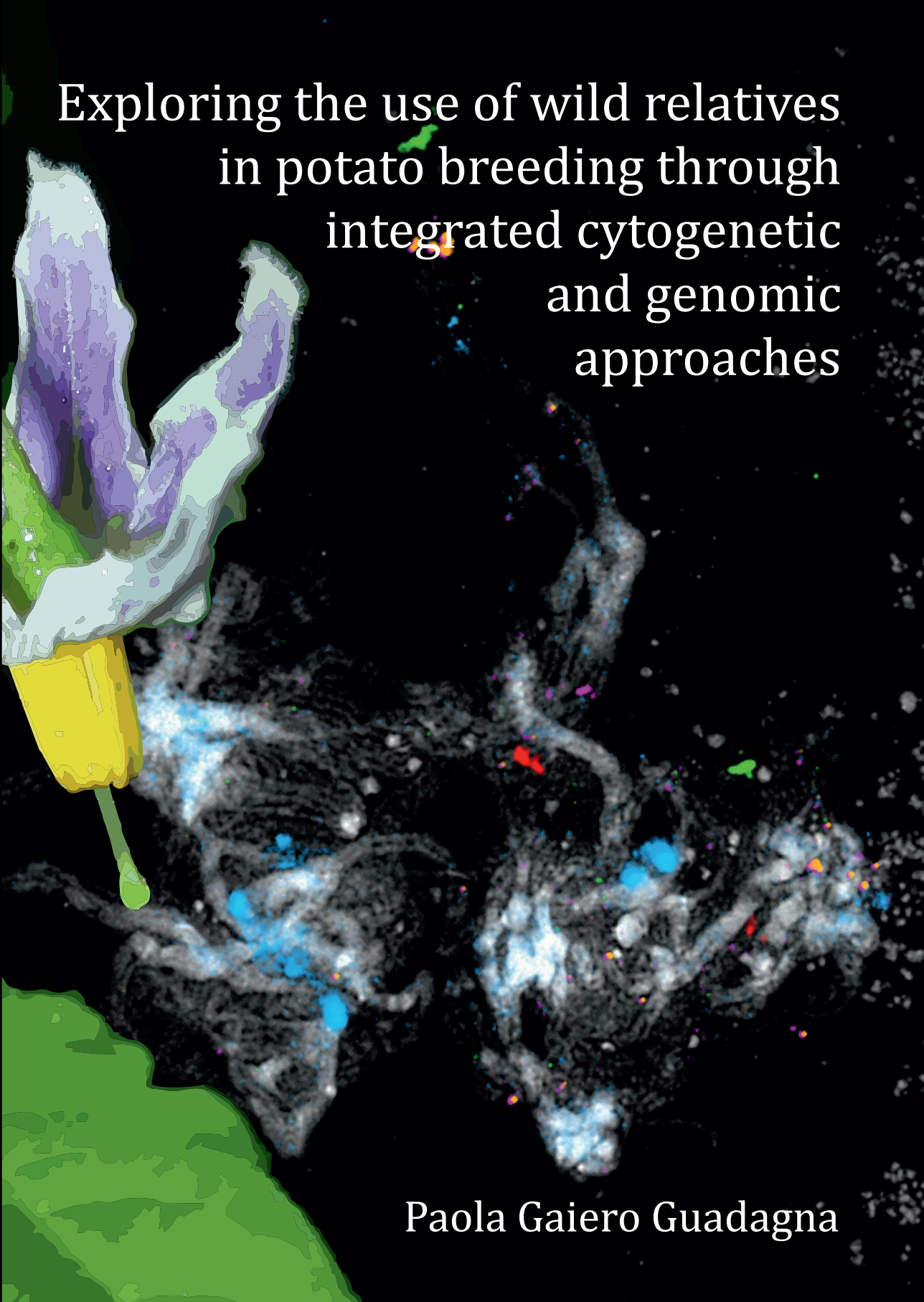
Exploring the use of wild relatives in potato breeding through integrated cytogenetic and genomic approaches



Paola Gaiero Guadagna

Propositions

1. Advanced improvement in domesticated potato in the future depends on investing effort and resources into potato wild relatives –omics data now.
   (this thesis)

2. Genomic studies in *Solanum* crops and wild relatives fall short in interpretation without knowledge of chromosome biology.
   (this thesis)

3. Although sequencing platforms and assembly algorithms are constantly improving, even finished genome assemblies should be taken with a certain degree of scepticism.

4. Special protocols for the exchange of living plant material for research purposes should be implemented to reduce costly and time-consuming bureaucracy, without losing sight of safety and conservation of diversity.

5. If countries that are centres of diversity do not bridge the gap between generating information about Crop Wild Relatives (CWRs) and using it, they will either lose sovereignty over their germplasm or act as the dog in the manger.

6. The high demands in academia drive many researchers to give up their family life in order to have a successful career.

Propositions belonging to the thesis entitled “Exploring the use of wild relatives in potato breeding through integrated cytogenetic and genomic approaches”

Paola Gaiero Guadagna
Wageningen, 11 December 2018

Exploring the use of wild relatives in potato breeding through integrated cytogenetic and genomic approaches

Paola Gaiero Guadagna

Thesis committee

Promotor
Prof. Dr J.H.S.G.M. de Jong
Personal chair at the Laboratory of Genetics, Wageningen University & Research

Co-promotors
Prof. Dr M.E. Schranz
Professor of Biosystematics, Wageningen University & Research

Prof. Dr P.R. Speranza
Professor of Plant Breeding at the Department of Plant Biology, Faculty of Agronomy, University of the Republic, Uruguay

Other members
Prof. Dr M.A.M. Groenen, Wageningen University and Research, Wageningen
Prof. Dr F.P.M. Govers, Wageningen University and Research, Wageningen
Dr C.W.B. Bachem, Wageningen University and Research, Wageningen
Dr P. Lindhout, Solynta R & D, Wageningen

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences (EPS)

Exploring the use of wild relatives in potato breeding through integrated cytogenetic and genomic approaches

Paola Gaiero Guadagna

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 11 December 2018
at 4 p.m. in the Aula

Paola Gaiero Guadagna
Exploring the use of wild relatives in potato breeding through integrated cytogenetic and genomic approaches,
194 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summaries in English, Dutch and Spanish
ISBN: 978-94-6343-346-4
DOI: https://doi.org/10.18174/458827

To Pablo
To Andrés and Lara, our masterpieces

A Pablo
A Andrés y Lara, nuestras obras maestras

Papa
te llamas
papa
y no patata,
no naciste castellana:
eres oscura
como
nuestra piel,
somos americanos...
Pablo Neruda
("Oda a la papa", 1954)

Contents

CHAPTER 1

General introduction

Potato is the third most important food crop after rice and wheat and the number one vegetable crop (Birch et al. 2012; Jansky et al. 2013). Its production has increased dramatically in the developing world in the past few decades, acting as the staple or main carbohydrate supplier (Bradshaw 2007a; Bradshaw and Bonierbale 2010; Ramsay and Bryan 2011; Birch et al. 2012). It produces more dry matter, proteins and calories per hectare and has a higher water productivity than any of the major cereal crops (Burton 1989; Birch et al. 2012). Fresh potato is virtually fat- and cholesterol-free, high in dietary fibre and rich in antioxidants and vitamins (Bradshaw and Bonierbale 2010; Bradeen and Haynes 2011; Birch et al. 2012). The crop is an important resource to fight malnutrition because it provides significant amounts of proteins, minerals and micronutrients. It contributes enormously to UN's Millennium Development Goals of providing food security and eradicating poverty (Bradshaw and Ramsay 2009; Ramsay and Bryan 2011).

Physiologically, potato has a very high harvest index (proportion of the plant's dry-weight which is harvestable tuber) of 0.80. Consequently, as long as its photosynthetic apparatus is functional, a potato plant will efficiently transform energy into tubers. Therefore, gains in yield will come mostly from extending functionality of the plant's aerial vegetative organs, by avoiding damage by pests and diseases, and avoiding or tolerating abiotic stresses (Bradshaw 2007a; Bradshaw 2009; Kloosterman et al. 2013). However, factors that increase biotic and abiotic stress combined with the negative effects of climate change jeopardize stable production, and so put a serious and constant risk for food security (Bradshaw 2007b). As for other crops including potato, breeders can resort to a rich pool of wild relatives that may become valuable sources of economically important genes if efficiently tackled (Jansky 2009).

Cultivated potato and its wild relatives represent a more diverse and accessible germplasm resource than that of any other crop (Ross 1986; Hanneman 1989; Peloquin et al. 1989; Hawkes 1990), whose diversity can be used to (re-)introduce specific traits. This can be done by introgressive hybridization, in which chromatin carrying a gene of interest from a wild relative is integrated into the genome of the crop by interspecific hybridization. During the subsequent backcrossing generations, gene(s) of interest are incorporated into the crop chromosomes by homeologous recombination. The offspring are then selected for the desired trait while the original cultivated genetic background is recovered by backcrossing and selection as far as possible. Potato wild relatives display various advantages over cultivated germplasm (Jansky and Peloquin 2005), such as resistance to late blight, the main potato disease caused by *Phytophthora infestans* and other diseases caused by bacteria and viruses (Jansky 2000; Simko et al. 2009), but also confer tolerance to cold, frost and other environmental stresses. The efficient use of wild relatives now requires extensive knowledge of their allelic variation and genomic structure, including the screening for desirable traits. To minimize the occurrence of linkage drag when introgressed chromatin still contains closely linked wild traits from

the ancestral donor, knowledge on the genomic structure of crop and donor species is indispensable (Ramsay and Bryan 2011).

Overcoming hybridization barriers with *Solanum commersonii* and *S. chacoense*

In spite of the widely available diversity in germplasm collections worldwide, only about 10 % of potato species have been explored for use in breeding programmes (Bradshaw 2007a). Some of the diploid species that have been used are *Solanum chacoense* and *S. commersonii*. To cross them with cultivated potato, traditional methods to overcome hybridization barriers have been used (reviewed by Janksy 2006 and Bradshaw and Bonierbale 2010). Such methods involve ploidy manipulation and hybridization through either somatic fusion or sexual hybridization (Ortiz 1998; Jansky 2006; Ortiz et al. 2009). Somatic fusion circumvents sexual hybridization barriers and incompatibilities, but it does not take advantage of the increased variability carried by gametes due to recombination and assortment during meiosis. Somatic hybrids with *Solanum commersonii* have been used to introgress frost tolerance and cold adaptation (Cardi et al. 1993; Carputo et al. 1998) and to introgress resistance to tuber soft rot caused by *Pectobaterium carotovorum* (Carputo et al. 2000b). *Solanum chacoense* was used to produce somatic hybrids with potato to introduce resistance to bacterial wilt (Chen et al. 2013). Although the results have been promising in terms of resistance and yield, researchers and breeders have turned to sexual hybridization as source of variation and new allele combinations.

In the case of sexual hybridization, ploidy manipulation and bridge crosses are usually necessary. Endosperm balance number (EBN) is the most important post-zygotic hybridization barrier in potaoes (Camadro et al. 2004). EBNs for the different potato species were determined empirically by crossing with *S. chacoense* and evaluating the success in the offspring (Johnston et al. 1980). Cultivated potato is a tetraploid with 4EBN (Endosperm Balance Number), while *S. commersonii* and *S. chacoense* are diploids with 1EBN and 2EBN, respectively (Johnston et al. 1980). One possible solution is to lower the ploidy level of cultivated potato by haploid production. Haploids which have the chromosome number and EBN of the parental gamete can be obtained by parthenogenetic development of the egg cells using pollination with particular clones of diploid group Phureja (Upadhya and Cabello 1997; Panahandeh 2010; Murovec and Bohanec 2012) or by gametophytic embryogenesis through *in vitro* anther culture (Veilleux 2005; Lightbourn and Veilleux 2007; Rokka 2009; Germanà 2011). These haploids have been crossed directly with *S. chacoense* and their progenies were evaluated for agronomic and yield traits (Bani-Aameur et al. 1993; Santini et al. 2000) and for tuber chipping quality traits (Hamernik et al. 2009; Zhao et al. 2013). For *S. commersonii*, these dihaploids have been used extensively in combination with an increase of the ploidy level of the wild donor mitotically through somatic duplication (Carputo et al. 1995; Carputo et al. 1997). Their progenies were evaluated for fertility and traits of interest (Cardi et al. 1993; Tucci et al. 1996; Carputo et al. 1998; Carputo 1999; Carputo et al. 2002; Carputo et al. 2003a; Carputo

2003; Carputo et al. 2003b; Caruso et al. 2008; Carputo et al. 2009; Iorizzo et al. 2011). To maximize the genetic diversity captured from the wild species, its ploidy level can also be increased meiotically by taking advantage of the production of unreduced 2n gametes from several naturally occurring recessive meiotic mutations (den Nijs and Peloquin 1977; Mendiburu and Peloquin 1977; Camadro and Peloquin 1980; Masuelli et al. 1992; Carputo et al. 2000; Jansky 2009).

Unreduced gametes can be formed through different mechanisms, which will determine their genetic consequences. If they are formed by failures in the second meiotic division such as parallel or tripolar spindles, the genetic consequences are equivalent to first division restitution (FDR) (Mok and Peloquin 1975), retaining an average 80% of the original heterozygosity present in the parent (Carputo et al. 2000a; Carputo and Frusciante 2011). When 2n gametes are formed by improper anaphase II chromatid disjoin or premature cytokinesis, genetic consequences are equivalent to second division restitution (SDR) (Mok and Peloquin 1975), which means that in the situation of potato an average of 40% of the parental diversity is retained (Carputo et al. 2000a; Carputo and Frusciante 2011).

The significance of *Solanum commersonii* and *S. chacoense* for potato breeding

Solanum commersonii has been crossed with potato with the aim to increase genetic diversity and to introgress resistance to bacterial wilt (Galvan et al. 2006; González 2010). The hybridization success, viability and fertility of the hybrids previously obtained through sexual polyploidization in these crosses have been reported in (Novy and Hanneman 1991; Masuelli and Camadro 1997). Despite the many efforts to cross this diploid species with cultivated potato, little is known about what happens with the transmission and homoeologous recombination of the alien chromosomes and chromatin after the hybridization. The success of such processes depends to a greater part on the differentiation of genomes between the species that are involed in the introgressive hybridisation. Such a genome differentiation between wild potato species does not seem to operate as a postzygotic isolating mechanism (Dvorak 1983; Camadro et al. 2004) and consequently lacks particular scientific or commercial attention. For potato breeders, information about homoeologous pairing and crossovers in the hybrids between wild and cultivated potatoes is essential, as it determines germplasm accessibility for breeding purposes. In order to introgress traits from a wild donor into tetraploid potato, an intermediate F1 triploid hybrid needs to be produced. In the meiosis of a created allotriploid, obtained from a tetraploid potato and a diploid wild relative, analysis of homeologous pairing and recombination can be derived from chromosome behaviour at meiotic prophase, where preferential pairing between the homologous potato chromosomes competes with the alien homeologue. If the chromosomes do form trivalents in the meiosis of triploids, then introgression is possible (Jansky 2006). A few studies describe the cytological behaviour in the backcrosses of hybrids with *S. commersonii* (Carputo et al. 2003a; Carputo

2003), both through cytological observations and molecular markers (Barone et al. 1999; Barone et al. 2001; Iovene et al. 2004). However, most of them focused on the aneuploid backcross progenies except for one which described allotriploid hybrids (Carputo et al. 1995). However, information on early stages of meiotic prophase I was not available, so it is not known if there is complete or partial homoeologous pairing in these allotriploids. Another crucial phenomenon that needs to be regarded is structural hybridity, caused by chromosome rearrangements between related species that cause specific meiotic problems of their interspecific hybrids leading to spore abortion and sterility. Such rearrangements between the homoeologues represent an important limitation in introgressive breeding due to the unintended retention of large blocks of DNA surrounding a gene of interest (Jacobsen and Schouten 2007). Their impact is even greater when regions linked to the trait of interest have a negative effect on agronomic performance. Although there are no reports of rearrangements among potato and its wild relatives, there are detailed descriptions of such karyotype evolution processes among Solanaceous crops (Tang et al. 2008; Iovene et al. 2008; Lou et al. 2010; Peters et al. 2012; Szinay et al. 2012). Using multi-colour chromosome painting combining BAC FISH technology and fluorescence microscopy it is possible to hybridize many probes in a single experiment, reducing the examination of chromosome integrity to only a few experiments (Tang et al. 2009; Szinay et al. 2010). Under adapted conditions for the FISH, cross-species painting of tomato or potato probes can be performed, in which homoeologous chromosomal positions in related *Solanum* species can be displayed. For potato, the RHPOTKEY BAC library was developed for diploid clone RH89-039-16 (Borm 2008) and anchored to AFLP markers in the ultrahigh density (UHD) genetic map (van Os et al. 2006). From this library, 60 BACs were selected as a cytogenetic painting set in order to anchor the Ultra High Density (UHD) linkage map to the pachytene karyotype, thus providing a useful, yet underused tool to establish collinearity between potato and its wild relatives (Tang et al. 2009; Achenbach et al. 2010; de Boer et al. 2011).

The available genetic and genomic knowledge on wild potatoes is relatively limited compared to that of tomato wild relatives (Szinay et al. 2012; Aflitos et al 2014; Bolger et al 2014). Genome divergence caused by repetitive sequences can be tested by analysing the nature, abundance and dynamics of the repetitive fractions of their genomes. Divergence in the repetitive fraction of the genome usually underlies divergence in genome homology, expansion (considering the larger genomes in the tomato clade) and the occurrence and impact of structural rearrangements.

The use of a very large number of BAC FISH probes, together with information from genetic and physical maps (restriction and optical maps) allowed for the correction of misassemblies in the tomato genome, identifying scaffolds that were not in the correct order or orientation (or both) (Shearer et al. 2014). The potato assembly was similarly improved to achieve a reference genome at the level of pseudomolecules (equivalent to chromosomes) (Sharma et al. 2013; Hardigan et al. 2016). The only potato wild relative

that has such integration of technologies in its genome assembly is *Solanum chacoense* (Leisner et al. 2018). In the case of *Solanum commersonii* assembly to the level of pseudomolecules was achieved by mapping against the reference potato genome (Aversano et al. 2015). This approach does not cater for structural variation between the two species. There are only a few studies that have utilized sufficient high-quality sequence data needed to reveal fine-scale structural differences related to introgression barriers (Peters et al. 2012; Aflitos et al. 2014; Aflitos et al. 2015; de Boer et al. 2015). Although whole genome sequence data for *S. chacoense* and *S. commersonii* are available, the only comparative analysis performed so far used DArT markers (Traini et al. 2013). In order to use these genome assemblies as tools for gene discovery and to assess collinearity between a potential donor and cultivated potato, breeders need reference level genomes and comparative analyses.

Scope and outline of this thesis

The aim of this thesis was to broaden my knowledge in the use of potato wild relatives in breeding by generating information on their genome organization that will assist breeders in their choice of materials as donors. In order to do that, I have used a variety of cytogenetic, genetic and genomic tools, ranging from conventional to cutting edge technologies.

In chapter 2 we review the state of the art in the use of cytogenetic, genetic and genomic tools to trace past introgressions from wild relatives into cultivated potato and to plan new introgressive hybridization schemes. We propose that although these approaches are still expensive and difficult to apply to everyday introgressive hybridization breeding, at the current rate of technological advance, they may soon be implemented routinely. Therefore, it seems possible to envision the fulfillment of the promises of the use of potato wild relatives through new technological approaches that facilitate their exploration and efficient exploitation.

In chapter 3 we looked at the suitability of *Solanum commersonii*, a diploid wild relative of potato, as a donor of resistance to biotic threats. We analysed homoeologous pairing in male meiocytes of allotriploid hybrids obtained through spontaneous fertilization of unreduced eggs from *S. commersonii* with reduced pollen from a diploid potato (*S. tuberosum* Group Phureja). These hybrids behaved as near autotriploids, in the sense of extensive homoeologous pairing and recombination. We also examined meiosis in pollen mother cells (PMC) from the successive progenies resulting from backcrosses with tetraploid potato (*S. tuberosum* Group Tuberosum). We used genome painting (Genomic *in situ* hybridization or GISH) to distinguish the chromosomes or chromosome segments of the wild donor species and to follow their fate in the successive backcrosses. This technique could not discriminate the chromosomes coming from the different parental species, indicating low divergence at the repetitive sequences level. Nevertheless, we identified each chromosome with chromosome-specific BAC probes and showed that there were

no specific chromosomes that failed in pairing with their homoeologues, which suggests that there is high genome similarity between the homoeologues, rendering *S. commersonii* as a promising species for introgressive hybridization breeding.

In chapter 4 we tested whether there was collinearity among the chromosomes of potato and two of its diploid wild relatives, *Solanum commersonii* and *S. chacoense*. Firstly, we described and compared their chromosome morphology and heterochromatin distribution at pachytene in PMC complements. We also performed cross-species BAC FISH experiments using probes with known positions in the potato chromosomes and hybridized them on cell spreads of pollen mother cells at pachytene from the wild species. We discovered that all BAC probes belonging to each linkage group in potato mapped to the same chromosome in the wild species. They also showed the same order and very similar positions, indicating high collinearity at the chromosome level without large-scale rearrangements. This high level of synteny has important implications in introgressive hybridization breeding, because it suggests that, at least at large scale, linkage drag is not expected.

Chapter 5 deals with genomic comparisons of various *Solanum* species, including the two major crops potato and tomato and their wild relatives. From the comparison of publicly available genomic information and that of our results, we concluded that genome differentiation in the sense of repeat composition and organisation worked differently between the tomato and the potato clades. In order to describe genome differentiation at the repetitive sequences level among potato and its wild relatives, we compared their genomes to those of tomato and its wild relatives, which are a well-studied sister group. We compared the repetitive fractions from the genomes within each clade, using low-pass next generation sequencing data and analysing it with the pipeline RepeatExplorer. We also wanted to know if repeats grouped *Solanum etuberosum* with the potatoes or the tomatoes. Results indicate that the repeat profiles of potatoes and tomatoes differ in their relative abundance of repetitive element families and that *S. etuberosum* is more similar to the potato clade according to the repetitive fraction of its genome.

Chapter 6 focuses on whole-genome structural comparative genomics among *Solanum commersonii*, *S. chacoense* and cultivated potato. To achieve such comparisons, a high quality genome assembly was needed, so I performed a hybrid assembly of the *S. commersonii* genome combining short and long read data. This hybrid assembly was then anchored to a SNP-based genetic map. The resulting assembly at the pseudomolecule level was suitable for pairwise genome comparisons which showed that although many assembly artifacts impeded conclusive results on rearrangements, some microsynteny breaks could be confirmed that should be further explored to assess their impact in introgressive hybridization breeding.

In chapter 7 my attention shifts to the more methodological aspects of this thesis. We described modifications made to a flow sorting protocol in order to obtain large amounts of pure high molecular weight (HMW) DNA from different *Solanum* species.

The HMW DNA isolated through this method is the most suitable for next generation sequencing and mapping strategies, including BIONANO® genome mapping. This technology will allow us to close genome assemblies for various *Solanum* species and will enable structural genome comparisons.

In Chapter 8 the results presented in this thesis are integrated and discussed in the light of the progress made and of their usefulness in potato hybridization breeding. We highlighted the most relevant findings that may become tools for breeders to make decisions when using potato wild relatives in breeding and we mentioned the work which is underway and the perspectives for future studies.

Literature cited

Achenbach UC, Tang X, Ballvora A, et al (2010) Comparison of the chromosome maps around a resistance hot spot on chromosome 5 of potato and tomato using BAC-FISH painting. Genome 53:103–110. doi: 10.1139/G09-086

Aflitos SA, Sanchez-Perez G, de Ridder D, et al (2015) Introgression browser: High-throughput whole-genome SNP visualization. Plant J 82:174–182. doi: 10.1111/tpj.12800

Aflitos SA, Schijlen E, de Jong H, et al (2014) Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. Plant J 80:136–148. doi: 10.1111/tpj.12616

Aversano R, Contaldi F, Ercolano MR, et al (2015) The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. Plant Cell 27:954–968. doi: 10.1105/tpc.114.135954

Bani-Aameur F, Lauer FI, Veilleux RE (1993) Enhancement of diploid *Solanum chacoense* Bitt. using adapted clones of *Solanum phureja* Juz. & Buk. Euphytica 68:169–179. doi: 10.1007/BF00029870

Barone A, Sebastiano A, Carputo D (1999) Chromosome pairing in *Solanum commersonii* - *S. tuberosum* sexual hybrids detected by commersonii-specific RAPDs and cytological analysis. Genome 42:218–224. doi: 10.1139/gen-42-2-218

Barone A, Sebastiano A, Carputo D, et al (2001) Molecular marker-assisted introgression of the wild *Solanum commersonii* genome into the cultivated *S. tuberosum* gene pool. TAG Theor Appl Genet 102:900–907. doi: 10.1007/s001220000498

Birch PRJ, Bryan G, Fenton B, et al (2012) Crops that feed the world 8: Potato: are the trends of increased global production sustainable?

Bolger A, Scossa F, Bolger ME, et al (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nat Genet 46:1034–1038. doi: 10.1038/ng.3046

Borm TJA (2008) Construction and use of a physical map of potato. PhD Thesis, Wageningeng UR, The Netherlands

Bradeen JM, Haynes KG (2011) Introduction to Potato. In: Bradeen J, Kole C (eds) Genetics, Genomics and Breeding of Potato. CRC Press, pp 1–19

Bradshaw JE (2007a) Potato breeding strategy. In: Vreugdenhil D, Bradshaw JE, Gebhardt C, Govers F, et al. (eds) Potato Biology and Biotechnology: Advances and Perspectives. Elsevier B.V., pp 157–177

Bradshaw JE (2009) A genetic perspective on yield plateau in potato. Potato J 36:79–94.

Bradshaw JE (2007b) Breeding potato as a major staple crop. In: Kang M, Priyadarshan P (eds) Breeding major food staples. Blackwell Publishing, Oxford, pp 277–332

Bradshaw JE, Bonierbale M (2010) Potatoes. In: Bradshaw JE (ed) Root and Tuber Crops. Handbook of Plant Breeding 7. Springer New York, New York, NY, pp 1–52

Bradshaw JE, Ramsay G (2009) Potato Origin and Production. In: Advances in Potato Chemistry and Technology, First Edit. Elsevier Ltd., pp 1-15.

Burton WG (1989) The potato., 3rd edn. Longman, London

Camadro E, Peloquin S (1980) The Occurrence and Frequency of 2n Pollen in Three Diploid *Solanums* from Northwest Argentina. Theor. Appl. Genet. 15:11–16.

Camadro EL, Carputo D, Peloquin SJ (2004) Substitutes for genome differentiation in tuber-bearing *Solanum*: interspecific pollen-pistil incompatibility, nuclear-cytoplasmic male sterility, and endosperm. Theor Appl Genet 109:1369–76. doi: 10.1007/s00122-004-1753-2

Cardi T, D'Ambrosio E, Consoli D (1993) Production of somatic hybrids between frost-tolerant *Solanum commersonii* and *S. tuberosum*: characterization of hybrid plants. Theor Appl Gen 87:193–200.

Carputo D (2003) Cytological and breeding behavior of pentaploids derived from 3x x 4x crosses in potato. Theor Appl Genet 106:883–8. doi: 10.1007/s00122-002-1165-0

Carputo D (1999) Post-zygotic gametic selection due to endosperm balance number explains unusual chromosome numbers of 3x x 2x progeny in *Solanum*. Sex Plant Reprod 12:27–31. doi: 10.1007/s004970050168

Carputo D, Aversano R, Barone A, et al (2009) Resistance to *Ralstonia solanacearum* of sexual hybrids between *Solanum commersonii* and *S. tuberosum*. Am J Potato Res 86:196–202. doi: 10.1007/s12230-009-9072-4

Carputo D, Barone A, Cardi T, et al (1997) Endosperm balance number manipulation for direct *in vivo* germplasm introgression to potato from a sexually isolated relative (*Solanum commersonii* Dun.). Proc Natl Acad Sci U S A 94:12013–7. doi: VL - 94

Carputo D, Barone A, Frusciante L (2000a) 2N Gametes in the Potato: Essential Ingredients for Breeding and Germplasm Transfer. TAG Theor Appl Genet 101:805–813. doi: 10.1007/s001220051547

Carputo D, Basile B, Cardi T, Frusciante L (2000b) *Erwinia* resistance in backcross progenies of *Solanum tuberosum* x *S. tarijense* and *S. tuberosum* (+) *S. commersonii* hybrids. Potato Res 43:135–142.

Carputo D, Cardi T, Frusciante L, Peloquin SJ (1995) Male fertility and cytology of triploid hybrids between tetraploid *Solanum commersonii* (2n = 4x = 48, 2EBN) and Phureja - Tuberosum haploid hybrids. Euphytica 83:123–129.

Carputo D, Frusciante L (2011) Classical genetics and traditional breeding. In: Bradeen J, Kole C (eds) Genetics Genomics and Breeding of Potato. CRC Press, pp 20–40

Carputo D, Frusciante L, Monti L, et al (2002) Tuber quality and soft rot resistance of hybrids between *Solanum tuberosum* and the incongruent wild relative *S. commersonii*. Am J Potato Res 79:345–352.

Carputo D, Garreffa P, Mazzei M, et al (1998) Fertility of somatic hybrids *Solanum commersonii* (2x, 1EBN) (+) *S. tuberosum* haploid (2x, 2EBN) in intra-and inter-EBN crosses. Genome 781:776–781.

Carputo D, Parisi M, Consiglio F, et al (2003a) Aneuploid hybrids from 5x – 4x crosses in potato: Chromosome number, fertility, morphology and yield. Am J Potato Res 80:93–101. doi: 10.1007/BF02870208

Carputo D, Terra A, Barone A, et al (2003b) Glycoalkaloids and acclimation capacity of hybrids between *Solanum tuberosum* and the incongruent hardy species *Solanum commersonii*. Theor Appl Genet 107:1187–94. doi: 10.1007/s00122-003-1356-3

Caruso I, Castaldi L, Caruso G, et al (2008) Breeding potential of *Solanum tuberosum* – *S. commersonii* pentaploid hybrids: fertility studies and tuber evaluation. Euphytica 164:357–363. doi: 10.1007/s10681-008-9673-x

Chen L, Guo X, Xie C, et al (2013) Nuclear and cytoplasmic genome components of *Solanum tuberosum* + *S. chacoense* somatic hybrids and three SSR alleles related to bacterial wilt resistance. Theor Appl Genet 126:1861–72. doi: 10.1007/s00122-013-2098-5

Datema E, Mueller LA, Buels R, et al (2008) Comparative BAC end sequence analysis of tomato and potato reveals overrepresentation of specific gene families in potato. BMC Plant Biol 8:34. doi: 10.1186/1471-2229-8-34

de Boer JM, Borm TJA, Jesse T, et al (2011) A hybrid BAC physical map of potato: a framework for sequencing a heterozygous genome. BMC Genomics 12:594. doi: 10.1186/1471-2164-12-594

de Boer JM, Datema E, Tang X, et al (2015) Homologues of potato chromosome 5 show variable collinearity in the euchromatin, but dramatic absence of sequence similarity in the pericentromeric heterochromatin. BMC Genomics 16:374. doi: 10.1186/s12864-015-1578-1

den Nijs TPM, Peloquin SJ (1977) 2N Gametes in Potato Species and Their Function in Sexual Polyploidization. Euphytica 26:585–600. doi: 10.1007/BF00021684

Dvorak J (1983) Evidence of genetic suppression of heterogenetic chromosome pairing in polyploid species of *Solanum*, sect. Petota. Can J Genet Cytol 25:530–539.

Galvan G, Franco Fraguas L, Quirici L, et al (2006) *Solanum commersonii*: una especie con gran potencial para el mejoramiento genético de papa por resistencia a *Ralstonia solanacearum*. In: Procisur (ed) Avances de investigación en recursos genéticos en el cono sur II. Montevideo, pp 87–101

Germanà MA (2011) Anther culture for haploid and doubled haploid production. Plant Cell Tissue Organ Cult 104:283–300. doi: 10.1007/s11240-010-9852-z

González M (2010) La resistencia a la marchitez bacteriana de *Solanum commersonii* Dun. y su utilización en el mejoramiento genético de papa. Tesis de Maestría en Ciencias Agrarias, Facultad de Agronomía, Universidad de la Republica, Uruguay

Hamernik AJ, Hanneman RE, Jansky SH (2009) Introgression of wild species germplasm with extreme resistance to cold sweetening into the cultivated potato. Crop Sci 49:529. doi: 10.2135/cropsci2008.04.0209

Hanneman RE (1989) The potato germplasm resource. Am Potato J 66:655–667.

Hardigan MA, Crisovan E, Hamiltion JP, et al (2016) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. Plant Cell 28:TPC2015-00538-RA. doi: 10.1105/tpc.15.00538

Hawkes JG (1990) The potato: evolution, biodiversity and genetic resources. Belhaven, London, UK

Iorizzo M, Aversano R, Bradeen J, et al (2011) Fertilization fitness and offspring ploidy in 3x × 2x matings in potato. Plant Biosyst 146:317–321. doi: 10.1080/11263504.2011.620640

Iovene M, Barone A, Frusciante L, et al (2004) Selection for aneuploid potato hybrids combining a low wild genome content and resistance traits from *Solanum commersonii*. Theor Appl Genet 109:1139–46. doi: 10.1007/s00122-004-1741-6

Iovene M, Wielgus SM, Simon PW, et al (2008) Chromatin structure and physical mapping of chromosome 6 of potato and comparative analyses with tomato. Genetics 180:1307–1317. doi: 10.1534/genetics.108.093179

Jacobsen E, Schouten HJ (2007) Cisgenesis strongly improves introgression breeding and induced translocation breeding of plants. Trends Biotechnol. 25:219–223.

Jansky SH (2000) Breeding for Disease resistance in potato. In: Janick J (ed) Plant Breeding Reviews. 19, pp 69–152

Jansky SH (2006) Overcoming hybridization barriers in potato. Plant Breed 125:1–12.

Jansky SH (2009) Breeding, Genetics, and Cultivar Development. In: Singh J, Kaur L (eds) Advances in Potato Chemistry and Technology, First Edit. Elsevier Ltd., pp 27–62

Jansky SH, Dempewolf H, Camadro EL, et al (2013) A Case for Crop Wild Relative Preservation and Use in Potato. Crop Sci 53:746. doi: 10.2135/cropsci2012.11.0627

Jansky SH, Peloquin SJ (2006) Advantages of Wild Diploid *Solanum* Species Over Cultivated Diploid Relatives in Potato Breeding Programs. Genet Resour Crop Evol 53:669–674. doi: 10.1007/s10722-004-2949-7

Johnston S, Nijs T Den, Peloquin S, Hanneman R (1980) The significance of genic balance to endosperm development in interspecific crosses. Theor Appl Gen 57:5–9.

Kloosterman B, Abelenda JA, Gomez MDMC, et al (2013) Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495:246–50. doi: 10.1038/nature11912

Leisner CP, Hamilton JP, Crisovan E, et al (2018) Genome sequence of M6, a diploid inbred clone of the high glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. Plant J. doi: 10.1111/tpj.13857

Lightbourn GJ, Veilleux RE (2007) Production and Evaluation of Somatic Hybrids Derived from Monoploid Potato. Am J Potato Res 84:425–435. doi: 10.1007/BF02987188

Lou Q, Iovene M, Spooner DM, et al (2010) Evolution of chromosome 6 of *Solanum* species revealed by comparative fluorescence in situ hybridization mapping. Chromosoma 119:435–442. doi: 10.1007/s00412-010-0269-6

Masuelli RW, Camadro EL (1997) Crossability relationships among wild potato species with different ploidies and Endosperm Balance Numbers (EBN). Euphytica 94:227–235. doi: 10.1023/A:1002996131315

Masuelli RW, Camadro EL, Mendiburu AO (1992) 2n gametes in *Solanum commersonii* and cytological mechanisms of triplandroid formation in triploid hybrids of *Solanum commersonii* x *Solanum gourlayi*. Genome 35:864–869. doi: 10.1139/g92-131

Mendiburu AO, Peloquin SJ (1977) Bilateral sexual polyploidization in potatoes. Euphytica 26:573–583. doi: 10.1007/BF00021683

Mok DWS, Peloquin SJ (1975) Three mechanisms of 2n pollen formation in diploid potatoes. Can J Genet Cytol 17:217–225.

Murovec J, Bohanec B (2012) Haploids and Doubled Haploids in Plant Breeding. Plant Breed 88–106. doi: 10.5772/29982

Novy R, Hanneman R (1991) Hybridization between Gp. Tuberosum haploids and 1EBN wild potato species. Am Potato J 151–169.

Ortiz R (1998) Potato breeding via ploidy manipulations. Plant breeding Reviews (16): 15-86. https://doi.org/10.1002/9780470650110.ch2

Ortiz R, Simon P, Jansky SH, Stelly D (2009) Ploidy manipulation of the gametophyte, endosperm and sporophyte in nature and for crop improvement: a tribute to Professor Stanley J. Peloquin (1921-2008). Ann Bot 104:795–807. doi: 10.1093/aob/mcp207

Panahandeh J (2010) Dihaploid induction ability of three clones of *Solanum phureja* (2n = 2x = 24) in interploidy cross with *S. tuberosum* (2n = 4x = 48). Acta Agron Hungarica 58:49–54. doi: 10.1556/AAgr.58.2010.1.6

Peloquin SJ, Jansky SH, Yerk GL (1989) Potato cytogenetics and germplasm utilization. Am Potato J 66:629–638.

Peters SA, Bargsten JW, Szinay D, et al (2012) Structural homology in the Solanaceae: analysis of genomic regions in support of synteny studies in tomato, potato and pepper. Plant J 71:602–14. doi: 10.1111/j.1365-313X.2012.05012.x

Ramsay G, Bryan G (2011) *Solanum*. In: Kole C (ed) Wild Crop Relatives: Genomic and Breeding Resources, Vegetables. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 259–271

Rokka V-M (2009) Potato haploids and breeding. In: Touraev A (ed) Advances in Haploid Production in Higher Plants. Springer Netherlands, pp 199–208

Ross H. (1986) Potato breeding–problems and perspectives. Advances in plant breeding. J Plant Breed Supplement:75–82.

Santini M, Camadro E, Marcellan O, Erazzu L (2000) Agronomic characterization of diploid hybrid families derived from crosses between haploids of the common potato and three wild Argentinian tuber-bearing species. American Journal of Potato Research 77 (4): 211–218. 10.1007/BF02855788

Sharma SK, Bolser D, de Boer J, et al (2013) Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3 Genes, Genomes, Genet 3:2031–47. doi: 10.1534/g3.113.007153

Shearer LA, Anderson LK, de Jong H, et al (2014) Fluorescence *in situ* hybridization and optical mapping to correct scaffold arrangement in the tomato genome. G3 Genes, Genomes, Genet 4:1395–1405. doi: 10.1534/g3.114.011197

Simko I, Jansky S, Stephenson S, Spooner DM (2009) Genetics of resistance to pests and disease. In: Vreugdenhil D (ed) Potato Biology and Biotechnology: Advances and Perspectives. pp 117–155

Szinay D, Bai Y, Visser R, De Jong H (2010) FISH applications for genomics and plant breeding strategies in tomato and other solanaceous crops. Cytogenet Genome Res 129:199–210. doi: 10.1159/000313502

Szinay D, Wijnker E, van den Berg R, et al (2012) Chromosome evolution in *Solanum* traced by cross-species BAC-FISH. New Phytol 195:688–698. doi: 10.1111/j.1469-8137.2012.04195.x

Tang X, De Boer JM, Van Eck HJ, et al (2009) Assignment of genetic linkage maps to diploid *Solanum tuberosum* pachytene chromosomes by BAC-FISH technology. Chromosom Res 17:899–915. doi: 10.1007/s10577-009-9077-3

Tang X, Szinay D, Lang C, et al (2008) Cross-species bacterial artificial chromosome-fluorescence *in situ* hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics 180:1319–1328. doi: 10.1534/genetics.108.093211

Traini A, Iorizzo M, Mann HS, et al (2013) Genome Microscale Heterogeneity among Wild Potatoes Revealed by Diversity Arrays Technology Marker Sequences. Int J Genomics 2013:257218. doi: 10.1155/2013/257218

Tucci M, Carputo D, Bile G, Frusciante L (1996) Male fertility and freezing tolerance of hybrids involving *Solanum tuberosum* haploids and diploid *Solanum* species. Potato Res 39:345–353. doi: 10.1007/BF02357938

Upadhya MD, Cabello R (1997) Interaction of maternal clones of potato (*Solanum tuberosum*) with pollinator clones of *S. phureja* for induced parthenogenesis. Impact on a changing world Program. Report 1997–98. International Potato Center, Lima, pp 179–184

van Os H, Andrzejewski S, Bakker E, et al (2006) Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics 173:1075–87. doi: 10.1534/genetics.106.055871

Veilleux R (2005) Cell and tissue culture of potato (Solanaceae). In: Razdan M, Mattoo A (eds) Genetic improvement of Solanaceous Crops Volume I: Potato. Science Publishers, Enfield, NH, USA, pp 185–208

Zhao Q, Zhao B, Zhang Q, et al (2013) Screening for Chip-Processing Potato Line from Introgression of Wild Species' Germplasms with Post-Harvest Storage and Chip Qualities. Am J Potato Res 90:425–439. doi: 10.1007/s12230-013-9316-1

Honrada eres
como
una mano
que trabaja en la tierra,
(...) enemiga del hambre,
en todas las naciones
se enterró su bandera
vencedora
(...) harina de la noche
subterránea,
tesoro interminable
de los pueblos.
Pablo Neruda
("Oda a la papa", 1954)

Chapter 2

Introgressive Hybridization in Potato Revealed by Novel Cytogenetic and Genomic Technologies

Authors:
Paola Gaiero[1], Pablo Speranza[1], Hans de Jong[2]

1. Department of Plant Biology, Facultad de Agronomia, Universidad de la Republica. Garzon 780, PC 12900, Montevideo, Uruguay
2. Laboratory of Genetics, Wageningen University & Research, Droevendaalsesteeg 1, P.O. Box 16, 6708 PB, Wageningen, The Netherlands.

Published in:
American Journal of Potato Research, 1–15. doi:10.1007/s12230-018-9669-6.

Abstract

Potato is the third most important food crop in the world and is crucial to ensure food security. However, increasing biotic and abiotic stresses jeopardize its stable production. Fortunately, breeders count on a rich pool of wild relatives that provide sources for disease resistance and tolerance to environmental stresses. To use such traits effectively, breeders require tools that facilitate exploration and exploitation of the genetic diversity of potato wild relatives. Introgression programs to incorporate alien chromatin into the crop have so far relied on cytogenetic and genetic studies to tap desired traits from these wild resources. The available genetic and cytogenetic tools, supplemented with more recent genomic technologies, can assist in the use of potato relatives in pre-breeding. This information can also facilitate cisgenesis and genome editing to improve potato cultivars. Despite the abundant and rapidly growing genomic information of potato, that of its wild relatives is still limited.

The use of wild relatives in potato breeding

Potato is one of the major crops in the world and is viewed as a key source to ensure food security of its fast-growing population. The crop can produce high yields with limited inputs, and supplies at the same time a good source of energy and health-promoting nutrients (Birch et al. 2012). However, increasing biotic and abiotic stresses represent a serious and constant risk for food security and so jeopardize stable production (Bradshaw 2007a).

The genetic diversity of cultivated potato that may provide allelic resources for controlling such stresses has been substantially reduced in the process of domestication and selection. Only a few clones of tetraploid cultivated *Solanum tuberosum* from the Andes were introduced to Europe and though they must have contained a lot of genetic variation, the available biodiversity was only partially captured (Hawkes 1990; Spooner et al. 2005; Ríos et al. 2007; J.M. Bradeen and Haynes 2011; Ramsay and Bryan 2011; Birch et al. 2012; Kloosterman et al. 2013). This limited genetic diversity was further reduced due to genetic bottlenecks during photoperiod adaptation and losses resulting from viruses and the late blight epidemics of 1845-1846 (Bethke et al. 2017). However, cultivated potato and its wild relatives signify a more diverse (Figure 1) and accessible germplasm resource than that of any other crop (Ross 1986; Hanneman 1989; Peloquin et al. 1989; Hawkes 1990). Their value as breeding material is given by their wide geographical distribution and great range of ecological adaptation (Figure 1) (Hawkes 1994), together with their availability through the Inter-genebank Potato Database (IPD) (http://germplasmdb.cip.cgiar.org) established by the CIP (International Potato Centre) and the Association for Potato Intergenebank Collaboration. To use potato wild relatives (WR) efficiently to expand its genetic base, breeders require tools that facilitate exploration and exploitation of their genetic diversity.

This diversity coming from potato WR can transfer specific traits to potato by introgressive hybridization. It involves the introduction of alien chromatin carrying a gene

Figure 1. Diversity in flowers, fruits, tubers, plants and habitats of sympatric potato wild relatives. a) Typical diploid *Solanum commersonii*, b) Triploid *S. commersonii*. c) Intermediate morphotype, possibly a triploid hybrid between *S. commersonii* and *S. chacoense*, d) Typical diploid *S. chacoense*.

of interest from a wild relative to the crop genome. After the interspecific hybridization and repeated backcrossings, the selected gene(s) of interest are incorporated into the crop chromosomes by homoeologous recombination. The offspring are then selected for the desired trait while the wild genetic background is removed by selection in consecutive backcross generations as far as possible. Linkage drag may occur when the introgressed chromatin still contains tightly linked wild traits from the ancestral donor that cannot be removed by recombination (Ramsay and Bryan 2011).

An alternative approach is genetic base broadening (Bradshaw 2016), which favours allelic variation besides incorporating genes of interest, and thus maximizes the heterozygosis and epistasis required for yield improvement (Mendoza and Haynes 1974), but completely loses the genetic background of the original cultivar. Base broadening, which is often the underlying objective of breeders (Bradshaw 2007b), uses the broadest possible starting material and depends on recombination between the parental genomes in the hybrid. It is then followed by weak selection in target environments but requires enough time to produce advanced backcrosses of improved material that can be crossed with elite germplasm without negative effects on yield and agronomic performance. This process results in improved genotypes that can be used as parents in breeding programs (Bradshaw 2016).

Determining existing introgression events in potato cultivars

Several reports of natural hybrids suggest that potato WR readily hybridize in the wild (Spooner and Hijmans 2001; Camadro 2012; Spooner et al. 2014). Examples of such events include the triploid hybrids between *Solanum commersonii* and *S. chacoense*, or *S. commersonii* and *S. gourlayi* (Masuelli and Camadro 1992; Ortiz 1998). When samples are collected from natural populations, these may carry introgressions from other wild species (Camadro 2012; Spooner et al. 2014; Bethke et al. 2017). Such introgressed segments represent a source of variability through new allele combinations but also a challenge for the *ex situ* conservation and utilization of potato WR.

Potato wild relatives like diploid *Solanum bulbocastanum*, *S. stoloniferum* and *S. chacoense* or hexaploid *S. demissum* (Pavek and Corsini 2001) have been extensively used in potato introgressive hybridization breeding (Hanneman 1989; Peloquin et al. 1989; Watanabe et al. 1994; Jansky 2000; Pavek and Corsini 2001; Bradshaw et al. 2006; Bradshaw 2007a; Bradshaw 2007b; Bradshaw and Ramsay 2009; Jansky 2009; Bradshaw and Bonierbale 2010; Ramsay and Bryan 2011). Such taxa not only display various advantages over cultivated germplasm (Jansky and Peloquin 2005), such as resistance to the potato late blight, caused by *Phytophthora infestans* and other diseases caused by bacteria and viruses (Jansky 2000; Simko et al. 2009b), they also provide the genetic basis for tolerance to cold, frost and other environmental stresses. It is widely accepted that many modern cultivars have wild species donors in their pedigrees (Love 1999). Andean farmers allow wild populations of potato species to grow on their fields, so wild germplasm is introduced into both diploid and tetraploid cultigens (Ugent 1970). Moreover, the use of potato WR in introgressive hybridization breeding before the existence of common pedigree records implies that the original introgression events have not been documented and that the sources of certain desirable traits are unknown (Love 1999; Leisner et al. 2018).

One of the direct methods to demonstrate introgressed alien chromatin in the crop chromosomes is comparative chromosome painting by Fluorescent *in situ* Hybridization (FISH), that establishes the structural and numerical comparisons of chromosome sets between species of the genus *Solanum* (Tang et al. 2008; Iovene et al. 2008; Szinay et al. 2008; Szinay et al. 2010; Lou et al. 2010; Verlaan et al. 2011; Szinay et al. 2012). Under low stringency conditions, it is possible to use tomato or potato probes in these experiments to perform cross-species chromosome painting and to display homoeologous chromosomal positions in related *Solanum* species. In this way, many hitherto unknown inversions could be described (Tang et al. 2008; Lou et al. 2010; Szinay et al. 2010; Peters et al. 2012; Szinay et al. 2012). BAC-FISH also allowed the accurate mapping of the *Ty*-1 gene introgressed from *S. chilense* into cultivated tomato and provided an explanation for observed linkage drag resulting from suppression of recombination (Verlaan et al. 2011). There are no such studies in potato cultivars, although there are many reports of intro-

gressions based on molecular markers (Hosaka 1995; Bryan et al. 1999; Provan et al. 1999; van der Voort et al. 1999; Gebhardt et al. 2004; Flis et al. 2005; Sokolova et al. 2011). Resequencing studies in tomato have identified polymorphisms related to introgressions (Causse et al. 2013; Aflitos et al. 2014), while in potato, these have been shown in some diploid and tetraploid landraces as well as in cultivars (Hardigan et al. 2017). Bioinformatic tools like iBrowser (Aflitos et al. 2015) have been developed to use SNPs identified from the increasing genome sequence data available to pinpoint past undescribed introgressions from wild relatives in the genomes of cultivated *Solanum* species. These approaches together with other modern technologies will also prove useful when designing new introgressive hybridization schemes.

Tools for establishing introgressive hybridizations

In spite of the widely available diversity in germplasm collections worldwide, only 10 % of the potato species have been explored for use in breeding programs (Bradshaw 2007a). This is a rather low percentage, bearing in mind that by manipulation of ploidy and other biotechnological interventions, virtually any potato species can be used in introgressive hybridization breeding (Ortiz 1998; Jansky 2006; Ortiz et al. 2009). Moreover, the few species that have been employed in breeding programs to provide specific traits have not been investigated systematically.

Knowledge of genome organization and divergence between potato and its wild relatives is most helpful to create new introgressive hybridization schemes. Before choosing a wild relative as donor, it is important to know if there are inversions or translocations that will impede introgression or cause linkage drag (Figure 2). Another key aspect is to always take into account hybridization barriers that have been thoroughly reviewed elsewhere (Camadro et al. 2004; Jansky 2009; Bethke et al. 2017). The great potential recognized in these wild relatives encouraged scientists to develop strategies for overcoming such barriers (Jansky 2006; Bradshaw and Bonierbale 2010; Bethke et al. 2017). Once the crossing barriers are overcome, stabilizing the introgression in the potato genotypes still represents a challenge due to its tetraploid inheritance. Additionally, inbreeding depression forces breeders to use different genotypes as recurrent parents for backcross progenies. Despite all these obstacles to recover a superior cultivated background after hybridization with a wild species, the value of these potato WR makes it worth the effort. A wide variety of cytogenetic, genetic and genomic tools can be used to assist in these efforts.

Classical cytogenetics tools

Classical cytogenetics should be the first tool to study potato WR to be used as donors and their hybrids with potato. It helps to establish ploidy levels and Endosperm Balance Number (EBN) of the interspecific hybrids (Peloquin et al. 1989; Jansky 2009; Ono and Hosaka 2010) and to assess the effects of ploidy changes in the parental species and their hybrids (Mok and Peloquin 1975; Adiwilaga and Brown 1991; Carputo et al. 1997; Ortiz 1998; Carputo et al. 2000; Jansky 2006; Lightbourn and Veilleux 2007; Jansky 2009; Ortiz et al. 2009). Direct cytogenetic analysis is also the most direct approach to observe meiotic chromosome pairing behaviour in interspecific hybrids and their progenies (de Jong et al. 1993; Carputo et al. 1995; Masuelli and Tanimoto 1995; Barone et al. 1999; Carputo 2003; Chen et al. 2004; Gaiero et al. 2017, chapter 3).

Genome differentiation between wild potato species is assumed to play a minor role as an isolation mechanism (Dvorak 1983; Camadro et al. 2004). For potato and its relatives, genomic formulas were proposed by Matsubayashi (1991) who distinguished five genomes in Section Petota (A, B, C, D and P) through classical genome analysis of meiotic behaviour and pollen fertility in interspecific hybrids. Genomes identified with different letters show little or no pairing in the meiotic prophase I of their amphidiploid hybrids, which display pollen sterility. The most common genome is type A, which presents different degrees of structural variants depending on the scale of the chromosomal rearrangements. Genome E is proposed for the closely related non-tuber-bearing species of the Section Etuberosum. For potato breeders, homoeologous pairing and crossovers in their hybrids with cultivated potatoes is of more interest than their phylogenetic relationships because it predicts their success for breeding via crossing. The most stringent test for pairing between homoeologous chromosomes is the analysis of meiosis of triploid hybrids. If the chromosomes form trivalents in the meiosis of triploids, then introgression is possible (Jansky 2006). The chances of alien chromatin introgression are greater as homoeologous pairings are more likely to occur in the triploid compared to the meiotic pairing in tetraploids in which potentially there is always a homolog for each chromosome which may pair preferentially (Sybenga 1996; Jansky 2006). Although the factors that determine homeologous pairing are not clear yet, homology in repetitive sequences seems to play an important role. Genome divergences caused by repeats have been assessed within and between the potato and tomato clades, characterizing the abundance and dynamics of the repetitive fractions of their genomes (Gaiero et al. in prep, chapter 5). These can have significant consequences in genome homology, genome expansion and in the occurrence and impact of structural rearrangements.

Molecular cytogenetics tools

Molecular cytogenetics has been one of the major instruments in tracing the course of alien chromosomes in introgressive hybridization breeding. It has been applied for most

major crop species (Benavente et al. 2008), including potato (Yeh and Peloquin 1965; Mok et al. 1974; Pijnacker and Ferwerda 1984; Visser and Hoekstra 1988; Mohanty et al. 2004; Gavrilenko 2007). It allows identification of whole chromosome sets or of specific chromosome pairs and it also enables comparisons of the chromosomal positions of markers or regions of interest across related species.

Genome painting or GISH (Genomic *in situ* hybridization) is a powerful FISH technique used for tracing homoeologous chromosome pairing, recombination and transmission. It consists in labelling genomic DNAs from one or both parental species as probe(s) to

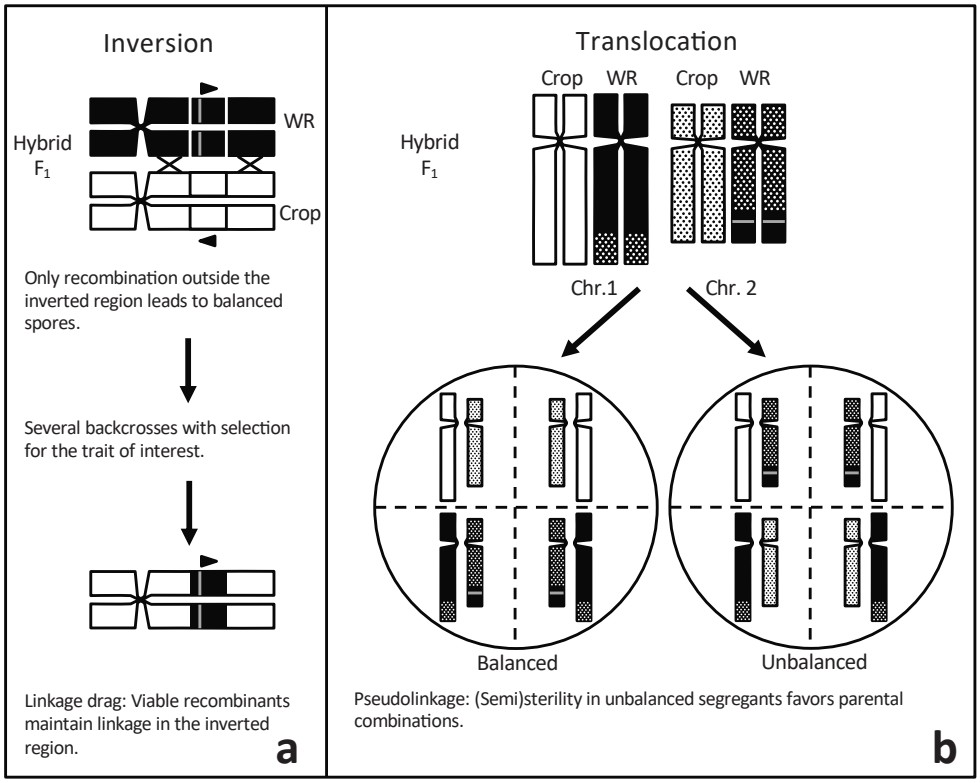


Figure 2 shows two simplified examples of structural chromosome rearrangements and the consequences it may have for plant genetics and breeding. a) In the case of a (paracentric) inversion the end of the long arm is inverted. In the hybrid in which one chromosome contains the inversion, the end chromosomes fail to pair in this region or form a loop structure. Any crossover that may arise will result in a sterile spore, so that this region remains non-recombinant in the next generation, a genetic phenomenon known as linkage drag. b) In the case of a plant with a translocation complex, fragments between two non-homologous chromosomes are swapped. In a meiotic division two cases can be considered: a balanced segregation in which the two normal (non-translocated) chromosomes and the two translocated chromosomes go to opposite poles, resulting in balanced viable spores. In the second case a normal chromosome and a translocated chromosome arrive in the same daughter cell, which is unbalanced and hence results in a sterile spore. In both cases partial sterility is the result.

hybridize on chromosome slides of the interspecific hybrid and its progeny (Figure 3a). If the genomes of the parental species (especially their dispersed and tandem repeats) have diverged sufficiently, chromosomes of the two species can be easily discriminated in the hybrid nuclei through different fluorescent dyes. In nuclei from interspecific (sexual or somatic) hybrids between potato (genome A) and non-tuber bearing relatives (genomes E, B or P) GISH has been successful in discriminating chromosomes (Dong et al. 1999; Dong et al. 2001; Gavrilenko et al. 2002; Gavrilenko et al. 2003; Dong et al. 2005).) In wider hybrids such as *S. nigrum* (+) *S. tuberosum* and its backcrosses (Horsman et al. 2001), alien chromosomes are easily distinguishable. These studies provide further evidence of genome differentiation, in the sense that genomes identified with different letters not only do not pair in the meiosis of their hybrids but also can be discriminated by GISH. When divergence is not so high, contrast in the hybridization differentiation can be improved by adjusting washing stringency and proportion of blocking (unlabelled) DNA in the FISH experiments (Jiang and Gill 1994). However, there is a technical limit to what can be discriminated by GISH. As an example, the technology has not been successful in studies of hybrids between potato and its closer A-genome tuber-bearing wild relatives, with the exception of *S. bulbocastanum*, a diploid (1EBN, A^b genome) Mexican species (Iovene et al. 2007). Hybrids between *S. commersonii* and *S. tuberosum* Group Phureja behaved as near autopolyploids during male meiosis and it was not possible to discriminate the chromosomes coming from each parental species through GISH (Gaiero et al. 2017, chapter 3). These results suggest that repetitive sequences have not diverged much among the genomes of cultivated and wild potatoes.

Chromosome rearrangements between related species can cause specific problems at different meiotic stages of their interspecific hybrids (Figure 2). Typical examples are heterozygosity for paracentric inversions which can cause anaphase I (and or II) bridges and hence sterility or aneuploidy (Figure 2a) or reciprocal translocations, which can lead to semi-sterility or aneuploidy (Figure 2b). Such rearrangements between the homoeologues represent an important limitation in introgressive breeding due to the unintended retention of large blocks of DNA surrounding a gene of interest (Figure 2a), genetically described as linkage drag (Jacobsen and Schouten 2007). There are no reports of large scale chromosome rearrangements in potato and its wild relatives, in contrast to some detailed descriptions of rearrangements among Solanaceous crops (Iovene et al. 2008; Tang et al. 2008; Lou et al. 2010; Peters et al. 2012; Szinay et al. 2012). Linkage drag represents a limitation for introgressive hybridization breeding because blocks of alien chromatin surrounding the gene of interest can be retained even after many generations of backcrossing. The impact of genetic drag is even greater when the linked regions have a negative effect on agronomic performance (Figure 2a). Selection against such undesired blocks can be simplified by marker-assisted breeding and genomic selection, but these approaches are still challenging (Warschefsky et al. 2014), especially in autotetraploid genotypes. While the literature on comparative FISH analysis in *Solanum*

is vast, there are only a few studies that have utilized sufficient high-quality sequence data needed to reveal fine-scale structural differences related to introgression barriers (Datema et al. 2008; Peters et al. 2012; Causse et al. 2013; Aflitos et al. 2014; Aflitos et al. 2015; de Boer et al. 2015).

A considerable body of literature on FISH applications for breeding is available for tomato as reviewed by Szinay et al. (2010), with reports of many structural rearrangements among tomato and its wild relatives that have significant impacts on breeding (Anderson et al. 2010; Verlaan et al. 2011; Szinay et al. 2012). With multi-colour fluorescence microscopy it is possible to hybridize many probes each labelled with a different fluorescent dye in a single experiment, reducing the examination of any chromosome set to only a few experiments (Tang et al. 2009; Szinay et al. 2010). For potato, a set of chromosome-specific cytogenetic DNA markers (CSCDM) made from DNA probes selected from a *S. bulbocastanum* library was developed to associate all twelve linkage groups to potato chromosomes and build a reference karyotype (Dong et al. 2000; Song et al. 2000). The RHPOTKEY BAC library was developed for the diploid potato clone RH89-039-16 (Borm 2008). The positions of the BACs in this library were anchored to AFLP markers in the ultrahigh density (UHD) genetic map (van Os et al. 2006). From this library, a set of 60 BACs with known positions in the Ultra High Density (UHD) linkage map were selected for localization on pachytene chromosomes, thus providing a useful tool to study collinearity between potato and its wild relatives (Tang et al. 2009; Achenbach et al. 2010; de Boer et al. 2011). Gaiero et al. (2016, chapter 4) used this BAC set to build cytogenetic maps for *S. commersonii* and *S. chacoense* and to compare them to that of cultivated potato. Their results indicate a high collinearity at the chromosomal scale between the three species which makes them promising donors in introgressive hybridization schemes. They also used them to identify specific chromosome pairs in triploid hybrids between *S. commersonii* and *S. tuberosum* Group Phureja (Figure 3b).

Using a higher number of BAC probes which are located closer together in linkage maps, high resolution cytogenetic mapping has been employed to describe rearrangements in chromosome 6 coming from potato and tomato (Iovene et al. 2008; Tang et al. 2008). Such fine mapping has also been useful to design strategies for the sequencing projects of these crops (Potato Genome Sequencing Consortium. 2011; The Tomato Genome Consortium 2012). They have helped by identifying the boundaries between the highly condensed heterochromatin and euchromatin, which is easier to sequence and assemble (Szinay et al. 2008; Tang et al. 2009; Szinay 2010; Tang et al. 2014). Thus, the sequencing and assembly efforts could be better directed to the euchromatic regions. Because these crops were sequenced using BAC libraries, cytogenetic maps have also been used to construct the backbone for sequence assembly of tomato (Szinay et al. 2008) and potato (Visser et al. 2009). The seed BAC clones that were chosen to start the assembly were confirmed through BAC-FISH, the BAC positions on genetic and physical maps were verified and gaps in the assembly were identified and sized (Iovene et al. 2008; Szinay et

al. 2008; Tang et al. 2008; Tang et al. 2009). The use of a very large number of BAC-FISH probes, coupled with information from physical maps (restriction and optical maps) allowed for the correction of miss-assemblies in the tomato genome and identification of scaffolds that were not in the correct order or orientation, or both (Shearer et al. 2014). The potato assembly was similarly improved using information from physical and genetic maps to achieve a reference genome at the level of pseudomolecules, which are equivalent to chromosomes (Sharma et al. 2013; Hardigan et al. 2016). Such integration of technologies has only been applied to date for *S. chacoense* among the potato WR (Leisner et al. 2018).

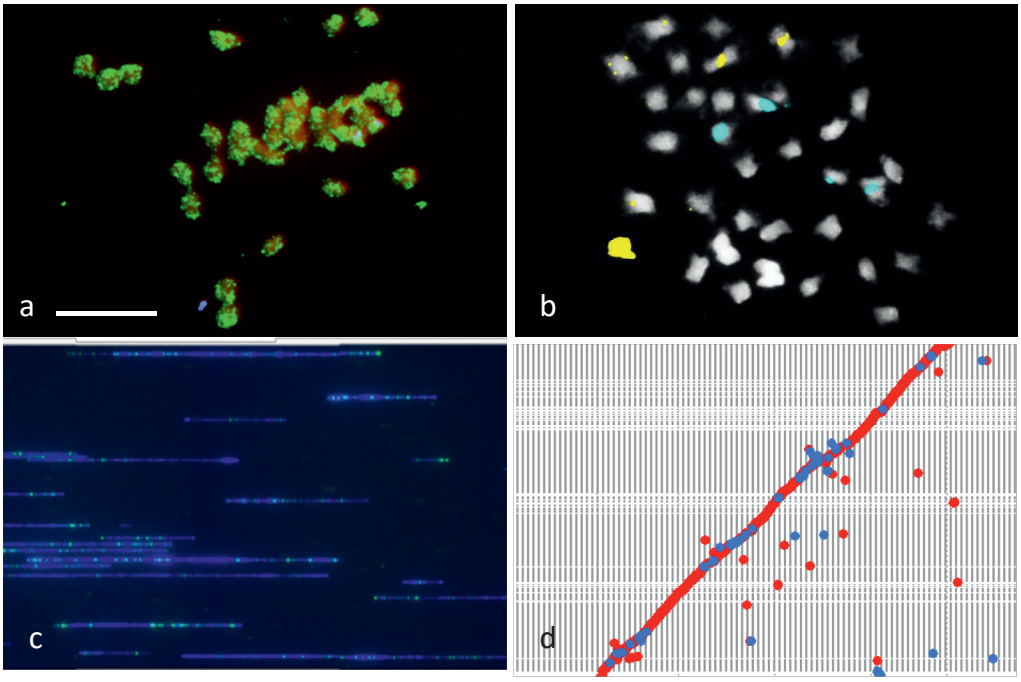


Figure 3. Examples of the various technologies available to assist in introgressive hybridization breeding. a) genome painting (GISH) with *S. commersonii* genomic DNA probe (green) on the chromosomes of a triploid interpecific hybrid (*S. commersonii* x *S. tuberosum* Group Phureja) in a meiotic cell complement. Notice that all chromosomes are hybridized with the probe. b) Chromosome identification through BAC-FISH on the chromosomes of a triploid interpecific hybrid (*S. commersonii* x *S. tuberosum* Group Phureja) in a meiotic cell complement. Three chromosomes from pair 1 are identified with a yellow probe and three chromosomes from pair 2 are identified with a blue probe. c) High molecular weight *Solanum* DNA molecules stained with YOYO (blue) and showing sequence-specific single strand nicks (green), stretched by moving along a nanochannel array. Millions of such images are integrated to build a consensus genome map. d) Dot-plot comparison of the genome assemblies of *Solanum commersonii* and *S. tuberosum* (DM) obtained through the software MUMmer. It shows a high degree of collinearity with a few small inversions (inverted stretches of dots across the diagonal).

Mapping tools

The cutting edge technology of genome mapping through nanochannels (Lam et al. 2012; Cao et al. 2014) is the ideal tool for completing genome assemblies and even identifying, spanning and assembling repeated sequences. In addition to assisting in genome assembly, genome mapping can assess structural variation among related species or genotypes within a species (Cao et al 2014). This technology uses nicking enzymes to create DNA sequence-specific nicks that are subsequently labelled by a fluorescent nucleotide analogue (Xiao et al. 2007). The DNA is linearized by confinement in a nanochannel array (Das et al. 2010) and then photographed (Figure 3c). The DNA loading and imaging cycle can be automatically repeated many times, so data can be obtained at high throughput and high resolution (Hastie et al. 2013). Genome mapping using nanochannels has been used only recently for genome assembly in higher plants such as spinach (Xu et al. 2017), subterranean clover (Kaur et al. 2017), maize (Jiao et al. 2017), quinoa (Jarvis et al. 2017) and bread wheat (Staňková et al. 2016). In the genus *Solanum*, the related method known as optical mapping (Zhou et al. 2004) was used for whole genome analysis in tomato (Shearer et al 2014) and also in other crops like rice (Zhou et al. 2007), maize (Zhou et al. 2009) and for crop relatives such as *Medicago truncatula* (Young et al. 2011). One of the limitations for its use in the higher plant genomics community, is the challenge of obtaining sufficient amounts of high molecular weight nuclear DNA (HMW DNA) due to the thick cell walls and cytoplasmic polyphenols and polysaccharides.

Moving from physical to genetic mapping, considerable effort has been put into mapping traits of interest on the genetic maps of the few potato WR that have been used in potato breeding. Understandably, most attention has been devoted to mapping resistance to late blight (*Phytophthora infestans*), the most important potato disease and responsible for the infamous Irish Potato Famine in 1845-46. The most remarkable source for resistance to *P. infestans* is *S. bulbocastanum* (Naess et al. 2001; Lokossou et al. 2010). Resistance to *P. infestans* was also mapped in *Solanum demissum* (Jo et al. 2011), *S. venturii* (Pel et al. 2009), *S. pinnatisectum* (Kuhl et al. 2001), *S. avilesii* (Verzaux et al. 2011), *S. paucisectum* (Villamon et al. 2005) and in *S. phureja* x *S. stenotomum* (Costanzo et al. 2005; Simko et al. 2006). Different alleles from a single locus on chromosome 8 from *S. bulbocastanum*, which carries resistance to late blight, have been the subject of physical mapping and positional cloning (Bradeen et al. 2003; Song et al. 2003; Van Der Vossen et al. 2003). Resistance to important potato viruses like PVX and PVY has also been the focus of mapping efforts (Cockerham 1970; Solomon-Blackburn and Barker 2001; Flis et al. 2005; Y.-S. Song et al. 2005; Sato et al. 2006; Simko et al. 2009a), together with other traits of interest (Anithakumari et al. 2011) and with pyramiding of resistance genes (Tan 2008). Genome-wide association studies (GWAS) have been particularly useful for complex traits in cultivated potato (Ewing et al. 2004; Gebhardt et al. 2004; Simko 2004; Simko et al. 2004; Simko et al. 2006; Visser et al. 2015), but only one study includes the use of wild relatives (Hardigan et al. 2017). All these efforts have allowed the use of tightly

linked molecular markers to select resistant genotypes or to select against donor genome in backcross progenies from an introgression scheme, in the so-called marker-assisted breeding (reviewed by Barone, 2004; Tiwari et al. 2013). Most literature on selection against wild genome comes from studies on the introgression of *S. commersonii* into a *S. tuberosum* tetraploid background (Barone et al. 2001; Carputo et al. 2002; Barone 2004; Iovene et al. 2004). The greatest impact of these molecular breeding technologies has been on pre-breeding and parental development (De Koeyer et al. 2011) and also on the exploration of germplasm resource (Bamberg and del Rio 2013; Carputo et al. 2013; Manrique-Carpintero et al. 2014; Warschefsky et al. 2014).

Genomics tools

The available genomic knowledge on wild potatoes is relatively limited compared to that of tomato WR (Szinay et al. 2012; Aflitos et al 2014; Bolger et al 2014). The tendency now is to slowly move to more sophisticated genomics of WR, elucidating the available diversity and desirable traits (Bradeen and Haynes 2011; Ramsay and Bryan 2011). The increasing number of molecular markers and DNA sequence data to be generated will allow for faster progress in breeding by simultaneously selecting genes/QTLs while selecting against wild species genome content (Bradshaw 2007b).

With the development of high-throughput DNA sequencing, genome assemblies for tomato (The Tomato Genome Consortium 2012), potato (Potato Genome Sequencing Consortium 2011) and several of their WR (Aflitos et al. 2014; Bolger et al. 2014; Aversano et al. 2015; Leisner et al. 2018) have become available. Concerted genomics and bioinformatics efforts have improved genome assemblies (Sharma et al. 2013; Shearer et al. 2014; Hardigan et al. 2016). However, only a few studies have utilized sufficient high-quality physical maps needed to reveal structural differences (Figure 3d) related to introgression barriers (Peters et al. 2012; Aflitos et al. 2014; Aflitos et al. 2015; de Boer et al. 2015). Although sequence data for some wild species are available (*e.g.*, *S. chacoense*, *S. commersonii*), the only comparative structural analysis performed so far used DArT markers, finding microscale genome sequence variation (Traini et al. 2013). A vast survey of genome-wide sequence variation across a diversity panel of cultivated and wild potato species was performed by Hardigan et al. (2017), finding more variation than in any other crop resequencing project. In most cases of crop wild relatives (CWR) only a draft genome is available and it is of limited use, depending on the quality of the assembly (Pérez-de-Castro et al. 2012). Such is the case of the whole genome draft sequence available for *Solanum commersonii* (Aversano et al. 2015). Assembly to the level of pseudomolecules is achieved when mapping against the reference potato genome. This approach does not cater for structural variation between the two species. In the case of *S. chacoense*, the genotype that was sequenced (M6) was an inbred clone, so increased homozygosity facilitated genome assembly. The construction of pseudomolecules was achieved including information from genetic maps using M6 as parent of the segregating population,

so it does not assume collinearity with a reference genome (Leisner et al. 2018). Ideally, breeders should count on fully assembled and well annotated reference genomes for potato WR to assist in gene discovery and dissection of the genetic basis of a trait.

Making the most of biotechnological approaches through wild relatives

One might argue that resorting to potato WR as donors of desirable traits through introgressive hybridization seems no longer necessary with modern technologies such as cisgenesis or the CRISPR-Cas9 genome editing, as it allows researchers to transfer directly the gene of interest or to change the native sequence into a tailor-made version, respectively. Nevertheless, it is first necessary to identify the original genes conferring the trait of interest and to mine their allele diversity in order to isolate them, clone them and accurately modify them or transfer them into targeted cultivars. This is possible through newly developed genetic and genomic tools (Cardi 2016).

Knowledge on the physical position of the genes of interest is useful to isolate them and transfer them to cultivated potato. The identification, mapping, cloning and the techniques to use resistance genes against *Phytophthora infestans* coming from potato WR was reviewed by Park et al. (2009). Most mapped and cloned genes come from *S. demissum* (Jo et al. 2011) or *S. bulbocastanum* (Naess et al. 2000; Naess et al. 2001; Bradeen et al. 2003; J. Song et al. 2003; Lokossou et al. 2010), although using an interspecific candidate gene approach, Pel et al. (2009) were able to map and clone a dominant allele from an alternative donor (*S. venturii*). These genes have already been used or are in the pipeline for cisgenesis into cultivated potato backgrounds (Haverkort et al. 2008; Park et al. 2009; Zhu et al. 2015) .

Recently, all known major *R* genes in potato have been sequenced and an 'omics' approach was used to recognize the genes responsible for late bight resistance (Van Weymers et al. 2016). The previously developed SolRgene database provides easy access to the sequences of *R* genes across *Solanum* section Petota, allowing the cloning of many of those genes for downstream biotechnological uses (Vleeshouwers et al. 2011). New sources of resistance have been identified and their genes cloned using the latest third generation sequencing technologies. These new variants are now available for biotechnological applications (Witek et al. 2016). The genome sequence and transcriptomes of potato WR like *S. commersonii* and *S. chacoense* have allowed the identification of pathogen-receptor genes and to describe non-acclimated and cold-acclimated gene expression as well as to get insights on tuberization and glycoalkaloid production (Narancio et al. 2013; Aversano et al. 2015; Leisner et al. 2018). A large resequencing effort across potato cultivars and landraces together with potato WR shed light on the kinds of traits and genes that were under selection during the domestication process and provided a useful catalogue of genomic variation within the potato genepool (Hardigan et al. 2017). Microsatellite markers (SSR) transferred from potato to its wild relatives can be used to

screen for genetic variability. An example of this is the evaluation of 10 accessions from *S. chacoense* using 15 SSR markers developed for potato, which showed high levels of heterozygosity in the collection (Haynes et al. 2017). Using sequence data, new SSR markers can be specifically devised for wild species, increasing their amplification success and polymorphic information content. This is what happened for a diversity panel of *S. commersonii* accessions and for a biparental population both screened with SSR markers developed from short read sequence data (Sandro et al. 2016). Adding value to collected samples in gene banks through all this genetic and genomic information and mining allele variation from natural populations or *ex situ* collections will be critical for the efficient use of potato WR in the genomics era.

Breeders can use potato WR to introduce new genes in a commercial cultivar or to select superior alleles to replace their cultivated counterparts through cisgenesis. They can also use structural and functional genomic information on potato WR to adopt as templates to target specific sites and edit gene sequences in elite cultivars. In the case of genome editing, knowing the target genome sequence is essential to prevent targeting of repeated sequences dispersed throughout the genome and to respond to regulatory demands (Cardi 2016). However, most of the time breeders do not aim at transferring only one gene of interest but to broaden the genetic base of a potato cultivar (Bradshaw 2007c; Bradshaw 2016) and to introduce adaptability and hardiness from potato WR usually growing in a wide variety of environments (Bethke et al. 2017). Such a time-consuming process depends on many backcrosses to recover the cultivated background that was lost with the initial hybridization. It is also claimed that the undesirable traits that come from the potato WR are hard to remove, especially in a tetraploid potato background. An idea that is gaining popularity is the use of diploid inbred lines in potato breeding (Lindhout et al. 2011; Endelman and Jansky 2016; Jansky et al. 2016). These allow for easier genetic mapping with increased resolution and simplify genetic analysis because of their disomic inheritance (Endelman and Jansky 2016). In breeding, they can be used to create F1 hybrid seed with enhanced heterosis that can be propagated through true potato seed (Lindhout et al. 2011; Jansky et al. 2016). Potato WR have a role to play both in the development of diploid inbred lines and in their use as breeding material. One of the most frequently used strategies to achieve diploid inbred lines is through the crossing with a *S. chacoense* genotype carrying a dominant self-incompatibility inhibitor allele called *Sli* (Hosaka and Hanneman, Jr. 1998; Phumichai et al. 2005; Lindhout et al. 2011; Jansky et al. 2016). After the diploid inbred lines are obtained, they can generally be crossed directly with diploid potato WR facilitating introgression at the diploid level (Jansky 2006; Jansky et al. 2016).

Many of the limitations in introgressive hybridization breeding can be overcome by an efficient use of new genomic technologies and approaches. These will allow prediction of homology and collinearity to anticipate the degree of pairing, recombination and linkage drag expected in any interspecific cross, together with mining of existing

variation in natural populations and optimal choice of the genotypes to start introgression schemes. Genomics will not only facilitate marker-assisted selection for the traits of interest but also against the wild donor chromatin. To the question posed by Bethke et al. (2017) in their review paper: Are we getting better at using Wild Potato Species in Light of New Tools? The answer is clearly Yes, but the possibilities are still endless. The approaches we are now developing may still seem expensive and difficult to apply to routine breeding; however, information is accumulating fast. At the current rate of technological advance in the automation of data acquisition and analysis, it does not seem impossible to envision the fulfillment of the promises of the use of PWR in the near future, as long as we keep going in that direction.

Acknowledgements

P. Gaiero was supported by grant Proyecto 720 and CSIC Recursos Humanos, University of the Republic. We thank Alejandro Vaco and Germán Abad for kindly providing some of the photographs in Figure 1. We are grateful to Francisco Vilaró for useful comments during the development of this work. We are also thankful to four anonymous reviewers for their comments to improve this work.

Literature cited

Achenbach, U.C., X. Tang, A. Ballvora, H. de Jong, and C. Gebhardt. 2010. Comparison of the chromosome maps around a resistance hot spot on chromosome 5 of potato and tomato using BAC-FISH painting. Genome 53: 103–110. doi:10.1139/G09-086.

Adiwilaga, K.D., and C.R. Brown. 1991. Use of 2n pollen-producing triploid hybrids to introduce tetraploid Mexican wild species germ plasm to cultivated tetraploid potato gene pool. Theoretical and Applied Genetics 81: 645–652. doi:10.1007/BF00226732.

Aflitos, S.A., G. Sanchez-Perez, D. de Ridder, P. Fransz, M.E. Schranz, H. De Jong, and S.A. Peters. 2015. Introgression browser: High-throughput whole-genome SNP visualization. The Plant Journal 82: 174–182. doi:10.1111/tpj.12800.

Aflitos, S.A., E. Schijlen, H. de Jong, D. de Ridder, S. Smit, R. Finkers, J. Wang, et al. 2014. Exploring genetic variation in the tomato (*Solanum* section Lycopersicon) clade by whole-genome sequencing. The Plant Journal 80: 136–148. doi:10.1111/tpj.12616.

Anderson, L.K., P.A. Covey, L.R. Larsen, P. Bedinger, and S.M. Stack. 2010. Structural differences in chromosomes distinguish species in the tomato clade. Cytogenetic and genome research 129: 24–34. doi:10.1159/000313850.

Anithakumari, A.M., O. Dolstra, B. Vosman, R.G.F. Visser, and C.G. Linden. 2011. *In vitro* screening and QTL analysis for drought tolerance in diploid potato. Euphytica 181: 357–369. doi:10.1007/s10681-011-0446-6.

Aversano, R., F. Contaldi, M.R. Ercolano, V. Grosso, M. Iorizzo, F. Tatino, L. Xumerle, et al. 2015. The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. The Plant cell 27: 954–968. doi:10.1105/tpc.114.135954.

Bamberg, J., and A. del Rio. 2013. Selection and Validation of an AFLP Marker Core Collection for the Wild Potato *Solanum microdontum*. American Journal of Potato Research 91: 368–375. doi:10.1007/s12230-013-9357-5.

Barone, A. 2004. Molecular marker-assisted selection for potato breeding. American Journal of Potato Research 81: 111–117. doi:10.1007/BF02853608.

Barone, A., A. Sebastiano, and D. Carputo. 1999. Chromosome pairing in *Solanum commersonii - S. tuberosum* sexual hybrids detected by commersonii-specific RAPDs and cytological analysis. Genome 42: 218–224. doi:10.1139/gen-42-2-218.

Barone, A., A. Sebastiano, D. Carputo, F. della Rocca, and L. Frusciante. 2001. Molecular marker-assisted introgression of the wild *Solanum commersonii* genome into the cultivated *S . tuberosum* gene pool. TAG Theoretical and Applied Genetics 102: 900–907. doi:10.1007/s001220000498.

Benavente, E., M. Cifuentes, J.C. Dusautoir, and J. David. 2008. The use of cytogenetic tools for studies in the crop-to-wild gene transfer scenario. Cytogenetic and genome research 120: 384–95. doi:10.1159/000121087.

Bethke, P.C., D.A. Halterman, and S.H. Jansky. 2017. Are we getting better at using wild potato species in light of new tools? Crop Science 57: 1241–1258. doi:10.2135/cropsci2016.10.0889.

Birch, P.R.J., G. Bryan, B. Fenton, E.M. Gilroy, I. Hein, J.T. Jones, A. Prashar, M.A. Taylor, L. Torrance, and I.K. Toth. 2012. Crops that feed the world 8: Potato: are the trends of increased global production sustainable? Food Security. Vol. 4. doi:10.1007/s12571-012-0220-1.

de Boer, J.M., T.J.A. Borm, T. Jesse, B. Brugmans, L. Wiggers-Perebolte, L. de Leeuw, X. Tang, et al. 2011. A hybrid BAC physical map of potato: a framework for sequencing a heterozygous genome. BMC Genomics 12: 594. doi:10.1186/1471-2164-12-594.

de Boer, J.M., E. Datema, X. Tang, T.J.A. Borm, E.H. Bakker, H.J. van Eck, R.C.H.J. van Ham, H. de Jong, R.G.F. Visser, and CWB Bachem. 2015. Homologues of potato chromosome 5 show variable collinearity in the euchromatin, but dramatic absence of sequence similarity in the pericentromeric heterochromatin. BMC genomics 16. BMC Genomics: 374. doi:10.1186/s12864-015-1578-1.

Bolger, A., F. Scossa, M.E. Bolger, C. Lanz, F. Maumus, T. Tohge, H. Quesneville, et al. 2014. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nature Genetics 46: 1034–1038. doi:10.1038/ng.3046.

Borm, T.J.A. 2008. Construction and use of a physical map of potato. PhD Thesis, Wageningeng UR, The Netherlands.

Bradeen, J.M., and K.G. Haynes. 2011. Introduction to Potato. In Genetics, Genomics and Breeding of Potato, ed. JM Bradeen and C. Kole, 1–19. CRC Press.

Bradeen, J.M., S.K. Naess, J. Song, G.T. Haberlach, S.M. Wielgus, C.R. Buell, J. Jiang, and J.P. Helgeson. 2003. Concomitant reiterative BAC walking and fine genetic mapping enable physical map development for the broad-spectrum late blight resistance region, RB. Molecular genetics and genomics : MGG 269: 603–11. doi:10.1007/s00438-003-0865-8.

Bradshaw, J.E. 2007a. Breeding potato as a major staple crop. In Breeding major food staples, ed. MS Kang and PM Priyadarshan, 277–332. Oxford: Blackwell Publishing.

Bradshaw, J.E. 2007b. Potato-breeding strategy. In Potato Biology and Biotechnology: Advances and Perspectives, ed. D. Vreugdenhil, JE Bradshaw, C. Gebhardt, F. Govers, DKL Mackerron, MA Taylor, and HA Ross, 157–177. Elsevier.

Bradshaw, J.E. 2016. Use of Sexual Reproduction in Base Broadening and Introgression. Plant Breeding: Past, Present and Future: Springer. doi:10.1007/978-3-319-23285-0.

Bradshaw, J.E., and M. Bonierbale. 2010. Potatoes. In Root and Tuber Crops. Handbook of Plant Breeding 7, ed. J. E. Bradshaw, 1–52. New York, NY: Springer New York. doi:10.1007/978-0-387-92765-7.

Bradshaw, J.E., G.J. Bryan, and G. Ramsay. 2006. Genetic Resources (Including Wild and Cultivated *Solanum* Species) and Progress in their Utilisation in Potato Breeding. Potato Research 49: 49–65. doi:10.1007/s11540-006-9002-5.

Bradshaw, J.E., and G. Ramsay. 2009. Potato Origin and Production. In: Advances in Potato Chemistry and Technology. Edited by J Singh and L Kaur. Advances in Potato Chemistry and Technology. First Edit. Elsevier Ltd. doi:10.1016/B978-0-12-374349-7.00001-5.

Bryan, G.J., J. McNicoll, G. Ramsay, R.C. Meyer, and W.S. De Jong. 1999. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. Theor. Appl. Genetics 99: 859–867. doi:10.1007/s001220051306.

Camadro, E.L. 2012. Relevance of the genetic structure of natural populations, and sampling and classification approaches for conservation and use of wild crop relatives: potato as an example. Botany 90: 1065–1072. doi:10.1139/b2012-090.

Camadro, E.L., D. Carputo, and S.J. Peloquin. 2004. Substitutes for genome differentiation in tuber-bearing *Solanum*: interspecific pollen-pistil incompatibility, nuclear-cytoplasmic male sterility, and endosperm. Theor. Appl. Genet. 109: 1369–76. doi:10.1007/s00122-004-1753-2.

Cao, H., A.R. Hastie, D. Cao, E.T. Lam, Y. Sun, H. Huang, X. Liu, et al. 2014. Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. GigaScience 3: 34. doi:10.1186/2047-217X-3-34.

Cardi, T. 2016. Cisgenesis and genome editing: Combining concepts and efforts for a smarter use of genetic resources in crop breeding. Plant Breeding 135: 139–147. doi:10.1111/pbr.12345.

Carputo, D, T Cardi, L Frusciante, and SJ Peloquin. 1995. Male fertility and cytology of triploid hybrids between tetraploid *Solanum commersonii* (2n = 4x = 48, 2EBN) and Phureja - Tuberosum haploid hybrids. Euphytica 83: 123–129.

Carputo, D. 2003. Cytological and breeding behavior of pentaploids derived from 3x x 4x crosses in potato. Theoretical and applied genetics 106: 883–8. doi:10.1007/s00122-002-1165-0.

Carputo, D., D. Alioto, R. Aversano, R. Garramone, V. Miraglia, C. Villano, and L. Frusciante. 2013. Genetic diversity among potato species as revealed by phenotypic resistances and SSR markers. Plant Genetic Resources: 1–9. doi:10.1017/S1479262112000500.

Carputo, D., A. Barone, T. Cardi, A. Sebastiano, L. Frusciante, and S.J. Peloquin. 1997. Endosperm balance number manipulation for direct *in vivo* germplasm introgression to potato from a sexually isolated relative (*Solanum commersonii* Dun.). Proceedings of the National Academy of Sciences of the United States of America 94: 12013–7. doi:VL - 94.

Carputo, D., A. Barone, and L. Frusciante. 2000. 2N Gametes in the Potato: Essential Ingredients for Breeding and Germplasm Transfer. Theor. Appl. Genet. 101: 805–813. doi:10.1007/s001220051547.

Carputo, D., L. Frusciante, L. Monti, M. Parisi, and A. Barone. 2002. Tuber quality and soft rot resistance of hybrids between *Solanum tuberosum* and the incongruent wild relative S. commersonii. American Journal of Potato Research 79: 345–352.

Causse, M., N. Desplat, L. Pascual, M.-C. Le Paslier, C. Sauvage, G. Bauchet, A. Bérard, et al. 2013. Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. BMC Genomics 14: 791. doi:10.1186/1471-2164-14-791.

Chen, Q., D. Lynch, H.Y. Li, Y. Shi, and H.J. Li. 2004. Interspecific crossability and cytogenetic analysis of sexual progenies of Mexican wild diploid 1EBN species *Solanum pinnatisectum* and *S. cardiophyllum*. American Journal of Potato Research 81: 159–169.

Cockerham, G. 1970. Genetical studies on resistance to potato viruses X and Y. Heredity 25: 309–348.

Costanzo, S., I. Simko, B.J. Christ, and K.G. Haynes. 2005. QTL analysis of late blight resistance in a diploid potato family of *Solanum phureja* x *S. stenotomum*. TAG. Theor. Appl. Genet. 111: 609–17. doi:10.1007/s00122-005-2053-1.

Das, S.K., M.D. Austin, M.C. Akana, P. Deshpande, H. Cao, and M. Xiao. 2010. Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. Nucleic Acids Research 38: 1–8. doi:10.1093/nar/gkq673.

Datema, E., L.A. Mueller, R. Buels, J.J. Giovannoni, R.G.F. Visser, W.J. Stiekema, and R.C.G.J. van Ham. 2008. Comparative BAC end sequence analysis of tomato and potato reveals overrepresentation of specific gene families in potato. BMC Plant biology 8: 34. doi:10.1186/1471-2229-8-34.

Dong, F., J.M. McGrath, J.P. Helgeson, and J. Jiang. 2001. The genetic identity of alien chromosomes in potato breeding lines revealed by sequential GISH and FISH analyses using chromosome-specific cytogenetic DNA markers. Genome 44: 729–734. doi:10.1139/gen-44-4-729.

Dong, F., R.G. Novy, J.P. Helgeson, and J. Jiang. 1999. Cytological characterization of potato - *Solanum etuberosum* somatic hybrids and their backcross progenies by genomic *in situ* hybridization. Genome 42: 987–992. doi:10.1139/gen-42-5-987.

Dong, F., J. Song, S.K. Naess, J.P. Helgeson, C. Gebhardt, and J. Jiang. 2000. Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. TAG Theoretical and Applied Genetics 101: 1001–1007. doi:10.1007/s001220051573.

Dong, F., A.L. Tek, A.B.L. Frasca, J.M. McGrath, S.M. Wielgus, J.P. Helgeson, and J. Jiang. 2005. Development and characterization of potato - *Solanum brevidens* chromosomal addition/substitution lines. Cytogenetic and Genome Research 109: 368–372.

Dvorak, J. 1983. Evidence of genetic suppression of heterogenetic chromosome pairing in polyploid species of *Solanum*, sect. Petota. Canadian Journal of Genetics and Cytology 25: 530–539.

Endelman, J.B., and S.H. Jansky. 2016. Genetic mapping with an inbred line-derived F2 population in potato. Theor. Appl. Genet. 129(5):935–943. doi:10.1007/s00122-016-2673-7.

Ewing, E.E., I. Simko, E.A. Omer, and P.J. Davies. 2004. Polygene mapping as a tool to study the physiology of potato tuberization and dormancy. American Journal of Potato Research 81: 281–289. doi:10.1007/BF02871770.

Flis, B., J. Hennig, D. Strzelczyk-Zyta, C. Gebhardt, and W. Marczewski. 2005. The *Ry-fsto* gene from *Solanum stoloniferum* for extreme resistant to Potato virus Y maps to potato chromosome XII and is diagnosed by PCR marker GP122718in PVY resistant potato cultivars. Molecular Breeding 15: 95–101. doi:10.1007/s11032-004-2736-3.

Gaiero, P., J. van de Belt, F. Vilaró, M.E. Schranz, P. Speranza, and H. de Jong. 2016. Collinearity between potato (*Solanum tuberosum* L.) and wild relatives assessed by comparative cytogenetic mapping. Genome 60: 228–240. doi:10.1139/gen-2016-0150.

Gaiero, P., C. Mazzella, F. Vilaró, P. Speranza, and H. de Jong. 2017. Pairing analysis and in situ Hybridisation reveal autopolyploid-like behaviour in *Solanum commersonii* × *S. tuberosum* (potato) interspecific hybrids. Euphytica 213: 137. doi:10.1007/s10681-017-1922-4.

Gavrilenko, T. 2007. Potato cytogenetics. In Potato Biology and Biotechnology: Advances and Perspectives, ed. D. Vreugdenhil, 72:203–216. Elsevier. doi:10.1016/B978-044451018-1/50052-X.

Gavrilenko, T., J. Larkka, E. Pehu, and V.-M. Rokka. 2002. Identification of mitotic chromosomes of tuberous and non-tuberous *Solanum* species (*Solanum tuberosum* and *Solanum brevidens*) by GISH in their interspecific hybrids. Genome 45: 442–449. doi:10.1139/g01-136.

Gavrilenko, T., R. Thieme, U. Heimbach, and T. Thieme. 2003. Somatic hybrids of *Solanum etuberosum* (+) dihaploid *Solanum tuberosum* and their backcrossing progenies: relationships of genome dosage with tuber development. Euphytica 131: 323–332.

Gebhardt, C., A. Ballvora, B. Walkemeier, P. Oberhagemann, and K. Schüler. 2004. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: A case study for potatoes with quantitative variation of resistance to late blight and maturity type. Molecular Breeding 13: 93–102. doi:10.1023/B:MOLB.0000012878.89855.df.

Hanneman, R.E. 1989. The potato germplasm resource. American Potato Journal 66: 655–667.

Hardigan, M.A., E. Crisovan, J.P. Hamiltion, J. Kim, P. Laimbeer, C.P. Leisner, N.C. Manrique-Carpintero, et al. 2016. Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. The Plant Cell 28: TPC2015–00538–RA. doi:10.1105/tpc.15.00538.

Hardigan, M.A., F.P.E. Laimbeer, E. Parker, L. Newton, E. Crisovan, J.P. Hamilton, B. Vaillancourt, et al. 2017. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proceedings of the National Academy of Sciences: 201714380. doi:10.1073/pnas.1714380114.

Hastie, AR, L Dong, A Smith, J Finklestein, ET Lam, N Huo, H Cao, et al. 2013. Rapid genome mapping in nanochannel arrays for highly complete and accurate *de novo* sequence assembly of the complex *Aegilops tauschii* genome. PloS one 8: e55864. doi:10.1371/journal.pone.0055864.

Haverkort, A.J., P.M. Boonekamp, R. Hutten, E. Jacobsen, L.A.P. Lotz, G.J.T. Kessel, R.G.F. Visser, and E.A.G. Vossen. 2008. Societal Costs of Late Blight in Potato and Prospects of Durable Resistance Through Cisgenic Modification. Potato Research 51: 47–57. doi:10.1007/s11540-008-9089-y.

Hawkes, J.G. 1990. The potato: evolution, biodiversity and genetic resources. Belhaven, London, UK.

Hawkes, J.G. 1994. Origins of cultivated potatoes and species relationships. In Potato Genetics, 3–42. CAB International.

Haynes, K.G., H.E.M. Zaki, C.T. Christensen, E. Ogden, L.J. Rowland, M. Kramer, and L. Zotarelli. 2017. High Levels of Heterozygosity Found for 15 SSR Loci in *Solanum chacoense*. American Journal of Potato Research 94: 638–646. doi:10.1007/s12230-017-9602-4.

Horsman, K., T. Gavrilenko, M. Bergervoet, D.J. Huigen, A.T.W. Joe, and E. Jacobsen. 2001. Alteration of the genomic composition of *Solanum nigrum* (+) potato backcross derivatives by somatic hybridization: selection of fusion hybrids by DNA measurements. Plant breeding 120: 201–207.

Hosaka, K. 1995. Successive domestication and evolution of the Andean potatoes as revealed by chloroplast DNA restriction endonuclease analysis. Theoretical and Applied Genetics 90: 356–363. doi:10.1007/BF00221977.

Hosaka, K., and R.E. Hanneman, Jr. 1998. Genetics of self-compatibility in a self-incompatible wild diploid potato species *Solanum chacoense*. 1. Detection of an *S* locus inhibitor (*Sli*) gene. Euphytica 99: 191–197. doi:10.1023/A:1018353613431.

Iovene, M., A. Barone, L. Frusciante, L. Monti, and D. Carputo. 2004. Selection for aneuploid potato hybrids combining a low wild genome content and resistance traits from *Solanum commersonii*. Theoretical and applied genetics 109: 1139–46. doi:10.1007/s00122-004-1741-6.

Iovene, M., S. Savarese, T. Cardi, L. Frusciante, N. Scotti, P.W. Simon, and D. Carputo. 2007. Nuclear and cytoplasmic genome composition of *Solanum bulbocastanum* (+) *S . tuberosum* somatic hybrids. Genome 450: 443–450. doi:10.1139/G07-024.

Iovene, M., S.M. Wielgus, P.W. Simon, C.R. Buell, and J. Jiang. 2008. Chromatin structure and physical mapping of chromosome 6 of potato and comparative analyses with tomato. Genetics 180: 1307–1317. doi:10.1534/genetics.108.093179.

Jacobsen, E., and H.J. Schouten. 2007. Cisgenesis strongly improves introgression breeding and induced translocation breeding of plants. Trends in Biotechnology 25 (5): 219-223.

Jansky, S.H. 2000. Breeding for Disease resistance in potato. In Plant Breeding Reviews: Volume 19, ed. J. Janick, 19:69–152. Oxford, UK.: John Wiley & Sons, Inc. doi:doi: 10.1002/9780470650172.ch4.

Jansky, S.H. 2006. Overcoming hybridization barriers in potato. Plant Breeding 125: 1–12.

Jansky, S.H. 2009. Breeding, Genetics, and Cultivar Development. In Advances in Potato Chemistry and Technology, First Edit, 27–62. Elsevier Ltd. doi:10.1016/B978-0-12-374349-7.00002-7.

Jansky, S.H., A.O. Charkowski, D.S. Douches, G. Gusmini, C. Richael, P.C. Bethke, D.M. Spooner, et al. 2016. Reinventing Potato as a Diploid Inbred Line–Based Crop. Crop Science 56. The Crop Science Society of America, Inc.: 1412. doi:10.2135/cropsci2015.12.0740.

Jansky, S.H., and S.J. Peloquin. 2006. Advantages of Wild Diploid *Solanum* Species Over Cultivated Diploid Relatives in Potato Breeding Programs. Genetic Resources and Crop Evolution 53: 669–674. doi:10.1007/s10722-004-2949-7.

Jarvis, D.E., Y.S. Ho, D.J. Lightfoot, S.M. Schmöckel, B. Li, T.J.A. Borm, H. Ohyanagi, et al. 2017. The genome of *Chenopodium quinoa*. Nature: 1–6. doi:10.1038/nature21370.

Jiang, J., and B.S. Gill. 1994. Nonisotopic *in situ* hybridization and plant genome mapping: the first 10 years. Genome 37: 717–725. doi:10.1139/g94-102.

Jiao, Y., P. Peluso, J. Shi, T. Liang, M.C. Stitzer, B. Wang, M.S. Campbell, et al. 2017. Improved maize reference genome with single-molecule technologies. Nature 546: 524–527. doi:10.1038/nature22971.

Jo, K-R, M Arens, T-Y Kim, MA Jongsma, RGF Visser, E Jacobsen, and JH Vossen. 2011. Mapping of the *S. demissum* late blight resistance gene *R8* to a new locus on chromosome IX. TAG. Theoretical and applied genetics. 123: 1331–40. doi:10.1007/s00122-011-1670-0.

de Jong, J.H., A.M.A. Wolters, J.M. Kok, H. Verhaar, and J. Eden. 1993. Chromosome pairing and potential for intergeneric recombination in some hypotetraploid somatic hybrids of *Lycopersicon esculentum* (+) *Solanum tuberosum*. Genome 36: 1032–1041.

Kaur, P., P.E. Bayer, Z. Milec, J. Vrána, Y. Yuan, R. Appels, D. Edwards, et al. 2017. An advanced reference genome of *Trifolium subterraneum* L. reveals genes related to agronomic performance. Plant Biotechnology Journal 15: 1034–1046. doi:10.1111/pbi.12697.

Kloosterman, B., J.A. Abelenda, M.D.M.C. Gomez, M. Oortwijn, J.M de Boer, K. Kowitwanich, B.M. Horvath, et al. 2013. Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495: 246–50. doi:10.1038/nature11912.

De Koeyer, D., H. Chen, and V. Gustafson. 2011. Molecular breeding for potato improvement. In Genetics Genomics and Breeding of Potato, ed. J. M. Bradeen and C. Kole, 41–67. Genetics, Genomics and Breeding of Crop Plants. CRC Press.

Kuhl, J.C., R.E. Hanneman, and M.J. Havey. 2001. Characterization and mapping of *Rpi*1 , a late-blight resistance locus from diploid (1EBN) Mexican *Solanum pinnatisectum*. Molecular Genetics and Genomics 265: 977–985. doi:10.1007/s004380100490.

Lam, E.T., A.R. Hastie, C. Lin, D. Ehrlich, S.K. Das, M.D. Austin, P. Deshpande, et al. 2012. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nature biotechnology 30. Nature Publishing Group: 771–6. doi:10.1038/nbt.2303.

Leisner, C.P., J.P. Hamilton, E. Crisovan, N.C. Manrique-Carpintero, A.P. Marand, L. Newton, G.M. Pham, et al. 2018. Genome sequence of M6, a diploid inbred clone of the high glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. The Plant Journal. doi:10.1111/tpj.13857.

Lightbourn, G.J., and R.E. Veilleux. 2007. Production and evaluation of somatic hybrids derived from monoploid potato. American Journal of Potato Research 84: 425–435. doi:10.1007/BF02987188.

Lindhout, Pim, Dennis Meijer, Theo Schotte, Ronald C. B. Hutten, Richard G. F. Visser, and Herman J. van Eck. 2011. Towards F1 Hybrid Seed Potato Breeding. Potato Research 54: 301–312. doi:10.1007/s11540-011-9196-z.

Lokossou, A.A., H. Rietman, M. Wang, P. Krenek, H. van der Schoot, B. Henken, R. Hoekstra, et al. 2010. Diversity, distribution, and evolution of *Solanum bulbocastanum* late blight resistance genes. Molecular plant-microbe interactions : MPMI 23: 1206–16. doi:10.1094/MPMI-23-9-1206.

Lou, Q., M. Iovene, D.M. Spooner, C.R. Buell, and J. Jiang. 2010. Evolution of chromosome 6 of *Solanum* species revealed by comparative fluorescence *in situ* hybridization mapping. Chromosoma 119: 435–442. doi:10.1007/s00412-010-0269-6.

Love, S.L. 1999. Founding clones, major contributing ancestors, and exotic progenitors of prominent North American potato cultivars. American Journal of Potato Research 76: 263–272. doi:10.1007/BF02853624.

Manrique-Carpintero, N.C., J.G. Tokuhisa, I. Ginzberg, and R.E. Veilleux. 2014. Allelic variation in genes contributing to glycoalkaloid biosynthesis in a diploid interspecific population of potato. TAG. Theoretical and applied genetics. 127: 391–405. doi:10.1007/s00122-013-2226-2.

Masuelli, R.W., and E.L. Camadro. 1992. Cytological Analysis and Fertility of *Solanum commersonii* Dun. x *Solanum gourlayi* Haw. Triploid hybrids. Cytologia 57: 161–166.

Masuelli, R.W., and E.Y. Tanimoto. 1995. Irregular meiosis in a somatic hybrid between *S. bulbocastanum* and *S. tuberosum* detected by species-specific PCR markers and cytological analysis. Theoretical and Applied Genetics 91: 401–408.

Matsubayashi, M. 1991. Phylogenetic relationships in the potato and its related species. In Chromosome Engineering in Plants: Genetics, Breeding, Evolution, Part B., ed. T. Tsuchiya and PK Gupta, 93–118. Elsevier.

Mendoza, H.A., and F.L. Haynes. 1974. Genetic relationship among potato cultivars grown in the United States. Hortsci. 9: 328-330.

Mohanty, I.C., D. Mahapatra, S. Mohanty, and A.B. Das. 2004. Karyotype analyses and studies on the nuclear DNA content in 30 genotypes of potato (*Solanum tuberosum*) L. Cell biology international 28: 625–33. doi:10.1016/j.cellbi.2004.05.004.

Mok, D.W.S., H.K. Lee, and S.J. Peloquin. 1974. Identification of potato chromosomes with Giemsa. American Potato Journal 51. Springer-Verlag: 337–341. doi:10.1007/BF02851508.

Mok, D.W.S., and S.J. Peloquin. 1975. Three mechanisms of 2n pollen formation in diploid potatoes. Canadian Journal of Genetics and Cytology 17: 217–225.

Naess, S.K., J.M. Bradeen, S.M. Wielgus, G.T. Haberlach, J.M. McGrath, and J.P. Helgeson. 2000. Resistance to late blight in *Solanum bulbocastanum* is mapped to chromosome 8. Theoretical and Applied Genetics 101 (5–6): 697–704.

Naess, S.K., J.M. Bradeen, S.M. Wielgus, G.T. Haberlach, J.M. McGrath, and J.P. Helgeson. 2001. Analysis of the introgression of *Solanum bulbocastanum* DNA into potato breeding lines. Molecular Genetics and Genomics 265: 694–704. doi:10.1007/s004380100465.

Narancio, R., P. Zorri.lla, C. Robello, M. Gonzalez, F. Vilaró, C. Pritsch, and M. Dalla Rizza. 2013. Insights on gene expression response of a characterized resistant genotype of *Solanum commersonii* Dun. against *Ralstonia solanacearum*. European Journal of Plant Pathology. doi:10.1007/s10658-013-0210-y.

Ono, S., and K. Hosaka. 2010. Efficient Chromosome Number Estimation Using Flow Cytometry in the Backcross of *Solanum demissum* (2n = 6x = 72) to *S. tuberosum* (2n = 4x = 48). American Journal of Potato Research 87: 553–556. doi:10.1007/s12230-010-9159-y.

Ortiz, R. 1998. Potato breeding via ploidy manipulations. Plant breeding Reviews (16): 15-86. https://doi.org/10.1002/9780470650110.ch2

Ortiz, R., P. Simon, S.H. Jansky, and D. Stelly. 2009. Ploidy manipulation of the gametophyte, endosperm and sporophyte in nature and for crop improvement: a tribute to Professor Stanley J. Peloquin (1921-2008). Annals of botany 104: 795–807. doi:10.1093/aob/mcp207.

van Os, H., S. Andrzejewski, E. Bakker, I. Barrena, G.J. Bryan, B. Caromel, B. Ghareeb, et al. 2006. Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics 173: 1075–87. doi:10.1534/genetics.106.055871.

Park, T.H., V.G.A.A. Vleeshouwers, E. Jacobsen, E. Van Der Vossen, and R.G.F. Visser. 2009. Molecular breeding for resistance to *Phytophthora infestans* (Mont.) de Bary in potato (*Solanum tuberosum* L.): A perspective of cisgenesis. Plant Breeding 128. Wiley/Blackwell (10.1111): 109–117. doi:10.1111/j.1439-0523.2008.01619.x.

Pavek, J.J., and D.L. Corsini. 2001. Utilization of potato genetic resources in variety development. American Journal of Potato Research 78: 433–441. doi:10.1007/BF02896375.

Pel, M.A., S.J. Foster, T.-H. Park, H. Rietman, G. van Arkel, J.D.G. Jones, H.J. Van Eck, E. Jacobsen, R.G.F. Visser, and E.A.G. van der Vossen. 2009. Mapping and cloning of late blight resistance genes from *Solanum venturii* using an interspecific candidate gene approach. Molecular plant-microbe interactions : MPMI 22: 601–15. doi:10.1094/MPMI-22-5-0601.

Peloquin, S.J., S.H. Jansky, and G.L. Yerk. 1989. Potato cytogenetics and germplasm utilization. American Potato Journal 66: 629–638.

Pérez-de-Castro, A.M., S. Vilanova, J. Cañizares, L. Pascual, J.M. Blanca, M.J. Díez, J. Prohens, and B. Picó. 2012. Application of genomic tools in plant breeding. Current genomics 13: 179–195.

Peters, S.A., J.W. Bargsten, D. Szinay, J. van de Belt, R.G.F. Visser, Y. Bai, and H. de Jong. 2012. Structural homology in the Solanaceae: analysis of genomic regions in support of synteny studies in tomato, potato and pepper. The Plant journal 71: 602–14. doi:10.1111/j.1365-313X.2012.05012.x.

Phumichai, C., M. Mori, A. Kobayashi, O. Kamijima, and K. Hosaka. 2005. Toward the development of highly homozygous diploid potato lines using the self-compatibility controlling Sli gene. Genome 48: 977–984. doi:10.1139/g05-066.

Pijnacker, L.P., and M.A. Ferwerda. 1984. Giemsa C-banding of potato chromosomes. Canadian Journal of Genetics and Cytology 26: 415–419. doi:10.1139/g84-067.

Potato Genome Sequencing Consortium. 2011. Genome sequence and analysis of the tuber crop potato. Nature 475: 189–95. doi:10.1038/nature10158.

Provan, J., W. Powell, H. Dewar, G. Bryan, G.C. Machray, and R. Waugh. 1999. An extreme cytoplasmic bottleneck in the modern European cultivated potato (*Solanum tuberosum*) is not re-

flected in decreased levels of nuclear diversity. Proceedings. Biological Sciences / The Royal Society 266: 633–639. doi:10.1098/rspb.1999.0683.

Ramsay, G., and G. Bryan. 2011. *Solanum*. In Wild Crop Relatives: Genomic and Breeding Resources, Vegetables, ed. Chittaranjan Kole, 259–271. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-20450-0.

Ríos, D., M. Ghislain, F. Rodríguez, and D.M. Spooner. 2007. What Is the Origin of the European Potato? Evidence from Canary Island Landraces. Crop Science 47: 1271. doi:10.2135/cropsci2006.05.0336.

Ross H. 1986. Potato breeding–problems and perspectives. Advances in plant breeding. Journal of Plant Breeding Supplement: 75–82.

Sandro, P.A., M.I. Rebollo, P. Gaiero, M. Vaio, and P. Speranza. 2016. Diseño de microsatélites para Solanum commersonii a partir de información genómica. In Journal of basic and applied genetics Suppl vol XXVII (1), 276. Buenos Aires: Argentine Society of Genetics.

Sato, Masatoshi, Kazuko Nishikawa, and Kazuyoshi Hosaka. 2006. Potato virus Y resistance gene, Rychc, mapped to potato chromosome 9. Euphytica 149: 367–372. doi:10.1007/s10681-006-9090-y.

Sharma, S.K., D. Bolser, J. de Boer, M. Sønderkær, W. Amoros, M.F. Carboni, J.M. D'Ambrosio, et al. 2013. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3: Genes, Genomes, Genet. 3: 2031–47. doi:10.1534/g3.113.007153.

Shearer, L.A., L.K. Anderson, H. de Jong, S. Smit, J.L. Goicoechea, B.A. Roe, A. Hua, J.J. Giovannoni, and S.M. Stack. 2014. Fluorescence *in situ* hybridization and optical mapping to correct scaffold arrangement in the tomato genome. G3: Genes, Genomes, Genet. 4: 1395–1405. doi:10.1534/g3.114.011197.

Simko, I. 2004. One potato, two potato: Haplotype association mapping in autotetraploids. Trends in Plant Science 9. Elsevier Current Trends: 441–448. doi:10.1016/j.tplants.2004.07.003.

Simko, I., S. Costanzo, K. G. Haynes, B. J. Christ, and R. W. Jones. 2004. Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theoretical and Applied Genetics 108: 217–224. doi:10.1007/s00122-003-1431-9.

Simko, I., S. Costanzo, V. Ramanjulu, B.J. Christ, and K.G. Haynes. 2006. Mapping polygenes for tuber resistance to late blight in a diploid *Solanum phureja* x *S. stenotomum* hybrid population. Plant Breeding 125: 385–389. doi:10.1111/j.1439-0523.2006.01232.x.

Simko, I., S. Jansky, S. Stephenson, and D.M. Spooner. 2009. Genetics of resistance to pests and disease. In Potato Biology and Biotechnology: Advances and Perspectives, ed. D. Vreugdenhil, 117–155. doi:10.1016/B978-044451018-1/50049-X.

Sokolova, E., A. Pankin, M. Beketova, M. Kuznetsova, S. Spiglazova, E. Rogozina, Isol'da Yashina, and Emil Khavkin. 2011. SCAR markers of the R-genes and germplasm of wild *Solanum* species for breeding late blight-resistant potato cultivars. Plant Genetic Resources 9: 309–312. doi:10.1017/S1479262111000347.

Solomon-Blackburn, RM, and H Barker. 2001. A review of host major-gene resistance to potato viruses X, Y, A and V in potato: genes, genetics and mapped locations. Heredity 86. Wiley/Blackwell (10.1111): 8–16. doi:10.1046/j.1365-2540.2001.00798.x.

Song, J., J.M. Bradeen, S.K. Naess, J.A. Raasch, S.M. Wielgus, G.T. Haberlach, J. Liu, et al. 2003. Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late

blight. Proceedings of the National Academy of Sciences of the United States of America 100: 9128–33. doi:10.1073/pnas.1533501100.

Song, J., F. Dong, and J. Jiang. 2000. Construction of a bacterial artificial chromosome (BAC) library for potato molecular cytogenetics research. Genome 43: 199–204.

Song, Y-S, L. Hepting, G. Schweizer, L. Hartl, G. Wenzel, and A. Schwarzfischer. 2005. Mapping of extreme resistance to PVY (*Ry sto*) on chromosome XII using anther-culture-derived primary dihaploid potato lines. Theoretical and Applied Genetics 111: 879–887. doi:10.1007/s00122-005-0010-7.

Spooner, D.M., M. Ghislain, R. Simon, S.H. Jansky, and T. Gavrilenko. 2014. Systematics, Diversity, Genetics, and Evolution of Wild and Cultivated Potatoes. Botanical Review 80: 283–383. doi:10.1007/s12229-014-9146-y.

Spooner, D.M., and R.J. Hijmans. 2001. Potato systematics and germplasm collecting, 1989–2000. American Journal of Potato Research 78: 237–268. doi:10.1007/BF02875691.

Spooner, D.M., K. McLean, G. Ramsay, R. Waugh, and G.J. Bryan. 2005. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. Proceedings of the National Academy of Sciences of the United States of America 102: 14694–9. doi:10.1073/pnas.0507400102.

Staňková, H., A.R. Hastie, S. Chan, J. Vrána, Z. Tulpová, M. Kubaláková, P. Visendi, et al. 2016. BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. Plant Biotechnology Journal 14: 1523–1531. doi:10.1111/pbi.12513.

Sybenga, J. 1996. Chromosome pairing affinity and quadrivalent formation in polyploids: do segmental allopolyploids exist? Genome 39: 1176–1184. doi:10.1139/g96-148.

Szinay, D. 2010. The development of FISH tools for genetic, phylogenetic and breeding studies in tomato (*Solanum lycopersicum*). PhD thesis, Wageningen UR, The Netherlands. ISBN: 9789085856351.

Szinay, D., Y. Bai, R. Visser, and H. de Jong. 2010. FISH applications for genomics and plant breeding strategies in tomato and other solanaceous crops. Cytogenetic and Genome Research 129: 199–210. doi:10.1159/000313502.

Szinay, D., S.-B. Chang, L. Khrustaleva, S.A. Peters, E. Schijlen, Y. Bai, W.J. Stiekema, R.C.H.J. van Ham, H. de Jong, and R.M. Klein Lankhorst. 2008. High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. The Plant journal 56: 627–637. doi:10.1111/j.1365-313X.2008.03626.x.

Szinay, D., E. Wijnker, R. van den Berg, R.G.F. Visser, H. de Jong, and Y. Bai. 2012. Chromosome evolution in *Solanum* traced by cross-species BAC-FISH. The New phytologist 195: 688–698. doi:10.1111/j.1469-8137.2012.04195.x.

Tan, M.Y.A. 2008. Genetic mapping and pyramiding of resistance genes in potato. PhD thesis, Wageningen University & Research. ISBN 9789085049777.

Tang, X., J.M. De Boer, H.J. Van Eck, C.W.B. Bachem, R.G.F. Visser, and H. De Jong. 2009. Assignment of genetic linkage maps to diploid *Solanum tuberosum* pachytene chromosomes by BAC-FISH technology. Chromosome Research 17: 899–915. doi:10.1007/s10577-009-9077-3.

Tang, X., E. Datema, M.O. Guzman, J.M. de Boer, H.J. van Eck, C.W.B. Bachem, R.G.F. Visser, and H. de Jong. 2014. Chromosomal organizations of major repeat families on potato (*Solanum tuberosum*) and further exploring in its sequenced genome. Molecular Genetics and Genomics 289: 1307–1319. doi:10.1007/s00438-014-0891-8.

Tang, X., D. Szinay, C. Lang, M.S. Ramanna, E.A.G. van der Vossen, Er. Datema, R.K. Lankhorst, et al. 2008. Cross-species bacterial artificial chromosome-fluorescence *in situ* hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics 180: 1319–1328. doi:10.1534/genetics.108.093211.

The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485: 635–641. doi:10.1038/nature11119.

Tiwari, J.K., S. Siddappa, B.P. Singh, S.K. Kaushik, S.K. Chakrabarti, V. Bhardwaj, and P. Chandel. 2013. Molecular markers for late blight resistance breeding of potato: an update. Edited by P. Wehling. Plant Breeding 245: n/a–n/a. doi:10.1111/pbr.12053.

Traini, A., M. Iorizzo, H.S. Mann, J.M. Bradeen, D. Carputo, L. Frusciante, and M.L. Chiusano. 2013. Genome Microscale Heterogeneity among Wild Potatoes Revealed by Diversity Arrays Technology Marker Sequences. International journal of genomics 2013: 257218. doi:10.1155/2013/257218.

Ugent, D. 1970. The Potato. Science 170. American Association for the Advancement of Science: 1161–1166. doi:10.2307/1730629.

Verlaan, M.G., D. Szinay, S.F. Hutton, H. de Jong, R. Kormelink, R.G.F. Visser, J.W. Scott, and Y. Bai. 2011. Chromosomal rearrangements between tomato and *Solanum chilense* hamper mapping and breeding of the TYLCV resistance gene *Ty*-1. The Plant journal 68: 1093–1103. doi:10.1111/j.1365-313X.2011.04762.x.

Verzaux, E., D. Budding, N. Vetten, R.E. Niks, V.G.A.A. Vleeshouwers, E.A.G. Vossen, E. Jacobsen, and R.G.F. Visser. 2011. High Resolution Mapping of a Novel Late Blight Resistance Gene *Rpi-avl*1, from the Wild Bolivian Species *Solanum avilesii*. American Journal of Potato Research 88: 511–519. doi:10.1007/s12230-011-9218-z.

Villamon, F.G., D.M. Spooner, M. Orrillo, E. Mihovilovich, W. Pérez, and M. Bonierbale. 2005. Late blight resistance linkages in a novel cross of the wild potato species *Solanum paucissectum* (series Piurana). TAG. Theoretical and applied genetics. 111: 1201–14. doi:10.1007/s00122-005-0053-9.

Visser, R.G.F., C.W.B. Bachem, J.M. de Boer, G.J. Bryan, S.K. Chakrabati, S. Feingold, R. Gromadka, et al. 2009. Sequencing the Potato Genome: Outline and First Results to Come from the Elucidation of the Sequence of the World's Third Most Important Food Crop. American Journal of Potato Research 86: 417–429. doi:10.1007/s12230-009-9097-8.

Visser, R.G.F., C.W.B. Bachem, T. Borm, J. de Boer, H.J. van Eck, R. Finkers, G. van der Linden, et al. 2015. Possibilities and Challenges of the Potato Genome Sequence. Potato Research 57: 327–330. doi:10.1007/s11540-015-9282-8.

Visser, R.G.F., and R. Hoekstra. 1988. *In situ* hybridization to somatic metaphase chromosomes of potato. Theoretical and Applied Genetics (TAG) 76: 420–424.

Vleeshouwers, V.G.A.A., R. Finkers, D. Budding, M. Visser, M.M.J. Jacobs, R. van Berloo, M. Pel, et al. 2011. SolRgene: an online database to explore disease resistance genes in tuber-bearing *Solanum* species. BMC Plant Biology 11: 116. doi:10.1186/1471-2229-11-116.

van der Voort, J.R., K. Kanyuka, E. van der Vossen, A. Bendahmane, P. Mooijman, R. Klein-Lankhorst, W.J. Stiekema, D. Baulcombe, and J. Bakker. 1999. Tight Physical Linkage of the Nematode Resistance Gene *Gpa*2 and the Virus Resistance Gene *R*x on a Single Segment Introgressed from the Wild Species *Solanum tuberosum* subsp. *andigena* CPC 1673 into Cultivated Potato. Molecular Plant-Microbe Interactions 12: 197–206. doi:10.1094/MPMI.1999.12.3.197.

Van Der Vossen, E., A. Sikkema, B. Te Lintel Hekkert, J. Gros, P. Stevens, M. Muskens, D. Wouters, An. Pereira, W.J. Stiekema, and S. Allefs. 2003. An ancient R gene from the wild potato spe-

cies *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. Plant Journal 36: 867–882. doi:10.1046/j.1365-313X.2003.01934.x.

Van Weymers, P.S.M., K. Baker, X. Chen, B. Harrower, D.E.L. Cooke, E.M. Gilroy, P.R.J. Birch, et al. 2016. Utilizing "Omic" Technologies to Identify and Prioritize Novel Sources of Resistance to the Oomycete Pathogen *Phytophthora infestans* in Potato Germplasm Collections. Frontiers in plant science 7. Frontiers: 672. doi:10.3389/fpls.2016.00672.

Warschefsky, E., R.V. Penmetsa, D.R. Cook, and E.J.B. von Wettberg. 2014. Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. American journal of botany 101: 1791–800. doi:10.3732/ajb.1400116.

Watanabe, K., M. Orrillo, M. Iwanaga, R. Ortiz, R. Freyre, and S. Perez. 1994. Diploid potato germplasm derived form wild and land race genetic resources. American Potato Journal 71: 599–604.

Witek, K., F. Jupe, A.I. Witek, D. Baker, M.D. Clark, and J.D.G. Jones. 2016. Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. Nature Biotechnology 34. Nature Publishing Group: 1–8. doi:10.1038/nbt.3540.

Wu, F., and S.D. Tanksley. 2010. Chromosomal evolution in the plant family Solanaceae. BMC genomics 11: 182. doi:10.1186/1471-2164-11-182.

Xiao, M., A. Phong, C. Ha, T.-F. Chan, D. Cai, L. Leung, E. Wan, et al. 2007. Rapid DNA mapping by fluorescent single molecule detection. Nucleic acids research 35: e16. doi:10.1093/nar/gkl1044.

Xu, Chenxi, Chen Jiao, Honghe Sun, Xiaofeng Cai, Xiaoli Wang, Chenhui Ge, Yi Zheng, et al. 2017. Draft genome of spinach and transcriptome diversity of 120 *Spinacia* accessions. Nature Communications 8. Nature Publishing Group: 1–10. doi:10.1038/ncomms15275.

Yeh, B.P., and S.J. Peloquin. 1965. Pachytene Chromosomes of the Potato (*Solanum tuberosum*, Group Andigena). American Journal of Botany 52: 1014–1020.

Young, N.D., F. Debellé, G.E.D. Oldroyd, R. Geurts, S.B. Cannon, M.K. Udvardi, V.A. Benedito, et al. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480: 520–4. doi:10.1038/nature10625.

Zhou, S., M.C. Bechner, M. Place, C.P. Churas, L. Pape, S.A. Leong, R. Runnheim, et al. 2007. Validation of rice genome sequence by optical mapping. BMC genomics 8: 278. doi:10.1186/1471-2164-8-278.

Zhou, S., A. Kile, M. Bechner, M. Place, E. Kvikstad, W. Deng, J. Wei, et al. 2004. Single-molecule approach to bacterial genomic comparisons via optical mapping. In Journal of Bacteriology, 186:7773–7782. American Society for Microbiology. doi:10.1128/JB.186.22.7773-7782.2004.

Zhou, S., F. Wei, J. Nguyen, M. Bechner, K. Potamousis, S. Goldstein, L. Pape, et al. 2009. A single molecule scaffold for the maize genome. PLoS genetics 5: e1000711. doi:10.1371/journal.pgen.1000711.

Zhu, S., J.H. Vossen, M. Bergervoet, M. Nijenhuis, L. Kodde, G.J.T. Kessel, V.G.A.A. Vleeshouwers, R.G.F. Visser, and E. Jacobsen. 2015. An updated conventional- and a novel GM potato late blight *R* gene differential set for virulence monitoring of *Phytophthora infestans*. Euphytica 202: 219–234. doi:10.1007/s10681-014-1276-0.

The word, defining, muzzles; the drawn line
Ousts mistier peers and thrives, murderous,
In establishments which imagined lines

Can only haunt. Sturdy as potatoes,
Stones, without conscience, word and line endure,
Given an inch. Not that they're gross (although

Afterthought often would have them alter
To delicacy, to poise) but that they
Shortchange me continuously: whether

More or other, they still dissatisfy.
Unpoemed, unpictured, the potato
Bunches its knobby browns on a vastly
Superior page; the blunt stone also.

Sylvia Plath
("Poems, Potatoes", 1958)

Chapter 3

Pairing analysis and *in situ* Hybridisation reveal autopolyploid-like behaviour in *Solanum commersonii* x *S. tuberosum* (potato) interspecific hybrids

Authors:
Paola Gaiero[1,2], Cristina Mazzella[1], Francisco Vilaró[3], Pablo Speranza[1], Hans de Jong[2]

1. Department of Plant Biology, Facultad de Agronomía, Universidad de la República. Garzón 780, PC 12900, Montevideo, Uruguay.
2. Laboratory of Genetics, Wageningen University. Droevendaalsesteeg 1, P.O. Box 16, 6708 PB, Wageningen, The Netherlands.
3. Horticulture Unit, National Institute for Agricultural Research, Ruta 48 km 10, Las Brujas, Uruguay.

Published in:
Euphytica 213 (7): 645. doi:10.1007/s10681-017-1922-4.

Abstract

Wild potato relatives are rich sources of desirable traits for introgressive hybridization into cultivated potato. One of them, *Solanum commersonii* (2n = 2x = 24, 1EBN, endosperm balance number), is an important species belonging to the potato tertiary genepool. It can be used in potato breeding through bridge crosses and 2n gamete production. Triploid F_1 hybrids between *S. commersonii* (through spontaneous 2n egg formation) and diploid 2EBN *S. tuberosum* Group Phureja were crossed with *S. tuberosum* Group Tuberosum, resulting in successive backcross 1, 2 and 3 progenies. The main aim of this study was to determine if there are any barriers to homoeologous pairing and recombination in the allotriploid (*S. commersonii* x potato) hybrids and their backcrosses, and so to predict if *S. commersonii* chromosomes can be transmitted to the next generation and introgressed into their recipient potato chromosomes. Microscopic observations of spread pollen mother cells suggested no preferential pairing in the triploid hybrids, while chromosome transmission and segregation in further meiotic stages were fairly balanced. Fluorescent *in situ* hybridisation with BAC probes (BAC FISH) was used to obtain markers to trace the meiotic behaviour of specific chromosome pairs. Moreover, Genomic *in situ* Hybridisation (GISH) demonstrated no obvious differences in fluorescence signals between the homoeologues suggesting that repetitive sequences did not diverge much between the parental species. As a consequence, we were not able to trace the course of the *S. commersonii* chromosomes in the successive introgressive hybridisation backcross generations. Our results strongly point at a high genomic similarity between the homoeologous chromosomes promising high suitability of *S. commersonii* in introgressive hybridisation breeding of potato.

Abbreviation list

EBN	Endosperm balance number	Cy3	Cyanine 3-dUTP
BAC	Bacterial artificial chromosome	Cy3.5	Cyanine 3.5-dCTP
FISH	Fluorescent *in situ* hybridisation	Cy5	Cyanine 5-dUTP
GISH	Genomic *in situ* hybridisation	DEAC	Diethylaminocoumarin-5-dUTP
BC	Backcross	DAPI	4',6-diamidino-2-phenylindole
PMC	Pollen Mother Cells	FITC	Fluorescein isothiocyanate

Introduction

Cultivated and wild relatives of potato comprise a huge germplasm resource, more diverse and accessible than that of any other crop (Hawkes 1966; Ross 1986; Hanneman 1989; Peloquin et al. 1989; Hawkes 1990). The collection provides essential genetic traits for many of the biotic and abiotic threats to the crop (Jansky 2000; Solomon-Blackburn and Barker 2001; Jansky and Peloquin 2006; Hamernik et al. 2009), which can be transferred to selected cultivars by introgressive hybridisation. The rich gene pool also contributes to allelic diversity for breeding programmes, and so determines the prospect for selecting desired allele combination and maximizing heterozygosity, required for yield improvements (Mendoza and Haynes 1974).

However, crossing and zygotic barriers in the interspecific hybrids and backcross derivatives, particularly in those belonging to the secondary and tertiary genepool (Bradshaw 2007), may impede the introgression programmes and thus, considerable efforts have been devoted to overcoming such barriers (reviewed by Jansky 2006) by ploidy manipulation (reviewed by Ortiz 1998; Ortiz et al. 2009) and/or bridge crosses (Jansky and Hamernik 2009). The most prominent hybridisation barrier in the potato gene pool is Endosperm Balance Number (EBN). The EBN hypothesis (Johnston et al. 1980) proposes that each *Solanum* species has a specific empirical EBN and that for a cross to be successful a 2:1 maternal to paternal EBN ratio is required in the hybrid endosperm. EBN is not directly linked to ploidy level but ploidy manipulation helps overcome the EBN hybridisation barrier (Johnston and Hanneman 1982). Crossability is then routinely evaluated through pollen stainability, an indirect measurement of fertility in the hybrids and backcrosses.

One important representative of the potato tertiary genepool is *S. commersonii*. This species harbours resistances to various biotic stresses such as *Phytophtora infestans* (Micheletto et al. 2000), *Ralstonia solanacerum* (González et al. 2013) and other severe potato pathogens, such as *Pectobacterium*, *Verticillium*, *Alternaria*, and X and Y viruses (Laferriere et al. 1999; Carputo et al. 2000). Additionally, *S. commersonii* has long received attention because of its frost tolerance and cold acclimation capacity (Palta and Li 1979; Palta and Simon 1993; Vega et al. 2000), and it is therefore an outstanding source for broadening the genetic base of cold and drought adaptation (Chen et al. 1999). A few accessions of this species have been used in potato introgression breeding (Bamberg et al. 1994; Carputo et al. 1997; Laferriere et al. 1999; Chen et al. 1999; Carputo et al. 2000; Carputo et al. 2009). However, interactions at the cytogenetic level have not been described and thus the efficiency of introgression is not completely understood.

A strategy based on the production of unreduced gametes was designed to start a hybridisation programme to introgress resistance to *Ralstonia solanacearum* and broaden the genetic base of the potatoes cultivated in Uruguay (González 2010). This strategy differs from previous schemes such as the one presented by Carputo et al. (1997) in that it makes use of the natural production of unreduced gametes in *S. commersonii* and in that it takes advantage of the variability generated by meiosis, instead of using ploidy manipulations to overcome the EBN barrier (Figure 1). A diploid *Solanum tuberosum* Group Phureja clone was used in the initial cross to overcome incongruity between the species. The F_1 plants were used as female parents for backcross 1 (BC_1) progenies, again through spontaneous unreduced gametes and with a *S. tuberosum* Group Tuberosum clone as male parent. Successive backcrosses with different *S. tuberosum* Group Tuberosum genotypes produced advanced backcross progenies.

The degree of homoeologous pairing and recombination in interspecific hybrids can be assessed directly by analysing pollen mother cells at diakinesis and later meiotic stages. When combined with Genomic *in situ* hybridization (GISH) to discriminate

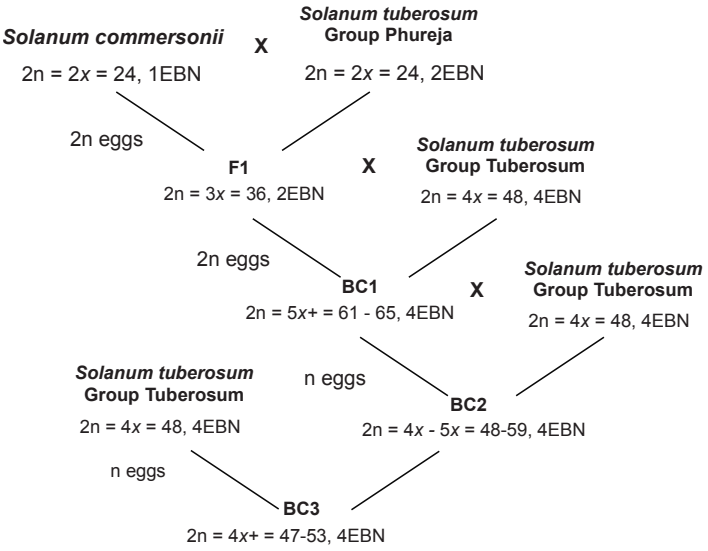


Figure 1. Introgressive hybridisation breeding crossing scheme between *Solanum commersonii* and *S. tuberosum* Group Tuberosum, using *S. tuberosum* Group Phureja as bridge cross species (González 2010). This crossing scheme produced the F_1, BC_1 and BC_2 progenies analysed in this study.

parental genomes, pairing analysis allows tracing alien chromosomes in the recipient background. This approach has been used in interspecific hybrids of various *Solanum* species, showing the parental chromosomes in *S. etuberosum* (+) potato (Dong et al. 1999; Gavrilenko et al. 2003), *S. nigrum* (+) potato (Horsman et al. 2001), *S. bulbocastanum* (+) potato (Iovene et al. 2007) and *S. brevidens* (+) potato hybrids (Dong et al. 2001; Gavrilenko et al. 2002; Dong et al. 2005) and chromosome substitution lines (Tek et al. 2004). Cross-species BAC-FISH painting can be useful to apply chromosome-specific markers from one species on the chromosomes of a relative (Dong et al. 2000). These markers have proved useful to analyse the behaviour of the corresponding chromosomes or chromosome regions during mitosis and meiosis (Schubert et al. 2001; Lysak et al. 2003; McKee 2004). They also constitute a powerful tool for karyotyping in species with small chromosomes or across related species (Dong et al. 2000) and to identify extra chromosomes in aneuploids (Ji 2014) or chromosomes responsible for resistance traits (Tek et al. 2004). When GISH was combined with chromosome-specific molecular markers the transmission and segregation of individual chromosomes could be traced in potato (+) tomato fusion hybrids (Jacobsen et al. 1995; Garriga-Calderé et al. 1997; Garriga-Calderé et al. 1998; Garriga-Calderé et al. 1999) and wide potato hybrids (Ono et al. 2016). This approach was also used to elucidate the genomic constitution of wild allopolyploid species (Pendinen et al. 2008; Pendinen et al. 2012). The limitation of the GISH technology in discriminating parental chromosomes in hybrids is that it depends on how much species-specific tandem and dispersed repeats have diverged in the progenitors and this is still unknown for *S. commersonii*, *S. tuberosum* Group Phureja and *S. tuberosum* Group Tuberosum.

According to Matsubayashi (1991), many diploid species in the genome classifications of section Petota share a common genome called A, based on regular meiosis and good pollen fertility in diploid hybrids. Lack of genome differentiation has been proposed for *Solanum* section Petota (Camadro et al. 2004), although some cryptic variants were described by Matsubayashi (1991). These variants are not visible in the meiotic behaviour of diploid hybrids and result in a more or less reduced pollen fertility. In amphiploids frequent preferential pairing was observed. Definite structural variants were also described, based on various meiotic irregularities in diploid hybrids and on low pollen fertility, around 10 % (Matsubayashi 1991). The factors underlying the observed meiotic behaviour and pollen fertility were not clarified with the approaches available at that moment. Although he proposes that *S. commersonii*, *S. tuberosum* Group Phureja and *S. tuberosum* Group Tuberosum share two A genomes, with two extra A_t genomes in tetraploid potato, it is not clear if there are cryptic or definite differences across these species and to what extent they affect homoeologous pairing and recombination. If such variants reflect quantitative or qualitative changes in the repetitive sequences it may be possible to visualise such genomic differentiation through GISH.

In this study, we focused on whether there was evidence of restrictions to pairing and recombination between the chromosomes of *Solanum commersonii* and *S. tuberosum* (both Group Phureja and Group Tuberosum). We analysed the male meiosis of triploid interspecific hybrids to test if there was preferential pairing between the two *S. commersonii* genomes and exclusion of the *S. tuberosum* Group Phureja chromosomes. We also attempted to follow the fate of specific chromosomes in these triploid hybrids and successive introgressive hybridisation backcrosses both through BAC FISH markers and genome painting (GISH).

Materials and Methods

Plant material and slide preparation

We described homoeologous pairing in the male meiosis of individuals from the introgressive hybridisation breeding programme described in Figure 1, namely two allotriploid interspecific hybrids code named 06.201.6 and 06.201.20 ($2n$ = 3x = 36, 2EBN) to test the scenario of preferential pairing and clones from the BC_1 and BC_2 progenies (Table 1), to look at pairing and follow the fate of alien chromosomes in the successive backcrosses.

For chromosome counts, root tips were pre-treated with 2 mM aqueous 8-hydroxyquinoline for 4 h at 20º C and 20 h at 4° C, and fixed in a 3:1 ethanol–acetic acid solution for 48 h, followed by 70 % ethanol and stored them at 4º C. We performed digestion with an enzyme mix containing 2 % pectinase (from *Aspergillus niger*, Sigma Aldrich, St. Louis, MO, USA, P-4716) and 2 % cellulase RS (Yakult 203033, Yakult Pharmaceutical, Tokyo, Japan) in 10 mM citrate buffer (pH 4.5) for 2 to 4 h at 37° C. Slides were prepared follow-

Table 1. Overview of the plant material used in this study. Pedigree, ploidy, chromosome number, 18-35S and 5S rDNA sites and pollen stainability of the parental, F1 hybrid, backcross 1 (BC1) and BC2 genotypes analysed. Genotypes indicated in bold were selected as parents for the following progenies.

Germplasm type	Genotype	Pedigree	Ploidy and chromosome number	18-35S/5S rDNA sites	Pollen stainability (%)
Wild species	**04.02.3**	cmm	2n = 2x = 24	2/2	89
Cultivated potatoes	94212.2	phu	2n = 2x = 24	2/2	95
	CIP 38228416	78A1-8 x G5264.1	2n = 4x = 48	4/4	82
	8809.2	Cupids x Bulk CIP	2n = 4x = 48	4/4	80-90
	Atlantic	Wauseon x Lenape	2n = 4x = 48	4/4	19
F1 hybrids	**06201.6**	04.02.3 x 94212.2	2n = 3x = 36	3/3	15
	06201.20	04.02.3 x 94212.2	2n= 3x = 36	3/3	18
Backcross 1	**08301.1**	06201.6 x 8416	2n = 5x = 63	4/6	40
	08302.2	06201.20 x 8416	2n = 5x = 61	5/6	36
	08302.4	06201.20 x 8416	2n = 5x = 65	6/6	39
Backcross 2	09.505.1	08302.2 x 8809.2	2n = 4x-5x = 54	5/4	20-30
	09.505.3	08302.2 x 8809.2	2n = 4x-5x = 53	4/5	20-30
	09.505.5	08302.2 x 8809.2	2n = 4x-5x = 55	4/6	20-30
	09.509.1	08302.4 x 8809.2	2n = 4x-5x = 59	4/5	20-30
	09509.2	08302.4 x 8809.2	2n = 4x-5x = 48	5/4	20-30
	09509.6	08302.4 x 8809.2	2n = 4x-5x = 56	4/6	38
	09.510.1	08.302.4 x BW	2n = 4x-5x = 52	nd	20-30

ing the squashing method in 45 % acetic acid and stained using with 5 µg.mL^{-1} DAPI. For meiosis analyses, we harvested young flower buds in the morning (11-12 am) and fixed them in 3:1 ethanol: acetic acid as described above. For the study of early meiotic stages, we prepared spread preparations of pollen mother cell (PMC) complements following the procedure described in Szinay et al. (2008) with minor modifications. For meiotic cells at diakinesis and later stages, we used the squashing method. Briefly, anthers previously selected for these stages were digested for 2-4 h at 37° C in an enzyme mix of 1 % pectolyase Y23 (pectolyase from *Aspergillus japonicus*, Sigma Aldrich, St. Louis, MO, USA, P-3026), 1 % cellulase RS (Yakult 203033, Yakult Pharmaceutical, Tokyo, Japan) and 1 % cytohelicase (cytohelicase from *Helix pomatia*, Sigma Aldrich, St. Louis, MO, USA, C8274) diluted 1:5 in 10mM citrate buffer (pH 4.5). Individual anthers were transferred to 30–40 µL 60 % acetic acid and squeezed carefully with fine needles to release the PMC. 8–10 µL of the cell suspension were dropped onto the clean slide and a 24 x 50 mm coverslip was put on top. We alternated treatment at 55° C for 30-60 sec, with 10-20 sec at about 20º C. This procedure was repeated as many times as necessary for about 10 min, adding 60 %

acetic acid to avoid preparation from drying. The preparation was then firmly squashed and the coverslip was removed after freezing in liquid nitrogen. Pollen stainability was measured by shaking mature flowers on slides and staining the pollen with 1 % aceto-carmine.

Probe and blocking DNA isolation and *in situ* Hybridisation

BAC clones used for FISH were obtained from the RHPOTKEY potato BAC library constructed from the RH clone RH89-039-16 and had previously been selected by Tang et al. (2009) for each of the twelve potato linkage groups. BAC DNA was isolated using the QIAGEN (Valencia, CA) plasmid minikit and amplified using the REPLI-g minikit (QIAGEN). Pools of BACs for each chromosome were either directly labelled with Cyanine 3-dUTP (Cy3, Enzo Life Sciences), Cyanine 3.5-dCTP (Cy3.5, GE Healthcare, Sweden) or Diethylaminocoumarin-5-dUTP (DEAC, Perkin Elmer Inc), or indirectly labelled biotin-16-UTP or digoxigenin-11-dUTP by standard nick translation reaction (Roche Diagnostic, Indianapolis). Probes from the 5S rDNA from the pCT4.2 plasmid (Campell et al. 1992) and 18-35S rDNA from the pTa71 plasmid (Gerlach and Bedbrook 1979) were used as chromo-

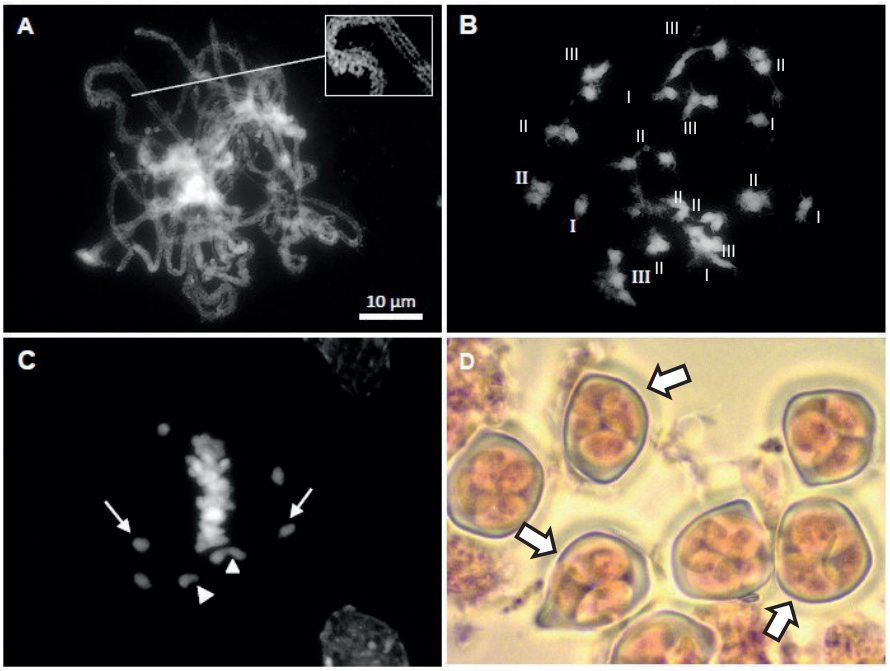


Figure 2. Homoeologous pairing in pollen mother cells (PMC) from *Solanum commersonii* and *S. tuberosum* Group Phureja triploids. A. Pachytene complement with bivalents and trivalents and a loop (see inset). B. Diakinesis showing five trivalents (III), eight bivalents (II) and eight univalents (I). C. Metaphase/early anaphase I complement showing precocious migration (arrows) and univalents and bivalents out of the metaphase plate (arrowhead). D. Tetrads and polyads (arrows), aceto-carmine stained cells, bright field microscopy.

some markers for pairs 1 and 2 respectively, and also as hybridisation controls. Genomic DNA from the *S. commersonii* maternal genotype 04.02.3 (Figure 1) was isolated following the protocol described by Jobes et al. (1995), starting from 5 g of ground leaf material to obtain larger DNA quantities and concentration and increasing the buffer volumes accordingly. We labelled it with digoxigenin-11-dUTP by standard nick translation reaction (Roche Diagnostic, Indianapolis). Additionally, we isolated genomic DNA from *S. tuberosum* cv Desirée also following Jobes et al. (1995) with the minor modifications described and we used it as blocking DNA (100 times probe concentration). FISH and GISH experiments were performed as described previously (Zhong et al. 1996), with the minor modifications introduced by Gaiero et al. (2016), Chapter 4. Hybridisation was carried out over three days to obtain enhanced signals. Hybridisation of the repetitive sequences in the BAC DNA was suppressed by adding unlabelled C_{ot}-100 (50 times probe concentration) which was prepared from *S. commersonii* and *S. tuberosum* genomic DNA as described by Tang et al. (2008). In GISH experiments, stringency was adjusted to 80-85 % by performing stringency washes with 50 % formamide/2 x SSC at 42° C for 15 min. Chromosomes were counterstained with 5 μg.mL^{-1} 4′,6-diamidino-2-phenylindole (DAPI) in Vectashield anti-fade (Vector Laboratories).

Image acquisition and processing

We examined the slides under a Zeiss Axioplan 2 imaging photomicroscope (http://www.zeiss.com) with epifluorescence illumination and filter sets for 4′,6-diamidino-2-phenylindole (DAPI), DEAC (blue), FITC (green), Cy3 (orange), Cy3.5 (red), and Cy5/Alexa Fluor 647 (far-red) fluorescence. Selected images were captured using a Photometrics Sensys 1305 x 1024 pixel CCD camera (Photometrics, http://www.photomet.com). Image thresholding was performed using image analysis software. We performed image adjustments with Adobe® Photoshop® software as follows: DAPI images were displayed in light grey and sharpened using a 7 x 7 pixel Hi-Gauss high-pass spatial filter to accentuate minor details and the heterochromatin morphology of the chromosomes. The remaining fluorescence images were pseudo-coloured and overlaid in multichannel mode. Brightness and contrast adjustments were performed using the Levels tool in Adobe® Photoshop® affecting all pixels equally.

Results

Analysis of homoeologous pairing in 3x hybrids and backcrosses

All F_1 hybrids were triploid ($2n = 3x = 36$), whereas all BC_1 have chromosome numbers varying from $2n = 61$ to 65. Their corresponding BC_2 progenies were 4x-aneuploid, ranging from 48 to 59 chromosomes (Table 1).

Detailed meiotic analysis was performed on pollen mother cells (PMC). Frequent trivalents were observed at pachytene, together with bivalents plus single chromosomes` (Figure 2A). Meiotic configurations were studied at diakinesis, when chromosome spreading was sufficient to distinguish the different pairing configurations (Figure 2B). In the F_1 allotriploids most chromosomes form trivalents, with an average frequency of 7.3 III, 4.9 II and 4.3 I per cell in a total of 61 cells (Figure 3). We observed chain, frying-pan and Y-shaped trivalents (Figure 2B), which are explained by combinations of homologous and homoeologous pairing and crossing over. There were some rare cases of cell complements containing a ring quadrivalent or quinquevalent, probably due to overlaps (data not shown). Later stages demonstrated rare cases of chromosome irregularities (Table 2), including precocious migration and chromosomes out of plate at metaphase I and early anaphase I. We also observed chromosome stickiness (bivalent interconnections) during diakinesis and metaphase I. Anaphase I and II were typical of odd ploidy genotypes with various unbalanced chromosome distributions (Figure 2C). We did not detect lagging chromosomes at anaphase I/II and only few cases of anaphase bridges were observed, that apparently resolved at later stages, as telophase I/II PMC did not contain micronuclei (Table 2). Due to an atypical alignment of meiotic spindles, about half of the meiotic products were polyads, while the other half produced tetrads (Figure 2D and Table 2). This, combined with the typical unbalanced segregation found in odd ploidy genotypes, meant that pollen stainability was on average 17 % (Table 1) so the proportion of fertile pollen was low.

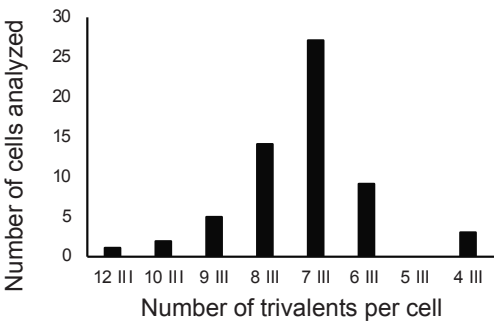


Figure 3. Distribution of the different numbers of trivalents (III) per cell in the male meiosis of the two F_1 triploid hybrids analysed together.

	Total	Normal	Precocious/ Out of plate	Bridges	Laggards	Micronuclei	Polyads
Metaphase I	131	87 (66 %)	44 (34 %)	-	-	-	-
Early Anaphase I	178	53 (30 %)	125 (70 %)	-	-	-	-
Late Anaphase I	15	11 (73 %)	-	4 (27 %)	0	-	-
Telophase I	17	13 (77 %)	-	4 (23 %)	0	-	-
Anaphase II	38	36 (95 %)	2 (5 %)	0	0	-	-
Telophase II	14	13 (93 %)	-	0	0	1 (7 %)	-
Cytokinesis	20	9 (45 %)	-	-	-	-	11 (55 %)
Total	413	222 (54 %)	171 (41 %)	8 (2 %)	0	1 (0.2 %)	11 (2.8 %)

Table 2. Number and frequency of meiotic irregularities found in the later stages of the male meiosis of *Solanum commersonii* x *S. tuberosum* Group Phureja triploid hybrids

Meiotic analysis was also performed on some BC_1 and BC_2 individuals. We observed a high proportion of PMCs at pachytene with configuration involving three or more chromosomes, and similar complex multivalents in cells at diakinesis (Figures 5C and D) and the typical segregation for odd ploidy genotypes in cells at further stages (data not shown). On average, we found 16.2 multivalents, 5.6 II and 2.4 I in PMCs at diakinesis from BC_1 individuals, while in BC_2 we found 12.9 multivalents, 4.3 II and 1 I on average (data not shown). The complexity of homoeologous pairing in these 4x- and 5x-aneuploids did not allow for detailed descriptions of the composition and configurations of the multivalents. Pollen stainability values are higher in these more advanced backcrosses, with an average of 38% stainable pollen. This pollen fertility was enough to use one of the BC_2 genotypes as male parent for the BC_3 progeny.

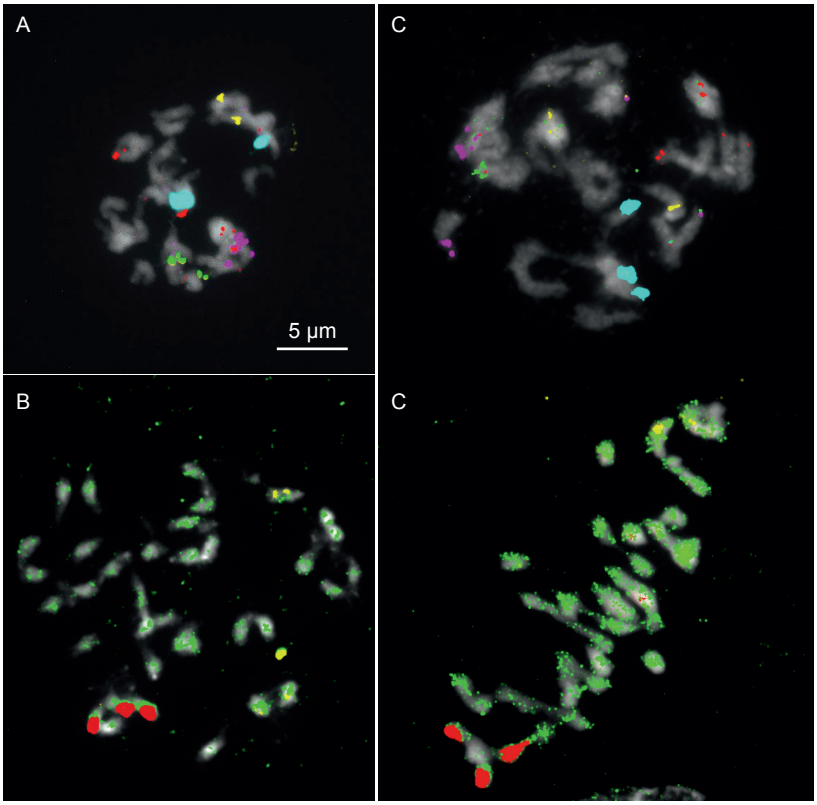


Figure 4. A. Homoeologous pairing in pollen mother cells (PMC) from 3x hybrids between *Solanum commersonii* and *S. tuberosum* Group Phureja hybridised with probes specific to chromosome pairs: 1 (yellow), 2 (blue), 3 (purple), 4 (red) and 6 (green) and B. 1 (yellow), 2 (blue), 3 (purple), 4 (red) and 7 (green). Chromosomes can be observed forming trivalents or bivalents/univalents. C and D. Genome painting on pollen mother cells (PMC) from 3x hybrids between *S. commersonii* and *S. tuberosum* Group Phureja using *S. commersonii* genomic DNA as probe (green), 18-35S rDNA (red) and 5S rDNA (yellow) as hybridisation controls in diakinesis and metaphase I, respectively.

Identification of meiotic chromosomes

We used chromosome-specific cytogenetic markers to follow the behaviour of the corresponding chromosomes during meiosis. We observed that all chromosomes identified with BAC FISH or rDNA signals were involved in trivalents in most of the complements assessed (Figures 4A and B). For example, chromosome pair 1 was forming a trivalent in 7 cell complements out of 15, while chromosome pair 2 was involved in a trivalent in 13 cells and appeared as a univalent plus bivalent in 2 cells (data not shown). We did not find any specific chromosomes that systematically failed to synapse with their homoeologues in the 15 cells quantified i.e., all chromosome pairs were forming trivalents in at least three of the 15 cells assessed, data not shown). Aneuploid individuals in the BC_1 and BC_2 show variable numbers of rDNA bearing chromosomes (Table 1), suggesting that imbalances may involve different chromosome pairs in each individual.

Genomic *in situ* Hybridisation (GISH)

To identify *S. commersonii* chromosomes in F_1 3*x* hybrids and to follow their fate in the successive backcrosses, we performed GISH using *S. commersonii* genomic DNA as probe and highly stringent blocking with unlabelled cultivated potato genomic DNA. All chromosomes show hybridisation with the *S. commersonii* genomic DNA probe, so it is not possible to discriminate those that come from each parental species (Figures 4C and D). The hybridisation signal shows the typical dispersed pattern of highly repetitive genome sequences and a stronger signal in the pericentromere region, coherent with the high stringency used in the experiments. These results did not allow us to visualize recombination events or introgressed chromosome segments.

The alien (*S. commersonii*) chromosomes could not be followed in the successive backcrosses through GISH in mitotic metaphases of BC_1 and BC_2 genotypes (Figure 5A and B). Although it could be argued that the pericentromere signal is stronger on some of the chromosomes, these quantitative differences cannot be interpreted as qualitative (presence/absence) differences and therefore do not allow to draw conclusions. Similar results were obtained when we performed GISH painting on diakinesis complements belonging to BC_1 and BC_2 genotypes (data not shown).

Discussion

The different approaches used here have allowed us to describe and quantify pairing between the chromosomes of *Solanum commersonii* and *S. tuberosum* (both Group Phureja and Group Tuberosum) to find out the extent of homology/homoeology between their genomes. We have not observed any barriers to the exchange of chromosomal segments through the stringent test of homoeologous pairing and recombination in triploid hybrids. We have also described the underlying factors that determine crossing success between these species at the cytogenetic level. We observed a high frequency of trivalents

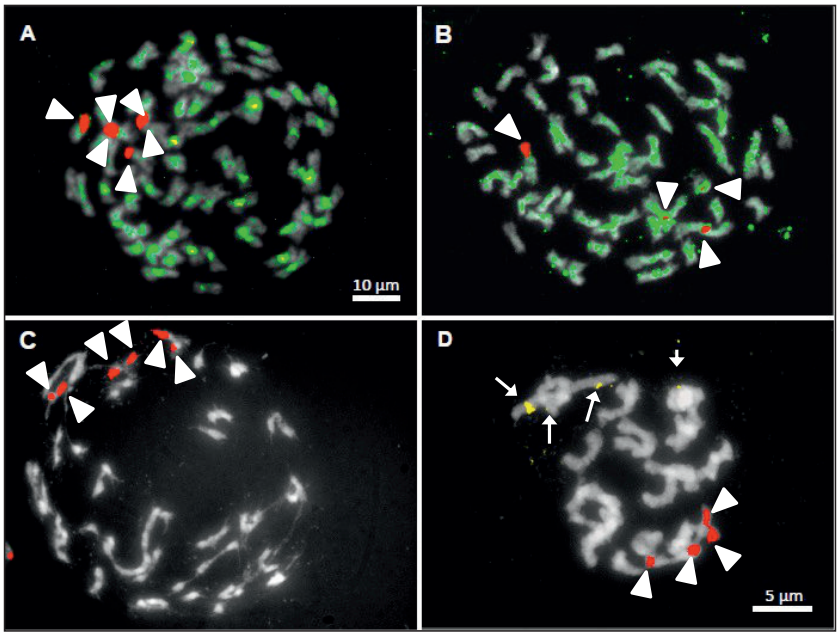


Figure 5. GISH in the backcrosses. A. Genome painting on a mitotic metaphase complement from backcross 1 genotype 08.301.1 (2n = 5x + 3 = 63) and B. on a mitotic metaphase from backcross 2 genotype 09.509.6 (2n = 4x-5x = 56) using *S. commersonii* genomic DNA as probe (green) and 18-35S rDNA (red) as hybridisation control. C. Homoeologous pairing in pollen mother cells (PMC) at diplotene/diakineses from backcross 1 genotype 08.302.4 (2n = 5x + 5 = 65) using *S. commersonii* genomic DNA (green) and 18-35S rDNA (red) as probes and D. Homoeologous pairing in PMC at diakinesis from backcross 2 genotype 09.509.2 (2n = 4x = 48) using 18-35S rDNA (red) and 5S rDNA (yellow) as probes. In all cases the identified chromosomes are forming multivalents. Arrow heads indicate 18-35S rDNA sites and arrows indicate 5S rDNA sites.

per cell, which may allow us to discard preferential pairing between genomes coming from *S. commersonii* and we did not find any evidence of exclusion of *S. tuberosum* Group Phureja or *S. tuberosum* Group Tuberosum chromosomes.

We have confirmed that there is frequent homoeologous pairing between the chromosomes of *S. commersonii* and *S. tuberosum* Group Phureja in the triploid hybrids. At pachytene we observed trivalents formed by different chromosome pairs. There was high degree of synapsis and some pairing breaks along the length of the trivalent, which are randomly distributed (Gaiero et al. 2016, Chapter 4). In some cases, we observed small loops (inset Figure 2A), which might represent small local rearrangements. However, a previous study has shown that at the large-scale there is high collinearity between these two species (Gaiero et al. 2016, Chapter 4). Although we have not analysed the female meiosis in these hybrids, its recombination rate is generally higher than in the male meiosis in hermaphroditic plants (reviewed by Lenormand and Dutheil 2005and by Wijnker and de Jong 2008). The recombination ratio between male and female meiosis ranged

between 0.84 in hybrids between tomato and a wild relative (de Vicente and Tanksley 1991), and 0.73 between potato and a wild relative (Kreike and Stiekema 1997) and 0.72 in a close wild relative, *S. chacoense,* (Rivard et al. 1996). Therefore it can be inferred that the ratio might be similar in the hybrids studied here, and thus that homoeologous pairing could be as frequent or even more frequent in the female meiosis.

At diakinesis, the pairing that was established in earlier stages was maintained and there was high frequency of trivalents (average 7.3 III, 4.9 II and 4.3 I per cell), with a range of 4III to 12III per cell (Figure 3). We could observe chiasmata, which confirm previous homoeologous pairing. Although trivalent formation was already reported for 4x *S. commersonii* x 2x *S. tuberosum* Group Tuberosum- Group Phureja allotriploids, the authors could not describe or quantify it (Barone et al. 1999). Homoeologous pairing frequencies found here are higher than those found for different allotriploid hybrids between species from different series within section Petota (Lange and Wagenvoort 1973; Masuelli and Camadro 1992). Moreover, our results were the same as those found for autotriploid *S. tuberosum,* with an average of 7.06 III (range 2-12 III) across different autotriploid plants (Lange and Wagenvoort 1973). We frequently found trivalents in Y-shaped or V-shaped configurations, with some frying pan and chain configurations as well. These results suggest recombination among homoeologues in earlier stages. In the case of Y-shaped and frying pan configurations, they point at low interference. These configurations require two crossing over events in the same arm (Singh 2003). This implies that the introgressed segments are smaller and therefore with lower linkage drag. According to our BAC FISH results, no chromosomes failed to pair systematically (i.e. no chromosomes appeared as a univalent plus a bivalent in all the cells assessed). These results coincide with the high structural collinearity found between *S. commersonii* and *S. tuberosum* by Gaiero et al. (2016), Chapter 4. Homoeologous pairing was also observed in PMC from the BC_1 and BC_2 progenies, showing mostly multivalents (16 on average in the BC1 and 13 in the BC2) with relatively few bivalents and very few univalents. It appears to be higher than that observed by Barone et al. (1999) and Carputo (2003). Put together, our chromosome pairing evidence indicates that the F_1 triploid hybrids between *S. commersonii* and *S. tuberosum* Group Phureja behave as autotriploids and that homoeologous recombination can take place among all their chromosomes. We have developed a set of *S. commersonii*-specific SSR markers (Sandro et al. 2016) and we are currently testing them in the advanced backcross progenies to follow the fate of *S. commersonii*-specific chromosomal segments and to find conclusive evidence for homoelogous recombination.

Chromosome segregation in further meiotic stages was fairly balanced, with fewer irregularities than those observed in previous studies on triploid hybrids between potato and wild relatives (Adiwilaga and Brown 1991; Masuelli and Camadro 1992; Carputo et al. 1995). These authors found high frequency of cells with laggards in anaphase I and II. In our case these events were absent, albeit the lower number of cells evaluated (Table 2). The most common irregularities found here were precocious migration/bivalents

out of plate, which is a common characteristic of anaphases in *Solanum* (Ramanna and Hermsen 1979; de Jong et al. 1993; Larrosa et al. 2012). These few irregularities were resolved further on. At telophase I and II most PMC did not contain micronuclei (Figure 2 and Table 2), while they were present in the 4x *S. commersonii* x 2x *S. tuberosum* Group Tuberosum- Group Phureja analysed by Carputo et al. (1995) and were very common in the *S. commersonii* x *S. gourlayi* allotriploids (Masuelli and Camadro 1992). According to our results, meiotic products were approximately 50 % polyads and 50 % tetrads (Figure 2D and Table 2) which could carry unbalanced chromosome numbers. These results are discouraging compared to those obtained by Carputo et al. (1995) and combined with unbalanced segregation, meant that pollen stainability was on average low (17 %). In the past, the only indication of crossability was pollen stainability. In such case, a highly suitable wild donor as *S. commersonii*, would have been discarded without looking at chromosome pairing. However, when pairing takes place, only a few successful crosses are enough to introduce the wild germplasm into the cultivated potato background. In our context this difficulty was overcome because these triploids were successfully used as female parents for the BC_1 progenies, through 2n gametes. These BC_1 and BC_2 progenies produce higher percentages of stainable pollen (Table 1), as expected from their more balanced genomic contributions.

Great variation in chromosome numbers was found in the BC_1 (2*n* = 5x + = 61-65) and BC_2 (2*n* = 4x- 5x = 48-59) progenies (Table 1). In our case BC_1 plants were aneuploid, while Barone et al. (1999) found that most were exact allopentaploids. The cases of 5x-aneuploids were explained in terms of occasional omission of chromosomes in the restitution nuclei forming 2n eggs (Carputo 2003). The most important post zygotic barrier in potatoes is EBN. It is not clear yet what factors determine EBN or where they are mapped in the potato genome. Aneuploid gametes with extra chromosomes bearing the EBN controlling factors might be positively selected for compatible crosses and thus favor the formation of genotypes with complements higher than the expected 2n = 5x = 60 (Carputo 1999; Henry et al. 2009). For the BC_2 plants Carputo et al. (2003) observed a narrower range of chromosome number variation, with a tendency towards tetraploid (2*n* = 4x = 48) numbers. In hybridisation programmes with *S. tuberosum* Group Tuberosum involving other species such as *S. demissum*, in pentaploid BC_1 progenies variations in chromosome number were directly related to the amount of wild germplasm in the BC_1 plant because the genome from *S. tuberosum* was equally inherited to the progeny (Ono et al. 2016). However, the ranges found here point at random segregation of both homologous and homoeologous chromosomes in the backcross progenies.

Genomic *in situ* hybridization (GISH) demonstrated hardly any differences in fluorescence signals between the homoeologues, so we were not able to discriminate chromosomes belonging to each species. GISH has been useful to discriminate potato chromosomes from those belonging to species from other *Solanum* sections like tomato (Jacobsen et al. 1995; Garriga-Calderé et al. 1997; Garriga-Calderé and Huigen 1998; Garriga-

Calderé et al. 1999) or *S. nigrum* (Horsman et al. 2001). Within section Petota, it has been successfully used for hybrids between *S. tuberosum* Group Tuberosum and non-tuber bearing potato relatives carrying the E genome (Matsubayashi 1991), like *S. brevidens* (Dong et al. 2001; Gavrilenko et al. 2002; Tek et al. 2004; Dong et al. 2005) or *S. etuberosum* (Dong et al. 1999; Gavrilenko et al. 2003). No successful GISH results have been reported for hybrids between potato and its closer A-genome tuber-bearing wild relatives, with the exception of *S. bulbocastanum*, a diploid (1EBN, A^b genome) Mexican species (Iovene et al. 2007). This suggests that repetitive sequences have not diverged much among their genomes, in spite of the estimated 2.3 million years divergence between the species (Aversano et al. 2015).

The low degree of genomic divergence among potato wild relatives contrasts with the results reported for the species of another well studied clade within the same genus: tomato wild relatives. Variable homoeologous recombination rates have been found through molecular markers and sequence data in introgression lines of *S. lycopersicoides* (Canady et al. 2006) or *S. pimpinellifolium* (Demirci et al. 2016) with tomato. Significant structural rearrangements have been described among species belonging to the tomato clade (Anderson et al. 2010; Verlaan et al. 2011; Szinay et al. 2012). GISH has been successfully applied to tomato hybrids with *S. peruvianum* or *S. lycopersicoides* (Parokonny et al. 1997; Ji and Chetelat 2003; Ji et al. 2004). Results of GISH experiments like those reported here show that the repetitive sequence fraction of the genome has low divergence among potato wild relatives. This compositional difference of the repetitive fractions in the genomes in the potato and tomato clades may itself (through repetitive DNA dynamics) underlie the mechanisms leading to different rates of molecular, pairing and structural chromosome differentiation among potato, tomato and their respective wild relatives. Analyses on the nature, abundance and dynamics of the repetitive fractions of their genomes are currently underway.

The results found in this study are evidence to reinforce the hypothesis of lack of genome differentiation within the potato clade, most likely maintained by (or being a by-product of) pre and post zygotic hybridisation barriers (Carputo et al. 1999; Camadro et al. 2004). Genome similarity between *S. commersonii* and potato must be 85 % or higher, because of the stringency used in GISH experiments and the high blocking conditions. To what extent this is related to homoeologous pairing and recombination is still unclear. A lot remains unknown about the mechanisms that guide homologous pairing (Bozza and Pawlowski 2008), let alone interactions among homoeologues. Although some authors give homology only indirect responsibility for initial pairing (Sybenga 1999), meiotic chromosome pairing is generally accepted to be based on DNA homology. However, it also depends on spatial location (Bozza and Pawlowski 2008). Initial recognition is related to simultaneous conformational changes in chromatin in both partners that can only be triggered by interactions between true homologues, so that they can only pair if they are in the same conformational state (Prieto et al. 2004). It also needs overcoming

of compact heterochromatin regions that might be difficult to access and of the spurious pairing that might be caused by repetitive sequences (Bozza and Pawlowski 2008). Similar heterochromatin distribution and condensation together with similar chromosome structures were found between the genomes of *S. commersonii* and *S. tuberosum* Group Phureja (Gaiero et al. 2016, Chapter 4). Therefore, the genome similarity found here between *S. commersonii* and *S. tuberosum* Group Tuberosum must be both at the local and global chromosomal homology levels to ensure the observed pairing.

Our results have important implications for the use of *Solanum commersonii* in potato introgressive hybridisation breeding. Allotriploids like the ones analysed here are the most stringent test for homoeologous pairing and recombination between two species. Our data show that homoeologous pairing is highly frequent and recombination can take place, as would rather be expected in autotriploids, making introgression of the desired traits possible. It was not possible to discriminate the chromosomes belonging to the parental species or to pinpoint the introgressed chromosomal segments in complements from advanced backcrosses, which would have been valuable tools to assist negative selection for alien chromatin and to confirm introgression. On the other hand, the high genome similarity observed, together with the high collinearity between the two genomes, suggest that introgression of *S. commersonii* chromosomal regions should be highly efficient and the potential of this species for potato breeding can be fully exploited.

Acknowledgements:

Research was supported by: Agencia Nacional de Investigación e Innovación (ANII) (grant code FMV_1_2011_ 1_6697) and Comisión Sectorial de Investigación Científica, University of the Republic, Uruguay (grant code CSIC I+D 2012 383). P. Gaiero was supported by grant Proyecto 720 and CSIC Recursos Humanos, University of the Republic. We are thankful to Elsa Camadro for her very useful suggestions for the project and to José van de Belt and Mariana Andino for their assistance during the development of the experiments. We are grateful to Alicia Castillo and the technical staff at INIA for kindly providing plant material.

Literature cited

Adiwilaga KD, Brown CR (1991) Use of 2n pollen-producing triploid hybrids to introduce tetraploid Mexican wild species germ plasm to cultivated tetraploid potato gene pool. Theor Appl Genet 81(5): 645-652. doi:10.1007/BF00226732

Anderson LK, Covey PA, Larsen LR, Bedinger P, Stack SM (2010) Structural differences in chromosomes distinguish species in the tomato clade. Cytogenet Genome Res 129:24–34. doi:10.1159/000313850

Aversano R, Contaldi F, Ercolano MR, et al. (2015) The *Solanum commersonii* Genome Sequence Provides Insights into Adaptation to Stress Conditions and Genome Evolution of Wild Potato Relatives. Plant Cell 27:954–968. doi: 10.1105/tpc.114.135954

Bamberg JB, Hanneman RE, Palta JP, Harbage JF (1994) Using disomic 4x (2EBN) potato species' germplasm via bridge species *Solanum commersonii*. Genome 37:866–70

Barone A, Sebastiano A, Carputo D (1999) Chromosome pairing in *Solanum commersonii*- *S. tuberosum* sexual hybrids detected by *commersonii*-specific RAPDs and cytological analysis. Genome 42:218–224. doi:10.1139/gen-42-2-218

Bozza CG, Pawlowski WP (2008) The cytogenetics of homologous chromosome pairing in meiosis in plants. Cytogenet Genome Res 120:313–9. doi:10.1159/000121080

Bradshaw J (2007) Potato breeding strategy. In Vreugdenhil D, Bradshaw J, Gebhardt C, Govers F, Mackerron DKL, Taylor MA, Ross HA (eds.) Potato Biology and Biotechnology: Advances and Perspectives, 1st edn. Elsevier Science, Amsterdam, pp 157–177

Camadro E, Carputo D, Peloquin S (2004) Substitutes for genome differentiation in tuber-bearing *Solanum*: interspecific pollen-pistil incompatibility, nuclear-cytoplasmic male sterility, and endosperm. Theor Appl Genet 109:1369–76. doi:10.1007/s00122-004-1753-2

Campell BR, Song Y, Posch TE, Cullis CA, Town CD (1992) Sequence and organization of 5S ribosomal RNA-encoding genes of Arabidopsis thaliana. Gene 112:225–228. doi:10.1016/0378-1119(92)90380-8

Canady MA, Ji Y, Chetelat RT (2006) Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. Genetics 174:1775–1788. doi:10.1534/genetics.106.065144

Cardi T, D'Ambrosio E, Consoli D (1993) Production of somatic hybrids between frost-tolerant *Solanum commersonii* and *S. tuberosum*: characterization of hybrid plants. Theor Appl Genet 87:193–200

Carputo D (1999) Post-zygotic gametic selection due to endosperm balance number explains unusual chromosome numbers of 3x x 2x progeny in *Solanum*. Sex Plant Reprod 12:27–31. doi: 10.1007/s004970050168

Carputo D (2003) Cytological and breeding behavior of pentaploids derived from 3x x 4x crosses in potato. Theor Appl Genet 106:883–888. doi:10.1007/s00122-002-1165-0

Carputo D, Aversano R, Barone A, Matteo A, Iorizzo M, Sigillo L, Zoina A, Frusciante L (2009) Resistance to *Ralstonia solanacearum* of Sexual Hybrids Between *Solanum commersonii* and *S. tuberosum*. Am J Potato Res 86:196–202. doi:10.1007/s12230-009-9072-4

Carputo D, Barone A, Cardi T, Sebastiano A, Frusciante L, Peloquin SJ (1997) Endosperm balance number manipulation for direct in vivo germplasm introgression to potato from a sexually isolated relative (*Solanum commersonii* Dun.). Proc Natl Acad Sci USA 94 (22):12013–7. PMCID: PMC23688

Carputo D, Basile B, Cardi T, Frusciante L (2000) *Erwinia* resistance in backcross progenies of *Solanum tuberosum* x *S. tarijense* and *S. tuberosum* (+) *S. commersonii* hybrids. Potato Res 43:135–142. doi:10.1007/BF02357954

Carputo D, Cardi T, Frusciante L, Peloquin SJ (1995) Male fertility and cytology of triploid hybrids between tetraploid *Solanum commersonii* (2n = 4x = 48, 2EBN) and Phureja - Tuberosum haploid hybrids. Euphytica 83(124): 123-129. http://link.springer.com/article/10.1007/BF01678039

Carputo D, Garreffa P, Mazzei M, Monti L, Cardi T (1998) Fertility of somatic hybrids *Solanum commersonii* (2x, 1EBN) (+) *S. tuberosum* haploid (2x, 2EBN) in intra-and inter-EBN crosses. Genome 41 (6):776–781. doi:10.1139/g98-082

Carputo D, Monti L, Werner JE, Frusciante L (1999) Uses and usefulness of endosperm balance number. Theor Appl Genet 98:478–484. doi:10.1007/s001220051095

Carputo D, Parisi M, Consiglio F, Iovene M, Caruso G, Monti L, Frusciante L (2003) Aneuploid hybrids from 5x − 4x crosses in potato: Chromosome number, fertility, morphology and yield. Am J Potato Res 80:93–101. doi:10.1007/BF02870208

Chen Y-KH, Palta JP, Bamberg JB (1999) Freezing tolerance and tuber production in selfed and backcross progenies derived from somatic hybrids between *Solanum tuberosum* L. and *S. commersonii* Dun. TAG Theor Appl Genet 99:100–107. doi:10.1007/s001220051213

de Jong J, Wolters A, Kok J, Verhaar H, Eden J (1993) Chromosome pairing and potential for intergeneric recombination in some hypotetraploid somatic hybrids of *Lycopersicon esculentum* (+) *Solanum tuberosum*. doi:10.1139/g93-138

de Vicente MC, Tanksley SD (1991) Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato. Theor Appl Genet 83:173–178. doi: 10.1007/BF00226248

Demirci S, van Dijk ADJ, Sanchez Perez G, Aflitos SA, Ridder D, Peters SA (2017) Distribution, position, and genomic characteristics of crossovers in tomato recombinant inbred lines derived from an interspecific cross between *S. lycopersicum* and *S. pimpinellifolium*. Plant J. 89 (3): 554–564. doi:10.1111/TPJ.13406

Dong F, McGrath JM, Helgeson JP, Jiang J (2001) The genetic identity of alien chromosomes in potato breeding lines revealed by sequential GISH and FISH analyses using chromosome-specific cytogenetic DNA markers. Genome 44:729–734. doi:10.1139/gen-44-4-729

Dong F, Novy RG, Helgeson JP, Jiang J (1999) Cytological characterization of potato - *Solanum etuberosum* somatic hybrids and their backcross progenies by genomic *in situ* hybridization. Genome 42:987–992. doi:10.1139/gen-42-5-987

Dong F, Song J, Naess SK, Helgeson JP, Gebhardt C, Jiang J (2000) Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. Theor Appl Genet 101:1001–1007. doi:10.1007/s001220051573

Dong F, Tek AL, Frasca ABL, McGrath JM, Wielgus SM, Helgeson JP, Jiang J (2005) Development and characterization of potato-*Solanum brevidens* chromosomal addition/substitution lines. Cytogenet Genome Res 109:368–372. doi:10.1159/000082421

Gaiero P, van de Belt J, Vilaró F, Schranz ME, Speranza P, de Jong H (2017) Collinearity between potato (*Solanum tuberosum* L.) and wild relatives assessed by comparative cytogenetic mapping. Genome 60(3): 228-240. doi:10.1139/gen-2016-0150

Garriga-Calderé F, Huigen D, Angrisano A, Jacobsen E, Ramanna M (1998) BC1 to BC2 progenies derived from backcrossing potato (+) tomato fusion hybrids to potato: the selection of single additions for seven different tomato chromosomes. Theor Appl Genet 96:155–163. doi:10.1007/s001220050722

Garriga-Calderé F, Huigen D, Filotico F, Jacobsen E, Ramanna M (1997) Identification of alien chromosomes through GISH and RFLP analysis and the potential for establishing potato lines with monosomic additions of tomato chromosomes. Genome 40:666–73. http://www.ncbi.nlm.nih.gov/pubmed/18464856

Garriga-Calderé F, Huigen DJ, Jacobsen E, Ramanna MS (1999) Prospects for introgressing tomato chromosomes into the potato genome: An assessment through GISH analysis. Genome 42:282–288. doi:10.1139/gen-42-2-282

Gavrilenko T, Larkka J, Pehu E, Rokka V-M (2002) Identification of mitotic chromosomes of tuberous and non-tuberous *Solanum* species (*Solanum tuberosum* and *Solanum brevidens*) by GISH in their interspecific hybrids. Genome 45:442–449. doi:10.1139/g01-136

Gavrilenko T, Thieme R, Heimbach U, Thieme T (2003) somatic hybrids of *Solanum etuberosum* (+) dihaploid *Solanum tuberosum* and their backcrossing progenies: relationships of genome dosage with tuber development. Euphytica 131:323–332. doi:10.1023/A:1024041104170

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7:1869–1885. doi:10.1093/nar/7.7.1869

González M (2010) Análisis de la resistencia a *Ralstonia solanacearum* en una progenie segregante de *Solanum commersonii*. MSc Dissertation, Universidad de la República, Uruguay

González M, Galván G, Siri MI, Borges A, Vilaró F (2013). Resistencia a la marchitez bacteriana de la papa en *Solanum commersonii*. Agrociencia Uruguay 7: 45–54. www.scielo.edu.uy/pdf/agro/v17n1/v17n1a05.pdf

Hamernik AJ, Hanneman RE, Jansky SH (2009) Introgression of Wild Species Germplasm with Extreme Resistance to Cold Sweetening into the Cultivated Potato. Crop Sci 49:529. doi:10.2135/cropsci2008.04.0209

Hanneman RE (1989) The potato germplasm resource. Am Potato J 66:655–667

Hawkes JG (1966) Modern taxonomic work on the *Solanum* species of Mexico and adjacent countries. Am Potato J 43:81–103. doi:10.1007/BF02861580

Hawkes, J. (1990). The potato: evolution, biodiversity and genetic resources. Belhaven, London, UK

Henry IM, Dilkes BP, Tyagi AP, Lin H-Y, Comai L (2009) Dosage and parent-of-origin effects shaping aneuploid swarms in A. thaliana. Heredity 103:458–468. doi: 10.1038/hdy.2009.81

Horsman K, Gavrilenko T, Bergervoet M, Huigen D, Joe A, Jacobsen E (2001) Alteration of the genomic composition of *Solanum nigrum* (+) potato backcross derivatives by somatic hybridization: selection of fusion hybrids by DNA measurements. Plant Breed 120:201–207. doi:10.1046/j.1439-0523.2001.00591.x

Iovene M, Savarese S, Cardi T, Frusciante L, Scotti N, Simon PW, Carputo D (2007) Nuclear and cytoplasmic genome composition of *Solanum bulbocastanum* (+) *S . tuberosum* somatic hybrids. Genome 50 (5):443–450. doi:10.1139/G07-024

Jacobsen E, de Jong H, Kamstra S (1995) Genomic *in situ* hybridization (GISH) and RFLP analysis for the identification of alien chromosomes in the backcross progeny of potato (+) tomato fusion hybrids. Heredity 74:250–257. http://www.nature.com/hdy/journal/v74/n3/abs/hdy199538a.html

Jansky SH (2000) Breeding for Disease resistance in potato. In: Janick J (ed) Plant Breeding Reviews 19. John Wiley & Sons, Inc., Oxford, pp 69–152

Jansky SH (2006) Overcoming hybridization barriers in potato. Plant Breed 125:1–12. doi:10.1111/j.1439-0523.2006.01178.x

Jansky SH, Hamernik A (2009) The introgression of 2× 1EBN *Solanum* species into the cultivated potato using *Solanum verrucosum* as a bridge. Genet Resour Crop Evol 56:1107–1115. doi:10.1007/s10722-009-9433-3

Jansky SH, Peloquin SJ (2006) Advantages of Wild Diploid *Solanum* Species Over Cultivated Diploid Relatives in Potato Breeding Programs. Genet Resour Crop Evol 53:669–674. doi:10.1007/s10722-004-2949-7

Ji Y, Chetelat RT (2003) Homoeologous pairing and recombination in *Solanum lycopersicoides* monosomic addition and substitution lines of tomato. Theor Appl Genet 106:979–89. doi:10.1007/s00122-002-1090-2

Ji Y, Pertuzé R, Chetelat RT (2004) Genome differentiation by GISH in interspecific and intergeneric hybrids of tomato and related nightshades. Chromosome Res 12:107–16

Ji X (2014) Numerical and structural chromosome aberrations in cauliflower (*Brassica oleracea* var . *botrytis*) and Arabidopsis thaliana. PhD Dissertation, Wageningen University, The Netherlands

Jobes D, Hurley D, Thien L (1995) Plant DNA Isolation: A Method to Efficiently Remove Polyphenolics, Polysaccharides, and RNA. Taxon 44:379–386. doi:10.2307/1223408

Johnston SA, Hanneman RE (1982) Manipulations of endosperm balance number overcome crossing barriers between diploid *Solanum* species. Science 217:446–8. doi:10.1126/science.217.4558.446

Johnston S, Nijs T Den, Peloquin S, Hanneman R (1980) The significance of genic balance to endosperm development in interspecific crosses. Theor Appl Genet 57:5–9. doi:10.1007/BF00276002

Kreike CM, Stiekema WJ (1997) Reduced recombination and distorted segregation in a *Solanum tuberosum* (2x) x *S. spegazzinii* (2x) hybrid. Genome 40:180–7. doi: 10.1139/g97-026

Laferriere LT, Helgeson JP, Allen C (1999) Fertile *Solanum tuberosum* + *S. commersonii* somatic hybrids as sources of resistance to bacterial wilt caused by *Ralstonia solanacearum*. Theor Appl Genet 98:1272–1278. doi:10.1007/s001220051193

Lange W, Wagenvoort M (1973) Meiosis in Triploid *Solanum tuberosum* L. Euphytica 22:8–18. doi:10.1007/BF00021550

Larrosa F, Maune J, Erazzú L, Camadro E (2012) Meiotic abnormalities underlying pollen sterility in wild potato hybrids and spontaneous populations. Plant Biol 14:223–33. doi:10.1111/j.1438-8677.2011.00470.x

Lenormand T, Dutheil J (2005) Recombination difference between sexes: A role for haploid selection. PLoS Biol 3:0396–0403. doi: 10.1371/journal.pbio.0030063

Lysak MA, Pecinka A, Schubert I (2003) Recent progress in chromosome painting of Arabidopsis and related species. Chromosome Res 11:195–204. http://www.ncbi.nlm.nih.gov/pubmed/12769287

Masuelli RW, Camadro EL (1992) Cytological Analysis and Fertility of *Solanum commersonii* Dun. x *Solunum gourlayi* Haw. Triploid hybrids. Cytologia 57:161–166. http://doi.org/10.1508/cytologia.57.161

Matsubayashi M (1991) Phylogenetic relationships in the potato and its related species. In: Tsuchiya T, Gupta P (eds) Chromosome Engineering in Plants: Genetics, Breeding, Evolution, Part B. Elsevier, pp 93–118

McKee BD (2004) Homologous pairing and chromosome dynamics in meiosis and mitosis. Biochim Biophys Acta - Gene Struct Expr 1677:165–180. doi:10.1016/j.bbaexp.2003.11.017

Mendoza HA Haynes FL (1974). Genetic relationship among potato cultivars grown in the United States. Hortscience 9:328– 330. http://agris.fao.org/agris-search/search.do?recordID=US201303178585

Micheletto S, Boland R, Huarte M (2000) Argentinian wild diploid *Solanum* species as sources of quantitative late blight resistance. Theor Applied Genet 101(5–6): 902–906. http://doi.org/10.1007/s001220051560

Ono S, Sanetomo R, Hosaka K (2016) Genetic transmission of *Solanum demissum* (2n = 6x = 72) chromosomes from a pentaploid hybrid of *S. tuberosum* (2n = 4x = 48) into the aneuploid BC1 progeny. Euphytica 207:149–168. doi:10.1007/s10681-015-1558-1

Ortiz R (1998) Potato breeding via ploidy manipulations. In: Janick J (ed) Plant Breeding Reviews 16, John Wiley & Sons, Inc., Oxford, pp 15-86. doi:10.1002/9780470650110.ch2

Ortiz R, Simon P, Jansky S, Stelly D (2009) Ploidy manipulation of the gametophyte, endosperm and sporophyte in nature and for crop improvement: a tribute to Professor Stanley J. Peloquin (1921-2008). Ann Bot 104:795–807. doi:10.1093/aob/mcp207

Palta JP, Li PH (1979) Frost hardiness related to leaf anatomy and *Solanum* spp distribution. Crop Sci. 19:665–671.

Palta JP, Simon G (1993) Breeding potential for improvement of freezing stress resistance: genetic separation of freezing tolerance, freezing avoidance, and capacity to cold acclimate. In: Li PH (ed) Advances in Plant Cold Hardiness. CRC Press, Boca Raton, Florida, pp 299–310

Parokonny AS, Marshall JA, Bennett MD, Cocking EC, Davey MR, Power JB (1997). Homoeologous pairing and recombination in backcross derivatives of tomato somatic hybrids [*Lycopersicon esculentum* (+) *L. peruvianum*]. Theor Appl Genet 94 (6-7): 713–723. doi:10.1007/s001220050470

Peloquin S, Jansky S, Yerk G (1989) Potato cytogenetics and germplasm utilization. Am Potato J 629–638. doi:10.1007/BF02853983

Pendinen G, Gavrilenko T, Jiang J, Spooner DM (2008) Allopolyploid speciation of the Mexican tetraploid potato species *Solanum stoloniferum* and *S. hjertingii* revealed by genomic *in situ* hybridization. Genome 51:714–20. doi:10.1139/G08-052

Pendinen G, Spooner DM, Jiang J, Gavrilenko T (2012) Genomic *in situ* hybridization reveals both auto- and allopolyploid origins of different North and Central American hexaploid potato (*Solanum* sect. Petota) species. Genome 55:407–15. doi:10.1139/g2012-027

Prieto P, Shaw P, Moore G (2004) Homologue recognition during meiosis is associated with a change in chromatin conformation. Nat Cell Biol 6:906–908. doi:10.1038/ncb1168

Ramanna MS, Hermsen JGT (1982) Gene transfer from non-tuberous to tuberous *Solanum* species: Evidence from meiosis in hybrids. Euphytica 31:565–572. doi:10.1007/BF00039194

Rivard SR, Cappadocia M, Landry BS (1996) A comparison of RFLP maps based on anther culture derived, selfed, and hybrid progenies of *Solanum chacoense*. Genome 39:611–621. doi: 10.1139/g96-078

Ross H (1986) Potato breeding–problems and perspectives. In: Parey P (ed) Advances in plant breeding, Verlag, Berlin-Hamburg, pp 65-94.

Sandro PA, Rebollo MI, Gaiero P, Vaio M, Vilaró F, Speranza P. Diseño de microsatélites para *Solanum commersonii* a partir de información genómica. (2016) ALAG 2016. Journal of Basic and Applied Genetics Suppl Vol XXVII (1): 276. http://www.alag2016.org/admin/files/alag2016/upload/files/V.XXVIII_2016_Suppl1_19092016.pdf

Schubert I, Fransz PF, Fuchs J, de Jong JH (2001) Chromosome painting in plants. Methods Cell Sci 23 (1-3):57–69. http://www.ncbi.nlm.nih.gov/pubmed/11741144

Singh RJ (2003) Plant cytogenetics. CRC Press, Boca Raton, Florida

Solomon-Blackburn RM, Barker H (2001) Breeding virus resistant potatoes (*Solanum tuberosum*): a review of traditional and molecular approaches. Heredity 86 (1):17–35. http://www.ncbi.nlm.nih.gov/pubmed/11298812

Sybenga J (1999) What makes homologous chromosomes find each other in meiosis? A review and a hypothesis. Chromosoma 108 (4):209–219. doi:10.1007/s004120050371

Szinay D, Chang S-B, Khrustaleva L, Peters SA, Schijlen E, Bai Y, Stiekema WJ, van Ham RCHJ, de Jong H (2008) High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. Plant J 56:627–637. doi:10.1111/j.1365-313X.2008.03626.x

Szinay DD, Wijnker E, van den Berg R, Visser RGF, de Jong H, Bai Y (2012) Chromosome evolution in *Solanum* traced by cross-species BAC-FISH. New Phytol 195:688–698. doi:10.1111/j.1469-8137.2012.04195.x

Tang X, Szinay D, Lang C, Ramanna MS, van der Vossen EAG, Datema E, Lankhorst RK, de Boer J, Peters SA, Bachem C, Stiekema W, Visser RGF, de Jong, H, Bai Y (2008) Cross-species bacterial artificial chromosome-fluorescence *in situ* hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics 180:1319–1328. doi:10.1534/genetics.108.093211

Tang X, De Boer JM, Van Eck HJ, Bachem C, Visser RGF, De Jong H (2009) Assignment of genetic linkage maps to diploid *Solanum tuberosum* pachytene chromosomes by BAC-FISH technology. Chromosome Res 17:899–915. doi:10.1007/s10577-009-9077-3

Tek AL, Stevenson WR, Helgeson JP, Jiang J (2004) Transfer of tuber soft rot and early blight resistances from *Solanum brevidens* into cultivated potato. Theor Appl Genet 109:249–254. doi: 10.1007/s00122-004-1638-4

Vega SE, Palta JP, Bamberg JB (2000) Variability in the rate of cold acclimation and deacclimation among tuber-bearing Solanum (potato) species. J Amer Soc Hort Sci 125 (2):205–211. http://journal.ashspublications.org/content/125/2/205

Verlaan MG, Szinay D, Hutton SF, de Jong H Kormelink R, Visser RGF, Scott JW, Bai Y (2011) Chromosomal rearrangements between tomato and *Solanum chilense* hamper mapping and breeding of the TYLCV resistance gene *Ty*-1. Plant J 68:1093–1103. doi:10.1111/j.1365-313X.2011.04762.x

Wijnker E, de Jong H (2008) Managing meiotic recombination in plant breeding. Trends Plant Sci 13:640–6. doi: 10.1016/j.tplants.2008.09.004

Zhong XB, de Jong H, Zabel P (1996) Preparation of tomato meiotic pachytene and mitotic metaphase chromosomes suitable for fluorescence *in situ* hybridization (FISH). Chromosome Res 4:24–28.

CHAPTER 4

Collinearity between potato (*Solanum tuberosum* L.) and wild relatives assessed by comparative cytogenetic mapping

Authors:
Paola Gaiero[1,2], José van de Belt[2], Francisco Vilaró[3], M. Eric Schranz[4], Pablo Speranza[1], Hans de Jong[2]

1. Department of Plant Biology, Facultad de Agronomía, Universidad de la República. Garzón 780, PC 12900, Montevideo, Uruguay.
2. Laboratory of Genetics, Wageningen University. Droevendaalsesteeg 1, P.O. Box 16, 6708 PB, Wageningen, The Netherlands.
3. Horticulture Unit, National Institute for Agricultural Research, Ruta 48 km 10, Las Brujas, Uruguay.
4. Biosystematics Group, Wageningen University, Wageningen, the Netherlands

Published in:
Genome 60 (3): 228–40. (2017) doi:10.1139/gen-2016-0150.

Abstract:

Introgressive hybridization has been widely used in potato breeding to transfer biotic and abotic resistances from wild relatives. A major bottleneck to this process is the lack of genome collinearity between the donor (alien) genome and the recipient crop genome. Chromosomal structural differences may create unbalanced segregation of bivalents or multivalents or cause unintended retention of large blocks of DNA surrounding the region of interest (a phenomenon known as linkage drag). To assess large-scale collinearity between potato and two of its wild relatives (*Solanum commersonii* and *S. chacoense*), we used multicolour BAC-FISH mapping of markers with known positions on the RH potato chromosomes and anchored on the ultrahigh density potato genetic map. BAC probes could successfully be hybridized to the *S. commersonii* and *S. chachoense* pachytene chromosomes, confirming their correspondence with linkage groups in RH potato. Our study shows that the order of BAC signals is conserved. Distances between BAC signals were quantified and compared; some differences found suggest either small-scale rearrangements or reduction/amplification of repetitive sequences. We conclude that *S. commersonii* and *S. chacoense* are collinear with cultivated *S. tuberosum* on the whole chromosome scale, making these amenable species for efficient introgressive hybridization breeding.

Introduction

Potato (*Solanum tuberosum* L.) is the third most important food crop worldwide (after rice and wheat) and is number one among vegetables (Birch et al. 2012; Jansky et al. 2013). As with other crops, potato growers face severe yield losses by pests and diseases (Haynes and Lu 2005), encouraging breeders to select for traits that confer specific resistance and tolerance within the wild relative gene pool (Jansky 2000). Strategies have been developed to overcome hybridization barriers and perform the necessary manipulations to introgress such resistances into potato cultivars (Jansky 2000; Jansky 2006). However, only few resistance genes have been successfully introgressed into cultivars, leaving most wild resources to various pests and disease resistances untapped (Jansky 2000; Solomon-Blackburn and Barker 2001; Simko et al. 2009; Birch et al. 2012). One of the reasons for this underutilization is lack of knowledge on the physical (co-)localization of genetic loci between cultivated potato and its close relatives that are potential donors for introgressive hybridization breeding initiatives. If there are physical differences among the chromosomes of potato and its donor relatives they can hinder homoeologous recombination, cause problems in segregation during the meiosis of hybrids and/or produce linkage drag (Young and Tanksley 1989). Such changes hamper introgression of selected traits into the recipient crop rendering the ultimate aim unachievable (Verlaan et al. 2011).

Genome-wide collinearity within Solanaceae has been established mostly by comparative genetic analysis using high-density molecular marker maps. More recently, cytogenetic studies have explored the use of Fluorescence *in situ* Hybridisation (FISH)-

based chromosome staining to establish structural and numerical comparisons of their karyotypes (Tang et al. 2008; Iovene et al. 2008; Szinay et al. 2008; Wu and Tanksley 2010; Anderson et al. 2010; Szinay et al. 2010; Lou et al. 2010; Verlaan et al. 2011; Szinay et al. 2012). With the development of high-throughput DNA sequencing, genome assemblies for tomato (The Tomato Genome Consortium 2012), potato (The Potato Genome Sequencing Consortium 2011) and several of their wild relatives (Bolger et al. 2014; Aflitos et al. 2014; Aversano et al. 2015) have become available. Concerted genomics and bioinformatics efforts have improved genome assemblies (Sharma et al. 2013; Shearer et al. 2014). However, only a few studies have utilized sufficient high-quality physical maps needed to reveal structural differences related to introgression barriers (Peters et al. 2012; Aflitos et al. 2014; Aflitos et al. 2015; de Boer et al. 2015). Cytogenetic information obtained by high-resolution FISH maps on extended chromosomes can provide important information on structural rearrangements.

There is a long history to the study of potato chromosome morphology, with a starting point being the detailed descriptions of gross morphology of pachytene bivalents by Yeh and Peloquin (1965) and Ramanna and Wagenvoort (1976) for dihaploid clones derived from different potato cultivars. Their studies enabled full identification of all chromosome pairs through morphological features such as length, centromere positions and heterochromatin patterns. The few discrepancies between their observations suggest intraspecific variation among different *Solanum tuberosum* genotypes (Yeh and Peloquin 1965; Ramanna and Wagenvoort 1976). However, since these studies are extremely laborious, such pachytene karyotypes are only available for one of the wild potato relatives, namely *Solanum clarum* (Marks 1969). Other types of structural genomic information for wild potato diploid species in Section Petota were initially obtained through meiotic pairing and fertility studies of their F1 hybrids. They share a common haploid genome (A), with the exception of species in series Etuberosa. Across most of the species, this A genome is only differentiated by cryptic variations, which are so small that pairing behaviour appears regular, or slightly more differentiated by definite structural differences, that are evidenced by meiotic irregularities (Matsubayashi 1991). More detailed and higher resolution chromosomal studies are needed to look at the genomes of potato wild relatives and assess whether these differences will have an effect on chromosome pairing, recombination and segregation. Such data can be obtained using light microscopy of long pachytene chromosomes (Tang et al. 2008; Iovene et al. 2008; Lou et al. 2010; Verlaan et al. 2011; Szinay et al. 2012), by electron microscopy of synaptonemal complexes (Anderson et al.2010) or by comparative chromosomal painting.

Fluorescence *In situ* Hybridization with Bacterial Artificial Chromosomes as probes (BAC-FISH) has proven to be a powerful diagnostic tool for high resolution genomic studies and introgression breeding (Verlaan et al. 2011). Various previously unknown inversions could be described in the *Solanum* crops using BAC-FISH analyses (Tang et al. 2008; Wu and Tanksley 2010; Szinay et al. 2010; Lou et al. 2010; Szinay et al. 2012). This strategy

was also used to identify potato chromosomes using a set of 12 chromosome-specific cytogenetic DNA markers (CSCDM) selected from a *S. bulbocastanum* library (Dong et al. 2000) and also to explain the suppression of recombination of an introgressed segment harbouring resistance to TYLCV in tomato (Verlaan et al. 2011). Using multi-colour fluorescence microscopy it is possible to hybridize several probes in a single experiment, simplifying the examination of any *Solanum* species chromosome set to only a few experiments (Tang et al. 2009; Szinay et al. 2010). Moreover, hybridization under lower stringency conditions allows cross-species painting of tomato or potato probes to display homoeologous chromosomal positions in related *Solanum* species. For example, the RHPOTKEY BAC library was developed for diploid potato clone RH89-039-16 (Borm 2008) and anchored to AFLP markers in the ultrahigh density (UHD) genetic map (van Os et al. 2006). From this library, 60 BACs were selected as a cytogenetic painting set with anchors for genetic map positions on the potato chromosomes, thus providing a useful tool to study collinearity between potato and its wild relatives (Tang et al. 2009).

For our study we focus on the diploid *Solanum commersonii* and *S. chacoense* (2n = 2x = 24), two potato relatives adapted to a variety of environmental and climate conditions (Hawkes 1990). Both are widely distributed in Brazil, Argentina and Uruguay, whereas *S. chacoense* is also distributed in Paraguay and Bolivia. Taxonomically they are included in section Petota, which comprises all tuber-bearing species including cultivated potatoes (Hawkes 1990). *Solanum commersonii* belongs to the potato tertiary gene pool, because it can be hybridized to potato through unreduced gametes and bridge crosses, whereas *S. chacoense* belongs to its secondary gene pool and can be directly crossed with tetraploid potato through unreduced gametes (Jansky 2006). Both species are excellent resources for potato breeding as they harbour resistances to many biotic stresses such as *Phytophtora infestans* (agent of late blight, the most important potato disease) (Micheletto et al. 2000), *Ralstonia solanacerum* (agent of bacterial wilt, the second most important potato disease) (González et al. 2013) and other severe potato pathogens, such as *Pectobacterium, Verticillium, Alternaria,* X and Y viruses (Laferriere et al. 1999). Additionally, they have been described as good reservoirs for broadening the genetic base to improve adaptation to cold and drought (Chen et al. 1999). A few accessions from these species have been used in potato introgression breeding (Carputo et al. 1997; Laferriere et al. 1999; Chen et al. 1999; Carputo et al. 2000; Chen et al. 2013; González et al. 2013). However, their full potential has not been realized due to lack of genetic and genomic information and the potential for homeologous incorporation without linkage drag. Although the genome sequence of *S. commersonii* has been published (Aversano et al. 2015), its fragmented assembly does not allow robust structural comparisons with potato. The aim of this work was to build a cytogenetic map of *Solanum commersonii* and *S. chacoense* with respect to cultivated potato (*Solanum tuberosum* L.) to discover large-scale translocations and/or inversions, if any, that could hamper introgressive breeding using these wild relatives as donors and could assist in the assembly of their genome sequences.

Materials and methods

Plant material and slides preparation

We used three clones of *Solanum commersonii* (02.04.1 from Rocha, southeast Uruguay, 04.02.3 from Colonia, southwest Uruguay and 05.05.2.4 from Canelones, southern Uruguay) and one clone of *S. chacoense* (07.01.7 from Salto, northwest Uruguay). In addition, we studied two full-sib triploid interspecific hybrids (*S. commersonii* (2n) x *S. tuberosum* Group Phureja (n)) developed by González (2010). Flower buds containing anthers at meiotic prophase I were collected in the morning (11-12 am), fixed in absolute ethanol: acetic

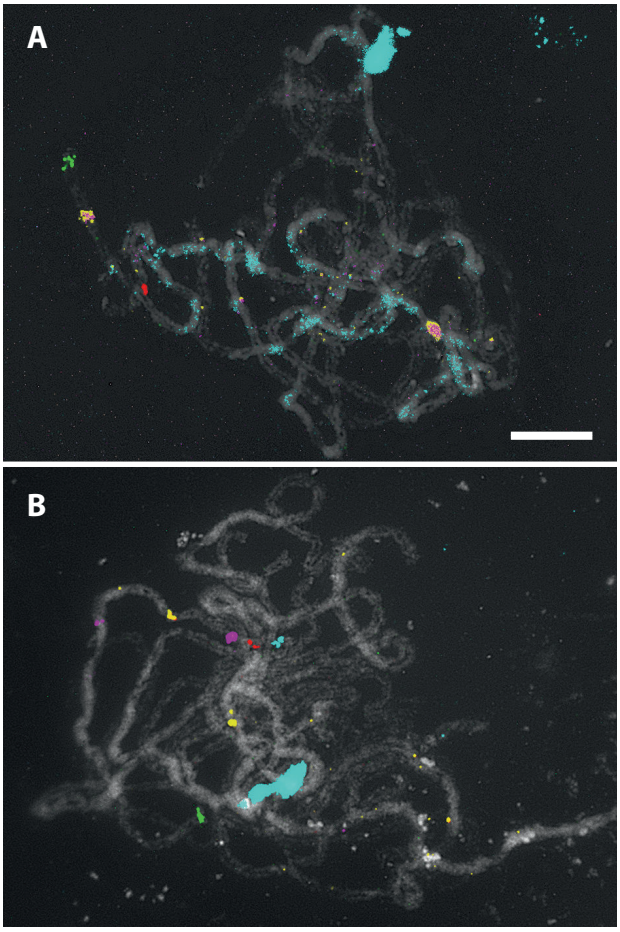


Figure 1: Representative images of pachytene complements in *S. commersonii* (A) and *S. chacoense* (B) hybridized with BAC probes for chromosomes 5 and 12 (chr 5 and chr 12). The green, yellow, blue, purple and red signals in A) correspond to BACs RH081B09, RH095I08, RH076O08, RH089A21, RH044A21, respectively, from Linkage group 5. The blue, red, purple, yellow and green signals in B) correspond to BACs RH106P06, RH016H06, RH183I02, RH043B17 and RH084C24, respectively, from Linkage group 12. The 45S rDNA (large blue signal) and 5S rDNA (large yellow signal) probes were used as hybridization controls. Note the dispersed blue signal on the chromosomes in (A) indicating hybridization of repetitive sequences not blocked by C0t-100. Scale bar = 10 µm.

acid (3:1) for one day and then stored in 70 % ethanol at 4º C until further use. Pachytene spreads preparation followed the procedure described in Szinay et al. (2008) with few minor modifications.

Probes, blocking and Fluorescence in situ Hybridization

We used the RHPOTKEY potato BAC library constructed from the RH clone RH89-039-16, previously selected by Tang et al. (2009), for each of the twelve potato linkage groups. BAC DNA was isolated with the QIAGEN (Valencia, CA) plasmid minikit and amplified with the REPLI-g minikit (QIAGEN, Valencia, CA). For direct labelling we used Cyanine 3-dUTP (Enzo Life Sciences), Cyanine 3.5-dCTP (GE Healthcare, Sweden) or Diethylaminocoumarin-5-dUTP (Perkin Elmer Inc), and for indirect labelling biotin-16-UTP or digoxigenin-11-dUTP, all of them incorporated in the probe DNA by standard nick translation reaction (Roche Diagnostic, Indianapolis). Probes from the 45S rDNA from the pTa71 plasmid (Gerlach and Bedbrook 1979) and 5S rDNA from the pCT4.2 plasmid (Campell et al. 1992) were used as hybridization controls. We performed FISH experiments as described previously (Zhong et al. 1996) with minor modifications. We carried out incubations of more than three days to obtain better hybridization of the BAC probe DNA. Hybridization of repetitive sequences in the BAC DNA was suppressed by adding unlabelled $C_{ot}100$ (50x probe concentration) which was prepared from genomic *S. commersonii* and *S. tuberosum* DNA as described by Tang et al. (2008). To enhance hybridization efficiency, we carried out post-hybridization washes under low stringency conditions for 3 x 5 min in 20% formamide, 2x SSC at 42º C. Chromosomes were counterstained with 5 $\mu g.mL^{-1}$ DAPI in Vectashield anti-fade (Vector Laboratories).

Image acquisition and processing

We examined slides under a Zeiss Axioplan 2 imaging photomicroscope (http://www.zeiss.com/) with epifluorescence illumination and filter sets for 4′,6-diamidino-2-phenylindole (DAPI), DEAC (blue), FITC (green), Cy3 (orange), Cy3.5 (red), and Cy5/Alexa Fluor 647 (far-red) fluorescence. Only late pachytene complements were considered. Selected images were captured using a Photometrics Sensys 1305 x 1024 pixel CCD camera (Photometrics, http://www.photomet.com). Image thresholding was performed using Genus image analysis software (Applied Imaging Corporation, http://www.aicorp.com). We performed image adjustments with Adobe® Photoshop® software as follows: DAPI images were displayed in light grey and sharpened using a 7 x 7 pixel Hi-Gauss high-pass spatial filter to accentuate minor details and the heterochromatin morphology of the chromosomes. The remaining fluorescence images were pseudo-coloured in blue (DEAC), green (FITC), orange (Cy3), red (Cy3.5, Texas Red) and purple (Cy5), and overlaid in multichannel mode. Brightness and contrast adjustments were performed using the Levels tool in Adobe Photoshop affecting all pixels equally.

Distance measurements and analysis

We straightened the images of late pachytene chromosomes using ImageJ (http://rsb.info.nih.gov/ij/plugins/straighten.html) and the cubic spline interpolation straighten-

ing plug-in of (Kocsis et al. 1991). We measured the total length of pachytene chromosomes, the centromere index (length of the short arm/total chromosome length) x 100 (Levan et al. 1964), the size of the pericentromere heterochromatin, the size and position of interstitial heterochromatin knobs and the relative position of BAC signals also using ImageJ. We used the chromosome numbering system employed by Tang et al. (2009), which corresponded to the potato linkage groups based on the molecular markers assigned to the UHD genetic map (van Os et al. 2006). We also followed Tang et al. (2009) in the alignment of chromosomes in the karyotype. For chromosome pairs 5, 7, 10, 11 and 12, we aligned the short arms to the south arms in the genetic map, contrary to the convention. We measured bivalents from different cell spreads and slides for each chromosome pair. In the case of *S. commersonii*, we analysed bivalents coming from different individuals, to account for intraspecific variation. Average lengths and standard deviations were calculated for all measurements.

We calculated the cytological position of each BAC clone as (S/T) ×100, where S is the distance (in µm) from the FISH site to the end of the of the chromosome arm aligned to the north and T is the total length of the chromosome (in µm). Because no genetic map is available for either *S. commersonii* nor *S. chacoense*, we could not convert positions to fraction

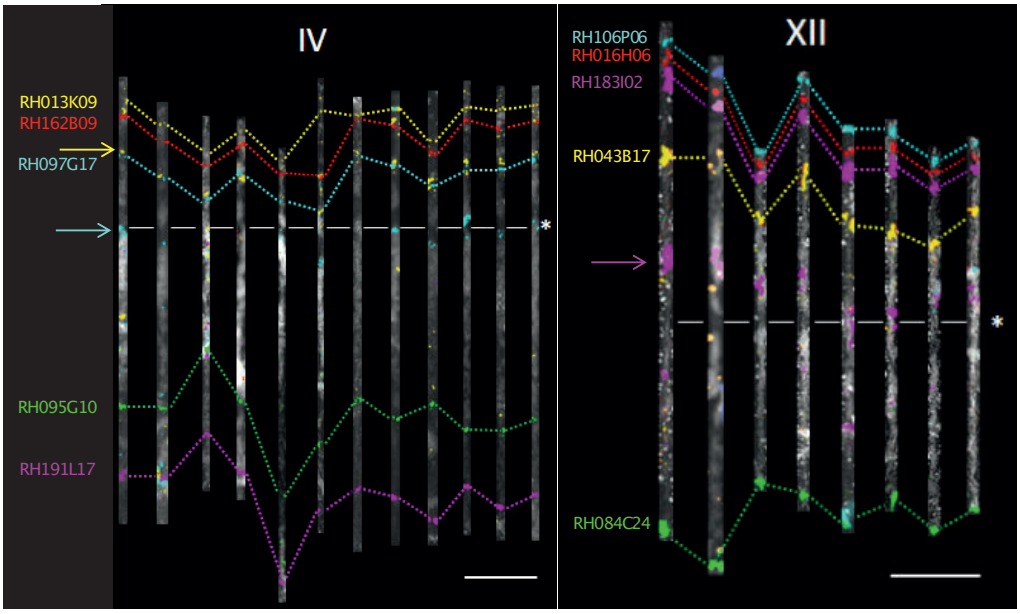


Figure 2: Examples of bivalents from chromosome 4 of *S. chacoense* and chromosome 12 of *S. commersonii*, with the colour scheme and the names of the five BAC clones used. Note the blue and yellow signals from BACs RH097G17 and RH013K09, respectively (yellow arrow) that always map together in chr 4 of *S. chacoense* and the blue and pink signals (blue and pink arrows) on the pericentromeric heterochromatin of chr 4 and chr 12 of *S. chacoense* and *S. commersonii* respectively, indicating that BACs RH097G17 and RH183I02 contained repeats that could not be blocked by C_{0t}-100 during the hybridization experiments. Scale bars = 10 µm

lengths (FL) for a direct comparison with the physical positions of BACs found in Tang et al. (2009). Thus, we converted the data in their work to relative positions as stated above.

Table 1. Absolute and relative length (with respect to total complement lenght), estimated DNA content, centromere index and heterochromatin content of the 12 chromosomes of *Solanum commersonii* (cmm) and *S. chacoense* (chc)

		Chromosome [a]					
		1	2	3	4	5	6
Absolute lenght (µm)	**cmm**	75.8±12.1	52.6±8.5	56.5±12.3	61.8±16.5	48.5±10.2	60.7±10.4
	chc	58.8±9.3	44.0±6.2	61.3±7.2	55.0±10.6	43.4±7.9	66.5±12.7
Relative lenght (%)	**cmm**	11.5±1.8	8.0±1.3	8.5±1.9	9.3±2.0	7.3±1.5	9.2±1.6
	chc	9.8±1.6	7.3±1.0	10.2±1.2	9.1±1.8	7.2±1.3	11.0±2.1
DNA content (Mb)[b]	**cmm**	95.2±15.2	66.1±10.6	71.0±15.5	77.6±20.8	60.9±12.8	76.3±13.0
	chc	60.1±9.6	45.1±6.3	62.7±7.4	56.3±10.9	44.5±8.1	68.1±13.0
Centromere index (%)[c]	**cmm**	23.0±4.9	13.7±3.6	18.9±2.6	37.8±9.8	41.0±7.0	21.1±5.0
	chc	27.8±8.4	14.6±3.1	23.8±1.1	28.5±6.3	43.4±8.6	27.8±7.0
Absolute pericentromere heterochromatin (µm)							
short arm	**cmm**	17.8±5.7	4.2±1.4	9.3±2.3	10.6±2.4	11.0±1.8	8.1±2.0
long arm		12.9±3.6	27.1±5.2	7.3±1.7	7.1±4.4	8.8±2.3	20.9±3.0
short arm	**chc**	12.6±5.8	3.7±0.8	9.5±2.1	4.4±0.9	7.7±2.0	12.0±4.1
long arm		16.7±7.0	22.9±2.5	7.6±3.4	19.6±2.8	11.7±2.3	15.9±6.7
Relative heterochromatin (%)	**cmm**	42.4±12.8	59.5±12.6	42.3±9.0	28.6±10.9	40.8±8.3	51.1±8.8
	chc	49.8±21.7	48.2±21.7	41.9±14.4	43.6±6.7	44.6±9.9	44.6±16.7

a) Chromosomes were ordered and numbered according to their corresponding linkage groups (van Os et al. 2006).
b) Estimation of the DNA content of each chromosome was based on relative length and the assumption that (1) the haploid genome size for potato is 840 Mb, for Solanum commersonii is 0.85 pg = 830 Mb (Aversano et al., 2015) and for S. chacoense 0.63 pg = 617 Mb (Bennett et al., 1976); (2) the euchromatic and heterochromatic regions contain equal amounts of DNA (Mb/µm).
c) Centromere index was calculated as percentage of short arm/total chromosome length (Levan et al. 1964).

Results

Morphology of pachytene chromosomes in *S. commersonii* and *S. chacoense*

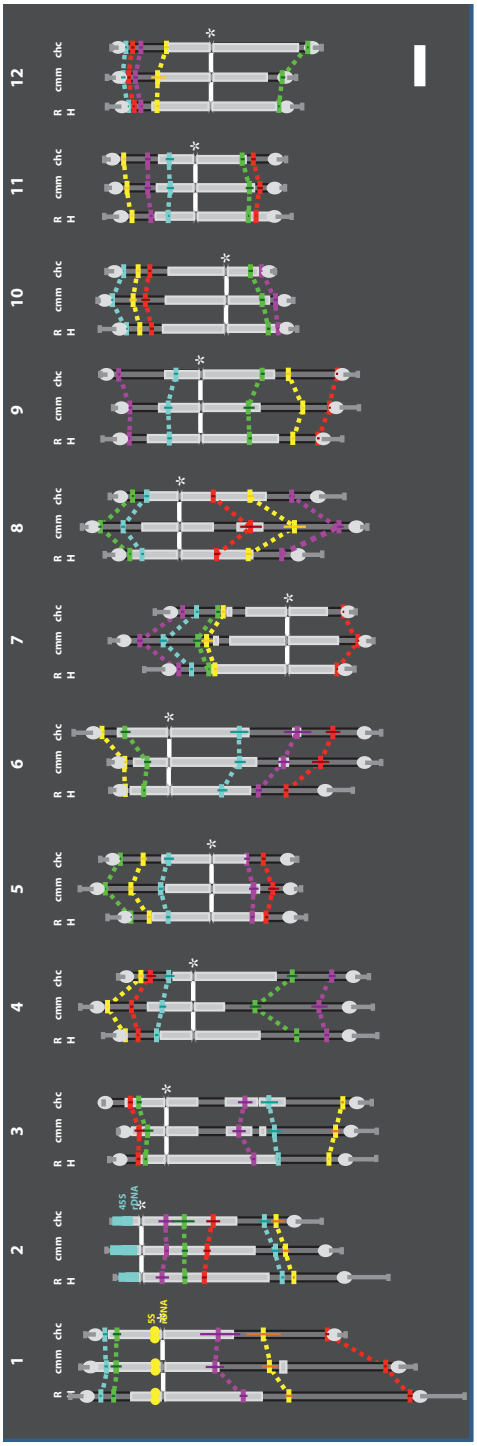
For each chromosome we measured the lengths in 2 to 12 different cell complements from *S. commersonii* and *S. chacoense* (see examples in Figure 1). Lengths ranged from 75.8±12.1 µm (chromosome pair 1, chr 1) to 40.2±5.6 µm (chr 11) in *S. commersonii* and from 66.5±12.7 µm (chr 6) to 37.5±6.0 µm (chr 10) in *S. chacoense,* from a total complement length of 660.9±116.0 µm and 602.7±113.1 µm, respectively (Table 1). Centromeres were unambiguously identified as small and faintly DAPI fluorescing regions, flanked by brightly fluorescing heterochromatin on both sides. According to their centromere indexes (i), most chromosome pairs in *S. commersonii* and *S. chacoense* were median (37.5<i<50) and submedian (25<i<37.5), with only pairs 2 and 3 having a subterminal (12.5<i<25) centromere position. DAPI counterstaining provided clear distinction between the brightly fluorescing heterochromatin in the pericentromeres, NORs and distal heterochromatin blocks, and the weaker fluorescence of euchromatin. The size of the pericentromeres ranged from 16.2±3.7 µm (chr 8) to 31.3±6.7 µm (chr 2) in *S. commersonii* and

7	8	9	10	11	12	Total
59.6±8.5	65.8±6.5	53.4±11.9	44.7±5.0	40.2±5.6	41.3±8.5	660.9±116.0
43.4±10.7	48.6±16.6	57.3±10.3	37.5±6.0	40.0±8.2	46.9±7.3	602.7±113.0
9.0±1.3	9.9±1.0	8.1±1.8	6.8±0.8	6.1±0.9	6.3±1.3	100
7.2±1.8	8.1±2.8	9.5±1.7	6.2±1.0	6.6±1.4	7.8±1.2	100
74.9±10.6	82.6±8.2	66.9±15.0	56.1±6.3	50.5±7.0	51.9±10.6	830
44.4±11.0	49.8±17.0	58.7±10.6	38.4±6.2	40.9±8.4	48.0±7.5	617
31.9±5.6	33.3±4.6	37.4±4.5	32.1±7.6	49.6±6.9	44.3±10.2	
34.2±7.1	30.2±8.8	39.5±6.0	28.3±6.1	50.9±8.4	52.5±8.2	
12.0±2.9	8.6±1.3	9.9±2.1	9.9±2.4	14.0±1.8	13.0±3.2	128.4±28.3
10.7±5.4	7.6±2.4	13.8±2.3	14.5±1.9	8.9±1.7	9.7±2.3	149.3±36.2
9.1±1.6	7.7±3.4	8.3±2.8	7.3±0.3	11.9±2.9	15.2±2.9	109.4±29.6
9.5±2.2	13.1±4.2	17.2±3.8	13.5±3.1	8.3±2.5	8.0±1.8	164.0±42.3
40.2±14.3	34.1±6.4	44.1±8.4	54.6±9.7	57.0±8.6	54.9±13.3	42.0±9.8
45.4±9.5	42.9±15.5	44.4±11.4	55.6±9.2	50.7±13.5	49.4±10.0	45.4±11.9

Table 2. Potato BACs used and their relative physical locations on *Solanum commersonii* (cmm) and *Solanum chacoense* (chc) chromosomes.

chromosome arm[a]	BAC[b]	Rel. Phys Location % (n)		chromosome arm[a]	BAC[b]	Rel. Phys Location %[c] (n)	
		cmm	chc			cmm	chc
1S	RH083J12	4.2±1.1 (4)	4.5±2.1 (5)	7L	RH075O19	7.3±2.1 (5)	6.4±1.6 (7)
1S	RH078E17	8.0±3.7 (4)	8.7±3.4 (5)	7L	RH078E15	17.1±2.6 (5)	14.4±1.4 (7)
1L	RH092C10	39.5±6.0 (4)	50.8±16.8 (5)	7L	RH170M01	31.1±4.1 (5)	26.1±2.3 (7)
1L	RH158N18	55.2±4.7 (4)	69.6±14.3 (5)	7L	RH093G10	34.7±4.8 (5)	29.5±2.3 (7)
1L	RH096H03	96.0±0.7 (4)	96.4±1.8 (5)	7S	RH186I02	98.2±0.5 (5)	97.1±0.9 (7)
2L	RH075N07	25.2±1.5 (6)	28.3±9.1 (3)	8S	RH122L16	3.7±1.2 (3)	6.5±1.3 (7)
2L	RH095G12	33.7±2.8 (6)	38.7±12.2 (3)	8S	RH055L21	11.0±1.4 (3)	13.6±2.2 (7)
2L	RH174A15	44.0±3.3 (6)	54.8±8.0 (3)	8L	RH184D07	59.6±8.2 (3)	47.0±1.9 (7)
2L	RH076J19	75.9±3.5 (6)	83.3±9.0 (3)	8L	RH122E19	76.0±8.3 (3)	65.6±3.2 (7)
2L	RH055P13	80.6±3.4 (6)	89.7±8.0 (3)	8L	RH127J02	92.5±2.9 (3)	86.6±2.7 (7)
3S	RH078O14	7.3±4.2 (4)	9.6±1.9 (4)	9S	RH135I22	4.3±1.8 (4)	4.2±1.0 (2)
3S	RH159O01	10.8±4.5 (4)	12.8±1.9 (4)	9S	RH061A13	22.1±3.8 (4)	28.6±1.6 (2)
3L	RH079E02	50.3±5.8 (4)	55.3±5.7 (4)	9L	RH101N09	59.0±5.1 (4)	65.6±1.5 (2)
3L	RH074E07	66.0±6.1 (4)	64.8±6.9 (4)	9L	RH168F09	84.1±0.7 (4)	76.8±2.0 (2)
3L	RH055M19	92.6±3.7 (4)	94.4±1.0 (4)	9L	RH079O06	96.8±0.9 (4)	97.6±0.8 (2)
4S	RH013K09	4.9±0.4 (5)	5.6±2.7 (12)	10L	RH049J10	5.2±1.3 (5)	4.8±0.6 (5)
4S	RH162B09	14.3±1.3 (5)	9.7±4.4 (12)	10L	RH048F15	16.6±4.1 (5)	14.3±1.8 (5)
4S	RH097G17	26.6±2.7 (5)	17.9±4.7 (12)	10L	RH184D02	23.2±4.9 (5)	21.9±1.8 (5)
4L	RH095G10	63.4±4.5 (5)	72.9±4.7 (12)	10S	RH106M22	87.4±2.8 (5)	87.9±2.9 (5)
4L	RH191L17	88.7±6.3 (5)	90.9±3.3 (12)	10S	RH178K07	94.7±1.9 (5)	95.2±1.0 (5)
5L	RH081B09	4.3±0.6 (7)	4.9±1.0 (7)	11L	RH204G21	6.3±2.5 (6)	7.1±1.7 (11)
5L	RH095I08	18.0±2.3 (7)	17.9±3.1 (7)	11L	RH097I18	20.0±4.0 (6)	21.6±3.5 (11)
5L	RH076O08	33.0±2.1 (7)	31.9±5.6 (7)	11L	RH162O21	32.8±4.4 (6)	35.0±5.2 (11)
5S	RH089A21	80.0±4.7 (7)	76.4±3.4 (7)	11S	RH058F17	82.6±4.9 (6)	79.6±2.4 (11)
5S	RH044A21	89.7±5.5 (7)	86.4±2.3 (7)	11S	RH042C12	89.1±3.4 (6)	86.6±2.6 (11)
6S	RH026H24	3.3±1.0 (6)	3.0±0.7 (5)	12L	RH106P06	2.5±1.0 (8)	2.8±1.1 (10)
6S	RH069B12	12.2±2.9 (6)	11.0±3.8 (5)	12L	RH016H06	6.6±0.6 (8)	6.2±1.7 (10)
6L	RH075G13	49.6±4.5 (6)	53.4±6.8 (5)	12L	RH183I02	10.5±1.0 (8)	10.7±1.7 (10)
6L	RH194M18	67.2±4.6 (6)	74.7±9.8 (5)	12L	RH043B17	23.8±2.9 (8)	24.1±2.5 (10)
6L	RH054L23	82.2±6.9 (6)	88.1±5.3 (5)	12S	RH084C24	97.8±1.0 (8)	97.8±1.0 (10)

a) Arabic numerals, chromosome number; S and L, short and long arm of potato chromosomes. The numbering system was according to corresponding linkage groups (van Os et al. 2006).
b) Selected by Tang et al (2009) according to their genetic position in the ultradense RH genetic map (van Os et al. 2006).
c) Relative physical location was calculated as (S/T)×100, where S=the distance in µm from the FISH hybridization site to the north end of the chromosome, T=the total length of the chromosome in µm.
d) The number of measurements.



from 17.0±5.5 µm (chr 3) to 29.3±12.8 µm (chr 1) in *S. chacoense*. Short stretches of euchromatin in the pericentromere heterochromatin were always noticeable on the long arms of chromosome pairs 1, 3 and 8 in *S. commersonii,* as well as interstitial heterochromatin knobs on the long arms of chromosome 6 and chromosome 7 for both species (Figure 3). The total area of heterochromatin comprises 45.8±10 % of the genome of *S. commersonii* and 46±14 % of the genome of *S. chacoense* (Table 1). Large BAC signals were observed on the pericentromere heterochromatin of chromosome pair 1 (hybridization from BAC RH158N18) pair 4 (hybridization from BAC RH097G17), pair 6 (hybridization from BAC RH069B12) and pair 12 (from BAC RH183I02) in both species, indicating the presence of repetitive sequences in those BACs complementary to the heterochromatin in those regions, whose hybridization was insufficiently suppressed by C_{0t}-100 DNA (Figure 2). Analysis of their sequences obtained from GenBank (http://www.ncbi.nlm.nih.gov/nuccore/168804211) or PGSC (http://solanaceae.plantbiology.msu.edu /data/Solanum_tuberosum.RH.bacs.zip) using Geneious R9 (http://www.geneious.com/, Kearse et al. 2012) revealed high content of repeats in these BACs. Distal heterochromatin blocks in both species were clearly distinguishable in all bivalents, showing brighter fluorescence and in most

Figure 3: Idiograms of the twelve pachytene bivalents of *S. tuberosum* (RH89-039-16), *S. commersonii* and *S. chacoense* hybridized with five BAC probes per linkage group in multicolour BAC-FISH experiments. The sizes and distances correspond to averages across cell complements measured and the bars represent the standard error for those measurements. The *S. tuberosum* idiogram was built based on the data presented by Tang et al. (2009). Scale bar = 10 µm.

cases two separate dots, one for each of the homologs. The fluorescence was brighter in the distal end of short arms.

Comparative cytogenetic mapping

All BACs in this cross-species FISH showed clear signals in both species at the selected stringency level (see examples in Figures 1 and 2). We first checked that all BACs from one linkage group in potato hybridized to the same chromosome, meaning that each chromosome pair from *S. commersonii* or *S. chacoense* corresponded to one linkage group in potato and was equivalent to one of its chromosome pairs. We were also able to demonstrate that the order of BAC positions in both wild species corresponds to the genetic markers of the potato map (Table 2 and Figure 3). Chromosome pair 1 was found to harbour the 5S rDNA site on the short arm pericentromeric heterochromatin and chromosome pair 2 harbours the 45S rDNA cluster on the large satellite and NOR region in the short arm. Some BACs hybridized on the boundaries between euchromatin and pericentromere heterochromatin, such as BAC RH092C10 on chromosome 1 of *S. commersonii* and BAC RH089A21 on chromosome 5 of *S. chacoense* (Figure 3), while BAC RH043B17 mapped

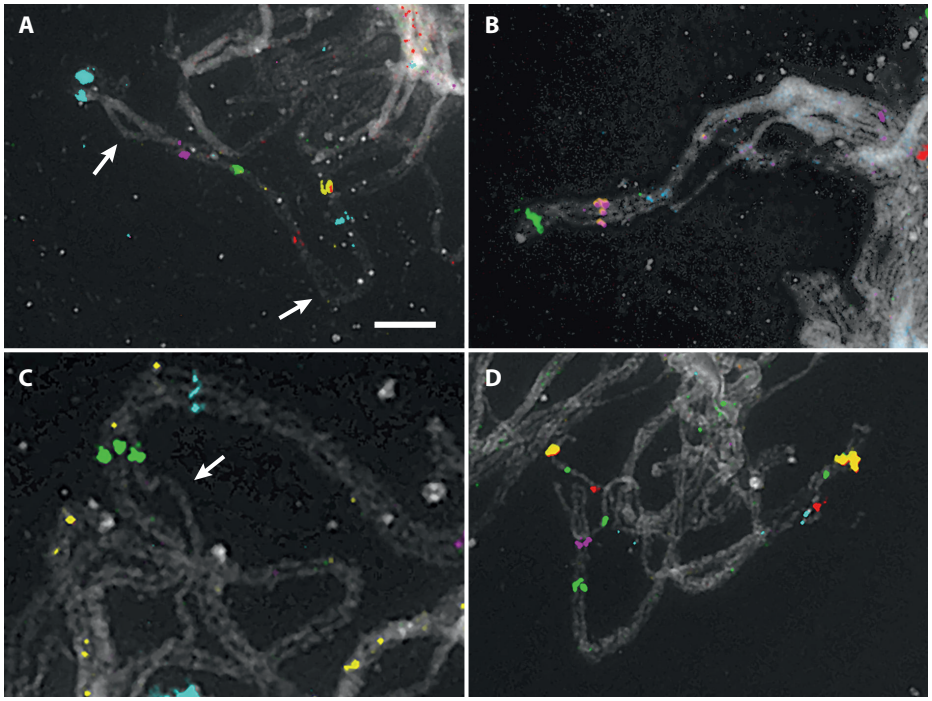


Figure 4: Examples of pachytene complements from two triploid interspecific hybrids (*S. commersonii* (2n) x *S. tuberosum* Group Phureja (n)) hybridized with five BAC probes belonging to A) chromosome 2 (chr 2), B) chr 5, C) chr 10, D) chr 11. The images show breakpoints in homoeologous pairing (white arrows) which have been quantified and are not correlated with any specific physical location. Scale bar = 10 µm.

on the boundary on the long arm of chromosome 12 for both species (see examples on Figures 1 and 2). BAC RH194M18 was located on the interstitial heterochromatic knob on the long arm of chromosome 6 in both species (Figure 3). It is worth noting that there is a clear double hybridization signal on chromosome pair 4 in *S. chacoense* – the blue signal from BAC RH097G17 is accompanied by a yellow signal from the distal BAC RH013K09 in all bivalents (Figure 2), pointing to the only rearrangement that may be observed in this species. The latter signal was not observed for *S. commersonii*.

In the pachytene complements of the two full-sib triploid CCT F_1 hybrids (*Solanum commersonii* x *S. tuberosum* Group Phureja) we observed that each BAC produced the expected hybridization signal on the same position of the *S. commersonii* homologues and potato homoeologues (Figure 4), although sometimes homoeologous pairing was broken and the two foci corresponding to the signals on the two homologs from *S. commersonii* were observed separately from the single focus on the *S. tuberosum* chromosome, assuming that it is the homoeologous chromosome that is failing to pair. More frequently, homoeologous pairing was discontinuous in between signals from contiguous BACs, although these pairing discontinuities were found all over the chromosomes (data not shown).

Physical distances and condensation patterns

Large variation was found in chromosome lengths across bivalents measured from different late pachytene complements (examples in Figure 2 and standard error bars in Figure 3). This variation was found both within and among individuals (from a variation of 11.2 % of the total chromosome length in chr 10 to 22.3 % in chr 9 in *Solanum commersonii* and from 11.7 % of the total chromosome length in chr 3 to 34.1 % in chr 8 in *S. chacoense*, see Table 1). Variation was also found in the relative position of the different markers (from a lower variation of 0.4 % in the position of BAC RH013K09 on chr 4 to a higher variation of 8.3 % in the position of BAC RH122E19 on chr 8 in *S. commersonii* and from a lower variation of 0.6 % in the position of BAC RH049J10 on chr 10 to a higher variation of ± 16.8 % in the position of BAC RH092C10 on chr 1 in *S. chacoense*, see Table 2). Even though we expressed physical distances between BAC signals as relative distances, there were only slight differences among the species (Figure 3).

Discussion

Chromosome morphology is similar between potato and two diploid wild relatives

The overall chromosome morphology of both *S. commersonii* and *S. chacoense* wild species in late pachytene cells was found to be highly similar to potato. Observed lengths varied considerably across observations in both species (see Table 1 and Figure 2). The extent of length variation of each chromosome in our study was similar to that reported

in other studies working with late pachytene chromosomes in potato (Yeh and Peloquin 1965; Ramanna and Wagenvoort 1976; Iovene et al. 2008; Tang et al. 2009). The ordering of chromosomes according to average lengths in both species differed from that reported for potato (Tang et al. 2009), albeit the fact that the longest and shortest chromosomes in *S. commersonii* corresponded to the same chromosome pairs as in potato (chr 1 and chr 11, respectively). In *S. chacoense*, however, the longest and shortest on average corresponded to chr 6 and chr 10, respectively (Table 1). Besides differences in condensation among replicates and studies, we expect that dissimilarities in total chromosome or chromosome arm length among closely related species result from changes in specific repetitive sequence distribution and abundance (Kubis et al. 1998; Sharma and Raina 2005).

The positions of the centromeres were very similar to those of potato (Yeh and Peloquin 1965; Ramanna and Wagenvoort 1976; Iovene et al. 2008; Tang et al. 2009) and the tuber-bearing *Solanum clarum* (Marks 1969). The characteristic pericentromere heterochromatin morphology in *S. commersonii* and *S. chacoense* (Figure 2 and Table 1) follows a similar pattern to that of cultivated potato (Tang et al. 2009), *S. clarum* (Marks 1969) and tomato (Chang et al. 2008). This brightly fluorescing pericentromere heterochromatin has been shown to match the distribution of hybridization with labelled C_{ot}-1000 in cultivated potato (Tang et al. 2014) and labelled C_{ot}-100 in tomato (Chang et al. 2008; Szinay 2010). Our measurements, however, indicate that there may be more heterochromatin in most chromosome pairs of both wild species (except for pairs 5 and 8), which is also seen in a higher relative heterochromatin percentage (Table 1) than in potato (Tang et al. 2009). Such differences in heterochromatin may evolve rapidly among related species and likely reflect large-scale variation in pericentromere-specific LTR retrotransposons (Rokka et al. 1998; Tek et al. 2005; Wang et al. 2006; Hemleben et al. 2007; Chang et al. 2008; Kejnovsky et al. 2012, He et al. 2013; Tang et al. 2014). Differences in the abundance of transposable elements causing intraspecific lack of collinearity in the pericentromere region, have also been directly observed for the heterochromatin haplotypes of RH (Group Tuberosum) and DM (Group Phureja) potato chromosome 5 in sequence based studies (de Boer et al. 2015).

Pericentromere heterochromatin blocks were sometimes interrupted by short stretches of euchromatin, as seen in the chromosomes 1, 3 and 8 for *S. commersonii.* Such uncondensed regions were previously described in potato (Yeh and Peloquin 1965; Ramanna and Wagenvoort 1976), *S. clarum* (Marks 1969), and in chromosome 6 of tomato, which may have resulted from an inversion event between tomato and potato involving pericentromere heterochromatin and euchromatin (Iovene et al. 2008). A second, albeit different heterochromatin pattern was the long arm interstitial knob observed in chromosome 6 of both *S. commersonii* and *S. chacoense*. Comparable knobs in potato were mentioned by Ramanna and Wagenvoort (1976) and Iovene et al. (2008), but were absent in the genotypes described by Yeh and Peloquin (1965) or Tang et al. (2008), suggesting intraspecific variation between the cultivars. We also found such a knob in the long

arm of chromosome 7 for both species, which so far was only reported for tuber-bearing *S. clarum* (Marks 1969). The nature of these interstitial knobs has been thoroughly described in tomato chromosome 7 as tandem arrays of the TGRI satellite repeat (Szinay 2010). Comparing the results of a BAC FISH and sequencing study, it was proposed that such interstitial islands of satellite repeats originated from other short arm telomere associated repeat arrays through transposition of extrachromosomal circular DNA during stages in which chromosomes are in the Rabl orientation. (Szinay 2010).

The distal ends of the *S. commersonii* and *S. chacoense* chromosomes follow the typical morphology of compact telomere heterochromatin blocks as described for potato (Yeh and Peloquin 1965; Ramanna and Wagenvoort 1976; Tang et al. 2009), and are in general larger and more conspicuous at the short arms of the chromosomes. Sequence analysis of their repeats revealed the REP3 repeat, which is a member of the PGR1 family and 99% identical to the CL14 subtelomere repeat (Torres et al. 2011) and seems specific to species of the potato clade (Tang et al. 2014). FISH of this REP3 probe on pachytene spreads from *S. commersonii* and the 3x hybrids demonstrated fluorescent foci on the distal ends of all short arms and a few of the long arms (data not shown), thus confirming the pattern in potato as described by Tang et al. (2014).

The BACs in our study were selected for their position in repeat-poor euchromatin regions in tomato and potato (Tang et al. 2009), yet some of them produced strong signals on the pericentromere heterochromatin in FISH experiments (Figure 1A, Figure 2). It is likely that *gypsy*, *copia* and other LTR retrotransposable elements are highly dispersed among stretches of single copy sequences. Such a repeat signal can be blocked in a FISH with excessive C_{ot}-100 but also results in overall hybridization and hence smaller / weaker fluorescent foci.

Solanum commersonii and *S. chacoense* are highly collinear with potato

Our study revealed a clear correspondence between the chromosomes of *Solanum commersonii* and *S. chacoense* to those in cultivated potato. Each chromosome pair from both wild species matches the corresponding chromosomes and linkage groups in potato and all BACs were in the same linear order (Figure 3). The distinct single foci with very strong signals obtained with most of the BAC probes in our cross-species BAC-FISH experiments, showed that genome differentiation among the three species at chromosomal scale is negligible if not absent at all. This is in line with the notion that diploid species in section Petota share a common genome A with only cryptic structural differences among species (Matsubayashi 1991). However, when comparing potato to species which are more distantly related such as tomato, eggplant or *S. etuberosum*, rearrangements are more prevalent (Tang et al. 2008; Iovene et al. 2008; Wu and Tanksley 2010; Lou et al. 2010; Szinay et al. 2012). In the tomato clade genome differentiation seems to work differently, since many inversions have been described between cultivated tomato and its close relatives (Seah et al. 2004; Van Der Knaap et al. 2004; Anderson et al. 2010; Verlaan

et al. 2011; Szinay et al. 2012; Aflitos et al. 2014) and shifts in centromere positions of homeologous bivalents have been observed in spread synaptonemal complexes of interspecific hybrids between tomato and wild relatives (Anderson et al. 2010).

Detection of smaller differences between homeologues are more difficult to demonstrate unequivocally. The small dissimilarities in neighbour BAC distances and the occurrence of a duplicated signal on chromosome pair 4 in *S. chacoense* might represent minor rearrangements or may point at local repeat dynamics across species in the euchromatic regions, where certain kinds of repetitive elements are more frequently located (Kidwell and Lisch 2001; Lamb et al. 2007; Kejnovsky et al. 2012 and references therein). Both explanations are congruent with the minor breaks in homeologous pairing observed in the triploid hybrids with potato. Fine-scale analysis of these regions requires a denser set of BAC probes (Tang et al. 2008; Iovene et al. 2008; Szinay et al. 2008; Achenbach et al. 2010; Verlaan et al. 2011), the use of smaller probes (Han et al. 2015; Aliyeva-Schnorr et al. 2016), advanced FISH with enhanced resolution and sensitivity (Kirov et al. 2015; Romanov et al. 2015; Khrustaleva et al. 2016) or comparative genomics including Mummer analyses (e.g., Peters et al. 2012).

Implications for breeding and genome evolution in *Solanum* section Petota

We have established cytogenetic maps for *Solanum commersonii* and *S. chacoense* using a reference set of five BACs per chromosome pair that had been selected and mapped on cultivated potato (Tang et al. 2009). No major rearrangements have been found that would be expected to hamper recombination and introgression or that could cause linkage drag or missegregation of homoeologues during meiosis in hybrids. One possible exception may be the duplicated signal on chromosome 4 in *S. chacoense*. However, in general terms, our results support the interpretation that structural differences at the large-scale chromosomal level do not seem to underlie genome evolution in species belonging to section Petota, while they appear to be important across sections in the genus (Szinay et al. 2012) and within the tomato clade.

The correspondence found here supports the application of genetic tools like the UHD genetic map of potato to analyse these wild species. The cytogenetic maps and description of the chromosome morphology developed here are valuable tools to assist in improvement of the genome sequences of *Solanum commersonii* and *S. chacoense*, thanks to the identification of euchromatin and heterochromatin and other landmarks at the cytological level and the establishment of the relative positions and distances of BACs with known sequence.

The nature and extent of genome differentiation has not been an important barrier for hybridization among tuber-bearing *Solanum* species (Camadro et al 2004) making the vast potato secondary and tertiary genepool readily accessible for breeding. The possible effects of the minor differentiation between *S. commersonii* and *S. chacoense* suggested by our results will become evident when high-density genetic

maps and genomic sequences become available for comparison in the near future.

Acknowledgements

Research was supported by: the Agencia Nacional de Investigación e Innovación (ANII) (grant code FMV_1_2011_ 1_6697); the Instituto Nacional de Investigación Agropecuaria (INIA) (grant Línea 4); and by the Comisión Sectorial de Investigación Científica, University of the Republic, Uruguay (grant code CSIC I+D 2012 383). P Gaiero was supported by grant Proyecto 720 and CSIC Recursos Humanos, University of the Republic. We are thankful to Dóra Szinay and Xianwen Ji for very useful suggestions during the development of the experiments. We are grateful to Alicia Castillo and technical staff at INIA for kindly providing plant material.

References

Achenbach, U.C., Tang, X., Ballvora, A., Jong, H. de and Gebhardt, C. 2010. Comparison of the chromosome maps around a resistance hot spot on chromosome 5 of potato and tomato using BAC-FISH painting. Genome, 53(2): 103–110. http://dx.doi.org/10.1139/G09-086. PMID: 20140028.

Aflitos, S., Schijlen, E., de Jong, H.,De Ridder, D., Smit, S., Finkers, R., et al. 2014. Exploring genetic variation in the tomato (*Solanum* section Lycopersicon) clade by whole-genome sequencing. Plant J. 80(1): 136–148. http://dx.doi.org/10.1111/tpj.12616. PMID: 25039268. [Accessed April 8, 2015].

Aflitos, S., Sanchez-Perez, G., Ridder, D. de, Fransz, P., Schranz, E., de Jong, H. and Peters, S. 2015. Introgression browser: High-throughput whole-genome SNP visualization. Plant J., 82(1): 174–182. http://dx.doi.org/10.1111/tpj.12800. PMID: 25704554. [Accessed June 5, 2015].

Aliyeva-Schnorr, L., Stein, N. and Houben, A. 2016. Collinearity of homoeologous group 3 chromosomes in the genus *Hordeum* and *Secale cereale* as revealed by 3H-derived FISH analysis. Chromosome Res. 24(2): 231-242. Available at: http://link.springer.com/10.1007/s10577-016-9518-8. [Accessed March 4, 2016].

Anderson, L.K., Covey, P.A., Larsen, L.R., Bedinger, P. and Stack, S.M. 2010. Structural differences in chromosomes distinguish species in the tomato clade. Cytogenet. Genome Res. 129(1-3): 24–34. http://dx.doi.org/10.1159/000313850.PMID: 20551606. [Accessed March 11, 2013].

Aversano, R., Contaldi, F., Ercolano, M.R., Grosso V., Iorizzo M., Tatino F. et al. 2015. The *Solanum commersonii* Genome Sequence Provides Insights into Adaptation to Stress Conditions and Genome Evolution of Wild Potato Relatives. Plant Cell, 27(5): 954–968. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25873387. [Accessed April 18, 2015].

Bennett, M.D., Smith, J.B. 1991. Nuclear DNA amounts in angiosperms. Philos. Trans. R. Soc. London Series B-Biological Sciences 334: 309-345. Available at: http://rstb.royalsocietypublishing.org/content/334/1271/309. [Accessed March 3, 2016].

Birch, P.R.J., Bryan, G., Fenton, B., Gilroy, E. M., Hein, I., Jones, J.T., et al. 2012. Crops that feed the world 8: Potato: are the trends of increased global production sustainable? Food Security

4(4): 477-508. Available at: http://www.springerlink.com/index/10.1007/s12571-012-0220-1. [Accessed March 4, 2013].

Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., et al. 2014. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nat. Genet. 46(9): 1034–1038. Available at: http://www.nature.com/doifinder/10.1038/ng.3046. [Accessed July 28, 2014]. PMID: 25064008.

Borm, T.J.A. 2008. Construction and use of a physical map of potato. Ph.D. thesis, Laboratory of Plant Breeding, Wageningeng UR, Wageningen,The Netherlands.

Camadro, E., Carputo, D., and Peloquin, S. 2004. Substitutes for genome differentiation in tuber-bearing *Solanum*: interspecific pollen-pistil incompatibility, nuclear-cytoplasmic male sterility, and endosperm. Theor Appl Genet 109 (7): 1369–76. doi: 10.1007/s00122-004-1753-2.

Campell, B.R., Song, Y., Posch, T.E., Cullis, C.A. and Town, C.D. 1992. Sequence and organization of 5S ribosomal RNA-encoding genes of Arabidopsis thaliana. Gene, 112 (2): 225-228. doi:10.1016/0378-1119(92)90380-8

Carputo, D., Barone, A., Cardi, T., Sebastiano, A., Frusciante, L. and Peloquin, S.J. 1997. Endosperm balance number manipulation for direct in vivo germplasm introgression to potato from a sexually isolated relative (*Solanum commersonii* Dun.). Proc. Natl. Acad. Sci. U. S. A. 94(22): 12013–7. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=23688&tool=pmcentrez&rendertype=abstract. PMID: 11038589

Carputo, D., Basile, B., Cardi, T. and Frusciante, L. 2000. *Erwinia* resistance in backcross progenies of *Solanum tuberosum* x *S. tarijense* and *S. tuberosum* (+) *S. commersonii* hybrids. Potato Res. 43(2): 135–142.doi: 10.1007/BF02357954.

Chang, S.-B., Yang, T.-J., Datema, E., van Vugt, J., Vosman, B., Kuipers, A., et al. 2008. FISH mapping and molecular organization of the major repetitive sequences of tomato. Chromosome Res. 16(7): 919–933. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18688733. [Accessed March 4, 2013].

Chen, L., Guo, X., Xie, C., He, L., Cai, X., Tian, L., et al.. 2013. Nuclear and cytoplasmic genome components of Solanum tuberosum + S. chacoense somatic hybrids and three SSR alleles related to bacterial wilt resistance. Theor. Appl. Genet. 126(7): 1861–1872. Available at: http://link.springer.com/article/10.1007/s00122-013-2098-5. [Accessed June 29, 2015]. PMID: 23580089.

Chen, Y.-K.H., Palta, J.P. and Bamberg, J.B. 1999. Freezing tolerance and tuber production in selfed and backcross progenies derived from somatic hybrids between *Solanum tuberosum* L. and *S. commersonii* Dun. Theor. Appl. Genet. 99(1): 100–107. Available at: http://link.springer.com/10.1007/s001220051213. [Accessed June 26, 2015].

de Boer, J.M., Datema, E., Tang, X., Borm, T.J., Bakker, E.H., van Eck, H.J., ,et al. 2015. Homologues of potato chromosome 5 show variable collinearity in the euchromatin, but dramatic absence of sequence similarity in the pericentromeric heterochromatin. BMC Genomics, 16(1): 374. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25958312. [Accessed May 17, 2015]. PMID: 25958312.

Dong, F., Song, J., Naess, S.K., Helgeson, J.P., Gebhardt, C. and Jiang, J. 2000. Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. Theor. Appl. Genet. 101(7): 1001–1007. Available at: http://link.springer.com/10.1007/s001220051573. [Accessed February 25, 2013].

Gerlach, W.L. and Bedbrook J.R. 1979. Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acid Res. 7(7): 1869–1885. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC342353. [Accessed May 10, 2016]. PMID: 342353

González, M. 2010. Análisis de la resistencia a *Ralstonia solanacearum* en una progenie segregante de *Solanum commersonii*. MSc Thesis, Faculty of Agronomy, University of the Republic, Montevideo, Uruguay.

González, M., Galván, G., Siri, M., Borges, A. and Vilaró, F. 2013. Resistencia a la marchitez bacteriana de la papa en *Solanum commersonii*. Agrociencia Uruguay 17(1):45–54. Available at http://www.fagro.edu.uy/agrociencia/index.php/directorio/article/viewFile/755/604. [Accessed May 28, 2016].

Han, Y., Zhang, T., Thammapichai, P., Weng, Y. and Jiang, J. 2015. Chromosome-Specific Painting in *Cucumis* Species Using Bulked Oligonucleotides. Genetics, 200(3): 771–779. Available at: http://www.genetics.org/content/200/3/771.full. PMID: 25971668.

Hawkes, J. 1990. The potato: evolution, biodiversity and genetic resources. Belhaven, London, UK

Haynes, K.G. and Lu, W. 2005. Improvement at the diploid species level. In Genetic improvement of Solanaceous Crops Volume I: Potato. Edited by M. Razdan and A. Mattoo, Science Publishers Inc, Enfield, NH, pp. 101–114.

He, L., Liu, J., Torres, G. a, Zhang, H., Jiang, J. and Xie, C. 2013. Interstitial telomeric repeats are enriched in the centromeres of chromosomes in *Solanum* species. Chromosome Res. 21(1): 5–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23250588. [Accessed July 28, 2014]. PMID: 23250588.

Hemleben, V., Kovarik, A., Torres-Ruiz, R., Volkov, R. and Beridze, T. 2007 Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. Syst. Biodivers. 5(3) 277–289. Available at: http://www.tandfonline.com/doi/pdf/10.1017/S147720000700240X. [Accessed April 1, 2016].

Iovene, M., Wielgus, S.M., Simon, P.W., Buell, C.R. and Jiang, J. 2008. Chromatin structure and physical mapping of chromosome 6 of potato and comparative analyses with tomato. Genetics, 180(3): 1307–1317. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2581936&tool=pmcentrez&rendertype=abstract. [Accessed March 18, 2013]. PMID: 18791232.

Jansky, S.H. 2000. Breeding for Disease resistance in potato. In Plant Breeding Reviews: Volume 19. Edited by J. Janick, John Wiley & Sons, Inc., Oxford, UK. pp. 69–152. doi: 10.1002/9780470650172.ch4.

Jansky, S.H. 2006. Overcoming hybridization barriers in potato. Plant Breed. 125(1): 1–12. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1439-0523.2006.01178.x/epdf. [Accessed August 14, 2014].

Jansky, S.H., Dempewolf, H., Camadro, E.L., Simon, R., Zimnoch-Guzowska, E., Bisognin, D. a. and Bonierbale, M. 2013. A Case for Crop Wild Relative Preservation and Use in Potato. Crop Sci. 53(3): 746-754. Available at: https://www.crops.org/publications/cs/abstracts/53/3/746. [Accessed June 29, 2015].

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647-1649. Available at: http://bioinformatics.oxfordjournals.org/content/28/12/1647.short. [Accessed April 1, 2016].

Kejnovsky, E., Hawkins, J.S. and Feschotte, C. 2012. Plant Transposable Elements: Biology and Evolution. In Plant Genome Diversity Volume 1. Edited by J.F. Wendel, Springer-Verlag, Vienna. pp. 17–34. Available at: http://link.springer.com/10.1007/978-3-7091-1130-7. [Accessed March 21, 2016].

Khrustaleva, L., Jiang, J. and Havey, M.J. 2016. High-resolution tyramide-FISH mapping of markers tightly linked to the male-fertility restoration (*Ms*) locus of onion. Theor. Appl. Genet. 129(3): 535–545. Available at: http://link.springer.com/10.1007/s00122-015-2646-2. [Accessed March 4, 2016]. PMID: 26704420.

Kidwell, M.G. and Lisch, D.R. 2001. Perspective: Transposable Elements, Parasitic DNA, and Genome Evolution. Evolution, 55(1): 1–24. Available at: http://doi.wiley.com/10.1111/j.0014-3820.2001.tb01268.x. [Accessed April 27, 2016]. PMID: 11263730.

Kirov, I.V., Laere, K. Van and Khrustaleva, L.I. 2015. High resolution physical mapping of single gene fragments on pachytene chromosome 4 and 7 of *Rosa*. BMC Genet. 16(1): 74. Available at: http://www.biomedcentral.com/1471-2156/16/74. [Accessed March 14, 2016]. PMID: 26134672..

Kocsis, E., Trus, B., Steer, C., Bisher, M. and Steven, A. 1991. Image averaging of flexible fibrous macromolecules: the clathrin triskelion has an elastic proximal segment. J. Struct. Biol. 107(1): 6–14. Available at: http://www.sciencedirect.com/science/article/pii/104784779190025R. [Accessed December 12, 2015]. PMID: 1817611.

Kubis, S., Schmidtj, T. and Pat, J.S. 1998. Repetitive DNA Elements as a Major Component of Plant Genomes. Ann. Bot. 82 (1): 45–55. Available at: http://aob.oxfordjournals.org/content/82/suppl_1/45. [Accessed July 1, 2013].

Laferriere, L.T., Helgeson, J.P. and Allen, C. 1999. Fertile *Solanum tuberosum* + *S. commersonii* somatic hybrids as sources of resistance to bacterial wilt caused by *Ralstonia solanacearum*. TAG Theor. Appl. Genet. 98(8): 1272–1278. Available at: http://link.springer.com/10.1007/s001220051193. [Accessed March 1, 2016].

Lamb, J.C., Meyer, J.M., Corcoran, B., Kato, A., Han, F. and Birchler, J.A. 2007. Distinct chromosomal distributions of highly repetitive sequences in maize. Chromosom. Res. 15(1): 33–49. Available at: http://link.springer.com/article/10.1007%2Fs10577-006-1102-1. [Accessed April 8, 2016]. PMID: 17295125.

Levan, A., Fredga, K. and Sandberg, A. 1964. Nomenclature for centromeric position on chromosomes. Hereditas, 52(2): 201–220. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1601-5223.1964.tb01953.x/epdf. [Accessed March 21, 2016].

Lou, Q., Iovene, M., Spooner, D.M., Buell, C.R. and Jiang, J. 2010. Evolution of chromosome 6 of *Solanum* species revealed by comparative fluorescence in situ hybridization mapping. Chromosoma, 119(4): 435–442. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20352244 [Accessed May 30, 2013]. PMID: 20352244.

Marks, G.E. 1969. The pachytene chromosomes of *Solanum clarum*, Caryologia 22(2): 161-167. doi: 10.1080/00087114.1969.10796334

Matsubayashi, M. 1991. Phylogenetic relationships in the potato and its related species. In Chromosome Engineering in Plants: Genetics, Breeding, Evolution, Part B. Edited by T. Tsuchiya and P.K. Gupta, Elsevier, pp. 93–118.

Micheletto, S., Boland, R. and ·Huarte, M. 2000. Argentinian wild diploid *Solanum* species as sources of quantitative late blight resistance. Theor Appl Genet 101(5):902–906. Available at: http://link.springer.com/article/10.1007/s001220051560 [Accessed June 30, 2016].

Peters, S.A., Bargsten, J.W., Szinay, D., van de Belt, J., Visser, R.G.F., Bai, Y. and de Jong, H. 2012. Structural homology in the Solanaceae: analysis of genomic regions in support of synteny studies in tomato, potato and pepper. Plant J. 71(4): 602–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22463056 [Accessed March 11, 2013]. PMID: 22463056.

Ramanna, M. and Wagenvoort, M. 1976. Identification of the trisomic series in diploid *Solanum tuberosum* L., Group Tuberosum. I. Chromosome identification. Euphytica, 25(1): 233–240. doi: 10.1007/BF00038929.

Rokka, V.M., Clark, M.S., Knudson, D.L., Pehu, E. and Lapitan, N.L. 1998. Cytological and molecular characterization of repetitive DNA sequences of *Solanum brevidens* and *Solanum tuberosum*. Genome, 41(4): 487–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9796097. [Accessed March 21, 2016]. PMID: 9796097.

Romanov, D., Divashuk, M., Havey, M.J., Khrustaleva, L. and Jenkins, G. 2015. Tyramide-FISH mapping of single genes for development of an integrated recombination and cytogenetic map of chromosome 5 of *Allium cepa*. Genome, 58(3): 111–119. Available at: http://www.nrcresearchpress.com/doi/10.1139/gen-2015-0019 [Accessed March 4, 2016]. PMID: 26158384.

Seah, S., Yaghoobi, J., Rossi, M., Gleason, C.A. and Williamson, V.M. 2004. The nematode-resistance gene, *Mi*-1, is associated with an inverted chromosomal segment in susceptible compared to resistant tomato. Theor. Appl. Genet. 108(8): 1635–1642. Available at: http://link.springer.com/article/10.1007%2Fs00122-004-1594-z. [Accessed February 21, 2016]PMID: 14963654.

Sharma, S. and Raina, S.N. 2005. Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. Cytogenet. Genome Res. 109(1-3): 15–26. Available at: http://www.karger.com/Article/Abstract/82377. [Accessed June 14, 2013]. PMID: 15753554.

Sharma, S.K., Bolser, D., de Boer, J., Sønderkær, M., Amoros, W., Carboni, M.F.., et al. 2013. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3: Genes, Genomes, Genet. 3(11): 2031–2047. Available at: http://g3journal.org/content/3/11/2031. [Accessed July 24, 2014]. PMID: 24062527.

Shearer, L.A., Anderson, L.K., de Jong, H., Smit, S., Goicoechea, J.L., Roe, B.A., et al. 2014. Fluorescence *in situ* hybridization and optical mapping to correct scaffold arrangement in the tomato genome. G3: Genes, Genomes, Genet. 4(8): 1395–1405. Available at: http://g3journal.org/content/4/8/1395. [Accessed July 10, 2014]. PMID: 24879607

Simko, I., Jansky, S., Stephenson, S. and Spooner, D.M. 2009. Genetics of resistance to pests and disease. In Potato Biology and and Biotechnology: Advances and Perspectives. Edited by D. Vreugdenhil, Elsevier, Amsterdam, pp. 117–155. Available at: http://www.sciencedirect.com/science/article/pii/B978044451018150049X. [Accessed April 26, 2013].

Solomon-Blackburn, R.M. and Barker, H. 2001. Breeding virus resistant potatoes (*Solanum tuberosum*): a review of traditional and molecular approaches. Heredity, 86(1): 17–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11298812. [Accessed April 26, 2013]. PMID: 11298812.

Szinay, D. 2010. The development of FISH tools for genetic, phylogenetic and breeding studies in tomato (*Solanum lycopersicum*). Ph.D. thesis, Laboratory of Genetics, Wageningeng UR, Wageningen,The Netherlands.

Szinay, D., Chang, S.-B., Khrustaleva, L., Peters, S.A., Schijlen, E., Bai, Y., et al. 2008. High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. Plant J. 56(4): 627–637. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2008.03626.x/epdf. [Accessed July 1, 2013]. PMID: 18643986.

Szinay, D., Bai, Y., Visser, R. and de Jong, H. 2010. FISH applications for genomics and plant breeding strategies in tomato and other solanaceous crops. Cytogenet. Genome Res. 129(1-3): 199–210. Available at: http://www.karger.com/Article/Abstract/313502. [Accessed June 29, 2015]. PMID: 20628252.

Szinay, D., Wijnker, E., van den Berg, R., Visser, R.G.F., de Jong, H. and Bai, Y. 2012. Chromosome evolution in *Solanum* traced by cross-species BAC-FISH. New Phytol. 195(3): 688–698. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2012.04195.x/epdf. [Accessed April 26, 2013]. PMID: 22686400.

Tang, X., Szinay, D., Lang, C., Ramanna, M.S., van der Vossen, E.A.G., Datema, E., et al. 2008. Cross-species bacterial artificial chromosome-fluorescence *in situ* hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics, 180(3) 1319–1328. Available at: http://www.genetics.org/content/180/3/1319. [Accessed February 21, 2016]. PMID: 18791231.

Tang, X., de Boer, J.M., Van Eck, H.J., Bachem, C., Visser, R.G.F. and de Jong, H. 2009. Assignment of genetic linkage maps to diploid *Solanum tuberosum* pachytene chromosomes by BAC-FISH technology. Chromosom. Res. 17(7): 899–915. Available at: http://link.springer.com/article/10.1007%2Fs10577-009-9077-3. [Accessed April 4, 2013]. PMID: 19774472.

Tang, X., Datema, E., Guzman, M.O., de Boer, J.M., van Eck, H.J., Bachem, C.W.B., et al. 2014. Chromosomal organizations of major repeat families on potato (*Solanum tuberosum*) and further exploring in its sequenced genome. Mol. Genet. Genomics 289(6): 1307–1319. Available at: http://link.springer.com/article/10.1007%2Fs00438-014-0891-8. [Accessed March 31, 2016]. PMID: 25106953.

Tek, A.L., Song, J., Macas, J. and Jiang, J. 2005. *Sobo*, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. Genetics, 170(3): 1231–8. Available at: http://www.genetics.org/content/170/3/1231. [Accessed August 14, 2014]. PMID: 15911575.

The Potato Genome Sequencing Consortium. 2011. Genome sequence and analysis of the tuber crop potato. Nature, 475(7355): 189–95. Available at: http://www.nature.com/nature/journal/v475/n7355/full/nature10158.html. [Accessed March 21, 2016]. PMID: 21743474

The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature, 485(7400): 635–641. Available at: http://www.nature.com/nature/journal/v485/n7400/full/nature11119.html. [Accessed April 26, 2013]. PMID: 22660326.

Torres, G.A., Gong, Z., Iovene, M., Hirsch, C.D., Buell, C.R., Bryan, G.J., et al. 2011. Organization and evolution of subtelomeric satellite repeats in the potato genome. G3: Genes, Genomes, Genet. 1(2): 85–92. Available at: http://g3journal.org/content/1/2/85. [Accessed July 16, 2014]. PMID: 22384321.

Van Der Knaap, E., Sanyal, A., Jackson, S.A. and Tanksley, S.D. 2004. High-resolution fine mapping and fluorescence *in situ* hybridization analysis of *sun*, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics, 168(4): 2127–2140. Available at: http://www.genetics.org/content/168/4/2127. [Accessed February 21, 2016]. PMID: 15611181.

van Os, H., Andrzejewski, S., Bakker, E., Barrena, I.,. Bryan G.J,, Caromel, B., et al. 2006. Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics, 173(2): 1075–

1087. Available at: http://www.genetics.org/content/173/2/1075 [Accessed March 26, 2013]. PMID: 16582432.

Verlaan, M.G., Szinay, D., Hutton, S.F., de Jong, H., Kormelink, R., Visser, R.G.F., et al. 2011. Chromosomal rearrangements between tomato and *Solanum chilense* hamper mapping and breeding of the TYLCV resistance gene *Ty*-1. Plant J. 68(6): 1093–1103. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2011.04762.x/epdf. [Accessed March 21, 2016]. PMID: 21883550.

Wang, Y., Tang, X., Cheng, Z., Mueller, L., Giovannoni, J. and Tanksley, S.D. 2006. Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome. Genetics, 172(4): 2529–2540. Available at: http://www.genetics.org/content/172/4/2529. [Accessed April 28, 2013]. PMID: 16489216.

Wu, F. and Tanksley, S.D. 2010. Chromosomal evolution in the plant family Solanaceae. BMC Genomics, 11(1): 182. Available at: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-182. [Accessed May 3, 2013]. PMID: 20236516.

Yeh, B. and Peloquin, S. 1965. Pachytene Chromosomes of the Potato (*Solanum tuberosum*, Group Andigena). Am. J. Bot. 52(10): 1014–1020. Available at; http://www.jstor.org/stable/2440131. [Accessed April 1, 2016].

Young, N.D. and Tanksley, S.D. 1989. RFLP analysis of the size of chromosomal segments retained around the *Tm*-2 locus of tomato during backcross breeding. Theor Appl Genet 77(3): 353-359. doi="10.1007/BF00305828.

Zhong, X.B., de Jong, H. and Zabel, P. 1996. Preparation of tomato meiotic pachytene and mitotic metaphase chromosomes suitable for fluorescence *in situ* hybridization (FISH). Chromosom. Res. 4(1): 24–28. Available at: http://link.springer.com/article/10.1007%2FBF02254940. [Accessed June 7, 2013]. PMID: 8653264.

Every hour of every day, I'm learning more
The more I learn, the less I know about before
The less I know, the more I want to look around
Digging deep for clues on higher ground.

UB40
(In: Higher ground, Promises and Lies, 1993)

Chapter 5

Comparative analysis of repetitive sequences among species from the potato and the tomato clades

Authors:
Paola Gaiero[1,2], Magdalena Vaio[2], Sander A. Peters[3], M. Eric Schranz[4], Hans de Jong[1], Pablo R. Speranza[2]

1. Laboratory of Genetics, Wageningen University & Research (WUR), Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
2. Laboratorio de Evolución y Domesticación de las Plantas, Facultad de Agronomía, Universidad de la República, Garzón 780, 12900 Montevideo, Uruguay
3. Applied Bioinformatics, Department of Bioscience, Wageningen University & Research (WUR), Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
4. Biosystematics Group, Wageningen University & Research (WUR), Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

PG and MV contributed equally to this work

Published in:
Annals of Botany (in press) https://doi.org/10.1093/aob/mcy186

Abstract:

- *Background and aims*: The genus *Solanum* includes important vegetable crops and their wild relatives. Introgressing their useful traits into elite cultivars requires effective recombination between hom(e)ologues which is partially determined by genome sequence differentiation. In this study we compared the repetitive genome fractions of wild and cultivated species of the potato and tomato clades in a phylogenetic context.
- *Method*: Genome skimming followed by a clustering approach was used as implemented in the RepeatExplorer pipeline. Repeat classes were annotated and the sequences of their main domains compared.
- *Key results*: Repeat abundance and genome size were correlated and the larger genomes of species in the tomato clade contained a higher proportion of unclassified elements. Families and lineages of repetitive elements were largely conserved between the clades, but their relative proportions differed. The most abundant repeats are *Ty*3/*Gypsy* elements. Striking differences in abundance were found in the highly dynamic *Ty*3/*Gypsy* Chromoviruses and *Ty*1/*Copia* Tork elements. Within the potato clade, early branching *S. cardiophyllum* showed a divergent repeat profile. There were also contrasts between cultivated and wild potatoes, mostly due to satellite amplification in the cultivated species. Interspersed repeat profiles were very similar among potatoes. The repeat profile of *S. etuberosum* was more similar to that of the potato clade.
- *Conclusions*: The repeat profiles in *Solanum* seem to be very similar despite genome differentiation at the level of collinearity. Removal of transposable elements by unequal recombination may have been responsible for structural rearrangements across the tomato clade. Sequence variability in the tomato clade is congruent with clade-specific amplification of repeats after its divergence from *S. etuberosum* and potatoes. The low differentiation among potato and its wild relatives at the interspersed repeats level may explain the difficulty in discriminating their genomes by GISH techniques.

Introduction

The genus *Solanum* includes various important vegetable crops, such as tomato and potato, and wild relatives that contain useful traits for introgression into elite crop cultivars (Bradshaw et al. 2006; Hajjar and Hodgkin 2007; Grandillo et al. 2011; Ramsay and Bryan 2011; Castañeda-Álvarez et al. 2016). However, genetic resources may not be directly usable due to limited crossability caused by pre- and post-zygotic hybridization barriers (Camadro et al. 2004; Jansky 2009; Grandillo et al. 2011). Once these barriers have been overcome and a fertile hybrid progeny obtained, the next challenge is that homoeologous chromosomes pair and recombine. Even then, local loss of collinearity may cause linkage drag, where undesirable alien traits remain completely linked with the traits of interest. These difficulties are largely related to the degree of genome differentiation between the crop and its wild relative, which means that the higher the differentiation, the harder it is to introgress genes of interest from the donor to the recipient genome.

Divergence between two genomes can be explained in terms of large-scale structural differences and of nucleotide-level differences, particularly of repetitive DNA sequences.

Structural differentiation of genomes with chromosome rearrangements such as inversions or translocations, may hinder recombination and increase linkage drag or cause (semi-)sterility. In addition, rapid evolution of tandem and interspersed repetitive elements can be a major factor in reduced pairing between homoeologous chromosomes in hybrids between related species (Dvorak 1983). Various aspects of genome differentiation between related species do not necessarily go hand in hand with their phylogenetic relationship.

Phylogenetic relationships within the genus *Solanum* have long been under debate. In particular, the tomato and potato clades, which diverged 7-8 million years ago (mya), are well-defined (Rodriguez et al. 2009; Särkinen et al. 2013). The tomato clade started diversifying only 2 mya, while the potato clade did so 7 mya (Särkinen et al. 2013). *Solanum etuberosum*, which is frequently included in phylogenetic analyses of these groups, has a debated position with respect to these two clades: originally it was included within Section Petota (Hawkes 1990) but later it was included in section Etuberosum together with other non-tuber bearing species (Spooner et al. 2014), a sister clade to both the tomato and potato clades (Rodriguez et al. 2009).

Despite their relatedness, the genomes in the tomato and potato clade species have evolved in different directions. Tomato and its close relatives exhibit more macro- and micro- genomic rearrangements (Seah et al. 2004; Van Der Knaap et al. 2004; Tang et al.

Table 1. Taxa sampled including taxonomic classification, three-letter code, accession details, genome size (http://data.kew.org/cvalues) and sequence data source.

Taxonomy	Species	Code	Accession	Genome size (1C)	Sequence data source
Subgenus Potatoe Section Petota	*Solanum tuberosum* Group Phureja	*phu*	DM	831 Mbp	http://solanaceae.plantbiology.msu.edu/
Series Tuberosa	*Solanum tuberosum* Group Tuberosum	*tbr*	RH	860 Mbp	http://solanaceae.plantbiology.msu.edu/
Series Commersoniana	*Solanum commersonii*	*cmm*	04.02.3	792 Mbp*	Gaiero et al. unpublished
	Solanum chacoense	*chc*	07.01.7	617 Mbp	Gaiero et al. unpublished
Series Bulbocastana	*Solanum cardiophyllum*	*cph*		675Mbp*	Biosystematics, WUR
Section Etuberosum	*Solanum etuberosum*	*etu*		763 Mbp	Biosystematics, WUR
Section Lycopersicon	*Solanum lycopersicum*	*lyc*	Heinz 1706	1002 Mbp	www.tomatogenome.net
	Solanum pimpinellifolium	*pim*	LA1584	831 Mbp	www.tomatogenome.net
Section Arcanum	*Solanum arcanum*	*arc*	LA2157	1125 Mbp	www.tomatogenome.net
	Solanum neorickii	*neo*	LA2133	not determined	www.tomatogenome.net
Section Neolycopersicon	*Solanum pennelli*	*pen*	LA0716	1200 Mbp	www.tomatogenome.net
Section Eriopersicon	*Solanum habrochaites*	*hab*	LYC4	905 Mbp	www.tomatogenome.net
	Solanum peruvianum	*per*	LA1954	1125 Mbp	www.tomatogenome.net

*genome size determined in this study.

2008; Anderson et al. 2010; Szinay et al. 2010, 2012; Verlaan et al. 2011), whereas the potatoes and some of their wild relatives have maintained higher chromosome collinearity (Lou et al. 2010; Gaiero et al. 2016; Chapter 4). Potato species are more syntenic with species belonging to other sections in *Solanum* and other genera of the Solanaceae (Lou et al. 2010; Peters et al. 2012; Szinay et al. 2012), which suggests that the species in the tomato clade present a more derived state of genome organization. Large scale chromosomal and small scale DNA rearrangements are caused by active transposable elements (TEs), which promote chromosomal breakages and subsequent rearrangements (McClintock 1946; Bennetzen 1996, 2000; Kidwell and Holyoake 2001; Raskina et al. 2008; Belyayev 2014; Bennetzen and Wang 2014), thus contributing to genome divergence. Their repeat profiles can give information on the phylogenetic relations within and between clades.

For evolutionary studies of the repetitive fractions of the genome, two strategies can be used as proxy. One of them is the ability of Genomic *in situ* hybridization (GISH) to discriminate parental genomes in hybrids, while the other is through differences in genome size. GISH has been successfully applied to hybrids between cultivated tomato and *S. peruvianum* or *S. lycopersicoides* (Parokonny et al. 1997; Ji and Chetelat 2003; Ji et al. 2004). Among potatoes, this genome painting strategy permits the distinction of parental chromosomes in interspecific hybrids between *S. tuberosum* Group Tuberosum and non-tuber bearing potato relatives carrying the E genome (Matsubayashi 1991), like *S. brevidens* (Dong et al. 2001, 2005; Gavrilenko et al. 2002; Tek et al. 2004) and *S. etuberosum* (Dong et al. 1999; Gavrilenko et al. 2003). GISH was not able to distinguish the parental chromosomes in hybrids between potato, *Solanum tuberosum* Group Phureja and its closer A-genome tuber-bearing wild relatives such as *S. commersonii* (Gaiero et al. 2017; Chapter 3). The lack of GISH differentiation suggests that a major part of the repetitive sequences among their genomes has not differentiated enough, in spite of the estimated 7-million-year divergence in the potato clade (Särkinen et al. 2013). The second proxy for the evolution of repetitive sequences is genome size. Genome size values are on average slightly higher for species in the tomato clade than in the potato clade (Table 1), although there is considerable variation among tomato species (Grandillo et al. 2011). The processes of genome size increase and reduction can be largely explained in terms of different dynamics of expansion/removal of repetitive elements (Feschotte et al. 2002; Leitch and Leitch 2013; Belyayev 2014). These dynamics may differ among related clades or individual species giving rise to variable degrees of divergence in the repeat composition of related genomes (Novák et al. 2014; Kelly et al. 2015; Macas et al. 2015).

The processes shaping the repeat composition of related plant genomes can be inferred by conducting a detailed study of the repetitive DNA in various species within a clade and across related clades. The availability of High Thoughput Sequencing (HTS) data for tomato, potato and their wild relatives allows us to compare their repetitive fractions. There are two classes of TEs: class I or retrotransposons with an RNA intermediate and a 'copy-and-paste' mechanism and class II or DNA transposons, with DNA

as intermediate and with a 'cut-and-paste' transposition mechanism. Class I is divided into two subclasses, those flanked by long terminal repeats (LTR retrotransposons) and those without or with short terminal repeats (non-LTR retrotransposons) (Finnegan 1989). The classes and subclasses are further divided hierarchically into order, superfamily, family and subfamily (Wicker et al. 2007; Piégu et al. 2015). TEs can thus be annotated and their relative abundances in related genomes determined.

TE classification and abundances carry phylogenetic signal (Dodsworth, Chase, et al. 2015) and have successfully been used to answer phylogenetic questions in the tomato clade (Dodsworth et al. 2016). From the structural point of view, the genome of *Solanum etuberosum* shows many rearrangements with respect to both potato and tomato (Lou et al. 2010; Szinay et al. 2012), but a recent analysis shows much greater genome collinearity with the potato lineage than with the tomato lineage (ME Schranz, WUR, Wageningen, The Netherlands, unpubl. res.). We expect TE analysis to provide further evidence of the relationship of this species to the tomato and potato clades.

The aim of this study is to elucidate differentiation of major repetitive sequence classes between and among species belonging to the tomato and potato clades of the genus *Solanum* in terms of their dynamics and evolutionary processes. We compared the classes of repetitive sequences of cultivated and wild species belonging to those clades and we assessed whether the composition of this genome fraction in *Solanum etuberosum* is more similar to that found in the tomato or in the potato clade.

Materials and methods

Taxa sampled, genome size determinations, DNA isolation and sequencing

We included 13 taxa from three sections of the genus *Solanum* including seven taxa from the tomato clade (section Lycopersicon), five from the potato clade (section Petota) and *S. etuberosum* (section Etuberosum). For some of the taxa we obtained sequence data from the 150 Tomato Genome ReSequencing project , www.tomatogenome.net (Aflitos et al. 2014), or from various research groups (Table 1). Genomic DNA of *S. commersonii* and *S. chacoense* was extracted from approximately 2.5 g of fresh etiolated leaf tissue samples using the nuclei enrichment protocol described by Bernatzky and Tanksley (1986), slightly modified. Libraries were prepared using the Nextera Library Preparation Kit (Illumina) and were sequenced on an Illumina HiSeq2000 sequencer at Applied Bioinformatics, Wageningen University and Research for *S. commersonii* (100-bp paired-end reads) and on an Illumina HiSeq4000 sequencer at The Beijing Genomics Institute (BGI) for *S. chacoense* (150-bp paired-end reads). Nuclear DNA measurements were performed according to Doležel and Göhde (1995). Propidium iodide was used to stain nuclei (PI, 50 $mg.mL^{-1}$) and tomato (2C=1.96; Doležel et al. 1992) was used as an internal stand-

ard. Three DNA estimations were carried out for each plant (5000 nuclei per analysis) in three different days. Nuclear DNA content (2C value) was calculated as sample peak mean/standard peak mean x 2C DNA content of the standard (pg).

Repeat identification from sequence data

We sampled the raw sequence data using the SEQTK command (https://github.com/lh3/seqtk) with a seed of 10 to reduce the genome coverage to 0.1x for all species, and different numbers of paired-end reads sampled depending on the genome size. All sequence reads were then trimmed to the same length (75 bp) and filtered by quality with 95 % of bases equal to or above the quality cut-off value of 10. We employed the similarity-based read clustering method for reads from each species compared to themselves as described by (Novák et al. 2010) as implemented in the REPEATEXPLORER pipeline (https://repeatexplorer-elixir.cerit-sc.cz/; (Novák et al. 2013). We used the pipeline default parameters and included a database of *Solanum* repeats which was available at that moment from The Plant Repeat Database (currently out of service; http://plantrepeats.plantbiology.msu.edu/index.html). The clustering was performed using the default settings of 90% similarity over 55% of the read length. This analysis resulted in the clustering of overlapping reads, and these clusters represented different families of repetitive sequences. Reads within individual clusters were also assembled to form contigs, representing sequence variants of corresponding repeats. For the comparative analyses we performed an all-to-all similarity comparison across all species following the same approach. Each set of reads was downsampled to represent 1% of each genome (i.e., coverage of 0.01) based on 1C values (Table 1). Samples from each species were identified with the three-letter prefixes mentioned above (Table 1), and concatenated to produce datasets as input for REPEATEXPLORER (Novák et al. 2013) for graph-based clustering. From these datasets, the pipeline retrieved 5,757,692 reads.

Repeat classification

We performed basic repeat classification using a combined approach that involved similarity searches with DNA and protein databases, as implemented in the REPEATEXPLORER pipeline (Novák et al. 2013), and improved by including a custom Solanum repeats database. Clusters that were not classified in that way could be annotated by the examination of cluster graph shape and by similarity searches using blastn and blastx against public databases (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The detection of subrepeats in assembled contigs was performed by similarity dot-plot analysis using a sliding window of 100 bp and similarity cut-off of 40%. All these sources were combined and used for final manual annotation and quantification of repeats from clusters that represented at least 0.01% of the investigated genomes. Overall repeat composition was

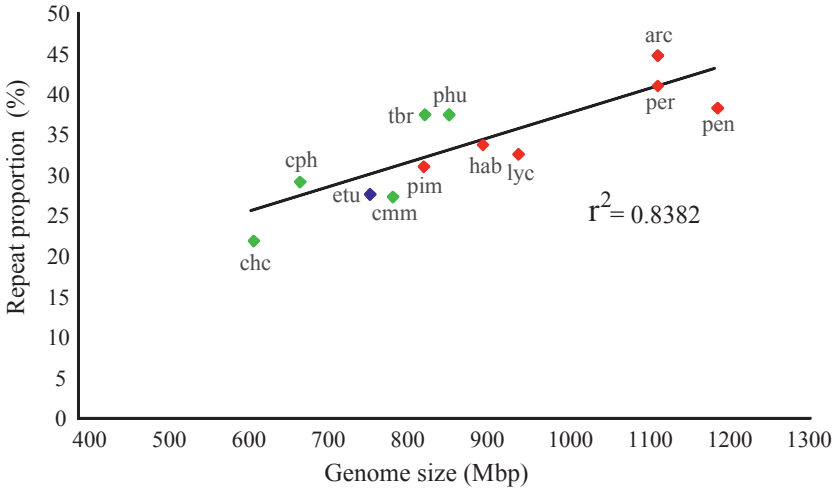


Figure 1. Correlation between repeat proportion and genome size in potato species (green dots), tomato species (red dots) and *Solanum etuberosum* (blue dot).

calculated excluding clusters of organelle DNA representing contamination of nuclear DNA preparations by chloroplast and mitochondrial DNA.

Sequence conservation across repeats

We compared the relative abundance of the largest clusters and we also investigated the graph representations of individual clusters with the SeqGrapheR programme (Novák et al. 2010) in order to identify protein domains and sequence variants derived from each species or clades as parallel paths in the graph.

RESULTS

Repeat proportion across all species

We estimated repeat proportions in the genomes of all species through comparative clustering in Repeatexplorer. Combined, the repeats identified for each species represent between 22.24 % (*S. cardiophyllum*) and 45.12 % (*S. arcanum*) of the total genome (Table 2). There is a high correlation (r^2= 0.84) between repeat proportion and genome size (Figure 1). There is also a clear grouping of potato clade species with lower genome sizes and tomato clade species with larger genomes.

Table 2. Genome proportions of different repeat classes and superfamilies in *Solanum* species. Total repeats refer to the proportion of reads clustered by RepeatExplorer. Percentages of different repeat classes and superfamilies are calculated on annotated clusters with a genome proportion higher than 0.01%.

Repeats		**Lineage/class**
LTR retroelements		
	Ty1/*Copia*	
		Maximus-SIRE
		Angela
		Tork
		Ale
		Ivana
		Bianca
		TAR
	Total Ty1/*Copia*	
	Ty3/Gypsy	
		Chromoviruses
		Ogre
		Athila
	Total Ty3/*Gypsy*	
Other	**Caulimovirus**	(Pararetrovirus)
	LINE	
	SINE	
	Helitron	
	DNA transposons	
		hAT
		CACTA
		Mutator
		Harbinger
	Total DNA transp	
	Tandem repeats	rDNA
		Satellites
	Total tandem repeats	
Annotated repetitive total		
Unclassified repetitive		
All repetitive total		

cph	*cmm*	*chc*	*phu*	*tbr*	*etu*	*neo*	*per*	*arc*	*pen*	*pim*	*hab*	*lyc*
0.118	0.135	0.123	0.087	0.072	0.106	2.127	1.913	1.819	2.310	1.872	1.967	1.993
0.069	0.124	0.139	0.067	0.068	0.108	0.170	0.180	0.177	0.215	0.232	0.210	0.225
0.075	0.128	0.123	0.131	0.152	0.087	0.576	0.530	0.535	0.544	0.547	0.627	0.605
0.048	0.019	0.014	0.005	0.010	0.005	0.009	0.008	0.007	0.015	0.011	0.007	0.008
0.185	0.247	0.248	0.175	0.231	0.190	1.153	1.261	1.188	1.004	0.983	1.077	1.104
0.177	0.152	0.172	0.186	0.209	0.080	0.156	0.137	0.158	0.167	0.149	0.141	0.155
0.014	0.017	0.010	0.011	0.010	0.021	0.115	0.092	0.090	0.101	0.125	0.097	0.107
0.201	0.274	0.250	0.114	0.181	0.244	0.600	0.588	0.512	0.646	0.686	0.688	0.708
0.647	0.871	0.772	0.526	0.609	0.676	1.253	1.166	1.166	1.044	1.273	1.264	1.229
1.416	**1.834**	**1.727**	**1.214**	**1.470**	**1.411**	**4.031**	**3.961**	**3.834**	**3.736**	**4.007**	**4.112**	**4.141**
3.627	4.536	4.414	5.146	5.412	3.748	3.151	2.948	2.760	3.731	2.733	3.111	2.900
8.392	13.29	13.45	16.22	17.02	14.94	9.558	16.34	12.48	10.70	10.09	10.61	11.39
0.386	0.505	0.523	0.918	0.840	0.214	0.513	0.522	0.506	0.415	0.456	0.540	0.504
1.260	2.922	3.403	4.412	4.418	1.980	4.746	3.946	4.846	4.099	3.397	4.794	3.868
13.67	**21.26**	**21.79**	**26.70**	**27.45**	**20.88**	**17.97**	**23.76**	**20.59**	**18.94**	**16.67**	**19.06**	**18.66**
1.072	1.121	1.017	0.170	0.190	1.991	0.723	0.484	0.576	0.794	0.727	0.529	0.514
0.168	0.552	0.458	0.888	1.130	0.033	0.496	0.469	0.394	0.577	0.382	0.440	0.432
0.016	0.027	0.020	0.051	0.030	0.016	0.013	0.010	0.011	0.008	0.014	0.012	0.011
<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.021	0.024	0.019	0.031	0.015	0.025	0.017
0.353	0.390	0.370	0.180	0.160	0.306	0.252	0.265	0.266	0.336	0.393	0.317	0.308
0.200	0.237	0.238	0.107	0.109	0.051	0.121	0.130	0.119	0.144	0.175	0.159	0.143
0.134	0.121	0.103	0.098	0.106	0.055	0.091	0.089	0.074	0.109	0.099	0.088	0.092
0.039	0.030	0.032	0.012	0.009	0.022	0.093	0.091	0.090	0.135	0.135	0.106	0.131
0.022	0.029	0.022	0.013	0.016	0.016	0.036	0.043	0.036	0.029	0.052	0.022	0.040
0.747	**0.807**	**0.765**	**0.411**	**0.399**	**0.450**	**0.593**	**0.618**	**0.585**	**0.753**	**0.854**	**0.691**	**0.714**
3.342	0.602	1.837	2.712	2.705	0.892	3.891	1.982	3.090	2.448	1.864	2.476	2.458
1.255	1.056	1.143	5.104	3.741	1.928	5.406	4.864	8.804	5.682	2.959	3.601	2.801
4.597	**1.658**	**2.980**	**7.816**	**6.446**	**2.820**	**9.927**	**6.846**	**11.894**	**8.130**	**4.823**	**6.077**	**5.259**
21.80	**27.40**	**28.88**	**37.33**	**37.19**	**27.71**	**35.27**	**38.08**	**39.72**	**35.28**	**29.37**	**32.91**	**31.74**
0.443	0.591	0.638	0.491	0.626	0.514	3.500	3.291	5.404	3.383	2.037	1.185	1.213
22.24	**27.99**	**29.52**	**37.83**	**37.81**	**28.23**	**38.77**	**41.37**	**45.12**	**38.67**	**31.41**	**34.10**	**32.95**

Figure 2. All-to-all similarity comparison of sequence reads from *Solanum* species based on cluster composition across all 13 species included in this study. The bar plots show the size of the repetitive fraction of the genome, represented as a percentage of each genome. Different colours represent different repeat families. a. Relative abundance of interspersed repeats. b. Relative abundance of tandem repeats.

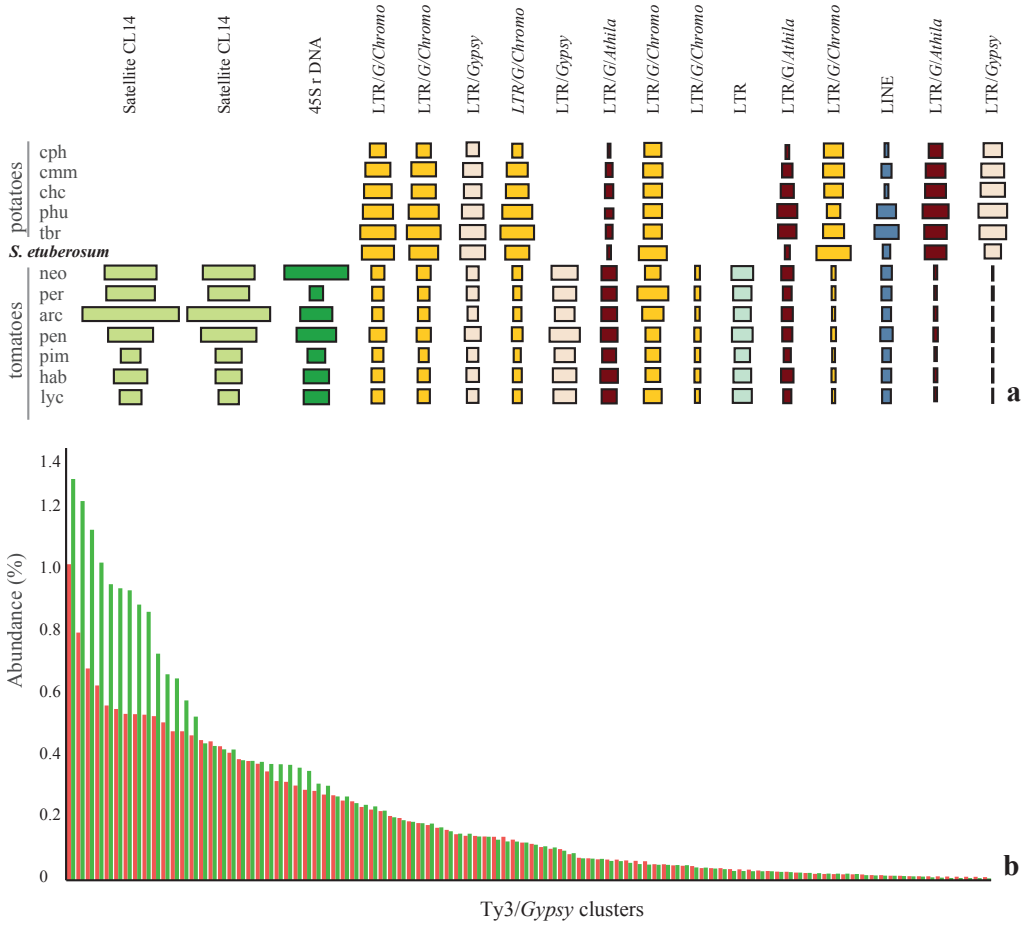


Figure 3. a. Sequence composition of the largest 17 clusters derived from the comparative analysis across all 13 species of *Solanum* included in this study. The size of the rectangle is proportional to the number of reads in a cluster for each species. Colours of the rectangles correspond to repeat type. b. Relative abundance of clusters containing LTR elements of the Ty3/*Gypsy* type, arranged from the largest to the smallest clusters in potato species (green bars) and tomato species (red bars).

Comparative analysis of major groups of interspersed repeats across and within clades

The repetitive fractions of the genomes of all species are composed mainly of LTR retrotransposons. A high proportion of these LTR elements remained unclassified in the tomato clade. Among those that we could annotate, *Ty3/Gypsy* elements were the most abundant (Figure 2a), mostly those belonging to the Chromovirus lineage (Table 2). Although these elements are highly prolific in all species, they show significant variation in abundance, with some species having as much as twice the relative content as the others such as *S. tuberosum* Group Tuberosum vs. *S. cardiophyllum* (Table 2). In the case

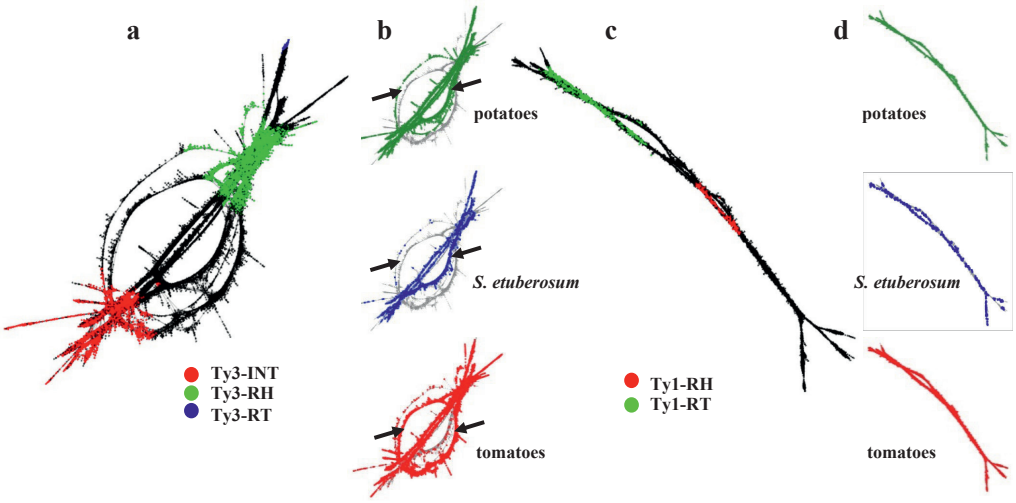


Figure 4. Repeat sequence differentiation across clades for LTR elements. a and c. Cluster graph layout (Novák et al. 2010) representation of a. *Ty3/Gypsy* Chromovirus cluster CL005 and c. of *Ty1/Copia* cluster CL025 from the comparative analysis across all species. Sequence reads are represented by nodes of the graph and reads with identity of at least 90% with minimal overlap of 55 % are connected by lines. Reads are coloured based on their similarity to conserved coding domains of LTR retrotransposons. b and d. Nodes of the graph are coloured based on their species of origin in the comparative analysis across all species. The parts of the graphs that represent the most variable sequence regions in b. the CL5 element and c. the CL25 element, which can differentiate between clades, are indicated by black arrows. These variants are evident as narrow parallel paths on the graph representation.

of tomato and its relatives, the most frequently observed were *Jinling* elements [Supplementary information – Table S1], which were almost undetectable in the potato clade.

When analysing individual sequence clusters, some of the largest represent *Ty3/Gypsy* elements, three of them belong to the Chromovirus lineage (Figure 3a). Several of the most abundant LTR elements including Chromoviruses occur in higher numbers in the potato than in the tomato clade; however several variants (clusters 8, 11 and 12) appear only in the tomato clade (Figure 3a). We plotted the relative abundance of all clusters annotated as Ty3/*Gypsy* in descending order from the largest to the smallest for potato species compared to tomato and its wild relatives (Figure 3b). We found that in potato species a higher proportion of Ty3/*Gypsy* repeats belong in a few large clusters, while in the tomato clade repeat sequences are more evenly distributed among smaller clusters (Figure 3b).

Some types of repeats are more variable among potato species, such as the Caulimovirus type of Pararetroviruses (Figure 2a). In terms of abundance Caulimoviruses represent only 0.17 and 0.19 % of the genome of cultivated *S. tuberosum* Group Phureja and Tuberosum, respectively, but their occurrence is almost 10x more prevalent in the wild potatoes (*S. cardiophyllum*, *S. commersonii* and *S. chacoense*, Figure 2a and Table 2). On the

other hand, they show comparable levels of abundance ranging from 0.50 to 0.70 % of the genome across all tomato species (Figure 2a and Table 2).

Comparative analysis of major groups of tandem repeats across and within clades

Wild potatoes have 3-6 times lower proportions of tandem repeats than the cultivated potatoes (including satellites, rDNA and telomeric repeats). This discrepancy is mostly caused by one satellite repeat that shows high homology with the satellite CL14 (Torres et al. 2011) when compared to the Solanum repeats database. This satellite is virtually absent in wild potatoes but it is conspicuously abundant in the tomato clade. The two largest clusters in our comparative analysis represent variants of the satellite element CL14 that are only present in the tomato clade (Figure 3a). Cluster 3 is a variant of the rDNA 45S tandem repeats only present in the tomato clade. Satellite St18 is far more abundant in cultivated potatoes than in the rest of the species (Figure 2b), while St3-58 has a much higher genomic proportion in tomatoes than in potatoes and notably absent in *S. etuberosum* and *S. cardiophyllum*. We also found some lineage-specific tandem repeats, such as a 334 bp satellite that is only present among tomato and its wild relatives, a 90 bp satellite that is more prevalent in wild potatoes and satellite element CL34 which is present in the potato clade except for *S. cardiophyllum* and the outgroup species *S. etuberosum* (Figure 2b and Supplementary table S1).

The relative abundances and the patterns of presence/absence of different repeat elements in the genome of S. etuberosum are more similar to those found in the potato clade. However, *S. etuberosum* does show some species-specific elements, such as two satellites with 163 bp and 260 bp repeat units representing 0.32 and 0.22 % of the total genome respectively (Figure 2b and Supplementary table S1).

Taxon-specific repeats

We identified a total of 58 clusters present in the tomato but not in the potato clade with a maximum genomic abundance of 4.2%. Among these, the single cluster classified as Helitron was only found among tomato species (Table 2). Tomato-clade specific repeats include many Chromoviruses belonging to supercluster 7 and among the Ty1/*Copia*, many Tork elements [Supplementary information – Table S1]. Twelve clusters found in the potato species could not be detected in tomato species; however, among these, the maximum genomic abundance was only 0.8%. *Solanum cardiophyllum* lacked some repeat types that were found in low abundances in other potato species. None of the species-specific repeats identified among the rest of the potato species were significantly abundant (Supplementary table S1).

Sequence divergence of the repeats across clades

We compared the sequences appearing in the tomato and potato clades and *S. etuberosum* in two of the most abundant shared clusters for which we could identify coding domains. Variants were evidenced by alternative paths in the cluster graph layouts (Novák et al. 2010). Cluster 5 (Fig. 3a) was the largest Ty3/*Gypsy* Chromovirus cluster for which we were able to identify the RT, RH and INT domains in the graph layout (Figure 4a). These domains were conserved across clades; however, we observed alternative narrow paths for the linking sequences in species belonging to the potato and tomato clades (Figure 4b). For the largest Ty1/*Copia* cluster (CL25) we identified reads coding for the RT and RH domains (Figure 4c), but no alternative clade-specific paths were observed in this case (Figure 4d). In both cases, the paths observed in *S. etuberosum* (blue dots) coincided with those of the potato species.

DISCUSSION

In this work we compared the classes, families and lineages of repetitive elements across representative wild and cultivated species of the tomato and potato clades using consistent sequence sampling strategies in order to generate equivalent datasets for each taxon. The combined dataset allowed us to interpret the different evolutionary dynamics that have shaped the present composition of the repetitive fraction of the genomes of these groups of species in the current phylogenetic context. The lack of abundant species-specific TEs among the potato species probably explains the difficulty to discriminate their genomes using genome painting techniques such as GISH (Gaiero et al. 2017; Chapter 3); however, our analysis has identified unique clusters in some tandem repeats across these clades which can be useful as cytogenetic markers.

Genome size variation and repeat content

The similarity between the genome sizes of species in the potato clade to the modal value of 600 Mbp for angiosperms (Dodsworth, Leitch, et al. 2015) was in sharp contrast with the values found in the tomato clade ranging from 905 to 1200 Mbp. The correlation between repeat content and genome size shown in Figure 1 is comparable to correlations published for other genera (Uozu et al. 1997; Piégu et al. 2006; Neumann et al. 2006; Zedek et al. 2010), tribes like Fabae (Macas et al. 2015) and across the angiosperms (Kidwell 2002; Vitte and Bennetzen 2006; Bainard and Gregory 2013; Lee and Kim 2014). The observed differences in repeat proportions indicate that the genomes in the tomato and potato clades must have reached a different balance between TE insertion and removal processes since their divergence from their common ancestor.

Tomato species contain more degraded or truncated elements (like solo LTRs) that were identified as LTRs without further classification or that remained simply unclassified. The resulting degraded repeats constitute what is sometimes called genomic 'dark

matter' and are the result of sequence removal from full length elements by ectopic recombination (Lee and Schatz 2012). For the species that had been analysed previously (*S. arcanum, S. habrochaites* and *S. pennellii*), Aflitos et al. (2014) suggested that the unique portion of their genomes is roughly the same. Our results show that different abundances of some satellite repeats and a significantly higher proportion of unclassified elements largely explain the rest of the genome size increase in the tomato species.

Interspersed repeats

The most abundant repeats in our study were the LTR type retrotransposons, particularly the Ty3/*Gypsy* elements. This higher abundance has already been reported using very different approaches for potato and tomato BAC-end sequences (Datema et al. 2008), tomato chromosome 6 (Peters et al. 2009) and the assembled genomes of tomato (The Tomato Genome Consortium 2012), *S. pennellii* (Bolger et al. 2014), potato (Potato Genome Sequencing Consortium 2011) *S. commersonii* (Aversano et al. 2015) and *S. chacoense* (Leisner et al. 2018). Tomato and potato LTRs are hypothesized to be the product of large-scale amplification events that took place about 2.8 mya (The Tomato Genome Consortium 2012), possibly as a result a large-scale epigenetic change and massive bursts of transposable element activity (Belyayev 2014).

Ty3/*Gypsy* elements were, on average, more frequent in the potato species than in the tomato species with the exception of the Jinling elements. These elements were the most abundant classified Ty3/*Gypsy* found in tomatoes. The presence and distribution of these TEs in tomatoes has already been described. *Jinling* elements are located in the pericentromere heterochromatin where they are thought to have spread five mya (Wang et al. 2006), during the radiation of the tomato clade after its divergence from the potato clade (Wang et al. 2006; Särkinen et al. 2013). The largest clusters classified as Ty3/Gypsy in potatoes were 30-50 % more prolific than in the tomato clade. In tomatoes, they were more evenly distributed across sequence clusters than in potatoes. This higher sequence divergence across Ty3/*Gypsy* elements in tomato species as a whole probably reflects different dynamics of this type of TEs in the two clades and independent amplification events of different sequence variants within each clade, as shown for Chromovirus CL5.

The Ty1/*Copia* elements were more abundant in tomato species than in potato species. Manetti et al. (2009) proposed that the *Copia* element insertion frequency, but not their abundance, may be correlated with the mating system. In the potato clade, diploid species are self-incompatible (Hawkes 1958). Within the tomato clade, although we did not find a clear relation between mating system and repeat content, selfing species like *S. lycopersicum*, *S. habrochaites* or *S. pimpinellifolium* had the lowest repeat abundances and consequently their genome sizes were the lowest among tomatoes and similar to those of potatoes.

Our study used unassembled sequences because we focused on building deliberately equivalent datasets for all the species analyzed to compare the relative abundance of

repeats. For several of these species, information is available about repetitive sequence distribution and insertion site preferences in those genomes that have been assembled and thouroughly studied cytogenetically. In potato pachytene chromosome complements, there is a large number of chromomeres in the euchromatin, while in tomato, euchromatin is relatively free of such chromomeres in most of the chromosomes (cf., Ramanna and Prakken 1967; Ramanna and Wagenvoort 1976; Wagenvoort 1988, own obervations). Chromomeres correspond to repeat-rich regions in the genome assemblies of potato (Potato Genome Sequencing Consortium 2011) and tomato (The Tomato Genome Consortium 2012). In the tomato chromosome 6, Ty1/*Copia* elements are more abundant in the gene-rich short-arm euchromatin and Ty3/*Gypsy* repeats are preferentially localized in the heterochromatin, both in the pericentromere and in small-sized chromomeres (Peters et al. 2009).

Interspersed repeats may have caused chromosomal breakages leading to structural rearrangements (Gaut et al. 2007; Belyayev 2014) in the genomes of the tomato clade. Our approach did not allow us to associate chromosome rearrangements to repeat localization; whereas large-scale changes followed by removal of repeats by unequal recombination (Gaut et al. 2007; Xu and Du 2014) in the tomato clade might have produced the large amounts of truncated and unclassified LTR elements we found. Peters et al. (2012) described such mobile elements at the synteny breakpoints with the potato and pepper genomes. Rearrangements have occurred between chromosomal fragments located in the pericentromere heterochromatin (Verlaan et al. 2011) and other repeat-rich regions (Seah et al. 2004). Lineage specific transpositional bursts and ectopic recombination might have been responsible for the chromosome rearrangements found among tomato species which have not taken place in potatoes and their wild relatives.

Tandem repeats

Tandem repeats, including satellites, occurred in the tomato clade at a higher abundance than in the potato clade (Figure 2b, Table 2). This class of repeats has been thoroughly described in both clades of *Solanum*, showing variation in location and abundance (Rokka et al. 1998; Tek and Jiang 2004; Tek et al. 2005; Chang et al. 2008; Zhu et al. 2008; Brasileiro-Vidal et al. 2009; Szinay 2010; Torres et al. 2011; Gong et al. 2012; He et al. 2013; Sharma et al. 2013; Tang et al. 2014). The patterns of occurrence are more evident when looking at the largest clusters, particularly satellite DNAs. The most abundant satellite in the tomatoes, (CL14, Torres et al. 2011), was originally described for potato and its relatives and has 99% sequence identity with the PGR1 repeat (Tang et al. 2014). In our results, the CL14 elements were much more frequent in the tomato clade and displayed a sequence variant that is not present in the potato clade. Although our analysis does not reveal major dissimilarities in the types of tandem repeats described across clades, the quantitative differences produced specific profiles for each clade consistent with the no-

tion of an ancestral "library" of satellite sequences, which were differentially amplified in each clade as proposed by (Fry and Salser 1977)

Phylogenetic context

Among potatoes, the distantly related *S. cardiophyllum* showed the most obvious divergence within the potato clade, which is coherent with its position as an early branching species in the 1EBN group. However, we found a sharp contrast between cultivated and wild potatoes. The most striking difference is the overall much higher proportion of tandem repeats in cultivated potatoes. Interspersed repeats also showed differences, with 5-6% more Ty3/*Gypsy* elements in cultivated potatoes and twice as much abundance of Caulimoviruses in wild potatoes. Amplification of a certain type of repeat can happen rapidly and even in a few marginal populations within a species (Belyayev 2014). It is possible that Caulimoviruses underwent amplification after the divergence of cultivated and wild potatoes, or that a selective bias against them (Kidwell and Lisch 2001) arose in domesticated potatoes. It remains to be tested whether domestication processes themselves underlie these differences.

The repeat profile of *S. etuberosum* was more similar to that of potato than tomato species although a few TE types show very unique patterns, particularly Caulimoviruses. The ten fold higher abundance of this type of TEs in *S. etuberosum* and possible sequence variants in other elements probably explain why GISH results have allowed discrimination of *S. tuberosum* and *S. etuberosum* chromosomes in hybrids (Dong et al. 1999; Gavrilenko et al. 2003). In terms of structural genome differentiation, *S. etuberosum* sometimes shares collinearity with potato species and sometimes with tomato species, while certain chromosome arms are entirely rearranged with respect to both clades (Lou et al. 2010; Szinay et al. 2012). Here we showed that the relative abundance and the patterns of presence/absence of repeats in *S. etuberosum* were more similar to those found in the potato than in the tomato clade. Moreover, *S. etuberosum* sequences were also more similar to those of potato species in the analysed TE clusters. Given the phylogenetic relationships among these clades, sequence similarity between TEs in potato and *S. etuberosum* is probably plesiomorphic. Tomato clade-specific sequence variants may have propagated by independent transposition after its divergence from the common ancestor of both clades and *S. etuberosum*.

Our results are congruent with the current phylogenetic hypotheses for these clades within the genus *Solanum*. At this point, one cannot establish causal relationships between the constitution of the repetitive fraction of the genome and the different paths genome evolution has taken in the tomato and potato clades. In spite of this, the patterns we observed and our current understanding support the notion that the dynamics of repetitive elements may be related to the underlying mechanisms that have driven tomato and potato genomes in different directions.

Supplementary information

Supplementary Table S1 is too big for display and can be downloaded from https://www.dropbox.com/s/4jb0i83chkcr2lo/Supplementary%20table%201.xlsx?dl=0&m=. It shows the output from the annotation of all the repeat classes and lineages across all species in the potato and tomato clade, plus *Solanum etuberosum*. The relative genome abundance for each cluster was calculated and the length of the cluster (in Mbp) was estimated where nuclear DNA content was known. All relative abundances for the same repeat type were added up for further comparisons across species and clades.

Funding

This work was supported by Comisión Sectorial de Investigación Científica, University of the Republic, Uruguay (grant code CSIC I + D 2012 383) and Instituto Nacional de Investigaciones Agropecuarias, Uruguay (grant Línea 4). P. Gaiero was supported by grant Proyecto 720 and CSIC Recursos Humanos, University of the Republic.

Acknowledgements

We are grateful to Dr. Francisco Vilaró, Dr. Alicia Castillo and technical staff at INIA for kindly providing *in vitro* plant material. We are also thankful to Bertus van der Laan and technical staff of the greenhouse facility of Wageningen University for kindly providing and taking care of the plants. We wish to thank Dr. Gustavo Folle, Federico Santiñaque and Beatriz López-Carro for their assistance in nuclear DNA content measurements. We also wish to thank Dr. Lidjia Berke, Henri van de Geest and Dr. Theo Borm for their help with sequence data transfer. We thank Dr. Saulo A. Aflitos and Henri van de Geest for patiently assisting us with bioinformatics questions.

Literature cited

Aflitos SA, Schijlen E, de Jong H, et al. 2014. Exploring genetic variation in the tomato (*Solanum* section Lycopersicon) clade by whole-genome sequencing. The Plant Journal 80: 136–148.

Anderson LK, Covey PA, Larsen LR, Bedinger P, Stack SM. 2010. Structural differences in chromosomes distinguish species in the tomato clade. Cytogenetic and genome research 129: 24–34.

Aversano R, Contaldi F, Ercolano MR, et al. 2015. The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. The Plant cell 27: 954–968.

Bainard JD, Gregory TR. 2013. Genome size evolution: patterns, mechanisms, and methodological advances. Genome 56: vii–viii.

Belyayev A. 2014. Bursts of transposable elements as an evolutionary driving force. Journal of Evolutionary Biology 27: 2573–2584.

Bennetzen JL. 1996. The contributions of retroelements to plant genome organization, function and evolution. Trends in Microbiology 4: 347–353.

Bennetzen JL. 2000. Transposable element contributions to plant gene and genome evolution. Plant Molecular Biology 42: 251–269.

Bennetzen JL, Wang H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65: 505–530.

Bernatzky R, Tanksley SD. 1986. Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112: 887–98.

Bolger A, Scossa F, Bolger ME, et al. 2014. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nature Genetics 46: 1034–1038.

Bradshaw JE, Bryan GJ, Ramsay G. 2006. Genetic Resources (Including Wild and Cultivated *Solanum* Species) and Progress in their Utilisation in Potato Breeding. Potato Research 49: 49–65.

Brasileiro-Vidal AC, Melo-Oliveira MB, Carvalheira GMG, Guerra M. 2009. Different chromatin fractions of tomato (*Solanum lycopersicum* L.) and related species. Micron (Oxford, England : 1993) 40: 851–9.

Camadro EL, Carputo D, Peloquin SJ. 2004. Substitutes for genome differentiation in tuber-bearing *Solanum*: interspecific pollen-pistil incompatibility, nuclear-cytoplasmic male sterility, and endosperm. TAG. Theoretical and applied genetics 109: 1369–76.

Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, et al. 2016. Global conservation priorities for crop wild relatives. Nature Plants 2: 16022.

Chang S-B, Yang T-J, Datema E, et al. 2008. FISH mapping and molecular organization of the major repetitive sequences of tomato. Chromosome Research 16: 919–933.

Datema E, Mueller LA, Buels R, et al. 2008. Comparative BAC end sequence analysis of tomato and potato reveals overrepresentation of specific gene families in potato. BMC Plant biology 8: 34.

Dodsworth S, Chase MW, Kelly LJ, et al. 2015. Genomic repeat abundances contain phylogenetic signal. Systematic Biology 64: 112–126.

Dodsworth S, Chase MW, Särkinen T, Knapp S, Leitch AR. 2016. Using genomic repeats for phylogenomics: a case study in wild tomatoes (*Solanum* section Lycopersicon : Solanaceae). Biological Journal of the Linnean Society 117: 96–105.

Dodsworth S, Leitch AR, Leitch IJ. 2015. Genome size diversity in angiosperms and its influence on gene space. Current Opinion in Genetics and Development 35: 73–78.

Doležel J, Göhde W. 1995. Sex determination in dioecious plants *Melandrium album* and *M. rubrum* using high-resolution flow cytometry. Cytometry 19: 103–106.

Doležel J, Sgorbati S, Lucretti S. 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. Physiologia Plantarum 85: 625–631.

Dong F, McGrath JM, Helgeson JP, Jiang J. 2001. The genetic identity of alien chromosomes in potato breeding lines revealed by sequential GISH and FISH analyses using chromosome-specific cytogenetic DNA markers. Genome 44: 729–734.

Dong F, Novy RG, Helgeson JP, Jiang J. 1999. Cytological characterization of potato - *Solanum etuberosum* somatic hybrids and their backcross progenies by genomic *in situ* hybridization. Genome 42: 987–992.

Dong F, Tek AL, Frasca ABL, et al. 2005. Development and characterization of potato-*Solanum brevidens* chromosomal addition/substitution lines. Cytogenetic and Genome Research 109: 368–372.

Dvorak J. 1983. Evidence of genetic suppression of heterogenetic chromosome pairing in polyploid species of *Solanum*, sect. Petota. Canadian Journal of Genetics and Cytology 25: 530–539.

Feschotte C, Jiang N, Wessler SR. 2002. Plant Transposable Elements: Where Genetics Meets Genomics. Nature Reviews Genetics 3: 329–341.

Finnegan DJ. 1989. Eukaryotic transposable elements and genome evolution. Trends in Genetics 5: 103–107.

Fry K, Salser W. 1977. Nucleotide sequences of HS-α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. Cell 12: 1069–1084.

Gaiero P, van de Belt J, Vilaró F, Schranz ME, Speranza P, de Jong H. 2016. Collinearity between potato (*Solanum tuberosum* L.) and wild relatives assessed by comparative cytogenetic mapping. Genome 60: 228–240.

Gaiero P, Mazzella C, Vilaró F, Speranza P, de Jong H. 2017. Pairing analysis and *in situ* Hybridisation reveal autopolyploid-like behaviour in *Solanum commersonii* × *S. tuberosum* (potato) interspecific hybrids. Euphytica 213: 137.

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. 2007. Recombination: an underappreciated factor in the evolution of plant genomes. Nature Reviews Genetics 8: 77–84.

Gavrilenko T, Larkka J, Pehu E, Rokka V-M. 2002. Identification of mitotic chromosomes of tuberous and non-tuberous *Solanum* species (*Solanum tuberosum* and *Solanum brevidens*) by GISH in their interspecific hybrids. Genome 45: 442–449.

Gavrilenko T, Thieme R, Heimbach U, Thieme T. 2003. Somatic hybrids of *Solanum etuberosum* (+) dihaploid *Solanum tuberosum* and their backcrossing progenies: relationships of genome dosage with tuber development. Euphytica 131: 323–332.

Gong Z, Wu Y, Koblízková A, et al. 2012. Repeatless and repeat-based centromeres in potato: implications for centromere evolution. The Plant cell 24: 3559–74.

Grandillo S, Chetelat RT, Knapp S, et al. 2011. *Solanum* section Lycopersicon In: Kole C, ed. Wild Crop Relatives: Genomic and Breeding Resources: Temperate Fruits. Springer-Verlag Berlin Heidelberg, 129–215.

Hajjar R, Hodgkin T. 2007. The use of wild relatives in crop improvement: A survey of developments over the last 20 years. Euphytica 156: 1–13.

Hawkes JG. 1958. Significance of wild species and primitive forms for potato breeding. Euphytica 7: 257–270.

Hawkes JG. 1990. The potato: evolution, biodiversity and genetic resources. Belhaven, London, UK.

He L, Liu J, Torres GA, Zhang H, Jiang J, Xie C. 2013. Interstitial telomeric repeats are enriched in the centromeres of chromosomes in *Solanum* species. Chromosome Research 21: 5–13.

Jansky SH. 2009. Breeding, Genetics, and Cultivar Development In: Singh J, Kaur L, eds. Advances in Potato Chemistry and Technology. Elsevier Ltd., 27–62.

Ji Y, Chetelat RT. 2003. Homoeologous pairing and recombination in *Solanum lycopersicoides* monosomic addition and substitution lines of tomato. TAG. Theoretical and applied genetics. 106: 979–89.

Ji Y, Pertuzé R, Chetelat RT. 2004. Genome differentiation by GISH in interspecific and intergeneric hybrids of tomato and related nightshades. Chromosome Research 12: 107–16.

Kelly LJ, Renny-Byfield S, Pellicer J, et al. 2015. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytologist 208: 596–607.

Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115: 49–63.

Kidwell MG, Holyoake AJ. 2001. Transposon-induced hotspots for genomic instability. Genome research 11: 1321–2.

Kidwell MG, Lisch DR. 2001. Transposable Elements, Parasitic DNA, and Genome Evolution. Evolution 55: 1–24.

Van Der Knaap E, Sanyal A, Jackson SA, Tanksley SD. 2004. High-resolution fine mapping and fluorescence *in situ* hybridization analysis of *sun*, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics 168: 2127–2140.

Lee S-I, Kim N-S. 2014. Transposable elements and genome size variations in plants. Genomics & informatics 12: 87–97.

Lee H, Schatz MC. 2012. Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score. Bioinformatics 28: 2097–2105.

Leisner CP, Hamilton JP, Crisovan E, et al. 2018. Genome sequence of M6, a diploid inbred clone of the high glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. The Plant Journal. 96 (2): 482. doi: 10.1111/tpj.13857

Leitch IJ, Leitch AR. 2013. Genome size diversity and evolution in land plants In: Plant Genome Diversity Volume 2. Vienna: Springer Vienna, 307–322.

Lou Q, Iovene M, Spooner DM, Buell CR, Jiang J. 2010. Evolution of chromosome 6 of *Solanum species* revealed by comparative fluorescence in situ hybridization mapping. Chromosoma 119: 435–442.

Macas J, Novak P, Pellicer J, et al. 2015. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. PLoS ONE 10: 1–23.

Manetti ME, Rossi M, Nakabashi M, Grandbastien MA, Van Sluys MA. 2009. The *Tnt1* family member Retrosol copy number and structure disclose retrotransposon diversification in different *Solanum* species. Molecular Genetics and Genomics 281: 261–271.

Matsubayashi M. 1991. Phylogenetic relationships in the potato and its related species In: Tsuchiya T, Gupta P, eds. Chromosome Engineering in Plants: Genetics, Breeding, Evolution, Part B. Elsevier, 93–118.

McClintock B. 1946. Maize genetics In: Carnegie Institution of Washington Year Book No. 45. Washington, D.C.,: Carnegie Institution of Washington, 176–186.

Neumann P, Koblížková A, Navrátilová A, Macas J. 2006. Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. Genetics 173: 1047–1056.

Novák P, Hřibová E, Neumann P, Koblížková A, Doležel J, Macas J. 2014. Genome-wide analysis of repeat diversity across the family Musaceae. PloS one 9: e98918.

Novák P, Neumann P, Macas J. 2010. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11: 378.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29: 792–793.

Parokonny A, Marshall J, MD B, R CM, Davey EC, Power JB. 1997. Homoeologous pairing and recombination in backcross derivatives of tomato somatic hybrids (*Lycopersicon esculentum* (+) *L . peruvianum*). Theoretical and Applied Genetics 94: 713–723.

Peters SA, Bargsten JW, Szinay D, et al. 2012. Structural homology in the Solanaceae: analysis of genomic regions in support of synteny studies in tomato, potato and pepper. The Plant journal 71: 602–14.

Peters SA, Datema E, Szinay D, et al. 2009. *Solanum lycopersicum* cv. Heinz 1706 chromosome 6: distribution and abundance of genes and retrotransposable elements. The Plant journal 58: 857–69.

Piégu B, Bire S, Arensburger P, Bigot Y. 2015. A survey of transposable element classification systems - A call for a fundamental update to meet the challenge of their diversity and complexity. Molecular Phylogenetics and Evolution 86: 90–109.

Piégu B, Guyot R, Picault N, et al. 2006. Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome Research 16: 1262–1269.

Ramanna MS, Prakken R. 1967. Structure of and homology between pachytene and somatic metaphase chromosomes of the tomato. Genetica 38: 115–133.

Ramanna M, Wagenvoort M. 1976. Identification of the trisomic series in diploid *Solanum tuberosum* L., Group Tuberosum. I. Chromosome identification. Euphytica 25: 233–240.

Ramsay G, Bryan G. 2011. Solanum In: Kole C, ed. Wild Crop Relatives: Genomic and Breeding Resources, Vegetables. Berlin, Heidelberg: Springer Berlin Heidelberg, 259–271.

Raskina O, Barber JC, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: Speciation-related events in plant genomes. Cytogenetic and Genome Research 120: 351–357.

Rodriguez F, Wu F, Ané C, Tanksley S, Spooner DM. 2009. Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? BMC evolutionary biology 9: 191.

Rokka VM, Clark MS, Knudson DL, et al. 1998. Cytological and molecular characterization of repetitive DNA sequences of Solanum brevidens and Solanum tuberosum. Genome 41: 487–94.

Särkinen T, Bohs L, Olmstead RG, Knapp S. 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evolutionary Biology 13: 214.

Seah S, Yaghoobi J, Rossi M, Gleason CA, Williamson VM. 2004. The nematode-resistance gene, *Mi-1*, is associated with an inverted chromosomal segment in susceptible compared to resistant tomato. Theoretical and Applied Genetics 108: 1635–1642.

Sharma SK, Bolser D, de Boer J, et al. 2013. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3: Genes, Genomes, Genet. 3: 2031–47.

Spooner DM, Ghislain M, Simon R, Jansky SH, Gavrilenko T. 2014. Systematics, Diversity, Genetics, and Evolution of Wild and Cultivated Potatoes. Botanical Review 80: 283–383.

Szinay D. 2010. The development of FISH tools for genetic, phylogenetic and breeding studies in tomato (*Solanum lycopersicum*). PhD Thesis. Wageningen University.

Szinay D, Bai Y, Visser R, de Jong H. 2010. FISH applications for genomics and plant breeding strategies in tomato and other solanaceous crops. Cytogenetic and Genome Research 129: 199–210.

Szinay D, Wijnker E, van den Berg R, Visser RGF, de Jong H, Bai Y. 2012. Chromosome evolution in Solanum traced by cross-species BAC-FISH. The New Phytologist 195: 688–698.

Tang X, Datema E, Guzman MO, et al. 2014. Chromosomal organizations of major repeat families on potato (*Solanum tuberosum*) and further exploring in its sequenced genome. Molecular Genetics and Genomics 289: 1307–1319.

Tang X, Szinay D, Lang C, et al. 2008. Cross-species bacterial artificial chromosome-fluorescence *in situ* hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics 180: 1319–1328.

Tek AL, Jiang J. 2004. The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. Chromosoma 113: 77–83.
Tek AL, Song J, Macas J, Jiang J. 2005. Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. Genetics 170: 1231–1238.
Tek AL, Stevenson WR, Helgeson JP, Jiang J. 2004. Transfer of tuber soft rot and early blight resistances from *Solanum brevidens* into cultivated potato. TAG. Theoretical and applied genetics. 109: 249–254.
The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485: 635–641.
Torres GA, Gong Z, Iovene M, et al. 2011. Organization and Evolution of Subtelomeric Satellite Repeats in the Potato Genome. G3: Genes, Genomes, Genet. 1: 85–92.
Uozu S, Ikehashi H, Ohmido N, Ohtsubo H, Ohtsubo E, Fukui K. 1997. Repetitive sequences: Cause for variation in genome size and chromosome morphology in the genus *Oryza*. Plant Molecular Biology 35: 791–799.
Verlaan MG, Szinay D, Hutton SF, et al. 2011. Chromosomal rearrangements between tomato and *Solanum chilense* hamper mapping and breeding of the TYLCV resistance gene *Ty-1*. The Plant journal 68: 1093–1103.
Vitte C, Bennetzen JL. 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proceedings of the National Academy of Sciences of the United States of America 103: 17638–17643.
Wagenvoort M. 1988. Spontaneous structural rearrangements in *Solanum* L. Chromosome identification at pachytene stage. Euphytica 9: 159–167.
Wang Y, Tang X, Cheng Z, Mueller L, Giovannoni J, Tanksley SD. 2006. Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome. Genetics 172: 2529–40.
Wicker T, Sabot F, Hua-Van A, et al. 2007. A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8: 973–982.
Xu Y, Du J. 2014. Young but not relatively old retrotransposons are preferentially located in gene-rich euchromatic regions in tomato (Solanum lycopersicum) plants. Plant Journal 80: 582–591.
Xu X, Pan S, Cheng S, et al. 2011. Genome sequence and analysis of the tuber crop potato. Nature 475: 189–95.
Zedek F, Smerda J, Smarda P, Bureš P. 2010. Correlated evolution of LTR retrotransposons and genome size in the genus *Eleocharis*. BMC plant biology 10: 265.
Zhu W, Ouyang S, Iovene M, et al. 2008. Analysis of 90 Mb of the potato genome reveals conservation of gene structures and order with tomato but divergence in repetitive sequence composition. BMC genomics 9: 286

...to explore strange new worlds,
to seek out new life and new civilizations
to boldly go where no one has gone before.

Gene Roddenberry
(In: Opening monologue, Star Trek: the next generation, 1987).

Chapter 6

Comparative genomics of potato (*Solanum tuberosum* L.) and the diploid wild relatives *Solanum commersonii* and *S. chacoense*

Authors:
Paola Gaiero[1,2], Pablo A. Sandro[1,3], Lucía Gutiérrez[1,3], Chaozhi Zheng[4], Xingtan Zhang[5], Haibao Tang[5], Henri van de Geest[6], Gabino Sánchez-Pérez[6], Sander A. Peters[7], Francisco Vilaró[1,8], M. Eric Schranz[9], Pablo R. Speranza[1], Hans de Jong[2]

1. Faculty of Agronomy, University of the Republic, Montevideo, Uruguay
2. Laboratory of Genetics, Wageningen University & Research, Wageningen, The Netherlands
3. Department of Agronomy, University of Wisconsin, Madison, Wisconsin, USA.
4. Biometris, Wageningen University & Research, Wageningen, The Netherlands
5. J. Craig Venter Institute, Rockville, Maryland, USA.
6. Genetwister B.V., Wageningen, The Netherlands
7. Applied Bioinformatics, Department of Bioscience, Wageningen University & Research, Wageningen, The Netherlands
8. National Institute for Agricultural Research, Las Brujas, Canelones, Uruguay.
9. Biosystematics Group, Wageningen University & Research, Wageningen, The Netherlands

Abstract

Comparative genomics can provide important information about syteny between the genomes of crop and related species and so assist in introgressive hybridization breeding. To make significant comparisons, the genomes to be compared need to be highly contiguous and complete. In order to do that, we have improved the genome assembly of wild potato relative *Solanum commersonii*. We produced a hybrid assembly combining short Illumina reads and long PacBio reads sequenced from a haploid clone which helped us eliminate heterozygosity. We used the 3,645 scaffolds obtained to anchor on the 12 linkage groups that we produced through mapping of 1,728 SNPs from a diploid biparental population. We managed to anchor 601 scaffolds into 12 pseudomolecules, at an anchor rate of 38.65 %. Their order was well defined but the orientation of most scaffolds was random. We used this assembly to make structural comparisons against the previously reported assemblies of DM potato and M6 *S. chacoense*. Genomes were overall highly collinear but there were discontinuities around the pericentromere heterochromatin and some inverted fragments. Most of these were assembly artifacts, although there were some supported rearrangements. The structural variants and copy number variants that we identified were mostly small and mostly located in the pericentromere region, so they should not hinder introgression. However, it is necessary to identify the coding sequences present in the rearranged segments and evaluate their phenotypic results. The high genome homology and collinearity found between *S. commersonii* and cultivated potato encourages its use in introgressive breeding.

Introduction

Comparative genomics can provide significant insights into synteny between the genomes of crop and related species, and so has practical relevance to plant breeding. When crossing related species in order to introgress traits from a donor wild relative into a crop, collinearity breaks between their genomes may occur, that lead to linkage drag and/or meiotic abnormalities that produce infertile gametes in the hybrids or backcrosses. Comparative genomics can directly identify the alien chromosome regions of the donor genome that possess polymorphism of DNA for the traits of interest. With this focus in mind several crops and their wild relatives were sequenced *de novo* (Imelfort and Edwards 2009; Abbott and Butcher 2012). Important examples are potato, tomato (Potato Genome Sequencing Consortium 2011; The Tomato Genome Consortium 2012) and some of their wild relatives (Bolger et al. 2014; Aflitos et al. 2014; Aversano et al. 2015), the *Petunia hybrida* progenitors (Istvánek et al. 2014; Bombarely et al. 2016), eggplant (Byrne et al. 2015), the peanut genome A donor (Chen et al. 2016), spinach (Xu et al. 2017) and coffee (Tran et al. 2018), among others. These studies have allowed to explore their gene space and know about their functional genomics (Appels et al. 2015).

However, only a few genomes show high quality *de novo* assembling, with euchromatin parts on the chromosome arms that are closed or nearly closed (Boetzer and Leiden 2013; Jiao et al. 2017a; Udall and Dawe 2017). Important efforts have been made to improve such assemblies even more, including that of tomato (Shearer et al. 2014), potato

(Sharma et al. 2013; Hardigan et al. 2016) and their wild relatives (Paajanen et al. 2017; Schmidt et al. 2017; Leisner et al. 2018), maize (Jiao et al. 2017c), subterranean clover (Kaur et al. 2017), quinoa (Jarvis et al. 2017), alfalfa (Tang et al. 2014) and turnip (Cai et al. 2017), to name a few. The main limitations in closing the (super) contigs of such genomes are posed by the nature of their sequence data. High-throughput sequencing reads often are too short to bridge complex genomic regions such as repeats which is a severe problem for the large-sized plant genomes in the Gb-range, while other genomic sequences are hard to capture into a sequence library. More problems of plant genomes may arise from polyploidy and high levels of heterozygosity. Resolving polyploids and highly heterozygous genomes requires genome complexity reduction techniques such as developing double haploids or the sequencing of inbred lines (Zhang et al. 2014). For the subsequent anchoring of the genomes, the abundant pericentromere heterochromatin in plant chromosomes poses extra challenges due to the lack of crossover recombination in these regions, which force researchers to find additional physical information to connect contigs and scaffolds, such as genome mapping (Tang et al. 2015a; Chaney et al. 2016; Udall and Dawe 2017). Such difficulties imply that large-scale genome rearrangements between related species across centromeres remain undetectable (Zapata et al. 2016; Paajanen et al. 2017). As a consequence, only few studies appeared with structural comparisons (Chaney et al. 2016) among crop and their wild relatives or related species (Hardigan et al. 2017; Montero-Pau et al. 2018).

Structural comparisons are important to study genome evolution and are helpful to predict the success of introgressive hybridization in chromosome regions with agronomically important traits. Models of chromosomal change could be developed for structural rearrangements among tomato, potato and pepper (Peters et al. 2012) or among grass genomes (Cardone et al. 2015). The most remarkable variation within a crop genepool was found for potato and wild relatives, both at the chromosomal level through large Copy Number Variants (CNVs) that could be detected through Fluorescent *in situ* hybridization (FISH) (Iovene et al. 2013) and at the sequence level (de Boer et al. 2015; Hardigan et al. 2017). Copy number variants (CNVs) were found to be prevalent among potato cultivars, landraces and wild relatives (Hardigan et al. 2017), even among a panel of 12 related double monoploid clones (Hardigan et al. 2016). It has been proposed that this variability is partly caused by extended wild introgressions into potato cultivars, which may have assisted the spread of potato by making it more suitable for colonization of new environments through alleles conferring tolerance to new ecological factors (Hardigan et al. 2017). Moreover, some important agronomic traits are directly caused by structural variations which often have significant functional impact (Tang et al. 2015a). It is also important to assess the degree of nucleotide variation through SNP analysis. Comprehensive nucleotide variation studies have been performed for potato and wild relatives (Kloosterman et al. 2015; Hardigan et al. 2016; Hardigan et al. 2017; Pham et al. 2017; Leisner et al. 2018), revealing a remarkable degree of diversity.

Diploid wild relatives of potato of great potential for breeding have been sequenced (Aversano et al. 2015; Paajanen et al. 2017; Leisner et al. 2018). The contiguity of the genome of *Solanum verrucosum* was improved by using long-range scaffolding technologies such as Bionano genome mapping (Lam et al. 2012; Hastie et al. 2013) and Hi-C Dovetail genomics (Belton et al. 2012; Reyes-Chin-Wo et al. 2017), increasing N50 by 6-fold with Hi-C and 2-fold with Bionano and producing a 9-fold increase combining both (Paajanen et al. 2017). However, the highly contiguous *S. verrucosum* assembly was not anchored into chromosome-scale pseudomolecules (Paajanen et al. 2017). Recently, the inbred line M6 (Jansky et al. 2014) *S. chacoense* was sequenced. Pseudomolecules were built independently from the DM potato reference by anchoring on two genetic maps, including the M6 genotype as parent, allowing structural comparisons (Leisner et al. 2018). The current *S. commersonii* draft genome (Aversano et al. 2015) is too fragmented for a meaningful genome comparison with cultivated potato or other wild relatives. A combination of sequencing technologies, hybrid assembly algorithms, long-range scale scaffolding technologies and anchoring on physical and genetic maps is necessary to support a comparative genomics study (Jiao et al. 2017b; Li and Harkess 2018) that can assess collinearity at the sequence level.

At a higher level, there are chromosome-scale comparisons between potato and diploid wild relatives, which are donors of traits of interest such as *S. commersonii* and *S. chacoense*. These comparisons showed no large (chromosomal) scale rearrangements (Gaiero et al. 2016, Chapter 4). However, when analyzing pachytene spreads of 3x hybrids between *S. commersonii* and potato, small loops or pairing breaks were revealed (Gaiero et al. 2017, Chapter 3). These pairing disruptions may be caused by small structural variations between homologous chromosomes in (F_1) hybrids that can be confirmed only by sequence comparisons and if they are surrounding the genes of interest, they may impair their proper introgression into the potato background.

The aim of this study is to improve genome assembly and assess the structural rearrangements between *S. commersonii* and potato and *S. commersonii* and close relative *S. chacoense*. Additionally, we aimed to evaluate sequence variation by detecting SNPs between genotypes of *S. commersonii* and also CNVs between *S. commersonii* and related species *S. tuberosum* and *S. chacoense*.

Materials and methods

Plant material, DNA purification and sequencing

We used a clone of *Solanum commersonii* (accession 04.02.3 from Colonia, southwest Uruguay). In order to avoid the pitfalls of heterozygosity for genome assembly, we generated a haploid clone from *S. commersonii* accession 04.02.3 through anther culture as described in Castillo et al. (2016). Tissue culture plantlets were grown using axillary

buds and apical shoots in MS media on light racks set to 16h/8h day/night photoperiod at 22 °C. Genomic DNA was extracted from approximately 2.5 g of fresh etiolated leaf tissue using the nuclei enrichment protocol described by Bernatzky and Tanksley (1986), with some modifications. Libraries were prepared using the Nextera Library Preparation Kit (Illumina). The insert size was 200-500 bp and the library was sequenced on one lane in an Illumina HiSeq2000 sequencer (100-bp paired-end reads). Additionally, a large-insert library (30 Kb) was prepared using the SMRTbell Template Prep Kit (Pacific Biosciences) and 18 SMRT cells were sequenced in a PacBio RS II sequencer with P4-C2 chemistry. Sequencing was performed at Applied Bioinformatics, Wageningen University and Research.

Genome assembly

We used the pipeline recommended by DBG2OLC (Ye et al. 2016) to perform hybrid assemblies. In this pipeline, we used SPARC (Ye and Ma 2016) to perform De Bruijn graph assembly on the Illumina reads assessed for quality and trimmed using FASTQC (v0.11.5) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We used DBG2OLC to align our PacBio reads to the De Bruijn graph assembly to produce a 'backbone', then, according to the DBG2OLC standard pipeline, we used the backbones to generate a consensus using the program BLASR (Chaisson and Tesler 2012) and polished the assembly with PBDAGCON (https://github.com/PacificBiosciences/pbdagcon).The assembly quality was evaluated using the QUAST package (v4.3) (Gurevich et al. 2013).

Table 1. Metrics of libraries used for the *de novo* hybrid assembly of haploid *S. commersonii* 04.02.3 genome.

Library id.	Type of library	# input read pairs	Total bases	Coverage	Paired reads distance (nt)	average read length (nt)	average fragment size (bp)
Illumina cmm	Paired end	210,642,856	29,792,017,695	35x	298	99	300
PacBio cmm	SMRT belt	235,221	3,214,000,000	9x	n/a	13,857	30,000

Genetic mapping

A biparental mapping population with heterozygous contrasting diploid *S. commersonii* genotypes as parents was used to anchor the scaffolds to the 12 linkage groups and to construct pseudomolecules. For the linkage groups, we genotyped 190 individuals and the two parents. DNA was isolated using the Quick-DNA Plant/Seed Miniprep Kit (Zymo Research, cat. number D6020). Genotyping by sequencing (GBS) was performed following the procedure described by Elshire et al. (2011). Briefly, 100 ng (10 ng/µL) of each genomic DNA sample was added to lyophilized adapter mix. Adapter and DNA mixtures

were subsequently digested using ApeKI (New England Biolabs) for 2 hours at 75°C in a total volume of 20 µL. Digested DNA and Adapters were used in subsequent ligation by T4DNA Ligase in a 50 µL reaction volume at 22°C for one hour followed by heat inactivation at 65°C for 30 minutes. Sets of 96 digested DNA samples, each of them with a different barcode adapter, were combined (10 µL each) and purified with Qiaquick PCR Purification columns (Quiagen). Purified pooled DNA samples were eluted in a final volume of 10 µL. DNA Fragments were amplified in 50 µL volume reactions containing 2 µL pooled DNA, 25 µL KAPA HiFi HotStart Master Mix (Kapa Biosystems), and 2 µL of both PCR primers (12.5 µM). PCR cycling consisted of 98°C for 30 seconds, followed by 18 cycles of 98°C for 30 seconds, 65°C for 30 seconds, 72°C for 30 seconds with a final extension for 5 minutes and kept at 4°C. Amplified libraries were purified as above but were eluted in 30 µL. Amplified library products were size selected using 2% agarose gel cassette on a blue pippin system (Sage Science), removing fragments smaller than 300 bp. Eluted size selected libraries were then purified by AmpureXP beads (Agencourt). Final libraries were used for clustering on one Illumina Paired End flowcell using a cBot (Illumina). Paired End sequencing (2*100nts) was performed on four flowcell lanes of an Illumina HiSeq2500 instrument at Applied Bioinformatics, Wageningen University and Research (WUR). A quality control and filtering process was carried out to identify segregating markers. SNPs were filtered in Tassel 5.0 (Bradbury et al. 2007) using our hybrid assembly of *Solanum commersonii* as reference genome. From this filtering we obtained a VFC file with 100,000 SNPs. Using R (v3.3) (R Core Team 2014) a marker matrix was ordered and SNPs incoherent with the parental genotypes were labelled as missing data. SNPs were then separated by the nucleotide variant according to the parental genotypes. On the remaining 15,000 SNPs left, we performed another filtering removing all SNPs that deviated 20 % from the expected segregation ratios, and also SNPs with more than 20 % missing data. A total of 2834 SNPs for 181 genotypes (including the parents) were produced as input for linkage mapping. For the genetic mapping we used the recently developed statistical software Rabbit (Reconstructing Ancestry Blocks BIT by bit), that has been extended for genotype imputation and linkage map construction (Zheng et al 2018, Biometris, WUR, manuscript in preparation). The software is based on identity by descent and Hidden Markov modelling algorithms.

Construction of pseudomolecules

Pseudomolecule construction was done using 1,728 SNPs from the cmm x cmm biparental population that matched to scaffold positions in the *S. commersonii* hybrid genome assembly. Scaffolds were anchored, ordered and oriented using Allmaps (Tang et al. 2015c). Final chromosome-scale pseudomolecules were constructed in Allmaps using the unequal weights2 parameters for a single Allmaps run for the entire genome.

Bioinformatics analyses

Comparison of the *S. commersonii* pseudomolecules was completed by performing whole genome sequence alignment of the 12 pseudomolecules to the 12 DM1-3 chromosomes (ver. 4.04) (Hardigan et al. 2016) with the NUCMER extension of the MUMMER package (v3.23) (Kurtz et al. 2004), using default parameters, followed by DELTA FILTERING to filter out alignments smaller than 5 Kbp, 50% identity and 10% uniqueness. Dot-plots were constructed using MUMMERPLOT, with the identification and orientation of the DM pseudomolecules as reference. This analysis was repeated against the *Solanum chacoense* M6 pseudomolecules (Leisner et al. 2018). We performed the same analysis for each pseudomolecule individually. The show-coords output was further analyzed to determine match length, synteny breakpoints and scaffold identity, breakpoint sequence at synteny breaks. Sequences at the breakpoints were then compared using BLAST against a database of *Solanum* repeats (The Plant Repeat Database; currently out of service). http://plantrepeats.plantbiology.msu.edu/index.html).

The hybrid assembly anchored scaffolds were aligned to the previous *S. commersonii* assembly, DM pseudomolecules, and the M6 *S. chacoense* pseudomolecules using BWA MEM (ver. 0.7.11r1034) and the alignments filtered with SAMTOOLS (v0.1.19) (Li et al., 2009) retaining alignments that were properly paired and had a mapping quality (MAPQ) > 30. The BAM files were then sorted and duplicates marked with PICARD (v2.1.1) (Picard, 2017). Reads were realigned around insertions/deletions (InDels) using GATK INDELREALIGNER (v3.7.0) (McKenna et al., 2010). Variant calling was performed using GATK INDELREALIGNER (v3.7.0) (McKenna et al., 2010). VCFTOOLS (v0.1.12b) (Danecek et al., 2011) was used to determine SNP counts and density, which was calculated as the number of SNPs per Kbp in non-overlapping 200-Kbp windows. Coverage was calculated using BEDTOOLS and BEDTOOLS GENOMECOV (ver. 2.25.0) and visualized in bedgraph format with IGV (v2.3) (Robinson et al. 2011). Copy number variations (CNVs) evidenced by increased read coverage (duplications) or decreased coverage (deletions) were identified using CNVNATOR (v3.02) (Abyzov et al. 2011).

Table 2. Assembly metrics of the de novo hybrid assembly of haploid *S. commersonii* 04.02.3 genome.

Estimated genome size (bp)	**830,000,000**
GC content (%)	34
# scaffolds	3,645
Total size of scaffolds (bp)	671,255,838
# scaffolds > 100 kbp	1,948
# scaffolds > 1 Mbp	44
N50 scaffold length	294,347
NG50 scaffold length	199,127
Longest scaffold (bp)	3,210,436
Shortest scaffold (bp)	2,854
Total anchored scaffold length	249,407,795
Total unanchored scaffold length	395,868,123
# anchored scaffolds	601
# unanchored scaffolds	3,044

Results

With the combination of sequencing approaches mentioned in Table 1, we achieved a coverage of 35x for Illumina reads and 9x for PacBio reads, with a balance between read length of 99 bp and almost 14 Kbp on average, respectively, and a base call quality that enabled the construction of a hybrid assembly from a haploid genome. The hybrid genome assembly has an N50 of over 294 Kbp, with only 3645 scaffold of lengths ranging from 3.2 Mbp to 2.8 Kbp (Table 2). Most scaffolds were above 100 Kbp (1948 out of 3645, 53 %) and 44 scaffolds were above 1 Mbp. The assembly covered 85 % of the total estimated genome size of *S. commersonii* (792 Mbp, Gaiero et al, in press, Chapter 5). The contiguity and completeness of this hybrid assembly allowed its improvement by integration with other technologies such as anchoring on the genetic map developed in this study. Although the anchoring rate is low (38.65 %), 249.6 Mbp of sequences belonging to 601 scaffolds were anchored onto the 12 pseudomolecules.

The size and genetic distances are reported for all 12 pseudomolecules and linkage groups (Table 3). The parental maps for each linkage group were averaged. In total 1728 markers could be mapped on 12 linkage groups, with a maximum of 216 loci on LG1 and a minimum of 99 loci on LG12. LG1 was also the linkage group with most anchored scaffolds (95) while LG10 had the fewest with only 30. Again, LG1 was the longest, with 137 cM in genetic distance and forming a pseudomolecule of 34.4 Mbp in size. On the other end, LG11 was the shortest linkage group (83 cM) while LG10 formed the shortest pseudomol-

Table 3. Summary of the *S. commersonii* genetic map generated from the 05.05.2.4 × 02.04.1 biparental population and its anchoring of the hybrid genome assembly. Linkage groups (LGs) are numbered according to homology with potato LGs.

LGs	# loci	# anchored scaffolds	pseudo-molecule size (Mbp)	genetic distance (cM)	Maximum scaffold size (Mbp)	Minimum scaffold size (Mbp)
LG1	216	75	34.4	137	2.1	0.036
LG2	169	51	25.8	86	3.2	0.052
LG3	169	60	24.7	109	2.6	0.041
LG4	124	53	20.3	106	2.3	0.025
LG5	119	52	17.2	87	1.4	0.067
LG6	205	55	21.9	95	1.2	0.029
LG7	137	51	23.1	90	1.4	0.040
LG8	105	40	18.0	92	2.6	0.039
LG9	120	45	17.8	109	1.2	0.099
LG10	114	30	13.6	122	2.0	0.046
LG11	151	50	18.8	83	2.0	0.106
LG12	99	39	14.0	91	1.9	0.059
Total	1728	601	249.6	1207	2.0	0.053

ecule with 13.6 Mbp. The total genetic distance of the map was 1207 cM with an average length of anchored scaffolds ranging from 2 Mbp to 0.053 Mbp.

The SNP density on each LG allowed the anchoring with a good correspondence between the LGs and the pseudomolecules, although there are some discrepancies between genetic and physical markers (Figure 1) represented by crosses between the green lines

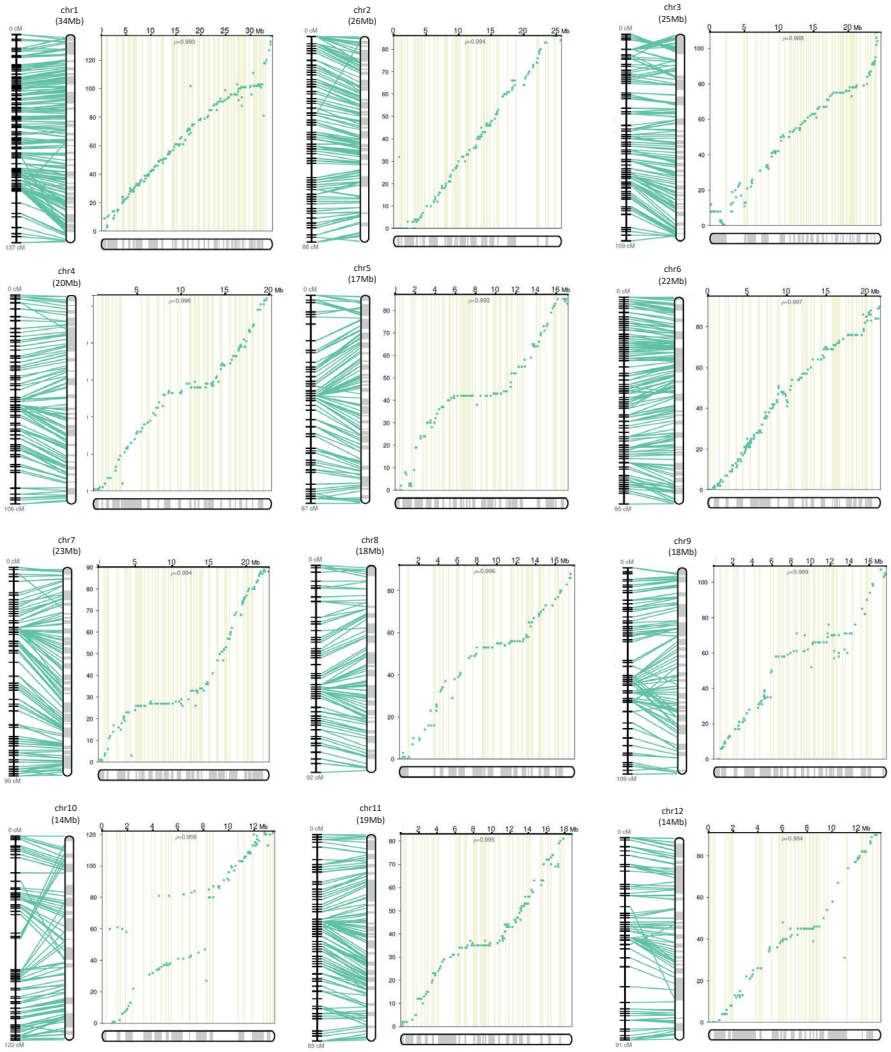
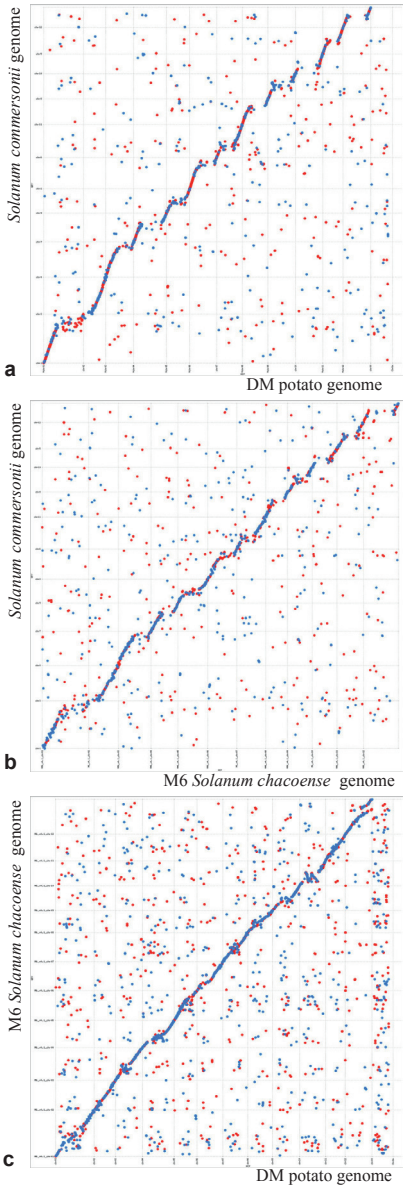


Figure 1. Anchoring of hybrid scaffolds on the linkage groups produced through genetic mapping with RABBIT (Zheng et al. 2018 in prep). The pseudomolecules were built using ALLMAPS (Tang et al. 2015). Each plot shows the matches (green lines) between the linkage group (left) and the pseudomolecule (right), and a biplot between the pseudomolecule (X axis) and the linkage group (Y axis), with green dots indicating anchored marker positions.

that link the positions of both types of markers. The LGs and pseudomolecules are presented in this figure in the order that they were obtained, but this numbering was then corrected to match the numbering of their homologous pseudomolecules in the DM potato reference genome (Hardigan et al. 2016). In this anchored assembly, the order of scaffolds is well supported. However, for most scaffolds (72 %) the orientation is assigned at random. This happens because many scaffolds are anchored with only one SNP marker or with many markers that are tightly linked. Only 28 % scaffolds had an assigned orientation, which matched the order of the map in 15 % of scaffolds and was inverted with respect to the map in 13 % of scaffolds.



A more detailed comparative analysis was obtained at pseudomolecule level. Pseudomolecules appear collinear in order but we observed specific differences in orientation. For pseudomolecule 1 as an example, we observed many scaffolds placed in a reverse order (Figure 3). Yet, it is not possible to identify true inversions and discriminate them from anchoring errors causing an aberrant scaffold orientation. More inverted scaffolds were observed comparing our assembly to the M6 assembly (Figure 3b) than comparing to the DM potato assembly (Figure 3a), suggesting accumulated assembly artifacts. Analyzing the collinearity

Figure 2. Whole genome NUCmer sequence alignment dot plots for the twelve pseudomolecules. Sequences aligned in forward and reverse orientations are represented by blue and red lines, respectively. Scaffold misplacements (or rearrangements) are shown as horizontal or vertical shifts in parts of the aligned blocks. a. Pseudomolecules of *Solanum commersonii* (plotted on y-axis) compared to DM pseudomolecules (ver4.04, plotted on x-axis). b. Pseudomolecules of *S. commersonii* (plotted on y-axis) compared to M6 *S. chacoense* (ver4.1, plotted on x-axis). c. Pseudomolecules of DM potato (ver4.04, plotted on x-axis) compared to M6 *S. chacoense* (ver4.1, plotted on y-axis). *S. commersonii* pseudomolecules were built by anchoring hybrid scaffolds on the genetic map. Genomes are plotted following the order and orientation of DM potato pseudomolecules.

within each scaffold revealed inversions. We consider these structural variations to be more supported rearrangements as the sequence order within scaffolds was confirmed by NGS hybrid assembly and genetic markers. Correspondingly, the structural variations (inversions) were marked (Figure 3) and appeared more abundant between *S. commersonii* and M6 *S. chacoense* than with DM potato. There were numerous cases of duplications, with one fragment on the diagonal and the other corresponding to fragments outside the diagonal (boxes in Figure 3), whereas small discontinuities within scaffolds suggest that sequences present in the reference are absent in the *S. commersonii* assembly, which may indicate deletions in the latter. These were more frequent when DM was used as reference. Pairwise comparisons with the three genomes across all twelve pseudomolecules show a similar scenario (Figure S1), with a clearly more broken assembly around the pericentromere in smaller pseudomolecules containing more mismatches when compared against the M6 *S. chacoense* pseudomolecules.

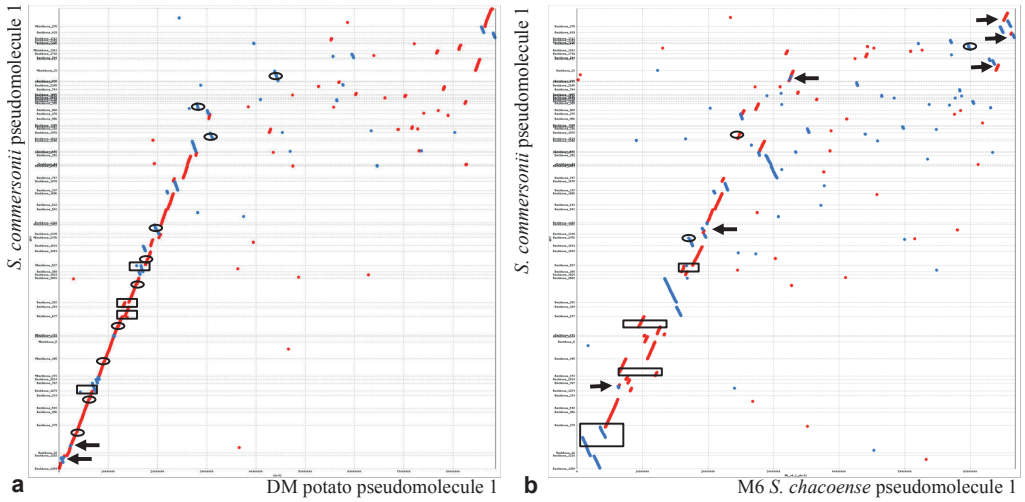


Figure 3. Mummerplots of pseudomolecule 1 from *Solanum commersonii* (cmm) against its homolog in DM potato (a) and M6 *chacoense* (b). The cmm pseudomolecule was built by anchoring hybrid scaffolds to the genetic map using at least one SNP marker per scaffold that was mapped to a linkage group. Scaffold order is coherent between pseudomolecules. Scaffold orientation in cmm is not conclusive, so inversions involving the whole scaffold are not considered. Only rearrangements within scaffolds were identified. Arrows indicate inversions, boxes indicate duplications and ovals indicate deletions.

We then examined the *S. commersonii* assembly and the reference DM potato genome in order to discriminate between assembly artifacts and ambiguities and supported synteny breaks, using pseudomolecule 2 as an example (Figure 4). First we ordered the scaffolds and produced bars of sizes proportional to the lengths obtained from the .coords file output from MUMmer. The overall size of the total matching scaffolds in *Solanum commersonii* is approximately 4x longer than in DM potato (data not shown). We repre-

sented the corresponding pseudomolecules as bars of equal total length (Figure 4a). We observed that many otherwise long scaffold alignments were broken and interrupted by smaller alignments that correspond to other scaffolds (Figure 4a, shaded bars). Also, small fragments from these long scaffolds were separated and mapped further away. We then analyzed the length distribution of those aligned scaffolds and scaffold fragments and found a noticeable gap in sequence lengths between approximately 10 and 100 Kbp (Figure 4b). When these fragment alignments of less than 10 Kbp are singled out (black dots in Figures 4b and 4c), there is an almost perfect matching between the DM reference (x-axis) and the *S. commersonii* scaffolds (y-axis). These small fragments, that distort the matching, may have resulted from assembly artifacts or ambiguous alignment. Furthermore, when these small fragments were removed, scaffold alignments were almost com-

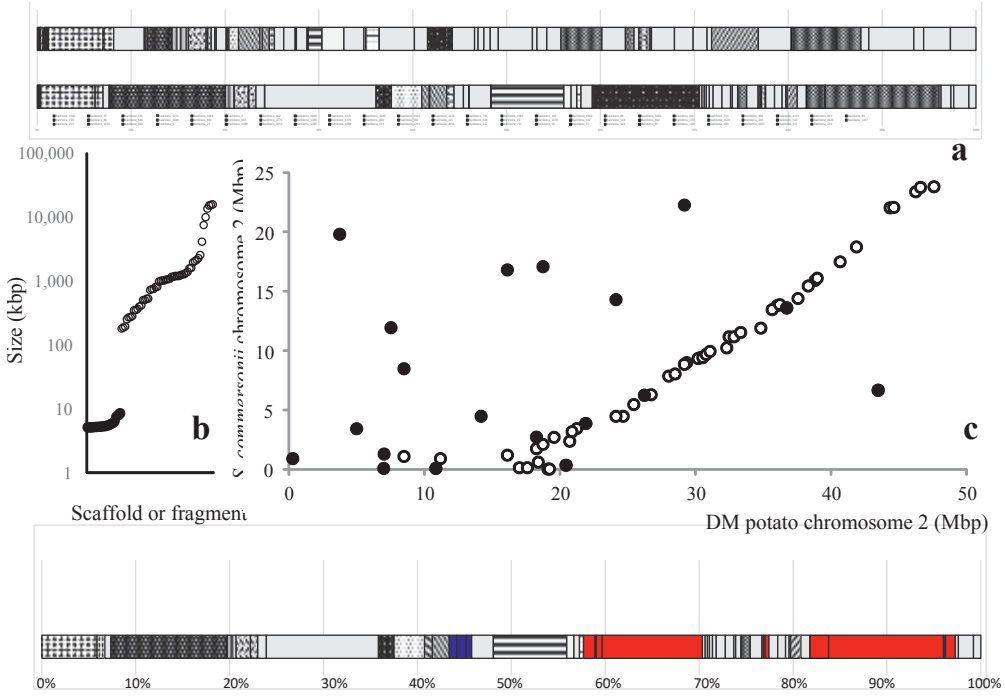


Figure 4. a. Schematic representation of all matching scaffolds from chromosome 2 in DM potato (top) and *Solanum commersonii* (bottom). Differences in size and position can be observed. Complete scaffolds are represented in solid light grey and split scaffolds are represented by shaded blocks, same pattern indicates same scaffold. b. Size distribution of the *S. commersonii* scaffolds and fragments represented in a. Scaffolds or fragments shorter than 10 kbp are shown as solid black dots. c. Dotplot of the start positions of all scaffolds from chromosome 2. The same scaffolds or scaffold fragments shorter than 10 kbp as in b are shown as black dots. d. Schematic representation of chromosome 2 in *S. commersonii*, after removing scaffold fragments that are <10 Kbp. Notice that some larger scaffolds that were split are now joined together. Confirmed inversions with respect to DM potato are shaded in blue and confirmed syntenic blocks are shaded in red.

pletely contiguous (Figure 4d). We then looked at the scaffolds with supported orientation which were either inverted (Figure 4d, blue blocks) or in the same orientation as the DM reference (Fig. 4d, red blocks). The first inverted (blue) block is formed by three consecutive scaffolds with known position, supporting its inverted state. In the BWA alignment of the hybrid assembly and the DM potato pseudomolecule 2, the corresponding order was scaffold 633, spanning a region from 20759 Kbp to 20938 Kbp, followed by scaffold 2004 running from 20943 to 21321 Kbp, and scaffold 891 mapping against the region from 21330 to 21690 Kbp of DM pseudomolecule 2 (data not shown). Inspection of the sequences of these inverted scaffolds revealed the presence of rDNA sequences in scaffold 633, which is anchored at a terminal position close to the NOR on chromosome 2. There were repeats homologous to LTR/*Jingling* repeats and to centromere-specific TGRIV and St3_58 and St3_238 satellite repeats in scaffolds 2004 and 891. For scaffold 144, which is present in the second inverted blue block and comprised within a red block of confirmed forward orientation, there were repeats homologous to St3_58 and St3_238 satellite repeats close to the point where the alignment breaks. However, we could not confirm the exact position nor link these repeats unequivocally to the synteny breakpoints in any of these cases.

Some features of the chromosome morphology can be related to the sequence coverage, anchoring or SNP density (Figure 5). When comparing the cytogenetic BAC-FISH map of *S. commersonii* chromosome 1 against its genome assembly and read coverage (Figure 5a, b and c), there seems to be higher coverage in some regions of the euchromatin and also in the pericentromere. The intraspecific SNP density distribution (comparing against the previous assembly) shows the same pattern as the read coverage (Figure 5c and d). However, this pattern is different when comparing against the DM potato or M6 *S. chacoense* pseudomolecule 1. There is a low SNP density along the whole chromosome when comparing with DM potato, with the exception of a small peak on the short arm heterochromatin (Figure 5e). The SNP density along pseudomolecule 1 is consistently lower when comparing against M6 *S. chacoense*, with only a few small peaks (Figure 5f). The distribution of SNP density along all pseudomolecules between both genome assemblies of *S. commersonii* accompanies read density (Supplementary Figure S2), while the SNP distribution patterns along pseudomolecules from DM potato (Figure S3) and M6 *S. chacoense* (Figure S4) are independent of read coverage and show a consistently low SNP density along all the pseudomolecules, with only a few peaks presumably representing high variation. The total number of SNPs and the number of SNPs per chromosome as well as the SNP rates reflect this, with a much higher SNP rate when comparing different accessions within *S. commersonii* (4.77 on average) than when comparing our genome against DM potato (1.22 on average) or M6 *S. chacoense* (average 0.47) (Table 4). The chromosome with most intraspecific SNPs is chromosome 1, which has the highest read coverage. However, the highest intraspecific SNP rate belongs to chromosome 5, which shows a moderate read coverage compared to the rest of the chromosomes. In the

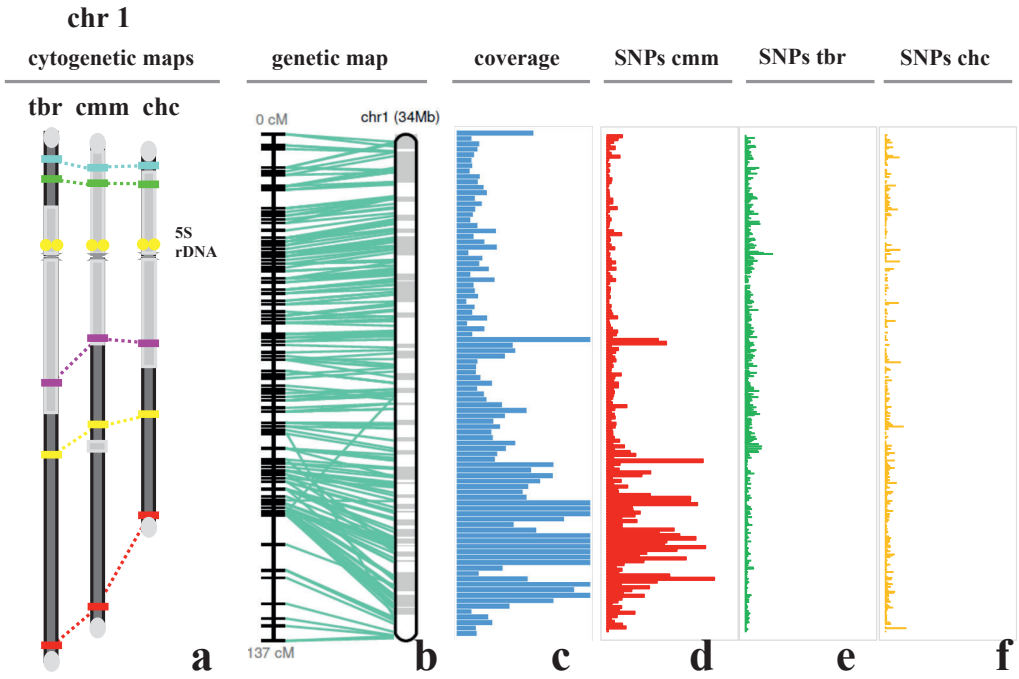


Figure 5. Example of the integration of all the different sources of information used on this improved version of the *Solanum commersonii* genome assembly for chromosome 1. a. Comparative cytogenetic maps of *Solanum tuberosum* (tbr), *S. commersonii* (cmm) and *S. chacoense* using five BACs (colour bands) and 5S rDNA as FISH probes. The main features of the chromosome morphology are indicated. Light grey circles at the ends: telomeres; light grey bands: heterochromatin; dark grey bands: euchromatin; light grey constriction: centromere (Gaiero et al. 2016, Chapter 4). b. Anchoring of the pseudomolecule 1 (right) on the SNP-derived genetic map (left). Green lines indicate the corresponding position of each SNP. Notice some discrepancies between the map and the assembly represented as diagonal crossing lines. c. Read coverage (blue bars) of the *S. commersonii* pseudomolecule 1 assembly produced by mapping all the Illumina short reads on the assembled and anchored pseudomolecule. d. SNP density (red bars) on *S. commersonii* pseudomolecule 1 compared to the previous genome assembly (Aversano et al. 2015). e. SNP density (green bars) on pseudomolecule 1 compared to the reference DM potato genome assembly (Hardigan et al. 2017). f. SNP density (yellow bars) on pseudomolecule 1 compared to the M6 *S. chacoense* genome assembly (Leisner et al. 2018).

interspecific comparison, most SNPs fall on chromosome 1 when comparing to potato and on chromosome 6 when comparing to M6 *S. chacoense*, while the highest interspecific SNP rates are for chromosome 12 with DM potato and chromosome 6 with M6 *S. chacoense*, albeit these values are relatively low (1.45 and 1.09 variants per Kbp, respectively). CNVs showed the opposite pattern, with 10x more between our assembly and M6 than between our assembly and DM potato (Table 4). The percentage of the genome involved in CNVs was also higher for M6 (386 Mbp, 46.8 % of the assembly) than for DM (57 Mbp, 7.8 % of the assembly). The size distribution of these CNVs was similar in both

Table 4. Coverage and intra and interspecific Single Nucleotide Polymorphism (SNP) and Copy Number Variation (CNV) per chromosome when comparing the *S. commersonii* sequence against the previous assembly (cmm, Aversano et al. 2015), against the reference DM potato genome (tbr, Hardigan et al. 2016) and the M6 *S. chacoense* genome. SNP rate expressed as variants per Kbp on average along each chromosome.

chr	# mapped reads	#SNPs cmm	SNP rate cmm	#SNPs tbr	SNP rate tbr	#CNVs tbr	#SNPs chc	SNP rate chc	#CNVs chc
1	8203188	179461	5.26	110201	1.24	110	26002	0.39	1150
2	4050515	93958	4.6	45524	0.94	64	14551	0.35	752
3	2814260	91860	3.6	70605	1.14	130	30097	0.64	867
4	2021543	85427	4.17	99190	1.37	81	20041	0.52	768
5	4328490	85289	6.65	67914	1.3	99	14780	0.35	767
6	6116082	106387	5.39	70056	1.18	45	45682	1.09	761
7	5908905	136403	5.8	61723	1.09	85	22919	0.55	791
8	2219945	65785	4.74	67561	1.19	77	14391	0.45	651
9	4164309	74229	5.2	77599	1.26	102	16062	0.45	809
10	1294546	53548	3.92	78120	1.31	46	9344	0.26	684
11	2152756	83868	4.21	55413	1.22	29	15710	0.41	763
12	1653004	53864	3.75	88790	1.45	75	10090	0.20	918
Total	**44927543**	**1110079**	**4.77**	**892696**	**1.22**	**943**	**239669**	**0.47**	**9681**

comparisons (Supplementary Figure S5), with similar medians of 21 Kbp, ranging from 1 Kbp to 632.8 Kbp in M6 (Suppl table 4) and from 0.4 Kbp to 782.6 Kbp in DM, but there were proportionally more CNVs under 1 Kbp in the comparison with DM potato (Supplementary table 3). When only CNVs less than 10 Kbp were considered, the medians differ greatly, with 3.2 Kbp for CNVs with DM and 5.6 Kbp for CNVs with M6 (Suppl tables 3 and 4). Chromosomes 1, 3 and 9 had similarly high numbers of CNVs whereas they were relatively lower in chromosomes 6, 10 and 11 (Table 4). Duplications were slightly more abundant (5233) than deletions (4446) in the comparison with M6 (Suppl table 4), while deletions were 2x more abundant (644) than duplications (298) when comparing against DM potato (Suppl table 3).

Discussion

Advantages over previous assembly

The main focus of this work was on structural genomics between potato and wild relatives *S. commersonii* and *S. chacoense*, so a stand-alone assembled and anchored reference genome of *S. commersonii* was essential to make structural comparisons with the genomes of DM potato and M6 *S. chacoense* which had been improved to the reference level (Sharma et al. 2013; Hardigan et al. 2016; Leisner et al. 2018). Although the gene

space for *S. commersonii* had been well described previously (Aversano et al. 2015), this draft assembly was highly fragmented and pseudomolecules were only achieved after mapping against the reference potato genome. We presented here a *S. commersonii de novo* reference genome that shows many improvements over the previous version (Aversano et al. 2015). The main improvement was in the planning of the sequencing effort, specifically in the choice of the genotype to be sequenced and the combination of libraries and sequencing platforms that were used. By sequencing a haploid genotype developed during this project (Castillo et al. 2016), we could avoid problems caused by heterozygosity, thus increasing assembly contiguity. With limited number of sequencing libraries and moderate sequence coverage, we managed to integrate high quality short sequences from Illumina with the long PacBio reads, achieving a less fragmented and more contiguous hybrid assembly (3,645 scaffolds vs 64,665; N50 294 Kb vs 44.3 Kb) with a more complete genome (85%) compared to the previous assembly. Our results at this stage are also comparable to the assembly metrics obtained in similar stages for the diploid self-compatible wild potato relative *S. verrucosum*, although the approach taken involved much higher coverage and combination of many more libraries and therefore the best assemblies that were chosen to continue with the long-range scaffolding were considerably more contiguous than our hybrid assembly. This hybrid assembly was further improved by including genetic map information.

We constructed a 1207 cM linkage map that could be condensed into the 12 linkage groups corresponding to the 12 chromosomes and some smaller groups. Not all the linkage group identifications corresponded to the homoeologous linkage groups in potato (Van Os et al. 2006), but they could be correctly identified following the potato ordering because of their high homoeology. We used this linkage map to integrate genetic and physical markers from the scaffolds, using a similar strategy as was chosen for potato (Sharma et al. 2013) and *S. chacoense* (Leisner et al. 2018), but with a different approach (Tang et al. 2015c). A high number of scaffolds were lost in the anchoring process (3044 out of 3645) and with an anchoring rate of under 39 %, a lot of the genome coverage was reduced. However, what remained was robust and reliable, and allowed us to build 12 pseudomolecules with scaffolds of confirmed order and mostly tentative orientation. When compared to the estimated size for each chromosome (Gaiero et al. 2016, Chapter 4), these pseudomolecules represent from 24 to 39 % of their corresponding chromosomes, coherent with the anchoring rate obtained. Compared to the M6 *S. chacoense* assembly (Leisner et al. 2018), the assembly presented here is less fragmented (3645 vs 8260 scaffolds) but covered less of the genome, had a lower N50 and showed lower anchor rate. Our *S. commersonii* reference genome was also less fragmented (3645 vs 66,254 scaffolds) than the DM potato reference (Sharma et al. 2013), had a similar coverage of the genome but a lower N50 and anchor rate. Nevertheless, this contiguous and robust assembly was sufficient to perform structural comparisons against potato and diploid wild relative *S. chacoense*.

Synteny and rearrangements among potato and wild relatives

As based on previously BAC-FISH study, there were no indications for large scale rearrangements among *S. commersonii*, *S. chacoense* and cultivated potato (Gaiero et al. 2016, Chapter 4). Nevertheless, BAC-FISH resolution is low and these experiments used only five markers per chromosome, leaving sufficient room for fine-scale structural variations that may involve traits of interest. The whole-genome level collinearity was confirmed indicating high homology across all three genomes. The whole-genome MUMmer comparison between DM potato and M6 *S. chacoense* performed here showed more mismatches than the one presented in the M6 genome paper (Leisner et al. 2018), probably due to differences in software parameters. It is highly likely that the higher contiguity of the present *S. commersonii* assembly resulted in a more contiguous alignment than when DM and M6 genomes were compared to each other.

It was expected to observe lack of contiguity in the pericentromeric heterochromatin for all pseudomolecules in all comparisons, as the high repeat content in this chromosome feature hinders sequencing and assembly (Paajanen et al. 2017) as well as anchoring on the linkage map due to reduced recombination so scaffolds remain unoriented or unordered. This was found when comparing DM potato chromosome 11 with *S. verrucosum*, even after long-range scaffolding (Paajanen et al. 2017) and also even between haplotypes of the exhaustively sequenced chromosome 5 homologues in potato (de Boer et al. 2015). We observed that these contiguity problems around the pericentromere were accentuated in smaller chromosomes, which is coherent with their higher proportion of repeat-rich pericentromeric heterochromatin (Peters et al. 2009).

The structural variants observed at the pseudomolecule level are mostly inversions. This was expected because these are the most common type of rearrangements between related species. However, when these inversions involve complete scaffolds, they should be divided into true and false positives, the latter caused by assembly artifacts resulting in incorrect random orientation of scaffolds. Currently, we have no means to differentiate these two types, but looking into scaffolds for pseudomolecule 2 into more detail, we could detect both true and false positives. Although wrong orientation and misassemblies were frequent, we have identified a few supported inversions, suggesting that these may be present on all pseudomolecules. Further integration of different sources of information such as genome maps (English et al. 2015; Tang et al. 2015b) will be needed to detect the full complement of rearrangements. Improvements on the assembly by anchoring on the genome map (Jiao et al. 2017a) or by using chromatin conformation capture (Udall and Dawe 2017; Li and Harkess 2018) will confirm putative structural variations. Sequences at the breakpoints of inverted regions have been related to repeats (Gaut et al. 2007; Belyayev 2014; Bennetzen and Wang 2014), and though repeats were found near or flanking synteny breaks between tomato, potato and pepper (Peters et al. 2012), in our case the repeats found could not be precisely mapped near the junctions. It still may

be that repeats are involved in the processes leading to the rearrangements observed, although further mapping is necessary to confirm their association.

Inversions or duplications within scaffolds can be regarded with more confidence. Various examples of these rearrangements could be found between homoeologues of pseudomolecule 1, which can be extended to the rest of the pseudomolecules. A detailed inspection of the sequences harboured on these segments should be performed to anticipate if they are going to represent an impediment to introgression and to shed light on the processes that may have caused them. We detected more duplications/deletions against the DM pseudomolecule than when comparing against the M6 pseudomolecule. This might be related to the higher incidence of copy number variation (CNV) in asexually propagated potato than in wild potatoes. Wild potato relatives have a higher rate of sexual reproduction and thus purge these variants through meiotic recombination (Hardigan et al. 2016). It is also possible that sequence information is missing in the *S. commersonii* pseudomolecule which is present in the DM homoeolog, and so appears as a deletion. Conversely, the M6 pseudomolecule could be not as well constructed as the DM pseudomolecule, showing more inverted regions due to misorientation of fragments. Although considerable effort has been put into improving the quality of the reference DM genome (Sharma et al. 2013; Hardigan et al. 2016), misassemblies, assembly artifacts and errors in order and orientation tend to be frequent and they are hard to discriminate from real structural variants.

Assembly-independent variants: SNPs and CNVs

Nucleotide variants, small indels (collectively identified as SNPs) and copy number variants (CNVs) were called by mapping the short Illumina reads for *S. commersonii* against the other assembled genomes so these variants do not depend on the quality of our assembly. Although there are still technical challenges in read alignments in repetitive or highly diverged regions of the genome, difficulties in read mapping do not seem to be causing the differences in SNP rate and SNP densities observed. Nevertheless, the intraspecific nucleotide diversity rate found here is more similar to the rate found among wild diploid species diversity panel (3.8%) while interspecific SNP rates were more similar to those found across diploid landraces (1.8%) (Hardigan et al. 2017). A population-wide SNP analysis within *S. commersonii* and comparing its diversity to *S. chacoense* would verify this higher intraspecific variability. A diverse collection from the distribution areas of these two species has been assembled and its analysis with SSR markers is underway, a genotyping by sequencing approach such as GBS (Elshire et al. 2011) could be implemented to confirm this hypothesis.

The incidence of Copy Number Variants (CNVs) was lower between our *S. commersonii* sequence and DM potato (7.8%) than what was reported for intraspecific CNVs in a panel of related double monoploids (30.2%) (Hardigan et al. 2016). Both the DM panel and our sequenced clone passed the haploid sieve through anther culture. Our haploid

clone survived both *in vitro* and *ex vitro* and even produced flowers (Castillo et al. 2016), so the CNVs that it carries are not lethal. It was not our aim to assess whether these CNVs fall in intergenic regions or on genes, but it seems safe to claim that they are part of the dispensable genome (Hardigan et al. 2016). These authors propose that asexually propagated plants have a higher rate of CNVs while wild potato relatives, which have a higher proportion of sexual reproduction should have CNVs purged through meiotic recombination events (Hardigan et al. 2016). However, this explanation does not reflect the high CNV proportion (46.8%) observed between *S. commersonii* and M6 *S. chacoense*, both wild diploids. This proportion is also higher than what was found among wild potatoes for gene-level CNVs (Hardigan et al. 2017). It is possible that inbreeding for many generations (Leisner et al. 2018) has fixed CNVs that are different from those in outcrossing *S. commersonii*. It is also possible that *S. chacoense* has more CNVs with respect to its relatives. In the DM panel, clone DM13 had similar numbers of CNVs (10532) and it is known that this clone has 50 % *S. chacoense* ancestry (Hardigan et al. 2016). Although most CNVs were short, with medians similar to those found within potato (3.0 Kbp) (Hardigan et al. 2016), the size range observed here was wider. Small CNVs can be explained by non-allelic homologous recombination in regions which contain segmental homology (Lu et al. 2012) or regions with low-copy repeats, even by transposition of repetitive elements (Morgante et al. 2007). Large CNVs (> 100 Kbp) are mostly duplications and mostly located in the pericentromere (Suppl tables 3 and 4), which is coherent with what has been found in potato diversity panels (Hardigan et al. 2016; Hardigan et al. 2017) and even between homologues in the same potato genotypes (de Boer et al. 2015). Although some large CNVs detectable through FISH were widespread in the euchromatin of potato cultivars (Iovene et al. 2013), the pericentromeric CNVs detected here are likely not to carry traits of interest and thus should not impair introgression.

Implications for breeding in *Solanum*

Overall, the genome assembly of the haploid genotype from wild potato species *S. commersonii* provides a high-quality representation of the estimated 792-Mb genome, with scaffolds anchored into 12 pseudomolecules. This improved version can assist exploration of traits, marker development and can guide decisions in introgressive hybridization schemes. However, for this assembly to achieve its full usefulness, it should be improved in its contiguity, as was done for wild relative *S. verrucosum* (Paajanen et al. 2017). An optimal combination of resources invested in long-range scaffolding technologies will see its contiguity increased enormously.

The high genome homology and collinearity found between *S. commersonii* and cultivated potato encourages its use in introgressive breeding. The structural variants and copy number variants that we could identify were mostly small (of the 100 Kbp order) and mostly located in the pericentromere region, so they should not hamper introgression of desirable traits. Nevertheless, more careful inspection of these structural vari-

ants is necessary to identify the coding sequences present and evaluate their phenotypic results. Crossing over (CO) rates should be evaluated in interspecific backcross lines to elucidate if there is reduced CO in particular domains that can be attributed to structural variants (Demirci et al. 2017) and thus assess the impact of both aspects in breeding.

Literature Cited

Abbott JC, Butcher SA (2012) Strategies towards sequencing complex crop genomes. Genome Biol 13:322. doi: 10.1186/gb4059

Abyzov A, Urban A, Snyder M, Gerstein M (2011) CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res 21:974–84. doi: doi: 10.1101/gr.114876.110.

Aflitos SA, Schijlen E, de Jong H, et al. (2014) Exploring genetic variation in the tomato (*Solanum* section Lycopersicon) clade by whole-genome sequencing. Plant J 80:136–148. doi: 10.1111/tpj.12616

Appels R, Nystrom J, Webster H, Keeble-Gagnere G (2015) Discoveries and advances in plant and animal genomics. Funct Integr Genomics 15:121–129. doi: 10.1007/s10142-015-0434-3

Aversano R, Contaldi F, Ercolano MR, et al. (2015) The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. Plant Cell 27:954–968. doi: 10.1105/tpc.114.135954

Belton J-M, McCord RP, Gibcus JH, et al. (2012) Hi–C: A comprehensive technique to capture the conformation of genomes. Methods 58:268–276. doi: 10.1016/j.ymeth.2012.05.001

Belyayev A (2014) Bursts of transposable elements as an evolutionary driving force. J Evol Biol 27:2573–2584. doi: 10.1111/jeb.12513

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65:505–530. doi: 10.1146/annurev-arplant-050213-035811

Bernatzky R, Tanksley SD (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887–98.

Boetzer M, Pirovano W (2012) Toward almost closed genomes with GapFiller. Genome Biol 13:R56. doi: https://doi.org/10.1186/gb-2012-13-6-r56

Bolger A, Scossa F, Bolger ME, et al. (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii.* Nat Genet 46:1034–1038. doi: 10.1038/ng.3046

Bombarely A, Moser M, Amrad A, et al. (2016) Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. Nat Plants In press:1–9. doi: 10.1038/nplants.2016.74

Bradbury PJ, Zhang Z, Kroon DE, et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples.

Byrne SL, Nagy I, Pfeifer M, et al. (2015) A synteny-based draft genome sequence of the forage grass *Lolium perenne*. Plant J 84:816–826. doi: 10.1111/tpj.13037

Cai C, Wang X, Liu B, et al. (2017) *Brassica rapa* Genome 2.0: A Reference Upgrade through Sequence Re-assembly and Gene Re-annotation. Mol Plant 10:649–651. doi: 10.1016/j.molp.2016.11.008

Cardone MF, D 'addabbo P, Alkan C, et al. (2015) Inter-varietal structural variation in grapevine genomes. doi: 10.1111/tpj.13274

Castillo A, Gaiero P, López-Carro B, Vilaró F (2016) Gametic embryogenic response in wild diploid *Solanum* species and its implications for genome sequencing projects and breeding. Plant Tissue Cult Biotechnol 26:159–173. doi: 10.3329/ptcb.v26i2.30566

Chaisson M, Tesler G (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics 13:238.

Chaney L, Sharp AR, Evans CR, Udall JA (2016) Genome Mapping in Plant Comparative Genomics. Trends Plant Sci 21:770–780. doi: 10.1016/j.tplants.2016.05.004

Chen X, Li H, Pandey MK, et al. (2016) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. Proc Natl Acad Sci. doi: 10.1073/pnas.1600899113

de Boer JM, Datema E, Tang X, et al. (2015) Homologues of potato chromosome 5 show variable collinearity in the euchromatin, but dramatic absence of sequence similarity in the pericentromeric heterochromatin. BMC Genomics 16:374. doi: 10.1186/s12864-015-1578-1

Demirci S, van Dijk ADJ, Sanchez Perez G, et al. (2017) Distribution, position and genomic characteristics of crossovers in tomato recombinant inbred lines derived from an interspecific cross between *Solanum lycopersicum* and *Solanum pimpinellifolium*. Plant J 89:554–564. doi: 10.1111/tpj.13406

Elshire RJ, Glaubitz JC, Sun Q, et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. doi: 10.1371/journal.pone.0019379

English A, Salerno W, Hampton O, et al. (2015) Assessing structural variation in a personal genome—towards a human reference diploid genome. BMC Genomics. doi: 10.1186/s12864-015-1479-3

Gaiero P, Mazzella C, Vilaró F, et al. (2017) Pairing analysis and *in situ* Hybridisation reveal autopolyploid-like behaviour in *Solanum commersonii* × *S. tuberosum* (potato) interspecific hybrids. Euphytica 213:137. doi: 10.1007/s10681-017-1922-4

Gaiero P, van de Belt J, Vilaró F, et al. (2016) Collinearity between potato (*Solanum tuberosum* L.) and wild relatives assessed by comparative cytogenetic mapping. Genome 60:228–240. doi: 10.1139/gen-2016-0150

Gaut BS, Wright SI, Rizzon C, et al. (2007) Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet 8:77–84. doi: 10.1038/nrg1970

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. doi: 10.1093/bioinformatics/btt086

Hardigan MA, Crisovan E, Hamiltion JP, et al. (2016) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. Plant Cell 28:TPC2015–00538–RA. doi: 10.1105/tpc.15.00538

Hardigan MA, Laimbeer FPE, Newton L, et al. (2017) Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proc Natl Acad Sci 201714380. doi: 10.1073/pnas.1714380114

Hastie A, Dong L, Smith A, et al. (2013) Rapid genome mapping in nanochannel arrays for highly complete and accurate *de novo* sequence assembly of the complex *Aegilops tauschii* genome. PLoS One 8:e55864. doi: 10.1371/journal.pone.0055864

Imelfort M, Edwards D (2009) *De novo* sequencing of plant genomes using second-generation technologies. Brief Bioinform 10:609–18. doi: 10.1093/bib/bbp039

Iovene M, Zhang T, Lou Q, et al. (2013) Copy number variation in potato - An asexually propagated autotetraploid species. Plant J 75:80–89. doi: 10.1111/tpj.12200

Istvánek J, Jaros M, Krenek A, Řepková J (2014) Genome assembly and annotation for red clover (*Trifolium pratense*; Fabaceae). Am J Bot 101:327–37. doi: 10.3732/ajb.1300340

Jansky SH, Chung YS, Kittipadukal P (2014) M6: A Diploid Potato Inbred Line for Use in Breeding and Genetics Research. J Plant Regist 8:195. doi: 10.3198/jpr2013.05.0024crg

Jarvis DE, Ho YS, Lightfoot DJ, et al. (2017) The genome of *Chenopodium quinoa*. Nature 1–6. doi: 10.1038/nature21370

Jiao W, Accinelli GG, Hartwig B, et al. (2017a) Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. Genome Res gr.213652.116. doi: 10.1101/gr.213652.116

Jiao W, Accinelli GG, Hartwig B, et al. (2017b) Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. 778–786. doi: 10.1101/gr.213652.116.Freely

Jiao Y, Peluso P, Shi J, et al. (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524–527. doi: 10.1038/nature22971

Kaur P, Bayer PE, Milec Z, et al. (2017) An advanced reference genome of *Trifolium subterraneum* L. reveals genes related to agronomic performance. Plant Biotechnol J 15:1034–1046. doi: 10.1111/pbi.12697

Kloosterman B, Abelenda JA, del Mar Carretero Gomez M, et al. (2015) Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495:246–250. doi: 10.1017/CBO9781107415324.004

Kurtz S, Phillippy A, Delcher AL, et al. (2004) Versatile and open software for comparing large genomes. Genome Biol 5:R12. doi: 10.1186/gb-2004-5-2-r12

Lam ET, Hastie AR, Lin C, et al. (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30:771–6. doi: 10.1038/nbt.2303

Leisner CP, Hamilton JP, Crisovan E, et al. (2018) Genome sequence of M6, a diploid inbred clone of the high glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. Plant J. doi: 10.1111/tpj.13857

Li FW, Harkess A (2018) A guide to sequence your favorite plant genomes. Appl Plant Sci 6:1–7. doi: 10.1002/aps3.1030

Lu P, Han X, Qi J, et al. (2012) Analysis of *Arabidopsis* genome-wide variations before and after meiosis and meiotic recombination by resequencing Landsberg *erecta* and all four products of a single meiosis. Genome Res 22:508–18. doi: 10.1101/gr.127522.111

Montero-Pau J, Blanca J, Bombarely A, et al. (2018) *De novo* assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the *Cucurbita* genus. Plant Biotechnol J 16:1161–1171. doi: 10.1111/pbi.12860

Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol 10:149–155. doi: 10.1016/j.pbi.2007.02.001

Paajanen P, Kettleborough G, Lopez-Girona E, et al. (2017) A critical comparison of technologies for a plant genome sequencing project. bioRxiv 201830. doi: 10.1101/201830

Peters SA, Bargsten JW, Szinay D, et al. (2012) Structural homology in the Solanaceae: analysis of genomic regions in support of synteny studies in tomato, potato and pepper. Plant J 71:602–14. doi: 10.1111/j.1365-313X.2012.05012.x

Peters SA, Datema E, Szinay D, et al. (2009) *Solanum lycopersicum* cv. Heinz 1706 chromosome 6: distribution and abundance of genes and retrotransposable elements. Plant J 58:857–69. doi: 10.1111/j.1365-313X.2009.03822.x

Pham GM, Newton L, Wiegert-Rininger, Krystle Vaillancourt B, et al. (2017) Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. Plant J. doi: doi: 10.1111/tpj.13706

Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–95. doi: 10.1038/nature10158

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Reyes-Chin-Wo S, Wang Z, Yang X, et al. (2017) Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. Nat Commun. doi: 10.1038/ncomms14953

Schmidt MH, Vogel A, Denton AK, et al. (2017) *De novo* Assembly of a new *Solanum pennellii* Accession Using Nanopore Sequencing. Plant Cell tpc.00521.2017. doi: 10.1105/tpc.17.00521

Sharma SK, Bolser D, de Boer J, et al. (2013) Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3 Genes, Genomes, Genet 3:2031–47. doi: 10.1534/g3.113.007153

Shearer LA, Anderson LK, de Jong H, et al. (2014) Fluorescence *in situ* hybridization and optical mapping to correct scaffold arrangement in the tomato genome. G3 Genes, Genomes, Genet 4:1395–1405. doi: 10.1534/g3.114.011197

Tang H, Krishnakumar V, Bidwell S, et al. (2014) An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. BMC Genomics 15:312. doi: 10.1186/1471-2164-15-312

Tang H, Lyons E, Town CD (2015a) Optical mapping in plant comparative genomics. Gigascience 4:3. doi: 10.1186/s13742-015-0044-y

Tang H, Zhang X, Miao C, et al. (2015b) ALLMAPS: Robust scaffold ordering based on multiple maps. Genome Biol 16:3. doi: 10.1186/s13059-014-0573-1

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641. doi: 10.1038/nature11119

Tran HTM, Ramaraj T, Furtado A, et al. (2018) Use of a draft genome of coffee (Coffea arabica) to identify SNPs associated with caffeine content. Plant Biotechnol J 16:1756–1766. doi: 10.1111/pbi.1291

Udall J, Dawe RK (2017) Is it ordered correctly? Validating genome assemblies by optical mapping. Plant Cell 30:tpc.00514.2017. doi: 10.1105/tpc.17.00514

Van Os H, Andrzejewski S, Bakker E, et al. (2006) Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics 173:1075–87. doi: 10.1534/genetics.106.055871

Xu C, Jiao C, Sun H, et al. (2017) Draft genome of spinach and transcriptome diversity of 120 *Spinacia* accessions. Nat Commun 8:15275. doi: 10.1038/ncomms15275

Ye C, Hill CM, Wu S, et al. (2016) DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. Sci Rep 6:31900. doi: 10.1038/srep31900

Ye C, Ma ZS (2016) Sparc : a sparsity-based consensus algorithm for long erroneous sequencing reads. 1–12. doi: 10.7717/peerj.2016

Zapata L, Ding J, Willing E-M, et al. (2016) Chromosome-level assembly of *Arabidopsis thaliana* L*er* reveals the extent of translocation and inversion polymorphisms. Proc Natl Acad Sci 113:E4052–E4060. doi: 10.1073/pnas.1607532113

Zhang H, Tan E, Suzuki Y, et al. (2014) Dramatic improvement in genome assembly achieved using doubled-haploid genomes. Sci Rep 4:6780. doi: 10.1038/srep06780

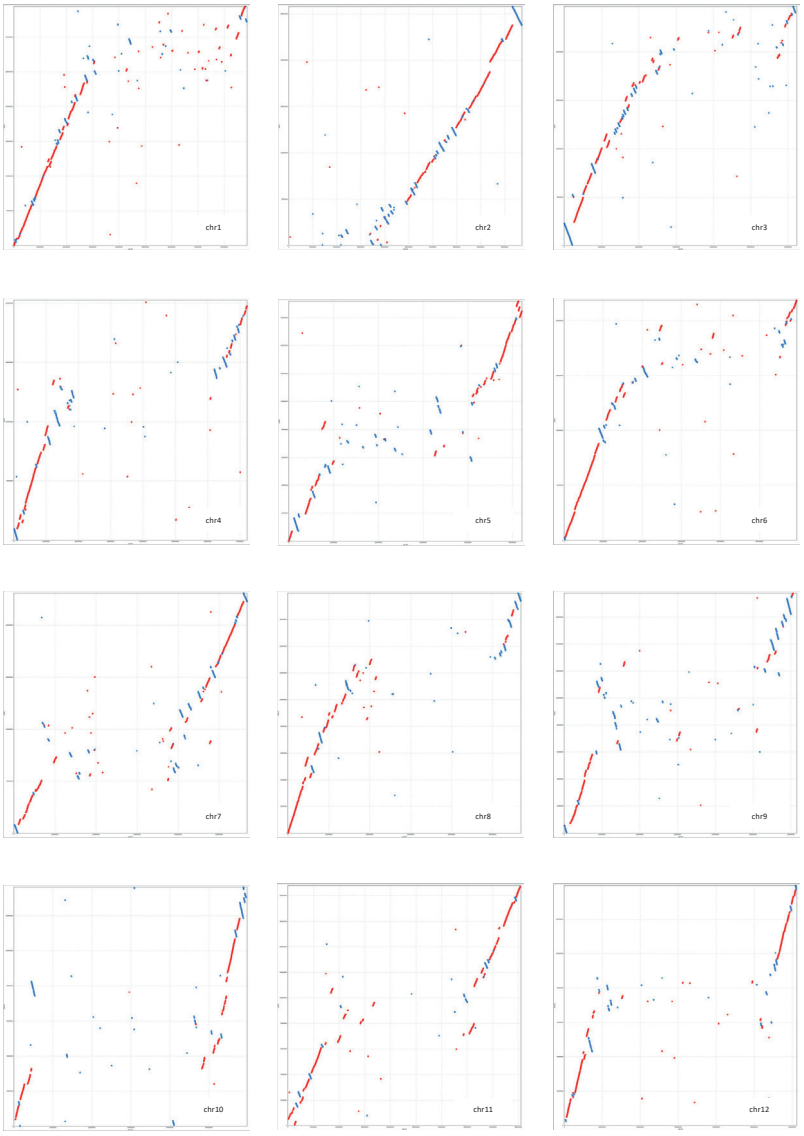


Figure S1. a. Mummerplots of each pseudomolecule from *Solanum commersonii* (cmm) against its homolog in DM potato. Each cmm pseudomolecule was built by anchoring hybrid scaffolds to the genetic map using at least one SNP marker per scaffold that was mapped to a linkage group. Identification of linkage groups and corresponding pseudomolecules was corrected following homology with DM pseudomolecules. The orientation of each pseudomolecule followed that of DM. Each scaffold within a cmm pseudomolecule was oriented using two or more SNP markers. When only one SNP marker was present, orientation is ambiguous.

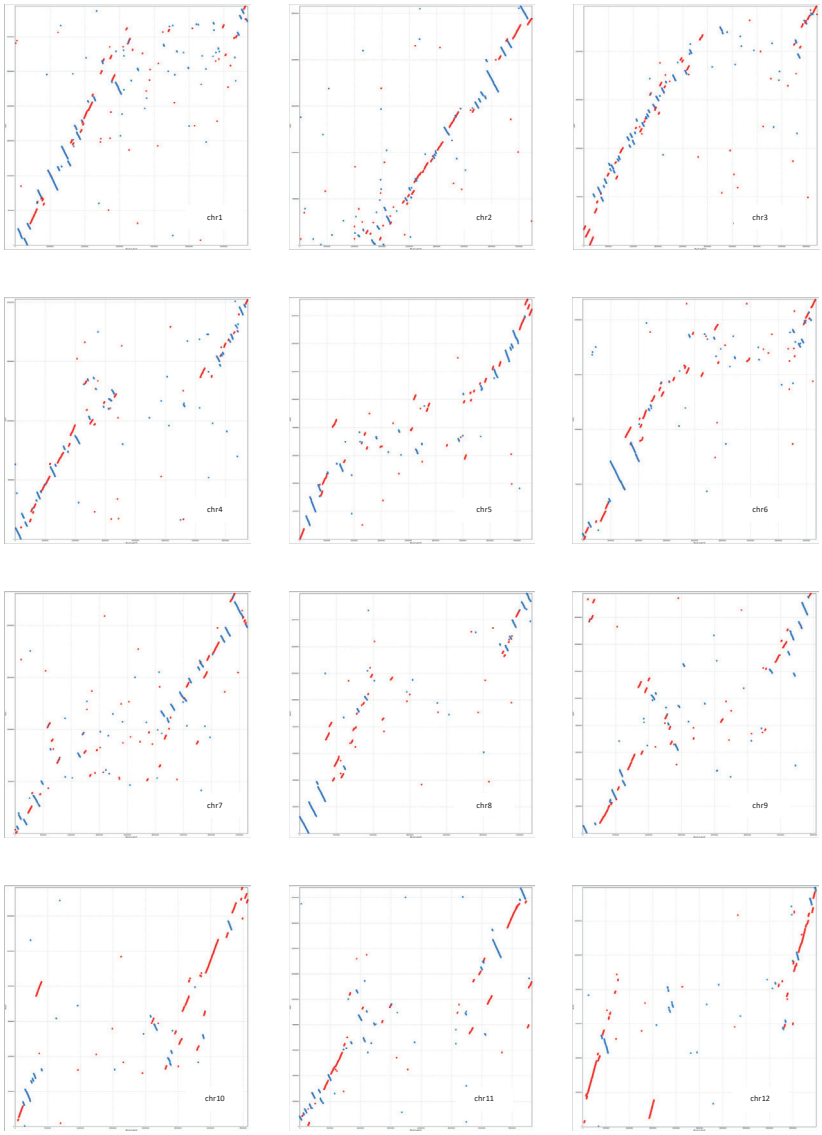


Figure S1. b. Mummerplots of each pseudomolecule from *Solanum commersonii* (cmm) against its homolog in M6 *S. chacoense*. Each cmm pseudomolecule was built by anchoring hybrid scaffolds to the genetic map using at least one SNP marker per scaffold that was mapped to a linkage group. Identification of linkage groups and corresponding pseudomolecules was corrected following homology with DM pseudomolecules. The orientation of each pseudomolecule followed that of DM. Each scaffold within a cmm pseudomolecule was oriented using two or more SNP markers. When only one SNP marker was present, orientation is ambiguous.

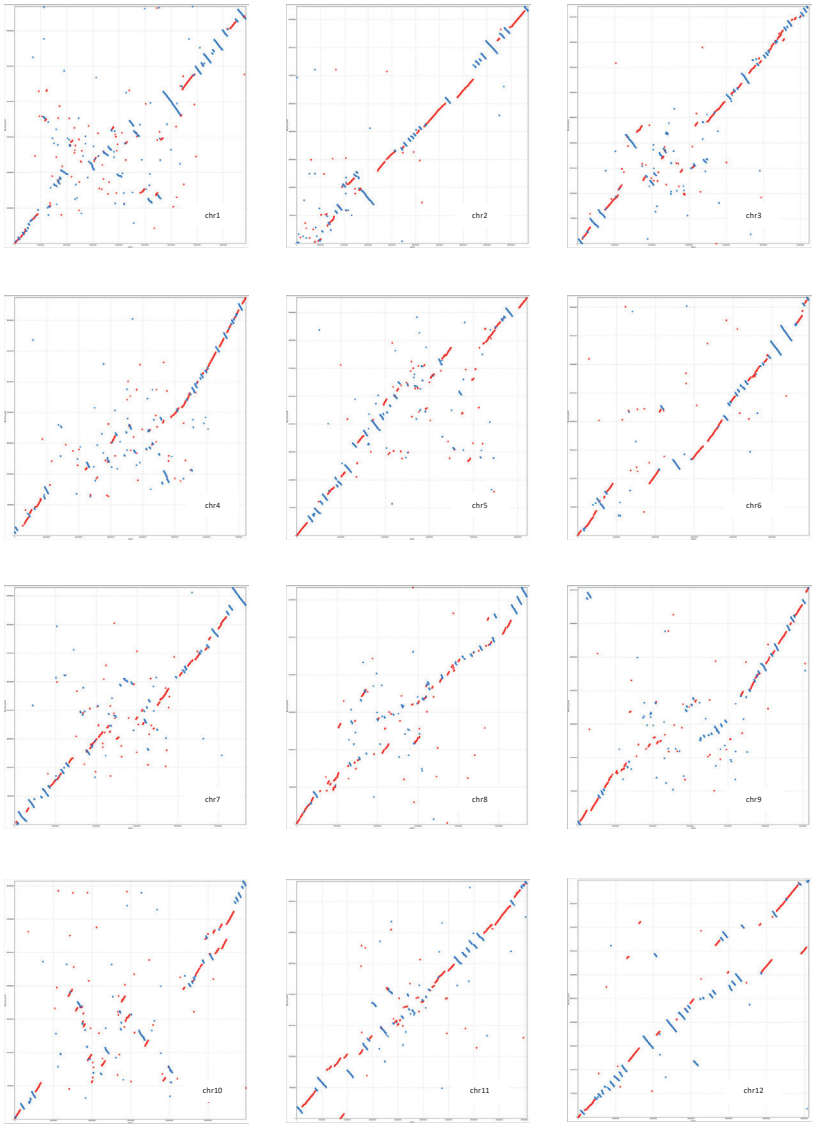


Figure S1. c. Mummerplots of each published pseudomolecule from M6 *Solanum chacoense* against its homolog in reference DM potato genome, using the same NUCmer parameters as in the previous comparisons.

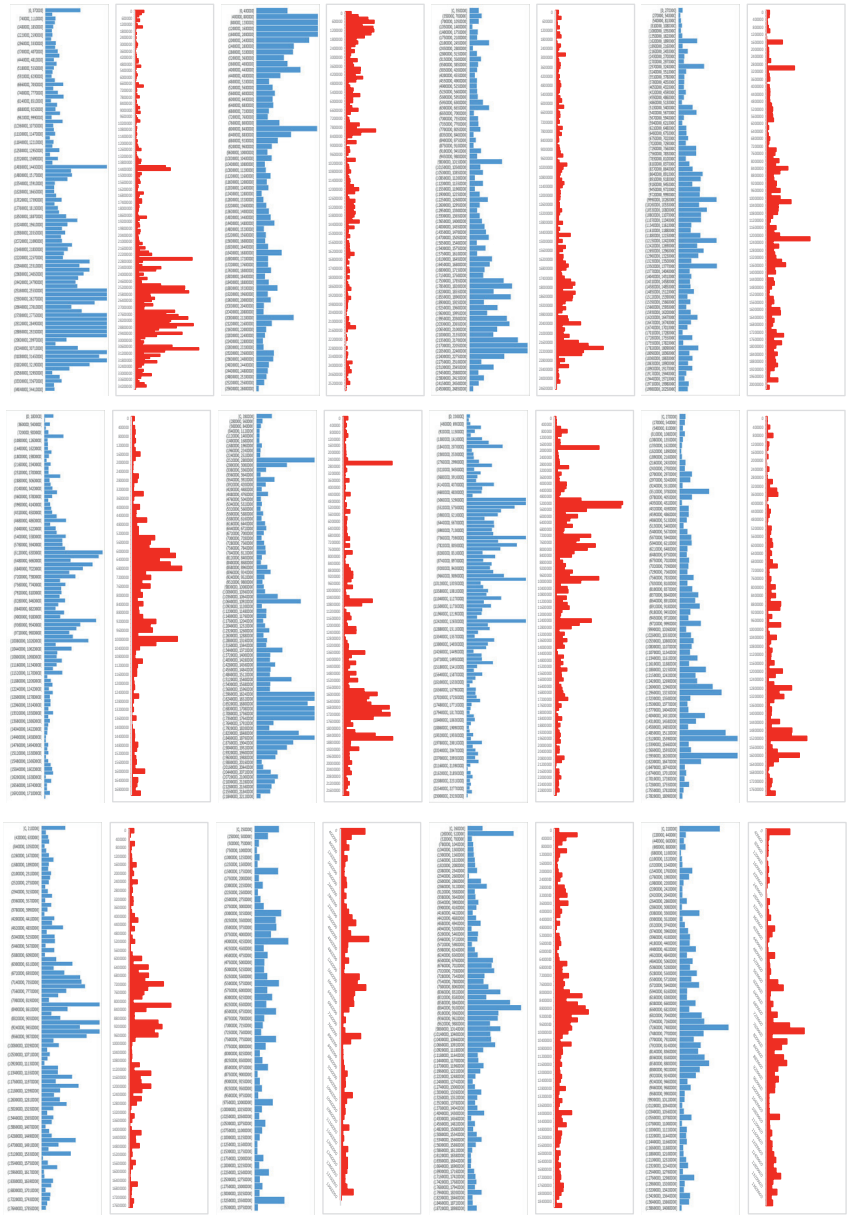


Figure S2. Read coverage (blue bars) of the twelve *S. commersonii* pseudomolecules produced by mapping all the Illumina short reads on the assembled and anchored pseudomolecules. SNP density (red bars) over 200-Kbp windows on *S. commersonii* pseudomolecules compared to the previous genome assembly (Aversano et al. 2015). Distances along the pseudomolecules are expressed in bp, maximum frequency of reads is 14000 and maximum SNP density is 8000 SNPs.

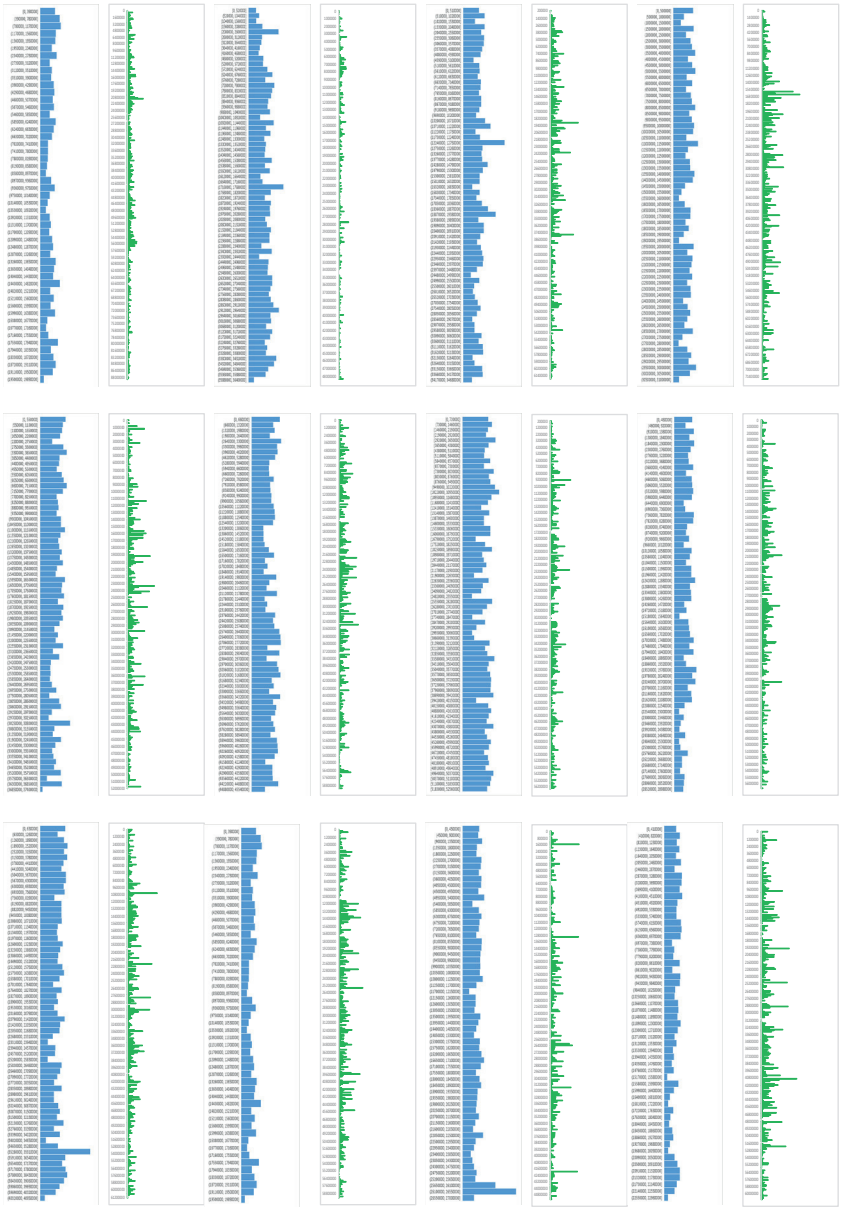


Figure S3. Read coverage (blue bars) of the twelve DM potato pseudomolecules assembly produced by mapping against the *S. commersonii* assembled and anchored pseudomolecules. SNP density (green bars) over 200-Kbp windows on the twelve reference DM potato pseudomolecules (Hardigan et al. 2016) compared to the *S. commersonii* genome assembly. Distances along the pseudomolecules are expressed in bp, maximum frequency of reads is 14000 and maximum SNP density is 4000 SNPs

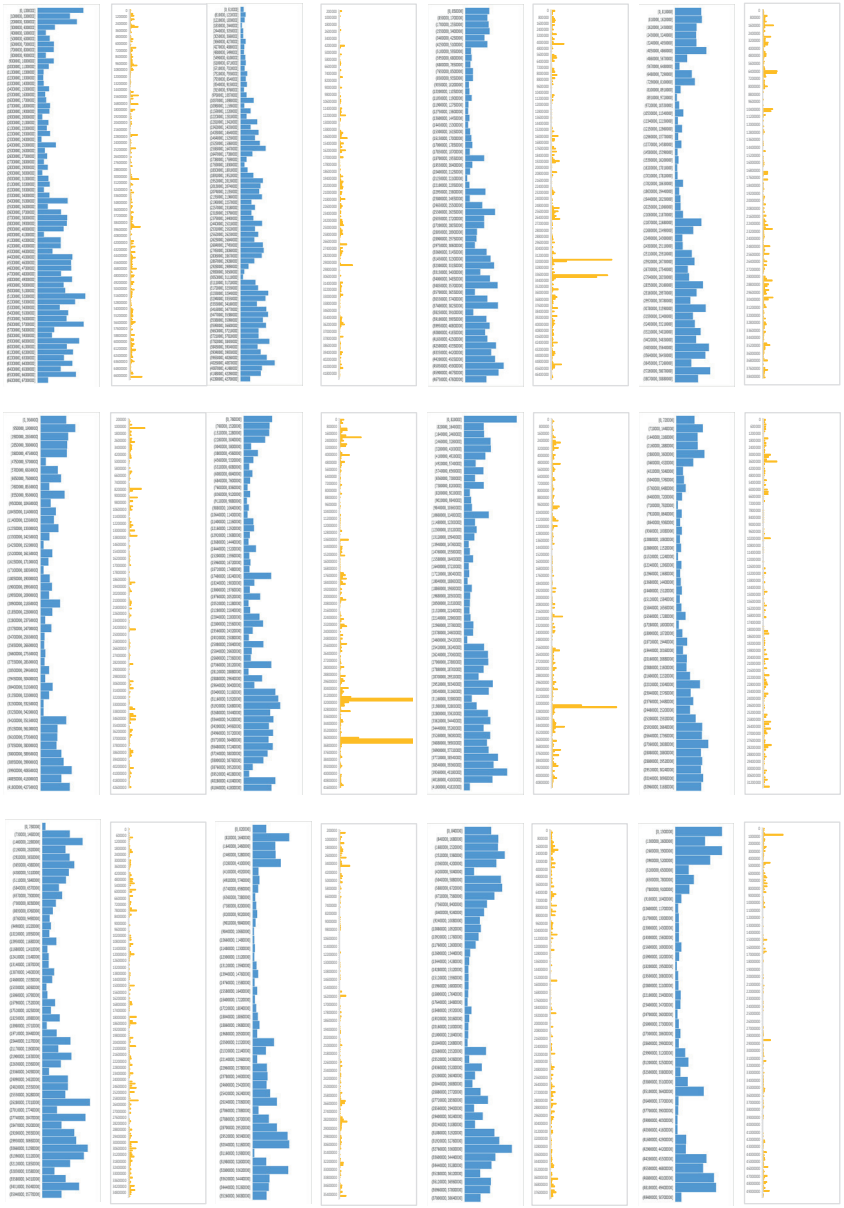


Figure S4. Read coverage (blue bars) of the twelve M6 S. chacoense pseudomolecules assembly produced by mapping against the *S. commersonii* assembled and anchored pseudomolecules. SNP density (yellow bars) over 200-Kbp windows on the twelve reference M6 *S. chacoense* pseudomolecules (Leisner et al. 2018) compared to the *S. commersonii* genome assembly. Distances along the pseudomolecules are expressed in bp, maximum frequency of reads is 14000 and maximum SNP density is 4000 SNPs.

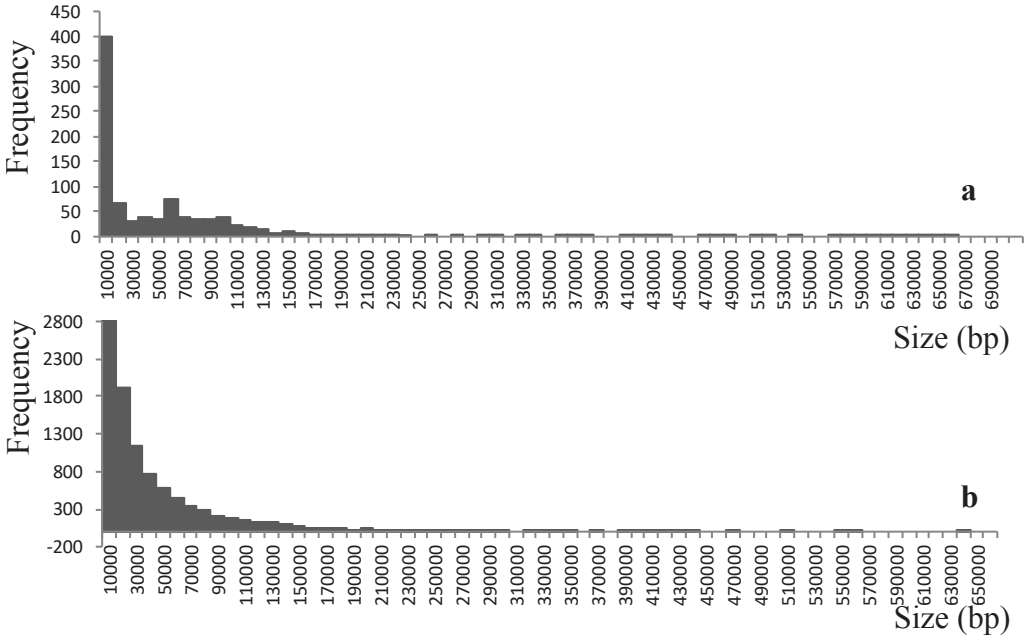


Figure S5. Size distribution of copy number variants between the present *S. commersonii* genome sequence and a. DM potato reference genome; b. M6 *S. chacoense* genome

Large-sized supplementary tables available from:

https://www.dropbox.com/sh/22d2fp116g01sc9/AAAU2_71e5y-gdzYEsoNAskAa?dl=0

Chapter 7

Intact DNA purified from flow-sorted nuclei unlocks the potential of next-generation genome mapping and assembly in *Solanum* species

Paola Gaiero[1,2], Hana Šimková[3], Jan Vrána[3], Federico F. Santiñaque[4], Beatriz López-Carro[4], Gustavo A. Folle[4], José van de Belt[2], Sander A. Peterse[5], Jaroslav Doležel[3], Hans de Jong[1]

1. Faculty of Agronomy, University of the Republic, Montevideo, Uruguay
2. Laboratory of Genetics, Wageningen University& Research, Wageningen, The Netherlands
3. Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany, Olomouc, Czech Republic
4. Flow Cytometry and Cell Sorting Core, Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), Montevideo, Uruguay
5. Applied Bioinformatics, Department of Bioscience, Wageningen University & Research, Wageningen, The Netherlands

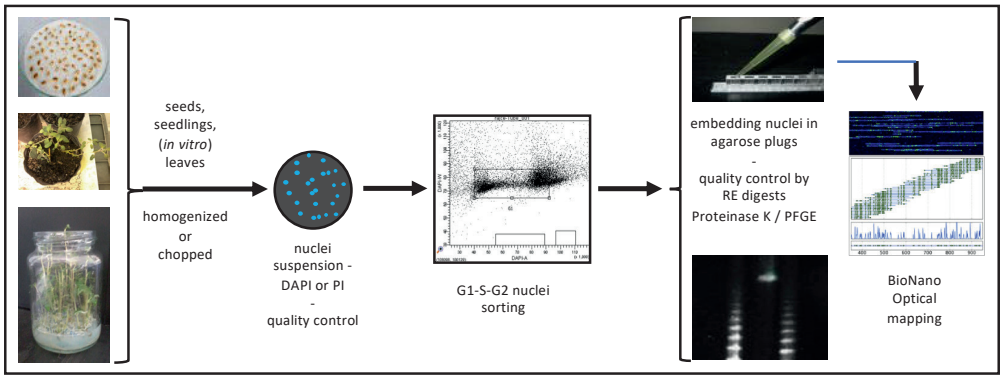
Published in:
MethodsX 5: 328-336 (2018) https://doi.org/10.1016/j.mex.2018.03.009

Abstract:

Next-generation genome mapping through nanochannels (Bionano optical mapping) of plant genomes brings genome assemblies to the 'nearly-finished' level for reliable detailed gene annotations and assessment of structural variations. Despite the recent progress in its development, researchers face the technical challenges of obtaining sufficient high molecular weight nuclear (HMW) DNA due to cell walls which are difficult to disrupt and to the presence of cytoplasmic polyphenols and polysaccharides that co-precipitate or are covalently bound to DNA and might cause oxidation and/or affect the access of nicking enzymes to DNA, preventing downstream applications. Here we describe important improvements for obtaining HMW DNA that we tested on *Solanum* crops and wild relatives. The methods that we further elaborated and refined focus on

- improving flexibility of using different tissues as source materials, like fast-growing root tips and young leaves from seedlings or *in vitro* plantlets.
- obtaining nuclei suspensions through either lab homogenizers or by chopping.
- increasing flow sorting efficiency using DAPI (4',6-diamidino-2-phenylindole) and PI (propidium iodide) DNA stains, with different lasers (UV or 488nm) and sorting platforms such as the FACSAria and FACSVantage flow sorters, thus making it appropriate for more laboratories working on plant genomics.

The obtained nuclei are embedded into agarose plugs for processing and isolating uncontaminated HMW DNA, which is a prerequisite for nanochannel-based next-generation optical mapping strategies.



Graphical abstract. Nuclei sorting and High Molecular Weight (HMW) DNA isolation workflow. Nuclei suspensions are prepared by homogenization or chopping of root tips from germinated seeds or leaves from seedlings or plants, either potted or *in vitro*. Nuclei are sorted by selecting the G1-S-G2 populations through a FACS Aria or FACS Vantage flow sorter. Flow sorted nuclei are embedded in agarose plugs and HMW DNA is purified with proteinase K and RNAse A and then isolated. Pulsed-field gel electrophoresis (PFGE) is used for DNA quality control. HMW DNA is subsequently labelled and analyzed on the Irys platform. Optical mapping (OM) data are then processed for *de novo* OM and hybrid assemblies.

Method details

1. Plant material and preparation of nuclei suspensions (for two plugs with 500,000 nuclei each)

The whole workflow for this method is summarized in the graphical abstract. The method follows Šimková et al. (2003) with the following modifications. First of all, we introduced variations in the starting material types. For tomato (*Solanum lycopersicum*), we used root tips, for *S. commersonii* young leaves from *in vitro* plantlets and for *S. etuberosum* young leaves from plants grown in pots. Second, we tested different options to obtain nuclei suspensions, namely homogenizing with a Polytron or chopping with a razor blade. The protocol for tomato root tips is as follows:

1.1.1 Germinate about 200 seeds on humidified filter paper in Petri dishes for about 4 days.

1.2.1 Add 1.1 µL β-mercaptoethanol per 1 mL of 1.5x isolation buffer (IB [1]: 15 mM Tris, 10 mM EDTA, 130 mM KCl, 20 mM NaCl, 1 mM spermine, 1 mM spermidine and 0.1 % Triton X-100, pH 9.4) just before use.

1.3.1 Transfer the seedlings to a 2 % formaldehyde solution (from stock solution 36.5-38 % in H2O, SIGMA F8775) in Tris buffer (10 mM Tris, 10 mM EDTA, 100 mM NaCl, 0.1 % v/v Triton X-100, pH 7.5). Incubate in a water bath at 4° for 20 min and wash three times in Tris buffer at 4° for 5 min.

1.4.1 Dissect 1-2 cm of the root tips on a glass Petri dish, divide the material between two 5 mL polystyrene tubes containing 1 mL ice-cold 1.5x IB with β-mercaptoethanol and keep on ice.

1.5.1 Homogenize samples using a Polytron PT1200 homogenizer at 15,000 rpm for 13 s (time and speed adjustable according to species).

The alternative protocol for *S. commersonii* and *S. etuberosum* young plant leaves is as follows:

1.1.2 Fix whole *in vitro* grown plantlets or detached leaves from potted plants in formaldehyde solution as described above.

1.2.2 After rinsing, place 0.5-1 g of leaves in a glass Petri dish with 1 mL of 1.5x IB buffer

1.3.2 Chop the tissues using a sharp razor blade until a soft homogenate is obtained. This should be formed by very small pieces of leaves in a green suspension. Continue from step 1.6.

1.6 Filter the crude homogenates through a 50 µm nylon mesh into a new polystyrene tube and a 25 µm nylon mesh (Silk & Progress, 130T EXTRA, www.silkandprogress.cz), respectively. Alternatively, samples can be filtered through a Falcon® 40 µm cell strainer (Corning Life Sciences, Oneonta, New York, Product #352340). Collect the filtered nuclei suspensions aliquots up to a volume of ~4 mL.

1.7 Add DAPI to a final concentration of 2 µg.mL-1. Check nuclei integrity and concentration under the fluorescence microscope equipped with appropriate excitation and emission filters. Nuclei should be round-shaped, not broken and at a density of 150-200 nuclei per mm2 (at 10x magnification).

1.8 Keep samples on ice until flow cytometric analysis and sorting.

2. Nuclei flow sorting

For flow sorting, we adjusted the protocols of Šimková et al. (2003) and Vrána et al. (2012) to allow the use of either the FACS Aria or the FACS Vantage flow sorters. We introduced the following modifications:

2.1.1 Stain the nuclei with DAPI (final concentration: 2 µg.mL-1)

2.2.1 Sort DAPI-stained nuclei using a FACSAria II SORP (BD Biosciences, Santa Clara, USA) with the following settings:

a) Solid-state laser in the UV range (355 nm, 100 mW); b) 70 µm nozzle, 70 psi; c) sorting speed: 300 events/s.

We performed data acquisition and analysis with the BD FACSDiva software (BD Bioscences, Santa José, CA, USA)

2.3.1 Select the G1, S and G2 nuclei for sorting using DAPI-A vs DAPI-W dot plots (Figure 1a).

In order to use the FACS Vantage (BD Biosciences, Santa Clara, USA), the following protocol for staining with PI was used:

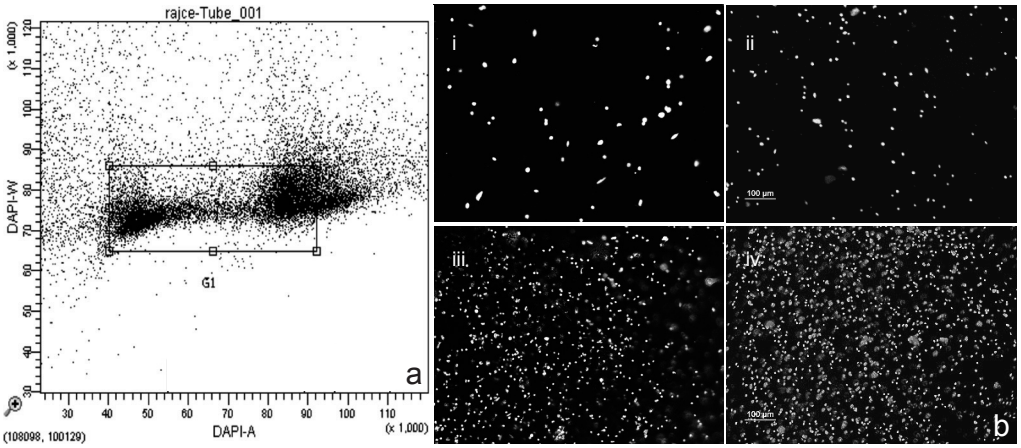


Figure 1: a. Dotplot of DAPI- A (x axis) vs DAPI-W (y axis), framed dot clouds are *Solanum* nuclei in G1 + S + G2 to be sorted. b. DAPI or PI stained *Solanum* nuclei from samples throughout the workflow under epifluorescence microscope: i. nuclei suspension after flow sorting, ii. pelleted nuclei after centrifugation at 500 g for 30 minutes, iii. pellet mixed with LMP agarose, iv. slice from a plug with nuclei already embedded. Scale bars represent 100 µm and apply to all panels in the figure.

2.1.2 Stain the nuclei with Propidium Iodide (PI, Sigma Aldrich, P4170) (final concentration: 50 µg.mL-1) in the dark for at least 10 min prior to the flow cytometry measurements.

2.2.2 Sort PI-stained nuclei using a FACSVantage cytometer operated with these settings:

a) Argon-ion Innova 304 Laser (Coherent, USA) (488 nm, 100 mW); b) 70 µm nozzle; c) sorting speed 500 events/s in counter sort mode.

We performed data acquisition and analysis with the CellQuest software (BD Biosences, Santa Clara, USA)

2.3.2 Select PI-stained G1, S and G2 nuclei populations for sorting using PI-A vs PI-W dot plots, avoiding the inclusion of debris (Figure 1a).

In both sorting platforms, use a 50 mM NaCl solution in MQ (milli-Q, Millipore Corporation) water as sheath fluid. Collection tubes for the sorted nuclei contain 400-500 µL of ice-cold 1.5x IB. This volume should be equal to the volume that comes with the sorted fraction and depends on the sorted-droplet volume and number of sorted nuclei. Keep the samples as well as collection tubes at 4° C during sorting using a precision refrigeration unit (± 0.2° C) connected to the flow sorter.

3. Agarose plugs preparation and quality controls

Plug preparation was performed following Šimková et al. (2003). We modified the centrifugation steps (speed and time) to make them more efficient to recover *Solanum* nuclei.

3.1 Pellet nuclei (400,000-500,000 per tube) at 500 g and 4° for 30 min.

3.2 Two fluorescence microscopic checkpoints can be optionally introduced, before and after pelleting, for nuclei integrity and concentration after staining with 2 µg.mL-1 DAPI or 50 µg.mL-1 PI (Figure 1b).

3.3 Discard supernatant keeping 15 µL in the tube with the pellet. Gently resuspend pelleted nuclei in the leftover supernatant.

3.4 Warm the nuclei suspension and keep at 52° for 3-5 min

3.5 Mix with 8.5 µL 2 % low-melting point (LMP) agarose (Bio-Rad, 1613111) dissolved in 1x IB.

3.6 Keep the mixture at 52°C for another 5 min and then slowly pour the mixture using a wide bore pipette tip in pre-warmed plug molds (Bio-Rad, 1703713).

3.7 Solidify the plugs at 4°C for 10 min

3.8 Push the plugs into a polystyrene tube with 500 µL per plug of lysis buffer C (0.5 M EDTA, 1 % N-lauroyl-sarcosine (Sigma Aldrich, L5000)). Add freshly prepared proteinase K (Sigma Aldrich, P6556) to reach a final concentration of 0.3 mg/mL (30 µL of stock solution per 1 mL). For the two washes, prepare 62 µL of stock solution per 1 mL of buffer. Weigh 62 µg and dilute in 62 µL of MQ (milli-Q, Mil-

lipore Corporation) water. Before weighing, let the proteinase K adapt to room temperature to avoid condensation on the powder. Add 30 µL for the first wash and keep the rest at 4° C for the second wash.

3.9 Incubate at 37° C for 24 h under gentle shaking (50 rpm) in an almost horizontal position.

3.10 Change for lysis buffer B (0.5 M EDTA, 1 % N-lauroyl-sarcosine (Sigma Aldrich, L5000), 0.3 mg.mL-1 proteinase K (Sigma Aldrich, P6556, pH 8.0) and incubate plugs for another 24 h under the same conditions.

3.11 After the proteinase K treatment, rinse and store agarose plugs in ET buffer (1 mM Tris, 50 mM EDTA, pH 8.0) at 4° C.

3.12 Check DNA quality using pulsed field gel electrophoresis (e.g. CHEF-DR II system, Bio-Rad or BluePippin, Sage Science) (Figure 2a and b).

The plugs obtained were ready for RNAse treatment, DNA release and labeling following the standard protocol recommended by the genome mapping platform manufacturer (BioNano Genomics).

Assessment of DNA quality

We included various quality checkpoints throughout the workflow. The integrity of nuclei was checked before and after flow sorting. Fluorescence microscopy revealed intact nuclei, with regular shape and a suitable density for isolating DNA, evenly distributed with about 1500-2000 nuclei per mm^2 at 10x magnification (Figure 1b, i and ii).

Following nuclei embedding in agarose plugs we checked nuclei features again (Figure 1b, iii and iv). High density of round, regular shaped nuclei was obtained. DNA quality and size after the proteinase K treatment was checked through Pulsed-Field Gel Electrophoresis (PFGE) (Figure 2a). DNA was protein-free (no fluorescence in the slots) and

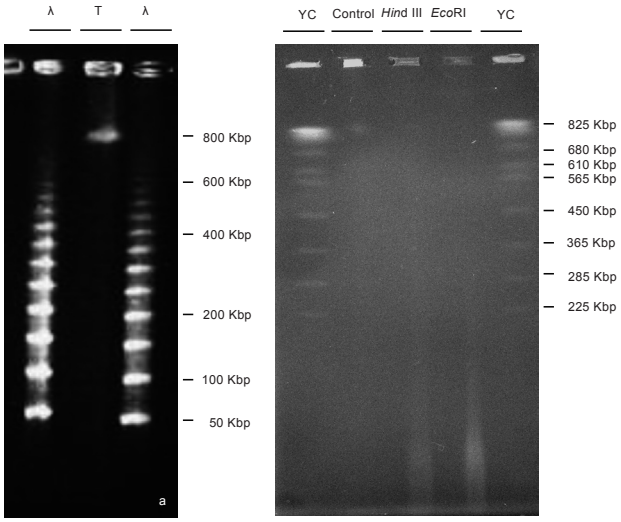


Figure 2: a. Quality check of *Solanum lycopersicum* HMW DNA (h) by PFGE. λ - Lambda Ladder, unit size 48.6 kb. b. PFGE of restriction enzyme digestion of *Solanum commersonii* HMW DNA for accessibility test (cropped image). YC- yeast chromosomes, Control - HMW DNA, 20 min at 37° C in digestion buffer without restriction enzymes, HindIII and EcoRI, 2 U for 20 min at 37° C.

with molecule size equal to or larger than 800 Kbp. DNA accessibility for enzymes was checked by digestion with restriction enzymes (HindIII and EcoRI, 2 U in Digestion buffer (DB, 1x enzyme buffer, 1mM DTT, 4 mM spermidine, 0.39 mg BSA) for 20 min at 37° C). DNA was readily accessible for restriction enzyme digestion even at low concentrations, confirming its suitability for physical mapping (Figure 2b).

Plugs with highly quality DNA were RNAse treated and DNA was released from the plugs and labelled following the standard protocol recommended by BioNano Genomics. The labeled DNA was imaged on the Irys platform (Figure 3a and b). Taking tomato cv Heinz 1706 as an example, we found single molecule N50 lengths of 290 kb and DNA quality allowing a labeling density of 7.7 sites per 100 Kbp from 12 sites per 100 Kbp predicted in silico. Additionally, there was no clogging of the chips, thus allowing a throughput of 1.3 Gbp per scan (size-filtered molecules > 150 Kbp) (Figure 3a and b).

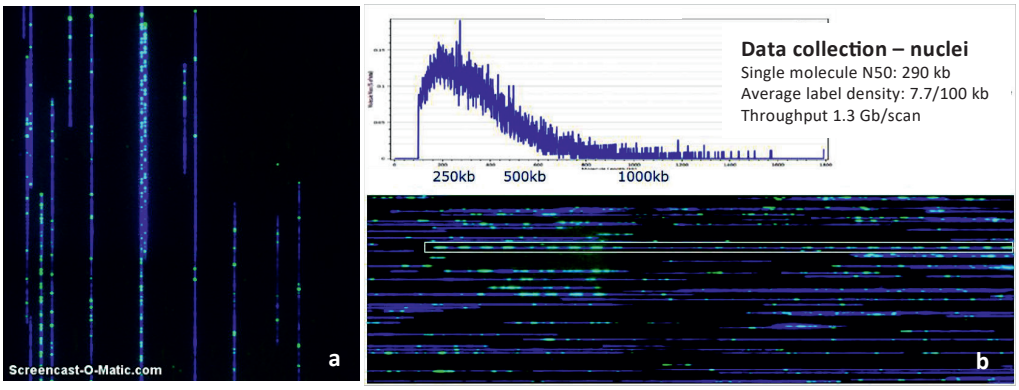


Figure 3: a. Snapshot of linearized labelled DNA fragments running in nanochannels for *Solanum lycopersicum*. b. Size distribution of molecules and N50 stat for *S. lycopersicum*. The white box highlights a tandem repeat of approximately 200 kbp.

Advantages over comparable methods

Production of HMW DNA of superior quality, i.e., molecules of several hundred kilobases, has been identified as the bottleneck in nanochannel-based genome mapping technologies (BioNano Genomics). Our method yields the longest DNA molecules on average (290 kbp, compared to the 160 Kbp molecules and 7.5 sites per 100 Kbp obtained through other methods, unpublished results). Run costs become lower, as longer molecules mean that less coverage is needed when doing genome mapping. Moreover, longer molecules with proper labeling frequency imply higher N50 in the consensus genome map. Complex regions in the genome can be spanned and resolved better, which contributes to the contiguity of the assembly.

Contamination by cytosol inclusions in different plant cell types represents another important impediment to the use of optical mapping. Such cytoplasmic compounds, mostly phenolic compounds, polysaccharides and other secondary metabolites co-precipitate with DNA and interfere with enzymes used in DNA labeling. When nuclei are

flow-sorted, the contamination with such substances is minimized, and so ultimately leads to fewer false negatives in optical mapping. Purity also reduces clogging of nanochannels, thus allowing for higher throughput since it extends chip lifetime. In terms of cytoplasmic contaminants, flow sorting yields HMW DNA of much higher purity compared to other methods, without cell debris and very low amounts of chloroplast and mitochondria contamination [5], which could represent a problem because in large leaves the amount of chloroplast DNA is often much higher than that of nuclear DNA [6].

The yield of pure, HMW DNA from the tomato cv Heinz 1706 material (genome size 950 Mbp) amounted about 1.6 x 106 nuclei in four agarose miniplugs, which was enough to produce 73 Gbp of size-filtered data with single molecule N50 of 290 Kbp on the Irys platform. We obtained this sample of nuclei in 3 hours of sorting in a FACSAria or 6 hours of sorting in a FACSVantage flow cytometer. These results confirm the excellent quality of nuclear plant DNA obtained through flow sorting which is similar to the quality obtained from mammalian cell cultures. Moreover, molecule size distribution was on par with human samples (personal communication, BioNano Genomics).

Implications for research and breeding

Flow sorted nuclei provide a good starting point for mapping and sequencing technologies where high purity and megabase-sized DNA is required. The protocol [1] that we modified was originally developed for construction of BAC libraries, but is equally suitable for optical mapping. We applied our optimized method to *Solanum* crops and wild relatives, in order to further improve the quality of genome sequencing and assembly, and for comparative structural genomics including related crops and wild relatives. We introduced relevant modifications that enhanced both efficiency and versatility of this method. The main adjustments are related to the use of different source materials (root tips or young leaves from seedlings or *in vitro* plantlets), different methods to obtain nuclei suspensions (homogenization or chopping), two DNA-specific fluorescence dyes (DAPI and PI) with their corresponding lasers (UV and 488 nm or 514 nm) and both classic and modern sorting platforms (FACS Vantage or Aria). With these modifications, we expect that the method is also successful in different sorting platforms and laser configurations, meaning that laboratories without access to the latest flow sorting technology still can have access to next-generation mapping. One point of extra attention when following this protocol using sorting platforms is that only standard UV lasers and DAPI can be used, as the formaldehyde in the fixed nuclei interferes with PI fluorescence. However, the histograms obtained in this work were clear and well defined, and had low CVs (coefficients of variation – ratio of the standard deviation to the mean) of DNA peaks. The possibility to use either mechanical homogenization or manual chopping allows for flexibility depending on the plant species. For example, nuclei from *S. commersonii* leaves better preserve their integrity and carry less debris when obtained by chopping with a

razor blade than using a Polytron whereas this kind of homogenizer is ideal for tomato root material.

In the case of elite or proprietary breeding material, seed propagation is often not possible or it is necessary to preserve the genotype to be analyzed so the only way to obtain enough material is through vegetative propagation. It would not be possible to isolate HMW DNA from this kind of materials from embryonic root tips. The modifications included in this method allow for isolation of nuclei from young leaf material, thus enabling the use of genome mapping for breeding lines that are propagated vegetatively.

In conclusion, the workflow proposed here involving the coupling of flow sorting with nanochannel-based mapping will allow this genome mapping technology to fulfill its potential in plant genomics and genomics-based breeding.

Acknowledgements:

Research was supported by: Comisión Sectorial de Investigación Científica, University of the Republic, Uruguay (grant code CSIC I+D 2012 383) and INIA (Línea 4). P. Gaiero was supported by grant Proyecto 720 and CSIC Recursos Humanos, University of the Republic. We are grateful to Alicia Castillo and technical staff at INIA for kindly providing *in vitro* plant material. We are also thankful to Bertus van der Laan and technical staff at UNIFARM (WUR) for kindly providing Heinz tomato seeds and taking care of the plants in the greenhouse. We wish to thank María Inés Siri and Rodrigo Achigar for kindly allowing us to use the PFGE equipment in the Microbiology Lab, Faculty of Chemistry, University of the Republic.

Supplementary material and/or Additional information:

High-throughput NGS technologies have enabled the de novo sequencing of an increasing number of plant species. However, nearly-finished well-assembled genomes are not easy to obtain. Issues related to order and orientation of contigs and distribution of repetitive sequences remain major challenges [7]. Genome sequences from non-model species, orphan crops or even main crops with larger or more complex genomes are still far from finished. Genome studies are lacking structural comparisons, since the focus of most resequencing efforts has been on SNP variation and, at best, on microsynteny [8]. Among several developments that aim to facilitate genome assembly (such as chromatin conformation capture or Hi-C [9] or Chicago libraries by Dovetail Genomics [10]), the new next-generation genome mapping technologies (BioNano Genomics Irys) [11,12] have provided significant improvements across a broad range of organisms. They can improve assembly metrics such as N50 or percentage of whole genome assembled, by sizing and/or closing gaps, scaffolding, joining scaffolds, correcting assembly errors and even identifying, spanning and assembling repeated sequences. In addition, genome mapping can

in its own right provide a comprehensive assay system for defining structural variation among related species or genotypes within a species [12].

The nanochannel-based genome mapping technology has been described extensively [11,12]. This technology uses nicking enzymes to create single strand DNA sequence-specific cuts that are subsequently labelled by a fluorescent nucleotide analog upon repair of the nicks by a DNA polymerase [13]. The nick-labeled DNA is stained with the intercalating dye YOYO-1, loaded onto the nanofluidic chip by an electric field, and imaged with high N.A. optics and a CCD camera. The DNA is linearized by confinement in a nanochannel array [14], resulting in uniform linearization and allowing precise and accurate measurement of the distance between nick-labels on DNA molecules comprising a signature pattern. Also, the DNA loading and imaging cycle can be repeated many times in a completely automated fashion; data can be obtained at high throughput and high resolution [15]. It builds on the earlier optical mapping technologies overcoming many of their limitations, particularly in terms of throughput, resolution and precision of distance measurements [15,16].

Despite all these advantages, nanochannel-based genome mapping has been used only recently for the assembly of DNA in higher plants such as spinach [17], subterranean clover [18], maize [19], quinoa [20] and bread wheat [21], with HMW DNA isolation in most cases as the bottleneck for its application. Previously, a related method called optical mapping [22,23] was used for whole genome analysis in crops like rice [24], maize [25] and tomato [7] and for crop relatives such as *Medicago truncatula* [26,27]. It has also been applied to validate assembly of a 2.1-Mb prolamin gene family region from the genome of *Aegilops tauschii* [15] and more recently to evaluate the quality of the whole genome hybrid assembly from this wheat progenitor [28].

Setting aside issues of genome size and complexity and computational limitations, one of the main bottlenecks for the application of nanochannel-based genome next-generation mapping to plant genomes is the requirement of high quality HMW DNA. Such DNA is easier to obtain from mammalian cells than from plant cells, because of many important differences in their composition. The rigid cell walls in plant cells demand for mechanical methods to disrupt them, which can cause shearing of the DNA. There are various contaminants in plants that are not found in mammalian cells, such as chloroplasts and a range of secondary metabolites which contaminate the DNA sample during the precipitation process [29]. Sometimes plants rich in secondary metabolites are the most interesting from the point of view of breeding, since these metabolites might be the breeding target and/or confer resistance to pests and diseases [29]. It was reported previously that nuclei and chromosomes purified by flow cytometric sorting provide quality HMW DNA even in species rich in secondary metabolites [1,30]. Finally, the higher prevalence of polyploidy in plants affects DNA yield.

References:

[1] H. Šimková, J. Číhalíková, J. Vrána, M.A. Lysak, J. Dolezel. Preparation of HMW DNA from plant nuclei and chromosomes isolated from root tips. Biol. Plant., 46 (2003), pp. 369–373. http://link.springer.com/article/10.1023/A:1024322001786

[2] J. Doležel, J. Vrána, J. Safář, J. Bartoš, M. Kubaláková, H. Simková. Chromosomes in the flow to simplify genome analysis. Funct. Integr. Genomics, 12 (2012), pp. 397–416. doi:10.1007/s10142-012-0293-0

[3] J. Doležel, J. Vrána, P. Cápal, M. Kubaláková, V. Burešová, H. Simková. Advances in plant chromosome genomics. Biotechnol. Adv., 32 (2014), pp. 122–136. doi:10.1016/j.biotechadv.2013.12.011

[4] J. Vrána, H. Simková, M. Kubaláková, J. Číhalíková, J. Doležel. Flow cytometric chromosome sorting in plants: the next generation. Methods, 57 (2012), 331–337. doi:10.1016/j.ymeth.2012.03.006

[5] J. Šafář, J.J.C. Noa-Carrazana, J. Vrána, J. Safár, J.J.C. Noa-Carrazana, J. Vrána, J. Bartos, O. Alkhimova, X. Sabau, H. Simková, F. Lheureux, M.-L. Caruana, J. Dolezel, P. Piffanelli. Creation of a BAC resource to study the structure and evolution of the banana (Musa balbisiana) genome. Genome, 47 (2004), pp. 1182–1191. doi:10.1139/G04-062

[6] C. Jope, A. Hirai, S. Wildman. Evidence that the amount of chloroplast DNA exceeds that of nuclear DNA in mature leaves. J. Cell Biol., 79 (1978), pp 631–636

[7] L. A. Shearer, L.K. Anderson, H. de Jong, S. Smit, J.L. Goicoechea, B. a Roe, A. Hua, J.J. Giovannoni, S.M. Stack. Fluorescence in situ hybridization and optical mapping to correct scaffold arrangement in the tomato genome. G3 Genes, Genomes, Genet., 4 (2014), pp. 1395–1405. doi:10.1534/g3.114.011197

[8] W. Sanseverino, E. Hénaff, C. Vives, S. Pinosio, W. Burgos-Paz, M. Morgante, S. Ramos-Onsins, J. Garcia-Mas, J. Casacuberta. Transposon insertion, structural variations and SNPs contribute to the evolution of the melon genome. Mol. Biol. Evol., 32 (10) (2015), pp. 2760-2774. doi:10.1093/molbev/msv152

[9] J.-M. Belton, R.P. McCord, J.H. Gibcus, N. Naumova, Y. Zhan, J. Dekker. Hi–C: A comprehensive technique to capture the conformation of genomes Methods, 58 (2012), pp. 268–276. doi:10.1016/j.ymeth.2012.05.001

[10] N.H. Putnam, B.L. O'Connell, J.C. Stites, B.J. Rice, M. Blanchette, R. Calef, C.J. Troll, A. Fields, P.D. Hartley, C.W. Sugnet, D. Haussler, D.S. Rokhsar, R.E. Green. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage Genome Res., 26 (2016), pp. 342–350 doi:10.1101/gr.193474.115

[11] E.T. Lam, A. Hastie, C. Lin, D. Ehrlich,S.K. Das, M.D. Austin, P. Deshpande, H. Cao, N. Nagarajan, M. Xiao, P.-Y. Kwok. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat. Biotechnol., 30 (2012), pp. 771–776. doi:10.1038/nbt.2303.

[12] H. Cao, A.R. Hastie, D. Cao, E.T. Lam, Y. Sun, H. Huang, X. Liu, L. Lin, W. Andrews, S. Chan, S. Huang, X. Tong, M. Requa, T. Anantharaman, A. Krogh, H. Yang, H. Cao, X. Xu. Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. Gigascience, 3 (2014), p. 34. doi:10.1186/2047-217X-3-34

[13] M. Xiao, A. Phong, C. Ha, T.-F. Chan, D. Cai, L. Leung, E. Wan, A.L. Kistler, J.L. DeRisi, P.R. Selvin, P.-Y. Kwok. Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Res., 35 (2007), p. e16. doi:10.1093/nar/gkl1044

[14] S.K. Das, M.D. Austin, M.C. Akana, P. Deshpande, H. Cao, M. Xiao. Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. Nucleic Acids Res., 38 (2010), p. e177. doi:10.1093/nar/gkq673

[15] A.R. Hastie, L. Dong, A. Smith, J. Finklestein, E.T. Lam, N. Huo, H. Cao, P.-Y. Kwok, K.R. Deal, J. Dvorak, M.-C. Luo, Y. Gu, M. Xiao. Rapid genome mapping in nanochannel arrays for highly complete and accurate de novo sequence assembly of the complex Aegilops tauschii genome. PLoS One, 8 (2013), p. e55864. doi:10.1371/journal.pone.0055864

[16] R. Appels, J. Nystrom, H. Webster, G. Keeble-Gagnere. Discoveries and advances in plant and animal genomics Funct. Integr. Genomics, 15 (2015), pp. 121–129. doi:10.1007/s10142-015-0434-3

[17] C. Xu, C. Jiao, H. Sun, X. Cai, X. Wang, C. Ge, Y. Zheng, W. Liu, X. Sun, Y. Xu, J. Deng, Z. Zhang, S. Huang, S. Dai, B. Mou, Q. Wang, Z. Fei, Q. Wang. Draft genome of spinach and transcriptome diversity of 120 *Spinacia* accessions. Nat. Commun., 8 (2017), p. 15275. doi:10.1038/ncomms15275

[18] P. Kaur, P.E. Bayer, Z. Milec, J. Vrána, Y. Yuan, R. Appels, D. Edwards, J. Batley, P. Nichols, W. Erskine, J. Doležel. An advanced reference genome of Trifolium subterraneum L. reveals genes related to agronomic performance. Plant Biotechnol. J., 15 (8) (2017), pp. 1034–1046. doi:10.1111/pbi.12697

[19] Y. Jiao, P. Peluso, J. Shi, T. Liang, M.C. Stitzer, B. Wang, M. Campbell, J.C. Stein, X. Wei, C.-S. Chin, K. Guill, M. Regulski, S. Kumari, A. Olson, J. Gent, K.L. Schneider, T.K. Wolfgruber, M. May, N. Springer, E. Antoniou, R. McCombie, G.G. Presting, M. McMullen, J. Ross-Ibarra, R.K. Dawe, A. Hastie, D.R. Rank, D. Ware. Improved maize reference genome with single-molecule technologies. Nature, (2017). doi:10.1038/nature22971

[20] D.E. Jarvis, Y.S. Ho, D.J. Lightfoot, S.M. Schmöckel, B. Li, T.J.A. Borm, H. Ohyanagi, K. Mineta, C.T. Michell, N. Saber, N.M. Kharbatia, R.R. Rupper, A.R. Sharp, N. Dally, B.A. Boughton, Y.H. Woo, G. Gao, E.G.W.M. Schijlen, X. Guo, A.A. Momin, S. Negrão, S. Al-Babili, C. Gehring, U. Roessner, C. Jung, K. Murphy, S.T. Arold, T. Gojobori, C.G. van der Linden, E.N. van Loo, E.N. Jellen, P.J. Maughan, M. Tester. The genome of *Chenopodium quinoa*. Nature., 542 (2017), pp. 307–312. doi:10.1038/nature21370

[21] H. Staňková, A.R. Hastie, S. Chan, J. Vrána, Z. Tulpová, M. Kubaláková, P. Visendi, S. Hayashi, M. Luo, J. Batley, D. Edwards, J. Doležel, H. Šimková. BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. Plant Biotechnol. J., 14 (2016), pp. 1523–1531. doi:10.1111/pbi.12513

[22] S. Zhou, J. Herschleb, D.C. Schwartz. Chapter 9. A Single Molecule System for Whole Genome Analysis. Perspect. Bioanal., 2 (2007), pp. 265–300. doi:10.1016/S1871-0069(06)02009-X

[23] E.T. Dimalanta, A. Lim, R. Runnheim, C. Lamers, C. Churas, D.K. Forrest, J.J. De Pablo, M.D. Graham, S.N. Coppersmith, S. Goldstein, D.C. Schwartz. A Microfluidic System for Large DNA Molecule Arrays. Anal. Chem., 76 (18) (2004), pp. 5293–5301. doi:10.1021/ac0496401.

[24] S. Zhou, M.C. Bechner, M. Place, C.P. Churas, L. Pape, S. a Leong, R. Runnheim, D.K. Forrest, S. Goldstein, M. Livny, D.C. Schwartz. Validation of rice genome sequence by optical mapping
BMC Genomics., 8 (2007), p. 278. doi:10.1186/1471-2164-8-278

[25] S. Zhou, F. Wei, J. Nguyen, M. Bechner, K. Potamousis, S. Goldstein, L. Pape, M.R. Mehan, C. Churas, S. Pasternak, D.K. Forrest, R. Wise, D. Ware, R. a Wing, M.S. Waterman, M. Livny, D.C. Schwartz. A single molecule scaffold for the maize genome. PLoS Genet., 5 (2009), p. e1000711. doi:10.1371/journal.pgen.1000711

[26] N.D. Young, F. Debellé, G.E.D. Oldroyd, R. Geurts, S.B. Cannon, M.K. Udvardi, V. a. Benedito, K.F.X. Mayer, J. Gouzy, H. Schoof, Y. Van de Peer, S. Proost, D.R. Cook, B.C. Meyers, M. Spannagl,

F. Cheung, S. De Mita, V. Krishnakumar, H. Gundlach, S. Zhou, J. Mudge, A.K. Bharti, J.D. Murray, M. a. Naoumkina, B. Rosen, K. a. T. Silverstein, H. Tang, S. Rombauts, P.X. Zhao, P. Zhou, V. Barbe, P. Bardou, M. Bechner, A. Bellec, A. Berger, H. Bergès, S. Bidwell, T. Bisseling, N. Choisne, A. Couloux, R. Denny, S. Deshpande, X. Dai, J.J. Doyle, A.-M. Dudez, A.D. Farmer, S. Fouteau, C. Franken, C. Gibelin, J. Gish, S. Goldstein, A.J. González, P.J. Green, A. Hallab, M. Hartog, A. Hua, S.J. Humphray, D.-H. Jeong, Y. Jing, A. Jöcker, S.M. Kenton, D.-J. Kim, K. Klee, H. Lai, C. Lang, S. Lin, S.L. Macmil, G. Magdelenat, L. Matthews, J. McCorrison, E.L. Monaghan, J.-H. Mun, F.Z. Najar, C. Nicholson, C. Noirot, M. O'Bleness, C.R. Paule, J. Poulain, F. Prion, B. Qin, C. Qu, E.F. Retzel, C. Riddle, E. Sallet, S. Samain, N. Samson, I. Sanders, O. Saurat, C. Scarpelli, T. Schiex, B. Segurens, A.J. Severin, D.J. Sherrier, R. Shi, S. Sims, S.R. Singer, S. Sinharoy, L. Sterck, A. Viollet, B.-B. Wang, K. Wang, M. Wang, X. Wang, J. Warfsmann, J. Weissenbach, D.D. White, J.D. White, G.B. Wiley, P. Wincker, Y. Xing, L. Yang, Z. Yao, F. Ying, J. Zhai, L. Zhou, A. Zuber, J. Dénarié, R. a. Dixon, G.D. May, D.C. Schwartz, J. Rogers, F. Quétier, C.D. Town, B.A. Roe. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature, 480 (2011), pp. 520-524. doi:10.1038/nature10625

[27] H. Tang, V. Krishnakumar, S. Bidwell, B. Rosen, A. Chan, S. Zhou, L. Gentzbittel, K.L. Childs, M. Yandell, H. Gundlach, K.F.X. Mayer, D.C. Schwartz, C.D. Town. An improved genome release (version Mt4.0) for the model legume Medicago truncatula. BMC Genomics, 15 (2014), p. 312. doi:10.1186/1471-2164-15-312

[28] A.V. Zimin, D. Puiu, M.-C. Luo, T. Zhu, S. Koren, J.A. Yorke, J. Dvorak, S. Salzberg. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res, 27 (2017), pp. 787-792. doi: 10.1101/gr.213405.116[29]

[29] M. Wink. Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. TAG Theor. Appl. Genet., 75 (1988), pp. 225–233

[30] E. Barghini, L. Natali, T. Giordani, R.M. Cossu, S. Scalabrin, F. Cattonaro, H. Šimková, J. Vrána, J. Doležel, M. Morgante, A. Cavallini. LTR retrotransposon dynamics in the evolution of the olive (*Olea europaea*) genome. DNA Res., 22 (2015), pp. 91–100. doi:10.1093/dnares/dsu042

"Tienes destellos de arados
de manos con callo y lodo
y eres cual un tesoro
para el labriego que sueña."
Salvador Cerpa
(In: "Embajadora de sueños", 2008)

Chapter 8

General discussion

In my thesis I have presented an overview of genetic, cytogenetic and genomic characteristics of *Solanum commersonii* or Commerson's wild potato, a close relative of potato and native to Uruguay and neighbouring countries. The work reflects a true evolution in scientific approaches making use of the progressive development of genomic technologies during the last decade. The basic idea is to create scientific tools to distinguish the genomes of *S. commersonii* and *S. tuberosum* (cultivated potato), along with *S. chacoense*, and trace their genomes in introgressive hybridization breeding programmes. In the following paragraphs I will highlight the main achievements in my search for genetic, cytogenetic, and genomic instruments in this fast evolving and exciting field of biology.

The first step involved the test for suitability of introgressive hybridization schemes in F1 hybrids. To this end a combination of strategies was used to overcome hybridization barriers and incongruities between *S. commersonii* and potato. While diploid *S. commersonii* has an Endosperm Balance Number (EBN) of 1, tetraploid cultivated potato has 4 EBN (Johnston et al. 1980). Prior to my work, a bridge cross strategy was used to hybridize *S. commersonii* clone 04.02.3 from south-west Uruguay to a diploid *S. tuberosum* from Group Phureja, which has 2EBN. Clone 04.02.3 was selected because of its moderate resistance to bacterial wilt caused by *Ralstonia solanacearum* and its ability to produce unreduced (2n) gametes. It was used as female parent to produce F1 hybrids with diploid *S. tuberosum* Group Phureja. Because of the EBN barrier, the only viable seeds are those produced through fertilization of a 2n egg cell from *S. commersonii* (2n = 2x = 24, 1EBN) and reduced pollen from a diploid from Group Phureja (n = x = 12, 1EBN) (González 2010). The resulting F1 hybrid progeny, though not numerous, is certain to have retained the maximum diversity possible, coming from the fertilization of two meiotic products. These F1 hybrids are obviously triploid, with two chromosome sets coming from the wild *S. commersonii* and one chromosome set from *S. tuberosum* Group Phureja (Gaiero et al. 2017, Chapter 3). In the first backcross, triploid hybrids that retained the ability to produce unreduced eggs (2n = 3x = 36, 2EBN) were used as female parents, producing viable seed with tetraploid potato (n = 2x = 24, 2EBN). The resulting 5x offspring were backcrossed to different tetraploid potato genotypes to avoid inbreeding depression. The backcross progenies then obtained were screened and selected for their resistance to bacterial wilt together with other agronomic and yield traits (Andino et al. in prep.).

The fate and pairing behaviour of alien chromosomes in a potato background.

Although the barriers to interspecific hybridization were overcome, the advanced lines derived from it have to recover the euploid condition bearing only the wild chromatin with the desirable traits. This can only be accomplished if homoeologous recombination is possible in the regions of interest. As loss of wild parent chromosomes was expected in the backcrossing process, I was ignorant as to the fate of these chromosomes in the subsequent generations. To this end, I presented in Chapter 3 the chromosome countings of offspring individuals and meiotic analyses, showing random pairing of the ho-

moeologous *S. commersonii* genomes and *S. tuberosum* Group Phureja chromosomes in the triploid hybrids, while chromosome transmission and segregation in further meiotic stages were fairly balanced. I also followed the fate of specific chromosomes in these hybrids and backcross derivatives with a combination of BAC FISH and genome painting (GISH). There were no obvious differences in fluorescence signals between the homoeologues suggesting that repetitive sequences did not diverge much between the parental species, and hence, I could not provide tools to select against wild chromatin in the advanced materials to be used as breeding lines nor pinpoint the chromosomal exchanges between homoeologues. However, the result did point at the low degree of genomic divergence between the species involved in terms of repetitive sequences and therefore on their potential for homoeologous pairing. Knowing that their chromosome pairing is saturated, I looked at the factors that potentially determine this chromosome pairing, namely homology in repetitive sequences and homology in chromosome structure.

Comparative repetitive sequences analysis between the potato and the tomato clades

In chapter 5 I looked at the repetitive fraction of the genomes of potato and its wild relatives compared to tomato and its wild relatives. Many reasons motivated this approach. There were studies with conspicuous differentiation of GISH painting in hybrids between tomato and relatives (Parokonny et al. 1997; Ji and Chetelat 2003; Ji et al. 2004), whereas various reports claim structural variation between tomato and wild species (Seah et al. 2004; Van Der Knaap et al. 2004; Anderson et al. 2010; Verlaan et al. 2011). Meanwhile, there are no data of such rearrangements between potato and its relatives (Lou et al. 2010; Gaiero et al. 2016, Chapter 4), suggesting different patterns of genome differentiation in these groups of species. To elucidate if the differences in genome differentiation observed among the tomato and potato clades of genus *Solanum* can be correlated to different dynamics in repetitive elements, my colleagues and I decided to compare the repetitive fractions of cultivated and wild species belonging to those clades, including the outgroup *Solanum etuberosum*. Short read sequence data were available for various species within both clades (http://www.tomatogenome.net, Aflitos et al. 2014; http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml, Potato Genome Sequencing Consortium 2011), that supplemented our sequence data for diploid potato relatives as well as for *S. etuberosum*. Representatives from both clades were chosen and their sequence data subsampled proportionally to genome sizes to emulate shallow sequencing. Using a clustering approach, we found that the classes, families and lineages of repetitive elements across the clades are largely conserved, but their abundances are different and the repeat profiles allowed us to discriminate the two clades. We also found that the repeat content of *S. etuberosum* is more similar overall to the potato clade. We observed that repeats are correlated to genome size and we could find preliminary evidence that

they are shaping genome differentiation in the tomato clade through a clade-specific balance of processes responsible for the amplification or removal of transposable elements. The identified taxa-specific repeats are likely constituting a great portion of sequences that allowed GISH to discriminate between parental species in the hybrids. All these tools provide information about possible crossability among these species in terms of their genome similarity. It remained to be clarified whether the chromosome structure was collinear between potato and its wild relatives.

Collinearity of *S. commersonii* and *S. chacoense* with respect to the cytogenetic map of *S. tuberosum*

Once it was clear that homoeologous chromosomal exchange was possible between potato and the wild relatives under study, the next question was whether introgression of specific regions of interest could take place with as little chromatin belonging to the wild species as possible. Addressing this issue in chapter 4 helped us to anticipate the existence of either suppression of recombination (Verlaan et al. 2011) or linkage drag (Jacobsen and Schouten 2007), which could hamper the introgression of only the desirable traits into a potato background and thus render the ultimate aim unachievable, unless the material linked to the desirable traits has neutral effects in the phenotypes of the offspring. I found that *S. commersonii* and *S. chacoense* are collinear with cultivated *S. tuberosum* on the whole chromosome scale. This study revealed a clear correspondence between the chromosomes of *S. commersonii* and *S. chacoense* to those in cultivated potato. BAC probes could successfully be hybridized to the *S. commersonii* and *S. chachoense* pachytene chromosomes, confirming their correspondence with linkage groups in RH potato. All BACs were in the same linear order. Microscopic distances between the BAC signals on the chromosomes were quantified and compared; some differences found suggest either small-scale rearrangements or reduction/amplification of repetitive sequences. No major rearrangements have been found that would be expected to hamper recombination and introgression or that could cause linkage drag or erratic segregation of homoeologues during meiosis in hybrids, making these amenable species for efficient introgressive hybridization breeding. However, the small-scale differences that were suspected from both distances between BAC signals and in from the pairing breakpoints in the pachytene trivalents of triploid F1 hybrids suggested that it was necessary to look at synteny at the sequence level. With the need to assess collinearity at a fine scale there was also a need to acquire expertise in analysing genome sequences and to move from cytogenetics into genomics.

Comparative genome homology/divergence at the sequence level between *S. commersonii*, *S. tuberosum* and *S. chacoense*

Microsynteny and small-scale rearrangements between potato and its wild relatives can only be detected at the sequence level (Peters et al. 2012) and their results can help to predict the success of such related species for their use in introgressive hybridization. It will also allow the use of molecular markers that have been developed for potato in their study and exploration (Milbourne et al. 1998; Ghislain et al. 2004; Feingold et al. 2005; Ghislain et al. 2009; Pérez-de-Castro et al. 2012; Warschefsky et al. 2014; Bethke et al. 2017; Gaiero et al. 2018, Chapter 2). The main challenge to be addressed in chapter 6 is the sequencing and assembly of a highly heterozygous (though fortunately diploid) plant, without relying on the other *Solanum* reference genomes for assembly, otherwise the chance for structural comparison will be cancelled. A strategy to solve this is to sequence haploid (or doubled-haploid) genotypes obtained through in vitro culture of microspores. I already had those haploid genotypes available (Castillo et al. 2016), derived from one of the resistant *S. commersoni*i plants selected for introgressive hybridization breeding (González 2010). I had to compromise between the financial resources available and the sequencing platforms that could be used, in order to get the best possible coverage and read length to make the assembly easier and of sufficient quality for comparative genome analyses. As platforms evolved and became cheaper, I was able to add long-read sequence data. However, it became necessary to integrate other sources of information to achieve a better assembly. Firstly, a genetic map was developed using SNP data generated through GBS from a biparental segregating population built by crossing contrasting genotypes for the traits to be introgressed. The genotyping data was integrated using Rabbit (Reconstructing Ancestry Blocks BIT by bit, Zheng et al. 2018, Biometris, WUR, ms in preparation) software. The genetic map thus obtained will be useful also to determine the genetic location of desirable traits found in *S. commersonii*. Then Allmaps (Tang et al. 2015b) was used, which is a bioinformatics tool developed to anchor assemblies to linkage maps. Using this combination of tools, I have achieved a far more contiguous assembly than the existing one (Aversano et al. 2015), anchored into 12 pseudomolecules built independently from any reference. The initial hybrid assembly presented in this thesis represented 85 % of the total genome of *S. commersonii*, with a modal size of scaffolds (N50) of 294 Kbp, ranging from 3.2 Mbp to 2.8 Kbp. These assembly metrics indicate that it is suitable for further improvement via integration of different sources of information such as genetic, cytogenetic and genome mapping data. With the use of anchoring techniques, my colleagues and I managed to integrate 601 scaffolds into the 12 pseudomolecules, with an anchor rate of only 38.65 %. This anchor rate will soon be further improved by using Bionano genome mapping (Lam et al. 2012; Hastie et al. 2013) and/or chromatin conformation capture technologies such as Hi-C (Belton et al. 2012) or Chicago libraries as implemented by Dovetail Genomics (Putnam et al. 2016) in the near future. Nevertheless, our anchored assembly was validated by abundant cover-

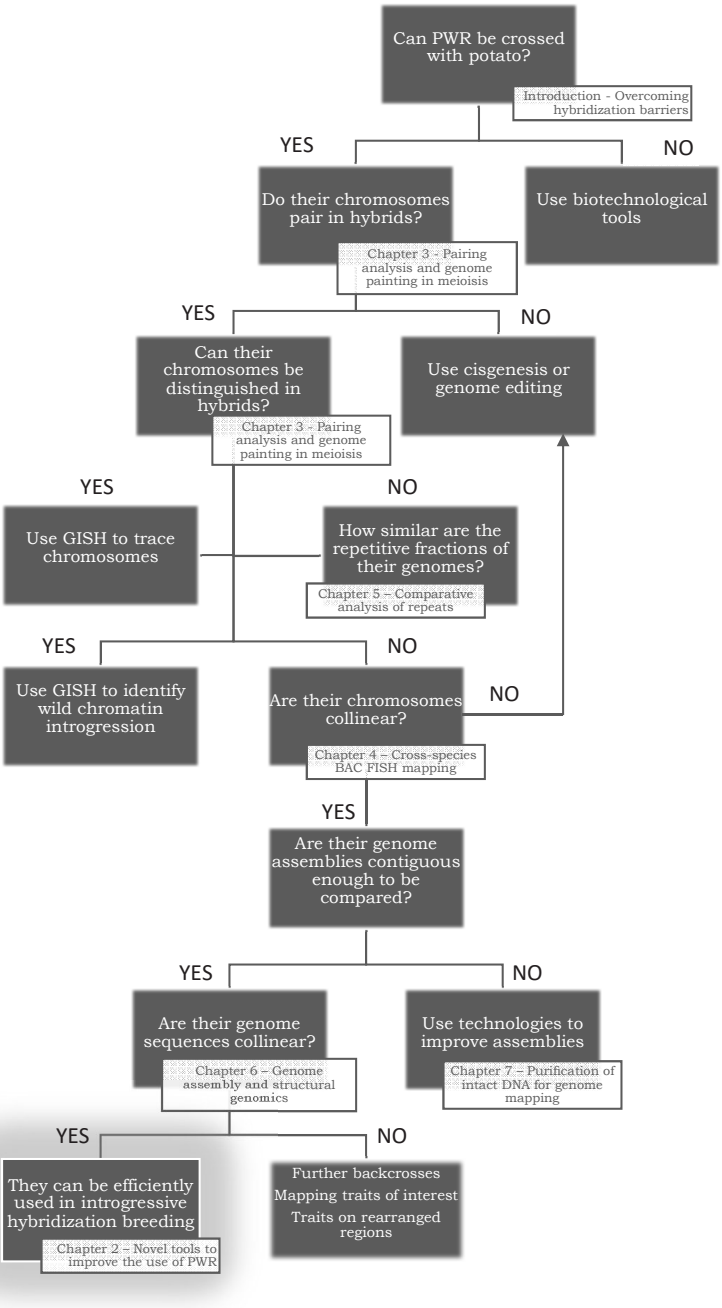


Figure 1. Flowchart of the decision-making process for the exploration and use of potato wild relatives (PWR) in breeding. GISH = Genomic *in situ* hybridization; BAC-FISH = Fluorescent *in situ* hybridization using Bacterial Artificial Chromosomes as probes.

age of the original reads and allowed me to perform preliminary structural comparisons with previously published genome assemblies.

The 12 pseudomolecules obtained for haploid *S. commersonii* showed high overall homology with the DM potato and M6 *S. chacoense* pseudomolecules. In some cases, the orientation of the whole pseudomolecule had to be corrected and in other cases the numbering of the linkage groups and corresponding pseudomolecules had to be changed to match the order assigned for the DM potato reference. When comparing the three genomes after these corrections, they show a high degree of collinearity at the whole genome level. Comparing each pseudomolecule separately, it is clear that the order of scaffolds is homologous among the three species. However, contiguity is lost in the pericentromere region, where scaffolds could not be anchored on the map because the genetic map had fewer markers to due to reduced recombination. Also in this region *Solanum* chromosomes display heterochromatin, which is repeat-rich and hinders assembly (Peters et al. 2009). Therefore, lack of collinearity is observed here. Small-scale rearrangements could not be conclusively distinguished from assembly or anchoring artifacts, with a few exceptions. In general terms, there is need for further improvement of contiguity and curation of the orientation of scaffolds in order to unequivocally detect rearrangements at the microsynteny scale.

Present views and future perspectives on the genomic relatedness between potato and PWR

We now know that there are a few microscale rearrangements between potato and its wild relatives, and that with a more contiguous stand-alone assembly we will be able to confirm or reject putative rearrangements identified so far. Once synteny breaks are confirmed, we need to find out if there are any coding sequences in the rearranged regions and if these can cause linkage drag around the traits of interest. In order to improve the assembly that we have obtained, we plan to integrate different sources of information (Tang et al. 2015a; Jiao et al. 2017; Udall and Dawe 2017). One of them is next-generation genome mapping through nanochannels (Bionano optical mapping (Lam et al. 2012; Hastie et al. 2013; Tang et al. 2015a). In order to use this platform, we need sufficient High Molecular Weight (HMW) nuclear DNA that is intact and pure. To achieve that, we developed important improvements to pre-existing methods (Šimková et al. 2003) to isolate HMW DNA from flow-sorted nuclei and we tested them on *Solanum* crops and wild relatives, as described in chapter 7. Using this workflow that we proposed involving the coupling of flow sorting with nanochannel-based mapping, complex regions in the genome can be spanned and resolved better, which contributes to the contiguity of the assembly (Udall and Dawe 2017). With high quality DNA and further resources, we will be able to apply genome mapping and to produce Chicago libraries to close the genome assembly of *Solanum commersonii* and finally answer the question of whether there are

small-scale rearrangements that may impede full introgression of desirable traits into potato.

The approach discussed here can be summarized in the decision-making process represented in Figure 1. This flowchart can guide anyone attempting to explore the use of a potato wild relative to decide what kind of information is necessary and what kind of tools can be used to obtain it or to achieve the successful transfer of traits into cultivated potato. The results that I obtained in my thesis represent important responses to the doubts that potato pre-breeders have had for decades and that have slowed down and even discouraged the use of *S. commersonii* and *S. chacoense* in introgressive hybridization breeding. Firstly, these species readily cross with diploid potato producing scant but viable offspring that can be used in backcrosses with cultivated potato. Secondly, wild chromosomes pair and recombine with their cultivated homoeologues, making introgression possible. Thirdly, for as far as I know, these species are highly collinear at the large and small scale and their genomes have not diverged much both at the level of repetitive sequences and at the level of synteny. All in all, these species are particularly promising for potato pre-breeding, not only because of their various traits of interest, their wide variability and their adaptability and hardiness, but also because of their high similarity with potato from the genetic, cytogenetic and genomic points of view.

References

Aflitos SA, Schijlen E, de Jong H, et al (2014) Exploring genetic variation in the tomato (*Solanum* section Lycopersicon) clade by whole-genome sequencing. Plant J 80:136–148. doi: 10.1111/tpj.12616

Anderson LK, Covey PA, Larsen LR, et al (2010) Structural differences in chromosomes distinguish species in the tomato clade. Cytogenet Genome Res 129:24–34. doi: 10.1159/000313850

Aversano R, Contaldi F, Ercolano MR, et al (2015) The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. Plant Cell 27:954–968. doi: 10.1105/tpc.114.135954

Belton J-M, McCord RP, Gibcus JH, et al (2012) Hi–C: A comprehensive technique to capture the conformation of genomes. Methods 58:268–276. doi: 10.1016/j.ymeth.2012.05.001

Bethke PC, Halterman DA, Jansky S (2017) Are we getting better at using wild potato species in light of new tools? Crop Sci 57:1241–1258. doi: 10.2135/cropsci2016.10.0889

Castillo A, Gaiero P, López-Carro B, Vilaró F (2016) Gametic embryogenic response in wild diploid *Solanum* species and its implications for genome sequencing projects and breeding. Plant Tissue Cult Biotechnol 26:159–173. doi: 10.3329/ptcb.v26i2.30566

Feingold S, Lloyd J, Norero N, et al (2005) Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). Theor Appl Genet 111:456–66. doi: 10.1007/s00122-005-2028-2

Gaiero P, Mazzella C, Vilaró F, et al (2017) Pairing analysis and *in situ* Hybridisation reveal autopolyploid-like behaviour in *Solanum commersonii* × *S. tuberosum* (potato) interspecific hybrids. Euphytica 213:137. doi: 10.1007/s10681-017-1922-4

Gaiero P, Speranza P, de Jong H (2018) Introgressive Hybridization in Potato Revealed by Novel Cytogenetic and Genomic Technologies. Am J Potato Res 1–15. doi: 10.1007/s12230-018-9669-6

Gaiero P, van de Belt J, Vilaró F, et al (2016) Collinearity between potato (*Solanum tuberosum* L.) and wild relatives assessed by comparative cytogenetic mapping. Genome 60:228–240. doi: 10.1139/gen-2016-0150

Ghislain M, Núñez J, Rosario Herrera M, et al (2009) Robust and highly informative microsatellite-based genetic identity kit for potato. Mol Breed 23:377–388. doi: 10.1007/s11032-008-9240-0

Ghislain M, Spooner DM, Rodríguez F, et al (2004) Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato. Theor Appl Genet 108:881–90. doi: 10.1007/s00122-003-1494-7

González M (2010) La resistencia a la marchitez bacteriana de *Solanum commersonii* Dun. y su utilización en el mejoramiento genético de papa. MSc Thesis, Facultad de Agronomia, Universidad de la Republica

Hastie A, Dong L, Smith A, et al (2013) Rapid genome mapping in nanochannel arrays for highly complete and accurate de novo sequence assembly of the complex *Aegilops tauschii* genome. PLoS One 8:e55864. doi: 10.1371/journal.pone.0055864

Jacobsen E, Schouten HJ (2007) Cisgenesis strongly improves introgression breeding and induced translocation breeding of plants. Trends Biotechnol. 25:219–223.

Ji Y, Chetelat RT (2003) Homoeologous pairing and recombination in *Solanum lycopersicoides* monosomic addition and substitution lines of tomato. TAG Theor Appl Genet 106:979–89. doi: 10.1007/s00122-002-1090-2

Ji Y, Pertuzé R, Chetelat RT (2004) Genome differentiation by GISH in interspecific and intergeneric hybrids of tomato and related nightshades. Chromosom Res 12:107–16.

Jiao W, Accinelli GG, Hartwig B, et al (2017) Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. 778–786. doi: 10.1101/gr.213652.116.Freely

Johnston S, Nijs T Den, Peloquin S, Hanneman R (1980) The significance of genic balance to endosperm development in interspecific crosses. Theor Appl Genet. 57:5–9.

Lam ET, Hastie AR, Lin C, et al (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30:771–6. doi: 10.1038/nbt.2303

Lou Q, Iovene M, Spooner DM, et al (2010) Evolution of chromosome 6 of *Solanum* species revealed by comparative fluorescence *in situ* hybridization mapping. Chromosoma 119:435–442. doi: 10.1007/s00412-010-0269-6

Milbourne D, Meyer RC, Collins a J, et al (1998) Isolation, characterisation and mapping of simple sequence repeat loci in potato. Mol Gen Genet 259:233–45.

Parokonny A, Marshall J, MD B, et al (1997) Homoeologous pairing and recombination in backcross derivatives of tomato somatic hybrids (*Lycopersicon esculentum* (+) *L. peruvianum*). Theor Appl Genet 94:713–723.

Pérez-de-Castro AM, Vilanova S, Cañizares J, et al (2012) Application of genomic tools in plant breeding. Curr Genomics 13:179–195.

Peters SA, Datema E, Szinay D, et al (2009) *Solanum lycopersicum* cv. Heinz 1706 chromosome 6: distribution and abundance of genes and retrotransposable elements. Plant J 58:857–69. doi: 10.1111/j.1365-313X.2009.03822.x

Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–95. doi: 10.1038/nature10158

Putnam NH, O'Connell BL, Stites JC, et al (2016) Chromosome-scale shotgun assembly using an *in vitro* method for long-range linkage. Genome Res 26:342–50. doi: 10.1101/gr.193474.115

Seah S, Yaghoobi J, Rossi M, et al (2004) The nematode-resistance gene, *Mi*-1, is associated with an inverted chromosomal segment in susceptible compared to resistant tomato. Theor Appl Genet 108:1635–1642. doi: 10.1007/s00122-004-1594-z

Šimková H, Číhalíková J, Vrána J, et al (2003) Preparation of HMW DNA from plant nuclei and chromosomes isolated from root tips. Biol Plant 46:369–373.

Tang H, Lyons E, Town CD (2015a) Optical mapping in plant comparative genomics. Gigascience 4:3. doi: 10.1186/s13742-015-0044-y

Tang H, Zhang X, Miao C, et al (2015b) ALLMAPS: Robust scaffold ordering based on multiple maps. Genome Biol 16:3. doi: 10.1186/s13059-014-0573-1

Udall J, Dawe RK (2017) Is it ordered correctly? Validating genome assemblies by optical mapping. Plant Cell 30:tpc.00514.2017. doi: 10.1105/tpc.17.00514

Van Der Knaap E, Sanyal A, Jackson SA, Tanksley SD (2004) High-resolution fine mapping and fluorescence *in situ* hybridization analysis of *sun*, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics 168:2127–2140. doi: 10.1534/genetics.104.031013

Verlaan MG, Szinay D, Hutton SF, et al (2011) Chromosomal rearrangements between tomato and *Solanum chilense* hamper mapping and breeding of the TYLCV resistance gene *Ty*-1. Plant J 68:1093–1103. doi: 10.1111/j.1365-313X.2011.04762.x

Warschefsky E, Penmetsa RV, Cook DR, von Wettberg EJB (2014) Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. Am J Bot 101:1791–800. doi: 10.3732/ajb.1400116

Zheng C, Boer M, van Eeuwijk F (2018) Rabbit (Reconstructing Ancestry Blocks BIT by bit). Construction of genetic linkage maps in multiparental populations. Manuscript.

So close, no m
atter how far
Couldn't be much more from the heart
Forever trusting who we are
And nothing else matters
...

Trust I seek and I find in you
Every day for us something new
Open mind for a different view
And nothing else matters
Metallica
(In: Nothing else matters, Metallica - The black album, 1992)

Summary

Cultivated potato and its wild relatives represent a more diverse and accessible germplasm resource than that of any other crop. Diversity coming from wild relatives can be used to introduce specific traits into the cultivated genetic background. Breeders have long been using introgressive hybridizations, in which chromatin carrying genes of interest from a wild relative is integrated into the genome of the crop by interspecific hybridization. During the subsequent backcrossing generations, genes of interest are incorporated into the crop chromosomes by homeologous recombination. Offspring plants with the desired traits are then selected, while the original cultivated genetic background is recovered by backcrossing and selection as far as possible. The efficient use of wild relatives now requires extensive knowledge of their allelic variation and genomic structure, including the screening for desirable traits. To minimize the occurrence of linkage drag when introgressed chromatin still contains closely linked wild traits from the ancestral donor, knowledge on the genomic structure of crop and donor species is indispensable. In this thesis, I have presented an overview of genetic, cytogenetic and genomic characteristics of *Solanum commersonii*, a South American wild relative of potato. The main aim was to create scientific tools to distinguish the genomes of *S. commersonii* and *S. tuberosum* (cultivated potato), along with related *S. chacoense*, and trace their genomes in introgressive hybridization breeding programmes. The application of these tools to pinpoint past introgressions into cultivated potato and to plan introgressive hybridization schemes is reviewed in **Chapter 2**, with emphasis on the need to make these technologies part of the routine toolkit for pre-breeding in order to exploit potato wild relatives at their fullest. Various questions related to the use of particular potato wild relatives in introgressions are addressed using different cytogenetic and genomic technologies. In **Chapter 3** pairing of the homoeologous *S. commersonii* genomes and *S. tuberosum* Group Phureja chromosomes in F1 interspecific triploid hybrids was analyzed, observing an almost autotriploid behaviour. Genome painting was used to follow the fate of specific chromosomes in these hybrids and backcross derivatives. Homoeologues could not be distinguished by fluorescence labels in the genome painting, so this approach did not provide tools to select against wild chromatin in the advanced materials nor was it possible to pinpoint chromosomal exchanges between the homoeologues. However, the results suggest a low degree of genomic divergence between the species involved in terms of repetitive sequences and highlight their potential for homoeologous pairing. A comparative cytogenetic mapping analysis was performed in **Chapter 4**, to test collinearity among genomes of the different species and thus anticipate the existence of either suppression of recombination or linkage drag. *Solanum commersonii* and *S. chacoense* are collinear with cultivated *S. tuberosum* on the whole chromosome scale. Some differences observed in microscopic distances between the fluorescent foci suggest either small-scale rearrangements or reduction/amplification of repetitive sequences. No major rearrangements have been found, making these amenable species for efficient introgressive hybridization breeding. In **Chapter 5** we looked at the repetitive fraction of the genomes of potato and its wild relatives compared to tomato and its wild relatives, in order to elucidate if differences in genome structures observed among the clades can be correlated with specific dynamics in repetitive elements. Using a statistical clustering approach with short read data, we found that the classes of repetitive elements across the clades are largely conserved, while their abundances are dissimilar so that repeat profiles allowed us to discriminate the two clades. We also

found that the repeat content of *S. etuberosum* is more similar overall to the potato clade. All these tools provide information about possible crossability among these species in terms of their genome similarity and collinearity in chromosome structure. However, the possibility of small-scale differences required a look at synteny at the sequence level. The main challenge to be addressed in **Chapter 6** was to achieve a contiguous *de novo* assembly that would allow for structural comparison, from the genome of a highly heterozygous species. This was solved by sequencing a haploid clone obtained through *in vitro* culture of microspores and by taking a hybrid approach to the assembly, integrating short Illumina reads with much longer PacBio sequence data. A genetic map was developed using SNP data generated through Genotype by sequencing (GBS) from a biparental segregating population. Then the hybrid assembly was anchored to the linkage map into 12 pseudomolecules that were built independently from any reference genome. The 12 pseudomolecules showed high overall homology with the DM potato and M6 *S. chacoense* pseudomolecules, with homologous scaffolds of the three species in the same order. However, contiguity in the pericentromeres was lost and so collinearity in these regions could not be established. Small-scale rearrangements could not be conclusively distinguished from assembly or anchoring artifacts, with a few exceptions. In general terms, there is need for further improvement of contiguity and manual correction of the orientation of scaffolds in order to unequivocally detect rearrangements at the microsynteny scale. This can be achieved by using Bionano genome mapping and/or chromatin conformation capture technologies in the near future. To apply these technologies, extremely pure High Molecular Weight (HMW) DNA is necessary, so in **Chapter 7** we adapted a pre-existing isolation protocol to optimize it for *Solanum* species. The results presented here and discussed in Chapter 8 address the doubts that have discouraged potato pre-breeders to use *S. commersonii* and *S. chacoense* in introgressive hybridization breeding. These species can be crossed with diploid potato producing viable offspring that can be used in backcrosses with cultivated potato. Wild chromosomes are able to pair and recombine with their cultivated homoeologues, making introgression possible. These species are highly collinear at the large and small scale and their genomes have not diverged much both at the level of repetitive sequences and at the level of synteny. With the advent of diploid potato breeding, these species are particularly promising because of their traits, diversity and their adaptability and hardiness, but also because of their high similarity with potato from the genetic, cytogenetic and genomic points of view.

Samenvatting

De gecultiveerde aardappel en haar wilde verwanten vertegenwoordigen een meer gevarieerd en toegankelijk kiemplasma dan welk ander gewas ook. Diversiteit dat van wilde verwanten komt kan gebruikt worden om specifieke eigenschappen toe te voegen aan het cultiveerde erfgoed. Veredelaars hebben sinds lang de introgressieve hybridisatie gebruikt, waarbij chromatine dat belangrijke genen bevatten van een wilde verwant kunnen overdragen naar het genoom van het cultuurgewas door middel van soortskruisingen. Gedurende de opeenvolgende terugkruisingsgeneraties worden de gewenste genen ingebouwd in de chromosomen van het gewas door middel van homoeologe recombinatie. In het nakomelingschap worden dan individuale planten met de gewenste eigenschappen geselecteerd terwijl de oorspronkelijke genetische informatie van het cultuurgewas zoveel mogelijk wordt behouden. Het efficiente gebruik van wilde verwanten vereist diepgaande kennis van hun allelenvariatie en genoomstructuur, inclusief de screening van de gewenste eigenschappen. Om het optreden van meegekomen ongewenste eigenschappen ("linkage drag") van de wilde verwant, die genetisch gekoppeld zijn aan het geselecteerde gen, is kennis van de genoomstructuur van gewas en wilde verwant onontbeerlijk. In dit proefschrift heb ik een overzicht gegeven van genetische, cytogenetische en genomische kenmerken van *Solanum commersonii*, een Zuid-Amerikaanse wilde verwant van de aardappel. Het belangrijkste doel was om wetenschappelijke gereedschappen te creeren, die onderscheid kunnen maken tussen de genomen van *S. commersonii* en *S. tuberosum* (de cultuuraardappel, samen met de verwante *S. chacoense*, en het lot van hun genomen te volgen in introgressieve hybridisatie programma's. Het literatuuroverzicht over het gebruik van dergelijke gereedschappen ter vaststelling van vroegere introgressies in cultuuraardappel en om introgressive hybridisatieschema's te plannen is beschreven in **Hoofdstuk 2.** De nadruk lag vooral op het nut van deze technologieën als onderdeel van gereedschappen die in het voortraject van een veredelingsprogramma kunnen worden ingezet. Diverse vragen over het gebruik van bepaalde wilde verwanten voor introgressie programma's voor de aardappel kunnen nu gesteld worden in het licht van cytogenetische en genomische technologieën. De analyse in **Hoofdstuk 3** van meiotische paring tussen de homoeologe chromosomen in F1 hybriden van *S. commersonii* en *S. tuberosum* groep Phureja laat een vrijwel autotriploid gedrag zien. Genoompainting werd vervolgens gebruikt om het lot van bepaalde chromosomen in deze soortshybride en haar terugkruisingsgeneraties te bepalen. De fluorescente labels in de genoompainting gaven geen verschillen tussen de homoeologe chromosomen, waardoor deze benadering niet geschikt is om het lot van het soortsvreemde chromatine te volgen, noch om de crossovers tussen de homoeologen aan te wijzen. De resultaten suggereerden echter wel een lage graad van genomische divergentie tussen de betreffende soorten in de zin van repetitieve sequenctie verschillen en de mogelijkheden van homoeologe paring. Een vergelijkbare cytogenetische kaartanalyse werd beschreven in **hoofdstuk 4**, waarbij de co-lineariteit tussen de genomen van de verschillende soorten centraal staat en dus vooruitloopt op het bestaan van onderdrukking van meiotische recombinatie of linkage drag. Op chromosoomniveau blijkt dat *S. commersonii* en *S. chacoense* co-lineair zijn ten opzichte van *S. tuberosum*. Wel zijn er kleine verschillen in de afstanden tussen de fluorescente foci, die het bestaan van kleine herrangschikkingen of reductie/amplificatie van repetitieve sequenties suggereren. Omdat er geen grote herrangschikkingen konden worden aangetoond lijken deze

soorten geschikt voor introgressive hybridisatie voor de veredeling. In **hoofdstuk 5** vergeleken we de repetitieve genoomfracties van aardappel samen met verwante soorten met die van de tomaat en verwante soorten. Het doel was helder te krijgen of genoomverschillen tussen beide claden gecorreleerd kunnen worden met de betreffende dynamiek van hun repetitieve elementen. Gebruikmakend van een clustering benadering met short read data vonden we dat de klassen van repetitieve elementen over beide claden grotendeels geconserveerd zijn, maar hun abundanties zijn anders terwijl de repeat profielen ons in staat stellen om de twee claden te onderscheiden. Al deze werktuigen bieden informatie over mogelijke kruisbaarheid tussen deze soorten in de zin van genoomgelijkheid en co-lineariteit van de chromosomen. Echter de mogelijkheid van kleinschalige verschillen vereisen een studie naar de syntenie op sequentieniveau. De grote uitdaging wordt nader beschreven in **Hoofdstuk 6**, waarbij het ons ging om een aaneengesloten *de novo* montage van het genoom van een heterozygote *S. commersonii* te verkrijgen, waarmee dan structurele vergelijkingen met andere soorten konden worden gemaakt. Hiervoor gebruikte we een haploide kloon verkregen gebruik makend van *in vitro* cultuur van microsporen. Voor de sequentie montage maakte we gebruik van een hybride benadering, waarbij korte Illumina reads en veel langere PacBio sequentie data worden gecombineerd. Voor de genetische kaart gebruikte we SNP data die door GBS (Genotyping by sequencing) van een biparentale segregerende populatie werden verkregen. Door combinatie van beide data sets konden de fysieke data verankerd worden aan de koppelingskaart waarmee 12 pseudomoleculen konden worden verkregen die onafhankelijk zijn van elke referentiegenoom. De 12 pseudomoleculen vertoonden een hoge mate van homologie met de pseudomoleculen van de DM aardappel en de M6 *S. chacoense* pseudomoleculen, inclusief de volgorde van de homologe scaffolds van de drie soorten. In de pericentromeres was de aaneengeslotenheid van de sequenties niet mogelijk en dus kon geen co-lineariteit in die gebieden worden vastgesteld. Afgezien van een paar kleinschalige herrangschikkingen waren we onzeker over de co-lineariteit tussen de soorten door mogelijke artefacten in de sequentiemontage of verankering met de genetische kaart. In algemene zin is het nodig om de sequentiemontage verder te verbeteren, terwijl handmatige correcties van de oriëntaties van scaffolds nodig zijn om herrangschikkingen op microsyntenie schaal ondubbelzinnig te kunnen vaststellen. Dit kan in de toekomst verkregen worden door gebruik te maken van Bionano genoom mapping en/of chromatin conformation capture technologieën. Voor het gebruik van deze technieken is extreem zuiver hoogmoleculair DNA nodig. In **Hoofdstuk 7** presenteren we een aanpassing van een bestaande techniek, nu geoptimaliseerd voor *Solanum* soorten. In **Hoofdstuk 8** richt ik mij op de twijfels van aardappelveredelaars op het gebruik van *Solanum commersonii* en *S. chacoense* als donor in introgressieve hybridizatie programma's. Deze soorten kunnen met diploid aardappel worden gekruist en produceren levensvatbaar nakomelingschap op dat kan worden ingezet voor terugkruisingen met cultuur aardappel. De chromosomen van de wilde verwanten kunnen paren en recombineren met hun homoeologe aardappel chromosomen en dus maken introgressie mogelijk. In feite zijn deze soorten in hoge mate co-lineair, terwijl de hun genomen weinig zijn gedivergeerd zowel in de zin van repetitieve sequenties als syntenie. Met de opkomst van diploide potato veredeling, zijn *S. commersonii* en *S. chacoense* buitengewoon veelbelovend vanwege hun eigenschappen, diversiteit, aanpasbaarheid en hardheid, maar ook vanwege hun sterke gelijkheid met de aardappel in de zin van genetische, cytogenetische and genomische aspecten.

Resumen

La papa cultivada y sus parientes silvestres representan una fuente de germoplasma más diversa y accessible que la de cualquier otro cultivo. La diversidad proveniente de parientes silvestres se puede usar para introducir características específicas de los parientes silvestres en un contexto cultivado. Los mejoradores han usado por mucho tiempo la hibridación introgresiva, en la cual la cromatina que lleva un gen de interés proveniente de un pariente silvestre se integra en el genoma del cultivo por hibridación interespecífica. Durante las generaciones de retrocruza subsecuentes, los genes de interés se incorporan a los cromosomas de la especie cultivada por recombinación homeóloga. La progenie es luego seleccionada usando la característica deseable como criterio, mientras que el contexto genético cultivado original se recupera en la medida de lo posible mediante retrocruzas y selección. El uso eficiente de los parientes silvestres requiere en la actualidad de un extenso conocimiento de su variación alélica y estructura genómica, incluyendo el tamizaje por caracteres deseables. Para minimizar la ocurrencia de arrastre por ligamiento cuando la cromatina introgresada aún contiene caracteres silvestres cercanamente ligados provenientes del donante ancestral, el conocimiento de la estructura genómica de las especies cultivada y donante es indispensable. En este trabajo he presentado una vista general de las características genéticas, citogenéticas y genómicas de *Solanum commersonii*, un pariente silvestre sudamericano de la papa cultivada. El objetivo general fue crear herramientas científicas para distinguir los genomios de *S. commersonii* y *S. tuberosum* (papa cultivada), junto con *S. chacoense*, y seguir sus genomios en programas de mejoramiento por hibridación introgresiva. El **Capítulo 2** es una revisión de la aplicación de estas herramientas para localizar introgresiones pasadas en la papa cultivada y para planificar nuevos esquemas de hibridación introgresiva, con un énfasis en la necesidad de hacer de estas tecnologías parte de la "caja de herramientas" de rutina del premejoramiento de manera de explotar al máximo a los parientes silvestres de la papa. Una variedad de preguntas relativas al uso de ciertos parientes silvestres de la papa en introgresión son abordadas en este trabajo usando diferentes tecnologías citogenéticas y genómicas. En el **Capítulo 3** se analizó el apareamiento genómico de los cromosomas homeólogos de *S. commersonii* y *S. tuberosum* Grupo Phureja en híbridos interespecíficos F1 triploides, observándose un comportamiento meiótico prácticamente autotriploide. Se utilizó pintura genómica para seguir cromosomas específicos en estos híbridos y sus derivados por retrocruza. Los homeólogos no se pudieron distinguir por fluorescencia, por lo tanto este enfoque no aportó herramientas para seleccionar en contra de cromatina silvestre en los materiales avanzados ni para localizar los intercambios cromosómicos entre homeólogos. Sin embargo, estos resultados sugieren que hay un bajo grado de divergencia genómica en términos de secuencias repetidas entre las especies involucradas y destacan su potencial para el apareamiento homeólogo. Un análisis de mapeo citogenético comparativo fue llevado a cabo en el **Capítulo 4**, para evaluar colinealidad entre los genomas de las especies y así anticipar la existencia ya sea de supresión de la recombinación o arrastre por ligamiento. *Solanum commersonii* y *S. chacoense* son colineares con la papa cultivada *S. tuberosum* a nivel de todo el cromosoma. Algunas diferencias que fueron encontradas en las distancias microscópicas entre señales sugieren o bien rearreglos cromosómicos a pequeña escala o reducción/amplificación de secuencias repetidas. No se observaron rearreglos importantes, lo que hace que estas especies se presten para una hibridación introgresiva eficiente.

En el **Capítulo 5** examinamos la fracción repetitiva de los genomas de la papa y sus parientes silvestres en comparación con la de los genomas del tomate y sus parientes silvestres, con el fin de elucidar si las distintas diferenciaciones genómicas observadas entre los clados pueden correlacionarse con distintas dinámicas de elementos repetidos. Usando un enfoque de agrupamiento con datos de lecturas cortas de secuencia, encontramos que las clases de los elementos repetidos están mayormente conservadas a través de los clados, pero que sus abundancias son diferentes y por tanto los perfiles de repetidos nos permitieron discriminar los dos clados. También hallamos que el contenido de repetidos en *S. etuberosum* es globalmente más similar al del clado de las papas. Todas estas herramientas proporcionan información sobre la posible capacidad de cruzamiento entre estas especies en términos de similitud genómica y colinealidad en la estructura cromosómica. No obstante, la posibilidad de diferencias a pequeña escala requirió una mirada a la sintenia a nivel de secuencias. El principal desafío a encarar en el **Capítulo 6** fue el de obtener un ensamblado *de novo* contiguo que permitiera comparaciones estructurales, a partir del genoma de una planta altamente heterocigota. Esto se resolvió al secuenciar un clon haploide obtenido a través de cultivo *in vitro* de microsporas y al optar por un abordaje híbrido del ensamblado, integrando lecturas cortas de Illumina con datos de secuencia mucho más largos de PacBio. Un mapa genético fue desarrollado usando datos de SNPs generados mediante GBS a partir de una población segregante biparental. Luego el ensamblado híbrido fue anclado en el mapa de ligamiento formando 12 pseudomoléculas construidas con independencia de cualquier genoma de referencia. Las 12 pseudomoléculas mostraron una gran homología global con las de la papa DM y M6 *S. chacoense*, con un orden homólogo de los scaffolds entre las tres especies. Sin embargo, la contigüidad se pierde en la región pericentromérica y por tanto, no se pudo evaluar la colinealidad. Los rearreglos a pequeña escala no se pudieron distinguir concluyentemente de los artefactos del ensamblado o el anclaje, con unas pocas excepciones. En términos generales, se necesitan mayores mejorías en la contiguidad y corrección manual de la orientación de los scaffolds de manera de detectar de forma inequívoca los rearreglos a escala de microsintenia. Esto puede ser logrado usando mapeo genómico de Bionano y/o tecnologías de captura de la conformación de la cromatina en un futuro cercano. Para aplicar estas tecnologías, es necesario obtener ADN de alto peso molecular extremadamente puro, por lo que en el **Capítulo 7** adaptamos un protocolo preexistente de extracción de ADN de alto peso molecular para optimizarlo para especies del género *Solanum*. Los resultados presentados en este trabajo y discutidos en el **Capítulo 8** abordan las dudas que han desalentado a los premejoradores de papa de usar a *Solanum commersonii* y *S. chacoense* en mejoramiento por hibridación introgresiva. Estas especies pueden ser cruzadas con papas diploides produciendo progenie viable que puede ser usada en retrocruzas con papa cultivada. Los cromosomas silvestres se aparean y recombinan con sus homeólogos cultivados, lo que hace posible la introgresión. Estas especies son altamente colineales a gran y pequeña escala y sus genomas no han divergido en gran manera tanto a nivel de secuencias repetidas como de sintenia. Con el surgimiento del mejoramiento de papa a nivel diploide, estas especies se vuelven particularmente prometedoras debido a sus caracteres de interés, su diversidad y su adaptabilidad y rusticidad, pero también debido a su gran similitud con la papa cultivada del punto de vista genético, citogenético y genómico.

Acknowledgements

Even though it seemed endless, this thesis has come to an end. There is light at the end of the tunnel and I find myself still very much fascinated by the research that I carried out during my PhD and looking forward to the challenges ahead. Looking back at my PhD, I find a very enjoyable time. During this thesis I went through the happiest and saddest times of my life and I have certainly grown and changed. Many people were by my side, through thick and thin, and I want to thank them all.

First of all, I want to thank all the teachers that planted the seed of critical thinking and curiosity in me, and those who developed in me a taste for biological questions. Also those who encouraged me to pursue a career in research, especially to my undergraduate and MSc mentor Cristina Mazzella. She supported me in my various professional and life projects and guided me with patience and wisdom, always having my best interest as top priority. She was the one who suggested I should contact Prof. Hans de Jong for the possibility of a Sandwich PhD at his lab and who suggested that I asked Prof. Pablo Speranza to be my co-promotor in Uruguay. To both of them I am enormously thankful for pushing me when I was ready for it and for pulling me out when I needed it the most.

My promotor Prof. Hans de Jong welcomed me in his lab and his office for exciting discussions about my PhD project and also welcomed me in his home for delicious meals shared with his wife Lak and many other wonderful people that he introduced to me. He shared with me the passion for cytogenetics and the beauty of this "art" that required patience and creativity, but also saw the need to push my limits and take me out of my comfort zone into the field of genomics. He foresaw the importance this would have for my development as a researcher and how this would impact my research group and my teaching in Uruguay. He also shared with me his high commitment to teaching and I benefitted greatly from his advice. He showed me the importance of collaboration for good research and gave me the means to network with many top researchers in both academia and industry. I also learned from him that having clear research questions and visualizing the end product were essential before even starting my project. In spite of the distance, the time difference and his many occupations, he guided me through the process of producing this thesis and taught me attention to detail in the preparation of figures and manuscripts. Most importantly, he supported me in many situations that I came across during my life in The Netherlands and was understanding and warm as I went through different challenges in my family life. Someone once told me that the greatest choice in a PhD is who will be your promotor and I think that I was very fortunate to have been accepted by Hans.

On the other side of the Atlantic there was Prof. Pablo Speranza, who always followed my project and my progress very closely. To him I must be thankful for introducing me to the world of potato wild relatives and for showing me that it would be a promising field for our lab and my future career. His bright mind is always restless with ideas that he

shared and challenged with me in enriching conversations and discussions. He always emphasized that research must be question- or problem-driven and that everything we do in a scientific study, from compiling literature to designing experiments or preparing figures and tables for a manuscript, must revolve around a hypothesis. In the many pitfalls that I found in my thesis, I always felt supported by him and always had a committed interlocutor to discuss ideas, doubts and to finally elaborate on the interpretation of results and the drafting of manuscripts. My writing developed noticeably thanks to his expert guidance and his sharp questions and my teaching was also improved by learning from his dedication and his approach. Above all, I am grateful to him for being there when I was surrounded by a dark cloud and for challenging me when I most needed it.

My co-promotor Prof. M. Eric Schranz provided interesting discussions and alternative perspectives to the problems I faced in my experiments and in the interpretation of my results. I admire the scope and depth of his knowledge and feel motivated by the ambitious questions he proposes. And I will always remember when he enquired about which chapter I felt was going to make the most impact from my thesis. That question motivated me to pursue the sequence assembly and comparative genomics approach in spite of the difficulties I found along the way.

It was great to find such a warm and welcoming environment in the Cytogenetics lab. It is a spotless and organized place where everything runs like clockwork and which invites precise and flawless work. All of that is thanks to my paranymph, José van de Belt. She taught me how to work safely in the lab, how to perform neat and efficient FISH experiments in our break-of-dawn sessions and together with Laurens introduced me to all the details of lab work and of Dutch coffee breaks. She continued helping me throughout my thesis and was key to make this Sandwich PhD format work. I have tried my best to bring many of her methods and her organization into our lab in Uruguay. I want to thank Xianwen for her thoughtfulness, her kindness and her example of hard work and dedication. I also want to thank Erik Wijnker for very interesting talks and discussions about meiosis, genetics and research, both in the lab and in social situations. His bright and active mind seems to always be eager to share ideas and he showed me that we should always strive to answer the big questions in our research field, to take a leap forward in knowledge. Dóra set the basis for the cytogenetics part of my thesis with her beautiful work on tomato and other Solanum species. She shared with me her experience both in the lab and in being a foreigner in a Dutch social context. Leila was always warm and honest and we had a special connection from the beginning. I was lucky to have met them both.

The Laboratory of Genetics is a vibrant and dynamic place where good conversations over coffee are always at hand and where scientific discussions, whether in the Genetics Seminar Series (GSS) or during coffee breaks, are certainly enlightening and helped me broaden the scope of my interests. That is thanks to Prof. Zwaan and all the staff in Genetics. To Corrie and Frank, my great appreciation for making everything run

smoothly with a helpful and cheerful attitude. Many thanks to my fellow PhD students in the lab, Ramón, Vanessa, Cris, Jelle, Magda, Florien, Xiaomin, Jianhua, Ana Carolina, Mina, Charles, Pádraic, Claudio, Valeria, Roxanne, Alex and Tina. Also to Aina, Tânia and Mark Z. My special thanks go to Wytske, always kind and thoughtful, making our lives easier, taking care of all the complicated procedures and organizing social activities in the department. Thank you for your warm words and for keeping an eye on how everybody is doing.

As my thesis progressed from cytogenetics into genomics, I needed more and more help from colleagues in Bioinformatics. First of all, I want to thank Sander Peters for great discussions throughout the past 6 years, from the planning of the project and the sequencing to the analysis and interpretation of the results. He played devil's advocate and so made my project improve enormously, always asking me to provide a valid biological question. Gabino Sánchez-Pérez helped me invaluably with the planning of the sequencing and the most efficient use of the resources I had. Our cultural similarity made negotiations a lot easier. My appreciation to Elio for his careful preparation of my samples and their sequencing. Thanks also to Henri for teaching me "hybrid assembly for dummies" via remote assistance and to Lidjia for making data so much easier to access. My most special thanks go out to Saulo, who patiently answered all my inexpert questions on bioinformatics, always with a suitable suggestion and in a great mood. In Uruguay there is a growing bioinformatics community and three of the pioneers, Silvia Garaycochea, Schubert Fernández and Akira Saito from INIA Las Brujas, together with Victoria Bonnecarrere were always there to help me from my first attempts to get access to the server and upload data to the most complex (and last minute) analyses, always with a prompt response and understanding words.

I had the pleasure to visit one of the best Plant Genomics labs in the world, the Centre of Plant Structural and Functional Genomics in Olomouc, Czech Republic. I was welcomed by Prof. Jaroslav Doležel and I could not believe I was meeting the great scientist whose papers I had read so many times. He turned out to be very approachable and made sure that I had a wide perspective of all the fascinating top science that was performed in his institute. I was lucky to work under the guidance of Hana Šimková, who meticulously taught me how to purify the precious DNA molecules that we needed so much. She helped me to make this protocol work in Uruguay and was always ready to answer my many questions with genuine interest and care. I want to thank Jan Vrána who taught me the details of nuclei sorting. My special thanks to Gustavo Folle, Federico Santiñaque and Beatriz López-Carro from IIBCE in Montevideo, for their commitment to making this whole effort applicable in our country, for the many years of collaboration (and many more to come exploring the possibilities opened by the new technology in their lab) and for their affection in some of my happiest times. I am also thankful to Alicia Castillo and her staff in the Biotechnology Unit at INIA Las Brujas for constantly providing me with all the in

vitro material that I so often and so urgently needed and for her enthusiasm and dedication in the development of the haploid materials to be sequenced.

Back in Uruguay the amazing people from the Laboratory of Evolution and Domestication of Plants (LEDP) were always there to hold the fort. Thanks to Alejandro, Mario, Pablo Sandro, Inés, Mariana and Manuela and now Germán for all the collaboration in the potato projects. Enormous thanks to Ana Nicola for her help with all the management of projects and funds and the ordering of lab products. Thanks to the greenhouse technicians Matías Antuoni and Julio Sburlatti for their care of the plants and to all the teachers in the Genetics course (Cristina, Magdalena, Susana, Jorge, Marianella, Crosa and Clara) for their support during my absences. Thanks to Clara for her constant encouragement and her sharp questions, gracias a Dora por siempre estar atenta a todo lo que podía necesitar a la distancia y por siempre encargarse de todo con una gran sonrisa, ha sido una gran tranquilidad. Thanks to all the staff in the Department of Plant Biology. Above all, thanks to my dear friend Magdalena, for everything. No need for words.

The open nature of our collaboration with Paco Vilaró made everything simpler and more exciting. Thanks Paco for always providing me with interesting questions and for generously sharing your many years of experience and knowledge on potato breeding. Your example in the love and passion that you have for your work and your deeply rooted belief in pre-breeding has left its mark in me. Together with Matías González, Guillermo Galván and Maria Julia Pianzolla, Maria Inés Siri and many reserchers from the Faculty of Chemistry, you started this promising research that I hope I can continue. I am very grateful to Guillermo for his encouragement when I first decided to go to Wageningen, for his advice on how to start this adventure and for introducing me to the kind Nico and Teresa. My gratitude also goes to Elsa L. Camadro for many deep discussions and exchanges of ideas that greatly improved this thesis.

Now memories and emotions overflow me as I reach the point of thanking the people who are dearest to me. I was lucky to share this journey with some brilliant PhD students (now doctors!) that made me feel that in their company I was growing both in knowledge and as a researcher. Thanks to Setareh, Jordi, Serena, Magda, Lauri, Marcela, Rafael, Johannes, Manon, and to my dear friends Alejandro and Hanna (you deserve a whole separate section Hannita!). I made some good friends along the way and I am proud to say that I have many dear people spread around the globe. Some of them are Uruguayans, like Gabriel who welcomed me in Wageningen and was always encouraging and ready for a good conversation, Pablo, Germán, Carolina, Rafael. To the ones that will come after, I wish a lot of success and a great experience, hopefully I can help the way I was helped. Thanks to the Lunch Club for healthy meals and good conversation. The lovely ladies Ana Paula, Nilma, Valeria and Greice (together with Denis) shared good fun and hard work with me. These friends also were my family away from home, when I could not ask for better company in some of the most important moments of my life. Valeria, you were so moved as I was when seeing Andrés and I know he will be in your heart forever. Hanna

and Daniela supported me through uncertain times (although we did not know it at the moment) and I could never thank them enough for that. Greice and Daniela were there with me when I met the love of my life, my baby Andrés, and I know your heartfelt happiness added to my elation at that moment. To my dear friend, my paranymph Daniela, I have enormous gratitude and affection. Thank you for shared mates and empanadas, for conversations that made me feel back in Rio de la Plata, for helping me with the intricacies of Dutch life (coming from such a similar culture to mine) and for introducing me to so many wonderful people like Natalia, Panam, Amaya, Mia, Paolo, Iliana, Victor, Ilse, and thus expanding my social life in Wageningen.

Thanks to my friends and study partners from FCien for 20 (!!!) years of friendship and support. My special thanks to Lu, Flor, Kari, Merce, Lore y Laura. Karina, thanks for always making sure that we got together and for always organizing gatherings and welcoming us in your home, especially to catch up when one of us came back from abroad. I wish to thank the TTC chicks, Lauris, Ces, Majo and their families. Together we learned many of the skills that have taken me here and they patiently endured my absences. Sorry for not being there in so many important moments, thanks for sharing my wanderlust and for the laughter and music throughout these years. Chichis, Laura, Verito, Picón, Carola, Lali, thank you for always being there for me and making me feel you were close all the time. Gracias por acortar las distancias con mensajes, llamadas, regalitos sorpresa. Gracias por hacer suyas nuestras alegrías y por sostenernos en nuestras tristezas. Agradezco infinitamente tenerlas en nuestras vidas, todos los recuerdos y anécdotas compartidas y espero que sigamos así hasta hacernos viejitas.

My life in the Netherlands wouldn't have been the same without my Dutch family. Rini and Gerrit, you welcomed me in your home to practice your Spanish and I got so much more out of it. Thank you for your cappuccinos and the long talks that came with them, sometimes also over delicious meals or a nice glass of wine. Thank you for helping me when the hard Dutch winter was bringing me down, for your care and your interest in my progress and my adaptation. It seemed that you always knew when I was too tired to cook and those evenings I always found a warm note saying that a nice homemade dinner (sometimes even my beloved pumpkin soup!) was waiting for me. Thank you for making me feel at home away from home and for being there in all the important moments, even across the ocean. To all the Breman family, mi cariño y gratitud.

Finally, I want to thank my family. To my brothers and my parents, I have only gratitude for their love and support. Enzo, thank you for covering for me when I was not around, thanks for your many talents that have enriched my life and for brushing up my English language and culture when they were too rusty. Bruno, Cata, Gero y Bauti, thanks for your constant support, for your sincere interest in my progress and your pride in my achievements. A mis padres, Ledo y Esther, que desde el principio me impulsaron a lograr lo que me propusiera, que priorizaron siempre nuestra educación a pesar de sus muchos sacrificios, que nos dieron ejemplo de trabajo, esfuerzo, amor y dedicación. Gracias por

enorgullecerse con mis logros, por sus horas de cansancio para que yo pudiera terminar este trabajo y por disfrutar conmigo todos estos viajes y experiencias. Gracias por darme las oportunidades e impulsarme a tomarlas, son los mejores padres del mundo. To my children, who had to share my time and attention with this thesis and who have been my greatest teachers. For them I have done everything I did and for them I wanted to make a small contribution to make this world a little bit better. I know Lara will understand everything that we went through to follow our dreams and that she will learn that with hard work and dedication, the sky is the limit. My final words go to my husband, Pablo. A lifetime will not be enough to thank him for his support through this years, from the decision to embark in this PhD till the late night writing sessions a few days ago. Gracias por ser mi compañero en este proyecto de vida, por poner siempre primero mi felicidad, aunque implicara grandes sacrificios para vos. Gracias por ser mi cable a tierra cuando los nervios me ganaban, por siempre tener un abrazo y darme una mano cuando estaba abrumada, por ser mi roca. Agradezco cada momento de nuestra vida juntos, gracias a vos todo esto fue posible y gracias a nuestra familia todo esto tiene sentido.
It has been an amazing rollercoaster ride.

Curriculum Vitae

Paola Gaiero Guadagna was born on 17th August 1979 in Montevideo, Uruguay. After attending Colegio Nacional José Pedro Varela school and graduating from Instituto Uruguayo de Educación Preparatoria (IUDEP), she started her studies at the Faculty of Sciences of the University of the Republic (UdelaR), Montevideo, Uruguay. In 2005 she obtained her Bachelor's degree in Biology, specializing in cytogenetics of native palm trees. She also obtained her Diploma in Teaching English as a Second Language at Instituto Cultural Anglo Uruguayo in 2004 and worked as an English Language Teacher, test designer and Academic Assistant until 2012.

In 2006 she started her MSc programme in Plant Sciences, in the Postgraduate School of the Faculty of Agronomy, UdelaR (Montevideo, Uruguay), while working as a teaching assistant for the Animal Breeding course and as a research assistant in the Laboratory of Genetics (later Laboratory of Evolution and Domestication of Plants), Faculty of Agronomy (Montevideo, Uruguay). In her MSc thesis, under the supervision of Dr. Cristina Mazzella and Dr. Marcelo Rossato (Universidade de Caxias do Sul, Brazil), she used a Conservation Genetics approach to assess genetic diversity and contribute to species delimitation of endangered species of native palm trees (genus *Butia*, *Syagrus romanzoffiana* and *Trithrinax campestris*), using molecular markers and cytogenetic analyses. This work gave her the opportunity of going on research internships in the Laboratory of Plant Cytogenetics, Federal University of Pernambuco (UFPE), Brazil, under the supervision of Prof. Marcelo Guerra and Dr. Andrea Pedrosa-Harand. After obtaining her MSc in 2010, she became part of the permanent staff of the Laboratory of Evolution and Domestication of Plants.

In 2012 Paola started her Sandwich PhD programme in the Laboratory of Genetics of Wageningen University (WU) under supervision of Prof. Hans de Jong, with the co-promotors Prof. Pablo Speranza from Uruguay and Prof. Eric Schranz from the WU Biosystematics Group. This work arose from our collaboration with Dr. Francisco Vilaró, potato breeder at National Institute of Agricultural Research (INIA), Uruguay. The results of this work are described in this thesis. The research focused on generating cytogenetic, genetic and genomic tools that would allow us to compare potato and its wild relatives, assess their degree of genomic differentiation and their suitability for introgressive hybridization breeding. This research was linked to many projects lead by Dr. Speranza, Dr. Mazzella or Dr. Vilaró, in which Paola was part of the team and to MSc theses projects in which she participated. Currently she works at the Laboratory of Evolution and Domestication of Plants where she leads a project and supervises students on topics related to the characterization of wild potato relatives in their natural distribution areas. She lives near the sea in El Pinar (Canelones, Uruguay), with her husband Pablo, their daughter Lara and their three dogs.

Publications

Gaiero P, Vaio M, Schranz ME, de Jong H, Speranza P. 2018. Comparative analysis of repetitive sequences among species from the potato and the tomato clades. Annals of Botany – in press. https://doi.org/10.1093/aob/mcy186

Gaiero P, Šimková H, Santiñaque FF, López-Carro B, Folle GA, Vrána J, van de Belt J, Peters SA, Doležel J, de Jong H. 2018. Intact DNA purified from flow-sorted nuclei unlocks the potential of next-generation mapping in *Solanum* species. MethodsX 5: 328-336. https://doi.org/10.1016/j.mex.2018.03.009

Gaiero P, Speranza P, de Jong H. 2018. Introgressive hybridization in potato revealed by novel cytogenetic and genomic technologies. American Journal of Potato Research – Online first https://doi.org/10.1007/s12230-018-9669-6

Gaiero P, Mazzella C, Vilaró F, Speranza P, de Jong H. 2017. Pairing analysis and in situ Hybridisation reveal autopolyploid-like behaviour in *Solanum commersonii* × *S. tuberosum* (potato) interspecific hybrids. Euphytica 213 (7): 137-152. https://doi.org/10.1007/s10681-017-1922-4

Gaiero P, van de Belt J, Vilaró F, Schranz ME, Speranza P, de jong H. 2016. Collinearity between potato (*Solanum tuberosum* L.) and wild relatives assessed by comparative cytogenetic mapping. Genome 60(3): 228-240. https://doi.org/10.1139/gen-2016-0150

Castillo A, Gaiero P, López-Carro B, Vilaró F. 2016. Gametic embryogenic response in wild diploid *Solanum* species and its implications for genome sequencing projects and breeding. Plant Tissue Culture and Biotechnology 26 (2): 159-173. http://dx.doi.org/10.3329/ptcb.v26i2.30566

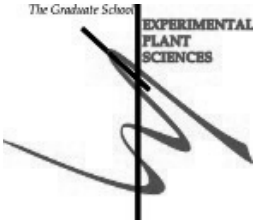
Mourelle D, Gaiero P, Speroni G, Millán C, Gutiérrez L, Mazzella C. 2015. Comparative pollen morphology and viability among endangered species of *Butia* (Arecaceae) and its implications for species delimitation and conservation. Palynology, v.: 40 (2): 160-171. https://doi.org/10.1080/01916122.2014.999955

Gaiero P, Mazzella C, Vaio M, Barros e Silva AE, Santiñaque FF, López-Carro B, Folle GA, Guerra M. 2012. An unusually high heterochromatin content and large genome size in the palm tree *Trithrinax campestris*. Australian Journal of Botany 60 (4):378-382. https://doi.org/10.1071/BT12029

Gaiero P, Mazzella C, Agostini G, Bertolazzi S, Rossato M. 2011. Genetic diversity among endangered Uruguayan populations of *Butia* Becc. species based on ISSR. Plant Systematics and Evolution 292 (1-2):105-116. https://doi.org/10.1007/s00606-010-0412-0

Mazzella C, Rodríguez M, Vaio M, Gaiero P, López-Carro B, Santiñaque FF, Folle GA, Guerra M. 2010. Karyological features of *Achyrocline* (Asteraceae, Gnaphalieae): stable karyotypes, low DNA content variation and rRNA genes linkage. Cytogenetic and Genome Research (128): 169-176. https://doi.org/10.1159/000290689

Education Statement of the Graduate School Experimental Plant Sciences



Issued to: Paola Gaiero Guadagna
Date: 11 December 2018
Group: Laboratory of Genetics
University: Wageningen University & Research

1) Start-up phase	Date
• **First presentation of your project**	
Exploring the use of wild species in potato breeding	02 Apr 2013
• **Writing or rewriting a project proposal**	
Exploring the use of wild germplasm in potato breeding through integrated genetic and genomic approaches	04 Apr 2013
• **Writing a review or book chapter**	
Introgressive hybridization in potato revealed by novel cytogenetic and genomic technologies, American Journal of Potato Research	2018
• MSc courses	
• Laboratory use of isotopes	
Subtotal Start-up Phase	8.5 credits*
• **EPS PhD student days** EPS PhD student day 2013, Leiden, the Netherlands 5th European Plant Science Retreat for PhD Students, Ghent, Belgium 6th European Plant Sciences Retreat for PhD Students, Amsterdam, the Netherlands	 29 Nov 2013 23-26 Jul 2013 01-04 Jul 2014
• **EPS theme symposia**	
EPS theme 4 symposium 'Genome Biology', Nijmegen, the Netherlands	07 Dec 2012
• **National meetings (e.g. Lunteren days) and other National Platforms**	
Annual meeting 'Experimental Plant Sciences', Lunteren, the Netherlands	11-12 Apr 2016
• **Seminars (series), workshops and symposia**	
EPS Flying seminar: Dr. Detlef Weigel	27 Feb 2013
Seminar: Genetic mapping in autopolyploids, with emphasis in sugarcane by Dr. Augusto Franco García	04 Oct 2013
WEES seminar: Dr. Eric Schranz	21 Nov 2013
Seminar: Genome-wide Marker-assisted Selection by Dr. Jeff Endelman	25 Jun 2014
Seminar: Genetic analysis in MAGIC: advantages and challenges by Dr. Emma Huang	25 Jun 2014
Seminar: New directions in Potato Breeding and Quantitative Genetics by Dr. Jeff Endelman	26 Jun 2014
TransPlant Workshop: Exploiting and understanding Solanaceous genomes, Wageningen, the Netherlands	13-14 Oct 2014

Symposium: All-inclusive Breeding - Integrating high-throughput science, Wageningen, the Netherlands	16 Oct 2014
• **Seminar plus**	
• **International symposia and congresses**	
IV Latinamerican Symposium of Cytogenetics and Evolution, Brazil	26-29 May 2013
Current Opinion Conference on Plant Genome Evolution, Amsterdam, the Netherlands	8-10 Sep 2013
III Meeting of the Uruguayan Society of Genetics	07-08 May 2014
19th Triennial Conference of the European Association of Potato Research, Brussels, Belgium	07-11 Jul 2014
Plant Molecular Cytogenetics in Genomic and Postgenomic Era, Katowice, Poland	23-24 Sep 2014
Workshop: The Potato-Ralstonia solanacearum pathosystem - Advances in breeding for resistance, Montevideo, Uruguay	22 Oct 2015
XXIV International Plant and Animal Genome Conference, San Diego, USA	09-13 Jan 2016
• Presentations	
Poster: IV Latin American Symposium on Cytogenetics and Evolution	26-29 May 2013
Poster: 5th European Plant Science Retreat for PhD Students	23-26 Jul 2013
Poster: Current Opinion Conference on Plant Genome Evolution	08-10 Sep 2013
Poster: 6th European Plant Sciences Retreat	01-04 Jul 2014
Poster: Plant Molecular Cytogenetics in Genomic and Postgenomic Era, Katowice, Poland	23-24 Sep 2014
Poster: XXIII International Plant & Animal Genome	10-14 Jan 2015
Poster: Plant Genomics Congress	11-12 May 2015
"Poster: 18th Triennial Meeting of the EAPR Section 'Breeding and Varietal Assessment' and the EUCARPIA Section 'Potatoes' , Vico Equense, Italia"	15-19 Nov 2015
Poster: XXIV International Plant and Animal Genome Conference	09-13 Jan 2016
Flash Presentation: All-inclusive Breeding - Integrating high-throughput science	16 Oct 2014
Talk: III Meeting of the Uruguayan Society of Genetics	07-08 May 2014
Talk: 19th Triennial Conference European Association of Potato Research	07-11 Jul 2014
Talk: The Potato-Ralstonia Solanacearum Pathosystem - Advances in Breeding for Resistance	22 Oct 2015
Talk: XVI Latin American Genetics Congress	Oct 2016
• **IAB interview**	
• **Excursions**	
Subtotal Scientific Exposure	25.1 credits*
• **EPS courses or other PhD courses**	
Systems biology: statistical analysis of -omics data	10-14 Dec 2012
NCBI advanced course: de novo assembly of NGS data	08-09 Jan 2013
Bioinformatics: a user's approach	04-08 Mar 2013
Next generation sequencing (NGS) data analysis	28-30 Aug 2013

Mixed model based genetic analysis in GenStat: from QTL mapping and association mapping to genomic prediction	02-04 Sep 2013
Introducción a la Bioinformática: aplicaciones en proyectos genómicos de mejoramiento genético Introduction to Bioinformatics: applications in breeding genomic projects	17-21 Mar 2014
• **Journal club**	
• **Individual research training**	
Research internship, Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany, Olomouc, Czech Republic. Training in flow sorting. Supervisor: Jaroslav Dolezel	14-18 Jul 2014
Subtotal In-Depth Studies	6.3 credits*
4) Personal development	
• **Skill training courses**	
Competence Assessment	22 Jan & 27 Feb 2013
EPS Career Day 'ExPectationS': Creativity and Inspiration in Science, Wageningen, the Netherlands	01 Feb 2013
Library Symposium: How to write a world-class paper, Wageningen, the Netherlands	17 Oct 2013
Project and Time Management	Sept-Oct 2013
Data Management	2014
Information Literacy including EndNote Introduction	2014
Social Dutch 1	Feb-Apr 2013
Library courses - EndNote demonstration	14 Feb 2013
Library courses - Mendeley demonstration	26 Feb 2013
Authors' Seminar - Elsevier LatinAmerica	25 Nov 2015
• **Organisation of PhD students day, course or conference**	
Collaboration in the organization of the XVI Latin American Genetics Congress	Sep 2015-Sep 2016
• **Membership of Board, Committee or PhD council**	
Editorial committee of the Plant Biology Department newsletter, Faculty of Agronomy, Montevideo, Uruguay	Aug 2015-Dec 2015
Subtotal Personal Development	6.3 credits*
TOTAL NUMBER OF CREDIT POINTS*	**46.9**
Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits. * A credit represents a normative study load of 28 hours of study.	

The research described in this thesis was carried out at the Laboratory of Genetics at Wageningen University and Research, Wageningen, the Netherlands and the Laboratory of Plant Evolution and Domestication at University of the Republic, Montevideo, Uruguay, and was financially supported by CSIC-University of the Republic, the National Research and Innovation Agency (ANII) and the National Institute of Agricultural Research (INIA), Uruguay.
Financial support from National Institute of Agricultural Research (INIA), Uruguay for printing this thesis is gratefully acknowledged.

Cover by Gabriela Jolochín (gjolochin@yahoo.com.ar).
Thesis layout by Hans de Jong.
Printed by DigiForce|| ProefschriftMaken.