Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*

Eric J. Belfield, Xiangchao Gan, Aziz Mithani, et al.

*Genome Res.* 2012 22: 1306-1315 originally published online April 12, 2012
Access the most recent version at doi:10.1101/gr.131474.111

---

**Supplemental Material**

http://genome.cshlp.org/content/suppl/2012/04/06/gr.131474.111.DC1.html

**References**

This article cites 37 articles, 17 of which can be accessed free at:
http://genome.cshlp.org/content/22/7/1306.full.html#ref-list-1

**Creative Commons License**

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

**Email Alerting Service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

---

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

© 2012, Published by Cold Spring Harbor Laboratory Press
Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of Arabidopsis thaliana

Eric J. Belfield, Xiangchao Gan, Aziz Mithani, Carly Brown, Caifu Jiang, Keara Franklin, Elizabeth Alvey, Anjar Wibowo, Marko Jung, Kit Bailey, Sharan Kalwani, Jiannis Ragoussis, Richard Mott, and Nicholas P. Harberd

Ionizing radiation has long been known to induce heritable mutagenic change in DNA sequence. However, the genome-wide effect of radiation is not well understood. Here we report the molecular properties and frequency of mutations in phenotypically selected mutant lines isolated following exposure of the genetic model flowering plant Arabidopsis thaliana to fast neutrons (FNs). Previous studies suggested that FNs predominantly induce deletions longer than a kilobase in A. thaliana. However, we found a higher frequency of single base substitution than deletion mutations. While the overall frequency and molecular spectrum of fast-neutron (FN)-induced single base substitutions differed substantially from those of “background” mutations arising spontaneously in laboratory-grown plants, G:C→A:T transitions were favored in both. We found that FN-induced G:C→A:T transitions were concentrated at pyrimidine dinucleotide sites, suggesting that FNs promote the formation of mutational covalent linkages between adjacent pyrimidine residues. In addition, we found that FNs induced more single base than large deletions, and that these single base deletions were possibly caused by replication slippage. Our observations provide an initial picture of the genome-wide molecular profile of mutations induced in A. thaliana by FN irradiation and are particularly informative of the nature and extent of genome-wide mutation in lines selected on the basis of mutant phenotypes from FN-mutagenized A. thaliana populations.

[Supplemental material is available for this article.]

 Ionizing radiation is pervasive in the environment and acts as a natural mutagen via its DNA-damaging properties (Friedberg et al. 2006). In addition, the mutagenic property of artificial ionizing radiation has been a mainstay of genetic research since the pioneering experiments of Müller (1928). However, while the effects of ionizing radiation on individual genes are now relatively well understood, its genome-wide effects are not. We therefore undertook an analysis of the genome-wide consequences of fast neutron (FN) irradiation in Arabidopsis thaliana, comparing our findings with those of recent studies documenting the frequency and molecular spectrum of spontaneous “background” mutations in the genomes of laboratory-grown “mutation accumulation” (MA) line Arabidopsis plants (Ossowski et al. 2010) and of Arabidopsis plants regenerated in vitro from root tissue explants (Jiang et al. 2011). We found that exposure of Arabidopsis to FNs induces a broader range of mutational lesions than previously suspected, and that both the incidence and spectrum of FN-induced mutations are distinct from those of “spontaneous” mutations. These discoveries have important consequences for the use of FNs in experimental plant mutagenesis and provide an indication of the likely mutagenic effects of environmental ionizing radiation on organisms living in the wild (Hinton et al. 2007).

Results

FN irradiation, mutant generation, and isolation

Our analyses began with a multiply mutant Arabidopsis line [predominantly Landsberg erecta (Ler) background carrying the mutations gai-t6, rga-t2, rgl1-1, rgl2-1, and rgl3-4; henceforth called progenitor] (see Methods) that lacks the growth-regulating DELLAs. Progenitor (M0) seeds were exposed to FN irradiation (see Methods), with the long-term aim (although not the focus of this study) of isolating and characterizing novel growth-regulation mutants in a DELLAs-deficient genetic background. Accordingly, M2 seedlings were screened for elongated hypocotyls, a mutant phenotype in which exaggerated growth of the hypocotyl (embryonic stem structure) results in a longer hypocotyl than that of progenitor controls. This class of mutant commonly contains loss-of-function mutations in genes that...
mediate the light inhibition of hypocotyl growth (Chen et al. 2004). We identified more than 300 independent putative elongated hypocotyl mutants and selected six of these (E71, E99, E125, E128, E138, and E216) to confirm heritability and for analysis in subsequent generations.

Creation of a progenitor reference sequence and alignment of progenitor reads to Col-0 reference

To identify mutations in the genomes of the six selected elongated hypocotyl mutants, we first needed to obtain a whole-genome reference sequence of the progenitor from which they were derived. We therefore obtained Illumina Genome Analyzer DNA sequence data from a single progenitor line plant and used these data to create a progenitor reference genome (see Methods). Initial mapping of the progenitor paired-end sequencing reads to the Arabidopsis Information Resource (TAIR 9) (http://www.arabidopsis.org) Col-0 reference genome (using MAQ v0.7.1) (Li et al. 2008) revealed single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs). This observation is consistent with the fact that only 3.2 Mb of the ~120-Mb progenitor genome is derived from Col-0 (Fig. 1), while the remaining ~97% of the genome is of Ler origin. To eliminate the progenitor versus TAIR 9 variants observed and to aid the identification of the genetic lesions caused by FN mutagenesis, we created our own progenitor reference genome sequence (see Methods).

Detection of mutations in FN-induced mutant lineages

We next generated Illumina Genome Analyzer DNA sequence data from single M1 plants homozygous for the phenotype-conferring mutations in the six selected mutant lineages. We were able to identify a near-complete set of homozygous DNA sequence variants (versus the progenitor reference) in the nonrepetitive fractions of these mutant genomes. A mutation (base substitution or INDEL) was called if 95%–100% of aligned reads in a sample differed from the progenitor reference sequence (see Methods). We detected a total of 108 homozygous mutations (see Fig. 2; Table 1; Supplemental Table 1). Of the 56 variants detected in lines E99 and E125, all but one was independently confirmed by direct PCR amplification and Sanger sequencing, indicating that our false-positive discovery rate was negligible. The unconfirmed variant in E125 was a G:C—to-A:T substitution that was difficult to confirm because it lies in a region of sequence that is represented repetitively in the Arabidopsis genome.

Estimation of original numbers of mutations in FN-irradiated plants

While the above 108 homozygous mutations were detected in M1 plant genomes, they had originally been induced by FN irradiation as heterozygous mutations in M1 plant genomes (Fig. 3). We therefore calculated the original mutation numbers in the following way. Assuming recessivity of the initial elongated hypocotyl mutation, each original M2 mutant seedling would have been homozygous for the mutation causing that phenotype (and any neighboring mutations) (see Fig. 3). In addition, there was an approximately one-fourth chance of the original M2 mutant seedling being homozygous for induced mutations (of unknown phenotypic consequence) in other regions of the genome. These regions of “fixed” homozygosity would have been uniformly transmitted to the subsequent M3 generation. Further mutations, heterozygous in the original M2 seedling, will have then become homozygous in M3 seedlings (the genome of one of which was sequenced). Thus, of the total 108 detected homozygous M2 mutations (Supplemental Table 1), 72 were likely homozygous in the M2, with the remaining quarter (36) resulting from an estimated 144 variants that had been heterozygous in the M2 and had then become homozygous in the M3 (see Fig. 3). Mutations heterozygous in the M2 generation were overwhelmingly likely to have been discarded from variant lists (and thus not detected) because of failure to reach the 95%–100% read difference criterion described above. In summary, we estimate that the 108 homozygous mutations detected in M3 plants (Fig. 2) imply an original number of ~288 heterozygous FN-induced mutations in M1 plants (Fig. 3).

Detection of spontaneous mutations in the progenitor line

The progenitor line lacks the five growth-regulating DELLA proteins and is substantially derived from the Ler genetic background. To be sure that the mutations detected in E71, E99, E125, E128, E138, and E216 were genuinely FN-induced and were not simply of spontaneous origin, we determined the frequency of spontaneous mutation in the progenitor line. We sequenced the genomes of single plants from three independent control progenitor lineages. Each of these three plants (F2-1 to -3) was an F2 generation plant (the product of two successive self-pollination generations) derived from a control, nonirradiated progenitor plant (and hence the control equivalent of the M2) (Fig. 3). Using our above-described...
mutation detection methods, we identified only one single mutation in these three genomes (an A:T > C:G substitution in the F2-2 line) (Supplemental Table 2).

In a separate determination of the frequency of spontaneous mutation in the progenitor, we sequenced the genomes of two individual M4 plants (progeny of E99 and E125 genome-sequenced M3 lines) and compared the mutations detected in the M4 with those detected in the M3 (Supplemental Fig. 1). While all mutations previously detected in the M3 were also detected in the M4, a total of seven previously undetected mutations were additionally detected in the E99 and E125 M4 data sets (Supplemental Table 3). Because these additionally detected mutations may have become detectable (homozygous) in the M4 having been undetectable (heterozygous) in the M3, we used Mendelian segregation laws (as in Fig. 3) to predict that there should have been 8.8 such newly detected homozygous variants, a figure slightly higher than the number (seven) actually observed (Supplemental Fig. 1; Supplemental Tables 3, 4). Thus spontaneous mutations do not significantly inflate the number of detected mutations in M4 plants beyond that predicted on the basis of Mendelian segregation of FN-induced heterozygous mutations in M3 plants.

These two separate control experiments indicate that the frequency of spontaneous mutations in the multiple DELLA knockout progenitor line is comparable to that previously observed in the Col-0 wild-type (WT) line (Ossowski et al. 2010). Because the frequency of mutations in FN-irradiated lineages (Figs. 2, 3) is one to two orders of magnitude higher than in the control lineages, we conclude that the mutations detected in our FN-irradiated lineages are indeed substantially caused by FN irradiation.

Chromosomal distribution and frequencies of FN-induced mutations

The 108 detected FN-induced mutations (Fig. 2) ranged in size from single base substitutions (SBSs) and insertions to deletions of one or two bases, to larger deletions (3 bases to 1 kb or greater) (Table 1; Supplemental Table 1). The distribution of mutations across chromosomes in each of the six lineages is shown in Figure 2. The number of mutations per line ranged from eight to 32 (Fig. 2; Table 1). These differences are significant (G-test, \( P = 1.7 \times 10^{-3} \)) and may reflect heterogeneity in radiation dosage or in water content of irradiated seeds (seed radiosensitivity is dependent on water content) (Kamra et al. 1960).

FN exposure increases both overall mutation frequency and the frequency of specific molecular mutation classes. Taking all detected single base mutations together (SBSs and INDELs), we estimated the mutation rate to be \( 359.7 \times 10^{-9} \) per site in the irradiated generation following correction for underlying spontaneous mutations (Table 2A; Supplemental Table 5). In essence, a 60-Gy FN dose caused a transient single-generational \( \approx 50 \)-fold increase in mutation rate above the spontaneous rate (7.1 \( \times 10^{-9} \) per site per generation) (Ossowski et al. 2010).
On a genome-wide scale, 35% of the 108 detected FN mutations were in genic regions (see Supplemental Fig. 2), which is higher than the expected 33% but not significant (G-test, \( P = 0.6 \)). This observation suggests that mutations are not preferentially located in gene-coding, gene regulatory, or intergenic regions (Fig. 2; Supplemental Table 1), despite previous indications of mutational bias associated with chromatin structural difference (Prendergast et al. 2007) and the possibility that local differences in chromatin architecture might differentially affect DNA repair mechanisms, and hence localized mutation rate.

FN-induced INDEL mutations

Although FN exposure was previously thought to induce a predominance of deletion mutations in plants (Bruggemann et al. 1996; Li et al. 2001), we, in fact, detected substantial impact on the relative frequencies of a range of specific mutational classes. Among INDELs, we detected 19 large deletions (>3 bp), ranging in size from 4 bp to 7.2 kb, and six smaller deletions (2–3 bp) (Fig. 4A) in our FN-exposed lineages. In addition, we detected 14 single base deletions, making this the most prominent class of deletion mutation detected (Fig. 4A), unlike previous suggestions that 2- to 4-kb deletions are the most frequent class of FN-induced deletion mutation (Bruggemann et al. 1996; Li et al. 2001). We also detected five single base insertions. The latter were all A or T insertions, suggesting a possible bias against G or C insertions and consistent with an apparent bias in favor of A:T composition (Table 2B) as previously reported in Col-0 MA lines (Ossowski et al. 2010).

Single base deletion and insertion mutations are often caused by replication slippage at homopolymer or polynucleotide repeat regions (Viguera et al. 2001). Consistent with this, of the 19 FN-induced single base INDELs, all five insertions and nine of the 14 (64%) deletions either occurred within or were adjacent to homopolymeric or polynucleotide repeats (Table 3A). In addition, six of the seven 3-bp or 4-bp deletions were located within adenine or thymine homopolymeric or polynucleotide repeats (although larger FN-induced deletions appear less likely to be associated with mononucleotide or dinucleotide microsatellites; seven of 15; 47%) (Table 3B). These observations suggest the possibility that ionizing radiation may promote the incidence of DNA replication slippage.

Onset of radiation frequency generates gross chromosomal aberration in plants (interstitial inversions, translocations, etc.) (Shikazono et al. 2005). To identify such events, we searched for “distant-pair” signatures (as previously described) (Korbel et al. 2007; Gan et al. 2011; Jiang et al. 2011), where a pair of reads from either end of a genome sequencing fragment that straddles a “breakpoint” (inversion or translocation chromosomal aberration) with respect to the progenitor reference genome align at an unexpected distance apart on that reference. Despite exhaustive searches, we did not detect any distant pairs, indicating the absence of gross chromosomal aberration in our six post-irradiation lineages. However, analysis of a further FN-induced M₃ mutant line, E88 (full mutation profile not included in this study), detected a chromosomal rearrangement mutation (an interstitial inversion of −215 kb), thus indicating that our methods are robust (Supplemental Fig. 3A). In addition, using our standard methods, an −21-kb deletion (larger than observed in the original six FN M₃ lines) was identified in the E88 M₃ line (Supplemental Fig. 3B).

### FN-induced single base substitutions

Exposure to FNs also induced a relatively high frequency of single base substitutions (SBs). SBs can be transitions (purine > purine; pyrimidine > pyrimidine) or transversions (purine > pyrimidine; pyrimidine > purine), and FNs induced both types of substitution (Fig. 4B; Supplemental Tables 1, 6). C>T transitions were the most common substitution in MA lines (Fig. 4B; Ossowski et al. 2010), consistent with C being the most spontaneously mutable base, and were also prominent in FN-exposed lines (Fig. 4C). Another form of radiation, UV, is known to induce the formation of covalent linkages between adjacent pyrimidine residues (where a C or T is adjacent to another C or T) in DNA sequence, resulting in a predominance of UV-induced C>T mutations at dipyrimidine sequences (Daya-Grosjean and Sarasin 2005). We identified 64 FN-induced SBs of which 51 (80%) were at pyrimidine dinucleotide sites, which is more than expected by chance (Fisher’s exact test, \( P = 1.9 \times 10^{-6} \)) (Supplemental Tables 1, 6). Of these SBs, 20 were C>T mutations and 14 of these (70%) were at pyrimidine dinucleotide sites (Supplemental Table 6); this was not found to be significant (Fisher’s exact test, \( P = 0.1 \)). We conclude that FN exposure induces C>T transitions (and other SBs) via sequence context-dependent mechanisms preferentially targeting pyrimi-
dine dinucleotides. While previously associated particularly with UV exposure, our results suggest that pyrimidine dinucleotide-associated C→T transitions are actually a feature of exposure to a wider range of radiation classes.

Another prominent class of FN-induced substitution was the G→T transversion (see Supplemental Table 1). Radiation-induced DNA damage is often attributed to the effects of reactive oxygen species (ROS) and hydroxyl radicals generated when water molecules absorb high-energy particles. In particular, ROS are thought to damage guanine residues, thus precipitating G:T mutations. Approximately one-quarter becoming homozgyous mutations (36; blue text) in the M2. Thus, a total of 108 homozgyous variants (detected by our methods) will have segregated in a Mendelian 1:2:1 (homozygous mutation:heterozygous mutation:nonmutant) ratio, with approximately one-quarter becoming homozgyous mutations (36; blue text) in the M2, half becoming heterozygous (144; blue text), with the remaining quarter being homozgyous nonmutant. These considerations permit calculation (see Supplemental Table 5) of FN-induced mutation rates.

![Figure 3. Estimation of original numbers of mutations in M1 plants. M1 seeds were irradiated with FNs, germinated, and resultant plants self-pollinated to yield M2 seed. Elongated hypocotyl mutants were sequenced.](image)

**Table 2.** Comparison of FN-exposure lineage versus MA line mutation rates

<table>
<thead>
<tr>
<th>Mutation</th>
<th>MA lines</th>
<th>FN lines</th>
<th>Fold-rate difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single base substitutions</td>
<td>6.5</td>
<td>264.6</td>
<td>40.7</td>
</tr>
<tr>
<td>Single base deletions</td>
<td>0.3</td>
<td>74.4</td>
<td>248.0</td>
</tr>
<tr>
<td>All mutations</td>
<td>7.1</td>
<td>359.7</td>
<td>50.7</td>
</tr>
</tbody>
</table>

B. A comparison of mutation rates per genome per generation calculated from Col-0 MA lines versus FN lines

<table>
<thead>
<tr>
<th>Mutation</th>
<th>MA lines</th>
<th>FN lines (minus spontaneous rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deletions</td>
<td>0.06</td>
<td>6.32</td>
</tr>
<tr>
<td>Large deletions (&gt;3 bp)</td>
<td>0.03</td>
<td>3.09</td>
</tr>
<tr>
<td>Single base deletions</td>
<td>0.03</td>
<td>2.23</td>
</tr>
<tr>
<td>A</td>
<td>0.02</td>
<td>0.27</td>
</tr>
<tr>
<td>T</td>
<td>0.01</td>
<td>0.48</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0.67</td>
</tr>
<tr>
<td>C</td>
<td>0.01</td>
<td>0.81</td>
</tr>
<tr>
<td>Single base insertions</td>
<td>0.03</td>
<td>0.73</td>
</tr>
<tr>
<td>A</td>
<td>0.01</td>
<td>0.48</td>
</tr>
<tr>
<td>T</td>
<td>0.03</td>
<td>0.25</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Single base substitutions</td>
<td>0.66</td>
<td>1.76</td>
</tr>
<tr>
<td>A→T=G→C</td>
<td>0.08</td>
<td>2.17</td>
</tr>
<tr>
<td>G→C&gt;A→T</td>
<td>0.39</td>
<td>1.19</td>
</tr>
<tr>
<td>A→T&gt;C→G</td>
<td>0.05</td>
<td>1.57</td>
</tr>
<tr>
<td>A→T=A→T</td>
<td>0.03</td>
<td>1.49</td>
</tr>
<tr>
<td>G→C&gt;T→A</td>
<td>0.06</td>
<td>0.51</td>
</tr>
<tr>
<td>G→C&gt;C→G</td>
<td>0.05</td>
<td>8.69</td>
</tr>
</tbody>
</table>

*The FN line data were corrected with the MA line data to account for underlying spontaneous mutation rates. Totals for all mutations are shown in bold. Standard errors of the means (SEM) are shown.*
shelter from environmental mutagenic agents, it is likely that any differences in mutational spectra reflect exposure to a mutagenic environment (in this case, a FN bombardment beam), and imply that the increased frequency of mutations due to FN exposure is not simply due to an accelerated occurrence of the same classes of spontaneous mutations as found in MA lines. There are similarities between the characteristics of FN-induced SBSs and those observed in in vitro regenerant Arabidopsis plants (Jiang et al. 2011), although why this should be the case is not clear.

FN irradiation of plants is traditionally considered to induce deletion mutations of 2–4 kb in size (Bruggemann et al. 1996; Li et al. 2001) and has been used for many years in forward genetics experiments to determine the function of unknown genes (Hoffmann et al. 2007; Hofer et al. 2009). An alternative mutagen is the alkylating agent ethylmethane sulphonate (EMS). While EMS mutagenesis produces both complete and partial loss-of-function mutant alleles (Bowman et al. 1991), almost exclusively via G:C→A:T point mutations (Greene et al. 2003; Martin et al. 2009) at frequencies of ~1000 mutations per Arabidopsis genome (Jander et al. 2003; Martin et al. 2009), FN mutagenesis was previously thought to produce full loss-of-function alleles by inducing a range of deletions (300 bp to 8 kb) (Bruggemann et al. 1996) and insertions (~3.4 kb) (Sun et al. 1992). However, our identification of 108 FN-induced genomic molecular variants indicate that these previous studies may have favored the detection of large-scale mutations. Of the 39 deletions that we detected, 38 (97%) were <56 bp in length, with single base deletions being the most frequently observed (36%, 14 of 39) (Table 3) and with the majority of single base deletions (64%, nine of 14) occurring at homopolymeric or dinucleotide sites (Table 3A). Thus the view that FNs predominantly induce relatively large-scale gene-inactivating deletion mutations seems less likely, although we cannot exclude the possibility that some larger deletions (up to megabases), which often have a reduced male transmission rate (Naito et al. 2005) or may cause lethality when homozygous, may have been under-represented in our genome-sequenced M3 lines. Indeed, analysis of an additional FN-induced mutant line, E88, did confirm that FNs can induce larger deletions (than those observed in the original six FN M3 lines) of >20 kb (see Supplemental Fig. 3B), and in addition, a chromosomal rearrangement mutation (an interstitial inversion of ~215 kb) was identified (see Supplemental Fig. 3A). Taken together, our findings suggest that the six FN lineages analyzed in this study provide a good overall picture of the kinds of variants caused by FNs, although the picture may be biased in favor of mutations less likely to have deleterious effects.

We have here described the establishment of six elongated hypocotyl mutant lineages from mutant homozygotes identified in the M2 generation. In each case, we were able to identify the likely phenotype-conferring mutation (see Fig. 2). However, a major previous concern in mutant screens of this nature has been the unknown frequency, chromosomal distribution, and phenotypic consequence of additional mutagen-induced mutations. This concern is the basis of the traditional approach of sequential backcrossing of novel phenotype-causal mutant alleles to the nonmutagenized progenitor, with the aim of segregating out unknown additional mutations (although mutations genetically linked to the phenotype-causal allele will tend to be retained). Because we analyzed the genomes of lineages selected on the basis of mutant phenotype, it is possible that the number of detected additional mutations is somewhat different from what would have been found in unselected M3 plants. However, our results show for the first time that additional mutations, although they
clearly occur, are not as frequent as might have been supposed, and that the traditional backcrossing approach is indeed a relatively effective means of "cleaning up" the genetic background of mutant lineages. Most importantly, our results identify the genome-wide spectrum of mutations that researchers can expect when using FN irradiation in reverse genetic screens. The high frequency of single base variants suggests that likely causal mutations may not be as readily identifiable using traditional deletion-detection approaches as previously supposed. In addition, our results provide an initial picture of the genome-wide mutational effect of exposure to ionizing radiation, identifying in particular an increase in the relative frequency of transversion mutations. The presence in the environment of cosmic radiation and natural and artificial radionuclides implies a consequential radiation exposure of all organisms. Genome-wide DNA sequence analysis can now be used as a measure of previous exposure to ionizing radiation. Next-generation sequencing of plant genomes could contribute to our understanding of the effect and consequences of environmental irradiation on plant populations, thus helping in the implementation of effective environmental protective measures (Copplestone et al. 2010).

Our analyses reveal for the first time the genome-wide consequences of FN irradiation on plant genomes and the spectrum and rate of mutations that researchers should expect to find in FN-mutagenized plant populations. In addition, our study provides a valuable benchmark for future studies of the effects of ionizing radiation on genome integrity.

### Methods

#### Plant material

All experiments used the Landsberg erecta (Ler) laboratory strain of A. thaliana as genetic background. The progenitor "global-DELLA" (gai-t6, rga-t2, rgl1-1, rgl2-1, and rgl3-4) line was isolated from the F3 progeny of a cross between Ler "quadruple-DELLA" (gai-t6, rga-t2, rgl1-1, and rgl2-1) (Achard et al. 2006) and a line carrying a T-DNA insertion in the A. thaliana Columbia-0 RGL3 locus (rgl3-4) obtained from the publicly available SAIL (Syngenta Arabidopsis Insertion Lines) collection. This rgl3-4 line had previously been backcrossed six times sequentially to the Ler background.

#### Plant mutagenesis, growth conditions, and mutant screening

Progenitor seeds were mutagenized by FN bombardment at a dose of 60 Gy at the KFKI Atomic Energy Research Institute (Budapest, Hungary). Mutants were identified on the basis of phenotypic criteria individually for each type of mutation.

Table 3. Genome-wide analysis of radiation-induced mutations in six FN M3 mutant lines

<table>
<thead>
<tr>
<th>Line</th>
<th>Loci</th>
<th>INDEL size (bp)</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>E125</td>
<td>Chr:17229771</td>
<td>-3</td>
<td>TCCCTCTTTTCTTTGCTGCCA</td>
</tr>
<tr>
<td>E138</td>
<td>Chr:2999448</td>
<td>-3</td>
<td>TTTTGGCTTAGTGGCAGAGA</td>
</tr>
<tr>
<td>E1126</td>
<td>Chr:2804349</td>
<td>-4</td>
<td>CCGATGCCGACATCGGAGT</td>
</tr>
<tr>
<td>E128</td>
<td>Chr:487491</td>
<td>-4</td>
<td>GCCTGCTCCTGACATAGGTC</td>
</tr>
<tr>
<td>E1025</td>
<td>Chr:19632942</td>
<td>-6</td>
<td>GTGGAGAAGTCATGGACAGA</td>
</tr>
<tr>
<td>E138</td>
<td>Chr:5266590</td>
<td>-6</td>
<td>AAGACGACGCGACCCATservicedd</td>
</tr>
<tr>
<td>E89</td>
<td>Chr:9194110</td>
<td>-6</td>
<td>AATACGACAGATGCTGGCACAAA</td>
</tr>
<tr>
<td>E125</td>
<td>Chr:2199360</td>
<td>-8</td>
<td>GACGCGCTCAGCAGCCGTTAGG</td>
</tr>
<tr>
<td>E99</td>
<td>Chr:4965535</td>
<td>-10</td>
<td>AATAGGCTCGGATGCTAGGACAGT</td>
</tr>
<tr>
<td>E126</td>
<td>Chr:13696668</td>
<td>-11</td>
<td>TTCTCACAGACATATGGCAGTCAGGATG</td>
</tr>
<tr>
<td>E125</td>
<td>Chr:13804210</td>
<td>-12</td>
<td>CAGGAAACACTACAGGCTGCTACAGGGAA</td>
</tr>
<tr>
<td>E99</td>
<td>Chr:2896321</td>
<td>-14</td>
<td>ATGCAAGACGACGCTACGGGACGAGGAGATA</td>
</tr>
<tr>
<td>E125</td>
<td>Chr:3727272</td>
<td>-17</td>
<td>ATCCGCTACGAAACGACGGCAAAATAAGGCTAAAT</td>
</tr>
<tr>
<td>E99</td>
<td>Chr:11340050</td>
<td>-28</td>
<td>CTGGAAACAGACGCTTAC...TGGCGGATTTACGCTGC</td>
</tr>
<tr>
<td>E125</td>
<td>Chr:19825067</td>
<td>-32</td>
<td>TGGCTGGATGAGAAAGATATAGTTACGAGT</td>
</tr>
<tr>
<td>E138</td>
<td>Chr:5723251</td>
<td>-43</td>
<td>TGGAGAGATCGTGGCTT...GATTTGATGAGGATG</td>
</tr>
<tr>
<td>E99</td>
<td>Chr:13714219</td>
<td>-55</td>
<td>CTGCGAGACAGGCGAG...AAGAAAGCAGATAGGGCG</td>
</tr>
<tr>
<td>E99</td>
<td>Chr:6220669</td>
<td>-7176</td>
<td>AGTGGCGACATTTAAG...ACAGCACAGACAGCAGCA</td>
</tr>
</tbody>
</table>

INDEL mutation locations and flanking sequences of (A) single base INDELS and (B) larger deletion mutations. Homopolymeric and polynucleotide stretches are shown underlined. Bases in bold are those that were inserted or deleted.
Hungary) and then sown on soil in glasshouses with a 16-h light/8-h dark photoperiod at 22°C–24°C (irradiance 120 μmol of photons/m² per second). The resultant M1 plants were allowed to self-pollinate in pools of about 1000 plants. M3 and M4 plants were grown in controlled environment rooms (CERs) with a 16-h light/8-h dark photoperiod at 22°C–24°C (irradiance 120 μmol of photons/m² per second).

Genetic screens of M2 seedlings for elongated hypocotyl mutants were performed by sowing seeds on soil in CERs and allowing them to germinate and grow for 5–7 d. Seedlings were screened visually for elongated hypocotyls. From a total of about 400,000 seeds, more than 300 mutant seedlings displaying elongated hypocotyls were isolated, grown, and self-pollinated. Segregation analysis based on hypocotyl length was performed on 40 M4 plants per line, identifying homozygous lines E71, E88, E99, E125, E128, E138, and E216.

**Genome sequencing**

Young leaf samples were taken from single nonirradiated progenitor F2, or FN M3 and M4 mutants as well as progenitor plants. DNA was isolated using a Plant DNeasy Mini kit (QIAGEN). Samples representing three independent nonirradiated progenitor derived F2 lines (F2-1 to -3; control plants equivalent to the M2 generation in Fig. 3), seven elongated hypocotyl M3 mutant lines (E71, E88, E99, E125, E128, E138, and E216), an individual E99 M4 and an E125 M4 mutant plant, and the progenitor itself were sequenced using the Illumina Genome Analyzer II platform according to the manufacturer’s instructions at the Wellcome Trust Center for Human Genetics, Oxford, UK, GeneServices, UK or the Beijing Genomics Institute, China. Several lanes of 36-, 51-, 75-, 76-, 90-, and/or 101-bp paired-end runs were produced for each line. A 24–85 (raw sequencing data) and 17–58 (aligned sequencing data) fold depth sequence coverage was obtained for each sample (Supplemental Table 7; sequencing data statistics for the E88 M3 line are not shown).

**Creation of progenitor reference sequence**

To identify DNA sequence variants in the genomes of the six FN-irradiated M3 mutant lines (E71, E99, E125, E128, E138, and E216; versus progenitor), it was necessary to create our own reference genome. We developed a software package called Iterative Read-Mapping and Realignment (IMR) v0.3.0 (http://mus.well.ox.ac.uk/19genomes/IMR-DENOMI/) (Gan et al. 2011), which was used to create the progenitor genome reference sequence. The algorithm underlying IMR iteratively modifies and corrects the reference, such that in each iteration, sequencing reads are mapped to the current version of the reference using MAQ v0.7.1 (Li et al. 2008) and variants identified with high confidence are used to update and improve the reference sequence. The reads are then remapped and the process repeated until convergence. For our data five iterations were run.

In each iteration base substitution calling is performed using an algorithm based on SAMtools v0.1.5c varFilter (Li et al. 2009). The default behavior of varFilter (which we follow) is to accept isolated base substitutions (with less than two other variants within 10 bp) if:

1. coverage ≥ 3 and <100; and
2. the root mean square (RMS) of the mapping qualities (Phred scores) of the aligned reads >25.

Short INDELs (up to 30 bp) reported by SAMtools (Li et al. 2009) were accepted. Under certain circumstances, SAMtools reported two different deletions at the same location. In such cases, the more likely deletion, usually the longer, was accepted.

One important issue that the IMR algorithm addresses is the problem of clusters of base substitutions and whether these are genuine or artifacts of undetected INDELs (often the alignments of the ends of sequencing reads generate base substitution calls in preference to INDELs). Since varfilter of SAMtools (Li et al. 2009) automatically filters out variant sites where three or more base substitutions occur within 10 bp, these cases were reconsidered as follows: the base substitutions with the greatest confidence (i.e., highest RMS) were accepted, and the two criteria (1) and (2) above were applied. Sites that passed these criteria were modified. Variants that were close to INDELs (i.e., within 10 bp) were ignored since they would be considered in the next iteration.

We also found a number of sites that could not be fixed with IMR because there were two or more different bases in the progenitor sequencing reads aligning to one position in the Col-0 reference sequence. To overcome this, we developed a filter to call base substitutions at these ambiguous sites in the progenitor whereby a base substitution was called at a position if:

1. at least 80% of reads had the same base; and
2. the number of sequencing reads was within ±20% of the average genome-wide read coverage.

In total, 169.5 million filtered paired-end progenitor reads were aligned to the original Col-0 reference sequence to create our own progenitor genome reference sequence of 119.6 Mb.

**Detection of DNA sequence variants**

The bioinformatic methods used in this study are the same as those used in Jiang et al. (2011), which found that neither data type nor method significantly altered observed relative frequency or spectrum of detected single base substitutions (SBSs) and small INDELs. In addition, the systematic comparisons performed in Jiang et al. (2011) showed that differences in types of data and data analysis methods between our methods and those used in Ossowski et al. (2010) had negligible effect on detected SBSs and INDELs, and that false-negative rates would have been at comparable small levels in both cases. These observations are also consistent with a recent publication indicating that different softwares are in general comparable (Lunter and Goodson 2011).

Sequencing reads from the three control nonirradiated plants that were derived from the progenitor and grown in independent lines over two successive generations (F2-1 to -3), seven FN-induced elongated hypocotyl mutants (E71, E88, E99, E125, E128, E138, and E216; M3 generation), and two M4 generation E99 and E125 plants were next mapped to the progenitor reference genome using the Burrows-Wheeler Aligner (BWA) v0.5.8a (Li and Durbin 2010). Reads with Phred-scaled mapping quality of 20 or less, reads that were not uniquely mapped, and reads for which the mate (the other read of a pair) did not map onto the reference genome were excluded from further analyses. Between 106.6 and 107.9 million nucleotide sites in the six FN M4 lines (E71, E99, E125, E128, E138, and E216) had sufficient read information (at least one read) for calling initial variants compared with the progenitor reference (an average of 107.5 million sites was used to determine the mutation rates per site calculated in Table 2).

The variant lists were generated for each data set using SAMtools (Li et al. 2009) after alignment with the progenitor reference genome sequence. The variant lists were then filtered to exclude putative false positives called at sites with low sequencing read coverage (which could lead to miscalling errors) or high coverage (that are associated with DNA sequences with a high degree of similarity to other sequences in the genome such as transposable elements). A minimum of seven and a maximum of 75 sequencing reads per site were set as the base sub-
tition threshold limits. For INDELs, at least two reads per site were required. Furthermore, base substitution and INDEL false positives called by SAMtools (Li et al. 2009) after re-alignment of the progenitor sequencing reads to the progenitor reference genome sequence using BWA were subtracted from our elongated hypocotyl mutant variant lists.

Following the filtering processes above, the base substitutions and INDEL lists were then checked manually by visually scanning the alignment files (BAM files) generated by SAMtools (Li et al. 2009) using a locally customized version of Integrated Genome Viewer (IGV) (Robinson et al. 2011), which allowed visual comparison of all aligned reads for each mutant data set with the progenitor reference. The six lines analyzed in this study were simultaneously visualized in IGV (for an example, see Supplementary Fig. 4). Based on comprehensive variant analysis using these data sets, the following filters were identified as those that reduced the number of false positives without excluding real mutations:

(a) Base substitutions: a Phred-scaled mapping quality of >63.
(b) Insertion: a minimum seven reads per site and at least 25% of reads had the extra nucleotides.
(c) Deletion: a minimum six reads per site and at least 30% of reads were missing the nucleotides.

Detection of large INDELs and other chromosomal aberrations

We also ran novel protocols (Jiang et al. 2011) to detect larger-scale variants such as insertions, inversions, and translocations. Novel codes were used to generate lists of “distant-pair” read pairs (where two paired reads [“mates”] align at unexpectedly distant regions of the reference sequence [progenitor]). Reads whose mates mapped >750 bp distant with respect to the reference (progenitor) were called, thus creating new (distant-pair) BAM files. Variant lists of covered regions (coverage >5) were generated and visually inspected with IGV. While no such mutations were detected in the six FN M3 lines (E71, E99, E125, E128, E138, and E216 lines), analysis of a further FN M3 line (E88) confirmed the reliability of these methods.

Further exhaustive searches were performed by systematically scanning each of the six FN M3 line BAM files with IGV. These searches did not reveal any genomic variants (large or small) in FN lines additional to the 108 mutations initially identified by the above-described mutant detection procedures.

Correction of FN-lineage mutation data with respect to spontaneous mutations

A total of 108 homozygous variants (SBs and INDELs) were identified in the genomes of six FN-irradiated M3 Arabidopsis lines. These variants are derived from a combination of both irradiation mutagenesis and also, to a far lesser extent, spontaneous mutational events. To determine the rate and spectrum of FN-induced mutations, the mutation rate and spectrum of spontaneously derived mutations needed to be deduced from our FN data. To do this, we used the results from the Col-0 MA lines experiment, which included a total of 98 substitutions, eight short deletions (ranging in size from 1 to 3 bp), four long deletions (ranging in size from 11 to 5445 bp), and five insertions (all 1 bp) across five MA lines over 30 generations (Ossowski et al. 2010).

To calculate the FN-specific mutation rate observed for single base variants in the different variant categories (substitutions and INDELs) and in the Ti/Tv profiles, we needed to correct the number of mutations observed in the M3 generation in each category twice: first for the spontaneous mutations that accumulated across six lines (E71, E99, E125, E128, E138, and E216) and second for the three generations of single seed descent from the reference strain (progenitor). To do this, we first calculated the frequencies of these events per MA line per generation, and then multiplied them by 18 (= 6 lines × 3 generations) to get the correction factors for different variant categories. These numbers were then subtracted from the number of homozygous mutations observed in M4 generations (e.g., see Table 2A) to obtain the number of homozygous mutations arising as a result of FN treatment.

Data access

The Illumina DNA sequencing data files analyzed in this manuscript have been submitted to the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) under accession number SRA051285.

Acknowledgments

This publication is based on work supported by Award No. KUK-I1-002-03, made by King Abdullah University of Science and Technology (KAUST) and Biological and Biotechnological Sciences Research Council (BBSRC) grant no. BB/F020759/1. We thank the IGV group (Broad Institute, Cambridge, MA) for help with implementation of IGV. We thank David Buck and colleagues at the Wellcome Trust Centre for Human Genetics (Oxford, UK) Genomics Core for advice, help, and performing most of the genomic sequencing, supported by the Wellcome Trust Core grant no. 090352/Z/09/Z.

References


Genomic consequences of fast-neutron irradiation


Received September 2, 2011; accepted in revised form April 4, 2012.