DOI: 10.1534/genetics.107.078709

Note

Evolutionary History and Positional Shift of a Rice Centromere

Jianxin Ma,* Rod A. Wing,† Jeffrey L. Bennetzen‡ and Scott A. Jackson*,[1]

**Department of Agronomy, Purdue University, West Lafayette, Indiana 47907, †Arizona Genomics Institute, University of Arizona, Tucson, Arizona 85721 and ‡Department of Genetics, University of Georgia, Athens, Georgia 30602*

Manuscript received July 11, 2007
Accepted for publication July 17, 2007

ABSTRACT

Rice centromere 8 was previously proposed to be an "immature" centromere that recently arose from a genic region. Our comparative genomics analysis indicates that *Cen8* was formed at its current location at least 7–9 million years ago and was physically shifted by a more recent inversion of a segment spanning centromeric and pericentromeric regions.

IN-DEPTH sequence analysis of the centromeric region of rice (*Oryza sativa*) chromosome 8 (*Cen8*) has provided valuable insights into the structure, organization, and evolutionary dynamics of a complex centromere, the first fully sequenced from any plant or animal species (Nagaki *et al.* 2004; Wu *et al.* 2004; Ma and Bennetzen 2006). One of the most intriguing observations in this region was the presence of at least four active genes in the ~750-kb CENH3-binding domain (Nagaki *et al.* 2004), the region that provides centromere segregation functions via its kinetochore association. The *Cen8* region also contains the fewest copies of the centromeric satellite, CentO, of all rice centromeres (Cheng *et al.* 2002). Also, the LTR retrotransposons identified in the CENH3-binding domain, including the centromeric retrotransposons of rice (Cheng *et al.* 2002), are similar in age to those elements located adjacent to this region (Nagaki *et al.* 2004; Ma and Bennetzen 2006). These results led Nagaki *et al.* (2004) to propose that *Cen8* may represent an intermediate stage in the evolution from *de novo* centromere formation at genomic regions, as in human neocentromeres, to fully mature centromeres that accumulate megabases of satellite arrays (Nagaki *et al.* 2004). However, because parts of the *Cen8* region have been rearranged and reshuffled dramatically (Ma and Bennetzen 2006; Ma *et al.* 2007), along with the rapid elimination of LTR retrotransposons in the region and the rest of the rice genome (Ma and Bennetzen 2004, 2006), the evolutionary status of *Cen8* may not be simply interpreted from its structural features. Hence, the formation time of this centromere remains to be elucidated.

Taking advantage of the physical map and BAC end sequence (BES) data generated by the *Oryza* Map Alignment Project (Wing *et al.* 2005; Ammiraju *et al.* 2006; http://www.omap.org), a comparative genomics approach to tracking the evolutionary history of rice *Cen8* has been developed by anchoring unique exonic portions of predicted genes embedded or surrounding rice *Cen8* to fingerprint contigs (FPCs) or BACs of wild Oryza species. Initially, three probes (P1, P2, P3) amplified from the coding regions of three single-copy genes identified in the *Cen8* region (Wu *et al.* 2004; International Rice Genome Sequencing Project 2005) (Figure 1) were used to screen the BAC library of *Oryza brachyantha*, a wild Oryza species that last shared an ancestor with rice ~7–9 MYA (Ge *et al.* 1999; Dawe 2005). Because the *O. brachyantha* BAC library represents a 14-fold genome coverage with an average clone-insert size of 131 kb (Ammiraju *et al.* 2006), it was expected that single BAC clones containing both P1 and P2, which are 86 kb apart in the *Cen8* region of rice, would be found in this library. Intriguingly, *O. brachyantha* BAC clones containing both P1 and P2 were not found. Instead, one BAC clone was found to contain both P2 and P3, which are 902 kb apart in the *Cen8* region of rice. Given that most LTR retrotransposons in the *Cen8* region of rice accumulated quite recently (Nagaki *et al.* 2004; Ma and Bennetzen 2006), it is possible, for example, that the *Cen8* region of rice and its orthologous region of *O. brachyantha* expanded or contracted differentially (Bruggmann *et al.* 2006), leading to tremendous variation of P2 and P3 intervals between the two species. Alternatively, a major

[1]*Corresponding author:* Department of Agronomy, Purdue University, 915 W. State St., West Lafayette, IN 47907.
E-mail: sjackson@purdue.edu

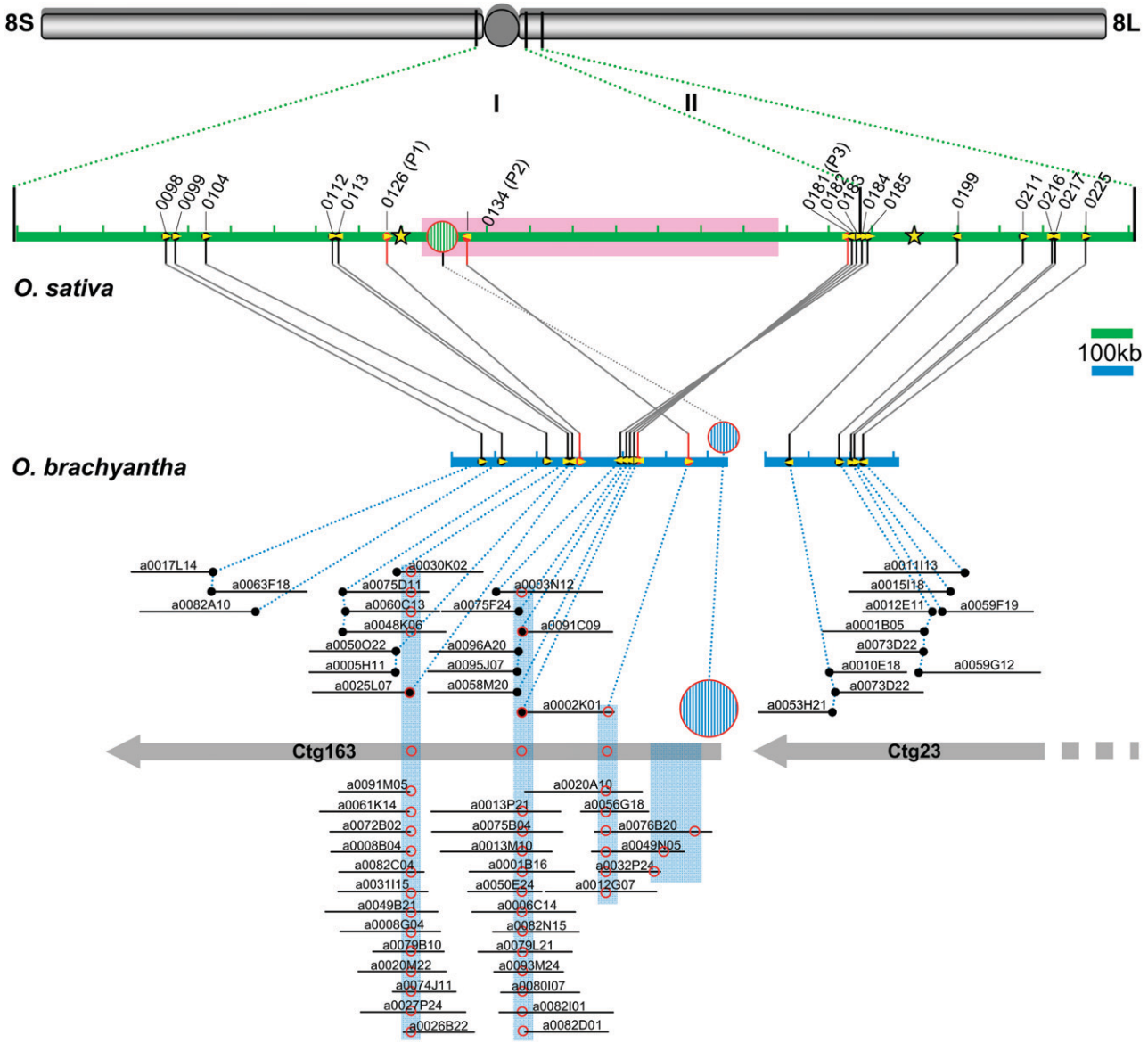

FIGURE 1.—Comparative physical maps of *Cen8* orthologous regions. The orthologous regions of *O. brachyantha* and rice (*O. sativa*) were identified on the basis of gene–BES alignments, hybridization anchors, and FPC maps. The green bar represents the *O. sativa* sequence and the blue bars represent the FPC maps of *O. brachyantha* orthologous to the *O. sativa* region. The scales of the green bar and the blue bars are equal. Dots on the green and blue bars represent orthologous genes and their orientations, and these orthologous genes are connected by gray lines. The gray bars with arrows show the FPC contigs at twice the scale and the orientation of the contigs (the dashed bar in Ctg23 indicates the region not included in the analysis). The horizontal lines above and below the gray bar represent *O. brachyantha* BAC clones in the FPC contigs with one or both ends (solid circle) matching the unique exonic portions of predicted genes in *O. sativa* (INTERNATIONAL RICE GENOME SEQUENCING PROJECT 2005) and BAC clones containing the hybridization anchors P1, P2, and P3, which were amplified from genes 0126, 0134, and 0181, respectively, by polymerase chain reaction (red circles). Circles outlined in red with vertical lines represent proposed *O. brachyantha* centromere satellite arrays and their physical locations detected by screening the *O. brachyantha* library using CentO_F (LEE *et al.* 2005) consensus sequence as probe, which was synthesized as two overlapping oligos. The pink region indicates the CENH3-binding domain (NAGAKI *et al.* 2004). Zone "I" and "II" represent rice *Cen8* (WU *et al.* 2004) and its adjacent pericentromeric region (INTERNATIONAL RICE GENOME SEQUENCING PROJECT 2005), respectively. Stars indicate the proposed breakpoints for the segmental inversion. Gene–BES alignments were conducted by CROSS_MATCH and BLAST. Probe labeling and Southern hybridization were performed as described earlier (HASS-JACOBUS *et al.* 2006).

chromosomal rearrangement may have occurred in the target region of either rice or *O. brachyantha*.

To identify the basis of the dramatic physical linkage variation observed, 71 single-copy genes interspersed in an ~2.6-Mb region (11,967,606–11,459,841 bp of chromosome 8) containing *Cen8* of rice were chosen from 158 genes (0016–0233, *i.e.*, 08_01_0076–08_02_0233 annotated by INTERNATIONAL RICE GENOME SEQUENCING

Project 2005) for further analysis. The unique exonic portions of these 71 genes were identified by comparison with the complete genomic sequence of rice (International Rice Genome Sequencing Project 2005) and were subsequently searched against 67,364 BESs generated by end sequencing 36,414 *O. brachyantha* clones (Wing *et al.* 2005; Ammiraju *et al.* 2006). Combining the data obtained from gene–BES sequence alignments and Southern hybridization experiments with the three gene probes (supplemental Table 1 at http://www.genetics.org/supplemental/) and the FPC maps of *O. brachyantha* (Wing *et al.* 2005; Ammiraju *et al.* 2006; http://www.omap.org), two *O. brachyantha* contigs, Ctg163 and Ctg23, representing two segments orthologous to the *Cen8* region of rice, were identified (Figure 1). On the basis of the order and orientations of genes aligned between these two species, an inversion of a segment containing P2 and P3 was discovered (Figure 1). In addition, 8 of 92,160 BAC clones of *Oryza officinalis* (a wild Oryza species more recently diverged from rice than *O. brachyantha*) containing P1 and P3 but not P2, and four containing P2 and P3 but not P1, were identified by Southern hybridization analysis. This result indicates that the genes P1, P2, and P3 are arranged in the order of P1—P3—P2 or P2—P3—P1 in *O. officinalis* (supplemental Table 2 at http://www.genetics.org/supplemental/), which is different from the order in rice, but most likely the same as in *O. brachyantha* (P1—P3—P2). Hence, the inversion appears to have occurred in rice after its more recent divergence from *O. officinalis* (Ge *et al.* 1999). This inversion spans both centromeric and pericentromeric regions, accounting for ~1 Mb of DNA in rice.

It is particularly interesting that CentO_F, the centromeric satellite DNA present only in centromeric regions of *O. brachyantha* on the basis of fluorescent *in situ* hybridization analysis (Lee *et al.* 2005), was found on three BAC clones assembled at or near the end of Ctg163 adjacent to Ctg23 (Figure 1). This suggests that the *Cen8*-orthologous region identified in *O. brachyantha* is also a centromeric region. This hypothesis is bolstered by the observation that CentO and CentO-F are located in orthologous positions in rice and *O. brachyantha*, although they are shifted physically by the inversion event (Figure 1). Together, these observations suggest that, despite its proposed "neocentromeric features" (Nagaki *et al.* 2004), rice *Cen8* formed at least before the divergence of rice and *O. brachyantha* from a common ancestor 7–9 MYA (Dawe 2005), followed by a more recent inversion event. Recent studies suggest that *O. brachyantha* is the species within the genus Oryza that is most diverged from *O. sativa* (R. A. Wing and S. A. Jackson, personal communication). Hence, rice *Cen8* may have been formed before the divergence of all Oryza species identified thus far (Ge *et al.* 1999). The identified inversion event reshaped the structure of the *Cen8* region, but whether it is responsible for the presence of "neocentromeric features" remains an intriguing question.

A hemicentric inversion of larger chromosomal segment with two breakpoints in the original centromere position and ~20 centiMcClintocks (cMC) on the long arm of maize chromosome 8 was identified in the maize line, knobless Tama flint (KTF) (Lamb *et al.* 2007). This inversion moved the site of the kinetochore-forming region, representing a molecular mechanism for the formation of neocentromeres. However, the new centromere in KTF may not show the proposed "neocentromeric features" (Nagaki *et al.* 2004) (*e.g.*, the presence of active genes), although it contains fewer copies of the centromere satellite repeats than the original centromere location (Lamb *et al.* 2007). Alternatively, these "features" may not be atypical of a mature centromere, as reflected by an additional observation that the copy numbers of centromeric satellites vary to extreme degrees among homologous chromosomes of different maize lines (Kato *et al.* 2004).

It appears that the *Cen8* orthologous regions have captured much more LTR retotransposon DNA in rice than in *O. brachyantha*, whose nuclear genome size is ~330 Mb (Ammiraju *et al.* 2006), slightly smaller than that of rice (389 Mb; International Rice Genome Sequencing Project 2005). For example, in the three comparable subregions, the distances between genes 0098 and 0126, genes 0134 and 0185, and genes 0199 and 0225 in rice are 515, 944, and 314 kb, respectively. In contrast, according to the FPC maps (Wing *et al.* 2005; Ammiraju *et al.* 2006; http://www.omap.org), the corresponding subregions in *O. brachyantha* are 248, 139, and 179 kb, respectively. The combination of the three intergenic regions accounts for 1773 kb in rice *vs.* 566 kb in *O. brachyantha* (Figure 1). This observation parallels the previous finding that LTR retrotransposons make up an exceptionally small portion of *O. brachyantha* centromeres (Lee *et al.* 2005). Differential expansion of orthologous pericentromeric regions of related Brassicaceae species has been previously described (Hall *et al.* 2006). Differences in the activity of mechanisms for LTR-retrotransposon regulation (Bennetzen *et al.* 2005) and DNA rearrangements, *e.g.*, segmental duplication as found in the *Cen8* and *Cen4* regions (Ma and Bennetzen 2006; Ma and Jackson 2006; Ma *et al.* 2007), could partially explain the rapid and dramatic size variation between these regions.

In summary, this study addresses the evolutionary history and dynamics of a rice centromere and provides the first molecular description of the positional shift of any higher eukaryotic centromere caused by a small chromosomal inversion. This study also demonstrates the value of physical mapping with BAC contigs for comparative and evolutionary analysis of complex genomic regions recalcitrant to other analytical approaches (*e.g.*, sequencing and assembly of repetitive DNA).

This work was partially supported by the National Science Foundation Plant Genome Research Program (grant nos. DBI-0227414 to S.A.J., DBI-0321678 to R.A.W. and S.A.J., and DBI-0501814 to J.L.B.) and Purdue University new faculty startup funds to J.M.

LITERATURE CITED

Ammiraju, J. S., M. Luo, J. L. Goicoechea, W. Wang, D. Kudrna *et al.*, 2006 The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. Genome Res. **16:** 140–147.

Bennetzen, J. L., J. Ma and K. M. Devos, 2005 Mechanisms of recent genome size variation in flowering plants. Ann. Bot. **95:** 127–132.

Bruggmann, R., A. K. Bharti, H. Gundlach, J. Lai, S. Young *et al.*, 2006 Uneven chromosome contraction and expansion in the maize genome. Genome Res. **16:** 1241–1251.

Cheng, Z., F. Dong, T. Langdon, S. Ouyang, C. R. Buell *et al.*, 2002 Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell **14:** 1691–1704.

Dawe, R. K., 2005 Centromere renewal and replacement in the plant kingdom. Proc. Natl. Acad. Sci. USA **102:** 11573–11574.

Ge, S., T. Sang, B. R. Lu and D. Y. Hong, 1999 Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. USA **96:** 14400–14405.

Hall, A. E., G. C. Kettler and D. Preuss, 2006 Dynamic evolution at pericentromeres. Genome Res. **16:** 355–364.

Hass-Jacobus, B. L., M. Futrell-Griggs, B. Abernathy, R. Westerman, J. L. Goicoechea *et al.*, 2006 Integration of hybridization-based markers (overgos) into physical maps for comparative and evolutionary explorations in the genus *Oryza* and in Sorghum. BMC Genomics **7:** 199.

International Rice Genome Sequencing Project, 2005 The map-based sequence of the rice genome. Nature **436:** 793–800.

Kato, A., J. C. Lamb and J. A. Birchler, 2004 Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc. Natl. Acad. Sci. USA **101:** 13554–13559.

Lamb, J. C., J. M. Meyer and J. A. Birchler, 2007 A hemicentric inversion in the maize line knobless Tama flint created two sets of centromeric elements and moved the kinetochore-forming region. Chromosoma **116:** 237–247.

Lee, H. R., W. Zhang, T. Langdon, W. Jin, H. Yan *et al.*, 2005 Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. Proc. Natl. Acad. Sci. USA **102:** 11793–11798.

Ma, J., and J. L. Bennetzen, 2004 Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. USA **101:** 12404–12410.

Ma, J., and J. L. Bennetzen, 2006 Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice. Proc. Natl. Acad. Sci. USA **103:** 383–388.

Ma, J., and S. A. Jackson, 2006 Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. Genome Res. **16:** 251–259.

Ma, J., R. A. Wing, J. L. Bennetzen and S. A. Jackson, 2007 Plant centromere organization: a dynamic structure with conserved functions. Trends Genet. **23:** 134–139.

Nagaki, K., Z. Cheng, S. Ouyang, P. B. Talbert, M. Kim *et al.*, 2004 Sequencing of a rice centromere uncovers active genes. Nat. Genet. **36:** 138–145.

Wing, R. A., J. S. Ammiraju, M. Luo, H. Kim, Y. Yu *et al.*, 2005 The Oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species. Plant Mol. Biol. **59:** 53–62.

Wu, J., H. Yamagata, M. Hayashi-Tsugane, S. Hijishita, M. Fujisawa *et al.*, 2004 Composition and structure of the centromeric region of rice chromosome 8. Plant Cell **16:** 967–976.

Communicating editor: J. A. Birchler