Regenerant Arabidopsis Lineages Display a Distinct Genome-Wide Spectrum of Mutations Conferring Variant Phenotypes

Caiifu Jiang,1 Aziz Mithani,1 Xiangchao Gan,1,2 Eric J. Belfield,1 John P. Klingler,2 Jian-Kang Zhu,2 Jiannis Ragoussis,3 Richard Mott,3 and Nicholas P. Harberd1,*

1Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK
2Plant Stress Genomics Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
3The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

Summary

Multicellular organisms can be regenerated from totipotent differentiated somatic cell or nuclear founders [1–3]. Organisms regenerated from clonally related isogenic founders might have priori have been expected to be phenotypically invariant. However, clonal regenerant animals display variant phenotypes caused by defective epigenetic reprogramming of gene expression [2], and clonal regenerant plants exhibit poorly understood heritable phenotypic (“somaclonal”) variation [4–7]. Here we show that somaclonal variation in regenerant Arabidopsis lineages is associated with genome-wide elevation in DNA sequence mutation rate. We also show that regenerant mutations comprise a distinctive molecular spectrum of base substitutions, insertions, and deletions that probably results from decreased DNA repair fidelity. Finally, we show that while regenerant base substitutions are a likely major genetic cause of the somaclonal variation of regenerant Arabidopsis lineages, transposon movement is unlikely to contribute substantially to that variation. We conclude that the phenotypic variation of regenerant plants, unlike that of regenerant animals, is substantially due to DNA sequence mutation.

Results

Regenerant Arabidopsis Lineages Display Somaclonal Variation

Arabidopsis is a genetic model plant with a condensed genome (~120 Mb). We reasoned that analysis of regenerant Arabidopsis plant genomes at single-base resolution might reveal genetic changes conferring somaclonal variation. In vitro regeneration of Arabidopsis is achieved via two-stage culture of root explants [8]. First, a pluripotent cell mass (callus) forms via activation of a lateral root genetic developmental program in pericycle cells of explants grown on auxin-rich medium [9, 10]. Subsequently, shoots or roots develop from callus grown on media containing specific auxin/cytokinin concentration ratios. The de novo induction of shoot meristems from small clusters of progenitor callus cells [11] leads eventually to the regeneration of an entire plant.

We first determined whether regenerant Arabidopsis lineages display heritable phenotypic somaclonal variation comparable with that seen in other species. Twenty-eight (R0) plants were regenerated from explants from a single Arabidopsis (Columbia laboratory strain; Col-0) root [8] (Figure 1A), and phenotypic variation was assessed in resultant self-pollination generated (R1) families (Figures 1A–1C; see also Figure S1 available online). Variant phenotypes were detected in 8 of the 28 lineages (Figures 1B and 1C; see also Table S1). Two of these phenotypes were not stably heritable (Figure 1B; see also Figure S1 and Table S1) and probably were conferred by unstable epigenetic change. In six further cases, phenotypes were stably heritable and segregated within R1 families, with segregation ratios approximating to Mendelian expectations for single-gene recessive mutations that had been heterozygous in the preceding R0 plant (Figure 1B; see also Table S1; data not shown). Thus in vitro regeneration of Arabidopsis plants results in a high frequency of heritable phenotypic variation and provides a general model for the study of plant somaclonal variation.

Regenerant Arabidopsis Lineages Display a Characteristic Spectrum of Genome-wide DNA Sequence Mutation

We reasoned that the somaclonal variation exhibited by regenerant Arabidopsis lineages might be associated with an increase in genomic DNA sequence mutation. We therefore next determined the genome-wide extent and molecular spectrum of DNA sequence mutations in regenerant Arabidopsis lineages, using 76 base pair paired-end Illumina “next-generation” DNA sequencing (Figures 2A–2E). DNA samples from the progenitor P1 plant (Figure 1A; progenitor genome) and from five individual R1 plants (Figure 1A; regenerant genomes) were sequenced to a coverage depth of between ~22x and ~30x per sample (Figure S2A). For each sample, high-quality uniquely mapped reads covered ~116 million of the ~120 million base pairs of the Col-0 TAIR9 reference genome (see Experimental Procedures), and most uncovered regions were located either in centromeres or telomeres (Figures S2A and S2B; Experimental Procedures). We detected 152 novel regenerant single-base substitution (SBSs) and short insertion or deletion (indel) mutations in R1 plants (mutations not present in the P1 progenitor) using a scheme summarized in Figure S2C. In addition, because previous work showed that somaclonal variation can be associated with larger-scale indels and chromosomal abnormalities (e.g., [12]), we exploited the “paired-end” property of Illumina Genome Analyzer data to detect such larger-scale events in R1 plants (see Experimental Procedures; Figures S2E–S2K). Despite exhaustive searches, no large indels or gross chromosomal abnormalities were detected in any of the five R1 plants representing regenerant Arabidopsis lineages.

The 152 regenerant DNA sequence mutations comprised 131 SBSs and 21 small indels (<2 bp) (“Detected mutations” in Figures 2A and 2B; Table S2A). Sample sets of detected SBSs and indels were confirmed by capillary sequencing (see Experimental Procedures). Regenerant mutations were apparently evenly spread between chromosomes (Figure 2A; see also Figure S2L). These de novo mutations would have been heterozygous when they first arose and hence had had a 25% chance of being homozygous in individual R1 plants of the subsequent (self-pollination generated)
generation. Because we were detecting homozygous mutations only (see Experimental Procedures and Figure S2C), and because we expected that only ~25% of regenerant mutations would have become homozygous in the R1 generation, we accordingly estimated the number of regenerant mutations by multiplying the detected mutations by four (“Theoretical mutations” in Figure 2B). We hence derive somaclonal mutation frequencies of between 4.2 × 10⁻⁷ and 24.2 × 10⁻⁷ mutations per site in the five regenerant lineages (Figure 2B; see Experimental Procedures). The mutation rate thus increased between 60× and 350× in the regenerant lineages versus the background “spontaneous” mutation rate observed in sexually propagated Arabidopsis (~7 × 10⁻⁹ mutations per site per generation [13]).

SBSs are the most common category of regenerant mutation (Figures 2B and 2C). However, the ratio of regenerant transitions:transversions (63:68 = 0.92) is very different to that (2.41) seen in sexually propagated plants [13], as are the relative frequencies (molecular spectrum) of individual base substitution classes (Figure 2D; see also Figures S2M–S2P). In addition, regenerant plants carried an elevated frequency of indel mutations (Figure 2E). Intriguingly, nearly all regenerant indels occurred in homopolymeric or poly dinucleotide stretches, which is not the case for small indels arising spontaneously in sexually propagated plants [13] (Figure 2E; Figures S2Q and S2R; see also Table S2B). Short read “next-generation” sequencing is relatively poor at reporting indels in simple sequence repeat regions (for further discussion see [13]), which might explain why the small indel mutation rates we determined are less than those previously determined by a different method for sexually propagated plants [14]. Nevertheless, our results show an increased frequency of small indels in regenerant plants at those sites we were able to assay (Table S2B; Figures S2Q and S2R). In summary, the overall distinctiveness of the regenerant mutational molecular spectrum (SBSs and indels) implies that the elevation in mutation frequency observed in regenerant plants cannot simply be attributed to an accelerated accumulation of the same kinds of mutations as arise spontaneously in sexually propagated plants.

be a major genetic cause of somaclonal variation. Indeed, one of the five genome-wide sequenced R1 plants, R1-19-2, carries a protein-truncating SBS in Cullin 3A (a locus at which mutant alleles confer late flowering [15]), and a late-flowering phenotype was indeed conferred by this SBS (Figures 3A and 3B). Further genetic complementation and molecular analysis of additional regenerant lines displaying long hypocotyl (R2-17-1; Figures 3C and 3D; Figure S2A) and late flowering (R2-6-3; Figures 3E and 3F; Figure S3B) phenotypes established that they respectively carried novel mutant HY1 (mutant alleles confer an elongated hypocotyl [16]) and FKF1 (mutant alleles confer late flowering [17]) alleles. Capillary sequencing of these mutant alleles identified nonsynonymous SBS mutations in the protein-encoding sequences of HY1 (Figure 3D) and FKF1 (Figure 3F). While a previous report identifies base substitution mutation as a source of regeneration-associated phenotypic change [18], our findings demonstrate that such mutations are actually a major contributor to somaclonal variation.

Transposed Mobile Genetic Elements Not Detected in Sampled Regenerant Arabidopsis Lineages

The Arabidopsis genome contains multiple mobile genetic elements (transposons), whose activity is normally held in check by DNA methylation and associated epigenetic regulation [19, 20]. Because increased transposon activity has been linked to DNA sequence mutations in rice (Oryza sativa) tissue culture [21], we next sought to detect transposed transposons in regenerant Arabidopsis lineages, with depth of read coverage as a measure of the increased copy number caused by transposon amplification (see Experimental Procedures). Comparative analysis of the read coverage representation of CACTA, COPIA, gypsy, hAT, non-LTR, and other transposon classes (representative of 3,321 A. thaliana transposons, retro-transposons, and other putative mobile elements; Table S3) in whole-genome sequence data from the five R1 regenerant plants and from an epigenetically compromised positive control met1/ nrpd2 mutant line (versus P1 data) revealed only the previously described novel AtCOPIA99 retrotransposition in met1/ nrpd2 [20] and no detectable novel transposon

Figure 1. Regeneration-Induced “Somaclonal” Variation in Arabidopsis

(A) A root explant from a single parental (P1) plant was the source (see [8]) of all R0 regenerant and subsequent generation plants (also see Table S1).

(B) Frequency (number and percentage) of R1 families segregating phenotypic variants (also see Figure S1).

(C) Selected phenotypic variant R1 plants (segregating variant plants or organs highlighted with red arrows). Scale bars represent 0.5 cm.

Regenerant Base Substitution Mutations Confer Somaclonal Variant Phenotypes

Of 29 regenerant SBSs affecting protein-coding sequence, 17 were nonsynonymous mutations that alter the amino acid sequence of proteins (Figure 2C; Table S2A). In contrast, none of the indels affected protein-coding sequence (Figure 2C), suggesting that SBSs may be a major genetic cause of somaclonal variation. Indeed, one of the five genome-wide sequenced R1 plants, R1-19-2, carries a protein-truncating SBS in Cullin 3A (a locus at which mutant alleles confer late flowering [15]), and a late-flowering phenotype was indeed conferred by this SBS (Figures 3A and 3B). Further genetic complementation and molecular analysis of additional regenerant lines displaying long hypocotyl (R2-17-1; Figures 3C and 3D; Figure S2A) and late flowering (R2-6-3; Figures 3E and 3F; Figure S3B) phenotypes established that they respectively carried novel mutant HY1 (mutant alleles confer an elongated hypocotyl [16]) and FKF1 (mutant alleles confer late flowering [17]) alleles. Capillary sequencing of these mutant alleles identified nonsynonymous SBS mutations in the protein-encoding sequences of HY1 (Figure 3D) and FKF1 (Figure 3F). While a previous report identifies base substitution mutation as a source of regeneration-associated phenotypic change [18], our findings demonstrate that such mutations are actually a major contributor to somaclonal variation.

Transposed Mobile Genetic Elements Not Detected in Sampled Regenerant Arabidopsis Lineages

The Arabidopsis genome contains multiple mobile genetic elements (transposons), whose activity is normally held in check by DNA methylation and associated epigenetic regulation [19, 20]. Because increased transposon activity has been linked to DNA sequence mutations in rice (Oryza sativa) tissue culture [21], we next sought to detect transposed transposons in regenerant Arabidopsis lineages, with depth of read coverage as a measure of the increased copy number caused by transposon amplification (see Experimental Procedures). Comparative analysis of the read coverage representation of CACTA, COPIA, gypsy, hAT, non-LTR, and other transposon classes (representative of 3,321 A. thaliana transposons, retro-transposons, and other putative mobile elements; Table S3) in whole-genome sequence data from the five R1 regenerant plants and from an epigenetically compromised positive control met1/ nrpd2 mutant line (versus P1 data) revealed only the previously described novel AtCOPIA99 retrotransposition in met1/ nrpd2 [20] and no detectable novel transposon
amplification in any of the regenerant lines (Figure 4; Figure S4). A second strategy for detection of novel transposon insertions in regenerant lines also detected nothing (see Experimental Procedures). We conclude that insertion of transposons into genes is unlikely to contribute significantly to the genetic variation underlying somaclonal phenotypic variation in regenerant Arabidopsis plants.

Discussion

Here we have shown that regenerant Arabidopsis plant lineages display extensive phenotypic somaclonal variation (Figures 1A–1C) and an elevated frequency and characteristic distribution of DNA sequence molecular mutation classes (Figures 2A–2E). We have also shown that the increased base substitution frequency in regenerant plants substantially explains somaclonal phenotypic variance (Figures 3A–3F). Although previous reports have described relatively high frequencies of gross chromosomal abnormality in somaclonal variant lines [12] and transposon movement during in vitro cell culture and/or in regenerant plants (e.g., [21, 22]), we detected neither in our regenerant Arabidopsis lineages (Figure 4). It is possible that species differences in genome architecture or the relatively short callus phase duration (~1 week) in our experiments can explain this apparent discrepancy. We found additional phenotypic variation in regenerant Arabidopsis lineages that is probably attributable to variable outcomes of epigenetic reprogramming (Figure 1B; Figure S1), indicating...
Thus, although DNA sequence change explains much of nuclear transfer does not detectably increase mutation rate sequence mutation, regeneration of animals via somatic cell variation is substantially due to an increased rate of DNA the underlying causes of this variability may differ between causes phenotypic variability, our observations suggest that they frequently adopt life cycle strategies that involve regeneration from somatic tissues.

Finally, we have highlighted increased frequencies of both base substitutions and small indels in regenerant plants. Regenerant indels mostly occur in homopolymeric or polydi-nucleotide regions and are probably DNA replication slippage mutations. Intriguingly, combined indels and base substitutions are characteristic both of the bacterial SOS response and of potential consequences for the evolution of plants, given that plant mutations distinguishes them from those arising spontaneously in sexually propagated plants [13]. There are two possible nonexclusive explanations for this observation. First, callus phase growth and/or in vitro regeneration from tissue culture might be inherently mutagenic. This first explanation may be supported by previous observations that somaclonal variant phenotype frequencies increase in proportion to the duration during which cells are maintained in tissue culture [26]. Second, mutations in regenerant plants might reflect somatic mutations that existed in the cells of the initial root explant prior to in vitro regeneration, with these somatic mutations comprising a molecular mutational spectrum differing from that of spontaneous “germline” mutations. This second explanation may be supported by the observation that somatic mutation rates are characteristically higher than germline rates in multicellular organisms [26] and has important particular potential consequences for the evolution of plants, given that they frequently adopt life cycle strategies that involve regeneration from somatic tissues.

We here show that the molecular spectrum of regenerant plant mutations distinguishes them from those arising spontaneously in sexually propagated plants [13]. There are two possible nonexclusive explanations for this observation. First, callus phase growth and/or in vitro regeneration from tissue culture might be inherently mutagenic. This first explanation may be supported by previous observations that somaclonal variant phenotype frequencies increase in proportion to the duration during which cells are maintained in tissue culture [26]. Second, mutations in regenerant plants might reflect somatic mutations that existed in the cells of the initial root explant prior to in vitro regeneration, with these somatic mutations comprising a molecular mutational spectrum differing from that of spontaneous “germline” mutations. This second explanation may be supported by the observation that somatic mutation rates are characteristically higher than germline rates in multicellular organisms [26] and has important particular potential consequences for the evolution of plants, given that they frequently adopt life cycle strategies that involve regeneration from somatic tissues.

Finally, we have highlighted increased frequencies of both base substitutions and small indels in regenerant plants. Regenerant indels mostly occur in homopolymeric or polydi-nucleotide regions and are probably DNA replication slippage mutations. Intriguingly, combined indels and base substitutions are characteristic both of the bacterial SOS response and of potential consequences for the evolution of plants, given that plant mutations distinguishes them from those arising spontaneously in sexually propagated plants [13]. There are two possible nonexclusive explanations for this observation. First, callus phase growth and/or in vitro regeneration from tissue culture might be inherently mutagenic. This first explanation may be supported by previous observations that somaclonal variant phenotype frequencies increase in proportion to the duration during which cells are maintained in tissue culture [26]. Second, mutations in regenerant plants might reflect somatic mutations that existed in the cells of the initial root explant prior to in vitro regeneration, with these somatic mutations comprising a molecular mutational spectrum differing from that of spontaneous “germline” mutations. This second explanation may be supported by the observation that somatic mutation rates are characteristically higher than germline rates in multicellular organisms [26] and has important particular potential consequences for the evolution of plants, given that they frequently adopt life cycle strategies that involve regeneration from somatic tissues.

We here show that the molecular spectrum of regenerant plant mutations distinguishes them from those arising spontaneously in sexually propagated plants [13]. There are two possible nonexclusive explanations for this observation. First, callus phase growth and/or in vitro regeneration from tissue culture might be inherently mutagenic. This first explanation may be supported by previous observations that somaclonal variant phenotype frequencies increase in proportion to the duration during which cells are maintained in tissue culture [26]. Second, mutations in regenerant plants might reflect somatic mutations that existed in the cells of the initial root explant prior to in vitro regeneration, with these somatic mutations comprising a molecular mutational spectrum differing from that of spontaneous “germline” mutations. This second explanation may be supported by the observation that somatic mutation rates are characteristically higher than germline rates in multicellular organisms [26] and has important particular potential consequences for the evolution of plants, given that they frequently adopt life cycle strategies that involve regeneration from somatic tissues.

We here show that the molecular spectrum of regenerant plant mutations distinguishes them from those arising spontaneously in sexually propagated plants [13]. There are two possible nonexclusive explanations for this observation. First, callus phase growth and/or in vitro regeneration from tissue culture might be inherently mutagenic. This first explanation may be supported by previous observations that somaclonal variant phenotype frequencies increase in proportion to the duration during which cells are maintained in tissue culture [26]. Second, mutations in regenerant plants might reflect somatic mutations that existed in the cells of the initial root explant prior to in vitro regeneration, with these somatic mutations comprising a molecular mutational spectrum differing from that of spontaneous “germline” mutations. This second explanation may be supported by the observation that somatic mutation rates are characteristically higher than germline rates in multicellular organisms [26] and has important particular potential consequences for the evolution of plants, given that they frequently adopt life cycle strategies that involve regeneration from somatic tissues.

Experimental Procedures

Arabidopsis Regeneration and Detection of Phenotypic Variance in Regenerant Lineages

Regeneration of R0 plants from root explants from a single (P1) Columbia (Col-0) laboratory strain plant was as described [8]. 28 R0 plants were self-pollinated to generate R1 families (Figure 1A). From each R1 family, 25–30 R1 plants were phenotypically characterized. Phenotypic stability was further monitored in R2 (self-pollination) descendents.

DNA Sequencing and Sequence Alignment

Genomic DNA from single individuals was sequenced (Illumina Genome Analyzer II technology) using libraries created from ~350 bp fragments and a single 76 bp paired-end run lane for each sample. Reads were aligned to the TAIR9 reference genome sequence (http://www.arabidopsis.org) using Burrows-Wheeler Aligner (BWA; http://bio-bwa.sourceforge.net/) (Figure S2C).

Detection of Base Substitutions, Short Insertions, and Deletions

After filtering out all nonuniquely mapped reads and reads with a Phred-scaled mapping quality of less than 20, individual lists of SBSs, short insertions, and short deletions in P1 and the five R1 samples (versus TAIR9) were detected using SAMTools v0.1.5c [30]. Many SBSs and indels were detected in P1 versus TAIR9 (see Figure S2C). As previously reported [31], most of these apparent P1 variants probably represent errors in TAIR9. We detected novel regenerant mutations (mutations in R1 samples and not in P1) using methods summarized in Figure S2C.

Detection of Large Indels, Chromosomal Inversions, and Translocations

We developed novel codes to detect larger-scale variants (see also [32–35]). Such variants (insertions, inversions, and translocations) will generate “distant-pair” read pairs (where two paired reads [“mates”] align with unexpectedly distant regions of TAIR9 [Figures S2E–S2H]). We therefore extracted reads for which mates mapped >750 bp distant with respect to TAIR9, thus creating new (distant-pair).bam files. Lists of covered regions (depth of coverage of at least five reads) were generated. Visual analysis...
Two strategies were used to search for transposed transposons in regenerant plants. First, variation in read coverage detected transposition-dependent amplification of transposon sequence (see Figure S4). For results shown in Figure 4, the average coverage for each of 3,321 transposable element genes (Table S3) was calculated by summing the coverage at all transposon sequences for which the Log2 coverage ratio score was expected to result in a log2 score value of 1. IGV scanning of all dots with value >0.75 (see Experimental Procedures) indicated the absence of detectable transposition during plant regeneration. See Figure S4 and Table S3.

Supplemental Information
Supplemental Information includes four figures and three tables and can be found with this article online at doi:10.1016/j.cub.2011.07.002.

Acknowledgments
We thank Olivier Mathieu, Jerzy Paszkowski, and Detlef Weigel for provision of met1+/nrpd2 sequence data. David Buck and colleagues in the Genomic Services Group at the Wellcome Trust Centre for Human Genetics (Oxford, UK) performed much of the genomic sequencing, supported by Wellcome Trust Core grant 090532/Z/09/Z. Additional genomic sequencing was performed by the Beijing Genomics Institute (China). We thank the IGV group at the Broad Institute (Cambridge, MA) for help with IGV implementation. This publication is based on work supported by Award No. KUK-11-002-03, made by King Abdullah University of Science and Technology and by Biotechnology and Biological Sciences Research Council (BBSRC) grants BB/F020759/1 and BB/F022697/1.

Received: March 10, 2011
Revised: May 27, 2011
Accepted: July 1, 2011
Published online: July 28, 2011

References