

REVIEW

Open Access



Rice: The First Crop Genome

Scott A. Jackson

Abstract

Rice was the first sequenced crop genome, paving the way for the sequencing of additional and more complicated crop genomes. The impact that the genome sequence made on rice genetics and breeding research was immediate, as evidence by citations and DNA marker use. The impact on other crop genomes was evident too, particularly for those within the grass family. As we celebrate 10 years since the completion of the rice genome sequence, we look forward to new empowering tool sets that will further revolutionize research in rice genetics and breeding and result in varieties that will continue to feed a growing population.

Keywords: Rice, Genome Sequencing, Translation, Gene Mapping

Introduction and Review: Rice as a Model Genome The Rice Genome

The rice genome was one of the few truly multinational plant genome projects, one with flags planted in chromosomes (Eckardt 2000)—Japan, chromosomes 1, 6, 7 & 8; US, chromosomes 3 & 10; China, chromosome 4; France, chromosome 12; Taiwan, chromosome 5; etc.—like early explorers claiming new territories. Arabidopsis had been sequenced and published in 2000 (Arabidopsis Genome I 2000) and rice, nearly four times the size, was next. These were heady days for plant genome researchers, a discipline still in its infancy. Researchers jockeyed to be part of the project and, like every good collaborative project, multitudinous meetings were held to plan, to execute and, finally, to celebrate the rice genome project. These meetings were often tense—deciding how chromosomes were distributed, discussing progress, or lack thereof—and the omnipresent funding agency representatives hovered about ensuring that their agencies funds were well spent, the most diminutive of whom could elicit fear in the most seasoned genome researcher.

Rice was one of the last clone-by-clone, Sanger-sequenced genomes. That is, BAC/PAC clones were sequentially selected for sequencing, independently assembled and then stitched together to form pseudochromosomes. One impact that the rice genome had, even in the midst of sequencing, was the first completely

sequenced, complex eukaryotic centromere on chromosome 8 (Nagaki et al. 2004), which are usually gaps in genome sequences as they are highly repetitive regions of the genome. This was possible due to its diminutive size, ~64 kbp of satellite repeats, even compared to even Arabidopsis centromeres. Thus, rice has been a model for studies of centromere structure and function. This was possible due to cytogenetic analysis of rice showing the diminutive centromere 8. Interestingly, if the US had heeded cytogenetic descriptions they may not have chosen chromosome 9 as a sequencing target as one arm is highly heterochromatic, full of repetitive DNA sequences, which complicated physical mapping and sequencing of that chromosome arm.

The public rice genome, which took advantage of whole genome shotgun sequenced genomes made available from Monsanto in 2000 and Syngenta in 2002 (Goff et al. 2002), was published in 2006 (International Rice Genome Sequencing P 2005) after which the arduous task of annotation took place. Rice was one of the few genomes to have competing annotations (Ohyanagi et al. 2006; Ouyang et al. 2007), which created some confusion in the community but was eventually resolved with a single unified annotation (Kawahara et al. 2013).

Rice as a Model Cereal Genome

One of the initial motivators for sequencing rice, besides the relatively small genome size, was that it could be used as a model for other cereal crops with larger genomes, such as maize and wheat. This was predicated somewhat on rice's small genome size and the realization

Correspondence: sjackson@uga.edu
Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30621, USA

from molecular mapping, e.g. RFLPs, of conserved markers and marker order. Early work had shown conserved synteny between sorghum and maize (Hulbert et al. 1990) and subsequent work had shown conserved synteny over even longer evolutionary timeframes leading to the model of concentric circle genomes based on conserved marker order among a number grasses (Moore et al. 1995).

Even though considerable genetic synteny was found in comparisons of other cereals to rice, sequencing of orthologous genomic regions uncovered considerable variation, in gene content, repeat structure and size. One of the first comparisons of rice, maize and sorghum, preceding the rice genome sequencing project, revealed extensive size variation, especially inflated in maize, and in conservation of low copy sequences among the three species (Chen et al. 1997). Subsequent studies have shown that while rice is generally predictive of gene content in related cereal genomes, many genes are missed as either absent in rice, or present in rice and absent in the comparator. For instance in a rice-barley comparison around a disease resistance locus, barley had six additional genes not found in the rice orthologous region (Moore et al. 1995). Thus, while useful for leveraging genetics in other cereal species, additional crop genome sequences were required to fully address the genetic underpinning of traits in those crops.

Contribution to Advancement of Rice Genetics and Improvement

Gene and QTL Cloning

Gene cloning and especially cloning genes underlying QTL can provide a deeper molecular understanding of a trait and in the case of breeding, markers linked directly with the causative DNA mutation/changes. In rice, from

1990 to 1994 there were 20–30 papers per year retrieved by using ‘gene cloning’ as a search term (Fig. 1). In 1995, the first rice BAC library was published (Wang et al. 1995) which was eventually used for cloning of the Xa-21 resistance gene (Song et al. 1995). Between 1995 and 2001, as more cloning tools became available, the number per year ranged in the 40–80. After the genome sequence, the number jumped to more than 100 per year and has been fairly stable since. Even though this analysis is very rough, no manual curation of the extracted references, just a count of returned citations, it seems obvious that the genomic resources of large insert DNA clones and eventually the genome sequence greatly advanced gene/QTL cloning in rice. An obvious example of where the genome facilitated cloning of a gene underlying a QTL is *SUB1*, the gene that confers submergence tolerance (Xu et al. 2006). Another outgrowth of the reference genome is the exploration of diversity at a genome scale across the entire *Oryza* genus (Ammiraju et al. 2008; Lu et al. 2009; Sanyal et al. 2010; Jacquemin et al. 2014), including exploration within relatives closely related to cultivated rice (AA genome species) (Wang et al. 2014; Zhang et al. 2014). Many of these explorations depended on resources that arose from the International Oryza Map Alignment Project (Wing et al. 2005; Ammiraju et al. 2006).

Contribution of the Rice Genome to Improvement

The bigger goal for the community is rice improvement. On several fronts the rice reference genome has greatly advanced rice improvement. First, the most immediate impact was in molecular markers in that the number was greatly increased, their physical order was understood and proximity to annotated genes was useful to predict gene-trait associations. In part, this has been

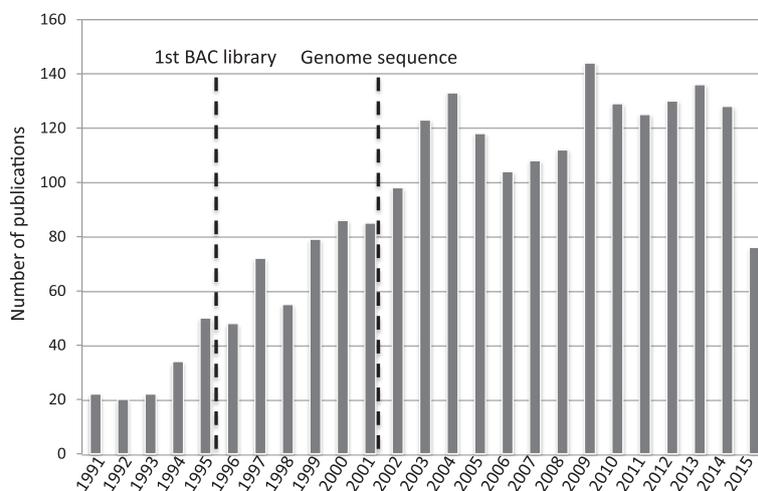


Fig. 1 Publications related to gene cloning in rice. ISI Web of Science (26-August-2015): search ‘rice’ and ‘gene cloning’ 1990–2015, retrieved 2,237 records

driven by reduction in DNA sequencing costs that have allowed researchers to resequence additional rice accessions and call nucleotide variations relative to the reference genome (McNally et al. 2009). Second, and related to the first, in-depth, sequence-based analysis of variation in cultivated and wild rice to allow breeders to better understand and exploit genetic variation, as proposed by McCouch and colleagues (McCouch et al. 2012). Third, molecular understanding of the genetic basis of traits such as N and P-use is allowing rice researchers to engineer 'Green Super Rice' that should help meet the challenge of the growing world population while requiring fewer inputs (Zhang 2007).

Conclusions

The availability of the rice genome, together with the community annotation and other resources that added functionality, transformed genetics research and rice breeding. This can be measured any number of ways, but the increase in gene/QTL cloning is one direct measure. Ongoing work to understand the evolution and domestication of rice and to use this and information on the genetic architecture underlying plant physiological responses and phenotypes is now be used to engineer superior rice strains to feed our growing population.

Abbreviations

BAC: bacterial artificial chromosome; bp: base pair; QTL: quantitative trait loci; RFLP: restriction fragment length polymorphisms.

Competing Interests

The author declares that he has no competing interest.

Author's Contributions

Written by SAJ.

Acknowledgements

None.

Received: 17 November 2015 Accepted: 16 March 2016

Published online: 22 March 2016

References

- Ammiraju JS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim H, Sisneros NB, Blackmon B 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(11):140–147
- Ammiraju JS, Lu F, Sanyal A, Yu Y, Song X, Jiang N, Pontaroli AC, Rambo T, Currie J, Collura K et al (2008) Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. *Plant Cell* 20(12):3191–3209
- Arabidopsis Genome I (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408(6814):796–815
- Chen M, SanMiguel P, de Oliveira AC, Woo SS, Zhang H, Wing RA, Bennetzen JL (1997) Microcolinearity in sh2-homologous regions of the maize, rice, and sorghum genomes. *Proc Natl Acad Sci U S A* 94(7):3431–3435
- Eckardt NA (2000) Sequencing the rice genome. *Plant Cell* 12(11):2011–2017
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296(5565):92–100
- Hulbert SH, Richter TE, Axtell JD, Bennetzen JL (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proc Natl Acad Sci U S A* 87(11):4251–4255
- International Rice Genome Sequencing P (2005) The map-based sequence of the rice genome. *Nature* 436(7052):793–800
- Jacquemin J, Ammiraju JS, Haberer G, Billheimer DD, Yu Y, Liu LC, Rivera LF, Mayer K, Chen M, Wing RA (2014) Fifteen million years of evolution in the *Oryza* genus shows extensive gene family expansion. *Mol Plant* 7(4):642–656
- Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S et al (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6(1):4
- Lu F, Ammiraju JS, Sanyal A, Zhang S, Song R, Chen J, Li G, Sui Y, Song X, Cheng Z et al (2009) Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proc Natl Acad Sci U S A* 106(6):2071–2076
- McCouch SR, McNally KL, Wang W, Sackville HR (2012) Genomics of gene banks: A case study in rice. *Am J Bot* 99(2):407–423
- McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE et al (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc Natl Acad Sci U S A* 106(30):12273–12278
- Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution. Grasses, line up and form a circle. *Curr Biol* 5(7):737–739
- Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. *Nat Genet* 36(2):138–145
- Ohyangagi H, Tanaka T, Sakai H, Shigemoto Y, Yamaguchi K, Habara T, Fujii Y, Antonio BA, Nagamura Y, Imanishi T et al (2006) The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Res* 34:D741–744
- Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L et al (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res* 35:D883–887
- Sanyal A, Ammiraju JS, Lu F, Yu Y, Rambo T, Currie J, Kollura K, Kim HR, Chen J, Ma J et al (2010) Orthologous comparisons of the Hd1 region across genera reveal Hd1 gene lability within diploid *Oryza* species and disruptions to microsynteny in Sorghum. *Mol Biol Evol* 27(11):2487–2506
- Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH et al (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* 270(5243):1804–1806
- Wang GL, Holsten TE, Song WY, Wang HP, Ronald PC (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. *Plant J* 7(3):525–533
- Wang M, Yu Y, Haberer G, Marri PR, Fan C, Goicoechea JL, Zuccolo A, Song X, Kudrna D, Ammiraju JS et al (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat Genet* 46(9):982–988
- Wing RA, Ammiraju JS, Luo M, Kim H, Yu Y, Kudrna D, Goicoechea JL, Wang W, Nelson W, Rao K et al (2005) The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol Biol* 59(1):53–62
- Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442(7103):705–708
- Zhang Q (2007) Strategies for developing Green Super Rice. *Proc Natl Acad Sci U S A* 104(42):16402–16409
- Zhang QJ, Zhu T, Xia EH, Shi C, Liu YL, Zhang Y, Liu Y, Jiang WK, Zhao YJ, Mao SY et al (2014) Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *Proc Natl Acad Sci U S A* 111(46):E4954–4962