Genetics: Published Articles Ahead of Print, published on July 29, 2007 as 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[*]

[*]Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA

[†]Arizona Genomics Institute, University of Arizona, Tucson, AZ 85721, USA

[‡]Department of Genetics, University of Georgia, Athens, GA 30602, USA

Running title: rice centromere evolution

Key words: centromere evolution, chromosomal rearrangement, comparative genomics, rice, wild rice.

Corresponding author:

Scott A. Jackson

Department of Agronomy

Purdue University

915 West State Street

West Lafayette, IN 47907

Tel: 1-765-496-3621

Fax: 1-765-496-7255

E-mail: sjackson@purdue.edu

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 recently 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 the rice centromeres (CHENG *et al.* 2002). Also, the LTR-retrotransposons identified in the CENH3 binding domain, including the centromeric retrotransposons of rice (CRRs, 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 and coworkers 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 and rapidly

(MA and BENNETZEN 2006; MA *et al.* 2007), plus rapid elimination and turnover of LTR-retrotransposons in the region and in the rice genome (MA *et al.* 2004; MA and BENNETZEN 2006), the evolutionary status of *Cen8* may not be reflected by its structural features. Hence, the formation time of this centromere remain to be elucidated.

Taking advantage of the physical map and BAC end sequence (BES) data generated by the *Oryza* Map Alignment Project (OMAP) (WING *et al.* 2005; AMMIRAJU *et al.* 2006; http://www.omap.org), a comparative genomics approach to track the evolutionary history of rice *Cen8* has been developed, by anchoring unique exonic portions of predicted genes embedded or surrounding rice *Cen8* to FPCs (fingerprint contigs) 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; IRGSP 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 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* regions of rice accumulated quite recently (NAGAKI *et al.* 2004; MA and BENNETZEN 2006), it would be possible, for instance, 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 two species. Alternatively, a major 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, seventy-one single-copy genes interspersed in an ~2.6-Mb region (11967606bp – 11459841bp of chromosome 8) containing *Cen8* of rice were screened and chosen from 158 genes (0016-0233, i.e., 08_01_0076 – 08_02_0233 annotated by IRGSP (2005), The unique exonic portions of these 71 genes were identified by comparison with the complete genomic sequence of rice (IRGSP 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 genes-BES sequence alignments and Southern hybridization experiments with the three gene probes (Supplemental Table 1), 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). Based on the order and orientations of genes aligned between these two species, an inversion of a segment containing P2 and P3 was identified (Figure 1). In addition, eight out 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, indicating 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), different from 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 FISH (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* also be 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*, though 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* be a species within the genus *Oryza* that is most diverged from *O. sativa* (WING RA and JACKSON SA, pers. comm.), hence, rice *Cen8* may have been formed before the divergence of all *Oryza* species identified thus far (GE *et al.* 1999). It is likely that the identified inversion event reshaped the structure of the *Cen8* region, but whether it is responsible for the presence of the "neocentromeric features" remains an intriguing question.

A hemicentric inversion of larger chromosomal segment with two breakpoints in the original centromere position and ~20 centi-McClintocks (cMc) on the long arm of maize chromosome 8, were identified in a maize line, knobless Tama flint (KTF) (LAMB *et al.* 2007), which moved the site of the kinetochore-forming region, representing a molecular

mechanism for the formation of neocentromeres. However, this 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 less copies of the centromere satellite repeats than the original centromere location (LAMB *et al.* 2007). Alternatively, these "features" may not be atypical to a mature centromere, as reflected by an additional observation that the copy numbers of centromeric satellites vary extremely among homologous chromosomes of different maize lines (KATO *et al.* 2004).

It appears that the *Cen8* orthologous regions have captured much more 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, IRGSP 2005). The three comparable subregions, for instance, the distances between genes 0098 and 0126, genes 0134 and 0185, and genes 0199 and 0225 in rice are 515 kb, 944 kb, 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 kb, 139 kb, and 179 kb, respectively. The combination of the three intergenic regions account for 1773 kb in rice versus 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, 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 the Purdue University new faculty startup funds to J.M.

SUPPLEMENTAL MATERIAL

Supplementary tables are available at *Genetics* online (http://genetics.org).

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. (Lond.) **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., KETTLER, G. C., 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. Cheomosoma **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., P. NEUMANN, D. ZHANG, S. OUYANG, C. R. BUELL *et al.*, 2005 Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. Mol. Biol. Evol. **22:** 845-855.

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.

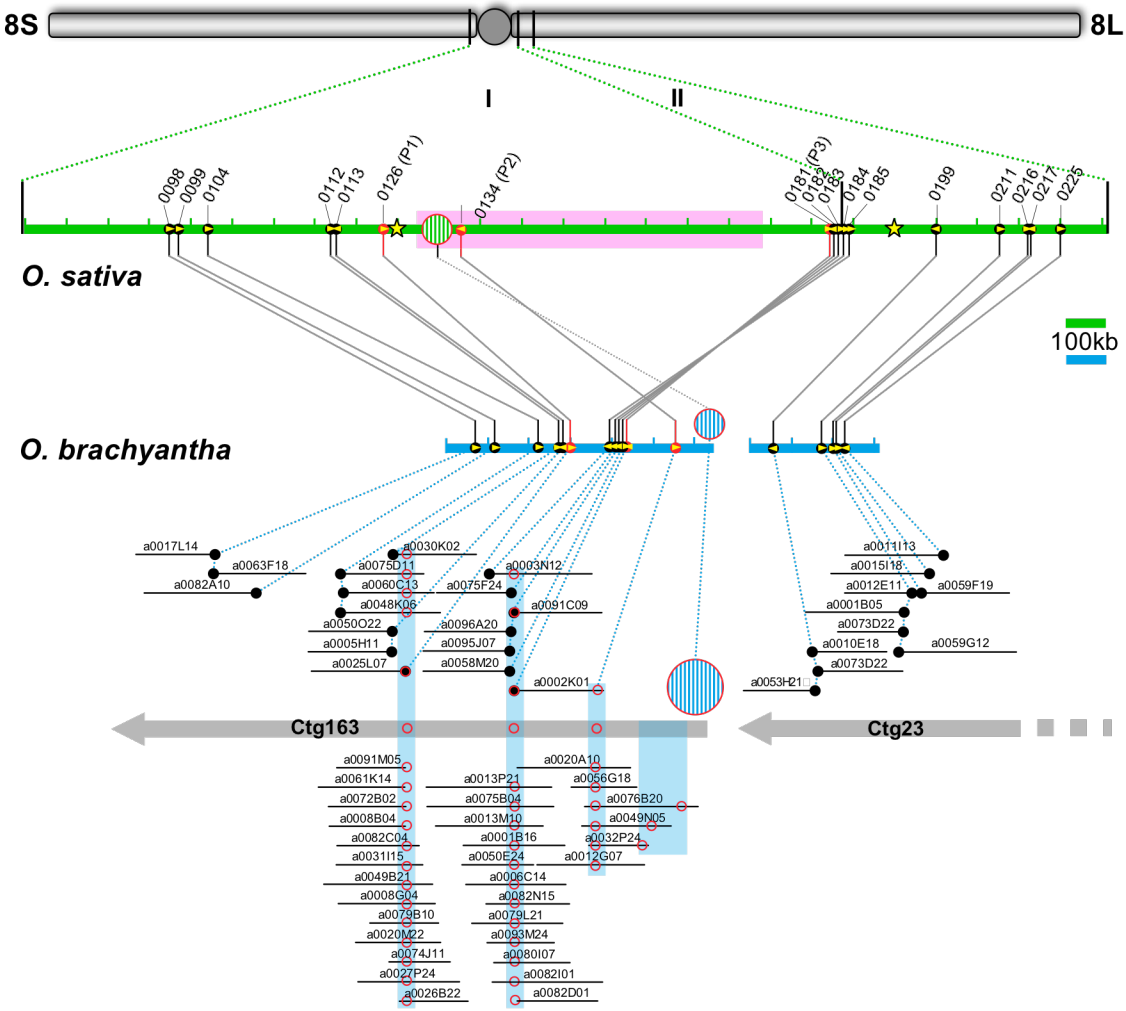
WU, J., H. YAMAGATA, M. HAYASHI-TSUGANE, S. HIJISHITA, M. FUJISAWA *et al.*, 2004 Composition and structure of the centromeric region of rice chromosome 8. Plant Cell **16:** 967-976.

FIGURE LEGEND:

FIGURE 1.—Comparative physical maps of *Cen8* orthologous regions. The orthologous regions of *O. brachyantha* and rice (*O. sativa*) were identified based on 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 bar are equal. Dots on the green and blue bars represent orthologous genes and their orientations, and these orthologous genes are connected by grey lines. The grey bars with arrows show the FPC contigs at 2× scale and the orientation of the contigs (the dashed area 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 (filled circle) matching the unique exonic portions of predicted genes in *O. sativa* (IRGSP 2005), and BAC clones containing hybridization anchors, P1, P2 and P3, which were amplified from genes 0126, 0134, and 0181, respectively, by polymerase chain reaction (red circles). Red circles with filled vertical lines represent proposed *O. brachyantha* centromere satellite arrays and their physical locations detected by screening *O. brachyantha* library using CentO_F (Lee *et al.* 2005) consensus sequence as probe, which was synthesized as two overlapping oligos. The pink shadowed 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 (IRGSP, 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).

FIGURE 1.


8S
8L
I
II
0098
0099
0104
0112
0113
0126 (P1)
0134 (P2)
0181 (P3)
0183
0184
0185
0199
0211
0216
0217
0225
O. sativa
100kb
O. brachyantha
a0017L14
a0063F18
a0082A10
a0030K02
a0075D11
a0060C13
a0048K06
a0050O22
a0005H11
a0025L07
a0003N12
a0075F24
a0091C09
a0096A20
a0095J07
a0058M20
a0002K01
a0011I13
a0015I18
a0012E11
a0059F19
a0001B05
a0073D22
a0010E18
a0059G12
a0073D22
a0053H21
Ctg163
Ctg23
a0091M05
a0061K14
a0072B02
a0008B04
a0082C04
a0031I15
a0049B21
a0008G04
a0079B10
a0020M22
a0074J11
a0027P24
a0026B22
a0013P21
a0075B04
a0013M10
a0001B16
a0050E24
a0006C14
a0082N15
a0079L21
a0093M24
a0080I07
a0082I01
a0082D01
a0020A10
a0056G18
a0076B20
a0049N05
a0032P24
a0012G07