Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in *Arabidopsis thaliana*

JOHN G. JELESKO*, RYAN HARPER*, MASAKI FURUYA†, AND WILHELM GRUISSEM*‡

*Department of Plant and Microbial Biology, 111 Koshland Hall, University of California, Berkeley, CA 94720-3102; and †Hitachi Advanced Research Laboratories, Hatoyama, Saitama, 350-0395, Japan

ABSTRACT Small, multigene families organized in a tandem array can facilitate the rapid evolution of the gene cluster by a process of meiotic unequal crossing-over. To study this process in a multicellular organism, we created a synthetic *RBCSB* gene cluster in *Arabidopsis thaliana* and used this to measure directly the frequency of meiotic, intergenic unequal crossing-over between sister chromatids. The synthetic *RBCSB* gene cluster was composed of a silent ΔRBCS1B::LUC chimeric gene fusion, lacking all 5′ transcription and translation signals, followed by *RBCS2B* and *RBC3B* genomic DNA. Expression of luciferase activity (luc⁺) required a homologous recombination event between the ΔRBCS1B::LUC and the *RBCS3B* genes, yielding a novel recombinant *RBCS3B/1B::LUC* chimeric gene whose expression was driven by *RBCS3B* 5′ transcription and translation signals. Using sensitive, single-photon-imaging equipment, three luc⁺ seedlings were identified in more than 1 million F2 seedlings derived from self-fertilized F1 plants hemizygous for the synthetic *RBCSB* gene cluster. The F2 luc⁺ seedlings were isolated, and molecular and genetic analysis indicated that the luc⁺ trait was caused by the formation of a recombinant chimeric *RBCS3B/1B::LUC* gene. A predicted duplication of the *RBCS2B* gene also was present. The recombination resolution break points mapped adjacent to a region of intron I at which a disjunction in sequence similarity between *RBCS1B* and *RBCS3B* occurs; this provided evidence supporting models of gene cluster evolution by exon-shuffling processes. In contrast to most measures of meiotic unequal crossing-over that require the deletion of a gene in a gene cluster, these results directly measured the frequency of meiotic unequal crossing-over ($\approx 3 \times 10^{-6}$), leading to the expansion of the gene cluster and the formation of a novel recombinant gene.

Genome organization can directly affect the evolution of a gene. Single-copy genes or dispersed members of a multigene family evolve independently. In contrast, members of a multigene family organized as a gene cluster can exploit this organization to generate further gene duplications and novel recombinant genes by a process of unequal crossing-over. For example, a single intergenic unequal crossover event in a gene cluster results in four simultaneous alterations: a deletion, a duplication, and two reciprocal, recombinant genes. The impact of such unequal crossover events evidently have been important in the evolution of complex loci such as *HOX* (1), amylase (2), globin (3), MHC (4), Ig (5), the maize *R-r* complex (6), *RBCS* (7), and plant disease-resistance loci (8–10). Although DNA sequencing of gene clusters provides information about past changes in a particular gene cluster, it can only estimate the rate of unequal crossing-over in terms of geological time scales.

In multicellular organisms, unequal crossing-over is implicated in several genetic disorders. This was demonstrated first with the *Drosophila bar* locus (11). Subsequent research with the *bobbed* locus demonstrated that deletions within the highly repeated rDNA gene cluster causes a range of mutant phenotypes proportional to the size of the rDNA segment deleted (12). Likewise, several human genetic disorders are caused by deletions between tandemly repeated homologous DNA sequences. These deletions are identified in homozygous individuals because of a null phenotype, e.g., β-globin thalessemias (13) and an inherited sensitivity to pressure palsy (14). Thus, in multicellular organisms most mutant-based screens for unequal crossing-over are inherently biased toward identifying the deletion product because these result in a null phenotype. In contrast, the evolutionarily important gene duplication and recombinant gene products from unequal crossing-over generally do not impart a mutant phenotype and, thus, likely go undetected. A more comprehensive understanding of the molecular mechanisms responsible for gene cluster evolution in a multicellular organism requires an effective genetic screen to identify meiotic unequal crossover events that form novel recombinant genes and/or gene duplications within a gene cluster.

In this study, we report a reproducible and highly sensitive genetic screen that specifically identifies a novel recombinant gene and a concomitant gene duplication resulting from meiotic unequal crossing-over in a model multicellular organism, *Arabidopsis thaliana*. This procedure allows the rapid screening of large populations of seedlings, thereby facilitating the identification and isolation of rare recombinant genes. Using this technique, we demonstrate that the frequency of meiotic intergenic unequal crossing-over occurs at a similar frequency as the overall spontaneous mutation rate. The evolutionary implications of this process will be discussed.

MATERIALS AND METHODS

Construction of Synthetic *RBCSB* Gene Cluster, Positive Control, and Transgenic Plants. Recombinant DNA constructs were made by using standard procedures (15). Plasmid pJGJ203 contained the synthetic *RBCSB* gene cluster shown in Fig. 1*b* and was constructed as follows. pJGJ188 consisted of a *Bam*HI–*Pfl*MI–*Hin*dIII–*Bsm*I–*Nar*I linker generated by PCR amplification by using pJGJ184 [promoterless firefly luciferase-NOS 3′ terminator subcloned in pBluescript SK(+)] as a template and oligonucleotide primers o1 (5′-TCCAGGAA-CCAGGGCGTATATCT-3′) and o2 (5′-CGGGATCCATT-CACTGGAAGCTTCAGTGAATGCAAGCTGGAAGAC-GCCAAAAACATAAA-3′) to generate a linker fragment that then was subcloned into pJGJ184 as a *Bam*HI–*Nar*I fragment. JGJ194 was a *Hin*dIII–*Bsm*I *RBCS1B* genomic DNA fragment from pATS17 subcloned into pJGJ188, which was cut similarly. Deletion of the 5′ *RBCS1B* regulatory sequences was

‡To whom reprint requests should be addressed. E-mail: gruissem@ nature.berkeley.edu.

accomplished by digesting pJGJ194 with *Pfl*MI and self-ligated to generate a Δ*RBCS1B::LUC-NOS* terminator gene fusion, pJGJ196. PJGJ193 consisted of a *Cla*I–*Sal*I *RBCS2B-RBCS3B* genomic DNA fragment reconstructed from a *Cla*I–*Sal*I *RBCS2B* subclone from pATS17 and a *Sal*I–*Sph*I *RBCS3B* subclone from pATS5, both introduced into pJGJ186 [pBluescript SK(+) with a modified polylinker to which a *Sph*I site was added by PCR using oligonucleotide primers o4 (5′-TT-GAGCTCATAACTTCGTATAGCATACATTATACGAA-GTTATCCACCGCGGTGGCGG-3′) and o3 (5-AACTCG-AGAATAACTTCGTATAATGTATGCTATACGAAGTT-ATGCATGCGTCGACGGTATCGATAAGC-3′) and pBluescript SK(+) as template]. The synthetic *RBCSB* gene cluster was generated by inserting the *Bam*HI–*Cla*I fragment from pJGJ196 into pJGJ193, which was similarly cut, yielding pJGJ200. The synthetic *RBCSB* gene cluster was subcloned into the T-DNA binary transformation vector pSLJ7292 (16) as a *Sst*I (partial)–*Xho*I fragment, yielding pJGJ203. pJGJ201 consisted of a 2.8-kb *Bam*HI–*Pfl*MI *RBCS1B* promoter containing a fragment from pATS18 subcloned into pJGJ194, which was similarly cut; the intact *RBCS1B* promoter-*RBCS1B::LUC* fusion was subcloned from pJGJ201 into pSLJ7292 on a *Sst*I–*Xho*I fragment, yielding pJGJ204. Transgenic plants were generated by mobilizing the recombinant binary transformation vectors by triparental mating with HB101/pRK2013 into *Agrobacterium tumefaciens* strain GV3101, which was used to perform vacuum-infiltration transformation of *A. thaliana* Col-0 (17).

Transgenic lines AtJGJ203.10 and AtJGJ203.15, AtJGJ204.7, and AtJGJ7292.9 contained a single transgenic locus derived from plasmids pJGJ203, pJGJ204, and pSLJ7292, respectively. A homozygous derivative of AtJGJ203.10, containing a single, random, synthetic *RBCSB* gene cluster insertion, was crossed with a control line, AtJGJ7292.9, generating an F1 population of approximately 10,000 plants hemizygous for the transgenic locus. This F1 population was allowed to self-fertilize, yielding an F2 population with more than 2 million seeds.

Imaging of F2 Seedlings. At least 7,500 F2 seedlings were germinated on a 20 cm × 20-cm field of Vermiculite, moistened with 1× Hoagland's solution, and grown under continuous white-light illumination. Five-day-old F2 seedlings were assayed for *in vivo* luciferase activity as follows. Twenty minutes before photon counting, seedlings were sprayed with 0.5 mM synthetic D-Luciferin (Biosynth, Basel)/0.01% Triton X-100 solution. An intensified charge-coupled device video camera (model C2400 47; Hamamatsu Photonics, Hamamatsu City, Japan) in conjunction with an Image Intensifier Controller (model M4314; Hamamatsu Photonics) and Image Processor (Argus 50; Hamamatsu Photonics) was used to image seedlings within a Hamamatsu Photonics imaging chamber (model A417) mounted with a Xenon CM 120 lens (Schneider, Bad Kreuznach, Germany). Each tray was imaged for 10 min in photon-counting mode; if a suspected luc+ signal appeared, the tray was imaged for an additional 10 min. ARGUS 50 software was used to collect and process digital images; photon-counting imaging was performed in slice/gravity mode whereas reflected green-light imaging was performed by using integration mode. The approximate position of a luc+ seedling first was approximated by superimposing a photon-counting image on a reflected green-light image of the tray containing tooth picks as position markers. Then, 20–40 seedlings were transferred to a water-agar plate and reimaged as above.

Characterization of Genomic DNA. Genomic DNA was isolated by using a CTAB miniprep protocol. Approximately 100 ng of genomic DNA was used in each PCR. Amplification consisted of one cycle of 94°C; 35 cycles of 1 min at 94°C, 1 min at annealing temperature appropriate to each oligonucleotide pair, 1 min/kb of expected size fragment at 72°C; followed by 7 min at 72°C extension. Various oligonucleotide pairs and

respective annealing temperatures utilized were as follows: o13 (5′-CAAAAGAAAGATAAGATAAGGGTGTCAA-3′) and o14 (5′-CCTTTCTTTATGTTTTTGGCGTCTTC-3′) at 53°C; o37 (5′-CCTTGTGATCCTTTTCCCTAC-3′) and o22 (5′-TCTTTATGTTTTTGGCGTCTT 3′) at 52°C. o13–o14 PCRs were analyzed by using 3.5% nondenaturing PAGE (15), whereas the other PCR products were resolved on 0.8% agarose genes. Approximately 2 μg of genomic DNA was cut with *Sph*I and subjected to DNA blot analysis. A full-length *LUC-NOS* probe was isolated as a 2.2-kb *Sac*I fragment from pJGJ184 and labeled with ^{32}P-dCTP (Amersham Pharmacia) by using a Random Primer DNA Labeling Kit version 2 (Takara Shuzo, Kyoto, Japan). Hybridization and washing were carried out at 65°C by using 0.2× standard saline citrate (SSC; 1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7) and 0.1% SDS. Digital imaging of the hybridization signal was done by using Bioimaging analyzer BAS-1000 (Fuji).

RESULTS

To identify meiotic unequal crossover events leading to a recombinant gene, an F1 population of transgenic plants hemizygous for the synthetic *RBCSB* gene cluster (AtJGJ203.10, Fig. 1*b*) was allowed to self-fertilize, yielding a large F2 population. Plants containing the synthetic *RBCSB* gene cluster did not show luciferase activity (luc−) because the Δ*RBCS1B::LUC* chimeric gene lacked all 5′ transcription and translation signals necessary for expression of the chimeric luciferase gene. During meiosis in the F1 generation, misaligned sister chromatids can undergo unequal crossing-over within the synthetic *RBCSB* locus, producing F2 progeny with a recombinant *RBCS3B/1B::LUC* chimeric reporter gene and a luc+ phenotype (Fig. 1*c*). Intact 5-day-old F2 seedlings were assayed for *in vivo* luciferase activity (photon emission) by using sensitive, single-photon counting and video-imaging equipment (18). The vast majority of F2 seedlings did not show photon emission levels above that observed with an empty imaging chamber (Fig. 1*d*). However, rare luc+ F2 seedlings were observed (Fig. 1*e*). The video-imaging equipment lacks sufficient image resolution to identify a single luc+ seedling in a dense field of plants; therefore, a two-step isolation procedure was used to isolate single luc+ seedlings (Fig. 1 *f* and *g*). Because only 10 min of photon counting was sufficient for initial detection of a luc+ seedling, this assay allowed the rapid screening of very large populations (>1.1 million seedlings in 25 hr of camera-imaging time). Three luc+ seedlings (4A1, 6F1, and 6G1), derived from two independent crosses, were isolated from a population of about 1 million observed F2 seedlings. These three lines showed an approximate 3:1 segregation ratio of the luc+ trait in the F3 generation, indicating that the gene responsible for the luciferase activity was heterozygous in the F2 generation.

To determine whether a homologous recombination event had positioned a *RBCSB* promoter upstream of the previously silent Δ*RBCS1B::LUC* reporter gene, genomic DNA from various lines were subjected to PCR analysis by using a *LUC*-specific oligonucleotide primer in conjunction with a primer that binds to a 14-bp sequence present in all *Arabidopsis RBCS* promoters (Fig. 1 *a* and *c*; o14 and o13, respectively). Genomic DNA from all luc+ lines yielded an ≈0.9-kb PCR fragment, whereas genomic DNA from luc− lines did not yield an equivalent-sized fragment (Fig. 2*a*). This indicated that the luc+ phenotype resulted from the fusion of a *RBCSB* promoter to the *LUC* reporter gene, suggesting that this fusion was the result of homologous recombination, rather than an illegitimate recombination event leading to a promoter/enhancer trap effect (19).

To determine which *RBCSB* promoter was responsible for the activation of the *LUC* reporter gene, *Sph*I-digested
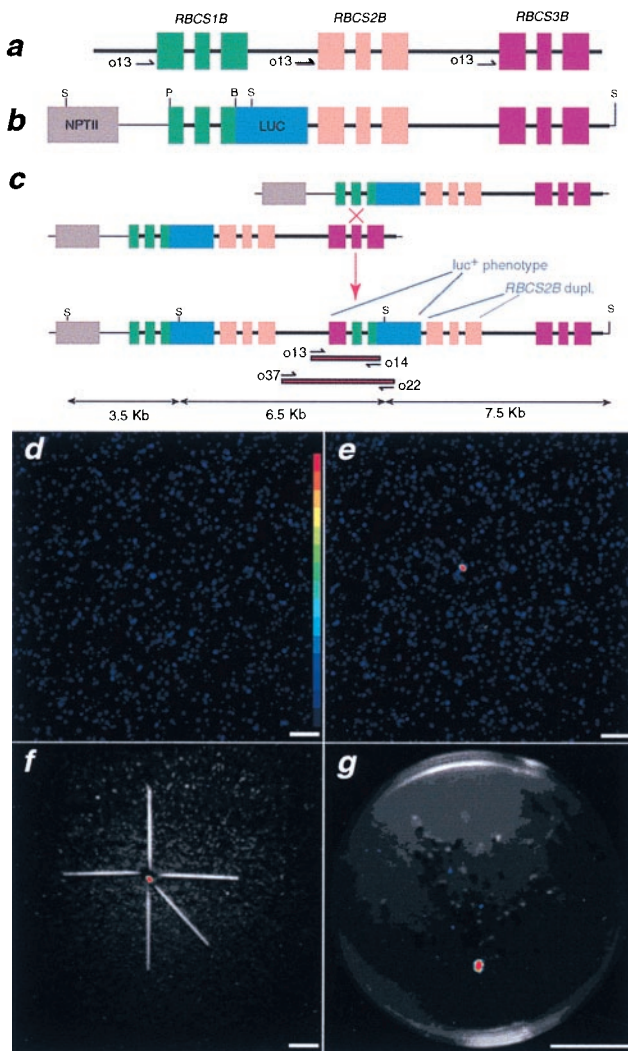
FIG. 1.   Genetic constructs and isolation of luc+ seedlings. (*a*) The *A. thaliana RBCSB* locus. The black line indicates noncoding genomic DNA. Exons shown in color: *RBCS1B* (green), *RBCS2B* (orange), and *RBCS3B* (red). Sizes of introns and exons are not to scale but rather represent genetic organization. Restriction enzymes: P, *Pfl*MI; B, *Bsm*I; and S, *Sph*I. Labeled black half-arrows indicate respective oligonucleotide primer-binding sites. (*b*) Synthetic *RBCSB* gene cluster construct. *NPTII* gene is shown in gray. Δ*RBCS1B::LUC* fusion consisted of *RBCS1B* sequences from the *Pfl*MI in exon I to the *Bsm*I site in exon III; firefly luciferase-*NOS* 3′ terminator (blue) was cloned in-frame 3′ to *RBCS1B* exon III. *RBCS2B-RBCS3B* sequences were positioned 3′ to the Δ*RBCS1B::LUC* fusion. (*c*) An unequal crossover event between sister chromatids containing the synthetic *RBCSB* gene cluster. Red boxes with black borders define the region of gDNA amplified by PCR with the respective oligonucleotide primers. (*d*) Ten-minute photon-counting image of empty imaging chamber. (*Inset*) Pseudocolor step gradient depicting low photon density (light blue) to high photon density (red). [Bar = 4 cm (*d–g*).] (*e*) Ten-minute photon-counting image of a tray containing approximately 7,500 F2 seedlings with 1 luc+ seedling (red spot). (*f*) Image from *e* superimposed on reflected green-light image of the same tray on which toothpicks were placed to approximate the location of the luc+ seedling. (*g*) Twenty-five seedlings transferred from the tray in *f* onto a 0.8% water agar Petri plate. A photon-counting image superimposed on a reflected green-light image allowed the unambiguous identification of a single luc+ seedling on the Petri plate.

genomic DNA, from luc+ and luc− seedlings, was subjected to Southern blot analysis by using a *LUC* probe. Two *LUC* hybridizing bands (3.5 and 7.5 kb) were observed in independent transgenic synthetic *RBCSB* gene cluster containing lines as well as the three luc+ isolates (Fig. 2*b*). A novel 6.5-kb *LUC*
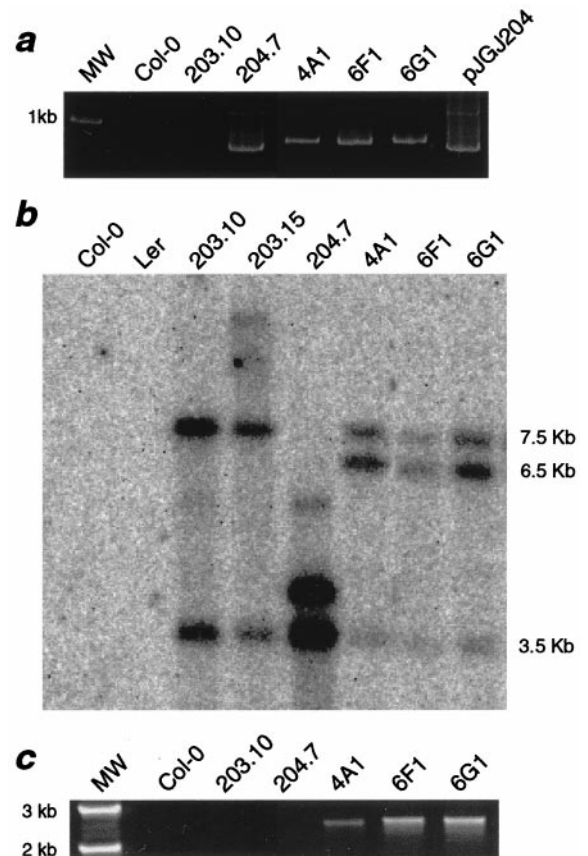

FIG. 2.   *RBCS3B/1B::LUC* responsible for luc+ trait. (*a*) Genomic DNA was isolated from plants and subjected to PCR by using o13 (*RBCSB* promoter-recognition) and o14 (*LUC*-specific) oligonucleotide primers. PCRs were separated on a 3.5% nondenaturing acrylamide gel. MW, molecular weight markers; Col-0, untransformed wild type; 203.10, luc− transgenic line containing the synthetic *RBCSB* gene cluster used to make F1 and F2 generations; 204.7, luc+ transgenic positive control line AtJGJ204.7; 4A1, 6F1, and 6G1, luc+ F2 isolates; pJGJ204, plasmid DNA used to make transgenic positive control line AtJGJ204.7. (*b*) Genomic DNA was cut with *Sph*I and subjected to DNA blot analysis by using a ³²P-dCTP-labeled *LUC-NOS* probe. DNA was loaded in each lane as indicated: Col-0, wild-type Col-0; Ler, wild-type Landsberg erecta; 203.10, luc− transgenic line containing synthetic *RBCSB* gene cluster used to make the F1 population; 203.15, independent luc− transgenic line containing a synthetic *RBCSB* gene cluster; 204.7, luc+ positive control line AtJGJ204.7; 4A1, 6F1, and 6G1, luc+ F2 isolates. (*c*) PCR amplification products from genomic DNA using an oligonucleotide primer pair specific for *RBCS2B-3B* intergenic region (o37) and *LUC* (o22) (see Fig. 1) were separated on a 0.8% agarose gel.

hybridizing fragment was observed only in genomic DNA from the luc+ F2 isolates (Fig. 2*b*). The 6.5-kb *Sph*I fragment is consistent with a homologous recombination event between two misaligned sister chromatids, such that the Δ*RBCS1B::LUC* gene on one sister chromatid crossed over with the *RBCS3B* gene present in the synthetic *RBCSB* gene cluster on the other sister chromatid (Fig. 1*c*). The 6.5-kb *Sph*I fragment indicated that a duplicate copy of *RBCS2B* also was present. To test the assertion that *RBCS3B* promoter sequences were responsible for expression of the *LUC* gene, genomic DNA was subjected to PCR analysis by using a primer specific to a unique *RBCS2B-3B* intergenic region in conjunction with a *LUC*-specific primer (Fig. 1*c*, o37 and o22). Consistent with the assertion that 5′ *RBCS3B* sequences were positioned upstream of the *LUC* sequences, only luc+ lines containing the 6.5-kb *LUC*-hybridizing *Sph*I fragment yielded the expected 2.8-kb PCR fragment (Fig. 2*c*).

In the postulated recombinant *RBCS3B/1B*::*LUC* chimera, the DNA sequences upstream of the *Pfl*MI site present in the original Δ*RBCS1B*::*LUC* fusion should be replaced with *RBCS3B* specific promoter, 5′ untranslated leader, and exon I sequences. The *RBCS3B* gene has a 21-bp insertion within the 5′ untranslated leader not present in the *RBCS1B* gene; consequently, a recombinant chimeric *RBCS3B/1B*::*LUC* gene should amplify a proportionally larger PCR fragment with primers o13 and o14 relative to a control *RBCS1B* promoter-*RBCS1B*::*LUC* transgene in line AtJGJ204.7. Consistent with this assertion, during nondenaturing PAGE conditions, the PCR fragments amplified from the F2 luc⁺ isolates showed a larger PCR fragment (Fig. 2*a*) relative to a *RBCS1B* promoter-*RCSB1B*::*LUC* control line (204.7). The PCR fragments were subcloned, sequenced, and aligned to genomic *RBCS1B* and *RBCS3B* sequences. Fig. 3 shows that *RBCS3B*-specific nucleotides were present in the promoter, 5′ untranslated leader, and exon I of the three luc⁺ recombinant lines. Using polymorphisms between *RBCS1B* and *RBCS3B*, the recombination resolution break points were localized to 3 bp in 4A1 (positions 447–449), 5 bp in 6F1 (positions 451–455), and 21 bp in 6G1 (positions 457–477). It is noteworthy that these recombination resolution break points were located close to the intron I–exon II boundary. Sequences 3′ to exon II showed a *RBCS1B*-specific DNA sequence pattern (data not shown). These data provided definitive evidence that a recombinant *RBCS3B/1B*::*LUC* chimera was responsible for the luc⁺ phenotype.

The reproducibility of these meiotic unequal crossover events was demonstrated in unrelated experiments in which four independent crosses with AtJGJ203.10 and two independent crosses with another synthetic *RBCSB* containing transgenic line (AtJGJ203.15) consistently yielded F2 populations with 3–4 luc⁺ seedlings per 1.1 million observed plants. Similarly, in these crosses, the luc⁺ phenotype was a result of a recombinant *RBCS3B/1B*::*LUC* chimeric gene (data not shown). Given the reproducible frequency of *RBCS3B/1B*::*LUC* recombinants, these results provided an accurate measure of meiotic unequal crossing-over at a (synthetic) gene cluster in *A. thaliana* (≈3 × 10⁻⁶).

DISCUSSION

It has been difficult to study unequal crossing-over events in multicellular organisms during the course of a single generation because of the lack of an effective genetic screen. Although many insights have been gained from analysis of unequal crossing-over in yeast (20–23), it is now clear that the yeast genome (with the exception of the rDNA genes) is largely devoid of multigene families organized as gene clusters of more than two paralogous genes (24). It appears that the yeast genome is a rearranged and reduced ancient tetraploid (25), suggesting that an economy of genome size may be advantageous. The situation in yeast is in marked contrast to many examples of gene clusters in multicellular organisms, which often have more than three genes (1–7, 9, 26–29). Multicellular organisms likely use gene-cluster organization to increase the number of genes in the cluster, and/or recombinant genes, providing additional copies free to evolve new functions or specificities. It is noteworthy that many examples of gene clusters are involved either with the evolution of new morphological structures (i.e., *HOX* complex) or with pathogen defense responses (e.g., Ig heavy chain, MHC, and plant disease-resistance genes). In short, recombination frequencies and mechanisms found in yeast may not accurately reflect those mechanisms responsible for the evolution of gene clusters in multicellular organisms.

Previous to this report, experimental measures of unequal crossing-over in multicellular organism have been inadequate to provide an in-depth understanding of how recombination contributes to the evolution of gene clusters. This is partly the result of the inherent bias toward identifying the deletion products (null alleles) rather than the potentially evolutionary, advantageous gene duplications and recombinant genes. Deletion alleles within a gene cluster are especially problematic to interpret because they also can form by an intrachromosomal looping-out mechanism. Investigations in the *Drosophila*
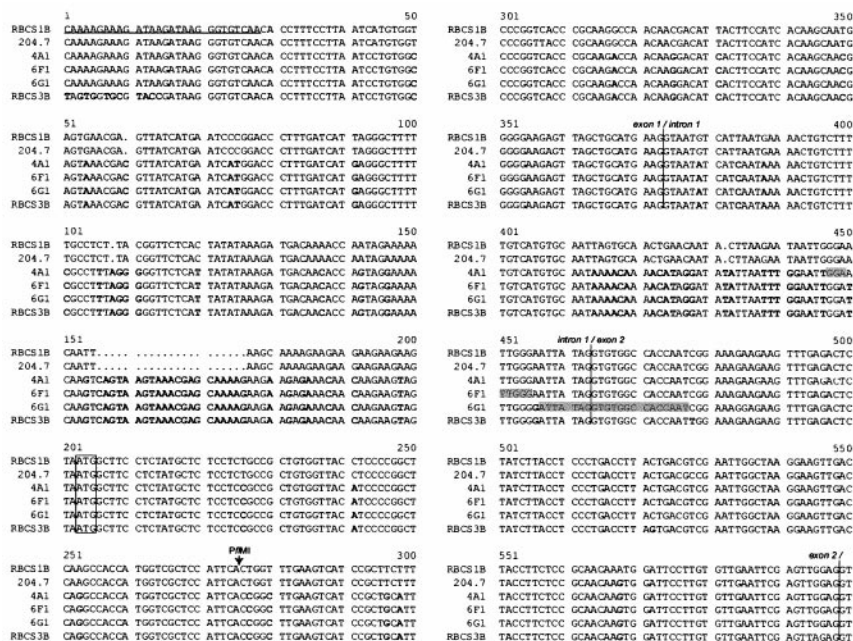

Fig. 3. 5′ DNA sequence of the chimeric *RBCS3B/1B*::*LUC* genes. 4A1, 6F1, 6G1, and AtJGJ204.7 PCR fragments shown in Fig. 2*a* were sequenced, and 600 bp of 5′ DNA sequence were aligned with genomic *RBCS1B* and *RBCS3B* DNA sequences. Underline indicates the o13 oligonucleotide sequence used in the PCR. Bases shown in bold type indicate *RBCS3B*-specific signatures. Box indicates translation initiation codon in exon I. Arrow indicates *Pfl*MI restriction site used to make the 5′ boundary of the Δ*RBCS1B*::*LUC* chimeric gene. Vertical lines define intron–exon boundaries. Shaded areas define the regions in which the recombination resolution break point must have occurred.

bobbed locus demonstrate that deletions and expansions within the large, tandem arrays of rDNA can be detected during the course of a single generation (30). However, the highly conserved rDNA sequence and large, tandem arrays of rDNA make it difficult either to predict or map specific recombination crossover sites, thus limiting the scope of molecular analysis. A better understanding of the evolution of gene clusters in multicellular organisms would be advanced by a detailed characterization of meiotic unequal crossing-over, leading to defined gene duplications and novel recombinant genes.

This report describes a genetic screen to detect recombinant genes produced by meiotic unequal crossing-over within a synthetic plant gene cluster. In contrast to previous measures of unequal crossing-over that required a mutant phenotype, this procedure used the activation of a chimeric reporter gene imparting a novel phenotype. *In vivo* firefly luciferase activity was a particularly good trait because there is no endogenous luciferase activity, and, by using single-photon-counting video-imaging equipment, even very low levels of luciferase activity could be identified easily. For example, as few as 23 apparent photons in 20 min of imaging time were sufficient to identify a plant containing a single *RBCS3B/1B::LUC* recombinant gene in a field of more than 7,500 seedlings. This allowed the rapid screening of millions of seedlings in a relatively short time period. Another advantage of this experimental system is the ability to isolate viable progeny containing the recombinant alleles, thus allowing further genetic analysis of the recombinant synthetic gene cluster. This is in contrast to PCR-based screens of sperm that also identify rare recombinant genes (31) but destroy the gametes in the process (32), thus precluding the possibility of further genetic characterization.

Many attempts to investigate DNA recombination in plants use overlapping mutant copies of a bacterial reporter gene, introduced either as transient (33–40) or stable transgenes (41–45), to provide measures of somatic DNA recombination leading to a functional copy of the bacterial transgene. Using similar bacterial transgenes, two reports (46, 47) measured the frequency of meiotic DNA recombination (6×10^{-6} and $<2 \times 10^{-5}$, respectively). In contrast to these reports, our study provided a measure of meiotic unequal crossing-over ($\approx 3 \times 10^{-6}$) between plant genes leading to an expanded synthetic *RBCSB* gene cluster. This provided a unique assay to recapitulate the recombination behavior of an endogenous plant gene cluster. The observed frequency of meiotic unequal crossing-over is considerably lower than that of yeast artificial gene duplications ($0.9–1.5 \times 10^{-2}$) (20). The reason for a 10,000-fold difference in the frequency of unequal crossing-over between yeast and plants is not known. Interestingly, the frequency of unequal crossing-over within the synthetic *RBCSB* gene cluster was well within the range of the spontaneous mutation rate observed in various eukaryotes, 10^{-5}–10^{-6} deleterious mutations per gene per generation (48). This raises the possibility that an advantage of organizing a small multigene family as a gene cluster may be the generation of additional gene copies at a similar frequency as they are "lost" because of spontaneous mutation. Such a compensatory mechanism would not be expected when individual members of a small, multigene family are positioned at different locations in the genome. Furthermore, our study provides quantitative evidence that unequal crossing-over within a gene cluster occurs more frequently than could be predicted from comparative DNA sequence analysis between related species, which measure the rate of change in terms of geological time scales.

Several lines of evidence support the assertion that meiotic unequal crossover events between sister chromatids were responsible for the formation of the *RBCS3B/1B::LUC* genes. Although it is formally possible that the endogenous *RBCS3B*

gene on chromosome 5 could recombine with the randomly integrated synthetic *RBCSB* gene cluster to form the *RBCS3/1B::LUC* gene, such an ectopic recombination event would yield a 11.3-kb *Sph*I *LUC* hybridizing band in the DNA gel blot analysis. However, an 11.3-kb LUC hybridizing fragment was not observed in Fig. 2*b*. Instead, a unique, 6.5-kb *Sph*I *LUC* hybridizing fragment was observed in all of the F2 luc$^+$ recombinants and was consistent with the restriction map of the *RBCS3B* gene present in the original synthetic *RBCSB* gene cluster. In addition, a *RBCS2B* gene duplication was present in the luc$^+$ recombinants and consistent with a meiotic unequal crossover event between misaligned sister chromatids.

DNA sequencing of PCR products from luc$^+$ lines definitively established that the luc$^+$ trait was caused by a recombinant *RBCS3B/1B::LUC* chimeric gene. DNA sequencing of the o13-o14 PCR products showed *RBCS3B*-specific signatures in the promoter, 5′ untranslated leader, exon I, and most of intron I. However, exon II DNA sequences showed a clear *RBCS1B* signature. This confirms the chimeric nature of the recombinant gene. Moreover, the recombination resolution break points mapped to unique positions in the 3′ region of intron I. This result confirmed that the three isolated *RBCS3B/1B::LUC* alleles were formed by independent meiotic recombination events. The recombination resolution break points in the three recombinants mapped very close to a region of relatively lower sequence similarity between *RBCS1B* and *RBCS3B*. From this limited sampling it is unclear whether the clustering of recombination resolution break points represents a recombination hot spot; alternatively, this grouping may reflect an inherent property of meiotic recombination.

The appearance of recombination resolution break points immediately adjacent to regions of lower DNA sequence similarity may provide important insights into how a particular aspect of recombination molecular biology might manifest an interesting predisposition for exons to shuffle as intact units during the evolution of gene clusters. The double-strand-break model of homologous recombination (49) predicts that DNA recombination begins by aligning homologous DNA sequences and the introduction of a double-strand DNA break in one of the paired chromosomes. This initiates a process whereby 3′ DNA strands are produced by an exonuclease activity and these single-stranded DNA molecules invade the opposite chromosome. The invading strands then are extended, forming a branched complex called a Holliday junction. The location of the Holliday junction then can move by a process of branch migration, and final resolution of the Holliday junction occurs by subsequent DNA cleavage and ligation steps (resulting in the recombination-resolution break point). The factors affecting branch migration and final recombination resolution are not well understood in eukaryotes. If recombination resolution tended to occur within regions of high DNA sequence similarity, then recombination-resolution break points generally should map within exons. However, the mapping of recombination-resolution break points to *RBCSB* intron I do not support a preference for recombination resolution within exons. Instead, our results suggest that recombination-resolution processes might be stimulated when migration of the Holliday junction encounters a transition from high to lower DNA sequence similarity (such as a promoter, 5′ untranslated, or intron). This might cause the Holliday junction to pause and complete final recombination resolution adjacent to regions where DNA sequence diverges between the parental DNA strands. This bias would have a simple, yet profound, effect on the final recombination products such that exons would tend to shuffle as intact units. Thus, information about the molecular details of meiotic recombination may provide important insights into how exon-shuffling processes could contribute to the evolution of gene clusters.

The genetic screen described in this report provides an effective assay to investigate meiotic unequal crossing-over in a model multicellular organism, *A. thaliana.* By using this system it will be possible to characterize further the molecular genetic mechanisms of meiotic recombination that recapitulate changes in gene organization that are important in the evolution of gene clusters.

1. Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T. & Pendleton, J. W. (1994) *Annu. Rev. Genet.* **28,** 423–442.
2. Gumucio, D. L., Wiebauer, K., Caldwell, R. M., Samuelson, L. C. & Meisler, M. H. (1988) *Mol. Cell. Biol.* **8,** 1197–1205.
3. Collins, F. S. & Weissman, S. M. (1984) *Prog. Nucleic Acid Res. Mol. Biol.* **31,** 315–462.
4. Hughes, A. L. & Yeager, M. (1997) *BioEssays* **19,** 777–786.
5. Matsuda, F. & Honjo, T. (1996) in *Advances in Immunology*, ed. Dixon, F. J. (Academic, San Diego), Vol. 62, pp. 1–29.
6. Walker, E. L., Robbins, T. P., Bureau, T. E., Kermicle, J. & Dellaporta, S. L. (1995) *EMBO J.* **14,** 2350–2363.
7. Dean, C., Pichersky, E. & Dunsmuir, P. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40,** 415–439.
8. Song, W. Y., Pi, L. Y., Wang, G. L., Gardner, J., Holsten, T. & Ronald, P. C. (1997) *Plant Cell* **9,** 1279–1287.
9. Hammond-Kosack, K. E. & Jones, J. D. G. (1997) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48,** 575–607.
10. Baker, B., Zambryski, P., Staskawicz, B. & Dinesh-Kumar, S. P. (1997) *Science* **276,** 726–733.
11. Sturtevant, A. H. (1925) *Genetics* **10,** 117–147.
12. Ritossa, F. M., Atwood, K. C. & Spiegelman, S. (1966) *Genetics* **54,** 819–834.
13. Metzenberg, A. B., Wurzer, G., Huisman, T. H. & Smithies, O. (1991) *Genetics* **128,** 143–161.
14. Murakami, T., Garcia, C. A., Reiter, L. T. & Lupski, J. R. (1996) *Medicine* **75,** 233–250.
15. Ausubel, F. A., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1996) *Current Protocols in Molecular Biology* (Wiley, New York).
16. Jones, J. D. G., Shlumukov, L., Carland, F., English, J., Scofield, S. R., Bishop, G. J. & Harrison, K. (1992) *Transgenic Res.* **1,** 285–297.
17. Bechtold, N., Ellis, J. & Pelletier, G. (1993) *Comptes Rendus de l'Academie des Sciences Serie III Sciences de la Vie* **316,** 1194–1199.
18. Furuya, M. & Inoue, Y. (1994) in *Photomorphogenesis in Plants*, eds. Kendrick, R. E. & Kronenburg, G. H. M. (Kluwer, Dordrecht, The Netherlands), pp. 29–40.
19. Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J. D. G., Dean, C., Ma, H. & Martienssen, R. (1995) *Genes Dev.* **9,** 1797–1810.
20. Jackson, J. A. & Fink, G. R. (1985) *Genetics* **109,** 303–332.
21. Petes, T. D. (1980) *Cell* **19,** 765–774.
22. Petes, T. D. & Pukkila, P. J. (1995) in *Advances in Genetics*, eds. Hall, J. C. & Dunlap, J. C. (Academic, San Diego), Vol. 33, pp. 41–62.
23. Gottlieb, S., Wagstaff, J. & Esposito, R. E. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 7072–7076.
24. Goffeau, A., *et al.* (1997) *Nature (London)* **387** (Suppl), 1–105.
25. Wolfe, K. H. & Shields, D. C. (1997) *Nature (London)* **387,** 708–713.
26. Dixon, M. S., Hatzixanthis, K., Jones, D. A., Harrison, K. & Jones, J. D. (1998) *Plant Cell* **10,** 1915–1926.
27. Meyers, B. C., Chin, D. B., Shen, K. A., Sivaramakrishnan, S., Lavelle, D. O., Zhang, Z. & Michelmore, R. W. (1998) *Plant Cell* **10,** 1817–1832.
28. Hulbert, S. H. (1997) *Annu. Rev. Phytopathol.* **35,** 293–310.
29. Ellis, J., Lawrence, G., Ayliffe, M., Anderson, P., Collins, N., Finnegan, J., Frost, D., Luck, J. & Pryor, T. (1997) *Annu. Rev. Phytopathol.* **35,** 271–291.
30. Locker, D. (1976) *Mol. Gen. Genet.* **143,** 261–268.
31. Tusie-Luna, M. T. & White, P. C. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10796–10800.
32. Li, H., Cui, X. & Arnheim, N. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4580–4584.
33. Offringa, R., De Groot, M. J. A., Haagsman, H. J., Does, M. P., Van den Elzen, P. J. M. & Hooykaas, P. J. J. (1990) *EMBO J.* **9,** 3077–3084.
34. Hrouda, M. & Paszkowski, J. (1994) *Mol. Gen. Genet.* **243,** 106–111.
35. Lyznik, L. A., McGee, J. D., Tung, P. Y., Bennetzen, J. L. & Hodges, T. K. (1991) *Mol. Gen. Genet.* **230,** 209–218.
36. Puchta, H. & Hohn, B. (1991) *Mol. Gen. Genet.* **230,** 1–7.
37. Puchta, H. & Hohn, B. (1991) *Nucleic Acids Res.* **19,** 2693–2700.
38. Puchta, H., Kocher, S. & Hohn, B. (1992) *Mol. Cell. Biol.* **12,** 3372–3379.
39. Puchta, H., Dujon, B. & Hohn, B. (1993) *Nucleic Acids Res.* **21,** 5034–5040.
40. Baur, M., Potrykus, I. & Paszkowski, J. (1990) *Mol. Cell. Biol.* **10,** 492–500.
41. Chiurazzi, M., Ray, A., Viret, J. F., Perera, R., Wang, X. H., Lloyd, A. M. & Signer, E. R. (1996) *Plant Cell* **8,** 2057–2066.
42. Lebel, E. G., Masson, J., Bogucki, A. & Paszkowski, J. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 422–426.
43. Peterhans, A., Schlupmann, H., Basse, C. & Paszkowski, J. (1990) *EMBO J.* **9,** 3437–3446.
44. Swoboda, P., Gal, S., Hohn, B. & Puchta, H. (1994) *EMBO J.* **13,** 484–489.
45. Puchta, H., Swoboda, P., Gal, S., Blot, M. & Hohn, B. (1995) *Plant Mol. Biol.* **28,** 281–292.
46. Tovar, J. & Lichtenstein, C. (1992) *Plant Cell* **4,** 319–332.
47. Assaad, F. F. & Signer, E. R. (1992) *Genetics* **132,** 553–566.
48. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998) *Genetics* **148,** 1667–1686.
49. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, R. W. (1983) *Cell* **33,** 25–35.