Cell division and cell enlargement during potato tuber formation. J Exp Bot. 1998;49:573-82.

26. Peterson L, Barker WG, Howarth MJ. Development and structure of tubers. In Potato physiology. Li PH, editors. Orlando, Florida: Academic Press. 1985. p.124-48.

27. Scholes DR, Paige KN. Plasticity in ploidy: A generalized response to stress. Trends Plant Sci.20:165-75.

28. Woolfe JA. Sweet potato: An untapped food resource. Cambridge University Press; 1992.

29. Nedunchezhiyan M, Byju G, Jata SK. Sweet potato agronomy. Fruit Veg Cereal Sci Biotech. 2012;6:1-10.

30. Villordon A, LaBonte D, Firon N. Development of a simple thermal time method for describing the onset of morpho-anatomical features related to sweetpotato storage root formation. Sci Hort. 2009;121:374-7.

31. Kloosterman B, Abelenda JA, Gomez Mdel M, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S, et al. Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature. 2013;495:246-50.

32. Ojala JC, Stark JC, Kleinkopf GE. Influence of irrigation and nitrogen management on potato yield and quality. Am Potato J. 1990;67:29-43.

33. Bourdon M, Frangne N, Mathieu-Rivet E, Nafati M, Cheniclet C, Renaudin J-P, Chevalier C. Endoreduplication and growth of fleshy fruits. In Progress in botany 71.

Lüttge U, Beyschlag W, Büdel B, Francis D, editors. Berlin, Heidelberg: Springer Berlin Heidelberg. 2010. p.101-32.

Chapter 2: Measuring Endoreduplication by Flow Cytometry of Isolated Tuber Protoplasts

Authors
F. Parker E. Laimbeer[1], Mellissa Makris[2] and Richard E. Veilleux[1]

Author's institutions:
[1]Department of Horticulture, Virginia Tech, Blacksburg, VA, USA
[2] Department of Biomedical Sciences and Pathobiology, Center for Molecular Medicine and Infectious Diseases, Virginia-Maryland Regional College of Veterinary Medicine, Virginia Tech, VA, USA

Corresponding author:
F. Parker E. Laimbeer
plaimbeer@gmail.com
Tel: (540) 231-6945

Email Addresses of Co-authors:
Mellissa Makris mmakris@vt.edu
Richard E. Veilleux potato@vt.edu

Keywords: C-Value, Endopolyploidy, *Solanum tuberosum*, DNA content, Potato, Solanaceae, Tissue Culture

Author's Contributions

FPEL conducted the research, analyzed the data and drafted the manuscript. FPEL and REV developed the protocol. MM performed the flow cytometry. All authors read and approved the final manuscript.

SHORT ABSTRACT

The protocol described herein is a method for measuring endoreduplication within tubers of potato (*Solanum tuberosum*). It includes plasmolysis and protoplast extraction steps to decrease the noise and debris in downstream flow cytometric analysis.

LONG ABSTRACT

Endoreduplication, the replication of a cell's nuclear genome without subsequent cytokinesis, yields cells with increased DNA content and is associated with specialization, development and increase in cellular size. In plants, endoreduplication seems to facilitate the growth and expansion of certain tissues and organs. Among them is the tuber of potato (*Solanum tuberosum*) which undergoes considerable cellular expansion in fulfilling its function of carbohydrate storage. Thus, endoreduplication may play an important role in how tubers are able to accommodate this abundance of carbon. However, the cellular debris resulting from crude nuclear isolation methods of tubers, methods that can be used effectively with leaves, precludes the estimation of the tuber endoreduplication index (EI). This article presents a technique for assessing tuber endoreduplication through the isolation of protoplasts while demonstrating representative results obtained from different genotypes and compartmentalized tuber tissues. The major limitations of the protocol are the time and reagent costs required for sample preparation as well as relatively short lifespan of samples after lysis of protoplasts. While the protocol is sensitive to technical variation, it represents an improvement over traditional methods of nuclear isolation from these large specialized cells. Possibilities for improvements to the protocol such as recycling enzyme, the use of fixatives, or and other alterations are proposed.

INTRODUCTION

Endoreduplication is the process by which a cell forgoes the typical cell cycle and instead undergoes an alternate course of development consisting of repeated rounds of DNA replication without cellular division. The resulting cell will have increased DNA content and nuclear size which is thought to play a role in cellular regulation, expansion, and specialization. Generally, a round of endoreduplication (termed an endocycle) and the corresponding increase in DNA content are associated with larger cell volume, an observation that precipitated the "karyoplasmic theory" that increased DNA content is required to properly regulate a larger, perhaps more complex, cell (Wilson 1925b). This phenomenon is common in higher plants, having been observed in a range of tissues including those with structural/defensive (trichomes)(Szymanski and Marks 1998; Melaragno et al. 1993), nutritive (maize endosperm)(Grafi and Larkins 1995; Schweizer et al. 1995), and sink/storage (tomato pericarp; potato tuber)(Chen and Setter 2003; Chen and Setter 2012; Cheniclet et al. 2005) functions. In fruit, it has been suggested that endoreduplication plays a role in facilitating the rapid expansion of the pericarp as evidenced by the negative relationship between endoreduplication and fruit developmental period (Pirrello et al. 2014). For instance cells with DNA content up to 512C (512 times the haploid genome) have been observed in the tomato pericarp (Cheniclet et al. 2005). Furthermore Chevalier et al. (2014) demonstrated that alterations in expression of cell cycle genes can lead to increases in endoreduplication levels within the pericarp which then results in larger fruit (Chevalier et al. 2014). Thus, alteration of genes promoting endoreduplication provides a potential target for improvement of biomass or yield through plant breeding or genetic manipulation. However, such improvement is contingent upon greater understanding of the causes and consequences of endoreduplication.

Endoreduplication is most often measured via flow cytometry whereby nuclei, released in crude tissue preparations(Galbraith et al. 1983), are incubated with a DNA-binding fluorophore, such as propidium iodide(Doležel et al. 2007b) (PI). The filtered samples are then passed by the laser of a flow cytometer where emission wavelengths specific to the fluorophore can be observed. The intensity of fluorescence in each event is directly correlated with the DNA-content of the particle. Thus, by comparing to a known standard, the relative and absolute DNA content of cells in a given sample may be calculated. Endoreduplication indices (EI) are determined by establishing the average number of endocycles per cell by observing cellular DNA content (C-value) where 1C is the DNA content of a haploid cell. For instance, in a diploid organism, the base DNA content of somatic cells is 2C. If a sample has few cells with 4C or greater it would have an EI near 0; however if nearly all the cells are 4C the EI would be approximately 1 as most cells would have undergone a single round of endoreduplication. However, as endoreduplication is common in higher plants real world EI values may be much greater. While calculating endoreduplication indexes may be relatively straightforward, certain species and tissues (including potato tubers) remain recalcitrant to typical preparations using a razor blade to release the nuclei directly into appropriate buffers.

In potato, the examination of endoreduplication within the tuber remains limited to just a few studies(Chen and Setter 2003; Chen and Setter 2012), perhaps due in part to a lack of a reliable protocol as the aforementioned crude preparations yield inconsistent results. While such approaches work well for leaves the tuber samples experience severe degradation and an abundance of noise in the form of debris, making differentiation of peaks comprised of nuclei with differing C-values nearly impossible. This is perhaps due to differences in cellular composition and a profusion of cellular debris within tuber cells (*e.g.* starch-storing

amyloplasts). To overcome this hurdle, we recently developed a more reliable protocol for acquiring distinguishable peaks in flow cytometry of potato tubers. The frequency of cells within each peak can then be used for calculating accurate endoreduplication levels for different genotypes or different tissues that comprise a tuber. The protocol, which employs a modified protoplast extraction method(Doke and Tomiyama 1980a; Davis and Currier 1986) antecedent previously described flow cytometry(Galbraith et al. 1983), showed considerable variation within potato relatives, cultivars, and tissues(Laimbeer et al. 2017). Here we present the protocol in detail by evaluating EI in the contexts of ploidy, tissue, and tuber size.

PROTOCOL

(NOTE: It is important to include appropriate controls, usually in the form of leaf tissue of the same ploidy as experimental samples. This is because the 2C peak of tuber samples may be small and difficult to identify)

1. **Preparations (The solutions listed here may be prepared ahead of time and stored at 4°C but others must be made fresh on the day of use)**

1.1) Make an appropriate volume of Plasmolysis Incubation Solution (PIS) (15 ml/sample): 0.55 M mannitol, 2 mM $CaCl_2$, 1 mM KH_2PO_4, 1 mM $MgCL_2$, 50 mM Tris buffer pH 7.5 (adjusted with HCl)

1.2) Make an appropriate volume of Plasmolysis Wash Solution (PWS) (15 ml/sample) which is identical to PIS except for 0.71 M Mannitol.

1.3) Make 250 mL of modified Galbraith's Flow Cytometry Buffer (Galbraith et al. 1983)(FCB): 13.6 mM sodium citrate (trisodium), 8 mM MOPS, 18 mM $MgCl_2$, 0.4% v/v Triton X-100. Stir for at least 30 min to distribute the Triton X-100.

1.4) Filter-sterilize PIS and PW solutions with 0.45 µm aPES filter. We use vacuum driven filtration units.

1.5) Prepare 106 µm mesh filters for lysed protoplasts. We prepare 1.5 ml microcentrifuge tubes which may be directly nested into collection tubes, however any method of passing suspensions through 106 µm filter may be used.

1.5.1) If using 1.5 ml microcentrifuge tubes, cut 1 cm off the tip each microcentrifuge tube. Using a hot plate or other heat source heat a 1 cm^3 piece 106 µm steel mesh on a piece of aluminum foil. Press the cut end of the tube into the mesh until they have fused.

2. Day 1: Plasmolyze tuber tissue. Control leaf samples can be included here to produce protoplasts or at step 5.2 if crude preparations are preferred. Note: This must be performed in an aseptic environment such as a laminar flow hood.

2.1) Surface sterilize potato tubers via immersion in 75% ethanol for at least 5 min.

2.2) Remove tubers from ethanol and air dry. This usually takes ~5 min, longer if the tuber is larger than 100g.

2.3) Prepare and label sterile 50 mL conical tubes for each sample and aliquot 15 ml filter sterilized PIS into each.

2.4) Sample approximately 1 cm^3 of the desired tissue from sterilized tubers.

2.4.1) Depending on the tissue being sampled one may use either a sterilized cork borer or knife/scalpel. For instance, if the pith is desired the cork borer may be used to take a core from the stolon to distal end of the tuber and the central fraction of the core may be sampled. However, due to the asymmetric internal morphology of potato tubers, it may be more precise to sample parenchyma or cortex by halving the tuber and excising the desired tissue. See **Figure 2.1** for a depiction of internal tuber morphology.

2.4.2) Coarsely chop tissue using a sterilized scalpel into approx. 3 mm^3 pieces and transfer tissue to conical tube containing PIS. Incubate overnight at 4 °C. This is to maximize surface area so that the PIS, and later the enzyme solution, may fully permeate the tissue resulting in more complete digestion.

3. Day 2: Generate potato protoplasts. Note: This must be performed in an aseptic environment such as a laminar flow hood.

3.1) Prepare and filter sterilize (0.45 µm aPES filter) an appropriate amount of Enzyme Solution (ES) (10 ml/sample): 0.71 M mannitol, 3 mM $CaCl_2$, 1 mM KH_2PO_4, 1 mM $MgCl_2$, 4% Onozuka R-10 cellulase, 0.8% macerozyme R-10, 1% hemicululase, 10 mM MES buffer, pH 5.8. It is important to ensure that appropriate filters are used in this step; filters with pore size smaller than 0.45 µm are probe to clogging while a low a low protein binding membrane, such as aPES, minimizes loss of enzyme during filtration.

3.2) Aspirate off PIS from samples in conical tubes using a sterile serological pipette.

3.3) Add 10 mL ES to each sample and invert 2-3 times.

3.4) Incubate samples overnight at 29 °C with 180 rpm horizontal shaking. Samples should incubate for at least 16 h.

4. Day 3: Harvest protoplasts and wash protoplasts. Note: aseptic conditions are no long necessary at this point.

4.1) Remove samples from shaker and allow them to settle for ~10 min.

4.2) Aspirate off as much of the ES solution as possible.

4.2.1) Remove as much of the ES solution as feasible with a serological pipette, taking care to avoid removing the digested tissue.

4.2.2) Using a micropipette remove any remaining ES solution, again avoiding the digested tissue.

4.3) Add 15 ml PWS to each sample and invert gently 2-3 times. Allow protoplasts to settle for 10 min.

4.4) Remove PIS with a serological pipette. Use a micropipette to remove any remaining liquid. Note: This step is absolutely critical to ensure the integrity of the sample nuclei during flow cytometry. **In troubleshooting, this step was found to be the most likely source of failure in a sample.**

5. Day 3: Prepare samples for flow cytometry. Note: The samples should be kept on ice from here on unless otherwise noted.

5.1) Prepare propidium iodide (0.4 mg propidium iodide/ml FCB) and RNase solutions (0.8 mg RNase A/ml FCB) by dissolving each in FCB. CAUTION: Propidium iodide is highly toxic. Avoid contact with skin or eyes. Wear appropriate PPE.

5.2) Add 1.5 ml of ice cold FCB to each sample. Briefly shake or vortex each sample to break up aggregated tissue. It is important to break up the clumps of tuber tissue so that the FCB may fully permeate the cells and release the nuclei. Different samples may require more or less agitation depending on tissue and genotype.

5.2.1) If control flow cytometry samples were not included in the protoplast generation step they may be included here. Finely chop a small leaf ~ 3 cm^2 in 1.5 ml of ice cold FCB using a razor blade. Control samples should be the same ploidy as the experimental samples, preferably the same genotype.

5.3) Pass 1 ml of the FCB/tissue suspension through a 106 µm mesh filter. We use a 1.5 ml microcentrifuge with the tip cut off and metal mesh melted to the bottom as described in step 1.5.1. This microcentrifuge tube may be nested directly into a 2 ml microcentrifuge tube and the sample passed directly through. If using another method of filtration, the filtrate may be collected in an ice cold petri plate and then transferred to a 2 ml tube.

5.4) Add 250 µl of the RNase solution to each sample. Invert and incubate for 30 min at RT. Note: this step is intended to remove RNA from the samples, leading to less noise during flow cytometry. However as the nuclei are short-lived, researchers may decide to decrease the incubation time or perform it on ice if they are experiencing severe degradation of their samples.

5.5) Add 125 µl of the propidium iodide solution to each sample. Invert and incubate on ice for 30 min. **Samples should be used for flow cytometry as soon as possible as degradation is apparent within 2 h, even on ice.**

6 Day 3: Flow Cytometry of potato tuber nuclei.

6.1) Create two dot plots using logarithmic scale of forward scatter vs. side scatter and propidium iodide (PI) vs side scatter. Also create a histogram with PI on the x-axis. Logarithmic scale is required to ensure all events are within scale as nuclei may differ vastly in fluorescence.

6.2) Load a known control sample tube and adjust the voltage so that all events are on scale. We use leaf tissue from a genotype being used. If samples of different ploidies are to be run, a control sample for each should be included.

6.2.1) Make note of the channel of the 2C peak in the control sample. The 2C peaks of the experimental samples should fall in the same location.

6.3) Load an experimental (tuber) sample and again ensure that all events are on scale. If adjustments are required repeat step 6.2 to identify the channel of 2C peaks.

6.4) Manually gate the protoplast nuclei using the side scatter vs PI plot.

6.5) Set the PI histogram to only show the gated protoplast nuclei region.

6.6) Collect the desired number of events from each sample. Frequently researchers use 10,000 gated events for flow cytometry; however we use 2,000 events for tuber protoplast samples to accommodate more samples and samples with low concentrations.

6.7) Calculate each sample's endoreduplication index from the PI histograms using the following formula:

$$\frac{4C + (2 * 8C) + (3 * 16C) + (4 * 32C) + (5 * 64C) + (6 * 128C) \ldots}{100}$$

where 4C is the percentage of nuclei which are 4C (representing a single round of endoreduplication), 8C is the percentage which are 8C, and so on. See **Figure 2.4** for examples of histograms and C-values of peaks.

REPRESENTATIVE RESULTS

Production of protoplasts

The generation of protoplasts is necessary to achieve repeatable flow cytometry results from potato tubers, the general morphology of which is displayed in **Figure 2.1**. Researchers may wish to ensure high quality protoplasts have been produced prior to the addition of the FCP, especially in the event that troubleshooting is required. The majority of the protoplasts should be spherical with minimal protrusions (**Figure 2.2**) and may differ greatly in size, which is perhaps reflective of differences in EI.

Evaluation of flow cytometry results

The success or failure of an experiment may be gauged from the width of peaks and their separation in the flow cytometry histograms. As measuring endoreduplication requires calculating relative abundance of nuclei in each peak, mere presence or absence of peaks is insufficient if there is too much noise to draw meaningful boundaries between them. **Figure 2.3**

displays the variation in results researchers may encounter in both the flow cytometry scatter plots and histograms. Potential causes for failed samples are presented within the discussion.

Endoreduplication differences between tissues

Previously we reported that tuber pith tissue has significantly greater levels of endoreduplication than cortex tissue. To confirm and expand on this we looked at endoreduplication levels of three different tissues in cv. Superior: pith, perimedullary parenchyma, and cortex, replicated in triplicate. Our results confirmed our earlier observation that pith tissue has significantly greater EI than cortical tissue with means of 1.79 and 1.12 endocycles per cell, respectively ($p= 0.018$). Somewhat surprisingly, the parenchyma tissue demonstrated a profile similar to cortex and was also significantly different from pith tissue (mean= 1.14; $p= 0.013$). These results are summarized in **Figure 2.4**.

Influence of tuber size and ploidy

We previously reported that, for a given genotype, tubers of different size did not display a corresponding difference in EI. We aimed to confirm this result as well as evaluate the relationship between ploidy and endoreduplication. For this experiment, we used three replicates of parenchyma tissue from three different genotypes: cv. Superior (4x), VT_SUP_19 (2x) which is a dihaploid extracted from cv. Superior by prickle pollination(Uijtewaal et al. 1987), and VT_SUP_19 4x which is a doubled dihaploid(Karp et al. 1984) isogenic to Sup19. We included a set of replicates for large (90-130 g) and small (<35 g) tubers for cv. Superior while tubers for the other two genotypes were 90-130 g. We observed a significant difference between VT_SUP_19 and its progenitor Superior ($p = 0.04$); however there was no significant difference

between VT_SUP_19 and VT_SUP_19 4x (p =0.69). This indicates that while there is a likely genetic component to endoreduplication, as unmasked by the genomic reduction, it is not dictated by ploidy, at least in this background. Lastly we once again observed no significant differences between large and small cv. Superior tubers as demonstrated in **Figure 2.5**.

DISCUSSION

The protocol presented herein provides researchers with a means to assess endoreduplication within potato tubers, whose modified cellular content and increased cell size seemingly preclude other flow cytometry preparations. The protocol relies upon protoplast generation as a means to reduce noise and debris while maintaining nuclear integrity. Previously, researchers have described similar preparations for particularly recalcitrant flow cytometry samples as well as utilized tuber protoplasts to study a variety of topics such pathogenesis(Doke 1983; Doke and Tomiyama 1980b). However, to our knowledge, none have combined the use of such tuber protoplasts with flow cytometry for the purpose of studying endoreduplication. Here we discuss the shortcomings of the protocol, potential pitfalls in its execution and sample preparation, and results of a typical experiment which employs it.

Drawbacks and Pitfalls

Despite the utility and repeatability of the tuber flow cytometry protocol, it does have a few weaknesses which should be discussed. To begin, the protocol is time intensive, requiring two overnight incubations. In our observation, it is possible to dispense with the first incubation (the plasmolysis step) and still extract protoplasts; however their integrity and abundance will suffer, which negatively impacts results. Furthermore the protocol requires some expensive reagents,

particularly the cellulase and macerozyme. Researchers intending to apply the protocol to many samples may consider modifying the enzyme concentrations and/or duration of the digestion step as different tissues and genotypes may vary in responsiveness. This includes the possibility of filtering and recycling the ES which has previously been demonstrated but not employed here(Saxena and King 1985). Additionally, the preparations are highly time-sensitive, degrading within a few hours, which may limit throughput within a single day, especially if many events are desired. Some flow cytometry approaches involve the use of fixatives (*e.g.,* formaldehyde) to preserve sample integrity(Sgorbati et al. 1986). This may be a useful means to both prevent deterioration and allow for sample storage rather than protoplast exaction and flow cytometry occurring on the same day as presented. Lastly, the protocol, while reliable, is sensitive to errors within sample preparation all of which seem to result in damage to the nuclei.

In our observation, the single most critical step is 4.4, the removal of all plasmolysis wash solution before the addition of the flow cytometry buffer (FCB). If even a small volume (< 10 µl) remains, the samples are much more likely to be degraded which can lead to a poor or even failed sample. This is likely due to impurities within the enzyme solution *(e.g.* nucleases, proteases)(Bhojwani and Dantu 2013; Flores et al. 1981) which quickly degrade the nuclei during the incubation periods and time between sample runs, even at low concentrations. Another important consideration is the duration of the PI and RNase incubation steps. As noted in the protocol, the samples do not remain stable for more than a few hours after addition of the FCB so researchers may decide to reduce the duration of these steps to accommodate more samples or lengthy flow cytometry runs resulting from low nuclei concentrations. This also requires the researcher to consider the tradeoff between number of events per sample and total number of samples to be run or consider the use of fixatives as previously mentioned. The last

potential pitfall is microbial contamination during the plasmolysis and protoplast generation steps (1-3.4). Sample contamination, while not always precluding the success of a sample, seems to dramatically decrease quality of obtained histograms, again likely due to damage to the nuclei.

Differences between Tissues and Genotypes

To provide results representative of the protocol we designed two simple experiments to confirm previously reported variation by tissue and examine the influence of ploidy and genotype. The results of the tissue experiment demonstrate that the protocol yields reproducible results, as pith tissue was once again found to have the highest EI. Somewhat surprisingly parenchyma tissue, which had not previously been evaluated, had an EI value similar to cortical tissue and significantly lower than pith. This was unanticipated as parenchyma cells are typically larger than either pith or cortical cells, at least at maturity. One possible explanation is that the tubers (90-130 g) were too immature for the parenchyma cells, which comprise the majority of tuber volume at maturity, to have fully expanded and reached their maximum C-values(Reeve et al. 1973). This may also explain why the dihaploid (VT_SUP_19) and the doubled dihaploid (VT_SUP_19 4x) demonstrated significantly greater EI than cv. Superior tubers; while tubers of approximately equal size were sampled from each genotype, the maximum size of tubers from either VT_SUP_19 or the isogenic tetraploid is much smaller than that of the progenitor. Thus, it may be that they displayed greater EI simply because they were larger relative to their maximum attainable size. Alternatively, the difference may be a consequence of the genetic complement VT_SUP_19 received from its tetraploid progenitor. Another possibility is that the genomic reduction that occurred on extraction of the dihaploid from the tetraploid may have unmasked deleterious alleles resulted in plant-wide stress, which has also been shown to contribute to

elevated EI(Scholes and Paige 2015). Nevertheless this demonstrates the careful consideration researchers must employ when selecting tubers and tissues to be used for comparison of EI values, especially between genotypes.

In conclusion, the protocol described in this article provides researchers with an important tool for understanding endoreduplication in potato. It may allow for studies into the genetic and environmental components of endoreduplication, the time-course of development across tuber tissues, and assessment of natural variation. Ultimately, endoreduplication may make a promising target for potato improvement, an undertaking which will require a reliable means of assessment.

DISCLOSURES

The authors declare that they have no competing financial interests.

ACKNOWLEDGMENTS

This research was funded by National Science Foundation Award Number 1237969, “Unraveling the Heterozygosity, Allelic Composition, and Copy Number Variation of Potato” and USDA Special Grant 2014-34141-22266 (University of Maine) to RV.

FIGURES

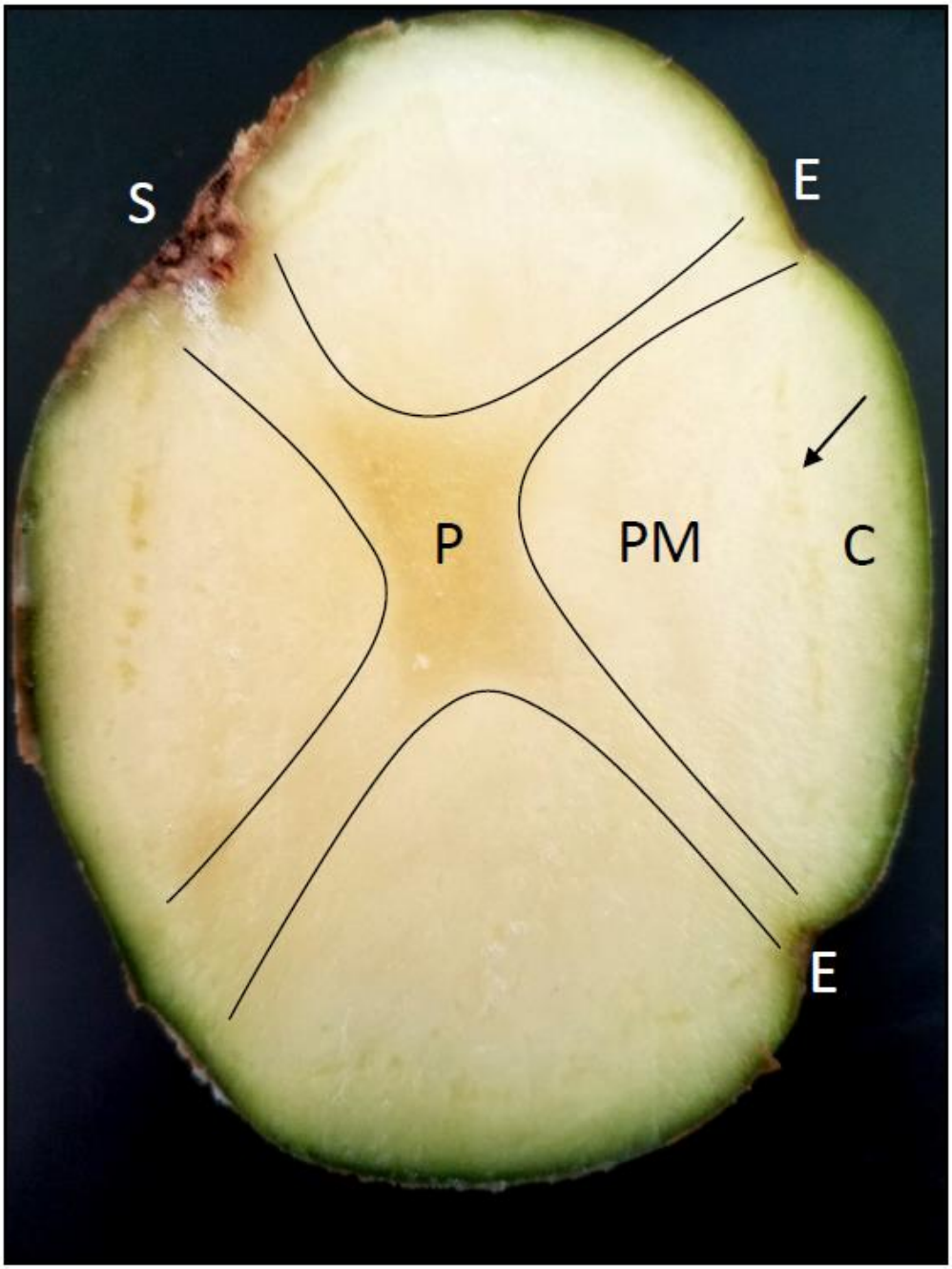


Figure 2.6: Internal morphology of a potato tuber. The separation between pith (P) and perimedullary parenchyma (PM) is illustrated with black lines. The vascular ring, which separates parenchyma from the cortex (C) is denoted with a black arrow. The stolon end (S) and tuber eyes (E) are labeled as well.

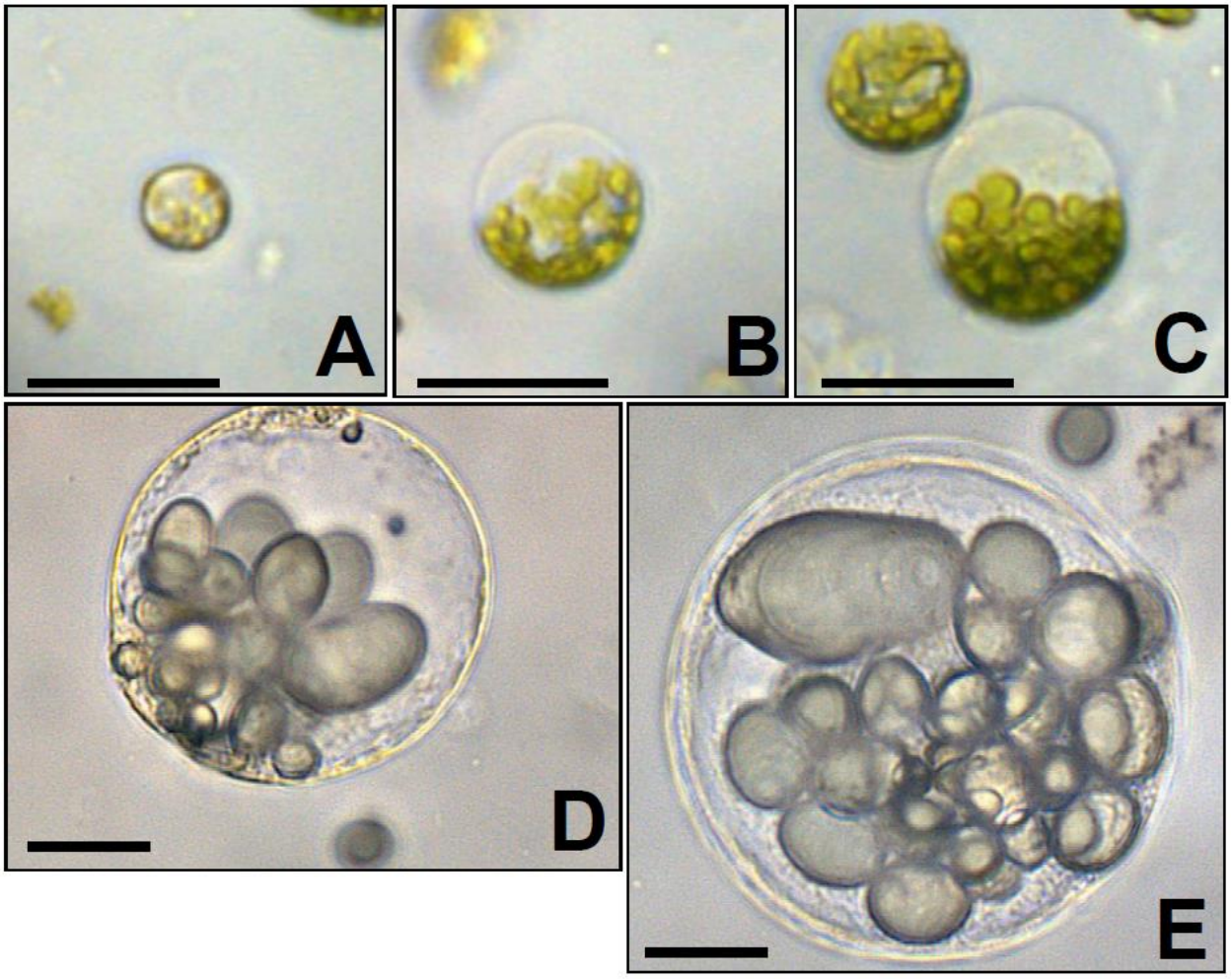


Figure 2.7: Representative leaf and tuber protoplasts acquired at step 4.4 of the protocol. Isolated protoplasts should be spherical and symmetrical with plasma membrane intact. Note the size difference between leaf (A-C) and tuber (D, E) protoplasts as well as the differences in size within a tissue which may indicate different levels of endoreduplication. Leaf protoplasts contain chloroplasts whereas amyloplasts are visible within the tuber protoplasts. Scale bars = 1000 µm.

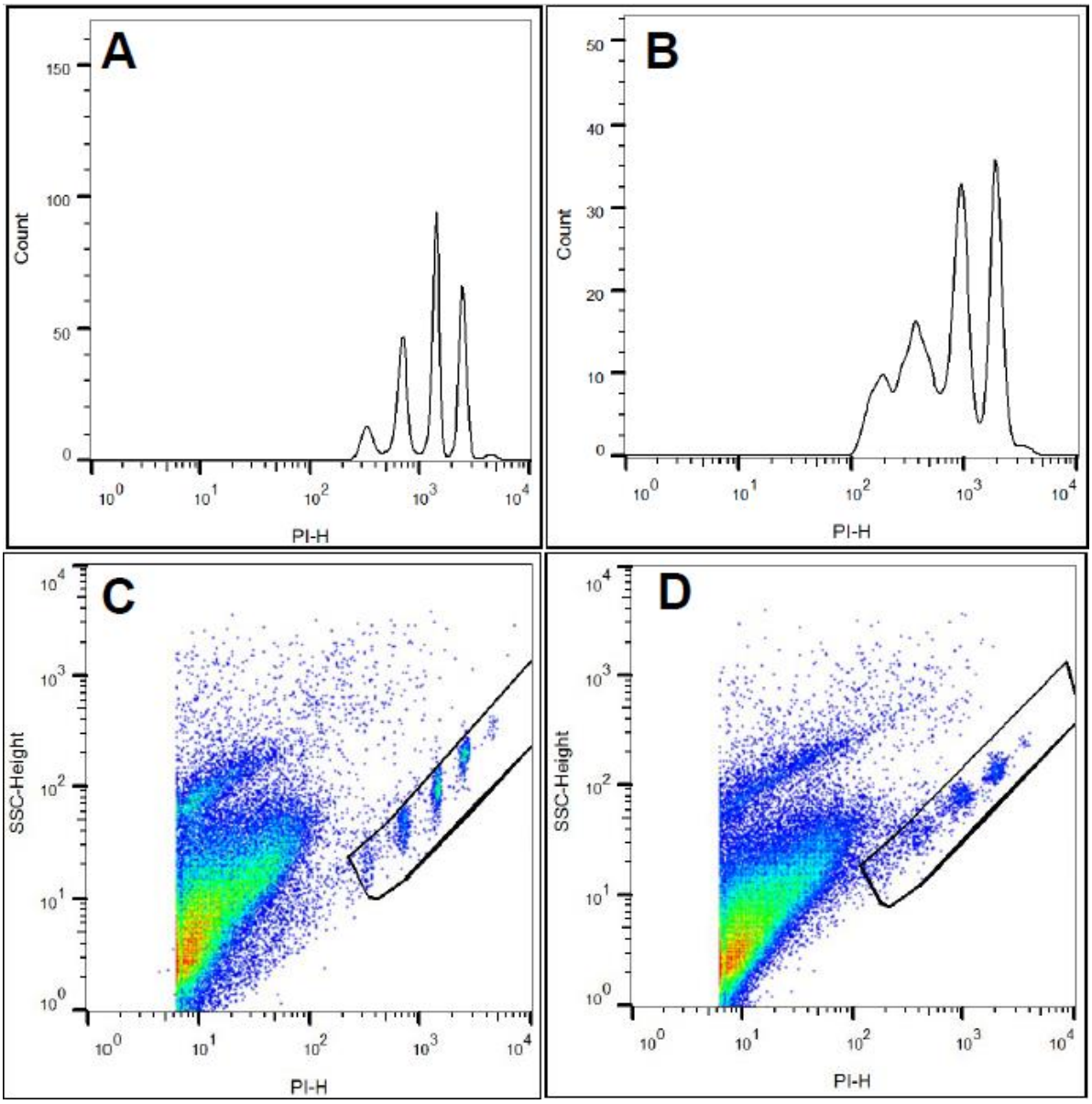


Figure 2.8: Representative results obtained from flow cytometry of protoplast nuclei of intact and degraded pith samples. Intact nuclei (A) show clean separation between peaks to quantify relative abundance of cells in each using the appropriate software whereas peaks in degraded samples (B) show shifted, wide and overlapping bases. In the scatterplots of intact (C) and degraded (D) samples, black boxes indicate nuclear event gating for histograms and the clustering of events is reflected in the width of the histogram peaks. Events outside the gates indicate debris, severely degraded nuclei, or other aggregates. PI-H corresponds to intensity of fluorescence which is measured in arbitrary units. Troubleshooting methods to prevent sample degradation are discussed within the text

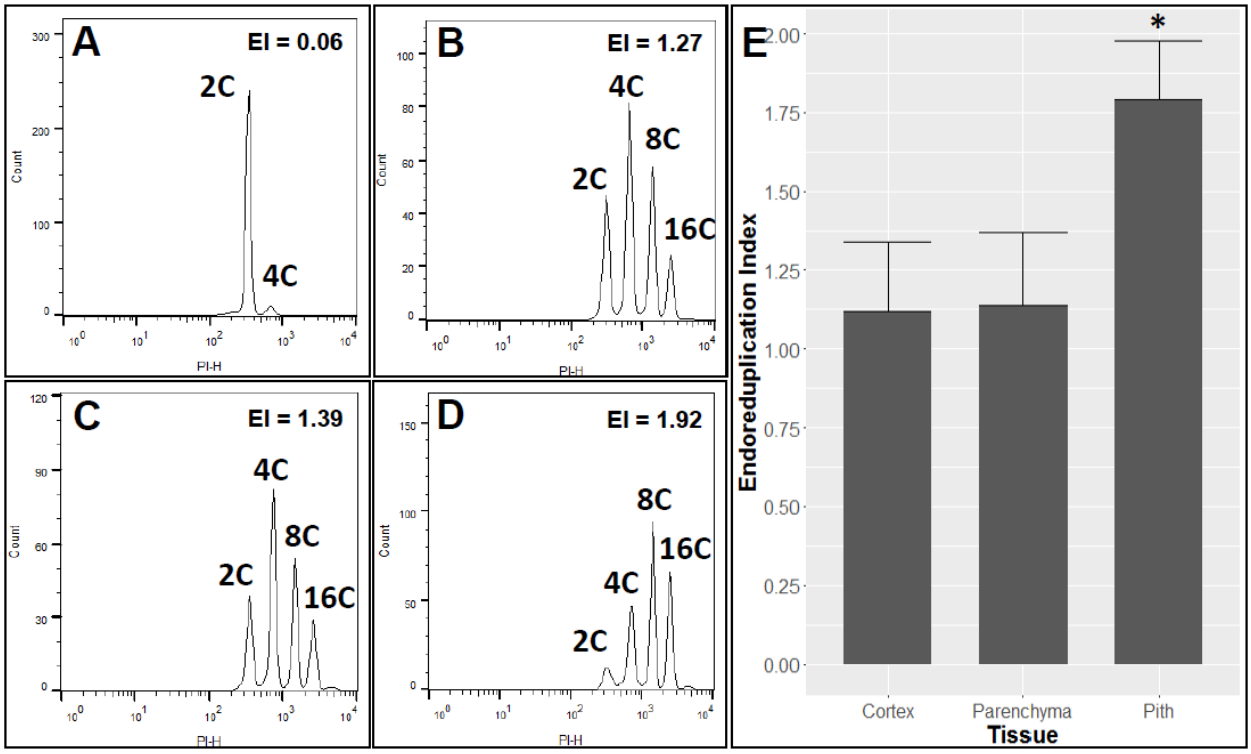


Figure 2.9: Endoreduplication in three tissues of cv. Superior**.** Histograms of leaf (A), tuber cortex (B), parenchyma(C), and pith (D) tissues are shown along with the EI value calculated from those histograms. Differences in relative abundance of nuclei comprising each peak are readily apparent, especially between leaf and tuber tissues. C-values for each peak are also displayed. PI-H corresponds to intensity of fluorescence which is measured in arbitrary units. Mean comparisons of EI of tuber tissues are displayed in E where asterisk indicates significant difference ($p < 0.05$).

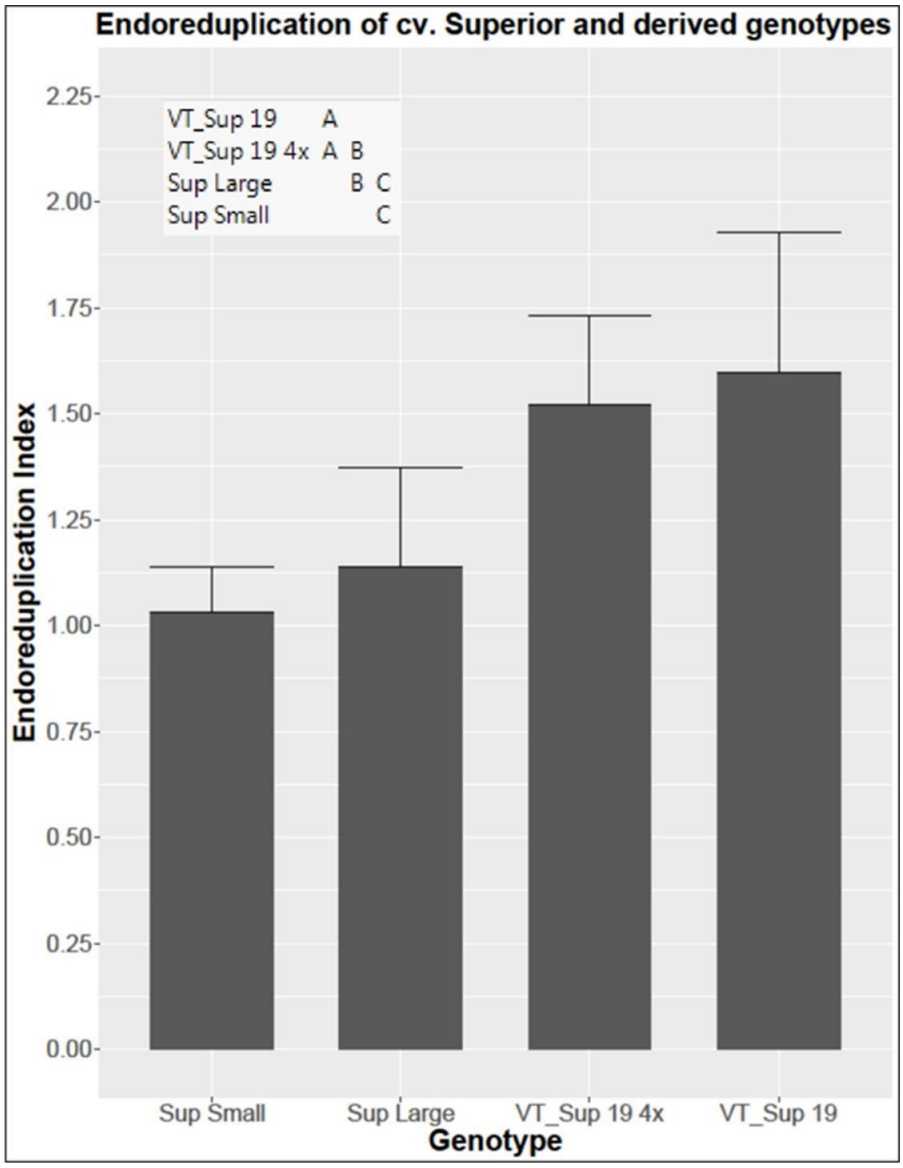


Figure 2.10: Endoreduplication in two sizes of tubers of cv. Superior and its diploid (VT_Sup_19) and tetraploid (VT_Sup_19 4x) derivatives. The barchart shows the mean for small (< 35 g) and large (90-130 g) cv. Superior tubers as well as large (90 – 130 g) tubers of the dihaploid VT_SUP_19 and the isogenic doubled dihaploid VT_SUP_19 4x. Significant differences ($p < 0.05$) are indicated by the connecting letter report.

Works Cited

1 Wilson, E. *The karyoplasmic ratio*. 3rd edn, Vol. 727 (The Macmillan Company, 1925).

2 Szymanski, D. B. & Marks, M. D. GLABROUS1 overexpression and TRIPTYCHON alter the cell cycle and trichome cell fate in Arabidopsis. *Plant Cell.* **10** (12), 2047-2062 (1998).

3 Melaragno, J. E., Mehrotra, B. & Coleman, A. W. Relationship between endopolyploidy and cell size in epidermal tissue of Arabidopsis. *Plant Cell.* **5** (11), 1661-1668, doi:10.1105/tpc.5.11.1661, (1993).

4 Grafi, G. & Larkins, B. A. Endoreduplication in maize endosperm: involvement of M phase--promoting factor inhibition and induction of S phase--related kinases. *Science.* **269** (5228), 1262-1264, doi:10.1126/science.269.5228.1262, (1995).

5 Schweizer, L., Yerk-Davis, G., Phillips, R., Srienc, F. & Jones, R. Dynamics of maize endosperm development and DNA endoreduplication. *Proc Natl Acad Sci.* **92** (15), 7070-7074 (1995).

6 Chen, C. T. & Setter, T. L. Response of potato tuber cell division and growth to shade and elevated CO^2. *Ann Bot.* **91** (3), 373-381, doi:10.1093/aob/mcg031, (2003).

7 Chen, C. T. & Setter, T. L. Response of potato dry matter assimilation and partitioning to elevated CO^2 at various stages of tuber initiation and growth. *Environ Exp Bot.* **80** 27 (2012).

8 Cheniclet, C. *et al.* Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. *Plant Physiol.* **139** (4), 1984-1994, doi:10.1104/pp.105.068767, (2005).

9 Pirrello, J. *et al.* How fruit developmental biology makes use of flow cytometry approaches. *Cytometry Part A.* **85** (2), 115-125, doi:10.1002/cyto.a.22417, (2014).

10 Chevalier, C. *et al.* Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. *J Exp Bot.* **65** (10), 2731-2746, doi:10.1093/jxb/ert366, (2014).

11 Galbraith, D. W. *et al.* Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science.* **220** (4601), 1049-1051, doi:10.1126/science.220.4601.1049, (1983).

12 Doležel, J., Greilhuber, J. & Suda, J. in *Flow Cytometry with Plant Cells* Ch. Flow Cytometry with Plants: An Overview, 41-65 (Wiley-VCH Verlag GmbH & Co. KGaA, 2007).

13 Doke, N. & Tomiyama, K. Effect of hyphal wall components from Phytophthora infestans on protoplasts of potato tuber tissues. *Phys Plant Path.* **16** (2), 169-176, doi:http://dx.doi.org/10.1016/0048-4059(80)90031-4, (1980).

14 Davis, D. A. & Currier, W. W. The effect of the phytoalexin elicitors, arachidonic and eicosapentaenoic acids, and other unsaturated fatty acids on potato tuber protoplasts. *Physiol Mol Plant Pathol.* **28** (3), 431-441, doi:http://dx.doi.org/10.1016/S0048-4059(86)80085-6, (1986).

15 Laimbeer, F. P. E. *et al.* Protoplast isolation prior to flow cytometry reveals clear patterns of endoreduplication in potato tubers, related species, and some starchy root crops. *Plant Methods.* **13** (1), 27, doi:10.1186/s13007-017-0177-3, (2017).

16 Uijtewaal, B. A., Huigen, D. J. & Hermsen, J. G. Production of potato monohaploids (2n=x=12) through prickle pollination. *Theor Appl Genet.* **73** (5), 751-758, doi:10.1007/bf00260786, (1987).

17 Karp, A., Risiott, R., Jones, M. G. K. & Bright, S. W. J. Chromosome doubling in monohaploid and dihaploid potatoes by regeneration from cultured leaf explants. *Plant Cell Tiss Org Cult.* **3** (4), 363-373, doi:10.1007/bf00043089, (1984).

18 Doke, N. Generation of superoxide anion by potato tuber protoplasts during the hypersensitive response to hyphal wall components of Phytophthora infestans and specific inhibition of the reaction by suppressors of hypersensitivity. *Phys Plant Path.* **23** (3), 359-367, doi:http://dx.doi.org/10.1016/0048-4059(83)90020-6, (1983).

19 Doke, N. & Tomiyama, K. Suppression of the hypersensitive response of potato tuber protoplasts to hyphal wall components by water soluble glucans isolated from Phytophthora infestans. *Phys Plant Path.* **16** (2), 177-186, doi:http://dx.doi.org/10.1016/0048-4059(80)90032-6, (1980).

20 Saxena, P. K. & King, J. Reuse of enzymes for isolation of protoplasts. *Plant Cell Rep.* **4** (6), 319-320, doi:10.1007/bf00269888, (1985).

21 Sgorbati, S., Levi, M., Sparvoli, E., Trezzi, F. & Lucchini, G. Cytometry and flow cytometry of 4′,6-diamidino-2-phenylindole (DAPI)-stained suspensions of nuclei released from fresh and fixed tissues of plants. *Physiol Plant.* **68** (3), 471-476, doi:10.1111/j.1399-3054.1986.tb03384.x, (1986).

22 Bhojwani, S. S. & Dantu, P. K. in *Plant Tissue Culture: An Introductory Text* Ch. Tissue and cell culture, 39-50 (Springer, 2013).

23 Flores, H. E., Kaur-Sawhney, R. & Galston, A. W. in *Advances in Cell Culture* Vol. Volume 1 (ed Maramorosch Karl) Ch. Protoplasts as Vehicles for Plant Propagation and Improvement, 241-279 (Elsevier, 1981).

24 Reeve, R. M., Timm, H. & Weaver, M. L. Parenchyma cell growth in potato tubers I. Different tuber regions. *Am Potato J.* **50** (2), 49-57, doi:10.1007/bf02855368, (1973).

25 Scholes, D. R. & Paige, K. N. Plasticity in ploidy: a generalized response to stress. *Trends Plant Sci.* **20** (3), 165-175, doi:10.1016/j.tplants.2014.11.007, (2015).

Chapter 3: Rampant Miniature Inverted-repeat Transposable Element (MITE) Activity is Associated with Variation in Gene Expression in Potato

Abstract

Miniature inverted-repeat transposable elements (MITEs) are non-autonomous, yet have inundated the genomes of many higher plants, including potato (*Solanum tuberosum*). Using a panel of 12 resequenced monoploid (2*n*=1*x*=12) potato accessions and representative sequences of 171 MITE families divided into five superfamilies retrieved from the P-MITE database, we identified a total of 197,116 MITE-related sequences, accounting for 39.44 Mb (4.47%) of the potato reference genome. As observed in other species, MITEs showed greater density in the euchromatic arms relative to the pericentromeric regions, an inclination for insertion into genic regions (~44% within 2 kb of an annotated gene), and a preference for upstream rather than downstream insertion. Comparison of gene expression data based on RNAseq of leaf and tuber tissues with MITE presence/absence calls from our panel demonstrated that ~2% of gene-proximate MITE insertions were associated with alterations in expression. Insertions upstream and within genes displayed a greater effect on expression than downstream insertions. Ontological analysis of MITE-proximate genes showed enrichment for plant stress response, secondary metabolic processes, and anthocyanin biosynthetic functions, while core biological functions such as ATP binding, DNA binding, and protein assembly were underrepresented. To infer the recent activity of particular families of MITEs, we compared the number of singlet insertions (insertions present in only one genotype of the panel) within each MITE family to the degree of homology of that family's insertions in the reference genome. This analysis indicated five of the 171 potato MITE families (DTM_Sot2, DTT_Sot1, DTT_Sot7, DTT_Sot9, DTA_Sot64) to be particularly active, accounting for approximately a quarter of genotype-unique insertions. This study emphasizes the degree to which these non-autonomous transposable elements have shaped the potato genomic landscape, contributed to genetic diversity, and altered gene expression.

Introduction

Transposable elements are ubiquitous in eukaryotic genomes, representing 44% and 85% of human and maize genomes, respectively (Schnable et al. 2009; International Human Genome Sequencing 2001). They are broadly classified by the molecular intermediate of transposition, where Type I transposons form an RNA intermediate prior to insertion in a 'copy and paste' process, Type II transposons instead mobilize through a DNA 'cut and paste' mechanism. TEs

are further classified into superfamilies based on similarities in structure, target site duplication (TSD) size and terminal inverted repeat (TIR) sequence (Finnegan 1989; Wicker et al. 2007). Type I elements may be present in high copy number, well into the thousands, whereas Type II transposons tend to present in much lower numbers (Lepetit et al. 2000; Feschotte et al. 2002). The singular exception is a specific class of non-autonomous DNA transposons known as Miniature Inverted Transposable Elements (MITEs). Lacking transposases of their own, MITEs are reliant on closely related autonomous elements to attain much higher copy numbers than fellow Type II elements. They are characterized by their diminutive size (typically <500 bp), absence of an ORF, hairpin secondary structure, proclivity for gene-proximate insertions, and similarity of TSD size and TIR sequence with the autonomous DNA transposon superfamily from which they are thought to be derived **(Figure 3.1)** (Yang et al. 2001; Yang and Hall 2003). Classification of MITEs is informed by dependence on autonomous Type II elements, wherein they are grouped based on progenitor superfamily (e.g., Mutator, hAT, etc) each of which has been assigned a three letter reference code (Wicker et al. 2007)

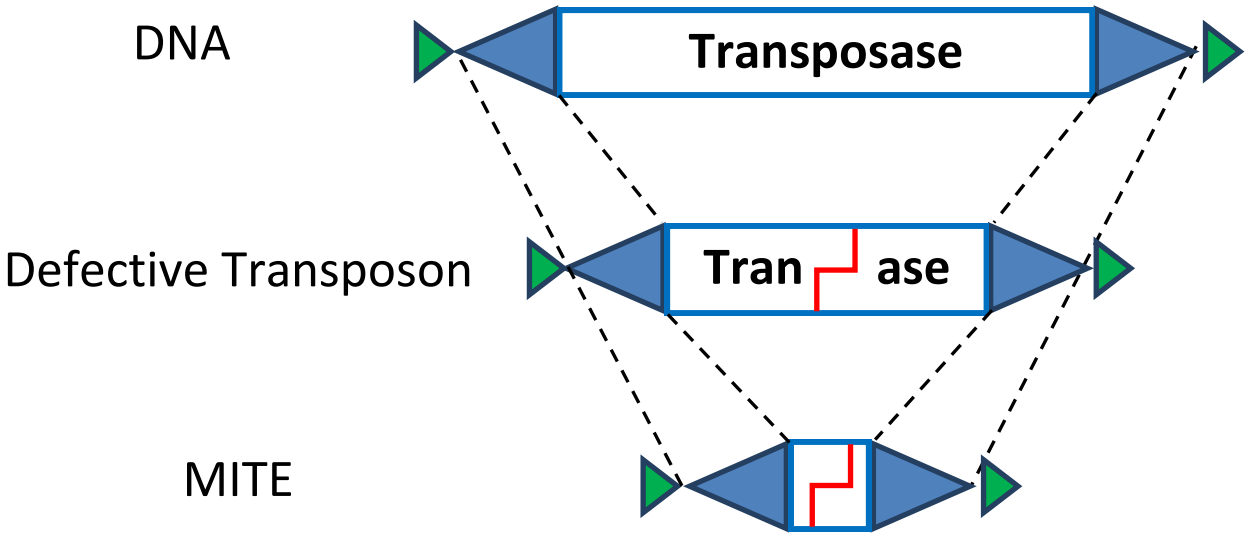


Figure 3.11: Schematic of MITE genesis. Green triangles represent target site duplications (TSD) while blue triangles correspond to terminal inverted repeats (TIRs). Figure adapted from Wessler (2006). Copyright (2006) National Academy of Sciences.

While sequence analysis based upon conservation of motifs may indicate the presence of MITE elements, it is not always clear whether a MITE is active (*i.e.*, still capable of transposing). The first described case of active MITE transposition was that of *mPing*, a MITE derived from the *Ping* element in the PIF/Harbinger superfamily, in rice cell culture (Jiang et al. 2003). *mPing* activity has since been observed in response various stress conditions including anther culture (Kikuchi et al. 2003), seed irradiation (Nakazaki et al. 2003), low temperature and hydrostatic pressure (Lin et al. 2006). In IM294, a radiogenic mutant of rice cultivar Gimbozu, *mPing* insertions were observed to create a recessive mutant allele of the *Rurm1* gene leading to the slender glume phenotype. However, fixation of this phenotype was precluded by repeated excision of the *mPing* element, often resulting in reversions to wildtype (Nakazaki et al. 2003). MITE insertions have since been implicated in variations in gene expression and plant mutant phenotypes. Naito et al. (2009) demonstrated that *mPing* insertions caused nearby genes to be stress inducible while other MITE insertions have been shown to influence drought tolerance in maize and yield in rice (Naito et al. 2009; Mao et al. 2015; Shen et al. 2017).

While rice and maize are the first and third most economically important food crops, respectively, and have received considerable investigation into how MITEs have shaped their genomic landscapes, potato (*Solanum tuberosum*), the fourth most important, has not. This may be due to the complexities of potato genetics; autopolyploidy, high heterozygosity, and widespread copy number variation may confound the investigation of genetic elements which are already present in high copy number. To the authors' knowledge, only a single investigation has been mounted focusing on the influence of MITEs in potato genetic regulation. Momose et al. (2010) observed active transposition and excision of a family of *Stowaway* MITEs into the coding region of flavonoid 3'5' hydoxylase, an anthocyanin biosynthetic gene, altering expression and, consequently, tuber color. Others, such as Chen et al. (2014) who developed P-MITE, a database of plant MITE sequences, have taken a broader descriptive approach which included, but did not focus on, potato.

According to Chen et al. (2014), the potato genome contains 171 MITE families distributed across five superfamilies, each with a three-letter designation: *CACTA* (DTC) , *hAT* (DTA), *PIF/Harbinger* (DTH), *Tc1/Mariner* (DTT), and *Mutator* (DTM). With the representative sequences of each family available from the P-MITE database, we set out to investigate the prevalence, distribution, genetic consequences, and activity of MITEs by leveraging previously published whole genome and transcriptome sequencing of a homozygous potato panel. We began by identifying all reference (those present in DM, the reference genome) and non-reference MITEs (those absent from DM) present in one or more of the panel genotypes.

Methods

MITE identification

Representative MITE sequences were retrieved from the P-MITE database and queried against the PGSC potato genome v4.04 (http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml) using RepeatMasker to identify all reference MITE insertions (Hardigan et al. 2016a). Whole genome sequencing and RNA sequencing reads for the monoploid panel were retrieved from the NCBI sequence read archive (Bioproject PRJNA287005). The pedigrees of the monoploid panel are displayed in **Supplementary Figure 3.1**. Sequencing adapters, low quality base calls (phred <20), and short reads (<50 bp) were trimmed using Trimmomatic v0.33 (Bolger et al. 2014). MITE presence/absence calls were performed using the McClintock metapipeline modules Retroseq and TE-locate for non-reference calls and the TEMP module for reference calls (Nelson et al. 2017; Platzer et al. 2012; Keane et al. 2013; Zhuang et al. 2014). Non-fixed reference and non-reference MITE calls were recoded as binary presence/absence and the phylogeny was produced via the maximum likelihood algorithm of META-PIGA (Helaers and Milinkovitch 2010) with 1000 bootstraps before visualization with FigTree v1.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Differential Expression Analysis

RNA-seq reads of leaf and tuber (with the exception of M6 which produces no tubers) samples were mapped to the PGSC genome v4.04 using HISAT2 (Kim et al. 2015). Read counts were normalized using the DESEQ2 package in R (Love et al. 2014). Independent t-tests for differential expression analysis were performed using the SciPy module in Python while

Benjamini/Hochberg p-value correction for multiple testing was performed using the StatsModels module. Differential expression analysis was only performed for those MITE-associated genes where a minimum of three of 12 genotypic samples were present or absent for the MITE (i.e., genes with 2 present : 10 absent were not tested). In an event where a gene was associated with multiple MITEs, samples were grouped based on the MITE closest to the gene or with the largest overlap.

MITE Verification

Some of the presence/absence calls output by the McClintock meta-pipeline were verified by PCR to ensure accuracy of calls. Twelve MITEs, six reference and six non-reference, were randomly selected for PCR verification, excluding those where no suitable flanking primer regions could be found due to polymorphism or structural variation. DNA was extracted using a modified CTAB/phenol method from apical leaflets of growth-chamber-grown plants frozen in liquid nitrogen (Devi et al. 2013). PCR conditions were as follows: 35 cycles of 98 C for 15 sec; 50 C for 30 sec; 68 C for 3 min; followed by a single 5 min extension at 68 C. Each 20 µl reaction contained 1x Primestar GXL buffer, 0.5 units Primerstar GXL taq, 0.2 µM each primer, 50 µM dNTPs, 50 ng DNA. Primers are provided in **Supplementary Table 3.1.** Longer product size in genotypes where the McClintock pipeline had indicated the presence of a MITE was taken as positive evidence.

Gene Ontology

Ontological terms were assigned by a BLAST of PGSC representative genes against the TAIR database (Lamesch et al. 2011). Each PGSC gene was assigned the ontological terms associated with the top TAIR hit (e-value threshold $1e^{-10}$). These ontological terms were combined with

terms from Blast2GO output kindly provided by John Hamilton at Michigan State University. Gene ontology term enrichment for MITE-associated genes was evaluated using a Fisher's exact test with association defined as an insertion within 2 kb of a gene.

Mite Activity

To infer recent MITE activity, we compared the quantity of genotype-unique insertions of each MITE superfamily to the number of sequences with high homology from that family present in DM. Only those unique insertions identified by both non-reference calling modules (Retroseq and TE-locate) were analyzed to reduce chances that false-positive calls could bias the results. The number of sequences from each family with high homology was identified as follows: First, all sequences in the reference with a Smith-Waterman alignment score (as compared to the MITE representative sequences) greater than 1000 were extracted from the reference genome. These sequences were then sorted by MITE family after which clustering was performed using CD-HIT-EST with a minimum similarity threshold of 95% of the shorter sequence (Li and Godzik 2006). A regression analysis was then used to ascertain the relationship between the number of unique insertions and cluster size.

Results

Distribution of MITEs in DM

We began by querying the potato genome with the P-MITE database which identified a total of 197,116 MITE sequences dispersed across the potato genome, accounting for 39.44 Mb or 4.47% of the potato genome, which is comparable to the observations made on the PGSC v3.0 genome by Chen et al. (2014)(**Table 3.1).** Of the five reported MITE superfamilies present in

potato, Mutator and CACTA were most and least prevalent, respectively, accounting for 35.8% and 2.2% of the total MITE landscape. All superfamilies showed similar patterns of distribution; characterized by low incidence in the pericentrimeric heterochromatin and increasing prevalence distally (**Figure 3.2A**). Analysis of MITE insertions showed a predisposition for insertions near genes (44% within 2 kb of a gene) and a preference for upstream, rather than downstream insertions ($p = 0.008$), with a paucity of insertions located in coding regions **(Figure 3.3)**; a trend that held true for all MITE superfamilies.

Table 3.2: Comparison of Observed MITEs versus those Reported in Chen et al. (2014)

Superfamily	Number of Families	Element Number (Chen et al.)	Element Number (Observed)	Total Length (bp)(Chen et al.)	Total Length (bp) (Observed)
Mutator (DTM)	65	58,462	69,027	14,123,376	14,100,879
Tc1/Mariner (DTT)	23	45,990	50,592	8,856,608	9,449,213
hAT (DTA)	64	43,895	47,367	10,554,906	10,055,322
PIF/Harbinger (DTH)	15	17,780	23,123	4,306,090	4,950,250
CACTA (DTC)	4	4,265	4,946	811,800	882,461
Total	**171**	**170,392**	**197,116**	**38,652,780**	**39,438,125**

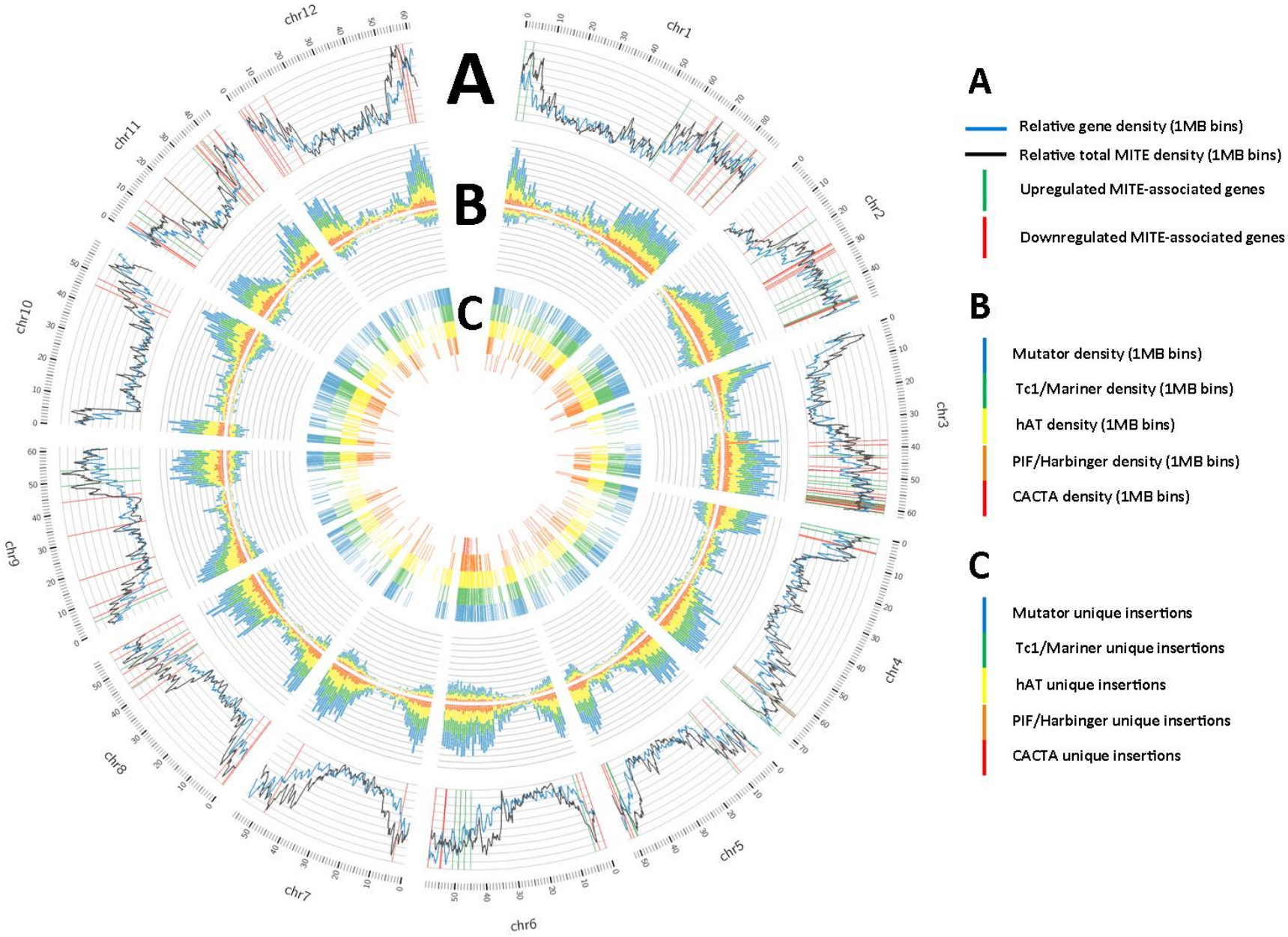


Figure 3.12: (A) Genome-wide distribution of total MITES (black) and genes (blue) in 1 MB bins with 200 kb overlaps. Differentially up-(green) and down-(red) regulated genes are highlighted. (B) Reference (outward) and non-reference (inward) MITE distribution by superfamily in 200 kb overlapping 1 MB bins. (C) Distribution of singlet insertions by superfamily shows regions of greater activity.

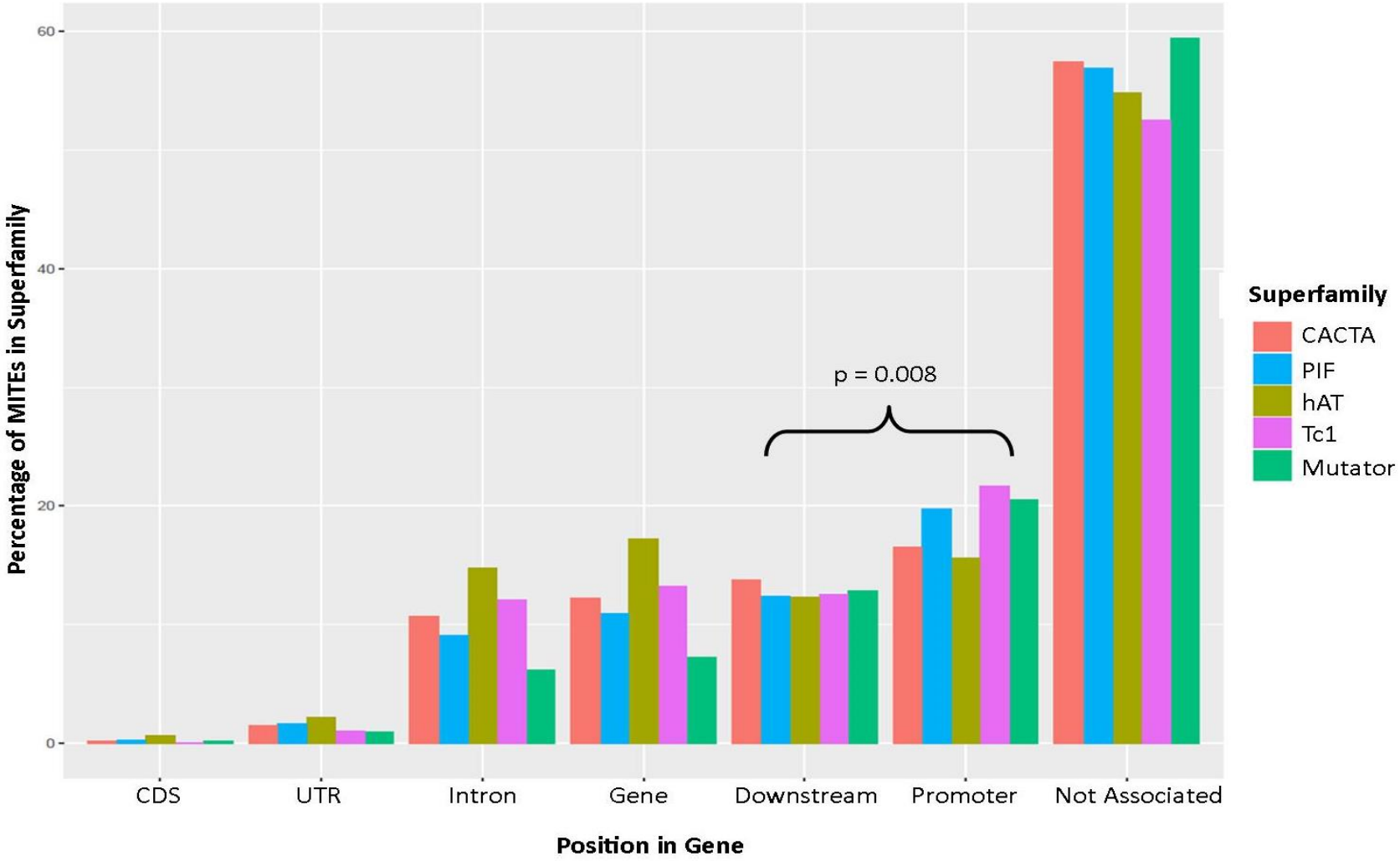


Figure 3.13: Distribution of reference MITEs relative to genes by superfamily. CDS: Coding sequence; UTR: Untranslated Region; Gene: Sum of CDS, UTR, and Intron; Downstream: within 2kb downstream of gene; Promoter: within 2kb upstream of gene; Not Associated: Not within gene or 2kb of gene.

Distribution of MITEs in Monoploid Panel

Next we evaluated the presence of reference and non-reference MITEs within the monoploid panel, whose leaves and tubers are displayed in **(Figure 3.4)**.

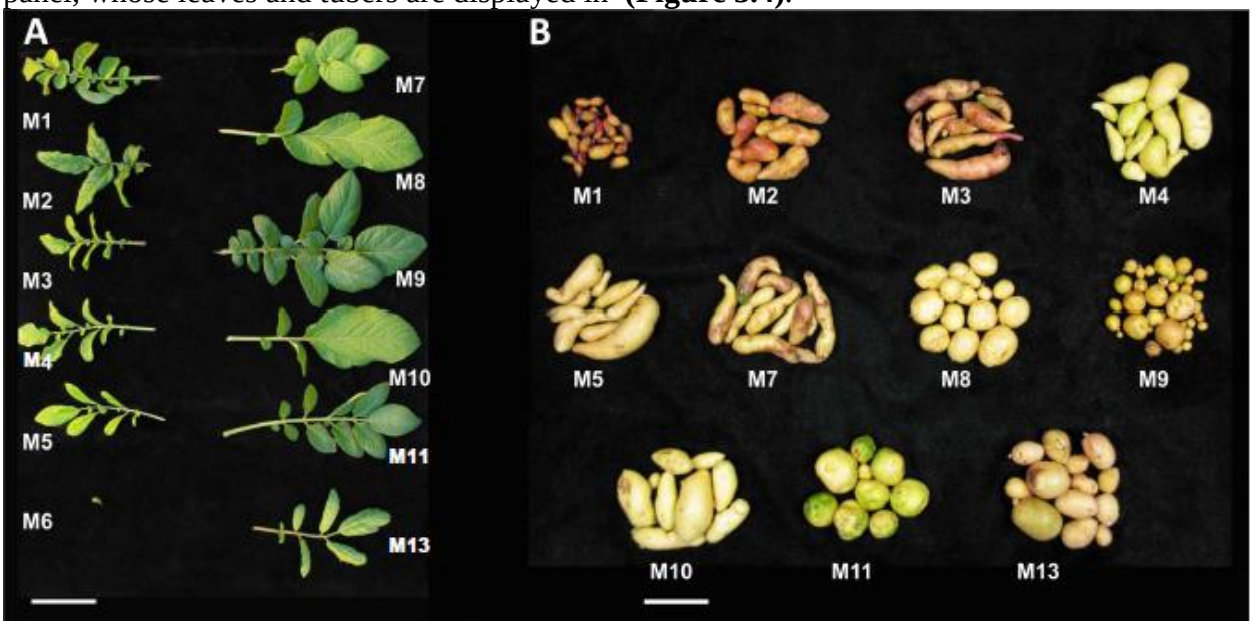


Figure 3.14: Leaves (A) and tubers (B) of the monoploid panel demonstrating variation of phenotypes. Taken from Hardigan et al. (2016), Copyright American Society of Plant Biologists . Scale bar = 10 cm.

Consistent with the expectation that the vast majority of MITE insertions are ancient, 90.1% of reference insertions were fixed in the panel (**Supplementary Figure 3.1)**. Non-reference MITEs displayed a pattern of enrichment in euchromatin arms similar to that of reference MITEs, further supporting the notion that these regions are the major targets of non-autonomous element activity **(Figure 3.2B; Supplementary Figure 3.2)**. Finally we observed a total of 4,676 genotype-unique insertions confirmed by both Retroseq and TE-locate **(Figure 3.2C)**. In order to assess the reliability of MITE calls in the monoploid panel, we created a MITE-based phylogeny and compared it to the Copy Number Variant (CNV) and SNP-based phylogenies presented in Hardigan et al. (2016). The three phylogenies were similar, with the major difference being: 1) longer relative lengths of M6, M9, and M13 branches; and 2) M1, M9, and M11 did not form a

separate clade in the MITE-based phylogeny **(Figure 3.5)**.

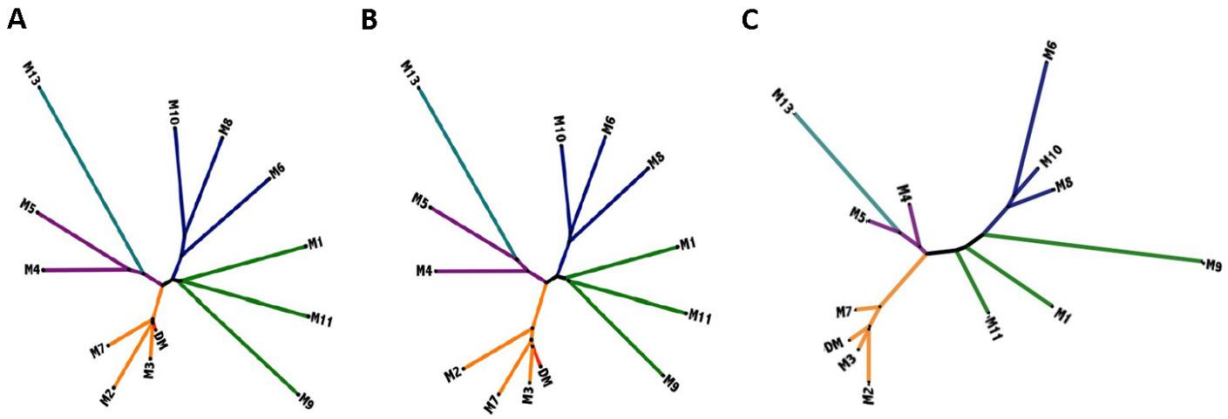


Figure 3.15: SNP (A) CNV (B) and MITE-based (C) phylogenies of the 12 genotypes of the monoploid potato panel (A and B from Hardigan et al. 2016, Copyright American Society of Plant Biologists). M numbers correspond to the similarly labeled phenotypes in Fig. 2.

PCR verification of MITE presence absence calls confirmed 96 and 100% of reference and non-reference MITEs, respectively, for the ten reactions which produced a single strong band **(Figure 3.6)**. Taken with the similarities between the MITE and SNP based phylogenies, this indicates that the transposon calls from the McClintock meta-pipeline are reliable.

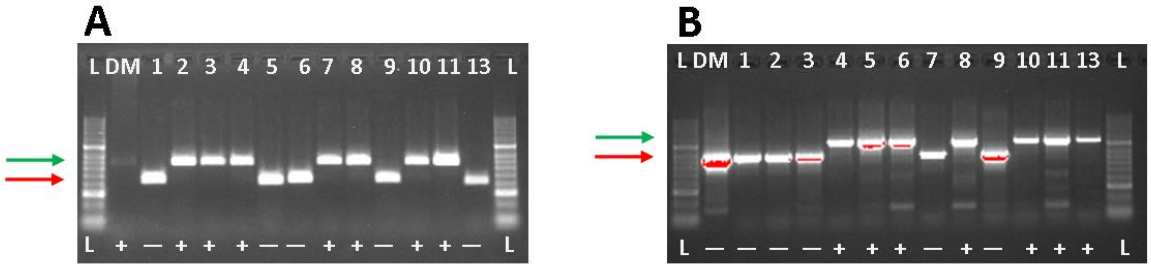


Figure 3.16: Examples of reference (A) and non-reference (B) MITE PCR-screens. Genotypes are listed along the top while calls from the McClintock pipeline are listed along the bottom. Red and green arrows show anticipated placement of MITE-absent and MITE-present bands, respectively. L : Invitrogen 1Kb + ladder.

Differential Expression

Considering that MITE insertions seem to preferentially target genic regions, we sought to evaluate how frequently nearby MITE insertions were associated with differential gene expression in leaves or tubers in the monoploid panel. To that end, we first identified all genes in the genome with nearby (<2 kb) MITE insertions polymorphic among the 12 genotypes which met criteria for minimum expression levels and samples numbers (see methods). There was a total of 2,067 genes with nearby reference insertions and 5,516 genes with nearby non-reference insertions meeting these criteria. Differential expression revealed 23 and 132 gene-tissue combinations to be differentially expressed in association with reference and non-reference MITE presence, respectively (FDR corrected p-value <0.05). Distributed across superfamilies proportionally to the number tested, 53 genes were up-regulated in association with MITE-presence, whereas 102 were down-regulated (**Table 3.2)**. A chi-squared test showed that there was no significant difference in propensity of each superfamily to up or down regulate genes.

	Mutator	**CACTA**	**PIF**	**Tc1**	**hAT**	**All**
Reference Tested	622	38	263	585	559	2067
Non-Reference Tested	2161	76	690	1213	1376	5516
Total Tested	2783	114	953	1798	1935	7583
Up-Regulated	22	0	5	13	13	53
Down-Regulated	35	2	11	29	25	102
Expect Up-Regulated	19.6	0.8	6.7	12.6	13.5	
Expect Down-Regulated	37.4	1.5	12.8	24.2	26.0	
Chi-Squared						***0.35***

Table 3.3: Distribution of differentially expressed genes across MITE families. Chi-Squared test compares observed number of differentially expressed genes versus expected based on number of tests performed.

A chi-squared analysis showed that insertions into genes (UTR, exon, or intron) were more likely to be associated with altered expression than those downstream (<2kb) **(TABLE 3.3).** Among the differentially expressed genes were genes for disease resistance, stress response, and secondary metabolic functions.

	Promoter	Internal	Downstream	All
Reference Tested	889	454	724	2067
Non-Reference Tested	1994	1900	1622	5516
Total Tested	2883	2354	2346	7583
Up-Regulated	18	25	10	53
Down-Regulated	38	38	26	102
Expect Up-Regulated	20.2	16.5	16.4	
Expect Down-Regulated	38.8	31.7	31.6	
Chi-Squared				***0.009***

Table 3.4: Distribution of differentially expressed genes across location relative to gene. Chi-Squared test compares observed number of differentially expressed genes versus expected based on number of tests performed.

Gene Ontology

In order to ascertain if certain functions predispose a gene to insertion by a MITE or certain insertions are retained through positive selection, we performed an ontological analysis, comparing the types of genes of DM where a MITE is present within or nearby (< 2 kb) with those genes where there are no proximate MITEs. Within the MITE-associated gene set, there was enrichment (Fisher's p-value < 0.05) for terms associated with plant stress response and secondary metabolism, including terms for hypersensitive response, regulation of defense response, response to fungal infection, signal transduction, and anthocyanin biosynthesis. Conversely, the MITE-associated gene set was depleted for terms corresponding to core biological functions including ATP binding and catalysis, nucleotide binding, and electron transport (**Supplementary Table 3.3**).

MITE Activity

As we observed considerable differences in MITE profiles across genotypes as well as an association with altered gene expression, we set out to infer the activity of these MITEs within potato. As MITEs often accrue mutation following insertion, a MITE with many insertions with high homology may be considered to be more active than a transposable element which is present in lower copy number or with less homology between copies. We therefore decided to compare each family's number of sequences with high homology in the reference genome to the number of individual insertions which were present in only one genotype of the panel. The latter were assumed to be more recent insertions than those which were fixed across all 12 genotypes in the panel. A regression analysis of these two proxies for MITE activity indicated a moderate positive correlation (R-squared= 0.55) and identified several outlier families which appear to have been more active in potato's recent evolutionary history (**Figure 3.7)**. To the author's knowledge, there has only been a single report observing active transposition of a MITE family within potato (Momose et al. 2010). A BLAST of the primers used to track this MITE's movement indicates it belongs to a member of the Tc1/Mariner superfamily, DTT_Sot9, which is expected to be one of most active families in potato based on this analysis.

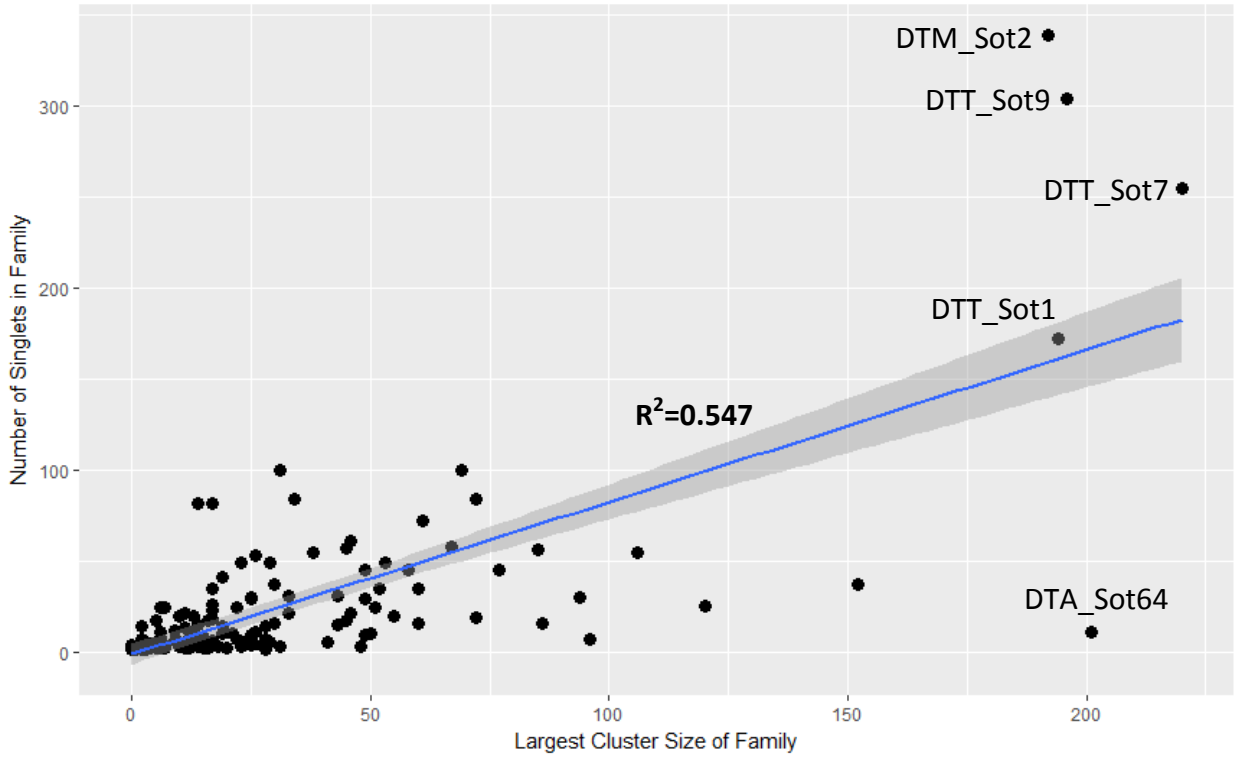


Figure 3.17: Activity of MITE families demonstrated by correlation between number of insertions in single genotypes of the monoploid panel and cluster size of the MITE family. Each point represents a single family. The points corresponding to the five most active families are labeled. Linear regression line with a 95% confidence interval and its R-squared value are displayed as well.

Discussion

MITE distribution in DM and the Panel

This study leverages the massive reduction in complexity afforded by homozygosity in both the reference genotype and a monoploid panel with the knowledge of MITE sequences made available by Chen et al. (2014) and the P-MITE database (http://pmite.hzau.edu.cn/) to reveal the scope of influence these non-autonomous elements have exerted on the potato genome. As expected, we observed a greater number of reference MITEs compared to Chen et al. (2014). This is likely due to the inclusion of superscaffolds chr00 and chrUn which were unavailable at the time of the previous study (Hardigan et al. 2016a). Furthermore as these

represent sequences which precluded assembly in previous iterations of the potato genome they are expected to be replete for low complexity sequences and repetitive elements which is reflected in a relatively consistent MITE density across the pseudomolecule rather than the heterochromatin enrichment observed in the majority of chromosome-representative pseudomolecules (**Supplementary Figure 3.2; Supplementary Figure 3.3)**. However, unexpectedly, the total length of MITE insertions only increased marginally by ~785 kb compared to that reported by Chen et al. (2014). This reflects an actual decrease in observed total length of insertions relative to Chen et al. (2014) in Mutator and hAT families. The observed decrease in length might be explained by the greater diversity of these elements, *i.e.*, larger numbers of constitutive families **(Table 3.1)**. As the MITE sequence library used to query the genome was representative rather than exhaustive it seems likely elements of these families would have been identified but the entirety of their length may have gone undetected due to less conserved homology to the representative sequences.

Nevertheless, the observed length of MITE sequences as a fraction of the total genome (4.47%), fell slightly below the 4.84% reported in Chen et al. (2014) but was still greater than related species such as tomato and grape with 3.44% and 3.02%, respectively. It is unclear why potato has a greater density of MITEs in its genome than tomato despite their recent divergence (Chen et al. 2014; Wu and Tanksley 2010). Asexual reproduction, which may decrease efficiency of purifying selection relative to sexual reproduction (Kondrashov 1982), makes a tempting scapegoat; however there are examples of sexually reproducing species such rice, 9.98%, and *Medicago truncatula*, 8.21%, which exhibit greater concentrations of MITEs in the genome than potato. As the reference genome, DM 1-3 R44/516, is a monoploid derived from

diploid Group Phureja germplasm, it is unlikely that polyploidy is responsible for the disparity in MITEs between potato and its close relatives.

In keeping with previous reports (Guo et al. 2017; Dai et al. 2015), we observed a close association between gene density and MITE density (**Figure 3.2A**) which may indicate a potentially adaptive role for MITEs by introducing regulatory motifs, MITE-derived siRNAs, alterations of nearby methylation, or contribution of partial coding sequences, thereby creating allelic diversity (Kuang et al. 2009; Oki et al. 2008; Ngezahayo et al. 2009; Castelletti et al. 2014). Reports of specific MITEs increasing expression levels under certain conditions (Yang et al. 2005; Naito et al. 2009) are seemingly contradictory with those of global reduction of expression in MITE-associated genes (Lu et al. 2012), a discrepancy which may indicate that the general distribution of MITE insertions is a consequence of greater chromatin accessibility while a small subset of elements may benefit the organism by permitting changes in transcriptional control. That is, similar to the conventional wisdom regarding genetic mutations, the majority of insertions may be neutral or detrimental while a minority may be leveraged to the organism's benefit. Our finding that MITEs are discriminatory in their insertions relative to genes is supported by previous research (Naito et al. 2009) and may provide some corroboration to this conclusion. The preferential targeting of upstream compared to downstream regions supports the notion that mite distribution patterns are in part due to differences in DNA accessibility, as upstream regions have been shown to have greater prevalence of DNase hypersensitive sites indicating a more open chromatin state (Marand et al. 2017). A consistent dearth of MITEs in CDS and UTRs across families is likely reflective of their disruptive capability while a deviation from this trend with regard to intronic insertions indicates a greater tolerance in these regions. One possible exception to this trend is the Mutator superfamily where only 6.1% of total

reference insertions fell within introns, a third lower than the next superfamily, PIF/Harbinger at 9.1% **(Figure 3.3)**. However, whether this is reflective of a diminished tolerance to intron insertions or an enhanced predilection for upstream insertions is unclear.

Investigation into the monoploid panel revealed that the vast majority of reference insertions were monomorphic, indicating ancient geneses. Likewise, the majority of non-reference insertions identified by both Retroseq and TE-locate was shared by multiple individuals (~70%), which may be expected considering the *a priori* knowledge of the relationships between panel genotypes **(Supplementary Figure 3.1)**. With respect to the five superfamilies, non-reference insertions displayed similar genomic distributions as reference, but with lower frequency despite being cumulative across 12 non-reference genotypes; providing further evidence that most MITEs in the genome are the result of ancient events. The concordance between the phylogenetic tree derived from MITE presence/absence and the SNP and CNV-based trees presented by Hardigan et al. (2016a) support the veracity of the McClintock pipeline calls in addition to the PCR verification presented in the results. The divergence that is apparent between the trees, such as increased distance of M13 from the rest of the panel, could perhaps be explained by differences in distribution of sampling; While the pericentromeric regions are largely devoid of MITEs, Hardigan et al. (2016a) reported that large CNVs were concentrated in these regions whereas SNPs selected for the phylogeny were distributed rather evenly across the genome. Therefore, a MITE-based phylogeny may be expected to overemphasize differences in the more dynamic chromosome arms while largely ignoring those features proximal to the centromere as compared to the other two methods. Sanseverino et al. (2015) observed similar differences between transposable element and SNP-based phylogenies in a resequencing study of seven melon (*Cucumis melo*) accessions. They

hypothesized that the increase in relative lengths of arms in the TE phylogeny may have been the result of greater and inconstant rates of transposition relative to point mutations, an explanation which merits consideration here as well.

Gene Ontology

An analysis of the MITE insertions within or nearby genes may elucidate those functions that are either accommodating of insertions or able to benefit from them. The enrichment of insertions in genes pertaining to defense response and secondary metabolism elicits two possible explanations; the first is that defense and stress related genes exhibit the flexibility in sequence and expression profiles. This explanation is supported by previous studies into the functional implications of MITEs. As previously mentioned, MITEs have been shown to alter regulation of stress response (Yongsheng et al. 2011). Other studies have shown that MITEs may have structural consequences such as contributions of novel exons, truncation due to ORF shifts, alteration of polyadenylation signals, and mobility of flanking sequences in coincidence with MITE transposition (Kuang et al. 2009; Oki et al. 2008; Ngezahayo et al. 2009; Castelletti et al. 2014). Since resistance genes are sometimes altered by other structural variations, such as CNV, to the plant's benefit, it may be that MITE transposition plays a role in the Red Queen Hypothesis, offering another means of genetic experimentation which occasionally gives rise to adaptive alleles (Knox et al. 2010; Van Valen 1973; Cook et al. 2012; Dixon et al. 1998). Alternatively, resistance and secondary metabolism genes are often expressed at low latent levels, with drastic increases in response to stimuli. As studies have shown a general trend of lower expression in MITE-associated genes, the observed enrichment and a consequential decrease in expression may only produce a fitness cost in the event of pathogen attack or urgent

need to produce a secondary metabolite; a pressure which may not prove consistent enough to guarantee purifying selection.

Differential Expression of Mite-associated Genes

We observed an approximate 2:1 ratio of down-regulated to up-regulated MITE-associated genes within the MITE panel, indicating that MITEs generally function to disrupt, rather than enhance nearby gene expression. Additionally, upstream and internal insertions were more likely to be associated with an expression change than downstream insertions while superfamily seemed to have a negligible effect **(Table 3.2; Table 3.3)**. This may indicate that, generally, the point of insertion is a greater determinant of effect than family of MITE. Alternatively, superfamilies may lack consistency of their consequences, with certain families having a disproportional effect on nearby genes. Furthermore, some of these differences are likely to be explained by relatedness of alleles rather than direct consequences of MITE insertions. Since insertions are the product of discrete events, haplotypes which share a MITE are likely to be more related than those that do not, with the exception of MITE excisions, and may therefore be expected to share other mutations which may impact expression. Finally, the transcriptomes examined herein were derived from plants grown under optimal conditions. As our gene ontology analysis demonstrated, stress response and metabolic genes are more likely to harbor elements; therefore, it seems likely that more signals of MITE influence could be detected in plants subjected to biotic or abiotic stress.

The use of homozygous germplasm could have altered the degree of observable MITE influence on gene expression. As these plants have advanced through the "monoploid sieve" they would no longer harbor any lethal or severely deleterious alleles. Thus, by correlating expression

levels with MITE presence in the monoploid panel we may observe their correlation without the complications of heterozygosity but lose the ability to detect those alleles which have been so severely disturbed they must be masked. We are therefore almost certainly underestimating: 1) the true amount of diversity created by non-reference MITE insertions; and 2) the degree to which those MITEs may negatively affect expression, especially with regard to indispensable genes.

Inference of Recent MITE Activity

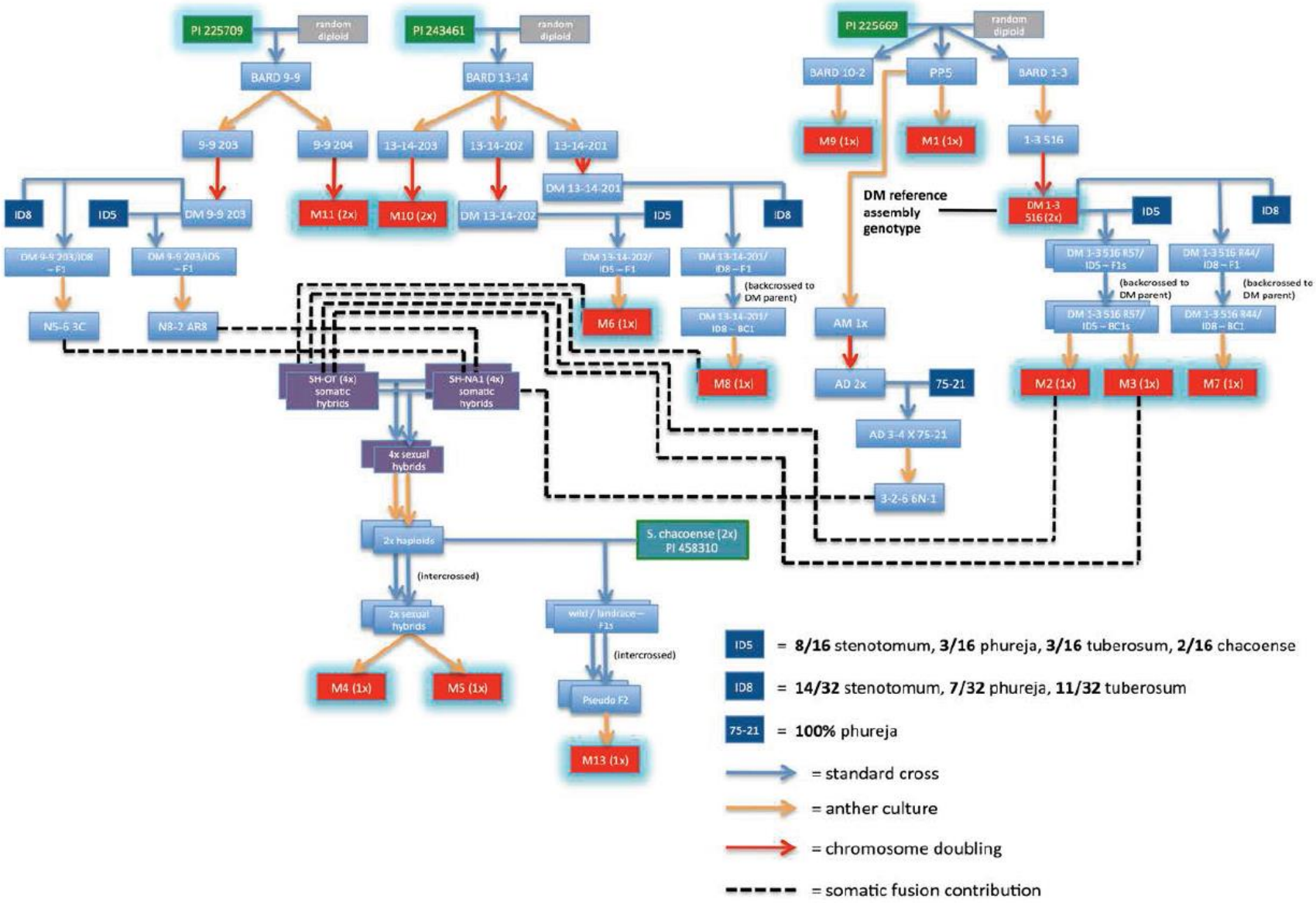
As a consequence of aforementioned observation that the majority of MITEs was present in multiple individuals of our panel, only a small fraction of MITE insertions was unique to single genotypes (**Figure 3.2C**). These 'singlet' insertions are likely to have arisen more recently due to their limited distribution within the panel, although it is also possible that some are restricted to a single genotype because they are particularly deleterious. Nevertheless, by combining frequency of a MITE family's singlet insertions with its number of highly conserved sequences in the reference, we may approach the question of recent MITE activity from two sides: scarcity and homology. As a new insertion would be identical to the sequence from which it was derived and restricted to a single genotype, concurrence between the two implies an actively transposing sequence. The regression between these metrics had a moderate R^2 value of 0.55, indicating some merit to this approach. While most MITE families appeared to have little or no recent activity, five stood out as particularly active: DTM_Sot2, DTT_Sot1, DTT_Sot7, DTT_Sot9, DTA_Sot64. These five families accounted for roughly a quarter of the 4,676 total singlet insertions observed and all but DTA_Sot64 had over 170 sequences from DM clustered with <95% homology.

There has only been a single documented instance of a MITE actively transposing in potato, where it interfered with the anthocyanin biosynthetic pathway, resulting in altered tuber color (Momose et al. 2010). A BLAST search of the primers used to track the *Stowaway*-like MITE highlighted in that study against the P-MITE database indicated that it is a member of the DTT_Sot9 family, thereby confirming that at least one of our five candidate active families is actually active and implying the others may be as well. One potential shortcoming of the homology-based approach employed here is that it only considers the size of the largest cluster of sequences. Therefore, if multiple distinct yet closely related elements fall within a family, there may not be enough homology between intra-family groups to meet the 95% similarity threshold. This seems to have been the case with DTA_Sot64 which displayed 204 singlet insertions, yet a major cluster size of only 11; on closer examination, it had nine clusters of five sequences or more, perhaps indicating the DTA_Sot64 is actually comprised of several closely-related and moderately-active elements, rather than a single highly active one.

Conclusions

MITEs have had a dynamic influence on the potato genome. Within the limited germplasm sampled (12 monoploids extracted from the adapted Group Phureja population), the landscape varied such that each genotype harbored from 153 (in a Phureja-derived individual close to DM) to 1285 (in a half *S. chacoense* accession) unique MITE insertions, reflecting differences in their genetic background and recent transpositions. Some 70% of the MITES examined appeared to be fixed in this germplasm, suggesting a more ancient origin. Under our stringent filtering conditions, it appeared that only 2% of MITE insertions near genes were associated with differences in expression, a figure that may underestimate true effects due to

homozygosity of the germplasm. Similar to other structural variants, MITEs contribute genetic variation which may fuel adaptation, preferentially targeting genic areas which suggests an outsized influence on genome evolution, especially as it relates to genes with MITE-enriched functions, such as stress response.

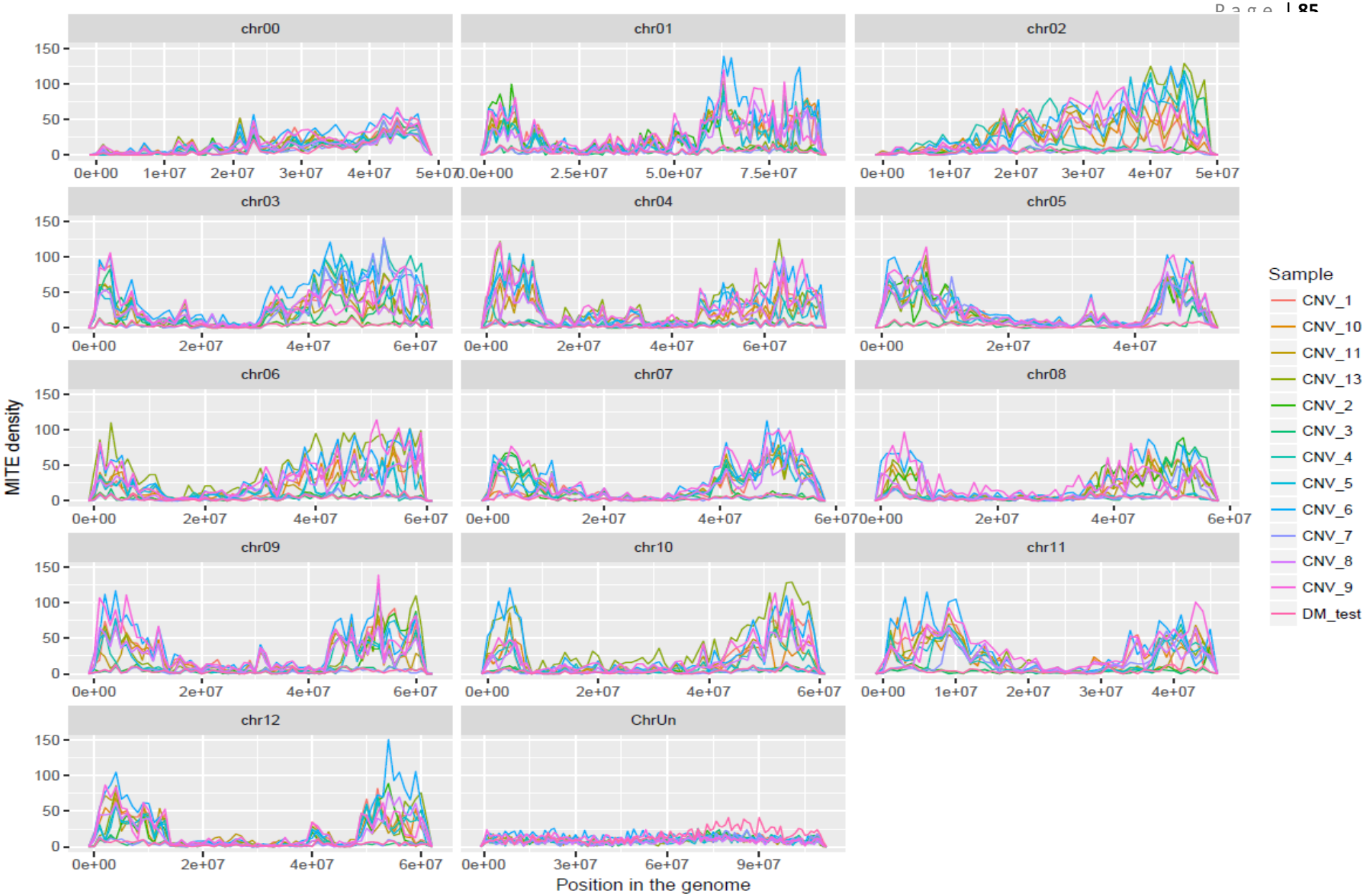


Supplementary Figure 3.2: Genetic relationships of monoploid panel. Taken from Hardigan et al. (2016). Copyright American Society of Plant Biologists.

Supplementary Table 3.1: Primer sequences used in MITE screen assay.

Gene	Chromosome	Location	MITE	PrimerL	PrimerR
PGSC0003DMT400003423	Chr02	45805185	DTM_Sot42	TCCAAGTGCTGGAATCAGTG	GGTTAGATTCAGCGGCAGAG
PGSC0003DMT400014700	Chr03	58711070	DTH_Sot1	TGCCGCTAGCTTTATGGATT	TTGTCTATCTTCATCATTCTTGTGC
PGSC0003DMT400012725	Chr04	66093675	DTM_Sot58	TTGCAAATAATTAAAATCCTCGT	TCCCATCTCGCATGTATCTG
PGSC0003DMT400039812	Chr12	226814	DTM_Sot9	CCATGAAAAGATTCATGTTGTG	AGCTGGGTAATGAGGTGTGA
PGSC0003DMT400014381	Chr03	58890741	DTT_Sot1	ATCAACAAGCATGTGGGTCA	TGTATATTCCCAGGGCCAAA
PGSC0003DMT400013691	Chr04	30735205	DTM_Sot8	CTGGAGCAGCATATCCAACA	TTTTGGCTTGGCCAGATTAG
PGSC0003DMT400032506	Chr08	405493	DTM_Sot38	TCATGTCCACCATCCAACAC	ACCCGTACCCAGATCAACAA
PGSC0003DMT400053419	Chr01	79669600	DTT_Sot1	TTGGTCGTGTGTATAATGCTGA	CCAGGCAAAAATGTTGAGAAA
PGSC0003DMT400060389	Chr05	51090974	DTM_Sot48	GAGCATGTATTTTTCCGGTGA	ACAGCACCCAAACACACAAA
PGSC0003DMT400079251	Chr04	64280195	DTM_Sot6	ACCGGGATTTACAGTGACCA	ACCAATTGAAACTCGGGAAA
PGSC0003DMT400014299	Chr03	59513488	DTH_Sot6	GAGGATCCAGATTGATGACGA	GGAATTGGGAAGAAGAGATGA
PGSC0003DMT400024331	Chr02	26927848	DTT_Sot9	TGCCTTTAGTTTGATAAATGGACA	TGAGGTAAACCGGCTTCTTG

Supplementary Figure 3.3: Distribution of reference MITEs across the potato genome by genotype in 1MB bins.



Supplementary Figure 3.4: Distribution of non-reference MITEs across the potato genome by genotype in 1MB bins.

Supplementary Table 3.2: MITE-associated differentially expressed genes. The gene, location of the MITE with respect to the gene, MITE superfamily, MITE family and tissue are given. Upreg and downreg indicate if the gene is upregulated or downregulated in association with the gene. Finally the uncorrected and corrected p-values are given as well as if the MITE is a reference or non-reference insertion.

Gene ID	MITE location	Superfamily	Family	Tissue	Up/Down	P-value	Corr. P-value	Ref/Nonref
PGSC0003DMT400010210	INTERNAL	DTA	Sot2	leaf	upreg	0.000170877	0.033906857	ref
PGSC0003DMT400037990	INTERNAL	DTA	Sot23	tuber	upreg	3.33E-09	3.65E-05	non_ref
PGSC0003DMT400035365	INTERNAL	DTH	Sot3	leaf	upreg	9.03E-06	0.009821371	ref
PGSC0003DMT400088898	PROMOTER	DTM	Sot59	leaf	upreg	5.57E-07	0.001164177	non_ref
PGSC0003DMT400010386	INTERNAL	DTM	Sot18	leaf	upreg	1.41E-06	0.003520382	ref
PGSC0003DMT400003423	PROMOTER	DTM	Sot42	tuber	upreg	1.69E-06	0.003520382	ref
PGSC0003DMT400006585	PROMOTER	DTM	Sot46	leaf	upreg	2.49E-05	0.009046263	non_ref
PGSC0003DMT400079251	PROMOTER	DTT	Sot9	leaf	upreg	9.43E-06	0.009821371	ref
PGSC0003DMT400066516	INTERNAL	DTT	Sot7	tuber	upreg	6.45E-06	0.005055571	non_ref
PGSC0003DMT400006887	INTERNAL	DTA	Sot27	leaf	upreg	5.21E-05	0.019727024	ref
PGSC0003DMT400050786	INTERNAL	DTM	Sot9	tuber	upreg	1.94E-05	0.008533371	non_ref
PGSC0003DMT400030634	INTERNAL	DTA	Sot5	tuber	upreg	1.75E-05	0.008329657	non_ref

PGSC0003DMT400004714	INTERNAL	DTM	Sot2	tuber	upreg	2.13E-05	0.00864913	non_ref
PGSC0003DMT400010386	INTERNAL	DTM	Sot18	tuber	upreg	1.92E-05	0.012557733	ref
PGSC0003DMT400039420	INTERNAL	DTT	Sot16	leaf	upreg	9.89E-06	0.005838885	non_ref
PGSC0003DMT400053693	DOWNSTREAM	DTM	Sot21	tuber	upreg	8.14E-05	0.01861236	non_ref
PGSC0003DMT400051416	PROMOTER	DTA	Sot44	tuber	upreg	2.62E-05	0.009046263	non_ref
PGSC0003DMT400003423	PROMOTER	DTM	Sot42	leaf	upreg	1.72E-05	0.012557733	ref
PGSC0003DMT400088259	PROMOTER	DTA	Sot9	tuber	upreg	6.07E-05	0.014909175	non_ref
PGSC0003DMT400025781	DOWNSTREAM	DTM	Sot13	leaf	upreg	2.11E-05	0.012557733	ref
PGSC0003DMT400018927	PROMOTER	DTM	Sot62	leaf	upreg	0.000129208	0.022876089	non_ref
PGSC0003DMT400014700	INTERNAL	DTH	Sot1	leaf	upreg	2.57E-05	0.01275631	ref
PGSC0003DMT400061577	INTERNAL	DTA	Sot41	tuber	upreg	0.000103899	0.020930663	non_ref
PGSC0003DMT400042614	DOWNSTREAM	DTT	Sot9	leaf	upreg	2.25E-05	0.008825542	non_ref
PGSC0003DMT400037174	INTERNAL	DTA	Sot23	leaf	upreg	0.000275296	0.034660508	non_ref
PGSC0003DMT400040753	INTERNAL	DTH	Sot2	leaf	upreg	2.76E-05	0.01275631	ref
PGSC0003DMT400003459	PROMOTER	DTT	Sot1	leaf	upreg	8.69E-05	0.024680342	ref
PGSC0003DMT400007360	INTERNAL	DTM	Sot8	leaf	upreg	3.81E-05	0.015896247	ref
PGSC0003DMT40001892	PROMOTER	DTM	Sot62	tuber	upreg	0.00016502	0.02695793	non_ref

7						1	4	
PGSC0003DMT400014425	PROMOTER	DTA	Sot2	tuber	upreg	8.88E-05	0.019121961	non_ref
PGSC0003DMT400014560	INTERNAL	DTM	Sot4	leaf	upreg	0.000410859	0.039561419	non_ref
PGSC0003DMT400036704	PROMOTER	DTA	Sot4	leaf	upreg	0.000116521	0.029141928	ref
PGSC0003DMT400042532	DOWNSTREAM	DTT	Sot1	tuber	upreg	0.000175937	0.026957934	non_ref
PGSC0003DMT400049308	PROMOTER	DTT	Sot7	leaf	upreg	0.000173448	0.026957934	non_ref
PGSC0003DMT400070805	PROMOTER	DTM	Sot42	tuber	upreg	0.0001081	0.020930663	non_ref
PGSC0003DMT400089705	PROMOTER	DTM	Sot8	tuber	upreg	0.000147104	0.025207738	non_ref
PGSC0003DMT400055259	DOWNSTREAM	DTH	Sot13	leaf	upreg	0.000141822	0.03110387	ref
PGSC0003DMT400025724	DOWNSTREAM	DTM	Sot3	tuber	upreg	0.000123194	0.022168898	non_ref
PGSC0003DMT400058960	INTERNAL	DTT	Sot11	leaf	upreg	0.000562684	0.049020467	non_ref
PGSC0003DMT400041420	INTERNAL	DTM	Sot1	tuber	upreg	0.000188619	0.02790672	non_ref
PGSC0003DMT400038571	DOWNSTREAM	DTT	Sot7	tuber	upreg	0.000135676	0.03110387	ref
PGSC0003DMT400046760	INTERNAL	DTM	Sot2	leaf	upreg	0.000199527	0.036149159	ref
PGSC0003DMT400063189	DOWNSTREAM	DTT	Sot2	leaf	upreg	0.000358497	0.037112012	non_ref
PGSC0003DMT400077329	PROMOTER	DTT	Sot8	leaf	upreg	0.000200365	0.02819759	non_ref
PGSC0003DMT400015365	DOWNSTREAM	DTH	Sot3	leaf	upreg	0.00011889	0.029141928	ref

PGSC0003DMT400029426	INTERNAL	DTM	Sot31	tuber	upreg	0.000197522	0.028158381	non_ref
PGSC0003DMT400019522	INTERNAL	DTT	Sot7	leaf	upreg	0.000485153	0.043740063	non_ref
PGSC0003DMT400072874	PROMOTER	DTA	Sot5	leaf	upreg	0.000334637	0.036733131	non_ref
PGSC0003DMT400043219	PROMOTER	DTT	Sot5	leaf	upreg	0.000187479	0.03551032	ref
PGSC0003DMT400037582	INTERNAL	DTA	Sot16	tuber	upreg	0.000352664	0.037112012	non_ref
PGSC0003DMT400042351	DOWNSTREAM	DTM	Sot16	tuber	upreg	0.000370688	0.037112012	non_ref
PGSC0003DMT400050893	INTERNAL	DTA	Sot29	leaf	upreg	0.000591147	0.049415313	non_ref
PGSC0003DMT400009386	INTERNAL	DTM	Sot2	leaf	upreg	0.000388465	0.038254857	non_ref
PGSC0003DMT400017187	PROMOTER	DTM	Sot42	leaf	downreg	0.000588467	0.049415313	non_ref
PGSC0003DMT400022096	INTERNAL	DTM	Sot21	leaf	downreg	0.000486134	0.043740063	non_ref
PGSC0003DMT400003398	INTERNAL	DTH	Sot14	leaf	downreg	0.000462101	0.042987117	non_ref
PGSC0003DMT400059387	DOWNSTREAM	DTM	Sot29	tuber	downreg	0.000590331	0.049415313	non_ref
PGSC0003DMT400037994	INTERNAL	DTT	Sot2	leaf	downreg	0.000393805	0.038254857	non_ref
PGSC0003DMT400068887	DOWNSTREAM	DTA	Sot22	leaf	downreg	0.000371898	0.037112012	non_ref
PGSC0003DMT400010389	DOWNSTREAM	DTA	Sot16	leaf	downreg	0.000500374	0.044295242	non_ref
PGSC0003DMT400031878	PROMOTER	DTH	Sot6	leaf	downreg	0.000333867	0.036733131	non_ref
PGSC0003DMT40005320	PROMOTER	DTT	Sot7	leaf	downreg	0.00058866	0.04941531	non_ref

9							3	
PGSC0003DMT400006502	DOWNSTREAM	DTM	Sot18	leaf	downreg	0.000356193	0.037112012	non_ref
PGSC0003DMT400008149	PROMOTER	DTH	Sot2	leaf	downreg	0.000579951	0.049415313	non_ref
PGSC0003DMT400014590	INTERNAL	DTT	Sot7	leaf	downreg	0.000447571	0.041991357	non_ref
PGSC0003DMT400075571	DOWNSTREAM	DTT	Sot3	leaf	downreg	0.000274022	0.034660508	non_ref
PGSC0003DMT400006798	DOWNSTREAM	DTA	Sot42	leaf	downreg	0.000266458	0.034660508	non_ref
PGSC0003DMT400048880	PROMOTER	DTT	Sot10	leaf	downreg	0.000355819	0.037112012	non_ref
PGSC0003DMT400066245	PROMOTER	DTM	Sot8	leaf	downreg	0.000328578	0.036733131	non_ref
PGSC0003DMT400080229	PROMOTER	DTA	Sot17	tuber	downreg	0.000497571	0.044295242	non_ref
PGSC0003DMT400053777	DOWNSTREAM	DTA	Sot42	tuber	downreg	0.00039338	0.038254857	non_ref
PGSC0003DMT400069025	PROMOTER	DTM	Sot35	tuber	downreg	0.000365771	0.037112012	non_ref
PGSC0003DMT400024400	DOWNSTREAM	DTT	Sot16	tuber	downreg	0.000308979	0.035894407	non_ref
PGSC0003DMT400040853	PROMOTER	DTM	Sot42	tuber	downreg	0.000365367	0.037112012	non_ref
PGSC0003DMT400082365	PROMOTER	DTM	Sot2	leaf	downreg	0.00019563	0.028158381	non_ref
PGSC0003DMT400036990	DOWNSTREAM	DTM	Sot27	leaf	downreg	0.000430409	0.040729272	non_ref
PGSC0003DMT400005432	PROMOTER	DTH	Sot6	tuber	downreg	0.000425697	0.040633701	non_ref
PGSC0003DMT400021450	PROMOTER	DTC	Sot4	leaf	downreg	0.00028418	0.034660508	non_ref

PGSC0003DMT400000803	INTERNAL	DTC	Sot1	leaf	downreg	0.000484919	0.043740063	non_ref
PGSC0003DMT400069059	DOWNSTREAM	DTT	Sot1	leaf	downreg	0.000594226	0.049415313	non_ref
PGSC0003DMT400083968	DOWNSTREAM	DTT	Sot2	tuber	downreg	0.000283049	0.034660508	non_ref
PGSC0003DMT400040528	DOWNSTREAM	DTA	Sot2	tuber	downreg	0.000242281	0.033243957	non_ref
PGSC0003DMT400048389	INTERNAL	DTA	Sot42	tuber	downreg	0.00032576	0.036733131	non_ref
PGSC0003DMT400082480	INTERNAL	DTM	Sot56	leaf	downreg	0.000283568	0.034660508	non_ref
PGSC0003DMT400097233	PROMOTER	DTT	Sot4	leaf	downreg	0.000190672	0.02790672	non_ref
PGSC0003DMT400041356	PROMOTER	DTH	Sot2	tuber	downreg	0.000299492	0.035733982	non_ref
PGSC0003DMT400047502	INTERNAL	DTH	Sot6	leaf	downreg	0.00027995	0.034660508	non_ref
PGSC0003DMT400024106	INTERNAL	DTM	Sot9	leaf	downreg	0.000185948	0.02790672	non_ref
PGSC0003DMT400032442	INTERNAL	DTM	Sot1	tuber	downreg	0.000307357	0.035894407	non_ref
PGSC0003DMT400060225	INTERNAL	DTA	Sot56	tuber	downreg	0.000522525	0.045886049	non_ref
PGSC0003DMT400058577	DOWNSTREAM	DTA	Sot23	leaf	downreg	0.000258328	0.034581291	non_ref
PGSC0003DMT400006365	INTERNAL	DTA	Sot56	tuber	downreg	0.000176466	0.026957934	non_ref
PGSC0003DMT400043791	DOWNSTREAM	DTM	Sot35	tuber	downreg	0.000280613	0.034660508	non_ref
PGSC0003DMT400041666	INTERNAL	DTH	Sot15	tuber	downreg	0.000163572	0.026957934	non_ref
PGSC0003DMT400018400	PROMOTER	DTH	Sot6	tuber	downreg	0.00020771	0.02886245	non_ref

0						9		
PGSC0003DMT400002341	INTERNAL	DTM	Sot8	tuber	downreg	0.000176822	0.026957934	non_ref
PGSC0003DMT400051632	PROMOTER	DTT	Sot7	tuber	downreg	0.00035707	0.037112012	non_ref
PGSC0003DMT400076032	PROMOTER	DTT	Sot1	leaf	downreg	0.000167762	0.026957934	non_ref
PGSC0003DMT400053777	DOWNSTREAM	DTA	Sot42	leaf	downreg	0.0001061	0.020930663	non_ref
PGSC0003DMT400063112	INTERNAL	DTA	Sot17	tuber	downreg	0.00025669	0.034581291	non_ref
PGSC0003DMT400063112	INTERNAL	DTA	Sot17	leaf	downreg	0.000113243	0.021068943	non_ref
PGSC0003DMT400010156	DOWNSTREAM	DTM	Sot27	leaf	downreg	8.40E-05	0.018814897	non_ref
PGSC0003DMT400054432	INTERNAL	DTM	Sot2	leaf	downreg	0.000310647	0.035894407	non_ref
PGSC0003DMT400037748	INTERNAL	DTM	Sot4	tuber	downreg	0.000149267	0.025207738	non_ref
PGSC0003DMT400075571	DOWNSTREAM	DTT	Sot3	tuber	downreg	0.000290806	0.035078822	non_ref
PGSC0003DMT400045787	INTERNAL	DTA	Sot2	leaf	downreg	9.78E-05	0.02027179	non_ref
PGSC0003DMT400079931	PROMOTER	DTH	Sot13	leaf	downreg	6.56E-05	0.015665729	non_ref
PGSC0003DMT400001636	INTERNAL	DTT	Sot2	leaf	downreg	5.45E-05	0.014234477	non_ref
PGSC0003DMT400014425	INTERNAL	DTT	Sot7	tuber	downreg	8.88E-05	0.024680342	ref
PGSC0003DMT400018927	PROMOTER	DTM	Sot58	tuber	downreg	0.000165021	0.033906857	ref
PGSC0003DMT400006475	PROMOTER	DTT	Sot7	tuber	downreg	8.62E-05	0.018928634	non_ref

PGSC0003DMT400018344	PROMOTER	DTT	Sot7	leaf	downreg	3.93E-05	0.011673429	non_ref
PGSC0003DMT400067714	PROMOTER	DTT	Sot11	leaf	downreg	4.50E-05	0.012845094	non_ref
PGSC0003DMT400024331	INTERNAL	DTT	Sot9	leaf	downreg	4.56E-05	0.012845094	non_ref
PGSC0003DMT400037964	DOWNSTREAM	DTM	Sot56	leaf	downreg	9.79E-05	0.02027179	non_ref
PGSC0003DMT400007734	DOWNSTREAM	DTT	Sot7	leaf	downreg	2.92E-05	0.009422526	non_ref
PGSC0003DMT400043791	DOWNSTREAM	DTM	Sot35	leaf	downreg	2.72E-05	0.009046263	non_ref
PGSC0003DMT400004562	INTERNAL	DTA	Sot22	leaf	downreg	0.00046723	0.043099062	non_ref
PGSC0003DMT400014299	PROMOTER	DTH	Sot6	leaf	downreg	2.69E-05	0.009046263	non_ref
PGSC0003DMT400014381	INTERNAL	DTM	Sot22	leaf	downreg	0.000146698	0.025207738	non_ref
PGSC0003DMT400039840	INTERNAL	DTM	Sot35	leaf	downreg	0.000112934	0.021068943	non_ref
PGSC0003DMT400035974	DOWNSTREAM	DTT	Sot4	leaf	downreg	6.75E-05	0.015754539	non_ref
PGSC0003DMT400027032	PROMOTER	DTA	Sot64	tuber	downreg	6.11E-05	0.014909175	non_ref
PGSC0003DMT400047305	DOWNSTREAM	DTA	Sot42	leaf	downreg	3.12E-05	0.009776777	non_ref
PGSC0003DMT400032506	INTERNAL	DTM	Sot38	tuber	downreg	3.46E-05	0.010549665	non_ref
PGSC0003DMT400076623	DOWNSTREAM	DTA	Sot64	leaf	downreg	0.000364648	0.037112012	non_ref
PGSC0003DMT400054373	PROMOTER	DTM	Sot25	leaf	downreg	5.88E-05	0.014909175	non_ref
PGSC0003DMT40007919	PROMOTER	DTA	Sot64	leaf	downreg	1.21E-05	0.00666006	non_ref

8								
PGSC0003DMT400057894	INTERNAL	DTM	Sot25	leaf	downreg	8.01E-05	0.024680342	ref
PGSC0003DMT400041777	INTERNAL	DTT	Sot1	tuber	downreg	5.20E-05	0.01403652	non_ref
PGSC0003DMT400001660	PROMOTER	DTT	Sot15	leaf	downreg	1.66E-05	0.008276347	non_ref
PGSC0003DMT400039812	PROMOTER	DTM	Sot9	tuber	downreg	5.68E-05	0.019727024	ref
PGSC0003DMT400039840	INTERNAL	DTM	Sot35	tuber	downreg	0.000120825	0.022104983	non_ref
PGSC0003DMT400046304	PROMOTER	DTM	Sot54	tuber	downreg	0.000333311	0.036733131	non_ref
PGSC0003DMT400020475	DOWNSTREAM	DTT	Sot7	leaf	downreg	1.01E-05	0.005838885	non_ref
PGSC0003DMT400011342	DOWNSTREAM	DTM	Sot22	tuber	downreg	2.04E-05	0.008622588	non_ref
PGSC0003DMT400077129	PROMOTER	DTT	Sot9	tuber	downreg	1.82E-05	0.008329657	non_ref
PGSC0003DMT400037748	INTERNAL	DTM	Sot4	leaf	downreg	1.32E-05	0.00691061	non_ref
PGSC0003DMT400026735	INTERNAL	DTT	Sot7	tuber	downreg	0.000108686	0.020930663	non_ref
PGSC0003DMT400079251	PROMOTER	DTM	Sot6	leaf	downreg	9.43E-06	0.005838885	non_ref
PGSC0003DMT400066226	PROMOTER	DTT	Sot9	leaf	downreg	2.94E-06	0.002687481	non_ref
PGSC0003DMT400001660	PROMOTER	DTT	Sot15	tuber	downreg	4.81E-06	0.004065422	non_ref
PGSC0003DMT400060389	PROMOTER	DTM	Sot48	leaf	downreg	1.10E-06	0.001212455	non_ref
PGSC0003DMT400082352	INTERNAL	DTA	Sot4	leaf	downreg	8.19E-07	0.001207678	non_ref

PGSC0003DMT400048157	PROMOTER	DTA	Sot25	leaf	downreg	8.77E-06	0.005838885	non_ref
PGSC0003DMT400003346	PROMOTER	DTA	Sot64	leaf	downreg	2.52E-05	0.009046263	non_ref
PGSC0003DMT400005432	PROMOTER	DTH	Sot6	leaf	downreg	1.24E-06	0.001235735	non_ref
PGSC0003DMT400003291	INTERNAL	DTA	Sot27	leaf	downreg	6.36E-07	0.001164177	non_ref
PGSC0003DMT400005684	INTERNAL	DTM	Sot29	leaf	downreg	1.07E-06	0.001212455	non_ref
PGSC0003DMT400046304	PROMOTER	DTM	Sot54	leaf	downreg	5.24E-05	0.01403652	non_ref
PGSC0003DMT400053419	INTERNAL	DTT	Sot1	leaf	downreg	8.80E-07	0.001207678	non_ref
PGSC0003DMT400009938	DOWNSTREAM	DTA	Sot9	leaf	downreg	1.74E-07	0.000638218	non_ref
PGSC0003DMT400089423	INTERNAL	DTA	Sot27	leaf	downreg	6.91E-06	0.005057795	non_ref
PGSC0003DMT400053419	INTERNAL	DTT	Sot1	tuber	downreg	2.42E-07	0.000665058	non_ref
PGSC0003DMT400032506	INTERNAL	DTM	Sot38	leaf	downreg	9.93E-09	5.45E-05	non_ref

Supplementary Table 3.3: Summary of enriched GO terms. For each GO term, the number of MITE-Associated genes with (present) or without (absent) is listed. The same is true for non-MITE-associated genes. The direction of enrichment indicates if the term is overrepresented in either the MITE-associated (with-enriched) or non-MITE-associated (without_enriched) gene set. Finally, the results of each Fisher's exact test are given.

GO Term	Present - MITE Associated	Absent - MITE Associated	Present - Non-Associated	Absent - Non-Associated	Direction of Enrichment	Fisher's Exact Test
GO:0046777	47	6684	150	11296	without_enriched	0.0001037614
GO:0016413	24	6707	10	11436	with_enriched	0.0001037815
GO:0031146	13	6718	2	11444	with_enriched	0.0001107688
GO:0004004	7	6724	48	11398	without_enriched	0.0001112665
GO:0005886	1313	5418	2509	8937	without_enriched	0.0001115353
GO:0009987	38	6693	128	11318	without_enriched	0.0001321335
GO:0009626	89	6642	85	11361	with_enriched	0.0001426228
GO:0008026	16	6715	72	11374	without_enriched	0.0001455892
GO:0005730	138	6593	341	11105	without_enriched	0.0001467818
GO:0009615	5	6726	40	11406	without_enriched	0.0001477692
GO:0006952	337	6394	439	11007	with_enriched	0.0001935665
GO:0007264	21	6710	84	11362	without_enriched	0.0002275281
GO:0004386	18	6713	74	11372	without_enriched	0.0003275258
GO:0042177	10	6721	1	11445	with_enriched	0.0003524464
GO:0043621	17	6714	72	11374	without_enriched	0.0003679334
GO:0031969	31	6700	105	11341	without_enriched	0.0004679395
GO:0016709	84	6647	221	11225	without_enriched	0.000504567
GO:0009941	199	6532	452	10994	without_enriched	0.0005078536

GO:0018024	4	6727	33	11413	without_enriched	0.000512992
GO:0006412	114	6617	281	11165	without_enriched	0.0006049861
GO:0032440	52	6679	150	11296	without_enriched	0.0007212949
GO:0005975	111	6620	274	11172	without_enriched	0.000756933
GO:0050660	35	6696	112	11334	without_enriched	0.0007703985
GO:0005506	168	6563	387	11059	without_enriched	0.0007996649
EC:2.3.1.0	5	6726	35	11411	without_enriched	0.0008452397
GO:0009570	239	6492	524	10922	without_enriched	0.0008516531
EC:2.4.1.0	20	6711	76	11370	without_enriched	0.0009155446
GO:0048446	7	6724	0	11446	with_enriched	0.0009529017
EC:1.3.1.74	50	6681	144	11302	without_enriched	0.0009680732
GO:0046982	35	6696	111	11335	without_enriched	0.001005781
GO:0005794	381	6350	790	10656	without_enriched	0.001013075
GO:0016571	1	6730	20	11426	without_enriched	0.001102855
GO:0052689	65	6666	61	11385	with_enriched	0.001104957
GO:0009536	186	6545	419	11027	without_enriched	0.001121745
GO:0006310	5	6726	34	11412	without_enriched	0.001300213
GO:0003676	164	6567	372	11074	without_enriched	0.001711458
EC:2.1.1.0	31	6700	99	11347	without_enriched	0.001824424
GO:0030125	0	6731	15	11431	without_enriched	0.001829945
EC:6.3.2.19	36	6695	110	11336	without_enriched	0.00185125
GO:0031201	7	6724	39	11407	without_enriched	0.001888828
GO:0048364	56	6675	153	11293	without_enriched	0.00189822

GO:0090333	4	6727	30	11416	without_enriched	0.001985729
GO:0006508	139	6592	322	11124	without_enriched	0.002053408
GO:0005737	1358	5373	2531	8915	without_enriched	0.002129048
GO:0080054	8	6723	1	11445	with_enriched	0.00212982
GO:0009791	13	6718	55	11391	without_enriched	0.002262261
GO:0008233	39	6692	115	11331	without_enriched	0.002451949
GO:0000413	8	6723	40	11406	without_enriched	0.002599659
GO:0005525	79	6652	199	11247	without_enriched	0.002616464
GO:0005789	123	6608	287	11159	without_enriched	0.002680691
GO:0046854	3	6728	25	11421	without_enriched	0.00282692
GO:0009630	6	6725	35	11411	without_enriched	0.003013945
GO:0003918	0	6731	13	11433	without_enriched	0.003050871
GO:0009785	1	6730	18	11428	without_enriched	0.003131684
GO:0031012	1	6730	18	11428	without_enriched	0.003131684
GO:0009860	29	6702	91	11355	without_enriched	0.003159877
GO:0006281	43	6688	122	11324	without_enriched	0.003430539
GO:0005768	94	6637	227	11219	without_enriched	0.003519104
GO:0003735	113	6618	265	11181	without_enriched	0.003616277
GO:0009817	18	6713	64	11382	without_enriched	0.004028391
GO:0000786	20	6711	69	11377	without_enriched	0.004037694
GO:0050794	20	6711	69	11377	without_enriched	0.004037694
GO:0009504	7	6724	36	11410	without_enriched	0.004060706
GO:0009826	40	6691	114	11332	without_enriched	0.004256839

GO:0009620	47	6684	44	11402	with_enriched	0.004545336
GO:0006184	17	6714	61	11385	without_enriched	0.004644452
GO:0043086	21	6710	71	11375	without_enriched	0.004646917
GO:0020037	193	6538	417	11029	without_enriched	0.004868938
EC:3.4.11.0	1	6730	16	11430	without_enriched	0.005308628
GO:0004177	1	6730	17	11429	without_enriched	0.005389352
GO:0006265	1	6730	17	11429	without_enriched	0.005389352
GO:0009625	1	6730	17	11429	without_enriched	0.005389352
GO:0016757	186	6545	403	11043	without_enriched	0.005466153
GO:0006897	7	6724	35	11411	without_enriched	0.005945866
GO:0031640	7	6724	35	11411	without_enriched	0.005945866
GO:0005618	234	6497	493	10953	without_enriched	0.006040081
GO:0046872	648	6083	1250	10196	without_enriched	0.006177892
GO:0031347	20	6711	13	11433	with_enriched	0.006437087
GO:0009414	102	6629	238	11208	without_enriched	0.006477594
GO:0071555	76	6655	186	11260	without_enriched	0.006724356
GO:0006470	29	6702	87	11359	without_enriched	0.006779006
GO:0006754	5	6726	29	11417	without_enriched	0.006828397
EC:1.10.3.1	5	6726	0	11446	with_enriched	0.006956335
GO:0000412	5	6726	0	11446	with_enriched	0.006956335
GO:0004097	5	6726	0	11446	with_enriched	0.006956335
GO:0010321	5	6726	0	11446	with_enriched	0.006956335
GO:0016127	5	6726	0	11446	with_enriched	0.006956335
GO:003443	5	6726	0	11446	with_enriched	0.006956335

4						
GO:0080143	5	6726	0	11446	with_enriched	0.006956335
GO:0009416	61	6670	155	11291	without_enriched	0.006990218
GO:0000287	47	6684	126	11320	without_enriched	0.007016383
GO:0090404	4	6727	26	11420	without_enriched	0.007062203
GO:0009408	54	6677	141	11305	without_enriched	0.007118508
GO:1902183	8	6723	2	11444	with_enriched	0.007167781
GO:0044550	75	6656	183	11263	without_enriched	0.007681305
GO:0009834	11	6720	44	11402	without_enriched	0.00775434
GO:0006810	124	6607	280	11166	without_enriched	0.007819989
GO:0016192	43	6688	116	11330	without_enriched	0.008267038
GO:0035556	54	6677	139	11307	without_enriched	0.008641402
GO:0046961	7	6724	34	11412	without_enriched	0.00872795
GO:0080022	7	6724	34	11412	without_enriched	0.00872795
GO:0005509	71	6660	174	11272	without_enriched	0.009248906
GO:0007112	0	6731	11	11435	without_enriched	0.009362126
GO:0009992	0	6731	11	11435	without_enriched	0.009362126
GO:0033320	0	6731	11	11435	without_enriched	0.009362126
GO:0007165	134	6597	169	11277	with_enriched	0.009839922
GO:0048765	5	6726	27	11419	without_enriched	0.01030519
GO:0000976	2	6729	18	11428	without_enriched	0.01033121
GO:0034968	2	6729	18	11428	without_enriched	0.01033121
EC:3.1.3.16	5	6726	28	11418	without_enriched	0.01034061
GO:000372	79	6652	189	11257	without_enriche	0.01066962

9					d	
GO:0042542	11	6720	43	11403	without_enriched	0.01071737
GO:0030176	2	6729	19	11427	without_enriched	0.01076243
GO:0009704	10	6721	4	11442	with_enriched	0.0110372
GO:0080118	10	6721	4	11442	with_enriched	0.0110372
GO:0003723	207	6524	434	11012	without_enriched	0.01109694
GO:0019825	127	6604	282	11164	without_enriched	0.01114056
GO:0042254	38	6693	104	11342	without_enriched	0.01121776
GO:0000139	100	6631	230	11216	without_enriched	0.01126512
GO:0005654	17	6714	58	11388	without_enriched	0.01132395
GO:0003779	15	6716	53	11393	without_enriched	0.0113283
GO:0009644	15	6716	53	11393	without_enriched	0.0113283
GO:0006888	8	6723	36	11410	without_enriched	0.01142585
GO:0009534	87	6644	204	11242	without_enriched	0.01197952
GO:0004872	53	6678	135	11311	without_enriched	0.01207976
GO:0009816	10	6721	40	11406	without_enriched	0.01225156
GO:0008237	7	6724	32	11414	without_enriched	0.0125146
GO:0015020	18	6713	12	11434	with_enriched	0.01279817
GO:0000149	7	6724	33	11413	without_enriched	0.01282101
GO:0005840	95	6636	218	11228	without_enriched	0.0131559
GO:0018105	16	6715	55	11391	without_enriched	0.01318032
GO:0009793	141	6590	308	11138	without_enriched	0.01327139
GO:0005802	82	6649	193	11253	without_enriched	0.01397812
GO:000939	1	6730	14	11432	without_enriche	0.01421335

9					d	
GO:0016717	1	6730	14	11432	without_enriched	0.01421335
GO:0050362	1	6730	14	11432	without_enriched	0.01421335
GO:0031977	11	6720	42	11404	without_enriched	0.01482376
GO:0004725	5	6726	26	11420	without_enriched	0.01510201
GO:0019898	5	6726	26	11420	without_enriched	0.01510201
GO:0042391	5	6726	26	11420	without_enriched	0.01510201
GO:0061025	5	6726	26	11420	without_enriched	0.01510201
GO:0016798	13	6718	47	11399	without_enriched	0.01524525
GO:0005774	215	6516	445	11001	without_enriched	0.01543598
GO:0009524	22	6709	68	11378	without_enriched	0.01555624
GO:0007169	67	6664	161	11285	without_enriched	0.01572358
GO:0009579	86	6645	199	11247	without_enriched	0.015843
GO:0005773	234	6497	480	10966	without_enriched	0.01589995
GO:0034765	4	6727	23	11423	without_enriched	0.01613369
GO:0007389	2	6729	17	11429	without_enriched	0.01628047
GO:0006012	3	6728	20	11426	without_enriched	0.01667504
GO:0009654	3	6728	20	11426	without_enriched	0.01667504
GO:0048278	3	6728	20	11426	without_enriched	0.01667504
GO:0000105	0	6731	10	11436	without_enriched	0.01695553
GO:0010929	0	6731	10	11436	without_enriched	0.01695553
GO:0071421	0	6731	10	11436	without_enriched	0.01695553
GO:0016705	79	6652	184	11262	without_enriched	0.01737813
GO:000973	84	6647	194	11252	without_enriche	0.01740379

5					d	
GO:0009058	32	6699	89	11357	without_enriched	0.01782391
EC:3.2.1.15	7	6724	31	11415	without_enriched	0.01789331
GO:0032580	7	6724	31	11415	without_enriched	0.01789331
GO:0008643	24	6707	71	11375	without_enriched	0.01870715
EC:3.5.1.1	4	6727	0	11446	with_enriched	0.01879255
GO:0004067	4	6727	0	11446	with_enriched	0.01879255
GO:0008798	4	6727	0	11446	with_enriched	0.01879255
GO:0016099	4	6727	0	11446	with_enriched	0.01879255
GO:0035264	4	6727	0	11446	with_enriched	0.01879255
GO:0035865	4	6727	0	11446	with_enriched	0.01879255
GO:0043161	47	6684	49	11397	with_enriched	0.01928223
GO:0048367	15	6716	50	11396	without_enriched	0.02021176
GO:0016772	35	6696	94	11352	without_enriched	0.02181159
GO:0005249	5	6726	25	11421	without_enriched	0.02220127
GO:0006032	5	6726	25	11421	without_enriched	0.02220127
GO:0010105	5	6726	25	11421	without_enriched	0.02220127
GO:0048046	173	6558	362	11084	without_enriched	0.02304303
EC:1.14.19.0	1	6730	13	11433	without_enriched	0.02370233
EC:3.6.3.8	1	6730	13	11433	without_enriched	0.02370233
GO:0004356	1	6730	13	11433	without_enriched	0.02370233
GO:0046340	1	6730	13	11433	without_enriched	0.02370233
GO:0052651	1	6730	13	11433	without_enriched	0.02370233
GO:0080097	1	6730	13	11433	without_enriched	0.02370233

GO:0005694	10	6721	38	11408	without_enriched	0.02373887
GO:0015986	10	6721	38	11408	without_enriched	0.02373887
GO:0010628	8	6723	3	11443	with_enriched	0.0239077
GO:0012511	8	6723	3	11443	with_enriched	0.0239077
GO:0044459	8	6723	3	11443	with_enriched	0.0239077
GO:0004497	90	6641	204	11242	without_enriched	0.02410226
GO:0005753	4	6727	22	11424	without_enriched	0.02417472
GO:0008287	4	6727	22	11424	without_enriched	0.02417472
GO:0009543	19	6712	58	11388	without_enriched	0.02458188
GO:0006950	45	6686	113	11333	without_enriched	0.02548029
GO:0005992	3	6728	19	11427	without_enriched	0.02560995
GO:0044260	3	6728	19	11427	without_enriched	0.02560995
GO:0016887	62	6669	148	11298	without_enriched	0.02566095
GO:0004575	2	6729	16	11430	without_enriched	0.02585482
GO:0009880	2	6729	16	11430	without_enriched	0.02585482
GO:0042773	2	6729	16	11430	without_enriched	0.02585482
GO:0051301	63	6668	150	11296	without_enriched	0.02668497
GO:0007047	15	6716	49	11397	without_enriched	0.0268711
GO:0045735	33	6698	32	11414	with_enriched	0.02795409
GO:0000049	10	6721	5	11441	with_enriched	0.02860173
GO:0009734	75	6656	172	11274	without_enriched	0.02866114
EC:1.13.12.7	5	6726	1	11445	with_enriched	0.02886427
GO:0043066	5	6726	1	11445	with_enriched	0.02886427

GO:0047077	5	6726	1	11445	with_enriched	0.02886427
GO:1903507	5	6726	1	11445	with_enriched	0.02886427
GO:0005643	6	6725	27	11419	without_enriched	0.0290456
GO:0010501	6	6725	27	11419	without_enriched	0.0290456
GO:0051555	6	6725	27	11419	without_enriched	0.0290456
GO:0022625	40	6691	102	11344	without_enriched	0.02908168
GO:0000165	18	6713	56	11390	without_enriched	0.02914438
GO:1990135	12	6719	7	11439	with_enriched	0.02940143
GO:0003824	137	6594	292	11154	without_enriched	0.02947046
EC:2.4.1.1	0	6731	8	11438	without_enriched	0.02985262
EC:2.4.1.224	0	6731	8	11438	without_enriched	0.02985262
EC:5.4.2.1	0	6731	8	11438	without_enriched	0.02985262
GO:0004619	0	6731	8	11438	without_enriched	0.02985262
GO:0004645	0	6731	8	11438	without_enriched	0.02985262
GO:0005471	0	6731	8	11438	without_enriched	0.02985262
GO:0005978	0	6731	8	11438	without_enriched	0.02985262
GO:0009247	0	6731	8	11438	without_enriched	0.02985262
GO:0009330	0	6731	8	11438	without_enriched	0.02985262
GO:0009745	0	6731	8	11438	without_enriched	0.02985262
GO:0010112	0	6731	8	11438	without_enriched	0.02985262
GO:0010428	0	6731	8	11438	without_enriched	0.02985262
GO:0010429	0	6731	8	11438	without_enriched	0.02985262
GO:0015696	0	6731	8	11438	without_enriched	0.02985262

GO:0015749	0	6731	8	11438	without_enriched	0.02985262
GO:0030527	0	6731	8	11438	without_enriched	0.02985262
GO:0040019	0	6731	8	11438	without_enriched	0.02985262
GO:0050508	0	6731	8	11438	without_enriched	0.02985262
GO:0006334	24	6707	68	11378	without_enriched	0.03029781
GO:0009736	16	6715	51	11395	without_enriched	0.03040079
GO:0004252	41	6690	104	11342	without_enriched	0.03064096
EC:2.6.1.27	0	6731	9	11437	without_enriched	0.03129032
EC:2.7.1.68	0	6731	9	11437	without_enriched	0.03129032
EC:4.2.3.21	0	6731	9	11437	without_enriched	0.03129032
EC:5.99.1.3	0	6731	9	11437	without_enriched	0.03129032
GO:0010726	0	6731	9	11437	without_enriched	0.03129032
GO:0016004	0	6731	9	11437	without_enriched	0.03129032
GO:0034003	0	6731	9	11437	without_enriched	0.03129032
GO:0072488	0	6731	9	11437	without_enriched	0.03129032
GO:0090421	0	6731	9	11437	without_enriched	0.03129032
GO:0010214	8	6723	31	11415	without_enriched	0.03165579
GO:0009958	8	6723	32	11414	without_enriched	0.03174039
GO:0005484	10	6721	36	11410	without_enriched	0.03234505
GO:0016791	10	6721	36	11410	without_enriched	0.03234505
GO:0047372	5	6726	24	11422	without_enriched	0.0326766
GO:0048359	5	6726	24	11422	without_enriched	0.0326766
GO:0008061	10	6721	37	11409	without_enriched	0.03295797

GO:0051607	10	6721	37	11409	without_enriched	0.03295797
GO:0003924	60	6671	142	11304	without_enriched	0.03336607
EC:3.6.1.3	17	6714	52	11394	without_enriched	0.03370186
GO:0006511	76	6655	173	11273	without_enriched	0.03433838
GO:0003677	500	6231	951	10495	without_enriched	0.03596793
GO:0048467	4	6727	21	11425	without_enriched	0.03628671
GO:0004197	7	6724	28	11418	without_enriched	0.03641509
GO:0015992	7	6724	28	11418	without_enriched	0.03641509
GO:0016998	7	6724	29	11417	without_enriched	0.03659028
GO:0042127	7	6724	29	11417	without_enriched	0.03659028
GO:0046933	7	6724	29	11417	without_enriched	0.03659028
GO:0051762	13	6718	43	11403	without_enriched	0.03685925
GO:0010114	18	6713	54	11392	without_enriched	0.03726057
GO:0051537	18	6713	54	11392	without_enriched	0.03726057
GO:0010224	27	6704	73	11373	without_enriched	0.03794248
GO:0030246	59	6672	139	11307	without_enriched	0.03802225
GO:0009873	64	6667	148	11298	without_enriched	0.03813901
GO:0019953	12	6719	8	11438	with_enriched	0.03846521
GO:0045454	42	6689	104	11342	without_enriched	0.03893539
EC:2.3.1.48	3	6728	18	11428	without_enriched	0.03944224
GO:0000266	3	6728	18	11428	without_enriched	0.03944224
GO:0010026	3	6728	18	11428	without_enriched	0.03944224
GO:0042644	3	6728	18	11428	without_enriched	0.03944224

GO:0023014	16	6715	49	11397	without_enriched	0.03952299
EC:2.4.1.15	1	6730	11	11435	without_enriched	0.03985756
EC:3.1.4.4	1	6730	11	11435	without_enriched	0.03985756
EC:3.2.1.26	1	6730	11	11435	without_enriched	0.03985756
GO:0001558	1	6730	11	11435	without_enriched	0.03985756
GO:0003825	1	6730	11	11435	without_enriched	0.03985756
GO:0006298	1	6730	11	11435	without_enriched	0.03985756
GO:0009231	1	6730	11	11435	without_enriched	0.03985756
GO:0009616	1	6730	11	11435	without_enriched	0.03985756
GO:0009955	1	6730	11	11435	without_enriched	0.03985756
GO:0030042	1	6730	11	11435	without_enriched	0.03985756
GO:0055062	1	6730	11	11435	without_enriched	0.03985756
GO:0070413	1	6730	11	11435	without_enriched	0.03985756
EC:3.4.22.0	1	6730	12	11434	without_enriched	0.03988109
GO:0006777	1	6730	12	11434	without_enriched	0.03988109
GO:0009963	1	6730	12	11434	without_enriched	0.03988109
GO:0010205	1	6730	12	11434	without_enriched	0.03988109
GO:0016308	1	6730	12	11434	without_enriched	0.03988109
GO:0030140	1	6730	12	11434	without_enriched	0.03988109
GO:0030515	1	6730	12	11434	without_enriched	0.03988109
GO:0080032	15	6716	11	11435	with_enriched	0.04023203
GO:0001708	2	6729	15	11431	without_enriched	0.04125691
GO:0009871	2	6729	15	11431	without_enriched	0.04125691

GO:0071323	2	6729	15	11431	without_enriched	0.04125691
GO:0004712	14	6717	44	11402	without_enriched	0.04170793
GO:0004568	6	6725	25	11421	without_enriched	0.04172692
GO:0009723	57	6674	134	11312	without_enriched	0.04178909
GO:0015991	6	6725	26	11420	without_enriched	0.04203495
GO:0008654	8	6723	30	11416	without_enriched	0.04371677
GO:0030170	49	6682	118	11328	without_enriched	0.0438461
GO:0015937	7	6724	3	11443	with_enriched	0.04553694
GO:0016021	1500	5231	2698	8748	without_enriched	0.04701501
GO:0016324	5	6726	22	11424	without_enriched	0.04752405
GO:0008171	11	6720	7	11439	with_enriched	0.04824801
GO:0009718	11	6720	7	11439	with_enriched	0.04824801
GO:0050662	21	6710	59	11387	without_enriched	0.04847372
GO:0010333	13	6718	41	11405	without_enriched	0.04897725
GO:0015031	91	6640	199	11247	without_enriched	0.04962881

Works Cited

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15):2114-2120. doi:10.1093/bioinformatics/btu170

Castelletti S, Tuberosa R, Pindo M, Salvi S (2014) A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL *Vgt1*. G3 (Bethesda) 4 (5):805-812. doi:10.1534/g3.114.010686

Chen J, Hu Q, Zhang Y, Lu C, Kuang H (2014) P-MITE: a database for plant miniature inverted-repeat transposable elements. Nucleic Acids Res 42 (Database issue):D1176-D1181. doi:10.1093/nar/gkt1000

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE (2012) Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. Science 338 (6111):1206-1209

Dai S, Hou J, Long Y, Wang J, Li C, Xiao Q, Jiang X, Zou X, Zou J, Meng J (2015) Widespread and evolutionary analysis of a MITE family Monkey King in Brassicaceae. BMC Plant Biol 15:149. doi:10.1186/s12870-015-0490-9

Devi KD, Punyarani K, Singh NS, Devi HS (2013) An efficient protocol for total DNA extraction from the members of order Zingiberales- suitable for diverse PCR based downstream applications. SpringerPlus 2:669. doi:10.1186/2193-1801-2-669

Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JD (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell 10 (11):1915-1925

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329+

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103-107. doi:https://doi.org/10.1016/0168-9525(89)90039-5

Guo C, Spinelli M, Ye C, Li QQ, Liang C (2017) Genome-wide comparative analysis of miniature inverted repeat transposable elements in 19 *Arabidopsis thaliana* ecotype accessions. Sci Rep 7:2634. doi:10.1038/s41598-017-02855-1

Hardigan MA, Crisovan E, Hamiltion JP, Kim J, Laimbeer P, Leisner CP, Manrique-Carpintero NC, Newton L, Pham GM, Vaillancourt B, Yang X, Zeng Z, Douches D, Jiang J, Veilleux RE, Buell CR (2016) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated Solanum tuberosum. Plant Cell. doi:10.1105/tpc.15.00538

Helaers R, Milinkovitch MC (2010) MetaPIGA v2. 0: maximum likelihood large phylogeny estimation using the metapopulation genetic algorithm and other stochastic heuristics. BMC Bioinformatics 11 (1):379

International Human Genome Sequencing C (2001) Initial sequencing and analysis of the human genome. Nature 409:860. doi:10.1038/35057062

https://www.nature.com/articles/35057062#supplementary-information

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. Nature 421:163. doi:10.1038/nature01214

https://www.nature.com/articles/nature01214#supplementary-information

Keane TM, Wong K, Adams DJ (2013) RetroSeq: transposable element discovery from next-generation sequencing data. Bioinformatics 29 (3):389-390. doi:10.1093/bioinformatics/bts697

Kikuchi K, Terauchi K, Wada M, Hirano H-Y (2003) The plant MITE *mPing* is mobilized in anther culture. Nature 421:167. doi:10.1038/nature01218

https://www.nature.com/articles/nature01218#supplementary-information

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12 (4):357-360. doi:10.1038/nmeth.3317

Knox AK, Dhillon T, Cheng H, Tondelli A, Pecchioni N, Stockinger EJ (2010) *CBF* gene copy number variation at *Frost Resistance-2* is associated with levels of freezing tolerance in temperate-climate cereals. Theor Appl Genet 121 (1):21-35

Kondrashov AS (1982) Selection against harmful mutations in large sexual and asexual populations. Genet Res 40 (3):325-332

Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, Jiang J, Buell CR, Baker B (2009) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. Genome Res 19 (1):42-56. doi:10.1101/gr.078196.108

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M (2011) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40 (D1):D1202-D1210

Lepetit D, Pasquet S, Olive M, Theze N, Thiebaud P (2000) Glider and Vision: two new families of miniature inverted-repeat transposable elements in *Xenopus laevis* genome. Genetica 108 (2):163-169

Li W, Godzik A (2006) CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22 (13):1658-1659. doi:10.1093/bioinformatics/btl158

Lin X, Long L, Shan X, Zhang S, Shen S, Liu B (2006) In planta mobilization of mPing and its putative autonomous element Pong in rice by hydrostatic pressurization. J Exp Bot 57 (10):2313-2323

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15 (12):550. doi:10.1186/s13059-014-0550-8

Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H (2012) Miniature Inverted–Repeat Transposable Elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. Mol Biol Evol 29 (3):1005-1017. doi:10.1093/molbev/msr282

Mao H, Wang H, Liu S, Li Z, Yang X, Yan J, Li J, Tran L-SP, Qin F (2015) A transposable element in a *NAC* gene is associated with drought tolerance in maize seedlings. Nat Commun 6:8326. doi:10.1038/ncomms9326

https://www.nature.com/articles/ncomms9326#supplementary-information

Marand AP, Jansky SH, Zhao H, Leisner CP, Zhu X, Zeng Z, Crisovan E, Newton L, Hamernik AJ, Veilleux RE, Buell CR, Jiang J (2017) Meiotic crossovers are associated with open chromatin and enriched with *Stowaway* transposons in potato. Genome Biol 18 (1):203. doi:10.1186/s13059-017-1326-8

Momose M, Abe Y, Ozeki Y (2010) Miniature inverted-repeat transposable elements of *Stowaway* are active in potato. Genetics 186 (1):59-66. doi:10.1534/genetics.110.117606

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461 (7267):1130-1134. doi:10.1038/nature08479

Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T (2003) Mobilization of a transposon in the rice genome. Nature 421 (6919):170

Nelson MG, Linheiro RS, Bergman CM (2017) McClintock: An integrated pipeline for detecting transposable element insertions in whole-genome shotgun sequencing data. G3 7 (8):2763-2778. doi:10.1534/g3.117.043893

Ngezahayo F, Xu C, Wang H, Jiang L, Pang J, Liu B (2009) Tissue culture-induced transpositional activity of mPing is correlated with cytosine methylation in rice. BMC Plant Biol 9 (1):91. doi:10.1186/1471-2229-9-91

Oki N, Yano K, Okumoto Y, Tsukiyama T, Teraishi M, Tanisaka T (2008) A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, Oryza sativa ssp. japonica. Genes Genet Syst 83 (4):321-329

Platzer A, Nizhynska V, Long Q (2012) TE-Locate: A tool to locate and group transposable element occurrences using paired-end next-generation sequencing data. Biology 1 (2):395-410. doi:10.3390/biology1020395

Sanseverino W, Hénaff E, Vives C, Pinosio S, Burgos-Paz W, Morgante M, Ramos-Onsins SE, Garcia-Mas J, Casacuberta JM (2015) Transposon Insertions, Structural Variations, and SNPs Contribute to the Evolution of the Melon Genome. Mol Biol Evol 32 (10):2760-2774. doi:10.1093/molbev/msv152

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J,

Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science 326 (5956):1112-1115. doi:10.1126/science.1178534

Shen J, Liu J, Xie K, Xing F, Xiong F, Xiao J, Li X, Xiong L (2017) Translational repression by a miniature inverted-repeat transposable element in the 3′ untranslated region. Nat Commun 8:14651. doi:10.1038/ncomms14651

https://www.nature.com/articles/ncomms14651#supplementary-information

Van Valen L (1973) A new evolutionary law. Evolutionary Theory

Wessler SR (2006) Transposable elements and the evolution of eukaryotic genomes. Proc Nat Acad Sci USA 103 (47):17600-17601. doi:10.1073/pnas.0607612103

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8 (12):973-982. doi:10.1038/nrg2165

Wu F, Tanksley SD (2010) Chromosomal evolution in the plant family Solanaceae. BMC Genomics 11 (1):182. doi:10.1186/1471-2164-11-182

Yang G, Dong J, Chandrasekharan M, Hall T (2001) *Kiddo*, a new transposable element family closely associated with rice genes. Mol Genet Genomics 266 (3):417-424. doi:10.1007/s004380100530

Yang G, Hall TC (2003) *MDM-1* and *MDM-2*: two *Mutator*-derived MITE families in rice. J Mol Evol 56 (3):255-264. doi:10.1007/s00239-002-2397-y

Yang G, Lee Y-H, Jiang Y, Shi X, Kertbundit S, Hall TC (2005) A two-edged role for the transposable element *Kiddo* in the rice *ubiquitin2* promoter. Plant Cell 17 (5):1559-1568. doi:10.1105/tpc.104.030528

Yongsheng Y, Yuman Z, Kun Y, Zongxiu S, Yaping F, Xiaoying C, Rongxiang F (2011) Small RNAs from MITE-derived stem-loop precursors regulate abscisic acid signaling and abiotic stress responses in rice. The Plant Journal 65 (5):820-828. doi:doi:10.1111/j.1365-313X.2010.04467.x

Zhuang J, Wang J, Theurkauf W, Weng Z (2014) TEMP: a computational method for analyzing transposable element polymorphism in populations. Nucleic Acids Res 42 (11):6826-6838. doi:10.1093/nar/gku323

Chapter 4: Structural variations at the *AN2* locus in potato contribute to differences in floral anthocyanin production

Abstract

Anthocyanins are a pigmented group of secondary metabolites which are produced via the flavonoid biosynthetic pathway and play important roles in plant stress response and consumer preference. Using a wide swath of germplasm, including a segregating population, inbred lines, and a panel of homozygous monoploids, we show that a homolog of the *AN2* locus is the master regulator of anthocyanin production in potato corollas. Bulk-segregant RNA-seq revealed the specific biosynthetic genes activated by *AN2* as well as expression differences in ripening, senescence, and primary metabolic pathways. Sequence analysis revealed a duplication in the *AN2* locus to be closely associated with gene expression, while a transgenic complementation approach indicated that expression difference in *AN2* is likely attributable to nearby genetic elements rather than deficiencies in either promoter or coding elements. Taken together, this research provides insight into the regulation of anthocyanin biosynthesis in potato while also highlighting how the dynamic nature of the *AN2* locus may affect expression more than sequence specific-variation.

Introduction

Anthocyanins are a group of secondary plant metabolites which provide a variety of benefits to the plant such as pollinator attraction and resistance to abiotic pressures including cold, UV, and oxidative stress (Schulz et al. 2016; Chalker-Scott 1999; Merzlyak and Chivkunova 2000; Steyn et al. 2002). Agronomically, they play to consumer preference in fruit and vegetables while also providing a source of dietary antioxidants which may confer health benefits, such as a reduction in cardiovascular diseases and diagnostic indicators for oxidative stress (de Pascual-Teresa et al. 2010; Alvarez-Suarez et al. 2014; Khoo et al. 2017). In potato, anthocyanins can accumulate in a wide range of tissues including leaves, stems, flowers, and, of course, tuber skin and flesh. Thus the study of anthocyanins in potato is important for its

implications in consumer health and appeal as well as the potential for stress mitigation which could affect important agronomic traits such as yield and storage quality.

Anthocyanins, glycosides (linked to sugars via glycosidic bonds) of anthocyanidins, are a broad class of flavonoids which may vary in states and types of glycosylation. The anthocyanidin

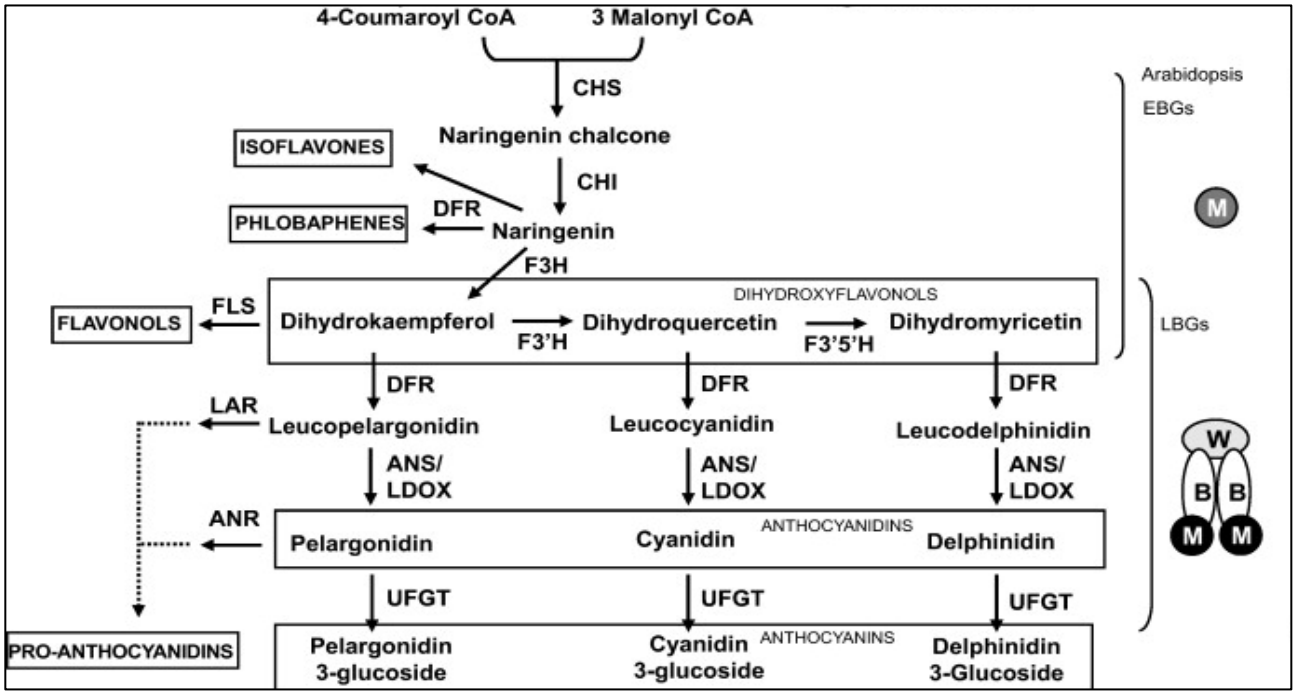


Figure 4.18: A simplified scheme of anthocyanin synthesis and regulation adapted from Petroni & Tonelli, 2011, with permission from Elsevier. Enzyme abbreviations are as follows: Chalcone synthase (*CHS*), chalcone isomerase (*CHI*), flavonol synthase (*FLS*), flavonoid 3 hydroxylase (*F3H*), flavonoid 3' hydroxylase (*F3'H*), flavonoid 3'5' hydroxylase (*F3'5'H*), dihydroflavonol reductase (*DFR*) anthocyanidin synthase/leucoanthocyanin oxidase (*ANS/LDOX*), anthocyanidin reductase (*ANR*), UDP- flavonol glucosyltransferase (*UFGT*). Transcription factor abbreviations are as follows: Myb transcription factor (M), WD 40 (W), basic helix-loop-helix (B). EBGs and LBGs refer to early and late biosynthetic genes, respectively.

precursors are produced through the phenylpropanoid pathway, which begins with the lysis of phenylalanine, before proceeding through the flavonoid pathway by way of chalcone (Tanaka et al. 2008; Petroni and Tonelli 2011). The remaining biosynthetic steps, and the model of their regulation, are summarized in **Figure 4.1.** Briefly, chalcone is converted to dihydrokaempferol

by a series of steps (chalcone isomerase; *CHI*, flavonoid 3-hydroxylase, *F3H*). Next, modifications may be made (by flavonoid 3' hydroxylase, *F3'H,* and flavonoid 3'-5' hydroxylase, *F3'5'H*) to alter the specific flavonoid precursor which will ultimately determine the type, and color, of anthocyanin it will become. Finally anthocyanin synthase and various glucosyl transferases complete the pathway, giving the anthocyanin its specific identity. While the structural enzymes are well conserved, their regulation differs by clade. In solanaceous species the early biosynthetic genes (EBGs), those leading to production of flavonoids, are regulated by R2R3 MYBs, which are often tissue specific. The late biosynthetic genes (LBGs), which include modification to the flavonoids and anthocyanins themselves, are regulated by a heterocomplex of MYB, basic helix loop helix (bHLH), and WD40 transcription factors (Spelt et al. 2000; de Vetten et al. 1997; Quattrocchio et al. 1999; de Pascual-Teresa et al. 2010; Feller et al. 2011).

In petunia flowers, which are a model organ for anthocyanin production, the EBGs are initiated by a MYB denoted as *PhAN2* which also controls expression of the bHLH, *PhAN1*, with which it regulates the LBGs (Quattrocchio et al. 1999; Spelt et al. 2000). In potato tubers, separate loci have been proposed to control tuber skin pigmentation: D, P and R, all of which are dominant. The D locus is required for any skin pigmentation whatsoever, while P is epistatic to R and they control purple and red skin color, respectively (Dodds and Long 1955). All three have been mapped and cloned. Subsequent analysis has shown that D encodes a MYB transcription factor on chromosome 10 with high homology to *AN2* in *Petunia* spp. (Jung et al. 2009). R encodes the potato homolog of *DFR* on chromosome 2 (Zhang et al. 2009a) (De Jong et al. 2004). P has been mapped to chromosome 11 and codes for *F3'5'H* (Jung et al. 2005). However the control of tuber flesh pigmentation is not as clear. A mapping study by Zhang et al.

(2009b) showed that a homolog of *AN2*, deemed *StAN1* but not homologous to the *PhAN1*, which is a bHLH transcription factor, played a significant role in, but was not wholly responsible for, anthocyanin accumulation in tuber flesh.

Similarly to tubers, three potato flower loci have been described: D, F, and P (van Eck et al. 1993). Unfortunately, the nomenclature is perplexing, as the D locus of flowers is responsible for red anthocyanin accumulation and does not correspond to the D locus of tubers which regulates all tuber anthocyanin production. Instead, the floral D maps to chromosome 2 and appears to correspond to the tuber R locus, encoding the enzyme *DFR*. The floral P locus, thankfully, is responsible for purple anthocyanin accumulation, localizes to chromosome 11 and seems to match the tuber P locus, which represents the *F3'5'H* gene. The same study by van Eck et al. (1993) used RFLPs to map the F, required for any floral pigment accumulation, locus to chromosome 10, nearby the tuber D locus which performs the same function of conferring tissue specific regulation. Therefore it is plausible that there are multiple homologs of *AN2* which arose via duplication on chromosome 10 on and dictate pigment accumulation in separate tissues independently.

As the genetic mechanisms of anthocyanin production have received considerably less investigation in potato flowers than in tubers we set out to investigate the nuances of the floral F locus. Using a segregating diploid population in conjunction with inbred and homozygous individuals derived from that population and an unrelated panel of homozygous monoploids, we identify a separate *AN2* homolog that underlies the F locus. An RNA-seq approach reveals the regulatory cascade caused by this homolog while analysis of the locus itself indicates implicates duplications that may underlie differences in gene expression and floral phenotypes.

Methods

DMxRH population, F_1-derived monoploids, and monoploid panel

The various sources of germplasms used in this study are displayed in **Figure 4.2**.

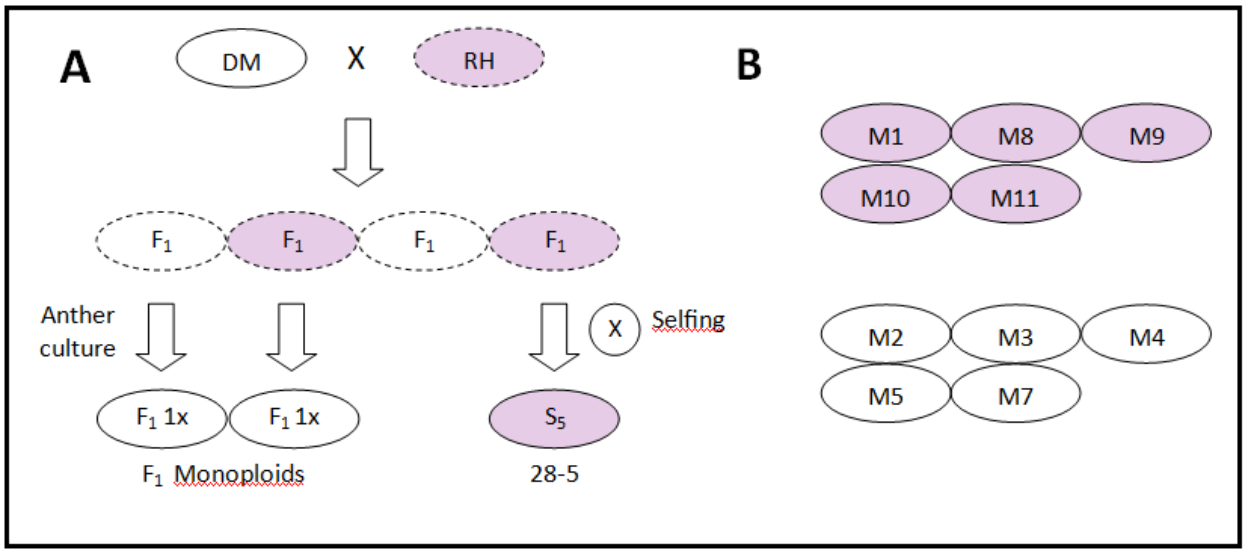


Figure 4.2: Overview of the germplasm utilized in this study. Each oval's color corresponds to genotype flower color, solid outlines indicate (near) homozygosity while dashed indicates heterozygous background. A) DMxRH population and its derivatives. B) The monoploid panel published in Hardigan et al. (2016)

The DMxRH population derives from a diploid cross between a white-flowered homozygous individual doubled monoploid (DM), which also happens to be the genotype that provides the reference potato genome, and a heterozygous purple-flowered individual, RH (PGSC 2011). The resulting F_1 population, comprised of 95 individuals, was grown in growth chambers at Virginia Tech (16 h photoperiod, 250 µE m^{-2} s^{-1}, with 22°C days and 18°C nights) and screened for flower color. DNA was extracted from immature leaflets using the CTAB method (Doyle 1991). The resulting DNA was quantified by NanoDrop™ spectrophotometer for genotyping on the Infinium 8303 Potato SNP Array using the Illumina iScan reader with alleles called in GenomeStudio. Immature flower buds of DMxRH F_1 plants were placed in anther culture and

regenerated plants screened by flow cytometry to identify monoploid individuals according to Teparkum and Veilleux (1998). Despite using both white and purple-flowered F_1 plants as anther donors, we obtained only white-flowered DMxRH monoploids. Whole genome sequencing was performed on these as well as on a monoploid panel consisting of ten independently derived individuals from anther culture as previously described (Hardigan et al. 2016b)**.** Lastly, as the DMxRH population has been demonstrated to segregate for self-fertility, an S_5 individual fixed for purple flowers, 28-5, was obtained through successive rounds of inbreeding as described in Peterson et al. (2016). While genomic sequencing was not performed on this S_5 plant, it provided a source for DNA extraction from an individual derived from DMxRH and homozygous for the purple allele. Whole genome sequencing alignments employed from Hardigan et al. (2016) were examined for variation at the *AN2* locus using Integrated Genome Viewer (Robinson et al. 2011)**.** Sequence and phylogenetic analysis was performed using the Lasergene suite (DNASTAR, Inc., Madison, WI) and sequences of *StAN1 (*AGC31676) and *PhAN2 (*A4GRV2) were retrieved from the UniProt database (The UniProt Consortium 2017).

Bulk Segregant RNA-seq

Four bulk samples were collected (2 purple, 2 white) by combining corollas of a single flower from each of ten separate F_1 individuals, sampling ten different individuals for each bulk of the same color. Bulk samples were immediately placed in liquid nitrogen prior to a hybrid trizol/Quiagen RNeasy mini kit extraction (www.microarray.adelaide.edu.au/protocols/rna/). The samples then underwent a DNase treatment (Ambion Catalog # AM 1906) followed by analysis on a Bioanalyzer to assess RNA quality. RIN values ranged from 10.0 to 9.9. Library preparation and sequencing were carried out by the Genomics Research Laboratory at the Biocomplexity Institute at Virginia Tech. The Illumina Hi-Seq 2500 platform was used to obtain 100 bp paired

end genomic sequencing reads with an insert size of ~200 bp wherein the four samples were barcoded and run in a single lane. All read quality control and subsequent analysis were performed using the CLC Genomics Workbench 7.5. Reads were trimmed to remove adapters and 13 bp from the 5' end to mitigate bias in the PCR amplification step of library preparation. Furthermore, all reads were trimmed to remove low quality base calls (phred <20) and short reads (<40 bp) were discarded. Reads were then mapped to the PGSC DM genome v 4.03 (Sharma et al. 2013) using the following parameters: mismatch cost = 2; insertion cost = 3; deletion cost = 3; length fraction = 0.9; similarity fraction = 0.8; Max number of hits per read = 10. Expression values were reported in RPKM (reads per thousand bp of gene per million total reads).

After mapping, the bulks were used as reps in an unpaired empirical analysis of differential gene expression and p-values were corrected for false discovery rate (FDR). Following this, the genes were filtered based on an FDR p-value < 0.05 and an absolute value fold-change > 2.0. This left 78 annotated genes classified as differentially expressed.

Transgenic complementation

Vector Construction

All constructs were created using the Pcambia 1305.1 vector, excising the 35s promoter and GUS sequences by restriction digest (BamHI and BstEII) and replacing them with the relevant promoter/gene combination. Gene, CDS and promoter sequences for transgenic complementation were cloned from a variety of sources. The *AN2* expressed cDNA sequence was cloned directly from the purple bulks used for RNA-sequencing. Genomic sequences of the purple haplotype promoter and gene sequences were cloned from an inbred S_5 DMxRH

individual (28-5) which is fixed for purple flowers (determined by lack of flower color segregation in the S_6 and S_7 derivatives) and was derived from inbred families described previously (Peterson et al. 2016)**.** Since no sequence data were available for 28-5, primers were designed to bind regions lacking polymorphism within the purple-flowered individuals of the monoploid panel described previously. The following primers were used to engineer compatible restrictions sites onto the promoter (BamHI) and coding (BstEII) sequences while a conserved Xba1 site located in the first exon was used to join promoters and coding sequences: An2pro2 (AATTAT**GgaTcc**TCTTGGTTTTTCTTTTCATATTTATAC), An2proXba1 (ACCAGCTCTAGAAGGAACAAGATGCC), An2CDSExon1 (TTGGGAGTGAGAAAAGGTTCATGG), and AN2CDSBstEII (ATATTA**ggtGacc**CCCTAGTACA AGTAGTAGTACAATACC). PCR conditions were as follows: 35 cycles of 98°C for 15 sec; 50°C for 30 sec; 68°C for 3 min; followed by a single 5 min extension at 68°C. Each 20 µl reaction contained 1x Primestar GXL buffer, 0.5 units Primerstar GXL taq (Takara # R050A), 0.2 µM each primer, 50 µM dNTPs, 50 ng DNA. PCR products were purified by gel purification using the NucleoSpin® Gel and PCR Clean-up kit (Machery-Nagel, Neumann Neander Str. 6-8 Duren, Germany). Purified products were verified by Sanger sequencing performed at the Biocomplexity institute at Virginia Tech. Verified products were ligated (T4 DNA Ligase NEB # M0202S) into the pJET 1.2 cloning vector (Thermo Scientific CloneJET PCR Cloning Kit #K1232) before transformation of α-Select Chemically Competent Cells (Bioline, BIO-85025) using the heat-shock method per manufacturer's instructions. Transformed cells were plated on LB medium (Acros # 611875000) with 100 ug/mL ampicillin selection and incubated overnight at 37°C. Resulting colonies were screened with the pJET1.2 forward (CGACTCACTATAGGGAGAGCGGC) and reverse

(AAGAACATCGATTTTCCATGGCAG) sequencing primers. Positive colonies were used to inoculate 5 ml of LB broth with 100 ug/mL ampicillin which were grown overnight at 37°C prior to plasmid extraction (Monarch Plasmid Miniprep Kit, NEB #T1010G). Extracted plasmids were digested with the appropriate enzyme (BamHI and XbaI for promoters; BstEII and XbaI for coding sequences; BamHI and BstEII for pCambia 1305.1 binary vector) prior to three piece ligation with a 1:3:3 ratio of vector:promoter:CDS. Ligations were again transformed into α-Select Chemically Competent Cells prior to an overnight incubation at 37°C and plasmid preparation. Plasmids were introduced to ElectroMAX *Agrobacterium tumefaciens* LBA4404 (Invitrogen #18313-015) cells through electroporation per manufacturer's instruction and plated on YM medium plates (20 g/l agar, 10 g/l glucose, 3 g/l yeast extract, 3 g/l malt extract, 5 g/l peptone, pH 6.2) with 100 µg/ml streptomycin and 50 µg/ml kanamycin.

Plant Transformation

In addition to DM, a white-flowered DMxRH F_1 individual, DMxRH 171, was selected for transformation to guard against the possibility of multiple defective loci due to inbreeding which may have contributed to DM's white-flowered phenotype; such loci could simply have been masked by either RH haplotype in the F_1 population. Plant transformations were carried out as described in Rooke and Lindsey (1998) with minor modification. The resulting shoots were allowed to root in basal MS media 4.43 g l^{-1} MS salts, 3% sucrose, 7 g l^{-1} agar, pH 5.7-5.8). Once established, the media was washed off and the shoots were transferred to peat pellets prior to being placed in 3.8 L pots on greenhouse ground beds for phenotyping. The greenhouse plants were grown with a 16 h photoperiod provided by supplemental lighting to encourage flowering and fertilizer was applied as needed.

Results

DMxRH SNP chip

The DMxRH F_1 population was derived from a cross between a white-flowered homozygous individual, DM, and a purple-flowered heterozygous Group Phureja individual, RH. The F_1 generation of DMxRH displayed an approximate segregation ratio of 1:1 for flower color, with all individuals being either white or purple-flowered. The 1:1 ratio (51:44; purple:white; χ^2=-0.38, p=0.54), and complete penetrance of the flower color phenotype suggested a single gene segregating from the heterozygous purple-flowered RH was the likely cause. Analysis of the SNP-chip data published by Peterson et al. (2016) indicated SNPs significantly linked to the phenotype on the distal end of chromosome 10, which is consistent with previous literature regarding the F locus for flower color (**FIGURE 4.3)**.

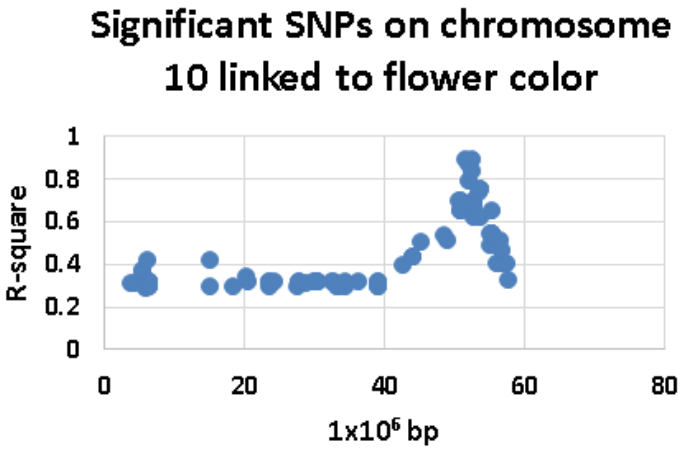


Figure 4.19: Significant SNPs linked to flower color in DMxRH are located on the distal end of chromosome 10.

Bulk-segregant RNA-seq

Bulk-segregant analysis of gene expression in the purple and white bulks uncovered 78 genes with a p-value <0.05 after FDR correction (**Supplementary Table 4.1)**. Among these genes, 13 were located on the distal end of chromosome 10, where the SNP-chip analysis indicated a region of high significance and the flower color locus is known to exist. One gene, *AN2*, which is an R2R3 MYB transcription factor, stood out with an expression level over 500 fold higher in the purple bulks. By contrast, the next highest gene-level expression difference was ~115-fold for a gene annotated as major latex protein, the relevance of which is not immediately clear. As a homolog of the petunia *AN2* which is known to regulate anthocyanin accumulation in petunia, this became our prime candidate gene. Furthermore, we observed that only a single AN2 sequence was expressed in the purple bulks. This fits with the observed 1:1 ratio in the F_1 generation: a single allele from RH is expressed (designated RH_p), while the other RH allele (RH_w) and the DM allele (DM) lack expression entirely.

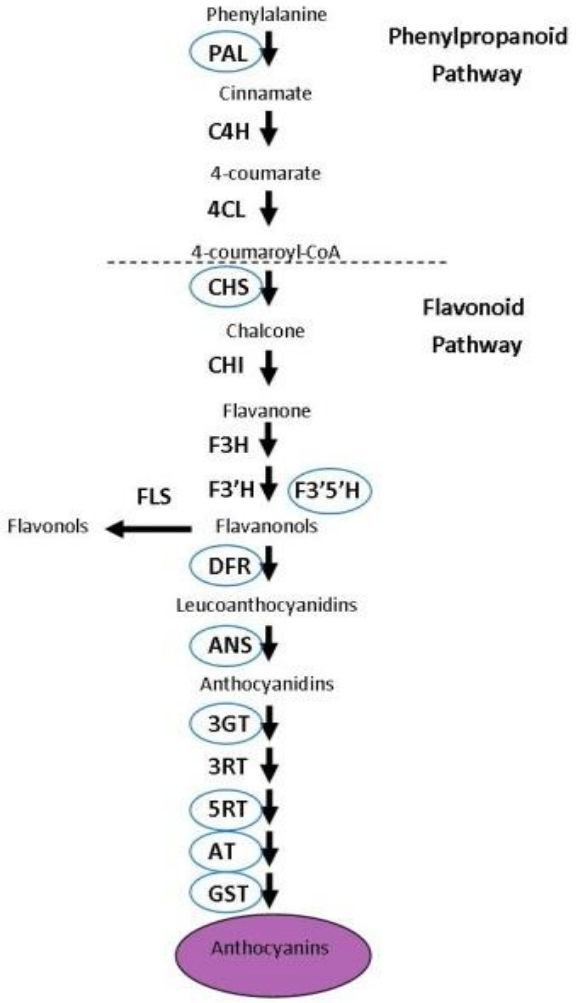


Figure 4.20: Steps of anthocyanin biosynthesis Circled genes are differentially upregulated in purple DMxRH flower bulks. Gene symbols are: Chalcone synthase (*CHS*), chalcone isomerase (*CHI*), flavonoid 3 hydroxylase (*F3H*), flavonoid 3' hydroxylase (*F3'H*), flavonoid 3'5' hydroxylase (*F3'5'H*), dihydroflavonol reductase (*DFR*) anthocyanidin synthase (*ANS*), UDP- flavonol glucosyltransferase (*UFGT*), anthocyanin 3-glucosyl transferase (*3GT*), anthocyanin 3-rhamnosyl transferase (*3RT*), anthocyanin 5-glucosyl transferase (*5GT*), anthocyanin acyl transferase (*AT*), glutathione-S-transferase (*GST*), anthocyanin methyl transferase (*MT*), flavonoid synthase (*FLS*).

In addition to *AN2*, 16 genes within the phenylpropanoid and anthocynanin synthesis pathways were found to be differentially regulated between the bulks. These genes are highlighted in **Supplementary Table 4.1** and those specific to the anthocyanin pathway are displayed in **Figure 4.4**. Additionally, a number of genes associated with fruit ripening were found to be up-regulated in the purple bulks, including four pectate lyases and three pectinesterases. Somewhat surprisingly, we observed up-regulation of a few primary metabolism genes in the white bulks, such as photosystem subunits, Ribulose bisphosphate carboxylase/oxygenase, and fructose-1,6-bisphosphatase.

Sequence analysis

Having identified *AN2* as the likely cause of differences in anthocyanin accumulation within the DMxRH population we performed sequence analysis to ascertain any differences between the alleles which may explain the stark expression contrast. While we did not have genomic sequencing data of any of the purple F_1 individuals, we did have access to sequencing from a variety of other homozygous or highly inbred sources which we used to inform our investigation (**Figure 4.2)**.

These include: a homozygous white-flowered monoploid derived from a white-flowered F_1 which represents the white-flowered RH haplotype (RH_w); DM, which provides the other white-flowered haplotype (DM) in our population; and a panel of ten homozygous, monoploid, Group Phureja potatoes, including five white and five purple-flowered individuals. Finally we had access to DNA, but not NGS, from a DMxRH-derived S_5 individual which is fixed for purple flowers and provides sequence of the purple RH haplotype (RH_p). This assortment of germplasm allowed us to examine the sequences of each haplotype (RH_w, RH_p, DM) of the DMxRH

population while also exploring the sequences of more-distantly related haplotypes provided by the monoploid panel.

We observed that the RH_p haplotype, as well as all of the purple-flowered individuals in the panel, contained a partial duplication of the *AN2* locus which was not present in the RH_w or DM haplotypes nor four of the five white flowered individuals in the monoploid panel, with M4 being the sole exception (**Figure 4.5 ; Figure 4.6)**. Furthermore, M5 and M6 appear to have an additional duplication of the promoter MITE. These two paralogs comprising the RH_p haplotype were designated as $RH_{p\text{-}E}$ and $RH_{p\text{-}NE}$, for the expressed and non-expressed paralog, respectively. Further examination of the *AN2* locus revealed that the reference genome, DM, harbored miniature transposable elements (MITEs) in both the promoter and first intron (both Mutator; DTM) of its *AN2* allele, both of which were absent from the RH_w and RH_p haplotypes; instead a MITE of a separate superfamily (Tc1/Mariner; DTT) was inserted within the first intron of the expressed paralog of the RH_p haplotype. Additionally, while the promoter sequences of RH_w and both RH_p paralogs exhibited high homology to the reference proximal to the reference MITE, the promoter region of one of the RH_p paralogs completely lacked homology to that of DM's distal to the promoter MITE, which manifests as a drop in genomic sequencing coverage in that region, apparent in the purple individuals of the monoploid panel, as well as a single band in RH_p PCR reactions spanning this MITE, rather than the two bands amplified from the CDS of RH_p.

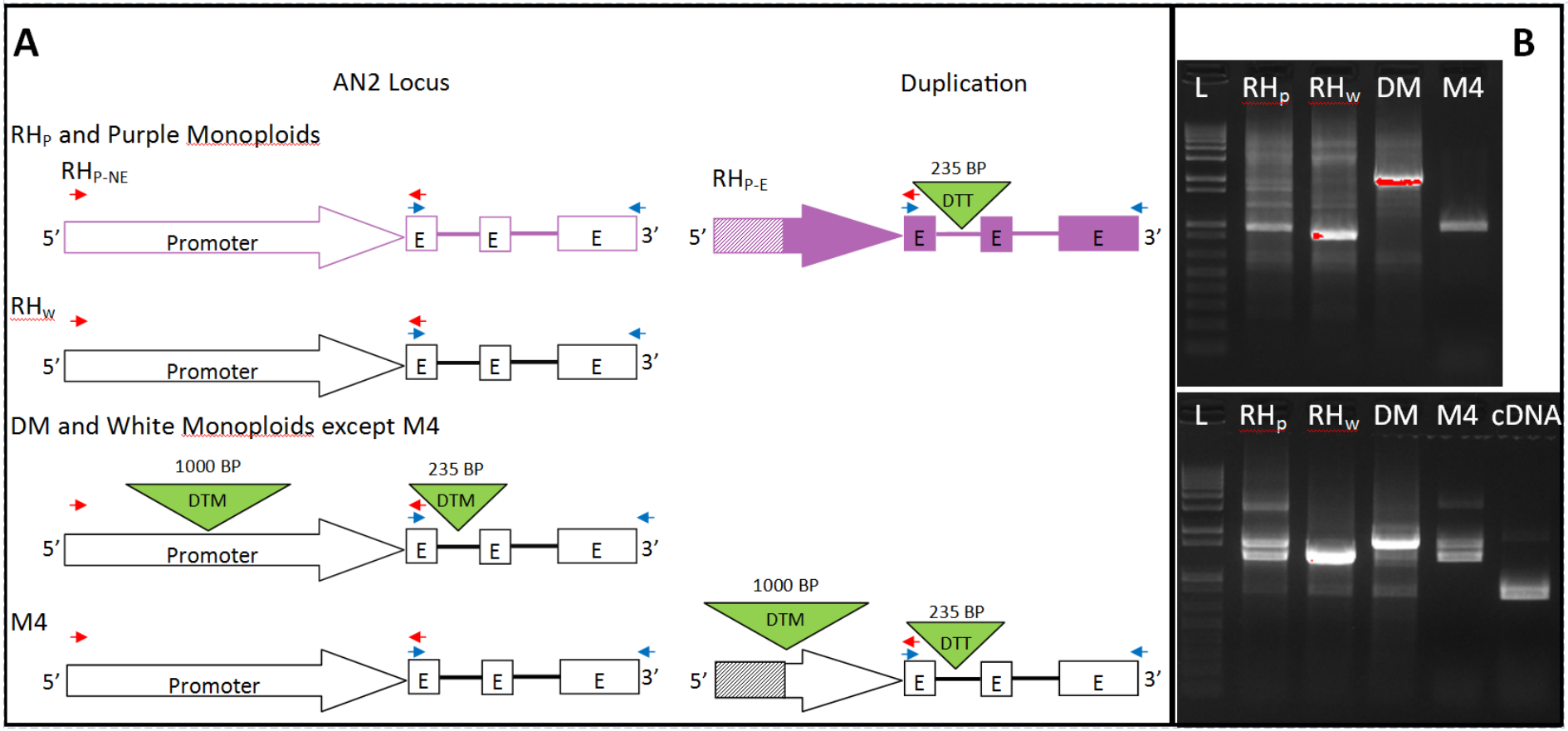


Figure 4.21: A) Scheme of AN2 locus in RHP, RHw, DM, and Monoploid Panel haplotypes. RHP includes both the non-expressed (RHP-NE) and expressed (RHP-E) paralogs. Arrows represent promoters while boxes and lines represent exons (E) and introns respectively. Green triangles depict MITEs inserted into either promoter or intronic sequences labeled by superfamily. Excluding transposons, shapes are filled to indicate expression in native germplasm while unfilled shapes indicate lack of expression. Striped fill indicates unknown sequence which lacks homology to the reference genome. Locations for primers used in cloning and PCR are displayed with red (promoter) and blue (CDS) arrows. B) PCR amplification of promoter (top) and coding sequences (bottom) of AN2 haplotypes observed in this study. L denotes Invtrogen 1kb+ ladder. Note: Differences in haplotypes ideograms in A are reflected in band size and number in B.

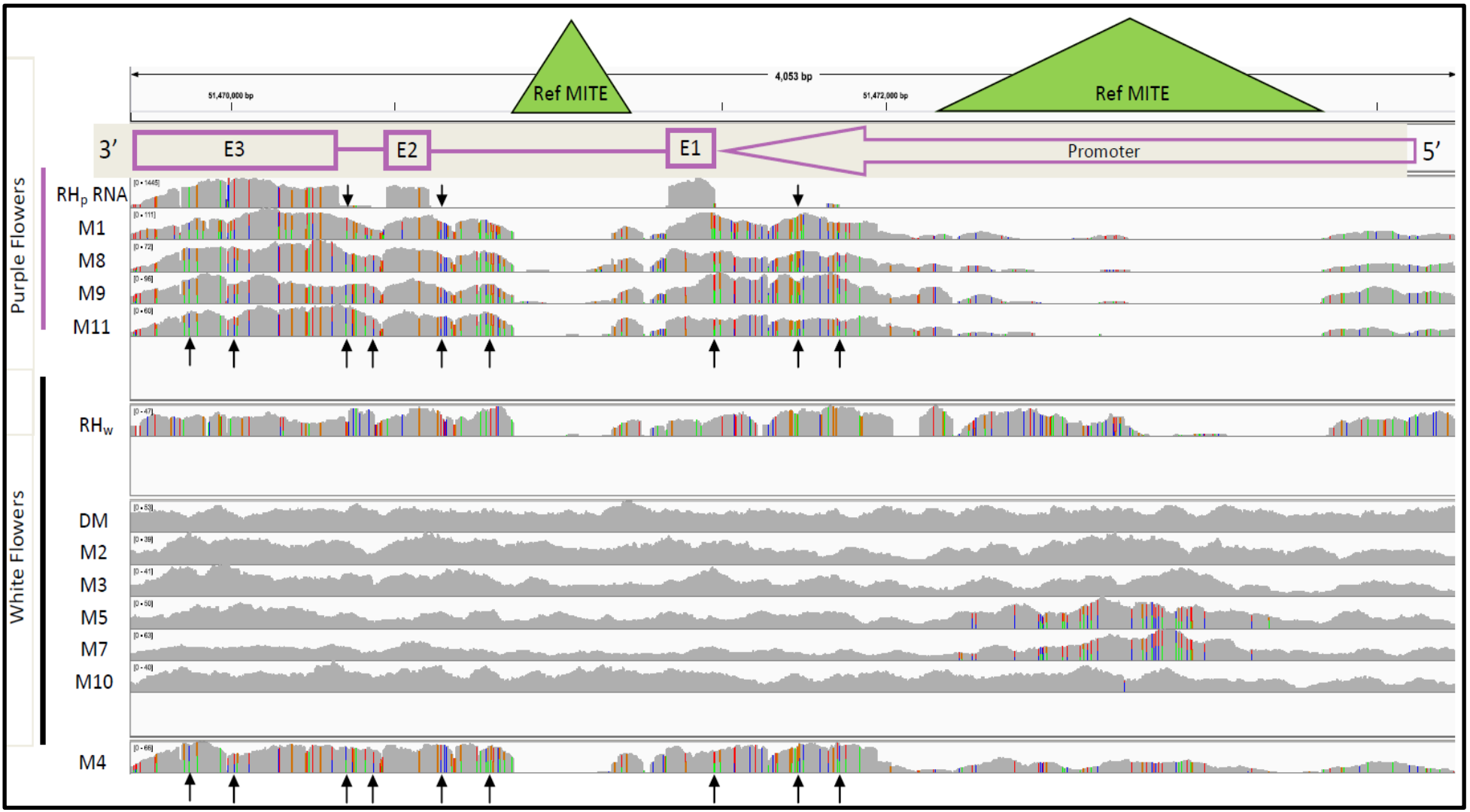


Figure 4.22: Sequencing coverage of the DMxRH and monoploid haplotypes sorted by flower color. RH_p RNA represents the RNA-seq from the purple DMxRH bulk while all other tracks are whole genome sequencing of their respective haplotypes. The gene structure of the negative strand, *AN2*, is displayed with a purple arrow representing the promoter and boxes representing each exon. Copy number variation is apparent in all purple tracks except for RH_p RNA (as only one paralog is expressed) as well as M4. This copy number variation manifests as multicolored bars which would otherwise be considered two separate SNP states in a heterozygous background and is indicated by black arrows. The locations of the two reference MITEs are displayed with green triangles.

A closer inspection of the coding sequences of each of the DM, RH_w, $RH_{p\text{-}NE}$, and $RH_{p\text{-}E}$ homologs indicated that all are theoretically functional, lacking premature stop codons or ORF shifts despite various point mutations and indels (**Figure 4.7)**. A comparison with the more intensively studied homologs, petunia *PhAN2,* and potato *StAN1*, which controls tuber color, revealed that the two RH_p paralogs were more closely related to the DM allele than the RH_w allele while all four were only distantly related to the orthologous *PhAN2* or paralogous *StAN1*. Considering this apparent lack of obvious functional deficiency in the coding sequences of each of the *AN2* alleles present within the DMxRH population, we decided to proceed with transgenic complementation of DM and a white-flowered F_1 with the hypothesis that variations in the promoter sequences precipitated the various expression levels: The MITE present within the promoter sequence of DM, and observable in the coverage of all white-flowered individuals of the panel, seemed to co-segregate perfectly with the white-flowered phenotype with the exception of the RH_w haplotype whose lack of expression could perhaps be due to another disruption within the promoter sequence. Furthermore, we hypothesized that the $RH_{p\text{-}NE}$ paralog lacked expression due to an incomplete duplication of its promoter, and that the $RH_{p\text{-}E}$ paralog retained the promoter which had greater homology to the DM allele distal to the MITE insertion.

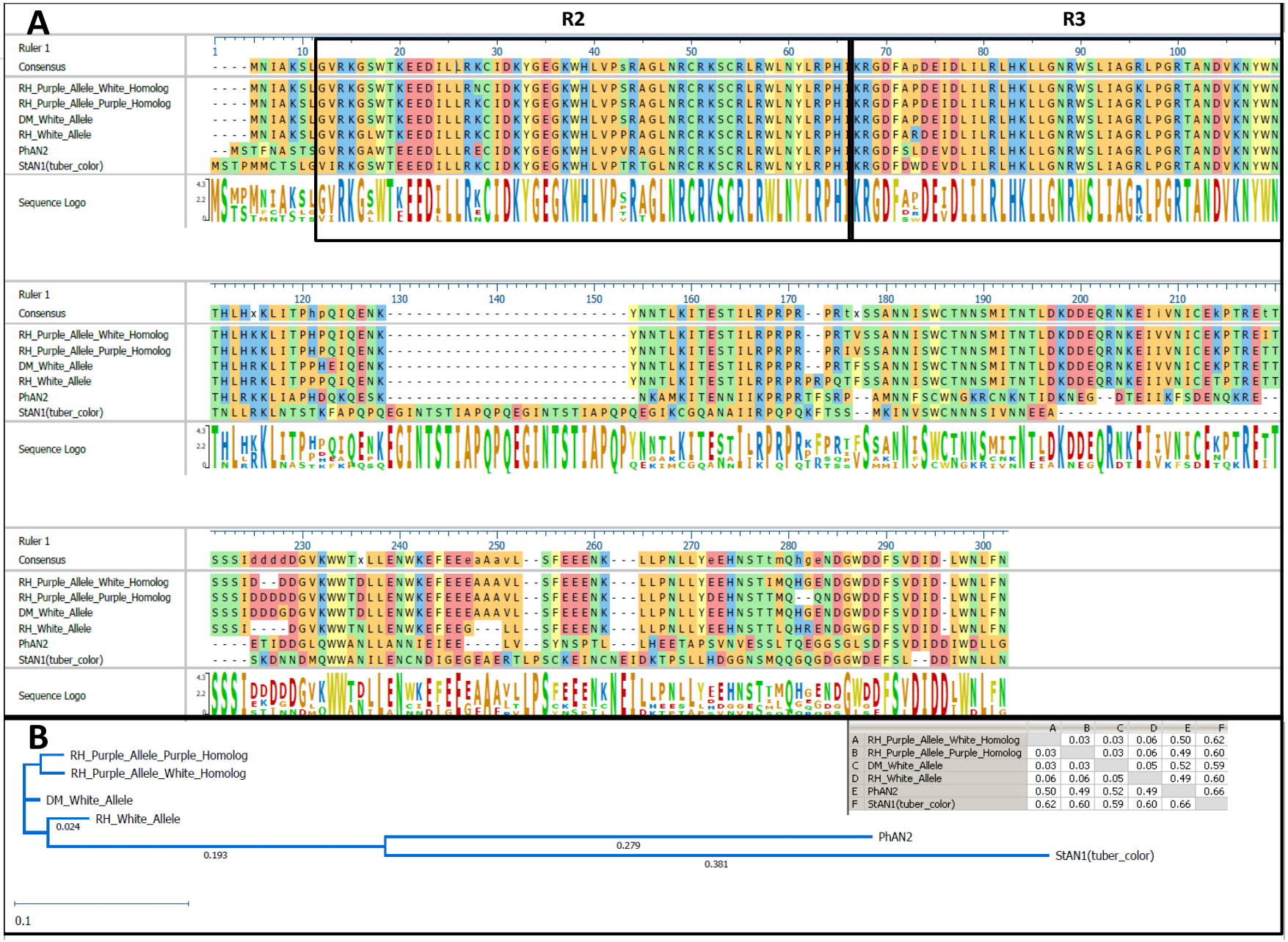


Figure 4.23: Comparison of *AN2* protein sequences in the DMxRH population with those of petunia (*PhAN2*) and the regulator of tuber anthocyanin production (*StAN1*). A) Protein alignment showing functionality of all *AN2* protein sequences and conservation at R1 and R2 regions. B) Phylogenetic tree of *AN2* (and related) protein sequences including distance table.

Transgenic complementation

Operating under the hypothesis that the aforementioned variations within the promoter sequences were responsible for phenotypic variation within the DMxRH population and the monoploid panel, we performed a series of promoter/CDS swaps to confirm that the DM promoter was indeed insufficient to induce expression. Along this line of inquiry, we created four separate constructs: the full length promoter of RH_p, with homology to DM on both sides of the MITE, driving cDNA from the purple-flower bulk (1); the same RH_p promoter driving the entire whole gene sequence (WGS), including introns, of the two RH_P paralogs, $RH_{P\text{-}NE}$ (2) and $RH_{P\text{-}E}$ (3); and finally the DM promoter, MITE included, driving the $RH_{P\text{-}E}$ WGS. To ensure our results would not be complicated by any background deficiencies in DM's anthocyanin biosynthetic pathway that may have been masked by both haplotypes of RH in the F_1 we transformed both DM and a white-flowered F_1 individual, DMxRH 171.

Table 4.5: Summary of transgenic complementation experiment on DM and a white-flowered F_1 individual (DMxRH 171). The source of the promoter is given before the "::" symbol while the source of coding sequence falls after. WGS indicates the entire whole gene sequence, including introns was used, while cDNA was amplified directly from the purple F_1 bulks of the bulk-segregant analysis experiment

Construct	Total Regenerates	Negative Regenerates	Positive Regenerates
	DM		
$RH_{p\text{-}NE}$ :: cDNA	7	0	7
$RH_{p\text{-}NE}$ :: WGS_{NE}	2	0	2
$RH_{p\text{-}NE}$ :: WGS_{E}	5	1	4
DM :: WGS_{E}	9	3	6
	DMxRH $F_1$171		
$RH_{p\text{-}NE}$ :: cDNA	6	0	6
$RH_{p\text{-}NE}$ :: WGS_{NE}	5	1	4
$RH_{p\text{-}NE}$ :: WGS_{E}	0	0	0
DM :: WGS_{E}	0	0	0
Total	34	5	29

Much to our surprise, all four constructs resulted in a purple-flowered phenotype in DM; the same was true for the two constructs for which we obtained regenerants in DMxRH 171. To further complicate matters, we amplified and Sanger sequenced the entire region (full length promoter and CDS) of the RH_p haploytype, ranging from a conserved region of the 3' end of the CDS to the promoter distal to the reference MITE, which we had assumed to be contiguous with $RH_{p\text{-}E}$. Instead, we found that the $RH_{p\text{-}NE}$ paralog retained the promoter with homology to DM on both sides of the MITE insertion, while the $RH_{p\text{-}E}$ promoter lost all homology with the DM allele shortly before the MITE insertion; and much of its promoter sequence is therefore unknown. Thus, we observed that the promoters of two unexpressed, but fully functional, copies of the *AN2* CDS were capable of inducing expression when transformed into the DM background (**Figure 4.8)**. As the region surrounding *AN2* is especially rife with transposable elements—there are 13 MITEs present within 5kb of *AN2* **(Table 4.2)** and a gap in the genome assembly a mere 1.2 kb

downstream of the CDS itself— we suspect the dynamic nature of the region may be the cause of this otherwise perplexing result, a consideration which will be expanded upon in the discussion.

Table 4.6: Reference MITEs within 5KB of AN2. The start and end of the insertion are given as well as the MITE Family itself. The intronic (blue) and promoter (red) MITEs are highlighted.

Chromosome	Start	End	MITE
chr10	51464846	51464908	Tc1_DTT_Sot4
chr10	51464962	51465031	Tc1_DTT_Sot4
chr10	51468461	51468631	Tc1_DTT_Sot7
chr10	51469069	51469164	Tc1_DTT_Sot4
chr10	51470636	51470815	hAT_DTA_Sot30
chr10	51470856	51471163	Mutator_DTM_Sot5
chr10	51472206	51472492	Mutator_DTM_Sot39
chr10	51473225	51473321	Mutator_DTM_Sot39
chr10	51473490	51473620	Tc1_DTT_Sot1
chr10	51474145	51474293	Tc1_DTT_Sot4
chr10	51474289	51474427	Tc1_DTT_Sot4
chr10	51475068	51475642	hAT_DTA_Sot44
chr10	51476153	51476409	hAT_DTA_Sot34

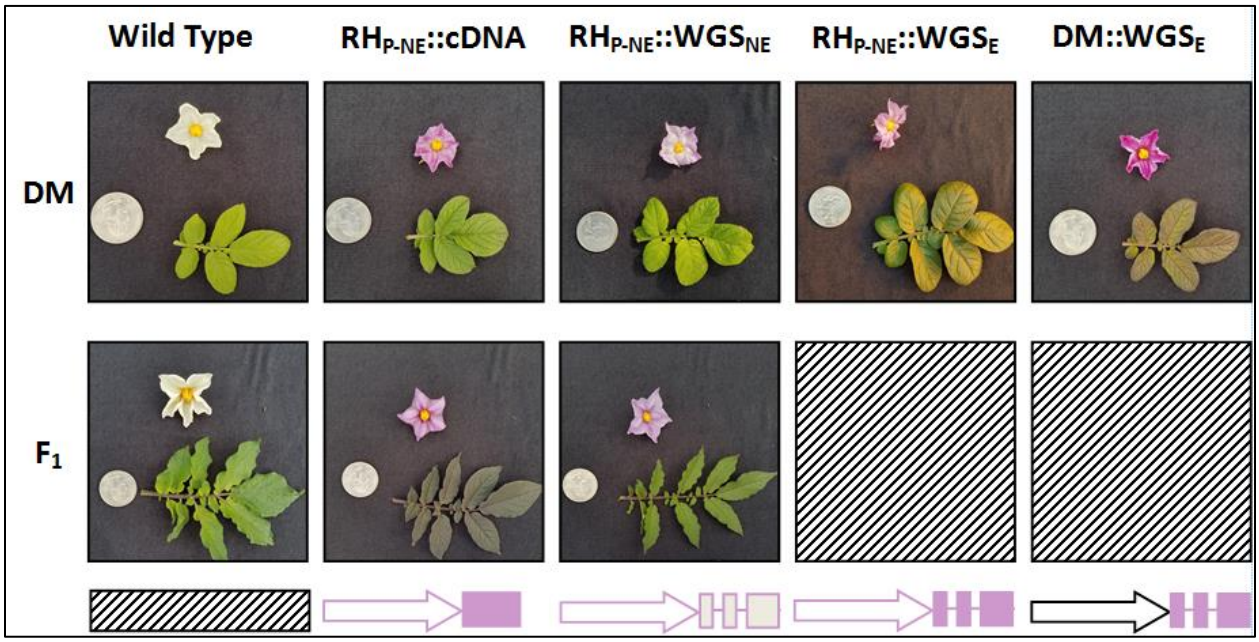


Figure 4.24: Exemplar phenotypes from transgenic complementation of DM and a white-flowered F_1. All constructs resulted in a purple flowered-phenotype. A schematic of the associated construct is presented below each image. Arrows represent promoters while boxes represent coding sequences and lines, introns. Shapes are colored to indicate the expression of their haplotype in the native DMxRH background.

Discussion

AN2 underlies the D locus for flower color in potato

With a combination of genetic mapping, RNA-seq, sequence analysis, and transgenic complementation, we show that an *AN2* homolog is the master regulator of floral anthocyanin production in potato, which has also been demonstrated in petunia, a close relative and model organism for anthocyanin production in flowers. The 1:1 segregation pattern for flower color in the F_1 combined with a presence/absence phenotype imply the flower color is due to segregation of a single regulatory gene; had there been a continuum of color or a more complex segregation pattern a biosynthetic gene might have made a more likely culprit as Śliwka et el. (2017) hypothesized in their recent study on flower color intensity. The results of the SNP-chip

indicating significance on the distal end of chromosome 10, combined with previous reports that this region harbors the requisite locus for flower color (van Eck et al. 1993), provide further support for this assertion. The bulk-segregant analysis shows *AN2* to be the most differentially regulated gene between purple and white genotypes, with a magnitude of difference (530-fold) that might be expected for a presence/absence phenotype. Finally, the complementation of two white-flowered backgrounds, using four separate constructs provides yet another level of support that *AN2* indeed controls expression of anthocyanin biosynthetic genes in potato corollas.

Regulatory effects exerted by *AN2*

The bulk-segregant RNA-seq experiment provides insight into just which genes are affected by the expression cascade initiated by *AN2*. These genes span the gamut of anthocyanin biosynthesis, starting with the initial flux of carbon from aromatic amino acids into the phenylpropanoid pathway catalyzed phenylalanine ammonia lyases (PGSC0003DMG400023458; PGSC0003DMG400031365) , to the first committed step of the flavonoid pathway, chalcone synthase (PGSC0003DMG400019110), and finally ending with the assortment of enzymes involved in anthocyanin structural modification, such as glucosyl, acyl and glutathione transferases (PGSC0003DMG400004573; PGSC0003DMG400024344; PGSC0003DMG401011536; PGSC0003DMG400016722) (Fraser and Chapple 2011). Such an observation in flowers is consistent with previous observations that accumulation of anthocyanins in potato tubers is accompanied by increased expression of both core phenylpropanoid genes and final committed steps (Payyavula et al. 2013). The dataset also confirmed previous reports of *AN2*'s influence on non-structural genes such as anthocyanin permease (PGSC0003DMG400010166) and anthocyanin, the homolog of petunia's *PhAN1* bHLH transcription factor which complexes with *AN2* to initiate transcription of metabolic genes

but is distinct from *StAN1* which is an alternative MYB homolog of *PhAN2* that plays a similar role in tuber skin. We also confirmed an observation presented by Payyavula et al. (2013) that expression of anthocyanin regulating tuber-specific MYBs is correlated with those genes within the phenylpropanoid pathway that do not directly contribute to production of anthocyanins, such as P-coumaroyl quinate/shikimate 3'-hydroxylase (PGSC0003DMG400007178) and caffeoyl-CoA O-methyltransferase (PGSC0003DMG400006448) (Payyavula et al. 2013). As these genes are believed to contribute to production of other phenylpropanoid-derived compounds such as lignin, their increased expression seems more attributable to an increased flux of precursors rather than the direct action of *AN2* itself (Fraser and Chapple 2011; Vanholme et al. 2010).

The remaining differentially expressed genes are less relevant to the actual production of anthocyanins but rather are likely to be the physiological consequences of anthocyanin production. As stated in the results section, excluding anthocyanin synthesis, two major functions stood out in the dataset; that is, up-regulation of genes involved in cell wall degradation or ripening in the purple bulks, and up-regulation of genes involved in photosynthesis and carbon fixation in the white bulks. The abundance of up-regulated pectinesterases (PGSC0003DMG400027840, PGSC0003DMG400024049, PGSC0003DMG400005208) and pectate lyases (PGSC0003DMG400005523, PGSC0003DMG400006938, PGSC0003DMG402017626, PGSC0003DMG400005522) in the purple bulk indicate the anthocyanin-containing corollas were further along in the ‘ripening’ process than those of the white bulk. While it is within the realm of possibility that a sampling error could be the root of this difference, this seems unlikely due to: 1) the number (ten) of individual samples pooled per bulk; 2) the concurrence between bulks and stringency of gene filtering; and 3) the care taken to sample only recently open, turgid flowers. Another possibility,

given the temporospatial relationship between anthocyanin production and ripening is that a certain degree of coordination occurs between the two pathways; however, to our knowledge, no such relationship has been demonstrated. Indeed, heightened anthocyanin has actually been shown to prevent over-ripening and extend shelf life in some fruits although this is more likely due to mitigation of oxidative and biotic stress than inhibition of ripening genes (Zhang et al. 2013). To that end we also observed a >20 fold reduction in polyphenoloxidase expression (PGSC0003DMG400022430) in the purple bulk. As polyphenoloxidases are known to degrade anthocyanins and cause tissue browning (Jiang 2000; Pifferi and Cultrera 1974), this result would likely extend shelf-life if translated to a fruit context. The protective function of anthocyanins may also serve as an explanation for lower expression of photosynthetic genes such as RuBisCo (PGSC0003DMG400019149) and photosystem subunits (PGSC0003DMG400020505, PGSC0003DMG400016504) in purple corollas. Anthocyanins have been shown to protect against photoinhibition by absorbance of light and limiting permeation into the leaf (Steyn et al. 2002). In photosynthetic tissues this is adaptive and may actually increase total photosynthetic capacity by reducing chlorophyll bleaching (Merzlyak and Chivkunova 2000). As the corolla is not typically a photosynthetic organ, a reduction in anthocyanin content may lead to greater light diffusion, thereby initiating a basal level of photosynthesis without concern for bleaching.

Structural variation at the *AN2* locus

The analysis of the floral *AN2* locus and CDS in the DMxRH population, its derivatives, and monoploid panel indicate an extremely dynamic local genetic terrain, plagued by both transposon activity and CNV, especially when placed in the context of other solanaceous *AN2* homologs. The study by Jung et al (2009) identified a tuber-specific *AN2* ortholog, originally

StAN2 but later retitled *StAN1*, as the master regulator of anthocyanin production in tuber skin. Subsequent analysis showed that the region harboring *StAN1* (PGSC0003DMG400013965), located approximately 300 kb distal to the *AN2* locus described here (PGSC0003DMG400019217;

uniquely annotated as *AN2* on the PGSC genome browser), to be replete with MYB homologues, including at least one other functional gene, denoted *StAN2* (PGSC0003DMG400013966; which is distinct from the floral *AN2* and regulates anthocyanin production in response to cold) as well as a multiple pseudogenes (D'Amelia et al. 2014; D'Amelia et al. 2018; André et al. 2009). Hence, the region, and the MYB homologues by extension, has been affected by so many duplications leading to subfunctionalizations that no less than three separate potato genes have been referred to as *AN2* at various points in time due to their shared homology with *PhAN2*; for clarity and continuity with the PGSC browser we will continue to refer to the floral locus as *AN2*. To further complicate matters, we report here that yet another duplication, now of the floral *AN2* RH_p haplotype, is responsible for segregation of corolla anthocyanin production in the DMxRH population, yielding hemizygous purple flowered individuals ($RH_{p\text{-}E}$, $RH_{p\text{-}NE}$/DM) and heterozygous (RH_w/DM) white-flowered individuals.

Sequence analysis shows *StAN1* is similar to the floral *AN2*, including both paralogs of the RH haplotype, $RH_{p\text{-}E}$, $RH_{p\text{-}NE}$, the RH_w allele, and the DM allele, but notably less similar to petunia *PhAN2* which is somewhat surprising considering the potato homologs shared background. Nevertheless, the anthocyanin inducing MYBs investigated here showed a strong conservation of the R2 and R3 repeats essential for DNA-binding with considerably more variation toward the 3' end (**Figure 4.8)**. A comparison of the alleles and paralogs within the DMxRH population and the monoploid panel shows the two paralogs comprising the *AN2* RH_p

haplotype to be the most closely related, which is unsurprising since they are also likely linked in tandem due to their co-segregation in DMxRH F_1 and all purple-flowered monoploids. Bombarely et al (2016) observed that the regions harboring *AN2* homologs in *Petunia inflata* and *P. axillaris* are also remarkably dynamic, with little synteny between the two despite the recent divergence of the species, perhaps due to high transposon density (Bombarely et al. 2016). Furthermore, tomato, a closer relative of potato than petunia, also contains a duplication of *AN2* homologs on chromosome 10; both homologs are functional but not redundant, with only one regulating fruit color (Kiferle et al. 2015). Thus, it is possible that complexities surrounding *StAN1* and *AN2* are the current manifestations of a region especially prone to structural variation for many of the Solanaceae due to a heightened density of repetitive elements and lack of pleiotropic effects of the *AN2* homologs themselves (Bombarely et al. 2016).

Promoter deficiencies are not responsible for lack of *AN2* expression

The fact that all *AN2* homologs maintain a functional ORF, devoid of frame shifts and stop codons, is somewhat surprising given that only one is expressed in the corollas. Even more surprising is that the promoters of two unexpressed alleles/paralogs, DM and $RH_{p\text{-}NE}$, are capable of driving *AN2* expression. Admittedly, we had intended to use the $RH_{p\text{-}E}$ promoter as a positive control with the DM promoter as a negative control, with the assumption being that the full length RH_p promoter with homology to DM was responsible for expression of $RH_{p\text{-}E}$ while the DM promoter was ineffective due to its MITE insertion. As transposable element insertions have been shown to alter expression of nearby genes, though siRNAs, sequence disturbance or altered methylation and chromatin states, this seemed a logical hypothesis (Wang et al. 2013; Hollister et al. 2011; Sigman and Slotkin 2016). As all constructs were effective at complementing the white-flowered phenotype with multiple events, and two in separate backgrounds, positional

effects with respect to the insertions are a dubious scapegoat. Instead, the competency of both the of $RH_{p\text{-}NE}$ and DM promoters when reintroduced into their native background implies a regional effect of the locus itself prevents expression in our white-flowered phenotypes. Since the A*grobacterium*-mediated insertions would re-establish these sequences removed from formerly nearby genetic elements, it seems likely that nearby genetic elements such as the plethora of transposons or a repressor may be restricting *AN2* expression. Furthermore, PCR amplification and subsequent cloning would remove any endemic methylation that repressed gene expression. This may also explain why only a single paralog of the RH_p haplotype is expressed despite their close similarity; the duplication of the locus may have dissociated the newly arisen gene from any repressive elements at the original locus. Further evaluation of the regions surrounding the *AN2* alleles/homologs, through approaches such as chromosome walking or long read whole genome sequencing, may further elucidate the underlying differences which contribute to their expression profiles.

Conclusion

This study combines genetic mapping, transgenic, and computational approaches to verify a homolog of *PhAN2* as performing a similar function in potato while exploring the causes and ramifications of variations in its expression. RNA-seq BSA revealed coordinated up-regulation of the entire anthocyanin biosynthetic pathway as a result of *AN2* expression in purple-flowered phenotypes in addition to changes in ripening and photosynthetic pathways in white-flowered phenotypes. Synthesis of results from transgenic complementation and analysis of gene structure in other homozygous germplasm indicated the *AN2* locus is highly dynamic, including a recent duplication for some haplotypes. Our expectation that differences in the

promoter regions of AN2 alleles with regard to the presence of a MITE in the promoter of white-flowered phenotypes was the ultimate cause of differential *AN2* expression was not borne out, as promoters and coding sequences which were silent in background germplasm were still capable of function when ectopically expressed in potato clones of similar genetic background. A targeted analysis of the epigenetics and structural variants of this region may serve to demonstrate just how a relatively unstable locus can give rise to new homologs and concomitant subfunctionalization.

Supplementary Table 4.4: Differentially expressed genes identified in the BSA analysis. The gene annotation, location, and PGSC_ID are listed; as are each FDR corrected p-value and fold change, where a positive fold change indicated upregulation in the purple bulks. Genes relevant to the flavonoid biosynthetic pathway are highlighted.

Annotation	Fold Change	FDR p-value	Chromosome	Start	End	PGSC ID
AN2	532.18	7.34E-41	Chr10	51469784	51470105	PGSC0003DMG400019217
Major latex_1	115.58	3.62E-09	Chr01	67618877	67620455	PGSC0003DMG400001807
Glutathione S-transferase_3	68.02	6.42E-69	Chr02	36961220	36962529	PGSC0003DMG400016722
Caffeoyl-CoA O-methyltransferase	37.08	1.64E-37	Chr09	54418734	54424093	PGSC0003DMG400006448
Anthocyanin 5-O-glucosyltransferase	36.62	1.14E-53	Chr12	59388654	59391021	PGSC0003DMG400004573
Anthocyanin synthase	28.36	1.39E-12	Chr08	53728301	53732590	PGSC0003DMG400022746
Anthocyanin permease_2	27.79	2.49E-48	Chr03	43777495	43780468	PGSC0003DMG400010166
Anthocyanin acyltransferase_26	18.19	1.10E-28	Chr12	56016125	56018266	PGSC0003DMG401011536
Flavonoid 3',5'-hydroxylase	17.19	8.71E-09	Chr11	39417578	39420629	PGSC0003DMG400000425

Acyl:coa ligase acetate-coa synthetase_1	13.81	7.28E-06	Chr02	36961117	36964883	PGSC0003DMG400016733
Anthocyanin 1	10.29	0.03	Chr09	47422300	47429221	PGSC0003DMG400012891
Homeobox protein_8	5.31	1.16E-03	Chr03	60116827	60123799	PGSC0003DMG400002627
D-alanine-D-alanine ligase	4.24	7.15E-03	Chr10	53674619	53681752	PGSC0003DMG400011105
Conserved gene of unknown function_3010	3.95	8.65E-10	Chr06	43585357	43588353	PGSC0003DMG400013816
Phenylalanine ammonia-lyase_3	3.91	6.66E-05	Chr05	51694755	51698709	PGSC0003DMG400023458
Aminotransferase family protein_4	3.22	2.10E-03	Chr10	2074803	2078104	PGSC0003DMG400021276
ASR4	2.66	7.10E-29	Chr04	60205626	60207072	PGSC0003DMG400006661
Proton P-ATPase	2.55	6.97E-04	Chr07	270950	275949	PGSC0003DMG400011839
Cytochrome P450 hydroxylase_15	2.44	2.52E-12	Chr10	54538508	54540606	PGSC0003DMG402010991
GPI-anchored protein_2	2.44	1.52E-03	Chr03	4985642	4986819	PGSC0003DMG400024379
Pectate lyase P59_3	2.41	0.03	Chr03	31505868	31507745	PGSC0003DMG400005523
SAUR family protein_35	2.41	4.95E-03	Chr10	55103086	55103856	PGSC0003DMG400028299
Pectate lyase P59_1	2.36	0.03	Chr02	26017219	26019334	PGSC0003DMG400006938
P-coumaroyl quinate/shikimate 3'-hydroxylase_3	2.35	1.03E-03	Chr10	59468653	59471590	PGSC0003DMG400007178
Pectate lyase_7	2.34	9.51E-03	Chr05	5914722	5917535	PGSC0003DMG402017626
L-ascorbate oxidase homolog	2.33	6.66E-05	Chr01	28857039	28858942	PGSC0003DMG400006516
Ntp201	2.31	3.63E-04	Chr01	1062928	1065071	PGSC0003DMG400032210
Chalcone synthase 2	2.31	1.19E-07	Chr05	48886959	48888729	PGSC0003DMG400019110
Pectinesterase_5	2.3	6.79E-05	Chr01	57206298	57208265	PGSC0003DMG400027840
Monosaccharide transporter 1	2.29	4.67E-03	Chr01	3931847	3936000	PGSC0003DMG400016362
Cullin 1A	2.27	1.37E-03	Chr06	1962708	1966421	PGSC0003DMG400014708
PGPS/D3_2	2.22	4.66E-03	Chr03	46614170	46615243	PGSC0003DMG400000576
Gene of unknown function_6940	2.22	0.02	Chr11	38321397	38324170	PGSC0003DMG400022083

Pectinesterase_37	2.22	0.01	Chr06	59454833	59456909	PGSC0003DMG400024049
Conserved gene of unknown function_2983	2.17	2.92E-05	Chr06	40610372	40612477	PGSC0003DMG400011815
Phenylalanine ammonia-lyase_9	2.16	1.82E-06	Chr10	51926200	51930606	PGSC0003DMG400031365
Anthocyanidin 3-O-glucosyltransferase	2.15	1.96E-08	Chr09	31520019	31521721	PGSC0003DMG400024344
KUP1_2	2.12	0.02	Chr09	52350668	52354458	PGSC0003DMG400011388
Gene of unknown function_2168	2.11	8.16E-07	Chr04	11393644	11394425	PGSC0003DMG400007627
Pectinesterase_7	2.1	1.10E-03	Chr01	63309078	63311172	PGSC0003DMG400005208
Pectate lyase P56	2.01	9.27E-03	Chr03	31546129	31547779	PGSC0003DMG400005522
Fasciclin-like arabinogalactan protein 3	2.00	1.24E-03	Chr08	2394607	2395637	PGSC0003DMG400009531
Ribulose bisphosphate carboxylase/oxygenase activase, Chloroplastic	-2.01	2.58E-09	Chr10	50945735	50948506	PGSC0003DMG400019149
Photosystem I reaction center subunit X psaK	-2.02	5.50E-03	Chr08	1301974	1310102	PGSC0003DMG400020505
S-adenosylmethionine-dependent methyltransferase_5	-2.03	1.11E-03	Chr04	12347557	12349360	PGSC0003DMG400010638
Cytochrome P450_203	-2.03	8.42E-09	Chr10	57175786	57177298	PGSC0003DMG400039791
21kD protein_1	-2.06	1.81E-03	Chr03	40102015	40102864	PGSC0003DMG400010809
Conserved gene of unknown function_4854	-2.06	4.04E-03	Chr10	51111644	51112600	PGSC0003DMG400019150
Conserved gene of unknown function_3754	-2.07	0.04	Chr08	5526693	5528435	PGSC0003DMG402005859
PSI-H_1	-2.07	3.63E-04	Chr06	49305821	49307605	PGSC0003DMG400016504
Beta-ketoacyl-coa synthase family protein_4	-2.12	7.71E-03	Chr08	40944428	40946611	PGSC0003DMG400007373
Beta-galactosidase_7	-2.12	5.59E-07	Chr03	60932522	60938528	PGSC0003DMG400002590
Desacetoxyvindoline 4-hydroxylase_1	-2.14	2.72E-05	Chr04	2625517	2628817	PGSC0003DMG400029517
16kDa membrane protein	-2.18	1.36E-08	Chr06	54752175	54753353	PGSC0003DMG400005890
Fructose-1,6-bisphosphatase_2	-2.24	5.25E-06	Chr09	739209	741850	PGSC0003DMG400020363
P-coumaroyl quinate/shikimate 3'-hydroxylase_5	-2.28	2.95E-06	Chr10	59481981	59484716	PGSC0003DMG400007180

Gene of unknown function_2136	-2.28	1.68E-04	Chr04	7644725	7646446	PGSC0003DMG400010761
Phosphoribulokinase	-2.34	1.37E-08	Chr08	48800791	48805673	PGSC0003DMG400009317
UDP-glucose:glucosyltransferase_17	-2.35	0.04	Chr05	7189487	7192188	PGSC0003DMG400015601
Fructose-bisphosphate aldolase_2	-2.39	0.02	Chr02	20680813	20683256	PGSC0003DMG400012012
Peroxidase 12	-2.46	0.02	Chr04	61089723	61091577	PGSC0003DMG401025083
Nb cell deth marker	-2.85	0.02	Chr03	49862776	49863532	PGSC0003DMG400010131
Zinc finger protein_82	-3.08	5.70E-04	Chr04	46475605	46480004	PGSC0003DMG400028701
CYP72A54_6	-3.08	0.04	Chr10	43658746	43666340	PGSC0003DMG400008267
Flavonoid glucoyltransferase UGT73E2_1	-3.67	3.56E-06	Chr10	52432622	52434505	PGSC0003DMG400017603
Conserved gene of unknown function_4918	-3.76	4.41E-19	Chr10	57306083	57307830	PGSC0003DMG400023700
Conserved gene of unknown function_3926	-3.83	4.11E-16	Chr08	41634497	41635721	PGSC0003DMG400002261
Conserved gene of unknown function_1306	-3.93	5.02E-04	Chr03	850686	851294	PGSC0003DMG400013467
Caryophyllene/alpha-humulene synthase_1	-4.07	1.36E-08	Chr06	41556836	41557802	PGSC0003DMG400013033
Patatin group O	-4.34	3.88E-04	Chr08	1649641	1654633	PGSC0003DMG400029247
Tropinone reductase I	-4.36	0.02	Chr10	56414553	56425560	PGSC0003DMG400028221
Pectin methlyesterase inhibitor protein 1	-4.87	1.13E-15	Chr03	39768603	39769328	PGSC0003DMG400012020
Senescence-specific cysteine protease	-4.99	2.15E-25	Chr02	32269516	32271942	PGSC0003DMG400010207
Cytokinin oxidase/dehydrogenase_1	-5.05	6.88E-03	Chr01	64568551	64571106	PGSC0003DMG400006764
Metallocarboxypeptidase inhibitor_1	-11.54	3.57E-06	Chr07	3198623	3200152	PGSC0003DMG400030731
Cysteine protease inhibitor 1_3	-11.56	5.66E-03	Chr03	49685390	49686264	PGSC0003DMG400010137
Polyphenoloxidase	-24.73	4.31E-07	Chr02	34338338	34340696	PGSC0003DMG400022430
Calcium ion binding protein_12	-	6.66E-05	Chr04	3712421	3714476	PGSC0003DMG400026764

Works Cited

Alvarez-Suarez JM, Giampieri F, Tulipani S, Casoli T, Di Stefano G, González-Paramás AM, Santos-Buelga C, Busco F, Quiles JL, Cordero MD (2014) One-month strawberry-rich anthocyanin supplementation ameliorates cardiovascular risk, oxidative stress markers and platelet activation in humans. The Journal of nutritional biochemistry 25:289-294

André CM, Schafleitner R, Legay S, Lefèvre I, Aliaga CAA, Nomberto G, Hoffmann L, Hausman J-F, Larondelle Y, Evers D (2009) Gene expression changes related to the production of phenolic compounds in potato tubers grown under drought stress. Phytochemistry 70:1107-1116

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208-218

Barow M (2006) Endopolyploidy in seed plants. Bioessays 28:271-281. doi:10.1002/bies.20371

Bengochea T, Dodds JH (1986) Plant protoplasts : a biotechnological tool for plant improvement. London : Chapman and Hall,

Bethke PC, Nassar AM, Kubow S, Leclerc YN, Li X-Q, Haroon M, Molen T, Bamberg J, Martin M, Donnelly DJ (2014) History and origin of Russet Burbank (Netted Gem) a sport of Burbank. Amer J Potato Res 91:594-609

Bhojwani SS, Dantu PK (2013) Tissue and cell culture. In: Plant Tissue Culture: An Introductory Text. Springer, pp 39-50

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120. doi:10.1093/bioinformatics/btu170

Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, Bliek M, Boersma MR, Borghi L, Bruggmann R, Bucher M, D'Agostino N, Davies K, Druege U, Dudareva N, Egea-Cortines M, Delledonne M, Fernandez-Pozo N, Franken P, Grandont L, Heslop-Harrison JS, Hintzsche J, Johns M, Koes R, Lv X, Lyons E, Malla D, Martinoia E, Mattson NS, Morel P, Mueller LA, Muhlemann J, Nouri E, Passeri V, Pezzotti M, Qi Q, Reinhardt D, Rich M, Richert-Pöggeler KR, Robbins TP, Schatz MC, Schranz ME, Schuurink RC, Schwarzacher T, Spelt K, Tang H, Urbanus SL, Vandenbussche M, Vijverberg K, Villarino GH, Warner RM, Weiss J, Yue Z, Zethof J, Quattrocchio F, Sims TL, Kuhlemeier C (2016) Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. Nature Plants 2:16074. doi:10.1038/nplants.2016.74

https://www.nature.com/articles/nplants201674#supplementary-information

Bourdon M, Coriton O, Pirrello J, Cheniclet C, Brown SC, Poujol C, Chevalier C, Renaudin JP, Frangne N (2011) *In planta* quantification of endoreduplication using fluorescent *in situ* hybridization (FISH). Plant J 66:1089-1099. doi:10.1111/j.1365-313X.2011.04568.x

Bourdon M, Frangne N, Mathieu-Rivet E, Nafati M, Cheniclet C, Renaudin J-P, Chevalier C (2010) Endoreduplication and growth of fleshy fruits. In: Lüttge U, Beyschlag W, Büdel B, Francis D (eds) Progress in Botany 71. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 101-132. doi:10.1007/978-3-642-02167-1_4

Bradshaw JE, Ramsay G (2005) Utilisation of the Commonwealth Potato Collection in potato breeding. Euphytica 146:9-19

Brush SB, Carney HJ, Huam, xe, n Z, xf, simo (1981) Dynamics of andean potato agriculture. Econ Bot 35:70-88

Castelletti S, Tuberosa R, Pindo M, Salvi S (2014) A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL *Vgt1*. G3 (Bethesda) 4:805-812. doi:10.1534/g3.114.010686

Chalker-Scott L (1999) Environmental significance of anthocyanins in plant stress responses. Photochem Photobiol 70:1-9

Chen CT, Setter TL (2003) Response of potato tuber cell division and growth to shade and elevated CO^2. Ann Bot 91:373-381. doi:10.1093/aob/mcg031

Chen CT, Setter TL (2012) Response of potato dry matter assimilation and partitioning to elevated CO^2 at various stages of tuber initiation and growth. Environ Exp Bot 80:27

Chen J, Hu Q, Zhang Y, Lu C, Kuang H (2014) P-MITE: a database for plant miniature inverted-repeat transposable elements. Nucleic Acids Res 42:D1176-D1181. doi:10.1093/nar/gkt1000

Cheniclet C, Rong WY, Causse M, Frangne N, Bolling L, Carde J-P, Renaudin J-P (2005) Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. Plant Physiol 139:1984-1994. doi:10.1104/pp.105.068767

Chevalier C, Bourdon M, Pirrello J, Cheniclet C, Gévaudant F, Frangne N (2014) Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. J Exp Bot 65:2731-2746. doi:10.1093/jxb/ert366

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE (2012) Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. Science 338:1206-1209

D'Amelia V, Aversano R, Batelli G, Caruso I, Castellano Moreno M, Castro-Sanz AB, Chiaiese P, Fasano C, Palomba F, Carputo D (2014) High AN1 variability and interaction with basic helix-loop-helix co-factors related to anthocyanin biosynthesis in potato leaves. Plant J 80:527-540. doi:10.1111/tpj.12653

D'Amelia V, Aversano R, Ruggiero A, Batelli G, Appelhagen I, Dinacci C, Hill L, Martin C, Carputo D (2018) Subfunctionalization of duplicate MYB genes in Solanum commersonii generated the cold-induced ScAN2 and the anthocyanin regulator ScAN1. Plant, cell & environment 41:1038-1051. doi:10.1111/pce.12966

Dai S, Hou J, Long Y, Wang J, Li C, Xiao Q, Jiang X, Zou X, Zou J, Meng J (2015) Widespread and evolutionary analysis of a MITE family Monkey King in Brassicaceae. BMC Plant Biol 15:149. doi:10.1186/s12870-015-0490-9

Davis DA, Currier WW (1986) The effect of the phytoalexin elicitors, arachidonic and eicosapentaenoic acids, and other unsaturated fatty acids on potato tuber protoplasts. Physiol Mol Plant Pathol 28:431-441. doi:http://dx.doi.org/10.1016/S0048-4059(86)80085-6

De Jong W, Eannetta N, De Jong D, Bodis M (2004) Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. Theor Appl Genet 108:423-432

de Pascual-Teresa S, Moreno DA, Garcia-Viguera C (2010) Flavanols and anthocyanins in cardiovascular health: a review of current evidence. Int J Mol Sci 11:1679-1703. doi:10.3390/ijms11041679

de Vetten N, Quattrocchio F, Mol J, Koes R (1997) The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. Genes Dev 11:1422-1434

Devi KD, Punyarani K, Singh NS, Devi HS (2013) An efficient protocol for total DNA extraction from the members of order Zingiberales- suitable for diverse PCR based downstream applications. SpringerPlus 2:669. doi:10.1186/2193-1801-2-669

Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JD (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell 10:1915-1925

Dodds KS, Long DH (1955) The inheritance of colour in diploid potatoes. Journal of Genetics 53:136-149. doi:10.1007/bf02981517

Doke N (1983) Generation of superoxide anion by potato tuber protoplasts during the hypersensitive response to hyphal wall components of Phytophthora infestans and specific inhibition of the reaction by suppressors of hypersensitivity. Phys Plant Path 23:359-367. doi:http://dx.doi.org/10.1016/0048-4059(83)90020-6

Doke N, Tomiyama K (1980a) Effect of hyphal wall components from Phytophthora infestans on protoplasts of potato tuber tissues. Phys Plant Path 16:169-176. doi:http://dx.doi.org/10.1016/0048-4059(80)90031-4

Doke N, Tomiyama K (1980b) Suppression of the hypersensitive response of potato tuber protoplasts to hyphal wall components by water soluble glucans isolated from Phytophthora infestans. Phys Plant Path 16:177-186. doi:http://dx.doi.org/10.1016/0048-4059(80)90032-6

Doležel J, Greilhuber J, Suda J (2007a) Estimation of nuclear DNA content in plants using flow cytometry. Nat Protoc 2:2233-2244. doi:10.1038/nprot.2007.310

Doležel J, Greilhuber J, Suda J (2007b) Flow Cytometry with Plants: An Overview. In: Flow Cytometry with Plant Cells. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp 41-65. doi:10.1002/9783527610921.ch3

Douches D, Maas D, Jastrzebski K, Chase R (1996) Assessment of potato breeding progress in the USA over the last century. Crop Sci 36:1544-1552

Doyle J (1991) DNA protocols for plants. In: Molecular techniques in taxonomy. Springer, pp 283-293

Feller A, Machemer K, Braun EL, Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J 66:94-116. doi:10.1111/j.1365-313X.2010.04459.x

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329+

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103-107. doi:https://doi.org/10.1016/0168-9525(89)90039-5

Flores HE, Kaur-Sawhney R, Galston AW (1981) Protoplasts as Vehicles for Plant Propagation and Improvement. In: Karl M (ed) Advances in Cell Culture, vol Volume 1. Elsevier, pp 241-279. doi:https://doi.org/10.1016/B978-0-12-007901-8.50014-1

Fraser CM, Chapple C (2011) The Phenylpropanoid Pathway in Arabidopsis. The Arabidopsis Book / American Society of Plant Biologists 9:e0152. doi:10.1199/tab.0152

Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E (1983) Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science 220:1049-1051. doi:10.1126/science.220.4601.1049

Grafi G, Larkins BA (1995) Endoreduplication in maize endosperm: involvement of M phase--promoting factor inhibition and induction of S phase--related kinases. Science 269:1262-1264. doi:10.1126/science.269.5228.1262

Guo C, Spinelli M, Ye C, Li QQ, Liang C (2017) Genome-wide comparative analysis of miniature inverted repeat transposable elements in 19 *Arabidopsis thaliana* ecotype accessions. Sci Rep 7:2634. doi:10.1038/s41598-017-02855-1

Hardigan MA, Crisovan E, Hamiltion JP, Kim J, Laimbeer P, Leisner CP, Manrique-Carpintero NC, Newton L, Pham GM, Vaillancourt B, Yang X, Zeng Z, Douches D, Jiang J, Veilleux RE, Buell CR (2016a) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated Solanum tuberosum. Plant Cell. doi:10.1105/tpc.15.00538

Hardigan MA, Crisovan E, Hamilton JP, Kim J, Laimbeer P, Leisner CP, Manrique-Carpintero NC, Newton L, Pham GM, Vaillancourt B (2016b) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated Solanum tuberosum. Plant Cell 28:388-405

Hare EE, Johnston JS (2011) Genome size determination using flow cytometry of propidium iodide-stained nuclei. In: Orgogozo V, Rockman MV (eds) Molecular Methods for Evolutionary Genetics. Humana Press, Totowa, NJ, pp 3-12. doi:10.1007/978-1-61779-228-1_1

Hawkes JG (1992) History of the potato. In: The potato crop. Springer, pp 1-12

Helaers R, Milinkovitch MC (2010) MetaPIGA v2. 0: maximum likelihood large phylogeny estimation using the metapopulation genetic algorithm and other stochastic heuristics. BMC Bioinformatics 11:379

Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, Gaut BS (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. Proc Nat Acad Sci USA 108:2322-2327. doi:10.1073/pnas.1018222108

International Human Genome Sequencing C (2001) Initial sequencing and analysis of the human genome. Nature 409:860. doi:10.1038/35057062

https://www.nature.com/articles/35057062#supplementary-information

Jansky SH, Charkowski AO, Douches DS, Gusmini G, Richael C, Bethke PC, Spooner DM, Novy RG, De Jong H, De Jong WS (2016) Reinventing potato as a diploid inbred line–based crop. Crop Sci 56:1412-1422

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. Nature 421:163. doi:10.1038/nature01214

https://www.nature.com/articles/nature01214#supplementary-information

Jiang Y (2000) Role of anthocyanins, polyphenol oxidase and phenols in lychee pericarp browning. J Sci Food Agric 80:305-310

Jones H, Karp A, Jones MG (1989) Isolation, culture, and regeneration of plants from potato protoplasts. Plant Cell Rep 8:307-311. doi:10.1007/BF00274137

Jung CS, Griffiths HM, De Jong DM, Cheng S, Bodis M, De Jong WS (2005) The potato P locus codes for flavonoid 3′, 5′-hydroxylase. Theor Appl Genet 110:269-275

Jung CS, Griffiths HM, De Jong DM, Cheng S, Bodis M, Kim TS, De Jong WS (2009) The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. TAG Theoretical and Applied Genetics Theoretische Und Angewandte Genetik 120:45-57. doi:10.1007/s00122-009-1158-3

Kaldy M (1972) Protein yield of various crops as related to protein value. Econ Bot 26:142-144

Karp A, Risiott R, Jones MGK, Bright SWJ (1984) Chromosome doubling in monohaploid and dihaploid potatoes by regeneration from cultured leaf explants. Plant Cell Tiss Org Cult 3:363-373. doi:10.1007/bf00043089

Keane TM, Wong K, Adams DJ (2013) RetroSeq: transposable element discovery from next-generation sequencing data. Bioinformatics 29:389-390. doi:10.1093/bioinformatics/bts697

Khoo HE, Azlan A, Tang ST, Lim SM (2017) Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. Food and Nutrition Research 61:1361779. doi:10.1080/16546628.2017.1361779

Kiferle C, Fantini E, Bassolino L, Povero G, Spelt C, Buti S, Giuliano G, Quattrocchio F, Koes R, Perata P, Gonzali S (2015) Tomato R2R3-MYB Proteins SlANT1 and SlAN2: Same Protein Activity, Different Roles. PLoS ONE 10:e0136365. doi:10.1371/journal.pone.0136365

Kikuchi K, Terauchi K, Wada M, Hirano H-Y (2003) The plant MITE *mPing* is mobilized in anther culture. Nature 421:167. doi:10.1038/nature01218

https://www.nature.com/articles/nature01218#supplementary-information

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357-360. doi:10.1038/nmeth.3317

Kloosterman B, Abelenda JA, Gomez Mdel M, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S, Visser RG, Bachem CW (2013)

Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495:246-250. doi:10.1038/nature11912

Knox AK, Dhillon T, Cheng H, Tondelli A, Pecchioni N, Stockinger EJ (2010) *CBF* gene copy number variation at *Frost Resistance-2* is associated with levels of freezing tolerance in temperate-climate cereals. Theor Appl Genet 121:21-35

Kolasa KM (1993) The potato and human nutrition. Am Potato J 70:375-384

Kondrashov AS (1982) Selection against harmful mutations in large sexual and asexual populations. Genet Res 40:325-332

Kotch GP, Ortiz R, Peloquin S (1992) Genetic analysis by use of potato haploid populations. Genome 35:103-108

Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, Jiang J, Buell CR, Baker B (2009) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. Genome Res 19:42-56. doi:10.1101/gr.078196.108

Laimbeer FPE, Holt SH, Makris M, Hardigan MA, Robin Buell C, Veilleux RE (2017) Protoplast isolation prior to flow cytometry reveals clear patterns of endoreduplication in potato tubers, related species, and some starchy root crops. Plant Methods 13:27. doi:10.1186/s13007-017-0177-3

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M (2011) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40:D1202-D1210

Larkins BA, Dilkes BP, Dante RA, Coelho CM, Woo YM, Liu Y (2001) Investigating the hows and whys of DNA endoreduplication. J Exp Bot 52:183-192. doi:10.1093/jexbot/52.355.183

Larson-Rabin Z, Li Z, Masson PH, Day CD (2009) *FZR2/CCS52A1* expression is a determinant of endoreduplication and cell expansion in arabidopsis. Plant Physiol 149:874-884. doi:10.1104/pp.108.132449

Lepetit D, Pasquet S, Olive M, Theze N, Thiebaud P (2000) Glider and Vision: two new families of miniature inverted-repeat transposable elements in *Xenopus laevis* genome. Genetica 108:163-169

Li W, Godzik A (2006) CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659. doi:10.1093/bioinformatics/btl158

Lin X, Long L, Shan X, Zhang S, Shen S, Liu B (2006) In planta mobilization of mPing and its putative autonomous element Pong in rice by hydrostatic pressurization. J Exp Bot 57:2313-2323

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. doi:10.1186/s13059-014-0550-8

Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H (2012) Miniature Inverted–Repeat Transposable Elements (MITEs) have been accumulated through amplification bursts and play

important roles in gene expression and species diversity in *Oryza sativa*. Mol Biol Evol 29:1005-1017. doi:10.1093/molbev/msr282

Mao H, Wang H, Liu S, Li Z, Yang X, Yan J, Li J, Tran L-SP, Qin F (2015) A transposable element in a *NAC* gene is associated with drought tolerance in maize seedlings. Nat Commun 6:8326. doi:10.1038/ncomms9326

https://www.nature.com/articles/ncomms9326#supplementary-information

Marand AP, Jansky SH, Zhao H, Leisner CP, Zhu X, Zeng Z, Crisovan E, Newton L, Hamernik AJ, Veilleux RE, Buell CR, Jiang J (2017) Meiotic crossovers are associated with open chromatin and enriched with *Stowaway* transposons in potato. Genome Biol 18:203. doi:10.1186/s13059-017-1326-8

Melaragno JE, Mehrotra B, Coleman AW (1993) Relationship between endopolyploidy and cell size in epidermal tissue of Arabidopsis. Plant Cell 5:1661-1668. doi:10.1105/tpc.5.11.1661

Merzlyak MN, Chivkunova OB (2000) Light-stress-induced pigment changes and evidence for anthocyanin photoprotection in apples. Journal of Photochemistry and Photobiology B: Biology 55:155-163

Momose M, Abe Y, Ozeki Y (2010) Miniature inverted-repeat transposable elements of *Stowaway* are active in potato. Genetics 186:59-66. doi:10.1534/genetics.110.117606

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461:1130-1134. doi:10.1038/nature08479

Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T (2003) Mobilization of a transposon in the rice genome. Nature 421:170

Nedunchezhiyan M, Byju G, Jata SK (2012) Sweet potato agronomy. Fruit Veg Cereal Sci Biotech 6:1-10

Nelson MG, Linheiro RS, Bergman CM (2017) McClintock: An integrated pipeline for detecting transposable element insertions in whole-genome shotgun sequencing data. G3 7:2763-2778. doi:10.1534/g3.117.043893

Ngezahayo F, Xu C, Wang H, Jiang L, Pang J, Liu B (2009) Tissue culture-induced transpositional activity of *mPing* is correlated with cytosine methylation in rice. BMC Plant Biol 9:91. doi:10.1186/1471-2229-9-91

Ochatt SJ (2008) Flow cytometry in plant breeding. Cytometry A 73:581-598. doi:10.1002/cyto.a.20562

Ojala JC, Stark JC, Kleinkopf GE (1990) Influence of irrigation and nitrogen management on potato yield and quality. Am Potato J 67:29-43. doi:Doi 10.1007/Bf02986910

Oki N, Yano K, Okumoto Y, Tsukiyama T, Teraishi M, Tanisaka T (2008) A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, Oryza sativa ssp. japonica. Genes Genet Syst 83:321-329

Owen HR VR, Levy D, Ochs DL (1988) Environmental, genotypic, and ploidy effects on endopolyploidization within a genotype of *Solanum phureja* and its derivatives. Genome:506-510

Payyavula RS, Singh RK, Navarre DA (2013) Transcription factors, sucrose, and sucrose metabolic genes interact to regulate potato phenylpropanoid metabolism. J Exp Bot 64:5115-5131. doi:10.1093/jxb/ert303

Peterson BA, Holt SH, Laimbeer FPE, Doulis AG, Coombs J, Douches DS, Hardigan MA, Buell CR, Veilleux RE (2016) Self-Fertility in a Cultivated Diploid Potato Population Examined with the Infinium 8303 Potato Single-Nucleotide Polymorphism Array. The plant genome 9

Peterson L, Barker WG, Howarth MJ (1985) Development and structure of tubers. In: Li PH (ed) Potato physiology. Academic Press, Orlando, Florida, pp 124-148

Petroni K, Tonelli C (2011) Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant science : an international journal of experimental plant biology 181:219-229. doi:10.1016/j.plantsci.2011.05.009

PGSC PGSC (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189-195

Pifferi P, Cultrera R (1974) Enzymatic degradation of anthocyanins: the role of sweet cherry polyphenol oxidase. J Food Sci 39:786-791

Pirrello J, Bourdon M, Cheniclet C, Bourge M, Brown SC, Renaudin J-P, Frangne N, Chevalier C (2014) How fruit developmental biology makes use of flow cytometry approaches. Cytometry Part A 85:115-125. doi:10.1002/cyto.a.22417

Platzer A, Nizhynska V, Long Q (2012) TE-Locate: A tool to locate and group transposable element occurrences using paired-end next-generation sequencing data. Biology 1:395-410. doi:10.3390/biology1020395

Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R (1999) Molecular Analysis of the *anthocyanin2* Gene of Petunia and Its Role in the Evolution of Flower Color. Plant Cell 11:1433-1444. doi:10.1105/tpc.11.8.1433

Reeve RM, Timm H, Weaver ML (1973) Parenchyma cell growth in potato tubers I. Different tuber regions. Am Potato J 50:49-57. doi:10.1007/bf02855368

Renault D, Wallender W (2000) Nutritional water productivity and diets. Agric Water Manage 45:275-296

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29:24

Rooke L, Lindsey K (1998) Potato Transformation. In: Foster GD, Taylor SC (eds) Plant Virology Protocols: From Virus Isolation to Transgenic Resistance. Humana Press, Totowa, NJ, pp 353-358. doi:10.1385/0-89603-385-6:353

Salaman RN, Burton WG (1985) The history and social influence of the potato. Cambridge University Press,

Sanseverino W, Hénaff E, Vives C, Pinosio S, Burgos-Paz W, Morgante M, Ramos-Onsins SE, Garcia-Mas J, Casacuberta JM (2015) Transposon insertions, structural variations, and snps contribute to the evolution of the melon genome. Mol Biol Evol 32:2760-2774. doi:10.1093/molbev/msv152

Saxena PK, King J (1985) Reuse of enzymes for isolation of protoplasts. Plant Cell Rep 4:319-320. doi:10.1007/bf00269888

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science 326:1112-1115. doi:10.1126/science.1178534

Scholes DR, Paige KN (2015) Plasticity in ploidy: a generalized response to stress. Trends Plant Sci 20:165-175. doi:10.1016/j.tplants.2014.11.007

Schulz E, Tohge T, Zuther E, Fernie AR, Hincha DK (2016) Flavonoids are determinants of freezing tolerance and cold acclimation in Arabidopsis thaliana. Sci Rep 6:34027. doi:10.1038/srep34027

Schweizer L, Yerk-Davis G, Phillips R, Srienc F, Jones R (1995) Dynamics of maize endosperm development and DNA endoreduplication. Proc Natl Acad Sci 92:7070-7074

Sgorbati S, Levi M, Sparvoli E, Trezzi F, Lucchini G (1986) Cytometry and flow cytometry of 4′,6-diamidino-2-phenylindole (DAPI)-stained suspensions of nuclei released from fresh and fixed tissues of plants. Physiol Plant 68:471-476. doi:10.1111/j.1399-3054.1986.tb03384.x

Sharma SK, Bolser D, de Boer J, Sønderkær M, Amoros W, Carboni MF, D'Ambrosio JM, de la Cruz G, Di Genova A, Douches DS, Eguiluz M, Guo X, Guzman F, Hackett CA, Hamilton JP, Li G, Li Y, Lozano R, Maass A, Marshall D, Martinez D, McLean K, Mejía N, Milne L, Munive S, Nagy I, Ponce O, Ramirez M, Simon R, Thomson SJ, Torres Y,

Waugh R, Zhang Z, Huang S, Visser RGF, Bachem CWB, Sagredo B, Feingold SE, Orjeda G, Veilleux RE, Bonierbale M, Jacobs JME, Milbourne D, Martin DMA, Bryan GJ (2013) Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps. G3 3:2031-2047. doi:10.1534/g3.113.007153

Shen J, Liu J, Xie K, Xing F, Xiong F, Xiao J, Li X, Xiong L (2017) Translational repression by a miniature inverted-repeat transposable element in the 3′ untranslated region. Nat Commun 8:14651. doi:10.1038/ncomms14651

https://www.nature.com/articles/ncomms14651#supplementary-information

Sigman MJ, Slotkin RK (2016) The First Rule of Plant Transposable Element Silencing: Location, Location, Location. Plant Cell 28:304-313. doi:10.1105/tpc.15.00869

Śliwka J, Brylińska M, Stefańczyk E, Jakuczun H, Wasilewicz-Flis I, Sołtys-Kalina D, Strzelczyk-Żyta D, Szajko K, Marczewski W (2017) Quantitative trait loci affecting intensity of violet flower colour in potato. Euphytica 213:254

Sopory SK, Munshi M (1996) Anther culture. In: In vitro haploid production in higher plants. Springer, pp 145-176

Spelt C, Quattrocchio F, Mol JNM, Koes R (2000) anthocyanin1 of Petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. Plant Cell 12:1619-1632

Steyn W, Wand S, Holcroft D, Jacobs G (2002) Anthocyanins in vegetative tissues: a proposed unified function in photoprotection. New Phytol 155:349-361

Szymanski DB, Marks MD (1998) *GLABROUS1* overexpression and *TRIPTYCHON* alter the cell cycle and trichome cell fate in Arabidopsis. Plant Cell 10:2047-2062

Tanaka Y, Sasaki N, Ohmiya A (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. The Plant Journal 54:733-749

Teparkum S, Veilleux RE (1998) Indifference of potato anther culture to colchicine and genetic similarity among anther-derived monoploid regenerants determined by RAPD analysis. Plant Cell Tiss Org Cult 53:49-58

The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158-D169. doi:10.1093/nar/gkw1099

Tiwari JK, Luthra SK, Kumar V, Bhardwaj V, Singh R, Sridhar J, Zinta R, Kumar S (2017) Genomics in True Potato Seed (TPS) Technology: Engineering Cloning Through Seeds. In: The Potato Genome. Springer, pp 297-305

Tracy WF, Goldman IL, Tiefenthaler AE, Schaber MA (2004) Trends in productivity of us crops and long-term selection. Plant Breeding Reviews: Long-term Selection: Crops, Animals, and Bacteria, Volume 24, Part 2:89-108

Uijtewaal BA, Huigen DJ, Hermsen JG (1987) Production of potato monohaploids (2n=x=12) through prickle pollination. Theor Appl Genet 73:751-758. doi:10.1007/bf00260786

Ulrich I, Ulrich W (1991) High-resolution flow-cytometry of nuclear-DNA in higher-plants. Protoplasma 165:212-215. doi:Doi 10.1007/Bf01322292

van Eck HJ, Jacobs JM, van Dijk J, Stiekema WJ, Jacobsen E (1993) Identification and mapping of three flower colour loci of potato (S. tuberosum L.) by RFLP analysis. Theor Appl Genet 86:295-300

Van Valen L (1973) A new evolutionary law. Evolutionary Theory

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin Biosynthesis and Structure. Plant Physiol 153:895-905. doi:10.1104/pp.110.155119

Villordon A, LaBonte D, Firon N (2009) Development of a simple thermal time method for describing the onset of morpho-anatomical features related to sweetpotato storage root formation. Sci Hort 121:374-377. doi:10.1016/j.scienta.2009.02.013

Walker JD, Oppenheimer DG, Concienne J, Larkin JC (2000) *SIAMESE*, a gene controlling the endoreduplication cell cycle in *Arabidopsis thaliana* trichomes. Development 127:3931-3940

Wang X, Weigel D, Smith LM (2013) Transposon Variants and Their Effects on Gene Expression in Arabidopsis. PLoS Genet 9:e1003255. doi:10.1371/journal.pgen.1003255

Wessler SR (2006) Transposable elements and the evolution of eukaryotic genomes. Proc Nat Acad Sci USA 103:17600-17601. doi:10.1073/pnas.0607612103

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973-982. doi:10.1038/nrg2165

Wilson E (1925a) The karyoplasmic ratio. The Macmillan Company, New York

Wilson E (1925b) The karyoplasmic ratio, vol 727. The Cell in Development and Heredity, 3rd edn. The Macmillan Company, New York

Woolfe JA (1992) Sweet potato: an untapped food resource. Cambridge University Press,

Wu F, Tanksley SD (2010) Chromosomal evolution in the plant family Solanaceae. BMC Genomics 11:182. doi:10.1186/1471-2164-11-182

Xu X, Vreugdenhil D, Lammeren AAMv (1998) Cell division and cell enlargement during potato tuber formation. J Exp Bot 49:573-582. doi:10.1093/jxb/49.320.573

Yang G, Dong J, Chandrasekharan M, Hall T (2001) *Kiddo*, a new transposable element family closely associated with rice genes. Mol Genet Genomics 266:417-424. doi:10.1007/s004380100530

Yang G, Hall TC (2003) *MDM-1* and *MDM-2*: two *Mutator*-derived MITE families in rice. J Mol Evol 56:255-264. doi:10.1007/s00239-002-2397-y

Yang G, Lee Y-H, Jiang Y, Shi X, Kertbundit S, Hall TC (2005) A two-edged role for the transposable element *Kiddo* in the rice *ubiquitin2* promoter. Plant Cell 17:1559-1568. doi:10.1105/tpc.104.030528

Yongsheng Y, Yuman Z, Kun Y, Zongxiu S, Yaping F, Xiaoying C, Rongxiang F (2011) Small RNAs from MITE-derived stem-loop precursors regulate abscisic acid signaling and

abiotic stress responses in rice. The Plant Journal 65:820-828. doi:doi:10.1111/j.1365-313X.2010.04467.x

Zhang Y, Butelli E, De Stefano R, Schoonbeek H-j, Magusin A, Pagliarani C, Wellner N, Hill L, Orzaez D, Granell A, Jones Jonathan D, Martin C (2013) Anthocyanins Double the Shelf Life of Tomatoes by Delaying Overripening and Reducing Susceptibility to Gray Mold. Curr Biol 23:1094-1100. doi:10.1016/j.cub.2013.04.072

Zhang Y, Cheng S, De Jong D, Griffiths H, Halitschke R, De Jong W (2009a) The potato R locus codes for dihydroflavonol 4-reductase. Theor Appl Genet 119:931-937. doi:10.1007/s00122-009-1100-8

Zhang Y, Jung CS, De Jong WS (2009b) Genetic analysis of pigmented tuber flesh in potato. Theor Appl Genet 119:143-150

Zhuang J, Wang J, Theurkauf W, Weng Z (2014) TEMP: a computational method for analyzing transposable element polymorphism in populations. Nucleic Acids Res 42:6826-6838. doi:10.1093/nar/gku323

Conclusions

This work described in this dissertation is grounded in a broad collection of biological disciplines; from the physiology of potato tuber formation and expansion, to the genomic scope of MITE insertions and influence, finally ending in a study of a particularly dynamic locus underlying organ specific anthocyanin production. At its heart, this work is contingent upon, and demonstrates the utility of, the ability to study and manipulate ploidy in crop species. Endoreduplication has been repeatedly shown to dictate cell, and by extension organ, size. The awareness of its variability in accessible germplasm combined with a capacity to evaluate it positions future researchers to harness it for meaningful crop improvement. Much in the same vein, we have exploited the ploidy-manipulation tools, previously conceived and perfected, to provide insight regarding the genesis and ramifications of new genetics variants, in both the generic case of transposons, and the specific instance of *AN2*.

If I have seen further it is by standing on the shoulders of giants.

-Sir Isaac Newton