Self-Fertility in a Cultivated Diploid Potato Population Examined with the Infinium 8303 Potato Single-Nucleotide Polymorphism Array


Abstract
Within a population of F1 hybrids between two genotypes (S. tuberosum L. Group Phureja DM 1-3 516 R44 [DM] and S. tuberosum L. Group Tuberosum RH89-039-16 [RH]) used in the potato genome sequencing project, we observed fruit set after self-pollination on many plants. Examination of pollen tube growth in self-fertile and self-unfruitful F1 plants after controlled self-pollinations revealed no difference in the ability of pollen tubes to reach the ovary. To identify genomic regions linked with self-fertility, we genotyped the F1 population using a genome-wide single-nucleotide polymorphism (SNP) array. Polymorphic and robust SNPs were analyzed to identify allelic states segregating with the self-fertile phenotype. All 88 highly significant SNPs occurred on chromosome 12. Seeds obtained after self-pollination of self-fertile individuals were used to advance the population for four generations. Genotyping 46 self-fruitful and 46 self-unfruitful S3 plants on the Infinium 8303 Potato SNP array revealed eight SNPs segregating with self-fertility on chromosomes 4, 9, 11, and 12. Three times more heterozygosity than expected was found in the S3 generation. Estimates of heterozygosity were influenced by copy number variation (CNV) in the potato genome leading to spurious heterozygous genotyping calls. Some spurious heterozygosity could be removed by application of a CNV filter developed from alignment of additional monoploid potato genomic sequence to the DM reference genome. The genes responsible for fruit set in self-fertile plants in the F1 generation were restricted to chromosome 12, whereas new genomic regions contributed to the ability of S3 plants to set fruit after self-pollination.

Core Ideas
- Self-fertility in diploid potato is independent of self-compatibility
- SNP genotyping may be confounded by copy number variation
- Genomic regions of a highly heterozygous species resist homozygosity during inbreeding

Diploid potato (2n = 2x = 24) is essential to genetic studies because of the simplicity it offers over traditionally cultivated highly heterogeneous tetraploids (2n = 4x = 48). One of the drawbacks to diploid potato propagation is that the production of inbred lines through self-pollination is typically limited by gametophytic self-incompatibility (Cipar et al., 1964). This limitation is of particular concern currently, as two diploid genotypes, S. tuberosum Group Phureja DM 1-3 516 R44 (DM) and S. tuberosum group Tuberosum RH89-039-16 (RH) were used to sequence the potato genome (The Potato Genome Sequencing Consortium, 2011). While RH, as...
Materials and Methods

Plant Material

The DRH population was developed by cross-pollination between DM 1-3 516 R44 (DM) as pistillate parent and RH89-039-16 (RH) as pollen parent. The DM parent has not been used successfully as a pollinator although it has shown considerable stainable pollen using acetocarmine staining (unpublished). Clone RH89-039-16 (Park et al., 2007; Van Os et al., 2006) was kindly provided by the Laboratory of Plant Breeding, Wageningen University. The RH parent was observed to set fruit after self-pollination under our growing conditions; hence, self-fertility, which segregated in the DRH F1 progeny, was believed to reflect a particular set of alleles or haplotypes inherited from the paternal genotype. Progeny of specific DRH F1 plants were advanced by self-pollination through the S2 generation (Fig. 1). We also used a panel of 11 diverse monoploid and doubled-monoploid genotypes developed by anther culture, nine of which were described by Hardigan et al. (2016), for identifying possibly spurious heterozygosity on the Infinium 8303 Potato SNP array. For microscopic observation of pollen tube growth in the style of a self-incompatible diploid potato, we used a single seedling from a backcross (DM 9-9 203/16 × N8-2) between DM and its F1 hybrid with a heterozygous diploid field selection known to have an intact self-incompatibility mechanism.

Evaluation of F1 Self-Fertility and Production of Inbred Lines

Unique F1 genotypes (n = 103) of the DRH population were planted in the greenhouse during the fall season with temperature settings of 22°C days and 16°C nights and photoperiod extension to 16 h by artificial light (Fig. 1). As each plant flowered, we removed a single anther per bud, extruded pollen on a glass slide, and stained with acetocarmine. Screening was performed at 40× using a Nikon Alphaphot YS compound microscope to estimate the approximate percentage of aborted pollen for each clone by identifying shriveled and transparent grains. Self-pollinations were made on all DRH genotypes that flowered not exceeding more than five pollinations per day per genotype. In the F1, S1, and S2 generations, we attempted to perform 25 pollinations per genotype. Because of expectations of reduced fruit and seed set in the advanced generations, we elected to not limit the number of pollinations in the S1 and S2 generations, allowing for >25 pollinations per genotype. After ~6 wk, the number of developed fruit per pollination was recorded, the ripened fruit harvested, and seeds extracted and counted. To confirm phenotypes, we replanted genotypes with poor flowering or sporadic fruit set the following fall season under similar conditions in the greenhouse using tubers harvested from plants in late winter and stored at 4°C for ~6 mo. Seed obtained from self-pollination of five DRH F1 plants (No. 16, 67, 76, 84, and 90) was used to advance to the S1
and S2 generations, but only three clones (DRH 16, 76, and 90) resulted in vigorous self-fertile plants after S2. Seedlings derived from self-pollination of the RH parent were likewise too weak to advance. Three hundred seeds comprised of 50 from each of two different self-fertile S3 selections from each of the three families (DRH 16, 76, and 90) were germinated in vitro on Murashige and Skoog (1962) basal medium to represent the S4 generation. After samples were taken for DNA extraction, the plants were acclimated to the greenhouse after first placing them in peat pellets in Magenta boxes, then moving them to 3.8-L pots in the greenhouse once visible rooting had occurred. Self-pollinations were conducted as before to evaluate self-fertility.

**Observation of Compatibility Reaction after Self-Pollination of Selected F1 Plants**

A selection of highly fertile and completely self-infertile DRH F1 individuals was grown in a Conviron Model CMP5090 controlled-environment, walk-in growth chamber for observation of pollen tube growth after self- and cross-pollinations. Flowers were self-pollinated by hand and collected 48 h later. Petals and anthers were removed and the remaining sepals, style, and ovary fixed for 24 h in 3:1 ethanol/acetic acid. After fixation, sepals were removed and fixed style and ovary tissues softened for 1 h in 8 N NaOH at 60°C, rinsed with water, and stained with 1% aniline blue in 0.15 M K2PO4. Tissues were gently squashed directly on a microscope slide with the cover slip and imaged using an Olympus SZX16 fluorescence stereomicroscope equipped with CFP excitation and emission filters. Photographs were taken with an Olympus DP71 camera and software.

**Infinium 8303 Potato Single-Nucleotide Polymorphism Array**

DNA was extracted from selected plants in each inbred generation using a modified hexadecyltrimethylammonium bromide protocol similar to that described by Doyle and Doyle (1987). The Infinium platform (Illumina, Inc.), 8303 SNP array for potato (Felcher et al., 2012) was used to genotype 103 F1 plants and 92 S3 plants. Genotyping calls were made using an Illumina iScan reader with the Infinium HD Assay Ultra and allele calls using GenomeStudio (Illumina, Inc.). Within the S3 generation, 46 self-fertile selections and 46 clones that did not yield fruit on selfing were genotyped.

The SNP data from the DM and RH parents, the F1 and S3 generations, as well as that of 11 monoploids and doubled monoploids were combined for the analysis. Before proceeding to statistical analyses, we used a variety of filters to remove poor quality and uninformative SNPs. The SNPs (n = 637) were removed, which were previously identified by SoLCAP as bad or questionable for three-cluster custom calling in Genome Studio (Hirsch et al., 2013). The SNPs classified under segregation were retained, as we did not know if their faulty segregation in other populations was relevant to DRH. The SNPs...
with two or more hits to the PGSC genome version 2.1.10 \((n = 399)\) were removed (Supplemental Spreadsheet S1: MSUPBG remove sheet). The parental lines DM and RH were genotyped in multiple runs and used to remove SNPs that yielded unexpected (DM heterozygous) or inconsistent reads for these control lines \((n = 247)\) (Supplemental Spreadsheet S1: DM or RH false sheet). The SNP markers with missing calls \((n = 326)\) \(\geq 10\%\) of the combined populations \((103\) DRH \(F_1 + 92\) \(S_1 + 11\) monoploids) were also removed (Supplemental spreadsheet S1: \(>10\%\) no calls sheet). After application of these filters, we identified an additional 22 SNPs that hit to greater or less than a single locus on the PGSC genome version 4.03 (Supplemental spreadsheet S1: more or less than 1 hit sheet), and these SNPs were removed.

As a result of an ongoing study of CNV in the potato genome (Hardigan et al., 2016), we were able to apply a CNV filter. The CNV in the RH genome could have resulted in the appearance of spurious heterozygosity on the Infinium 8303 Potato SNP array, where alternate SNP states in duplicated regions comprising an oligonucleotide probe would give the appearance of heterozygosity when in fact there were duplicate loci. As a test for CNV in the potato genome, we ran a set of 10 monoploid lines representing a diverse set of \(S.\ tuberosum\) Group Phureja germplasm on the Illumina 8303 Potato SNP array; nine of these lines were included in the study of CNV conducted by Hardigan et al. (2016), and the 10th line was derived by anther culture of an \(F_1\) hybrid between DM and a diploid heterozygous clone. In addition, we included a single monoploid clone of \(S.\ chacoense\) on the Infinium 8303 Potato SNP array. The illegitimate appearance of heterozygosity in these monoploids occurred for 185 of 8303 SNPs after application of filters to eliminate previously determined bad SNPs (Supplemental spreadsheet S1: CNV filter sheet). This spurious heterozygosity occurred in one to seven of the 11 monoploids. Any single monoploid had from eight to 91 heterozygous SNPs. The availability of \(30 \times\) to \(69 \times\) genomic sequence coverage of the monoploids (Hardigan et al., 2016) allowed us to observe the alignment of selected monoploids with spurious heterozygosity at a given SNP position against the DM genome version 4.03. Most of the SNPs with spurious heterozygosity in the monoploid set showed good evidence of CNV at the SNP position as exemplified in Supplemental Fig. S1. We excluded these 185 SNPs from our analysis by application of the CNV filter.

After examination of our SNP data, we removed 17 SNPs with illegitimate segregation, that is, RH was heterozygous, but the SNPs were nonsegregating in the \(F_1\) generation (Supplemental spreadsheet S1: AAxAB all hetero sheet). Finally, SNPs for which there was no polymorphism in the population were removed \((n = 4066\); Supplemental Spreadsheet S1: no polymorphism sheet). Filters for the \(F_1\) and \(S_1\) generations departed at this juncture. For the \(F_1\) population, 601 SNPs were completely heterozygous as a result of alternate homozygous states in the parents (Supplemental Spreadsheet S1: AAxBB all hetero sheet). These were removed in the \(F_1\) analysis, as they could not contribute to linkage with phenotype; however, they were reinstated for the \(S_1\), as they would be expected to segregate after selfing the \(F_1\). The total number of available SNPs for the \(F_1\) analysis was 1803. For the \(S_1\) analysis, the additional 601 SNPs heterozygous in the \(F_1\) population were reduced to 321 because 280 were nonpolymorphic in the three founder \(F_1\) individuals.

### Statistical Analyses

Segregation patterns for both fruit set and seed set were used to identify SNPs significantly correlated with the self-fertile phenotype. Genotypes for which no pollinations were performed, or for which no fruit production was observed and less than five pollinations were performed, were eliminated to avoid possible erroneous categorization of phenotypes. A bivariate fit of fruit-per-pollination rate vs. seeds per fruit was conducted to determine if these two traits were significantly correlated (JMP Pro 10). For the trait of fruit set categories, we assigned individuals into three classes of self-fertility: if a genotype produced zero fruits per pollination it was categorized as low, if a genotype produced more than zero but less than 0.25 fruits per pollination it was categorized as medium, and if a genotype produced equal to or more than 0.25 fruits per pollination it was categorized as high. These three categories were created because we believed that simply scoring genotypes as self-fertile and not self-fertile might be misleading because of the occurrence of some genotypes that only produced a few fruits after many pollinations. By creating categories that differentiated between medium and high fruit-per-pollination rates, we hoped to reduce the confusion of chance breakdown barriers to self-unfruitfulness from environmental conditions with the genetic factors that control this trait.

The statistics program R was used to test for correlation of segregating SNP genotypes with segregating self-fertility phenotypes (low, medium, and high categories) using a Fisher’s Exact Test (R Development Core Team, 2010). Only SNPs with a \(p\)-value \(\leq 0.0001\) were retained. The SNPs that met this criterion were subjected to contingency table analysis in JMP Pro 10 to identify the potential for Type 1 errors in the dataset (SAS). For each SNP, if an allelic state was represented by five or less individuals and that same allelic state was the sole cause for a SNP to be significant, then that SNP was discarded. The SNPs were also removed when two or more allelic states each represented <50% of a category (low, medium, or high). These filters helped to ensure that only well-represented allelic states that consistently segregated with a trait would lead to designating a SNP as significant. The SNPs located on unmarked or unanchored scaffolds were handled separately because of the inability to map these SNPs to search for candidate genes.

As an alternative approach to measuring the degree of self-fertility by fruit set, we also analyzed seed set for those genotypes that produced fruit. To avoid confusing SNPs segregating for fruit set with SNPs segregating for seed
set, we used only data from genotypes in the high fruit set category for this analysis. Self-unfruitful plants lacked any value of seed per fruit and were removed from the data set along with those genotypes that produced only a few fruits. Seed counts for each individual fruit were considered as unique events to capture the full range of potential seed set within a given genotype. One-way ANOVA was tested on the number of seed per fruit for each individual SNP (JMP Pro 10). Only SNPs with a p-value ≤ 0.0001 were retained in the analysis. The SNPs located on unmarked or unanchored scaffolds were handled as above.

For both fruit and seed datasets, we used p-values generated from the datasets to calculate the corresponding q-value using R to determine the false discovery rate. Lambda ranged from 0.0 to 0.9 by 0.5 while the smoother pi_0 method was applied (Dabney et al., 2014; Storey and Tibshirani, 2003).

To predict whether genotypes assigned to the medium category were truly self-fertile, we used the list of significant SNPs for fruit set to determine whether alleles of these specific SNPs segregated more or less with the alleles linked with self-fertility. If >75% of the alleles matched those alleles correlated with self-fertility, then that plant was considered a truly self-fertile plant. On the other hand, if >75% of the significant alleles matched those alleles correlated with infertility then the fruit on those genotypes was considered to have likely arisen from breakdown of fertility barriers.

Genotyping Single-Nucleotide Polymorphisms of Interest in the S1, S2, and S4 Generations

Prior to genotyping the S3 generation using the Infinium 8303 Potato SNP array, we performed biallelic discrimination assays to genotype specific loci in both the S1 and S2 generations to determine whether SNPs significantly correlated with self-fertility in the F1 generation maintained their significance in the inbred lines. Immature leaf tissue samples of the DRH S1 population were collected for DNA from 17 plants producing fruits from self-pollinations and also from 15 plants that did not produce fruit from self-pollinations. To ensure that both phenotypes were represented in each family, multiple tissue samples for both fruit-producing and non-fruit-producing plants were collected within each family. Likewise, in the S4 generation tissue was collected for DNA from all 13 fruit producing plants and also from 14 plants that did not produce fruit. Five allele specific primers were designed around SNPs to genotype the allelic state of these specific loci in the inbred lines (KBioScience) (Supplemental Table S1, rows 1–5). All SNPs used as primer targets were Infinium high-confidence SNPs from the 8303 SNP array and were chosen because these SNPs were located in regions of highly significant SNPs for self-fruit set. Calling dyes for the primers were FAM and VIC, while ROX was used for a passive reference. Parental lines DM and RH were used as control samples. Applied Biosystems software (ABI 7300 and ABI 7500 systems) was used to analyze the results of real-time PCR reads. Conditions for amplification can be found in Supplemental Table S2. Similar assays were conducted on S4 populations for 23 selected SNPs that remained heterozygous in the S4 generation, except, in this case, there was no selection for self-fertile plants and DNA was extracted from seedlings germinated in vitro to be able to include even those plants that eventually perished prior to flowering in the greenhouse, where plants were rated for vigor using a 1-to-5 scale with 1 representing the most robust plants and 5 representing plants that were dead or nearly dead. ANOVAs were conducted using allelic state as a source of variation and vigor rating as the dependent variable (JMP Pro 10).

Homozygosity Trends in the Inbred Lines

We performed comparative analyses using SNP genotyping data from both the F1 and S4 generations to observe trends imposed on the genome by selecting for self-fertility. We first examined the fate of 718 loci in the S3 for which all three F1 founder plants were heterozygous. We then examined the approach to homozygosity within the three different families. If loci exhibited more than four times the expected level of heterozygosity, they were considered to be excessively heterozygous. Loci exhibiting excessive heterozygosity were used to generate an ideogram using PhenoGram (Wolfe et al., 2013).

Results

Phenotypes Observed across the DRH Generations

In each selfed generation, several genotypes performed poorly and therefore could not be included in the analysis. This group included plants that produced buds with poor pollen shed, buds that did not open, or did not reach the flowering stage as a result of slow maturity or early death. Pollinations were not possible for those plants and therefore we lost the ability to classify them for self-fertility. Overall, the rate of flowering was lower in the inbred lines than in the F1 generation (Table 1). While the percentage of genotypes that produced selfed fruit in each population remained rather constant from generation to generation (range = 32 to 35%) the percentage of genotypes that were classified in the high category exhibited a greater decline in the inbred lines (Table 1). The percentage of self-fruitful genotypes in the high fruit set category dropped from 69% in the F1 generation to 38% in the S2 generation then rebounded to 53% in the S3 generation. In all generations, seed set varied greatly even within the same genotypes. The greatest seed set (361 seeds per fruit) was observed in the S4 generation with more than double the number of seeds observed in the F1 generation. In the F4 generation, the fruit set varied from 1 to 74 fruits per genotype and the seed set varied from 1 to 171 seeds per fruit. In the S4 generation, the fruit set varied from 1 to 13 fruits per genotype, and the seed set varied from 1 to 361 seeds per fruit. Although a significant correlation existed between fruit and seed set
in the F₁ generation, this correlation did not hold true on inbreeding (Table 1). Among all F₁ genotypes (both self-fertile and self-unfruitful) screened for aborted pollen (shriveled or transparent), none was found to have an average of >90% aborted pollen grains. Of the 12 S₂ genotypes screened for aborted pollen, four genotypes were found to have an average aborted pollen rate >90%. Of these four genotypes, one produced fruit on self-pollination and was considered self-fertile, one was inconclusive, and the other two were classified as self-unfruitful.

Of the 98 F₁ DRH genotypes pollinated, seven yielded inconclusive self-fertility phenotypes across plantings in different seasons and therefore were eliminated, leaving 91 genotypes to be included in statistical analyses. Fruit and seed set for all four generations are shown in Table 1. Although we recorded data for the number of fruit set per pollination, we often observed spontaneous fruit set on several genotypes. Spontaneous fruit set was classified as any fruit that set where we had not recorded pollinations via tagging. Seeds produced by the S₂ generation were bulked for all the fruit produced from a single plant were combined before counting. Where the seed set is 1, there was only a single seed in a single fruit.

### Pollen Tube Growth in Self-Fertile and Self-Unfruitful Plants

Pollen tube growth was examined in the styles of two self-fertile and 10 self-unfruitful DRH F₁ plants, for which the phenotypes had been confirmed by >45 self-pollinations under controlled environment conditions. The self-fertile selections for this study, DRH44 and DRH171, produced an average fruit set of 85% and 86 seeds per fruit. No difference was observed in the ability of pollen tubes to penetrate the style or reach the ovary in self-fruitful or self-unfruitful DRH plants (Fig. 2), thus ruling out classic gametophytic self-incompatibility as the mechanism preventing fruit set in the self-unfruitful plants. An unrelated diploid plant, a backcross hybrid (DM 9-9 203 × F₁ [DM 9-9 203 × ID8]) that displayed a strong self-incompatibility reaction with no fruit set after self-pollinations, was used as a control for visualizing the classic self-incompatibility reaction of pollen arrested in the style (Fig. 2). Additionally, male and female viability of self-unfruitful plants was tested by cross-pollinations. Pollen from three self-unfruitful DRH selections was used to pollinate two self-fertile DRH selections to verify pollen viability from self-unfruitful plants. For pollinations using pollen from self-unfruitful plants, we conducted at least 25 cross-pollinations, producing an average fruit set of 34% and seed set of 55 seeds per fruit. For evaluating the female fertility of self-unfruitful plants, we crossed them as female parents using pollen from the self-fruitful selections (at least 25 pollinations per combination), resulting in fruit and seed set in all combinations.

### Table 1. Observed phenotypes in the DRH population over four generations.

<table>
<thead>
<tr>
<th>Generation</th>
<th>No. of genotypes planted</th>
<th>Flowering rate</th>
<th>Fruit present</th>
<th>High fruit set</th>
<th>Seed set range</th>
<th>Correlation of fruit to seed</th>
</tr>
</thead>
<tbody>
<tr>
<td>F₁</td>
<td>103</td>
<td>95</td>
<td>35</td>
<td>69</td>
<td>1–171</td>
<td>0.0175</td>
</tr>
<tr>
<td>S₁</td>
<td>173</td>
<td>40</td>
<td>35</td>
<td>44</td>
<td>1–111</td>
<td>0.1399</td>
</tr>
<tr>
<td>S₂</td>
<td>149</td>
<td>37</td>
<td>32</td>
<td>38</td>
<td>1–86</td>
<td>0.3176</td>
</tr>
<tr>
<td>S₃</td>
<td>458</td>
<td>45</td>
<td>33</td>
<td>53</td>
<td>1–361</td>
<td>0.0624</td>
</tr>
</tbody>
</table>

† Frequency of plants that flowered.
‡ Frequency of plants that produced fruit on selfing.
§ Of those plants that set fruit, the frequency that yielded ≥0.25 fruit per pollination.
¶ Seeds per fruit of all plants that produced fruit; for the F₁, S₁, and S₂ generation, this is the average seeds per fruit for independently extracted fruit; for the S₃ generation, the seeds from all the fruit of a single plant were combined before counting. Where the seed set is 1, there was only a single seed in a single fruit.
# Bivariate fit of the number of fruits per pollination by the number of seeds per fruit.

Single-Nucleotide Polymorphism Analysis of the F₁ Generation

The filters described to remove poor and uninformative SNPs reduced the number of available SNPs for analysis of the DRH F₁ population from 8303 to 2404. Of these, 601 were completely heterozygous in the F₁ because of alternate homozygous states in the DM and RH parents. Of the remaining, 1803 that segregated in the F₁, a total of 88 SNPs was identified as significantly correlated with F₁ fruit set (p-value ≤ 0.0001, q-value < 0.01), three of which were on unanchored scaffolds and the remaining 85 SNPs all located on chromosome 12 (Fig. 3a, b; Supplemental Table S3). All 88 significant SNPs exhibited segregation distortion ranging from 1.8:1 to 3.8:1. The significant SNPs occurred from 11.7 to 61.2 Mbp along the DM genome; however, only 13 occurred in the proximal region of chromosome 12 (11.7–35.7 Mbp), whereas 72 occurred on the distal region between 40.1 and 61.2 Mbp in version 4.03 of the DM genome; however, only 13 occurred in the proximal region of chromosome 12 (11.7–35.7 Mbp), whereas 72 occurred on the distal region between 40.1 and 61.2 Mbp with 12 situated within a 1-Mbp range (55–56 Mbp). The occurrence of so many linked SNPs spread across chromosome 12 and associated with fruit set in the F₁ likely indicates low recombination around a locus controlling the trait as suggested for the 40-Mb linkage in coupling to a locus presumed to control wart (Synchytrium endobioticum (Schilb.), resistance on chromosome 12 in a tetraploid mapping population of potato (Obidiegwu et al., 2015). In the F₁ generation, a total of 39 SNPs was significantly linked with seed set (p-value ≤ 0.0001, q-value < 0.01), one on an unanchored scaffold, and the remaining 38 SNPs located on chromosomes 2 (n = 27), 4 (n = 8), and 12 (n = 3) (Supplemental Table S3). The 27 SNPs on chromosome 2 that were significantly linked with seed set in the F₁ all occurred on the distal region of the chromosome from 2.1 to 21.7 Mbp, the eight on chromosome 4 were tightly clustered between 71.8 and 72.0 Mbp, and the three on chromosome 12 occurred between 59.9 and 61.1 Mbp. The SNPs for seed set located on chromosome 12 were also significantly linked with fruit set.
Genotyping Assays in the S₁ and S₂ Generations
Of the five SNPs that we selected from the F₁ generation analysis for verification in the S₁ and S₂ generations by using allele-specific primers (Supplemental Table S1, rows 1–5), none was significantly linked with self-fertility (p-value > 0.01).

Single-Nucleotide Polymorphism Analysis in the S₃ Generation
Using the filters described to remove poor and uninformative SNPs in the F₁ population, we started the S₃ analysis with 2404 SNPs including the 601 that were completely heterozygous in the F₁ generation because of alternate homozygous allelic states for DM and RH. Of the 2404 SNPs, 280 were removed because, although they
were polymorphic in the DRH population, they were completely homozygous in the three founder DRH F1 plants advanced to the S3 and therefore did not segregate. This reduced the number of available SNPs for analysis from 8303 to 2124. For S3 fruit set, a total of eight significant SNPs was identified \((p\text{-value} \leq 0.0001, q\text{-value} < 0.01)\), all located on anchored scaffolds on chromosomes 4 \((n = 1)\), 9 \((n = 1)\), 11 \((n = 4)\), and 12 \((n = 2)\) (Supplemental Table S3). Of the eight SNPs significantly linked with selfed fruit set in the S3 generation, none had been significantly linked with selfed fruit set in the F1 generation \((0.04 < p\text{-values} < 0.81)\). For S3 seed set, a total of 99 significant SNPs \((p\text{-value} \leq 0.0001, q\text{-value} < 0.01)\) was identified, all located on anchored scaffolds on chromosomes 1, 3, 4, 8, 9, and 12 (Supplemental Table S3).

**Comparison of Heterozygosity in the F1 and S3 Generations**

The overall frequency of apparently heterozygous loci in the F1 generation was 37% \((2404\text{ loci that were either completely heterozygous} [n = 601], \text{exhibited} 1:1 \text{ segregation} [n = 1421] \text{ or distorted segregation} [n = 383] \text{ of 6487 loci after filtering}]\). Of the 2404 loci that were heterozygous in the F1 generation, 718 were heterozygous in all three F1 founder plants. Only 150 of these 718 loci \((21\%)\) were completely homozygous in the S3 compared with the expected 87.5% based on probability. In fact, 48 of these 718 SNP loci \((7\%)\) retained >50% heterozygosity, which is four times the expectation (Supplemental Table S4). Of the 48 highly heterozygous loci, 46 were distributed across four chromosomes with two on the unanchored scaffolds (chromosome 0), 20 on chromosome 1, nine on chromosome 2, 16 on chromosome 4, and one on chromosome 10 (Fig. 4). Alternatively, approach to homozygosity can be considered on a per family basis. In this regard, the frequency of loci with retained heterozygosity in the three families was 49% \((751\text{ with some heterozygosity of 1531 SNPs that were heterozygous in the F1 plant})\) for family DRH16, 55% \((620\text{ of 1119 SNPs})\) for family DRH76, and 52% \((675\text{ of 1278 SNPs})\) for family DRH90. Likewise, the...
frequency of loci with >50% retained heterozygosity in the S3 was 424 (28%), 304 (27%), and 134 (10%) for the three families, respectively. The per-family distribution of loci with high retained heterozygosity across the genome was not as compactly preserved (Fig. 4B–D) as when the 718 loci that were heterozygous in all three F1 founder plants were observed above.

Of the 48 highly heterozygous SNPs in the population as a whole, two (solcap_snp_c2_27677 and solcap_snp_52477, both on chromosome 1) were significantly linked with seed set in the S3 generation. For 137 SNP...
loci, the homozygous condition was found exclusively for the DM allele in the S₃, whereas there was only a single case (solcap_snp_c1_1944 on chromosome 12, also significant for selfed fruit set in the F₁) of exclusive homozygosity for the RH allele. The loci that were fixed for the DM allele were distributed across seven linkage groups excluding chromosomes 1, 2, 10, 11, and 12 (Fig. 4A; Supplemental Table S5); the preponderance of fixed DM loci occurred on chromosome 6 (n = 21; 15%), chromosome 7 (n = 26; 19%), and chromosome 8 (n = 58; 42%).

### Segregation of Selected Loci from S₃ to S₄ and Linkage of Specific Allelic States with Plant Vigor

The S₃ population consisted of six families: two each derived from DRH16, 76, and 90. In contrast to earlier generations, the S₄ was established first by germinating seeds in vitro, extracting DNA from in vitro plantlets, then acclimating the plants to the greenhouse during the fall 2013 growing season. Of 300 seeds planted, 146 plants were acclimated to the greenhouse. Plants were rated for vigor on a 1-to-5 scale with 1 representing the most vigorous plants and 5 the least. Mean vigor per family ranged from 2.2 to 3.5.

The S₄ population consisted of six families: two each derived from DRH16, 76, and 90. In contrast to earlier generations, the S₄ was established first by germinating seeds in vitro, extracting DNA from in vitro plantlets, then acclimating the plants to the greenhouse during the fall 2013 growing season. Of 300 seeds planted, 146 plants were acclimated to the greenhouse. Plants were rated for vigor on a 1-to-5 scale with 1 representing the most vigorous plants and 5 the least. Mean vigor per family ranged from 2.2 to 3.5.

### Discussion

#### Self-Fertility in the F₁ Generation

The wide range of both fruit set and seed set observed in our population of diploid potato supports the idea that self-fertility is not a strictly qualitative trait. The segregation pattern for self-fertile plants in the F₁ generation did not follow a one-to-one Mendelian ratio expected for segregation of a single gene in a homozygous × heterozygous diploid cross. While fruit and seed set were weakly correlated in the F₁ generation, these traits do not necessarily measure self-fertility to the same degree. Although self-fertility factors certainly influence the growth of the pollen tube there are many other factors to consider such as temperature, moisture, and the number of pollen grains that land on the stigma. Even in the most highly self-fruitful plants, not every self-pollination resulted in fruit and seed set.

Since seed set was inconsistent within genotypes, we chose to rely more heavily on fruit set for identifying genomic regions linked with self-fertility. Also, given that fruit set variation decreased from the F₁ to the S₃ generation while seed set variation increased, we believe that our efforts to select for the self-fertile phenotype had a greater impact on fruit than seed set. Perhaps, within ovaries, there was a wide range of ovule numbers that does not remain consistent and thereby can yield varying numbers.
of seeds. The helicoid cyme type of inflorescence characteristic of *Solanum* typically bears larger fruit with more seeds on the first flowers to bloom within the inflorescence as with tomato (*Lycopersicon esculentum* Mill.). We did not score fruit position on an infructescence. Factors such as temperature, resource competition, and position of a flower within an inflorescence can influence the success of fruit development (Diggle, 1995; Sato et al., 2000). Previous studies have used fruit set resulting from self-pollination as an indicator for degree of self-compatibility (Birhman and Hosaka, 2000; Lloyd and Schoen, 1992).

However, as we had no evidence of pollen tubes arrested in the styles of self-unfruitful plants after self-pollination, the typical pattern of gametophytic self-incompatibility, we have chosen to describe the trait we measured in our population as self-fertility. Presumably, some type of late acting self-incompatibility is occurring in our self-unfruitful plants (Seavey and Bawa, 1986).

The distal end of chromosome 12, where many significant SNPs for fruit set were found, also harbors the *S*-locus inhibitor (*Sli*) identified by Hosaka and Hanneman (1998a,b). The *Sli* gene was identified in a mutant of *S. chacoense* and acts sporophytically to induce self-compatibility. As *S. chacoense* is a weedy wild diploid species, the *Sli* gene is limited in usefulness to breeders unless considerable backcrossing is performed to return to cultivated types. Multiple successful efforts have been made to breed the *Sli* gene into self-incompatible stock to confer self-compatibility (Birhman and Hosaka, 2000; Lindhout et al., 2011; Phumichai et al., 2005). As the *Sli* gene functions to block *S*-allele interactions (Hosaka and Hanneman, 1998a; Hosaka and Hanneman, 1998b) we would expect that if the *Sli* gene was present in the DRH germplasm, we would not have observed pollen tube growth in self-unfruitful plants. Since pollen tubes grew freely in the styles of self-unfruitful plants, we believe that other factors contributed to the segregation of self-fertility in the DRH population.

The occurrence of open-pollinated fruit also posed a potential source of error in the experiment. Plants that consistently produced spontaneous fruit were generally considered self-fertile, even if a controlled self-pollination had not been performed, as has been assumed in previous studies of plants under controlled conditions (Stone et al., 2006). Since open-pollinated fruit generally appeared on plants that were already deemed self-fertile as a result of fruit production from intentional self-pollination, and since open pollinated fruit generally appeared consistently on some plants but not on others, it is likely that these fruit were the result of self-pollination.

By design of the Infinium 8303 Potato SNP array, we could only genotype RH SNPs that shared at least one allele with DM. Therefore, there is a possible loss of information as we were unable to score RH SNPs consisting of two alleles different from DM. It is possible that such SNPs might be linked to distorted loci linked with self-fertility.

**Changes Observed in the Inbred Lines**

The phenotypic trends from one generation to the next revealed the effects of inbreeding depression from repeated selfing. Selfing has long been known to cause severe inbreeding depression in potato (De Jong and Rowe, 1971). The skewed segregation that was observed in the DRH population may be linked with genes for plant development and vigor. Compared with the F1 generation, the inbred generations had more plants that could not be pollinated because of buds with poor pollen shed, buds that did not open, or plants that did not reach the flowering stage altogether as a result of slow maturity or early death. The most extreme reduction in flowering rate occurred during the transition from F1 to S3 as described previously (De Jong and Rowe, 1971). Since phenotypic classification could not be determined for many genotypes, they could not be included in the statistical analyses and their removal may have imposed a bias.

While fruit-set rates remained rather constant in all four generations, the quick decrease in percentage of plants producing high fruit set in the S1 and S2 generations was likely another indicator of inbreeding depression or else reflected reshuffling of either alleles at modifier genes that controlled an alternative route to self-fertility. In the S3 generation, the slight increase in percentage of plants producing high fruit set likely reflected the fact that we were selecting for self-fertility, and that despite inbreeding depression, the alleles favoring the self-fertile phenotype were slowly becoming more favored.

Over the four generations, seed set was unpredictable. Fruit set and seed set became less correlated over the course of inbreeding, indicating that there were multiple factors that controlled these traits. Surprisingly, seed set in the S3 generation surpassed that of the previous generations. All fruit with seed count greater than previously recorded were found on genotype SC-03, which was calculated to be ~85% homozygous. This finding demonstrates that through selection, it is possible to find high seed producing selfed plants with a highly homozygous background.

**Alternative Mechanisms for Determining Self-Fertility in the S3 Generation**

Because a lengthy section of chromosome 12 was so highly correlated with fruit set in the F1 (Fig. 3), and the RH allele for solcap_snp_c1_1944 was fixed in the selfed populations, the primary factor linked with self-fertility in the F1 likely resides near the position of this SNP. However, since the self-fertile phenotype continued to segregate in S1 through S4 plants, there must have been different genes determining self-fertility in the inbred lines. Further selfing and a continued decrease in heterozygosity would assist in the identification of specific genes and interactions necessary for continued fruit set in a nearly homozygous background.
Reduction in Heterozygosity

One of the goals of the project was to generate highly homozygous self-fertile inbred lines to develop a tool for future genetic studies. As demonstrated by the reduced percentage of heterozygosity in the S4 generation, it can be concluded that the inbred lines were overall more homozygous than the F1 generation. While traditional methods for calculating the rate of heterozygosity reduction would predict only 12.5% heterozygosity remaining by the S4 generation, we observed 70% heterozygosity remaining in the S5 in this study. While selfing diploid potato up to the S5 generation using Sli, Phumichai et al. (2005) observed a slower than expected approach to homozygosity by using restriction fragment length polymorphism and amplified fragment length polymorphism markers, possibly from selection for seed germination and vegetative vigor. We expect that many loci of maintained heterozygosity may yield a fitness advantage or otherwise prove beneficial in the selection for self-fertility.

The design of the Infinium 8303 Potato SNP array may also have influenced the apparent heterozygosity in our population, as the presence of duplicate loci with alternative alleles at the position of the SNP would be misconstrued as heterozygosity. The use of a CNV filter developed from a monoploid panel may inform us of loci that are susceptible to duplication, assuming that duplication in the monoploids reflects duplication in RH. Unique CNV in RH would not have been removed by our CNV filter and thus would have contributed to the inflation of heterozygosity estimated in the S5 generation. When we examined a subset of 718 loci that were heterozygous in all three of our F1 founder individuals, those that remained heterozygous in the S5 were generally restricted to a few genomic regions. However, the ideograms for the individual S5 families presented a more complex trend of retained heterozygosity. Selection for vegetative vigor or genes required for functional male and female gametes may have alternative strategies (Phumichai et al., 2005). The observation that some chromosomes bore many highly heterozygous SNPs in one family but none in others supports the concept that heterozygosity is not required for self-fertility and that achieving homozygous, self-propagating inbred lines is possible given that we can overcome the effects of inbreeding depression and genetic load. The common heterozygous regions in all three families may be more resistant to inbreeding.

The results showing that different families attained different numbers of unique fixed alleles demonstrates that not all families achieved homozygosity at the same rate. This effect could possibly be due to selection of a different combination of genes or haplotypes for self-fertility among the families. Our analysis of the S5 generation was conducted on all seedlings that managed to germinate in vitro rather than those that actually flowered sufficiently in the greenhouse to allow for selfing. In fact, some of the S5 seedlings that went into the KASP bioassays perished before they could be established ex vitro. Most of the SNP loci examined that were heterozygous in S5 segregated normally in S6 with only three showing a significant heterozygote advantage with regard to vigor. An examination of genes in genomic regions that remain stubbornly heterozygous through inbreeding may reveal functional causes for retained heterozygosity.

Conclusion

Through genome-wide SNP analysis of the F1 DRH generation, we identified chromosome 12 as harboring the gene or genes responsible for the segregation of self-pollinated fruit set. Continued segregation of this trait in advanced selfed generations indicated the contribution of other genomic regions. While a great reduction in heterozygosity occurred from the process of inbreeding for three generations, regions of heterozygosity still remained throughout the genome, some common to all three families of advanced inbreds and some unique. It remains to be determined whether a minimum level of heterozygosity is necessary to maintain self-fertility. Efforts to continue advancing the inbred generations to reduce heterozygosity will further clarify which genes are necessary for self-fertility and will help in the production of more reliably self-fertile diploid potatoes. The source of self-fertility in the germplasm employed in this study represents a valuable resource for generating homozygous building blocks for hybrid potato breeding at the diploid level. Self-fertility in S. tuberosum germplasm may facilitate introgression into other adapted germplasm. Crosses between advanced generation inbreds from this population and others derived from different diploid germplasm may restore sufficient vigor to allow derivation of improved inbred lines.

Supplemental Information Available

Supplemental information is available with the online version of this manuscript.

Author Contributions

BAP, SHH, FPEL, and AGD conducted the research, analyzed the data, and drafted the manuscript; JC and DSD conducted the SNP analyses and contributed to the manuscript; MH and CRB sequenced the monoploid clones and provided bioinformatics support; and RV designed the research, contributed to the manuscript, and assisted in the data analysis.

Acknowledgments

This research was supported in part by National Science Foundation Award number 12-1852-02, USDA Hatch Project VA-135853, the Department of Horticulture, Virginia Tech, and the SolCAP Solanaceae Coordinated Agricultural Project. During the conduct of this research, AGD was working at Virginia Tech on a sabbatical leave from his position in Greece. We would like to thank Jeff Burr for care and maintenance of greenhouse plants and Jonathan Stallings and Jing Niu for statistical consulting through the Laboratory for Interdisciplinary Statistical Analysis (LISA), Department of Statistics, Virginia Tech.