A Retrotransposon-Mediated Gene Duplication Underlies Morphological Variation of Tomato Fruit

Han Xiao, *et al.*

Science **319**, 1527 (2008);
DOI: 10.1126/science.1153040

A Retrotransposon-Mediated Gene Duplication Underlies Morphological Variation of Tomato Fruit

Han Xiao,[1] Ning Jiang,[2] Erin Schaffner,[1,3]* Eric J. Stockinger,[1] Esther van der Knaap[1]†

Edible fruits, such as that of the tomato plant and other vegetable crops, are markedly diverse in shape and size. *SUN*, one of the major genes controlling the elongated fruit shape of tomato, was positionally cloned and found to encode a member of the IQ67 domain–containing family. We show that the locus arose as a result of an unusual 24.7-kilobase gene duplication event mediated by the long terminal repeat retrotransposon *Rider*. This event resulted in a new genomic context that increased *SUN* expression relative to that of the ancestral copy, culminating in an elongated fruit shape. Our discovery demonstrates that retrotransposons may be a major driving force in genome evolution and gene duplication, resulting in phenotypic change in plants.

Most crop species display distinctly different morphologies in comparison with those of the relatives from which they were domesticated (*1*, *2*). For tomato (*Solanum lycopersicum*), cultivated types bear fruit of varying sizes with many diverse shapes (*2*). The locus *sun* comprises a major quantitative trait locus explaining 58% of the phenotypic variation in a cross that was derived from the *S. lycopersicum* variety "Sun1642," which has elongated shaped fruit, and its wild relative *S. pimpinellifolium* (accession LA1589), which has round fruit (*3*). Fine mapping indicates that *sun* resides in a region of the tomato genome on the short arm of chromosome 7 that carries small-scale insertions, deletions, and tandem duplications (*4*). One insertion, estimated to be 30 kb, is present in Sun1642 (but not in LA1589) and is linked to fruit shape (*4*).

We constructed a genetic and physical map of the *sun* locus in both parents [supporting online material (SOM) Materials and Methods and fig. S1, A and B]. Comparative sequence analysis of the locus showed that the size difference was due to the insertion of a 24.7-kb segment present in Sun1642 but absent from LA1589 (Fig. 1A). This insertion completely coassorted with the phenotype (fig. S1B), implying that it underlies the molecular basis of elongated fruit shape. A Southern blot showed that LA1589 contained only one copy of this sequence, whereas Sun1642 contained two copies (fig. S1C). Both accessions shared a copy of this locus on chromosome 10, suggesting that the inserted segment on chromosome 7 in Sun1642 originated from chromosome 10 (Fig. 1B and fig. S1C). Comparative sequence analysis of the 24.7-kb duplicates showed that *IQD12* was rearranged relative to the other sequences within the 24.7-kb segment (Fig. 1). Moreover, the only nucleotide divergence between the two copies was a 3–base pair (bp) mismatch present at the breakpoint of the rearrangement, which suggests a recent origin of the duplication. Close inspection of the sequences indicated that a *Copia-like* retrotransposon, which we named *Rider*, was present at both loci. *Rider* resembled a long terminal repeat (LTR) retrotransposon because it was flanked by identical 398-bp LTRs (designated in red as "1" and "2") and a 5-bp target site duplication (TSD) sequence GACCT (Fig. 1B). Similarly, the chromosome 7 region had the identical *Rider* element flanked by two LTRs (LTR1 and LTR2). One additional LTR (LTR3) was identified further downstream of *Rider* on chromosome 7, which (together with LTR1) flanked the entire duplicated fragment (Fig. 1A). At the site of the presumed integration, immediately flanking LTR1 and LTR3, a 5-bp motif ATATT was identified, resembling a TSD of a transposition event (Fig. 1A). Sequencing of LA1589 supported that the integration of the segment occurred at this 5-bp motif. These data, along with analyses of polypurine tracts (PPT) (SOM Text), suggest that the entire 24.7-kb fragment transposed via *Rider* from chromosome 10 to chromosome 7 as one large retrotransposon.

Because the fruit-shape phenotype at *sun* was entirely linked to the inserted segment (fig. S1B), we considered that the elongated fruit shape may be due to genes present on the transposed fragment or the disruption of a gene as a result of the insertion event. We identified five putative genes at the *sun* locus in Sun1642 as well as *Rider* (Fig. 1A). Four of these genes were *IQD12* [which belonged to the IQ67 domain-containing plant proteins (*5*)], an *SDL1-like* gene with high similarity to a tobacco and *Arabidopsis* gene (*6*, *7*), and two hypothetical genes *HYP1* and *HYP2* encoding a protein with weak similarity to CUC1 (*8*) and a 487–amino acid protein with high similarity to a potato protein of unknown function (AY737314), respectively. The fifth gene was *DEFL1*, encoding a putative secreted defensin protein (BT012682). The insertion into the intron of this gene strongly suggested that *DEFL1* in Sun1642 was disrupted (Fig. 1A). The *Rider* retrotransposon contained a single open reading frame, which encoded a 1307–amino acid polyprotein including integrase and reverse transcriptase.

To test the effect of the *sun* locus on fruit shape in a homogeneous background and to ascertain whether a difference in expression of one of the candidate genes might underlie the phenotype, we generated a set of near-isogenic lines (NILs) that differed at *sun*. The NIL in the LA1589 background showed negligible ovary-shape differences in floral buds 10 days before flower opening (Fig. 2A) ($n = 10$ buds, $P = 0.64$). At anthesis, ovary shapes began to show sig-

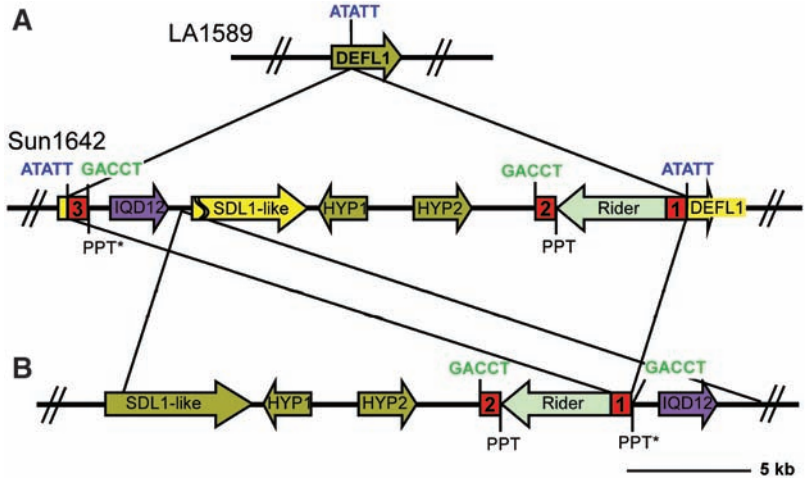

Fig. 1. Structural organization of the *sun* locus in LA1589 and Sun1642 (**A**) and the ancestral locus on chromosome 10 (**B**). Arrows show directionality of the predicted genes and pseudogenes. Dark green arrows indicate ab initio predicted genes, purple arrows indicate the rearranged *IQD12* gene, light green arrows indicate the LTR retrotransposon *Rider*, and yellow arrows indicate pseudogenes. Red numbered boxes identify *Rider*'s LTRs and are numbered according to the predicted order of transcription. The TSD sequences are indicated in green (chromosome 10) and blue (chromosome 7). The position of the original PPT and predicted actual PPT* are indicated (see SOM Text for more details).

[1]Department of Horticulture and Crop Science, Ohio State University/Ohio Agricultural Research and Development Center, Wooster, OH 44691, USA. [2]Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA. [3]College of Wooster, Wooster, OH 44691, USA.

*Present address: College of Medicine, University of Toledo, Toledo, OH 43606, USA.

†To whom correspondence should be addressed. E-mail: vanderknaap.1@osu.edu

nificant, albeit slight, differences ($n = 40$ ovaries, $P = 0.04$). The most significant differences in shape were found in developing fruit 5 days after anthesis (Fig. 2A) ($n = 35$ fruits, $P < 0.0001$), indicating that shape change primarily followed pollination and fertilization (*3*).

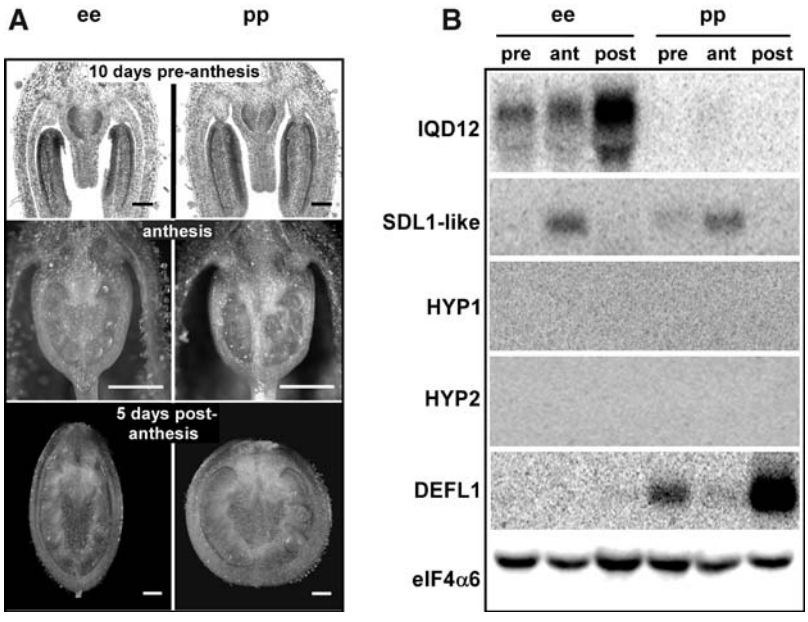

Fig. 2. *sun* affects fruit shape after anthesis. Data shown are from NILs in the LA1589 background. "ee" denotes homozygous Sun1642, whereas "pp" denotes homozygous LA1589 "pre," pre-anthesis; "ant,"-anthesis; "post," post-anthesis. (**A**) Representative ovaries at three developmental stages. Scale bars for each stage (from top to bottom) are 100, 500, and 500 μm, respectively. (**B**) Expression analysis of five candidate genes at *sun*. Total RNA was isolated from tissues of each genotype at the same developmental stages shown in (A), blotted, and hybridized with the probes indicated.

A much higher transcript level of *IQD12* was observed in the NIL carrying the Sun1642 allele in comparison with the NIL carrying the LA1589 allele (Fig. 2B). The highest transcript levels of *IQD12* were found in young developing fruit 5 days after anthesis. Transcript levels of the disrupted gene *DEFL1* showed an inverse expression pattern relative to *IQD12*: When *IQD12* was expressed, *DEFL1* was not (and vice versa) (Fig. 2B). Reverse transcription polymerase chain reaction failed to detect any *DEFL1* transcript in NILs that carried the chromosomal duplication, indicating that *DEFL1* expression was abolished as a result of *Rider*'s transposition into this gene (Fig. 1). Transcript levels of *SDL1-like*, *HYP1*, and *HYP2* were not altered in the NILs, or were undetectable, and thus were deemed less likely candidates for the *SUN* gene.

Fruit-shape phenotype was hypothesized to be dosage-dependent, because heterozygous *sun* NIL plants exhibit an intermediate fruit shape relative to that of the parents in both the LA1589 and Sun1642 backgrounds (fig. S2). Northern blots demonstrated that *IQD12* was expressed about twofold higher in individuals homozygous for the transposed fragment than in heterozygous plants. Similarly, *DEFL1* was expressed about twofold higher in homozygous individuals lacking the transposed fragment than in heterozygotes (fig. S2). Thus, the transposition event most likely placed *IQD12* under the cis-regulatory control of factors normally conferring

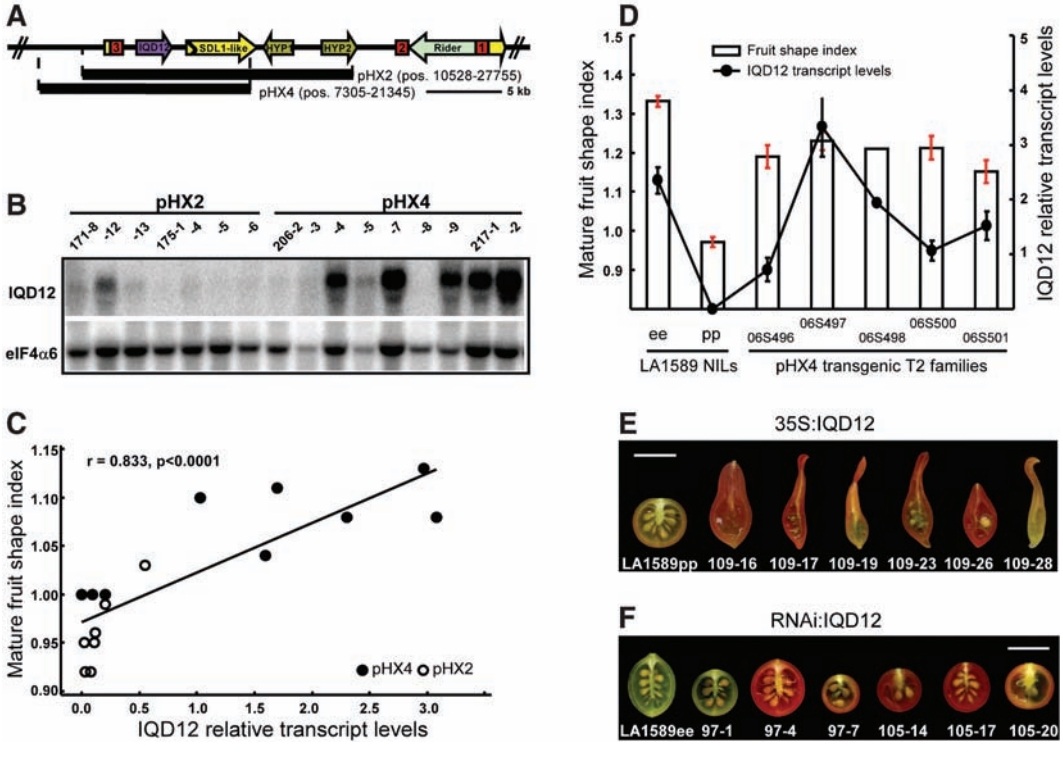

Fig. 3. The regulation of tomato fruit shape is controlled by increased transcription of *IQD12*. Data shown are in the LA1589 background. (**A**) Graphical depiction of the two genomic constructs used to complement *sun*. Features of the genomic region are described in Fig. 1. (**B**) Northern blot analysis of T1 plants transformed with pHX2 and pHX4 constructs with *IQD12* as probe with RNA isolated from fruit 5 days after anthesis. (**C**) Fruit-shape index of T1 lines significantly correlates with *IQD12* transcript levels. (**D**) Fruit-shape index and *IQD12* transcript levels of homozygous T2 plants derived from five independent pHX4 primary transformants. The average fruit-shape index (column) and *IQD12* transcript levels (line) were determined from the Northern blot data shown in (fig. S2B). Error bars denote SD. For comparison, fruit-shape index and *IQD12* transcript levels of the NIL were included. (**E**) Constitutive expression of *IQD12* in the round-fruited LA1589 background results in extremely elongated fruit. Each fruit represents an independently transformed plant. Fruit from the non-transformed round-fruited NIL (LA1589pp) is shown for comparison. (**F**) RNAi-mediated reduced expression of *IQD12* in the NIL carrying elongated fruit (LA1589ee) results in a notable reduction in fruit elongation. Each fruit represents an independently transformed plant. The fruit from the nontransformed NIL (LA1589ee) is shown for comparison. Scale bars in (E) and (F) represent 1 cm.

high levels of *DEFL1* expression in developing fruit.

The insert from two overlapping Sun1642 λ genomic clones was transformed into LA1589 and the round-fruited NIL in the Sun1642 background. These clones—pHX2 and pHX4—harbored the full-length *IQD12* gene, including the promoter shared with the ancestral gene and the LTR, but contained different 5′ as well as 3′ end points (Fig. 3A). In the LA1589 background, most plants transformed with the pHX4 construct expressed *IQD12* at high levels, whereas this gene was expressed at low levels in the pHX2 transformants (Fig. 3B and table S1). Moreover, pHX4 T1 plants exhibited a greater fruit-shape index (ratio of height to width), whereas those with the pHX2 construct did not. Regression analysis confirmed that there was a significant linear relationship between *IQD12* transcript levels and fruit-shape index, which was in turn correlated to transformation with the pHX4 construct (Fig. 3C). Transformation of the same constructs into the round-fruited NIL in the Sun1642 background produced similar results (table S2).

Increased expression of *IQD12* and increased fruit-shape index cosegregated with the presence of the pHX4 construct in the next generation (Fig. 3D and fig. S3A). Transcript levels of *IQD12* and fruit-shape index of homozygous pHX4 plants indicated that, for most transgenic plant families, the fruit-shape index and the *IQD12* transcript levels were nearly restored to the endogenous levels displayed by the NIL carrying the transposed segment (Fig. 3D, fig. S3, and table S1). Thus, control of fruit shape in tomato is regulated at the level of transcription of *IQD12*. In addition, important cis-element(s) in the 3.2-kb region upstream of *DEFL1* drive *IQD12* expression in the developing fruit, because this region was present in pHX4 and absent from pHX2 (Fig. 3, A to C). Additionally, the *SDL1-like* and *HYP1* genes probably do not affect fruit shape, because these genes were present on the pHX2 construct, which did not complement the phenotype (Fig. 3, A to C).

Although transcript levels and fruit-shape index were substantially increased in the transgenic lines carrying pHX4, they were not entirely restored to the levels of that in the NIL carrying the transposed segment (table S1). Thus, either additional cis-elements in the *DEFL1* 5′ region or other genes contribute to fruit-shape phenotype (Fig. 3D and fig. S3). To determine whether *IQD12* alone was sufficient to confer an elongated shape to tomato, we overexpressed this gene in LA1589 and used RNA interference (RNAi) to reduce expression in plants carrying the transposed segment exhibiting elongated fruit. Six of 13 LA1589 lines transformed with *IQD12* under control of the constitutive cauliflower mosaic virus (CaMV) 35S RNA promoter bore extremely elongated fruit and expressed *IQD12* at high levels (Fig. 3E, fig. S4A, and table S3), which cosegregated with the presence of the transgene in the next generation (fig. S4B and table S3). An *IQD12* inverted repeat construct—transformed into NILs carrying the transposed segment—resulted in rounder fruit similar to that found in the wild type, which was associated with decreased *IQD12* expression (Fig. 3F, fig. S4, C to E, and table S3). Thus, increased and reduced expression of *IQD12* confers changes to tomato fruit shape, which implies that this gene is necessary and sufficient in controlling shape at the *sun* locus. Hence, *IQD12* was renamed *SUN*.

All members of the IQD family contain the 67–amino acid IQ67 domain, which has only been found in plants (*5*). AtIQD1 binds calmodulin and is localized in the nucleus, and its overexpression increases glucosinolate levels in *Arabidopsis* (*9*). On the basis of the shared domain structure of the entire protein and phylogenetic analysis of the IQ67 motif, SUN is most closely related to AtIQD12, although the criterion of conservation of intron position suggests that it is more closely related to AtIQD11 (fig. S5A). Both *Arabidopsis* genes are members of the subfamily II of the IQD family of proteins that also include AtIQD13 and AtIQD14 (*5*) (fig. S5B). Extensive database searches identified members of this family from other dicot species and suggested that the IQD cluster containing SUN expanded to at least four members in tomato and potato, after the split from *Arabidopsis* (fig. S5B).

We propose that transcription of *Rider* on chromosome 10 began in LTR1. Instead of stopping in LTR2, readthrough occurred in the flanking genomic region (Fig. 4A, SOM Text, and fig. S7). In the first intron of the *SDL1-like* gene, the RNA polymerase switched to a region 3′ of *SUN*. The 5-bp direct repeat GCAGA on each side of the breakpoint suggests that a template switch occurred (Fig. 4A), resulting in an elongated transcript that terminated in LTR1 (Fig. 4B). The introns of the *SUN* and *SDL1-like* genes were retained in the antisense orientation with regard to *Rider* and were not recognized and removed by the splicing machinery after RNA synthesis. This large element inserted into chromosome 7 (Fig. 4C), resulting in the gene arrangement at the *sun* locus (Fig. 1A). *Rider*'s 3′ gene transduction and template switch are essential for a successful transposition, because two flanking LTRs are required for reverse transcription and double-stranded DNA synthesis of LTR retrotransposons (*10*). In addition, this model demonstrates that LTR retrotransposons can facilitate gene duplications, resulting in phenotypic change.

With the exception of AtIQD1 and SUN, the function of other IQ67 domain-containing proteins remains unknown. Unlike AtIQD1, SUN does not appear to be affecting glucosinolate levels in tomato, because these metabolites are not produced in solanaceous plants. However, AtIQD1 plays a role in the transcript accumulation of a small group of cytochrome P450 genes including

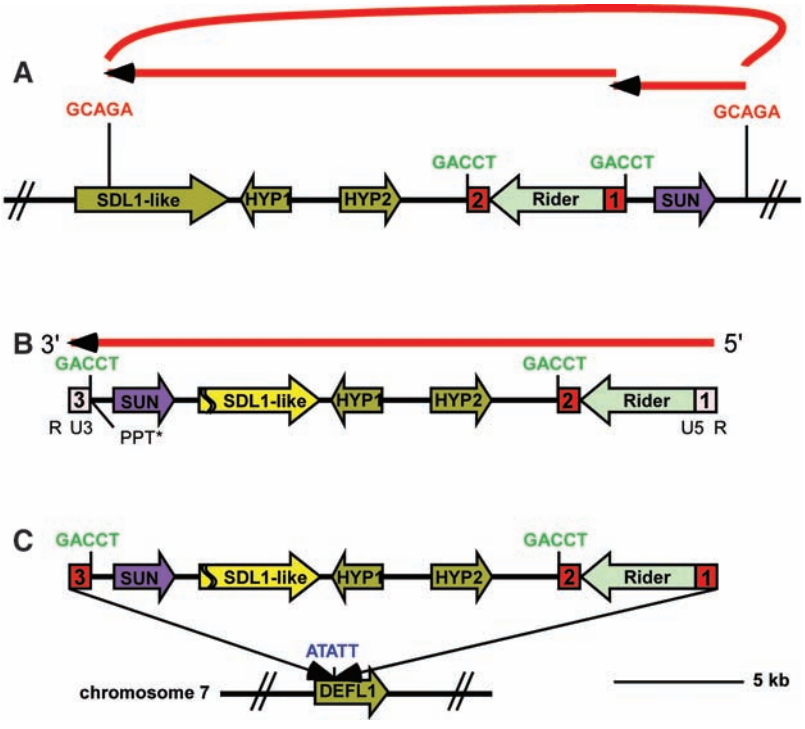

Fig. 4. A model for the segmental duplication and rearrangement at the *sun* locus. The features of the genomic region are described in Fig. 1. (**A**) Readthrough transcription of *Rider* on chromosome 10 to the signature sequence GCAGA in the first intron of *SDL1-like*. RNA polymerase switched the template to continue mRNA transcription from an identical signature sequence (GCAGA), located downstream of *SUN*. (**B**) Formation of one large retrotransposon mRNA. (**C**) After reverse transcription and second-strand cDNA synthesis, *Rider* integrated into chromosome 7 at the ATATT site located in the intron of *DEFL1*.

CYP79B3 and *CYP79B2* (*9*, *11*). The encoded proteins catalyze the conversion of tryptophan into indole-3-acetaldoxime in tryptophan-dependent auxin biosynthesis in *Arabidopsis* (*12*–*14*). In tomato, overexpression of *SUN* resulted in extremely elongated and often seedless fruit, reminiscent of the parthenocarpic, elongated, and pointed tomato fruit resulting from expression of the auxin biosynthesis gene *iaaM* controlled by the placenta and ovule-specific *DefH9* promoter (*15*, *16*). The extremely elongated fruit shape and lack of proper seed development when *SUN* is overexpressed, in addition to its potential biochemical function, suggest that SUN may affect auxin levels or distribution in the fruit (Fig. 3E). It is therefore plausible that involvement of SUN in shape variation is through regulation of plant hormone and/or secondary metabolite levels, thereby affecting the patterning of the fruit.

References and Notes

1. J. F. Doebley, B. S. Gaut, B. D. Smith, *Cell* **127**, 1309 (2006).
2. I. Paran, E. van der Knaap, *J. Exp. Bot.* **58**, 3841 (2007).
3. E. van der Knaap, S. D. Tanksley, *Theor. Appl. Genet.* **103**, 353 (2001).
4. E. van der Knaap, A. Sanyal, S. A. Jackson, S. D. Tanksley, *Genetics* **168**, 2127 (2004).
5. S. Abel, T. Savchenko, M. Levy, *BMC Evol. Biol.* **5**, 72 (2005).
6. K. Lertpiriyapong, Z. R. Sung, *Plant Mol. Biol.* **53**, 581 (2003).
7. A. Majira, M. Domin, O. Grandjean, K. Gofron, N. Houba-Herin, *Plant Mol. Biol.* **50**, 551 (2002).
8. S. Takada, K. Hibara, T. Ishida, M. Tasaka, *Development* **128**, 1127 (2001).
9. M. Levy, Q. Wang, R. Kaspi, M. P. Parrella, S. Abel, *Plant J.* **43**, 79 (2005).
10. B. Lewin, *Genes IX* (Jones and Bartlett, Sudbury, MA, ed. 9, 2008).
11. C. D. Grubb, S. Abel, *Trends Plant Sci.* **11**, 89 (2006).
12. A. K. Hull, R. Vij, J. L. Celenza, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2379 (2000).
13. M. D. Mikkelsen, C. H. Hansen, U. Wittstock, B. A. Halkier, *J. Biol. Chem.* **275**, 33712 (2000).
14. Y. Zhao *et al.*, *Genes Dev.* **16**, 3100 (2002).
15. N. Ficcadenti *et al.*, *Mol. Breed.* **5**, 463 (1999).
16. T. Pandolfini, G. L. Rotino, S. Camerini, R. Defez, A. Spena, *BMC Biotechnol.* **2**, 1 (2002).

17. We thank T. Meulia and A. Kaszas at the Molecular and Cellular Imaging Center in Wooster for microscopy and sequencing, as well as D. Francis, S. Kamoun, D. Mackey, and P. Springer for comments. This work was funded in part by NSF: DBI 0227541 and DBI 0400811 (to E.v.d.K.); DBI 0110124 (to E.J.S.). Accession numbers of deposited sequences: EF094939, EF094940, EF094941, and EU491503.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5869/1527/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Tables S1 to S4
References

15 November 2007; accepted 8 February 2008
10.1126/science.1153040

Evidence for Karyogamy and Exchange of Genetic Material in the Binucleate Intestinal Parasite *Giardia intestinalis*

Marianne K. Poxleitner,[1]* Meredith L. Carpenter,[1]* Joel J. Mancuso,[1] Chung-Ju R. Wang,[1] Scott C. Dawson,[2] W. Zacheus Cande[1]†

The diplomonad parasite *Giardia intestinalis* contains two functionally equivalent nuclei that are inherited independently during mitosis. Although presumed to be asexual, *Giardia* has low levels of allelic heterozygosity, indicating that the two nuclear genomes may exchange genetic material. Fluorescence in situ hybridization performed with probes to an episomal plasmid suggests that plasmids are transferred between nuclei in the cyst, and transmission electron micrographs demonstrate fusion between cyst nuclei. Green fluorescent protein fusions of giardial homologs of meiosis-specific genes localized to the nuclei of cysts, but not the vegetative trophozoite. These data suggest that the fusion of nuclei, or karyogamy, and subsequently somatic homologous recombination facilitated by the meiosis gene homologs, occur in the giardial cyst.

Giardia intestinalis (syn. *lamblia*, *duodenalis)* is a common intestinal protozoan parasite and a major cause of water-borne diarrheal disease worldwide (*1*). As a diplomonad, it has two apparently genetically identical, functionally equivalent diploid nuclei (*2*–*4*), and the sequenced genome shows little heterozygosity (*5*, *6*). Both nuclei are transcriptionally active and contain two complete copies of the genome (*2*–*4*). The parasite has two developmental stages: a binucleate, flagellated trophozoite that attaches to the upper small intestine, and the quadrinucleate infectious cyst, which is excreted from the body and persists in the water supply (*7*). In the encysting trophozoite, the two nuclei are thought to undergo a nuclear division without a subsequent cytokinesis, giving rise to a mature cyst with four nuclei (*8*). Upon excystation, the cell goes through one cellular division and then one nuclear and cellular division to form four binucleate trophozoites (*8*).

The two nuclei in the trophozoite remain physically and genetically distinct during mitosis, with two autonomous spindles segregating the parental nuclei into the daughter cells (*9*). Neither mating nor meiosis has been reported in *Giardia*. However, if *Giardia* is asexual, it should accumulate substantial allelic heterozygosity within and between the two nuclei, and this has not been observed (*5*, *6*, *10*).

To determine whether the cyst nuclei remain physically autonomous, as they do in trophozoites, or whether nuclei can exchange genetic material, we performed fluorescent in situ hybridization (FISH) on trophozoites and cysts (*9*). In *Giardia*, stably transfected plasmids (episomes) are found in only one of the two nuclei of trophozoites (Fig. 1A). This pattern persists throughout cell division and cytokinesis (*9*, *11*). When we use FISH to detect episomes in cysts, we see two distinct patterns ($n = 50$): 72% of the cysts have several episomes present in two of the four nuclei (Fig. 1B), whereas 28% of the cysts have episomes in three of the four nuclei (Fig. 1C). This result suggests that in roughly one-third of encysting cells, two nuclei come together and their nuclear envelopes fuse, resulting in plasmid transfer between the nuclei.

To ascertain whether nuclear fusion occurs in *Giardia* cysts, we performed transmission electron microscopy (TEM). Cysts were prepared for TEM with high-pressure rapid freezing–freeze substitution to achieve optimal cytological preservation and eliminate chemical fixation artifacts due to the impermeability of the cyst wall. Using this method, we observed nuclei joined by contiguous membranes (Fig. 1, D and E). These nuclear envelope (NE) fusion morphologies are not derived from an incomplete mitosis because there are no spindle microtubules associated with the NE, and the multiple axonemes associated with paired and fusing nuclei differ in location from the axoneme pairs found at each spindle pole in mitotic cells (*9*). Of the ~100 individual cysts examined, we observed 15 with contiguous NEs. Because these cells were collected at a single moment in time, the percentage of nuclear fusions is less than that seen by FISH, which represents the net outcome of nuclear fusions over time. Thus, the nuclei in the *Giardia* cyst fuse (undergo karyogamy), and this fusion likely facilitates the transfer of episomes between nuclei.

We next sought to determine when karyogamy takes place in the encystation process. We counted the number of nuclei in cysts and the total number of cysts present at six different time points (Fig. 2A). Cysts appeared 12 hours after induction of encystation; subsequently, multiple nuclear arrangements were seen in cysts at every time point, suggesting that encystation is not well synchronized and that cells at varying stages of

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. [2]Section of Microbiology, 255 Briggs Hall, One Shields Avenue, University of California, Davis, CA 95616, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: zcande@berkeley.edu