nature genetics

LETTERS

https://doi.org/10.1038/s41588-018-0319-1

The genetic basis of inbreeding depression in potato

Chunzhi Zhang[1,2,3,8], Pei Wang[1,8], Die Tang[2,8], Zhongmin Yang[4], Fei Lu[5], Jianjian Qi[6], Nilesh R. Tawari[7], Yi Shang[1], Canhui Li[1] and Sanwen Huang[2,3*]

Inbreeding depression confers reduced fitness among the offspring of genetic relatives. As a clonally propagated crop, potato (*Solanum tuberosum* L.) suffers from severe inbreeding depression; however, the genetic basis of inbreeding depression in potato is largely unknown. To gain insight into inbreeding depression in potato, we evaluated the mutation burden in 151 diploid potatoes and obtained 344,831 predicted deleterious substitutions. The deleterious mutations in potato are enriched in the pericentromeric regions and are line specific. Using three F_2 populations, we identified 15 genomic regions with severe segregation distortions due to selection at the gametic and zygotic stages. Most of the deleterious recessive alleles affecting survival and growth vigor were located in regions with high recombination rates. One of these deleterious alleles is derived from a rare mutation that disrupts a gene required for embryo development. This study provides the basis for genome design of potato inbred lines.

The adverse effects of inbreeding were first documented by Charles Darwin through extensive experiments on plants[1]. Inbreeding depression in clonally propagated plants is expected to be more severe than that in seed plants, as more deleterious mutations accumulate during asexual reproduction, whereas purging of these mutations by recombination is limited[2]. Potato (*Solanum tuberosum* L.) is the most important tuber crop, but its improvement is slow as a result of tetrasomic inheritance. Moreover, its clonal propagation bears a considerable carbon footprint[3]. To overcome these drawbacks, we and two other groups sought to re-domesticate potato into an inbred line-based diploid crop propagated by seeds[4–6]. Inbreeding depression has severely hampered the development of elite inbred lines, and there is therefore a need to understand its genetic basis.

Deleterious mutations are considered to be the major cause of inbreeding depression[7–9]. The detrimental effects of deleterious recessive alleles, which reduce the fitness of progeny, would be exposed through inbreeding. To better visualize deleterious alleles in potato, it would be ideal to develop selfing populations from diploid clones that would segregate for only two alleles at each locus. An obvious obstacle is that most diploid clones are self-incompatible, a trait that is conferred by the S-ribonucleases (S-RNases) specifically expressed in the style[10]. We searched over 200 diploid clones for natural self-compatible mutants and identified two clones, PG6235 and PG6359. We also used bud pollination to force selfing of PG6226, as the expression of *S-RNase* is low at the bud stage (Supplementary Fig. 1). In total, we obtained 104, 223, and 201 diploid progenies of PG6226, PG6235, and PG6359, respectively, which were subsequently re-sequenced and analyzed.

We first evaluated the mutation burden in 3 parental clones and 148 other diploid landraces (Supplementary Table 1) using amino acid conservation modeling[11]. We obtained 344,831 predicted deleterious substitutions, which were enriched in the pericentromeric regions (Fig. 1a and Supplementary Fig. 2), consistent with what is observed in fruit flies and maize[12–14]. In potato, there were 64.46% fewer homozygous deleterious mutations than heterozygous ones ($P = 2.15 \times 10^{-90}$; Fig. 1b). We analyzed the deleterious allele frequency and found that 50.87% of the deleterious mutations were rare, with deleterious allele frequency ≤0.05 (Fig. 1c). The wild species *Solanum candolleanum* was the progenitor of cultivated potatoes. Nearly 27% of deleterious mutations also exist in *S. candolleanum* and showed relatively higher deleterious allele frequency in landraces; thus, they are considered to be potentially ancestral. On average, 11.32% of predicted deleterious substitutions were overlapped between any two diploids (Fig. 1d). These results indicate that the mutation burden in potato is line specific.

To analyze the genotypes of the F_2 progeny, we developed a pipeline for parent-independent genotyping in potato (Supplementary Fig. 3). The bin maps were constructed using phased single-nucleotide polymorphisms (SNPs). In total, 4,255 bins were obtained across three populations, with an average length of 527 kb (Supplementary Table 2 and Supplementary Fig. 4). We captured 10,294 recombination events, corresponding to approximately 19.5 crossovers in each F_2 individual. This allowed us to analyze the segregations and effects of deleterious alleles in selfing populations.

We found that 15 blocks containing ~1,000 bins exhibited segregation distortions in the selfing populations (χ^2 test, $P < 0.001$; Supplementary Table 2), a percentage that is much higher than that reported in maize[15]. The total length of the distorted regions varied among three F_2 populations. PG6226 and PG6235 were highly heterozygous clones (Supplementary Fig. 5a,b), and they had more distorted bins (32.8% in PG6226 and 31.7% in PG6235). By contrast, contiguous homozygous segments (11.5–26.7 Mb) were detected on six chromosomes in PG6359 (Supplementary Fig. 5c), suggesting that this landrace originated from selfing or inbreeding between

[1]The AGISCAAS-YNNU Joint Academy of Potato Sciences, Yunnan Normal University, Kunming, China. [2]Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [3]Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of the Ministry of Agriculture, Sino-Dutch Joint Laboratory of Horticultural Genomics, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China. [4]College of Horticulture, Northwest Agriculture and Forest University, Yangling, China. [5]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [6]Inner Mongolia Potato Engineering and Technology Research Centre, Inner Mongolia University, Hohhot, China. [7]Computational and Systems Biology, Genome Institute of Singapore, Singapore, Singapore. [8]These authors contributed equally: Chunzhi Zhang, Pei Wang, Die Tang.
*e-mail: huangsanwen@caas.cn

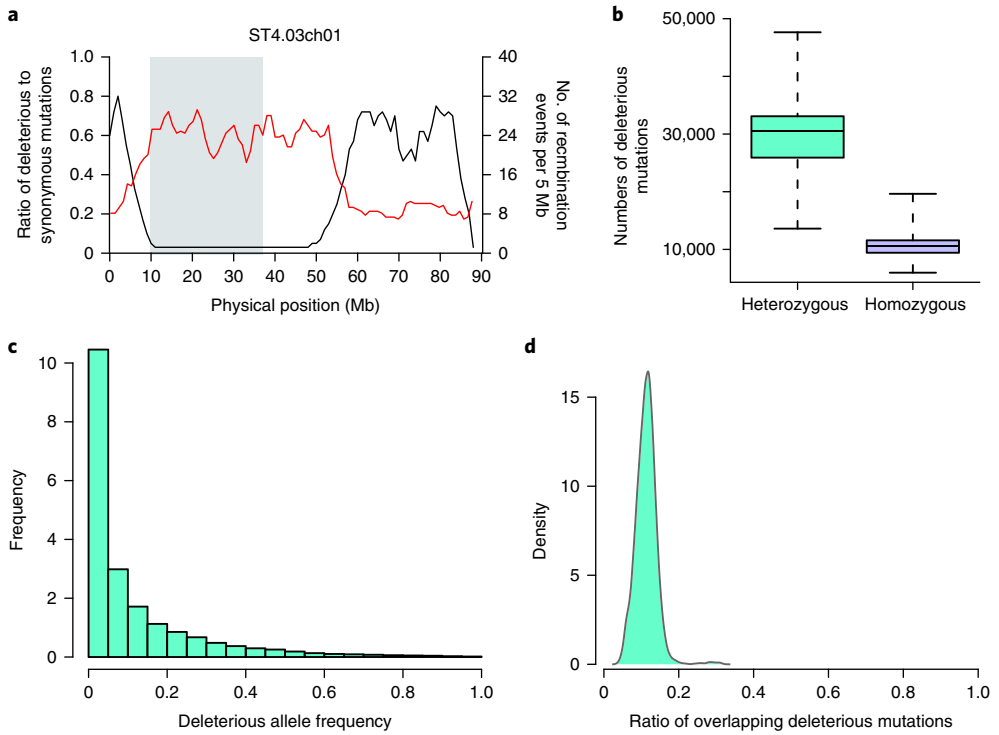

Fig. 1 | Identification of genome-wide deleterious mutations in potato. a, Deleterious mutations are enriched in pericentromeric regions. The *x* axis indicates the physical position. The red line indicates the ratio between deleterious substitutions and synonymous mutations (window size = 5 Mb, step = 1 Mb), and the black line indicates the number of recombination events per 5 Mb. The gray shaded box indicates the position of the pericentromeric region. **b**, Number of heterozygous (green) and homozygous (purple) deleterious mutations in 151 diploid clones. Each box represents the mean and interquartile range. The top whisker denotes the maximum value and the bottom whisker denotes the minimun value. **c**, Deleterious allele frequency in 151 diploid clones. **d**, High diversity of mutation burden in diploid potatoes. The *x* axis indicates the ratio of overlapping deleterious mutations in any two diploid clones, and the *y* axis indicates the density of the ratio.

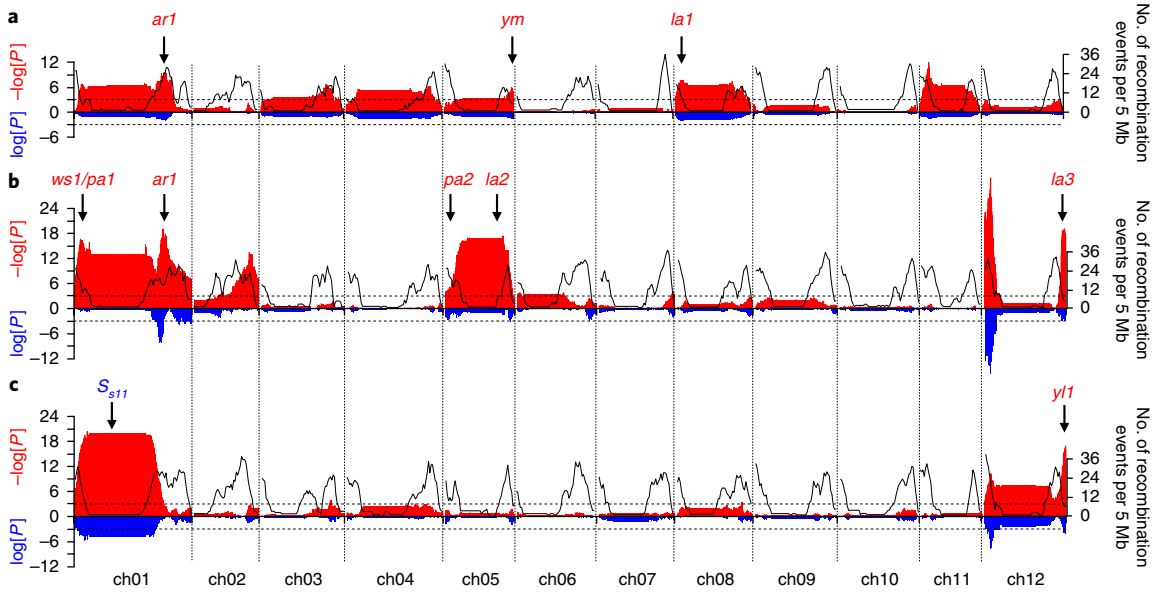

Fig. 2 | Genome-wide segregation distortions. The segregation distortions of bins at the seedling stage in the selfing populations of PG6226 (**a**), PG6235 (**b**), and PG6359 (**c**). The left *y* axis represents the −log[*P*] and log[*P*] for the χ^2 value of each bin at the zygotic (red) and gametic (blue) stages, respectively. The black curves indicate the number of recombination events per 5 Mb. The arrows indicate the positions of genes causing a deleterious phenotype.

genetic relatives in South America. The homozygous regions cover 15.82% of the genome, but contain only 3.87% of the deleterious mutations. We speculate that some large-effect deleterious mutations were purged during this process, and thus there were fewer distorted regions in the selfed progeny. Genome-wide segregation distortions have been widely reported in potato[16–20], and

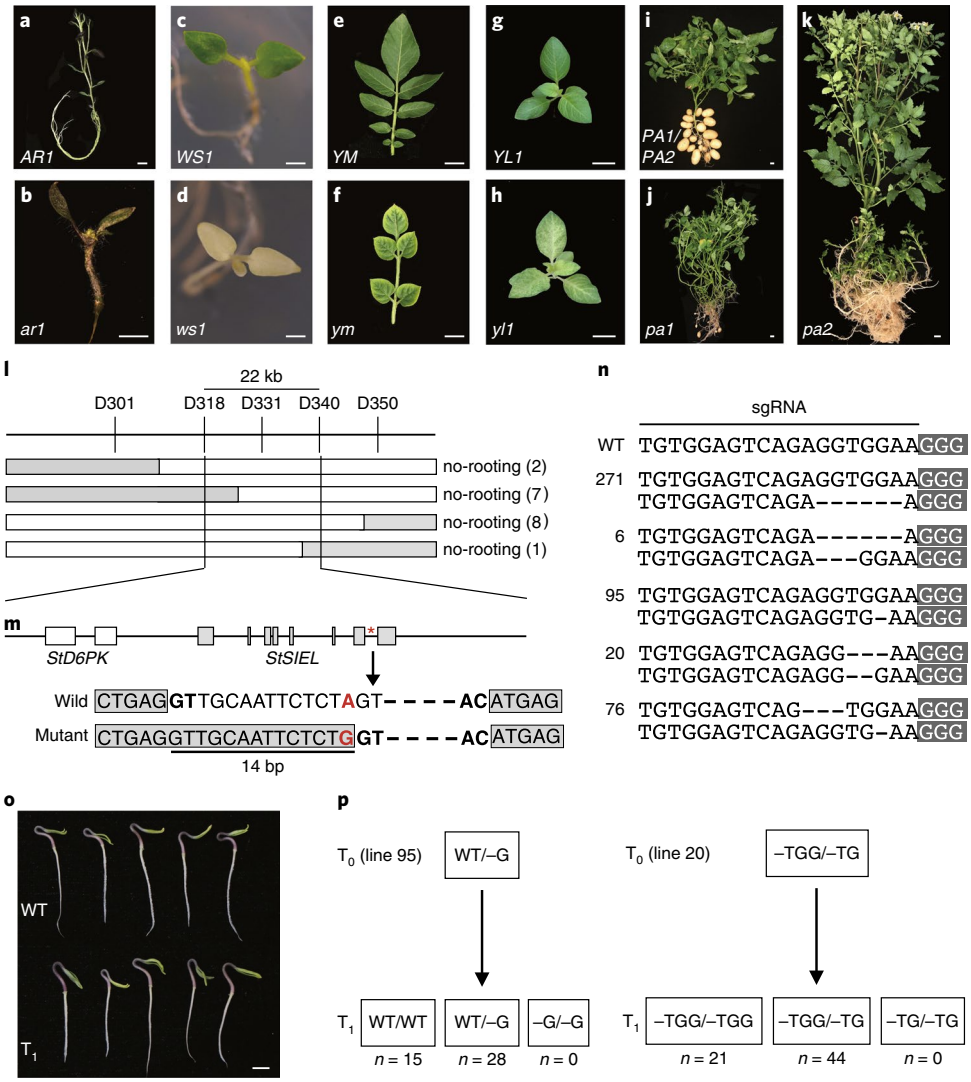

Fig. 3 | Identification of large-effect deleterious alleles. a,**b**,**c**,**d**,**e**,**f**,**g**,**h**,**i**,**j**,**k**, The phenotypes of normal rooting (**a**) and abnormal rooting (**b**), green seedling (**c**) and albino seedling (**d**), normal leaf (**e**) and yellow-margin leaf (**f**), normal leaf (**g**) and light-yellow leaf (**h**), normal branching (**i**) and increased branching (**j**), and strong vegetative growth (**k**). The photographs in **a**, **b**, **e**, and **f** show progeny of PG6226; those in **c**, **d**, **i**, **j**, and **k** show progeny of PG6235; and those in **g** and **h** show progeny of PG6359. Scale bars represent 0.5 cm in **a**–**d** and 2 cm in **f**–**k**. **l**, Fine mapping of *ar1*. Gray boxes indicate the heterozygous genotype and white boxes the homozygous recessive genotype of *ar1*. The numbers in parentheses indicate the number of recombinants with the same genotype. **m**, The candidate gene of *ar1*. Gray boxes indicate exons, and bold letters indicate splicing sites. The red letters indicate the only SNP site in *StSIEL* between normal-rooting and abnormal-rooting plants. **n**, Mutations generated by knockout of *StSIEL* in tomato. **o**, Rooting of T_1 seeds carrying mutant *StSIEL*. Scale bar represents 0.5 cm. **p**, Segregation of mutations in the T_1 generation. WT, wild type. The hyphens indicate deletions.

distorted regions vary according to the tested populations, which also demonstrates the diversity of deleterious mutations in potatoes.

Many factors affect distorted segregation, including gametic and zygotic selection. Among 15 blocks with strong segregation distortions, 4 were affected by gametic selection (χ^2 test, $P < 10^{-5}$; Fig. 2 and Supplementary Table 3). The *S* (self-incompatibility) locus on chromosome 1 is known as a common factor that causes distorted segregation by gametic selection[17,18]. In PG6359, two *S-RNase* alleles were obtained by de novo assembly of RNA-sequencing data from the style, and the expression of the S_{s12} allele was 100-fold greater than that of the S_{s11} allele. Pollen containing S_{s12} was rejected during self-pollination, resulting in the absence of the homozygous $S_{s12}S_{s12}$ genotype in the selfed progeny. Thus, the low expression level of S_{s11} in PG6359 may lead to semi-self-compatibility as well as the distorted segregations of flanking regions. The 11 remaining distorted blocks were a result of zygotic selection, which accounted for 70.2% of all distorted regions. Given that zygotic selection is the key factor resulting in inbreeding depression in potato, we then explored the deleterious alleles harbored in these regions.

Some large-effect deleterious alleles still exist in potato in the heterozygous state, which could result from the absence of recombination that might otherwise purge deleterious alleles. In this study, we identified five distorted blocks containing lethal mutations, which we named *lethal allele 1* (*la1*), *lethal allele 2* (*la2*), *lethal allele 3* (*la3*), *abnormal rooting 1* (*ar1*), and *white seedling 1* (*ws1*) (Fig. 2a,b and Supplementary Table 4). The homozygous recessive genotypes of *la1*, *la2*, and *la3* were absent in the selfed progeny (Supplementary Fig. 4). The *ar1*/*ar1* genotype had defects in lateral root initiation and shoot differentiation (Fig. 3a,b and Supplementary Fig. 6a) and survived only on Murashige and Skoog medium. The *ws1* mutant was deficient in chlorophyll and died a few days after germination (Fig. 3c,d and Supplementary Fig. 6a).

We also identified four deleterious alleles affecting growth vigor. Some progeny of PG6226 showed shorter shoots as well as yellow

leaf margins (Fig. 3e,f) as a result of the gene *yellow margin* (*ym*) on chromosome 5 (Supplementary Fig. 6c). The leaves of some progeny of PG6359 showed a virescent phenotype; the underlying gene was named *yellow leaf 1* (*yl1*) and was mapped to the end of chromosome 12 (Fig. 3g,h and Supplementary Fig. 6d). The growth vigor of the selfing population of PG6235 could be divided into three categories: group I had normal branching and tubers (Fig. 3i), group II had more branching and fewer tubers than normal plants (Fig. 3j), and group III had strong vegetative growth and no tubers (Fig. 3k). Bulked segregant analysis mapped increased branching (as a result of *plant architecture 1* (*pa1*)) and strong vegetative growth (as a result of *plant architecture 2* (*pa2*)) to chromosome 1 and chromosome 5, respectively (Supplementary Fig. 6e,f and Supplementary Table 4). All four deleterious alleles affecting growth vigor are located in the genomic regions with segregation distortions. The *pa2* allele is colocalized with the early-maturing gene *StCDF1*[21]. The late-maturing allele in Group III plants might lead to their strong vegetative growth.

To pinpoint the gene underlying the abnormal rooting of *ar1*, we sowed over 30,000 F_2 seeds derived from a cross between PG6226 and *Solanum chacoense* 34–28 and obtained ~2,000 seedlings that showed abnormal rooting. Genotyping of these plants narrowed down the *ar1* gene to a 22-kb region (Fig. 3l). The region contains two genes, PGSC0003DMG400026052 and PGSC0003DMG400026051, which correspond to the *Arabidopsis* genes *D6 PROTEIN KINASE* (*StD6PK* in potato) and *SHORT-ROOT INTERACTING EMBRYONIC LETHAL* (*StSIEL* in potato), respectively (Fig. 3m)[22,23]. Only one variation, SNP 01S079 (ST4.03ch01: 69339079) in the intron of *StSIEL*, was detected in this region. RNA sequencing showed that this SNP led to the alternative splicing of *StSIEL* (Fig. 3m). To confirm the function of *StSIEL*, we knocked out the homolog of this gene in tomato (*Solanum lycopersicum*) using clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein 9 (Cas9) (Fig. 3n). Analysis of the T_1 plants demonstrated that all of the homozygous mutants containing frameshift mutations were lethal (Fig. 3o,p). This result suggests that the *StSIEL* allele in potato, which is the candidate gene of *ar1*, should have the same function. The *ar1* allele was also identified in PG6235 (Fig. 2b). The level of nucleotide sequence identity of haplotypes flanking *ar1* in PG6226 and PG6235 is >99%, indicating that the *ar1* allele originated from a common ancestor (Supplementary Fig. 7). To evaluate the frequency of the *ar1* allele in potato germplasm, we searched for SNP 01S079 in the 148 remaining diploid clones. None of the clones contained the mutant SNP, suggesting that *ar1* is a rare allele.

In addition to the alleles with substantial deleterious effects, there were numerous mutations that had minor effects on fitness. For example, pollen viability segregated in all three selfing populations, but no major genetic locus for this trait was identified (Supplementary Fig. 8a–c,e–g). In addition, tuber weight showed continuous variation in the selfing population of PG6226, and bulked segregant analysis suggested that it was controlled by multiple loci with minor effects (Supplementary Fig. 8d,h). Accumulation of these minor-effect deleterious alleles can have a severe impact on fertility and yield; future work should aim to identify and eliminate these alleles.

Identifying the deleterious mutations that have accumulated during asexual propagation will allow breeders to effectively manage or purge them through sexual hybridization. The deleterious mutations in potato were enriched in the pericentromeric regions (Supplementary Fig. 2), and most of these mutations were line specific; thus, breeders can maintain the mutations in the heterozygous state by dedicatedly designed crosses. By contrast, all of the large-effect deleterious alleles identified in this study except *la3* on chromosome 5 are located in regions with high recombination rates (Fig. 2). Thus, it would be possible to purge these alleles via recombination. Notably, the short arm of chromosome 1 in PG6235 contains two tightly linked deleterious recessive alleles in repulsion phases (*ws1* in haplotype A and *pa1* in haplotype B), which is indicative of pseudo-overdominance. To stack all favorable alleles into one haplotype, a large segregating population is needed to screen for recombinants between these genes.

In this study, we analyzed the genome-wide distribution of deleterious mutations and examined their effects on phenotype in potato. Most of the deleterious mutations were line specific, which suggests that there will be a high degree of heterosis when two inbred lines from different lineages are crossed. Although some large-effect deleterious recessive mutations exist in potato, they can be effectively purged by recombination. Our data not only provide a basis for genome design of potato inbred lines, but they also contribute to our understanding of the genetic basis of inbreeding depression in clonally propagated plants.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0319-1.

Received: 20 June 2018; Accepted: 21 November 2018;
Published online: 14 January 2019

References

1. Darwin, C. *The Effects of Cross and Self Fertilization in the Vegetable Kingdom* (John Murray, London, 1876).
2. Ramu, P. et al. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat. Genet.* **49**, 959–963 (2017).
3. Vreugdenhil, D. et al. *Potato Biology and Biotechnology: Advances and Perspectives* (Elsevier Science, Amsterdam, 2007).
4. Lindhout, P. et al. Towards F_1 hybrid seed potato breeding. *Potato Res.* **54**, 301–312 (2011).
5. Li, Y., Li, G., Li, C., Qu, D. & Huang, S. Prospects of diploid hybrid breeding in potato. *Chin. Potato* **27**, 96–99 (2013).
6. Jansky, S. H. et al. Reinventing potato as a diploid inbred line–based crop. *Crop Sci.* **56**, 1412–1422 (2016).
7. Remington, D. L. & O'Malley, D. M. Evaluation of major genetic loci contributing to inbreeding depression for survival and early growth in a selfed family of *Pinus taeda*. *Evolution* **54**, 1580–1589 (2000).
8. Chen, Z. J. Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* **14**, 471–482 (2013).
9. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).
10. Takayama, S. & Isogai, A. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56**, 467–489 (2005).
11. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
12. Haddrill, P. R., Halligan, D. L., Tomaras, D. & Charlesworth, B. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome. Biol.* **8**, R18 (2007).
13. Mackay, T. F. C. et al. The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173–178 (2012).
14. Rodgers-Melnick, E. et al. Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc. Natl Acad. Sci. USA* **112**, 3823–3828 (2015).
15. McMulen, M. D. et al. Genetic properties of the maize nested association mapping population. *Science* **327**, 737–740 (2009).
16. Bonierbale, M. W., Plaisted, R. L. & Tanksley, S. D. RFLP maps based on a common set of clones reveals modes of chromosomal evolution in potato and tomato. *Genetics* **120**, 1095–1103 (1988).
17. Gebhardt, C. et al. RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theor. Appl. Genet.* **83**, 49–57 (1991).
18. Jacobs, J. M. E. et al. A genetic map of potato (*Solanum tuberosum*) integrating molecular markers including transposons and classical markers. *Theor. Appl. Genet.* **91**, 289–300 (1995).
19. Felcher, K. J. et al. Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS ONE* **7**, e36347 (2012).
20. Endelman, J. B. & Jansky, S. H. Genetic mapping with an inbred line-derived F_2 population in potato. *Theor. Appl. Genet.* **129**, 935–943 (2016).
21. Kloosterman, B. et al. Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* **495**, 246–250 (2013).

22. Zourelidou, M. et al. The polarly localized D6 PROTEIN KINASE is required for efficient auxin transport in *Arabidopsis thaliana*. *Development* **136**, 627–636 (2009).
23. Koizumi, K., Wu, S., MacRae-Crerar, A. & Gallagher, K. L. An essential protein that interacts with endosomes and promotes movement of the SHORT-ROOT transcription factor. *Curr. Biol.* **21**, 1559–1564 (2011).

Acknowledgements

We thank J. Yan and F. Tian for critical reading of the manuscript; G. Zhu for discussions and project coordination; Z. Peng, X. Xu, and S. Feng for phenotyping and genetic transformation; and Z. Wang, W. Xiao, and D. Zhang from Yinmore Group for greenhouse assistance. This work was supported by the Agricultural Science and Technology Innovation Program (ASTIP-CAAS to S.H.), the Agricultural Science and Technology Innovation Program Cooperation and Innovation Mission (CAAS-XTCX2016 to S.H.), Advanced Technology Talents in Yunnan Province (2013HA025 to S.H.), and National Natural Science Foundation of China (31601360 to C.Z.). This work was also supported by the Ministry of Agriculture and Rural Affairs of PRC and the Shenzhen municipal (The Peacock Plan KQTD2016113010482651 to S.H.) and Dapeng district governments.

Author contributions

S.H. and C.Z. designed the experiments and wrote the manuscript. C.Z., D.T., and F.L. performed the bioinformatics analyses. P.W. carried out the phenotyping assay. Z.Y. performed the cloning and functional analysis of *ar1*. J.Q. provided the genomic data for wild species *S. candolleanum*. N.R.T. predicted the deleterious mutations. Y.S. and C.L. coordinated the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0319-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Plant materials. Three diploid potato clones (PG6226, PG6235, and PG6359) were used to generate the selfing populations in this study. PG6226 (also named E) and PG6235 (also named RH) were kindly provided by the Department of Plant Breeding at Wageningen University, and PG6359 (also named C151) was kindly provided by the International Potato Center. PG6235 and PG6359 are natural self-compatible mutants, and the selfing population of PG6226 was generated by forced self-pollination at the young bud stage. In total, 104, 223, and 201 progenies from PG6226, PG6235, and PG6359, respectively, were subjected to re-sequencing. All of the sequenced individuals grew normally on Murashige and Skoog medium.

The 148 diploid landraces were kindly provided by the International Potato Center. The F_2 population used for fine mapping of *ar1* was derived from a cross between PG6226 and a wild potato clone, *S. chacoense* 34-28 (kindly provided by K. Hosaka of Obihiro University). Since the male parent 34-28 contains the *S-locus inhibitor* gene (*Sli*)[24], the self-compatible F_1 clone was used to generate the F_2 mapping population.

Identification of the *S-RNase* alleles. De novo assembly of RNA-sequencing data from styles was performed to identify the *S-RNase* alleles in the heterozygous diploid clone PG6359 (ref. [25]).

Genome sequencing, sequence alignment, and SNP calling. For whole-genome shotgun sequencing, total genomic DNA was isolated from fresh leaves using the cetyl trimethyl ammonium bromide method. Whole-genome re-sequencing was performed using the HiSeq X Ten platform. The insert size of the libraries was 250–300 base pairs (bp), and the read length was 150 bp. Short reads were aligned against the potato reference genome version 4.03 (ref. [26]) using Burrows–Wheeler Aligner (http://bio-bwa.sourceforge.net), and heterozygous SNPs in three parental clones were extracted using SAMtools (http://samtools.sourceforge.net) according to following criteria: (1) the sequencing quality of the nucleotide was ≥40; (2) the mapping quality of reads was ≥30; (3) the depth of the SNPs was ≥5 and ≤50; and (4) the index of SNPs was ≥0.3 and ≤0.7.

Prediction of deleterious mutations. Three parental clones (PG6226, PG6235, and PG6359), 20 *S. candolleanum* clones, and 148 diploid landraces (Supplementary Table 1) were used to predict the deleterious mutations. For each clone, ~10× genome sequences were used to call SNPs. The index of heterozygous SNPs was ≥0.3 and ≤0.7, and that of homozygous SNPs was ≥0.9 or ≤0.1. Amino acid substitutions and their effects on protein function were predicted with the SIFT algorithm[11]. The amino acid substitution is predicted to be deleterious if the score is ≤0.05 and tolerated if the score is >0.05. The ratio of overlapping deleterious mutations between any two clones was calculated using the following formula:

$$\text{ratio} = \frac{a + 0.5b + 0.25c}{d}$$

where *a*, *b*, and *c* indicate the number of deleterious mutations that are homozygous in both clones, homozygous in one clone and heterozygous in the other clone, and heterozygous in both clones, respectively; *d* indicates the total number of deleterious mutations in two clones.

Deduction of two haplotypes among the F_1 clones. The heterozygosity in the F_2 progeny was calculated based on the ratio between the number of heterozygous SNPs and the number of all SNPs using sliding window analysis (window size = 1 Mb, step = 100 kb). A genomic region with a ratio of <0.1 was considered to be homozygous. For each chromosome, the similarity between different homozygous regions was calculated following the formula:

$$\text{similarity} = \frac{\text{number of same SNPs} - \text{number of different SNPs}}{\text{number of overlapping SNPs}}$$

A similarity score of <0.01 indicated that the query region belonged to a different group from the reference, whereas a similarity score of >0.99 indicated that the query region belonged to the same group as the reference. All regions with a similarity score of 0.01–0.99 were discarded. Within each group, all SNPs with a consensus at each locus were combined to construct the haplotype. The phase with more SNPs was defined as haplotype B, and the other phase was defined as haplotype A.

Determining the breakpoints of crossovers and bin map construction. In a single F_2 individual, the genotype of each SNP was defined as 'a' if the SNP pattern in all reads was the same as in haplotype A, 'b' if the SNP pattern in all reads was the same as in haplotype B, and 'h' if the SNP was heterozygous. To estimate the breakpoint, sliding window analysis (window size = 200 SNPs, step = 20 SNPs) was performed for each chromosome by calculating the weighted value of genotype b (wv-b) as follows:

$$\text{wv} - \text{b} = \frac{\text{number of b} - \text{number of a}}{200}$$

If wv-b was ≤0.1 or ≥0.9, the window was defined as 'a' or 'b', respectively. If wv-b was between 0.1 and 0.9, the window was defined as 'h'. The transitions between different genotypes were defined as breakpoints. For each F_2 progeny, breakpoints <100 kb away from the flanking breakpoints were discarded.

For each chromosome, the breakpoints from all progeny were ordered. The genotype of the bin, the interval between two neighboring breakpoints, was estimated using the following formula:

$$\text{vb} = \frac{\text{number of b}}{\text{number of all SNPs}}$$

Where vb indicates the ratio of SNPs with genotype of phase B. Genotypes with a vb of ≥0.8 were designated as 'b'; those with a vb of ≤0.2 were designated as 'a'; and those with a vb of >0.2 and <0.8 were designated as 'h'. Redundant bins were filtered according to the following standards: (1) if two continuous bins were ≤50 kb, they were merged; (2) if the number of SNPs in a bin was ≤50, it was merged with the next bin; (3) if the genotype of a bin was the same as that of the next bin, it was merged with the next bin. A genetic map of bins for each population was constructed using Joinmap4 (ref. [27]).

Statistical tests. The significance was determined by two-tailed Student's *t* tests. Segregation distortions were analyzed with a χ^2 test for goodness of fit to the appropriate expected segregation ratio at a significance level of $P = 0.001$. The expected segregation ratio of zygotes is 1:2:1, and that of gametes is 1:1.

Bulked segregant analysis of inbreeding depression-related traits. The apparent defects in survival and growth vigor, such as abnormal rooting, white seedlings, yellow margin leaf, yellow leaf, and plant architecture, were observed with the naked eye. For these traits, we investigated the phenotype of F_2 individuals and used bulked segregant analyses to map the corresponding loci[28]. The positions of these traits were further localized based on the phenotype and the bin maps of sequenced F_2 individuals.

Since some progeny of PG6226 and PG6235 could not flower, to test the pollen viability, we grew 312 (containing 104 sequenced individuals) and 928 (containing 223 sequenced individuals) F_2 clones of PG6226 and PG6235, respectively. Only sequenced F_2 clones of PG6359 were used to test pollen viability. For each clone, three plants were grown. The pollen was dyed with triphenyl tetrazolium chloride, and the viability was graded as follows: 0, not viable; 1, 0–20% of pollen grains are viable; 2, 21–40% of pollen grains are viable; 3, 41–60% of pollen grains are viable; 4, 61–80% of pollen grains are viable; and 5, ≥80% of pollen grains are viable. In addition, the average tuber weight per plant of the selfing population of PG6226 was measured. Bulked segregant analyses were used to map the loci affecting fertility and tuber weight[28].

Fine mapping of *ar1*. To fine-map *ar1*, more than 30,000 F_2 seeds of the F_1 clone E-172 (derived from a cross between PG6226 and *S. chacoense* 34-28) were sown on culture medium, and only individuals showing abnormal rooting were subjected to genotyping. To develop polymorphic markers, the F_1 clone E-172 was sequenced, heterozygous indels were extracted, and primers were designed using Primer 3 (ref. [29]). The primers used for fine mapping are shown in Supplementary Table 5.

Transformation of tomato. For tomato transformation, the single guide RNA (sgRNA; 5′-TGTGGAGTCAGAGGTGGAA-3′) of the candidate gene was ligated to pKSE401. Heterozygous mutants of the T_0 generations were selfed to generate T_1 seeds. The T_1 seeds were sown on Murashige and Skoog medium, and roots of the resulting plants were observed.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability

The pipeline of parent-independent genotyping in potato was written using custom Python scripts. All codes are available from the corresponding author upon request.

Data availability

The sequencing data that support the findings of this study have been deposited in the Sequence Read Archive (SRA) under accession PRJNA471783. The deleterious mutations datasets are available from the following ftp link: ftp://ftp.agis.org.cn/~zhangchunzhi/.

References

24. Hosaka, K. & Hanneman, R. E. Jr. Genetics of self-compatibility in a self-incompatible wild diploid potato species *Solanum chacoense*. 2. Localization of an *S locus inhibitor* (*Sli*) gene on the potato genome using DNA markers. *Euphytica* **103**, 265–271 (1998).
25. Ye, M. et al. Generation of self-compatible diploid potato by knockout of *S-RNase*. *Nat. Plants* **4**, 651–654 (2018).

26. Sharma, S. K. et al. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *Genes Genomes Genet.* **3**, 2031–2047 (2013).
27. Van Ooijen, J. W. Joinmap4: Software for the Calculation of Genetic Linkage Maps in Experimental Populations (Kyazma BV, Wageningen, 2006).
28. Takagi, H. et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* **74**, 174–183 (2013).
29. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primers. *Bioinformatics* **23**, 1289–1291 (2007).

nature research

Corresponding author(s): Sanwen Huang

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect data.
Data analysis	We used BWA and Samtools for SNP calling, SIFT for prediction of deleterious substitutions, Joinmap4 for construction of linkage groups. All of them were mentioned in the Methods. The pipeline of parent-independent genotyping in potato were written using custom python scripts, and the codes are available from the corresponding author upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data that support the findings of this study have been deposited in NCBI Sequence Read Archive (SRA) under accession number PRJNA471783. The deleterious mutations datasets are available from the following ftp link: ftp://ftp.agis.org.cn/~zhangchunzhi/.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We selected over 100 selfed progeny from each population to calculate the segregation distortion. We analyzed the segregation distortion using 100 and 200 progeny, and the results were same. Thus we thought the number of selfed progeny we used in this study is enough.
Data exclusions	No data was excluded.
Replication	There are two main findings in this study. The first is the deleterious mutations in diploid potato are enriched in the pericentromeric regions and line-specific. To confirm this point, we analyzed 151 diploid accessions, and obtained the similar results. The second finding is the large-effect deleterious mutations are usually located in the regions with high recombination rate. In total, we identified 15 large-effect deleterious mutation, and 14 of them are in accordance with the pattern. The results demonstrated our findings have good repeatability.
Randomization	The allocation of samples was random.
Blinding	The investigators were blinded to the group allocations.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All the plant materials are available from the corresponding author upon request.