

Minireview

A fruitful outcome to the papaya genome project

Fusheng Wei and Rod A Wing

Address: Department of Plant Sciences and the Arizona Genomics Institute, University of Arizona, Tucson, AZ 85721, USA.

Correspondence: Rod A Wing. Email: rwing@Ag.arizona.edu

Published: 6 June 2008

Genome Biology 2008, **9**:227 (doi:10.1186/gb-2008-9-6-227)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2008/9/6/227>

© 2008 BioMed Central Ltd

Abstract

The draft genome sequence of a transgenic virus-resistant papaya marks the first genome sequence of a commercially important transgenic crop plant.

The nice thing about working with some plant genomes is that at the end of the day you can eat the fruits of your work. Originating from Central and South America, the papaya (*Carica papaya*) bears highly nutritious and delicious fruit and is also a source of papain - a protease used for centuries to tenderize meat. Papaya trees can grow 5-10 meters tall, with large leaves 50-70 cm in diameter and fruits 15-45 cm long and 10-30 cm in diameter (Figure 1). The papaya is grown as a crop in tropical and subtropical regions, but its cultivation has been severely hampered by the papaya ringspot virus (PRSV) (Figure 2). In Hawaii, papaya cultivation was almost completely destroyed by the virus until the introduction of virus-resistant transgenic lines in 1998. Now 80% of the Hawaiian papaya crop is transgenic [1]. A draft genome sequence and analysis of the transgenic papaya variety 'SunUp' has now been published by Ray Ming and co-workers [2].

In regard to genomics, the papaya is an ideal and interesting species to work with. It has a very small diploid genome of 372 Mb [3], slightly smaller than rice [4] and six times smaller than maize [5]. The papaya belongs to the order Brassicales, which includes the model plant *Arabidopsis* as well as the cabbage family; it shared a common ancestor with *Arabidopsis* approximately 72 million years ago [6]. The papaya can also be easily transformed [7] and has a generation time of 9-15 months. Also of interest is its primitive sex-chromosome system, which has interested evolutionary biologists for years [8].

The papaya genome was sequenced using a whole-genome shotgun approach by the traditional Sanger method to

approximately 3x coverage [1]. The majority of assembled contigs added up to about 271 Mb (73% of the genome) with scaffolds spanning 370 Mb. About 167 Mb of sequence (or 235 Mb of scaffolds) could be anchored to the integrated genetic and physical map of the papaya genome. More than half (52%) of the papaya genome comprises repetitive sequences, mainly long terminal repeat retrotransposons. Cytogenetic studies suggest that the genome is about 65-70% euchromatic and 35-30% heterochromatic. Various measures were used to assess the coverage of the draft genome, such as the percentage of unique genes (unigenes) and genetic markers matching the assembly. The authors estimate that approximately 90% of the euchromatin has been covered, containing 92.1% of the unigenes and 92.4% of the genetic markers.

Automated annotation of the genome combined with the genome coverage led the team to project a gene content of 24,746 genes. Compared with the other four sequenced plant genomes, this gene count is 11-20% less than that of *Arabidopsis* [9], 34% less than rice [4], 46% less than poplar [10] and 19% less than grape [11].

The indication that papaya contains the smallest number of genes of any plant yet sequenced was investigated further. First, all inferred non-redundant protein sequences from the five sequenced plant genomes were collapsed into 39,709 similarity groups, or 'tribes'. Then the numbers of genes found in each tribe were compared between papaya and each of the other genomes. In the papaya-*Arabidopsis* comparison, for example, 3,595 tribes out of 6,726 contained the same number of genes. However, for the remaining tribes,

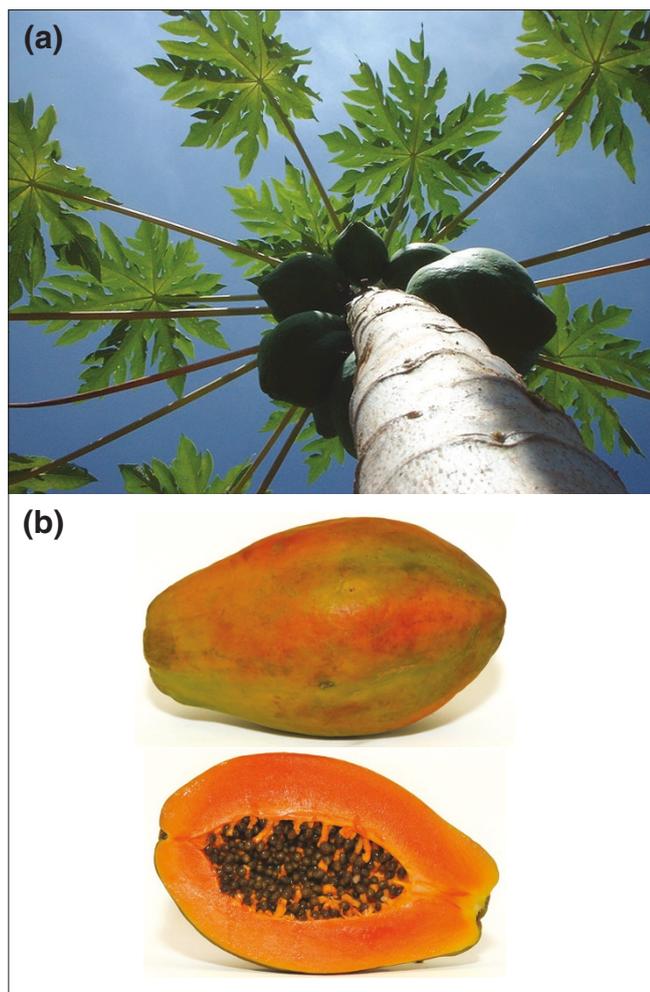


Figure 1
Papaya plant and mature fruit. **(a)** A papaya plant heavy with fruit. **(b)** Mature papaya fruit. (a) Photo courtesy of Wikicommons. (b) Photo courtesy Jayson Talag, University of Arizona.

Arabidopsis genes outnumbered papaya by two to one, and this trend was consistent with all the other plant genome sequences.

The team next asked what the minimum set of genes required for an angiosperm might be. By determining the genes shared across all the 39,709 tribes among the five sequenced genomes they estimated this minimum to be 13,311 genes. As papaya had the smallest numbers of genes over the most tribes, these data further supported the idea that it has the lowest gene count of any plant genome so far sequenced.

One possible explanation for the lower than expected number of genes is that the papaya genome did not undergo the two rounds of recent whole-genome duplication observed in *Arabidopsis* [12]. Analysis of syntenic blocks between papaya and *Arabidopsis* revealed that for single papaya genes,



Figure 2
Field trial of transgenic papaya. The disease free transgenic papaya plants (the right side) and the severely infected and stunted non-transgenic papaya plants (the left side) growing in adjoining plots. Image courtesy of [23].

Arabidopsis has two to four corresponding genes, but that each *Arabidopsis* gene only has one counterpart in papaya.

Interestingly, when syntenic blocks from the grape genome were included along with *Arabidopsis* in the analysis, Ming *et al.* [1] detected a possible ancient whole-genome triplication that occurred before the divergence of the three species, but after the separation of the monocotyledons and dicotyledons. This triplication event was first proposed on the evidence of the grape genome sequence [11] and is now supported by the papaya sequence.

Ming *et al.* [1] categorized several important gene families essential for papaya fitness. One surprise is the extremely small number of disease-resistance genes of the nucleotide-binding site leucine-rich repeat (NBS-LRR) class. *Arabidopsis* has more than 200 NBS-LRR genes [13] and rice more than 600 [14]. In contrast, there are only 55 NBS-LRR genes in papaya, but they are clustered in a similar fashion to those in *Arabidopsis* and rice. This dearth of NBS-LRR genes might suggest that papaya has developed alternative strategies of host defense, such as the evolution of other classes of resistance genes (for example, tomato *Cf*-like genes [15], rice *Xa-21* like receptor kinase [16] or maize *Hm1*-like detoxin protein [17]) or even of nonhost resistance [18], in which all members of a plant species exhibit resistance to all members of a given pathogen species.

The papaya genome has a similar number of genes to *Arabidopsis* and poplar for cellulose biosynthesis, cell wall and lignin syntheses, and ethylene biosynthesis, but fewer genes involved in cell-wall degradation and in light-induced and circadian rhythms. On the other hand, papaya has more genes associated with starch metabolism and the development of volatiles.

The papaya genome sequence also sheds new light on the primitive XY sex-chromosome system, where the Y chromosome contains a male-specific region (MSY) approximately 8 Mb in length [8,19]. Two scaffolds (totaling approximately 4.5 Mb) from the female papaya genome sequence determined by Ming *et al.* [1] aligned to a bacterial artificial chromosome (BAC) physical map of the X chromosome. The female region contained 254 genes, of which 75% were supported by expressed sequence tags. In contrast, only four expressed genes have so far been found among seven completely sequenced BACs (totaling 1.2 Mb in length) in the MSY region [20]. Using repeat data derived from the whole-genome sequence, Ming *et al.* were able to show that 85.6% of the 1.2 Mb MSY sequence is composed of repeats. Although complete sequence data are not yet available for the MSY from a male genome, the sequence generated from the female will provide essential comparative information to help unravel the mysteries of the evolution and function of sex chromosomes in plants.

An important point of the paper by Ming *et al.* is that the genome analyzed was from a transgenic inbred line. The PRSV coat protein transgene confers resistance to the virus and was introduced into papaya by particle bombardment. Particle bombardment can cause the construct to be fragmented, resulting in multiple integration events within the genome [21]. The genome sequence enabled the identification of multiple integration sites, three of which occurred in nuclear genomic regions that contained AT-rich DNA fragments from the chloroplast genome. From a regulatory viewpoint, precise identification of the insertion sites of a transgene is required by many countries in order to obtain permission to grow or import transgenic food crops. Now that these integration sites have been determined, a major hurdle to the introduction of transgenic papaya in other countries has been removed.

In summary, a new and interesting plant genome sequence is now publicly available for interrogation. The papaya genome provides basic plant research with an exciting new tool to better understand angiosperm evolution and sex-chromosome biology. It also provides clues to the minimum set of genes that are needed to be a flowering plant. On the practical side, the papaya genome sequence has yielded a vast set of molecular genetic markers that can be used to create higher yielding, more nutritious and hardier papaya varieties.

It is safe to assume that the genomes of all the major food and fiber crops will be sequenced within the next five years. This is an important goal for both the plant genomics community and the wider world if we are to meet the food and energy security needs of the future. Sequencing the papaya genome illustrates just how much we do not know about plant genomes and how important it will be to generate reference genome sequences at key nodes on the tree of life. One argument has recently been made to

sequence the genome of *Amborella trichopoda*, which lies at the base of angiosperm evolution [22]. Whoever coined the phrase 'post-genomics' has jumped the gun. Plant genomics has just touched the surface of new biological discovery and practical solutions in support of the next green revolution.

References

1. Stokstad E: **Papaya takes on ringspot virus and wins.** *Science* 2008, **320**:472.
2. Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, *et al.*: **The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus).** *Nature* 2008, **452**:991-996.
3. Arumuganathan K, Earle ED: **Nuclear DNA content of some important plant species.** *Plant Mol Biol Rep* 1991, **9**:208-218.
4. International Rice Genome Sequencing Project: **The map-based sequence of the rice genome.** *Nature* 2005, **436**:793-800.
5. Wei F, Coe E, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen M, Lee S, Fuks G, Sanchez-Villeda H, Schroeder S, Fang Z, McMullen M, Davis G, Bowers JE, Paterson AH, Schaeffer M, Gardiner J, Cone K, Messing J, Soderlund C, Wing RA: **Physical and genetic structure of the maize genome reflects its complex evolutionary history.** *PLoS Genet* 2007, **3**:e123.
6. Wikstrom N, Savolainen V, Chase MW: **Evolution of the angiosperms: calibrating the family tree.** *Proc R Soc Lond B* 2001, **268**:2211-2220.
7. Fitch MMM, Manshardt RM, Gonsalves D, Slightom JL, Sanford JC: **Virus resistant papaya plants derived from tissues bombarded with the coat protein gene of papaya ringspot virus.** *Bio/technology* 1992, **10**:1466-1472.
8. Liu Z, Moore PH, Ma H, Ackerman CM, Ragiba M, Yu Q, Pearl HM, Kim MS, Charlton JW, Stiles JL, Zee FT, Paterson AH, Ming R: **A primitive Y chromosome in papaya marks incipient sex chromosome evolution.** *Nature* 2004, **427**:348-352.
9. The Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**:796-815.
10. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhallerao RR, Bhallerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, *et al.*: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science* 2006, **313**:1596-1604.
11. The French-Italian Public Consortium for Grapevine Genome Characterization: **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla.** *Nature* 2007, **449**:463-467.
12. Bowers JE, Chapman BA, Rong J, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events.** *Nature* 2003, **422**:433-438.
13. Meyers BC, Morgante M, Michelmore RW: **TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes.** *Plant J* 2002, **32**:77-92.
14. Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D: **Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes.** *Mol Genet Genomics* 2004, **271**:402-415.
15. Jones DA TC, Hammond-Kosack KE, Balint-Kurti PJ, Jones JD: **Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging.** *Science* 1994, **266**:789-793.
16. Song WY WG, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P: **A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*.** *Science* 1995, **270**:1804-1806.
17. Johal GS, Briggs SP: **Reductase activity encoded by the HMI disease resistance gene in maize.** *Science* 1992, **258**:985-987.
18. Ellis J: **Insights into nonhost disease resistance: can they assist disease control in agriculture?** *Plant Cell* 2006, **18**:523-528.
19. Yu Q, Hou S, Feltus FA, Jones MR, Murray JE, Veatch O, Lemke C, Saw JH, Moore RC, Thimmapuram J, Liu L, Moore PH, Alam M, Jiang

- J, Paterson AH, Ming R: **Low X/Y divergence in four pairs of papaya sex-linked genes.** *Plant J* 2008, **53**:124-132.
20. Yu Q, Hou S, Hobza R, Feltus FA, Wang X, Jin W, Skelton RL, Blas A, Lemke C, Saw JH, Moore PH, Alam M, Jiang J, Paterson AH, Vyskot B, Ming R: **Chromosomal location and gene paucity of the male specific region on papaya Y chromosome.** *Mol Genet Genomics* 2007, **278**:177-185.
21. Sawasaki T, Takahashi M, Goshima N, Morikawa H: **Structures of transgene loci in transgenic *Arabidopsis* plants obtained by particle bombardment: junction regions can bind to nuclear matrices.** *Gene* 1998, **218**:27-35.
22. Soltis DE AV, Leebens-Mack J, Palmer JD, Wing RA, Depamphilis CW, Ma H, Carlson JE, Altman N, Kim S, Wall PK, Zuccolo A, Soltis PS: **The *Amborella* genome: an evolutionary reference for plant biology.** *Genome Biol* 2008, **9**:402.
23. **Hawaii Papaya Industry Association** <http://www.hawaiipapaya.com/>