Minireview

Brassica genomics: a complement to, and early beneficiary of, the *Arabidopsis* sequence

Andrew H Paterson*, Tien-hung Lan†, Richard Amasino‡, Thomas C Osborn§ and Carlos Quiros¥

Addresses: *Center for Applied Genetic Technologies, Riverbend Research Lab, University of Georgia, Athens, GA 30602, USA. †Structural Genomics, Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA92121, USA. ‡Department of Biochemistry, University of Wisconsin, 433 Babcock Drive, Madison, WI 53706-1544, USA. §Department of Agronomy, University of Wisconsin, Madison, WI 53706, USA. ¥Department of Vegetable Crops, University of California, Davis, CA 95616, USA.

Correspondence: Andrew H Paterson. E-mail: paterson@uga.edu

Published: 9 March 2001

Genome ***Biology*** 2001, **2(3)**:reviews1011.1–1011.4

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2001/2/3/reviews/1011

Abstract

Those studying the genus *Brassica* will be among the early beneficiaries of the now-completed *Arabidopsis* sequence. The remarkable morphological diversity of *Brassica* species and their relatives offers valuable opportunities to advance our knowledge of plant growth and development, and our understanding of rapid phenotypic evolution.

Beyond *Arabidopsis*

Less than 20 years after DNA-level studies [1] highlighted the small DNA content and simple organization of the genome of *Arabidopsis thaliana* - an ephemeral weed-like member of the *Cruciferae* family - an essentially complete sequence of its genome has been announced [2]. This, the first complete genome sequence for a flowering plant, provides the means to investigate in unprecedented detail the common features that plants share with one another and with other life forms, as well as the unique features that adapt particular taxa to diverse life histories (such as the largely sessile nature of plants). Many botanists are especially interested in studying how the interaction of natural and human selection has enabled a small subset of plants to be 'domesticated' to provide food, feed, and fiber. Such investigations should be greatly facilitated by the availability of plant genome sequences.

The economic and morphological importance of *Brassica*

Crops of the genus *Brassica* (tribe Brassiceae), which are in the same taxonomic family as *Arabidopsis thaliana,* are widely used in the cuisine of many cultures, due, in part, to the many choices of edible forms in the genus. Economically, *Brassica* is loosely categorized into oilseed, vegetable, and condiment crops. *B. napus*, *B. rapa* (formerly *campestris*), *B. juncea*, and *B. carinata* provide about 12% of the world-wide edible vegetable oil supplies [3]. *B. oleracea* and *B. rapa*, the so-called 'cole crops', comprise many of the vegetables in our daily diet. Several of these vegetables have extreme morphological characteristics. Examples of such morphologies include the enlarged inflorescence of cauliflower (*B. oleracea* subspecies *botrytis*) and broccoli (*B. oleracea* subspecies *italica*); the enlarged stem of kohlrabi (*B. oleracea* subsp. *gongylodes*) and marrowstem kale (*B. oleracea* subspecies *medullosa*); the enlarged root of turnip (*B. rapa* subspecies *rapifera*); the enlarged and twisted leaves of Pak-choi (*B. rapa* subspecies *chinesis*) and Chinese cabbage (*B. rapa* subspecies *pekinesis*); and the enlarged single apical bud of cabbage (*B. oleracea* subspecies *capitata*) or the many axillary buds of Brussels sprout (*B. oleracea* subspecies *gemmifera*) [4]. Finally, the seed of *B. nigra* is utilized as a condiment - mustard. *Brassica* species are a valuable source of dietary fiber, vitamin C, and other possible salubrious factors such as anticancer compounds [5]. Estimates of the

economic importance of *Brassica* species are conservative, because several cole crops such as collards are cultivated primarily for local or home use, but are nonetheless a dietary mainstay in low-income communities where other fresh vegetables can be prohibitively expensive.

The current state of *Brassica* genomics

The genomes of diploid *Brassica* species are 3-5 times the size of the *Arabidopsis* genome, ranging from 0.97pg/2C (468 Mb/1C; where C is haploid DNA per nucleus) for *B. nigra* to 1.37 pg/2C (662 Mb/1C) for *B. oleracea*. The amphidiploid genomes, which have two sets of chromosomes from each parent species, range from 2.29 pg/2C (1,105 Mb/1C) to 2.56 pg/2C (1,235 Mb/1C) [6,7]. Relationships among the three diploid species - *B. rapa* (syn. rapa; 2n = 20, genome AA), *B. nigra* (2n = 16, genome BB) and *B. oleracea* (2n = 18, genome CC) - and three amphidiploids - *B. napus* (2n = 38, genome AACC), *B. juncea* (2n = 36, genome AABB) and *B. carinata* (2n = 34, genome BBCC) - are well known [8]. Rapid-cycling strains of *B. oleracea* have been developed that have a life cycle as short as that of *Arabidopsis thaliana*, making them easier to study [9].

Neutral DNA polymorphisms are common among con-specific genotypes of most *Brassica* species, and at least 15 molecular maps have been produced, using at least 900 different publicly available *Brassica* and *Arabidopsis* DNA probes (which are each used with similar efficacy). Virtually all the probes hybridize to two or more loci, even in diploid *Brassica* species; however, only a subset of loci segregates for allelic variation in any one mapping population. The subset of genetically mapped loci can, therefore, be used to infer the locations of many additional sequence tagged sites (STSs) [10].

The comparative organization of the chromosomes of *Brassica* and *Arabidopsis* is well studied. On the basis of 186 corresponding loci, about 19 chromosome structural rearrangements differentiate *B. oleracea* and *A. thaliana* orthologs, suggesting that chromosomal tracts of about 20-25 cM remain largely colinear in the two taxa [10]. Microsynteny studies involving comparative genetic and physical mapping of specific chromosome segments have shown largely conserved gene order in *Arabidopsis* and *Brassica*, but some disruption in gene content by deletions or insertions [11,12]. Furthermore, duplicated partial gene clusters are commonly found in both *Arabidopsis* and *Brassica* species [13-16]. Comparative sequencing has permitted orthology assignment of some duplicated segments in *Arabidopsis* and *Brassica* species [17].

Expectations of comparative *Arabidopsis-Brassica* genomics

Comparative data help to highlight how tools from *Arabidopsis* might be used to identify genes that may directly account for *Brassica* quantitative trait loci (QTLs). An example of this from one of our labs involves the flowering-time genes. In *B. rapa*, a major QTL for flowering time was found in a region homologous to the top of chromosome 5 in *Arabidopsis* where several flowering-time genes are located [18]. After backcrossing, the *B. rapa* locus segregated as a discrete character, and, by comparative fine mapping, it was shown to correspond to the *Arabidopsis* flowering-time gene *flc* [19]. In *B. oleracea*, 15 QTLs controlling flowering time were aligned to *Arabidopsis* genomic regions containing flowering-time genes [20]. Ten of these QTLs occurred in regions corresponding to several segments of the *Arabidopsis* genome containing genes that affect flowering: the top of chromosome 5, which contains 7 flowering genes (*tfl1, flc, tfl2, co, fy, art1* and *emf1*); three regions of chromosome 1 containing *efs*, *fha*, and *gi*; and a region of chromosome 3 containing *hy2* and *vrn1*. Similarly abundant sets of *Arabidopsis* candidates have been found for *Brassica* QTLs that affect leaf and overall plant architecture [21]. The *Arabidopsis* genome sequence provides a valuable resource for identifying and further evaluating sets of candidate genes that may account for the genetic control of complex traits in *Brassica*.

New research opportunities from *Brassica* genomics

Comparative genomics has advanced the discovery and understanding of conserved genes, but most sequencing projects have also revealed many rapidly evolving 'orphan' genes of unknown function and evolutionary history. Such rapidly evolving genes in *Drosophila*, mammals, and several other species are important for reproductive success, cell-cell recognition, and cellular response to pathogens [22,23]. Because expressed sequence tag (EST) databases for *Brassica* remain small (as of 27 January 2001, a search of the National Center for Biotechnology Information (NCBI) databases [24] for 'Brassica' reveals only 5,868 matches), the abundance and possible importance of rapidly evolving genes in accounting for key differences between *Brassica* and *Arabidopsis*, or among different *Brassica* species and genotypes, remain unknown.

Naturally occurring *Brassica* genetic variants may prove useful for new investigations into plant development, particularly through the study of the effects of alleles that lack variation in *Arabidopsis*. Many *Brassica* phenotypes are under much more complex genetic control than similar phenotypes in *Arabidopsis*. For example, the discovery that the joint effects of two mutations in *Arabidopsis*, *CAULIFLOWER* and *APETALA1*, could produce a plant with a curd-like inflorescence [25] is in contrast to the much more complex genetic control of curd architecture reported for *Brassica* [20]. The high level of duplication in the *Brassica* genome [10] may permit mutations in loci that are under tight selective constraints in *Arabidopsis*. Although early hints that there may be large-scale chromatin duplications in

Arabidopsis [26,27] have been borne out by the complete sequence [2,28-30], gene loss or divergence has been substantial; most *Brassica* genes, however, are represented in two to three or even more copies. Investigations into the expression and function of redundant genes in *Brassica* species should shed light on the role of genome replication in the phenotypic divergence of the genus.

Brassica species also provide the opportunity to study rapid genome changes associated with polyploidy. Each of the amphidiploid species - *B. napus*, *B. juncea* and *B. carinata* - can be resynthesized by hybridizing the diploid species and then doubling the chromosomes [31]. This results in completely homozygous polyploid lines that should not segregate, but variation in DNA restriction fragments [32] and phenotype [33] has been observed among progeny derived from self-pollination of single resynthesized polyploids. As *Brassica* and *Arabidopsis* genes share, on average, 87% sequence identity [16], it may be possible to survey large numbers of *Brassica* genes for changes in expression using DNA microarrays based on *Arabidopsis* gene sequences. Comparisons of newly derived polyploids with their exact diploid progenitors should provide insight into the consequences of polyploidy on plant evolution.

Brassica is distinguished from many other plant taxa by its remarkable propensity to evolve new morphological variants rapidly, as is evident from the inter-fertility among common vegetables such as cabbage, cauliflower, broccoli, Brussels sprouts, and kohlrabi. In much the same manner that the dog genome project promises to contribute to the identification of genes conferring morphological differences among mammals [34], *Brassica* genomics might provide new insights into the size and shape of plants. To this end, the complete sequence of *Brassica*'s close relative, *Arabidopsis thaliana*, should prove to be a powerful tool.

Acknowledgements

We thank current and past members of our labs for valuable contributions, and the USDA-IFAFS and NSF Plant Genome Programs for financial support.

References

1. Leutwiler L, Hough-Evans B, Meyerowitz E: **The DNA of *Arabidopsis thaliana***. *Mol Gen Genet* 1984, **219:**225-234.
2. The *Arabidopsis* Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana***. *Nature* 2000, **408:**796-815.
3. Labana KS, Gupta ML: **Importance and Origin**. In *Breeding Oilseed Brassicas*. Edited by Labana KS, Banga SS, Banga SK. Berlin: Springer-Verlag Press; 1993, 1-20.
4. Kalloo G, Bergh BO: *Genetic Improvement of Vegetable Crops*. Oxford, Pergamon Press, 1993.
5. Fahey JW, Talalay P: **The role of crucifers in cancer chemoprotection**. In: *Phytochemicals and Health*. Edited by Gustine DL, Flores HE. Rockville: American Society of Plant Physiologists; 1995, 87-93.
6. Bennett MD, Smith JB: **Nuclear DNA amounts in angiosperms**. *Philos Trans R Soc Lond B Biol Sci*1976, **274:**227-74.
7. Arumuganathan K, Earle ED: **Nuclear DNA content of some important plant species**. *Plant Mol Biol Rep* 1991, **9:**208-218.
8. U N: **Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization**. *Jpn J Bot* 1935, **7:**389-452.
9. Williams PH, Hill C: **Rapid-cycling populations of *Brassica*.** *Science* 1986, **232:**1285-1289.
10. Lan, TH, Delmonte TA, Reischmann KP, Hyman J, Kowalski S, McFerson J, Kresovich S, Paterson AH: **EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana***. *Genome Res* 2000, **10:**776-788.
11. Conner JA, Conner P, Nasrallah ME, Nasrallah JB: **Comparative mapping of the Brassica S locus region and its homeolog in *Arabidopsis*: implications for the evolution of mating systems in the Brassicaceae**. *Plant Cell* 1998, **10:**801-812.
12. Grant MR, McDowell JM, Sharpe AG, de Torres Zabala M, Lydiate DJ, Dangl JL: **Independent deletions of a pathogen resistance gene in *Brassica* and *Arabidopsis***. *Proc Natl Acad Sci USA* 1998, **95:**15843-15848.
13. Sadowski J, Gaubier P, Delseny M, Quiros CF: **Genetic and physical mapping in *Brassica* diploid species of a gene cluster defined in *Arabidopsis thaliana***. *Mol Gen Genet* 1996, **251:**298-306.
14. Sadowski J, Quiros CF: **Organization of an *Arabidopsis thaliana* gene cluster on chromosome 4 including the *RPS2* gene, in the *Brassica nigra* genome**. *Theor Appl Genet* 1998, **41:**226-235.
15. Lagercrantz U, Lydiate D: **Comparative genome mapping in *Brassica***. *Genetics* 1996, **144:**1903-1910.
16. Cavell AC, Lydiate D, Parkin IAP, Dean C, Trick M: **Collinearity between a 30-centimorgan segment in *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome**. *Genome* 1998, **41:**62-69.
17. Wroblewski T, Coulibaly S, Sadowski J, Quiros CF: **Variation and phylogenetic utility of the *Arabidopsis thaliana rps2* homolog in various species of the tribe Brassiceae**. *Mol Phylogenet Evol* 2000, **16:**440-448.
18. Osborn TC, Kole C, Parkin IAP, Sharpe AG, Kuiper M, Lydiate DJ, Trick M: **Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana***. *Genetics* 1997, **146:**1123-1129.
19. Kole C, Quijada P, Michaels SD, Amasino RM, Osborn TC: **Evidence for homology of flowering-time genes *VFR2* from *Brassica rapa* and *FLC* from *Arabidopsis thaliana***. *Theor Appl Genet* 2001, in press.
20. Lan TH, Paterson AH: **Comparative evolution of QTLs sculpting the curd of *Brassica oleracea***. *Genetics* 2000, **155:**1927-1954.
21. Lan TH, Paterson AH: **Comparative mapping of QTLs determining the plant size of *Brassica oleracea***. *Theor Appl Genet* 2001, in press.
22. Yang Z, Nielsen R, Goldman N, Krabbe Pedersen AM: **Codon-substitution models for heterogeneous selection pressure at amino acid sites**. *Genetics* 2000, **155:**431-449.
23. Swanson WJ, Yang Z, Wolfner MF, Aquadro CF: **Positive Darwinian selection drives the evolution of reproductive proteins in mammals**. *Proc Natl Acad Sci USA*, 2001, **98:** 2509-2515.
24. **The National Center for Biotechnology Information Entrez Nucleotides**
[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide]
25. Kempin SA, Savidge B, Yanofsky MF: **Molecular basis of the cauliflower phenotype in *Arabidopsis***. *Science* 1995, **267:**522-525.
26. McGrath JM, Jancso MM, Pichersky E: **Duplicate sequences with a similarity to expressed genes in the genome of *Arabidopsis thaliana***. *Theor Appl Genet* 1993, **86:**880-888.
27. Kowalski, SD, Lan T-H, Feldmann KA, Paterson AH: **Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved gene order**. *Genetics* 1994, **138:**499-510.
28. Blanc G, Barakat A, Guyot R, Cooke R, Delseny M: **Extensive duplication and reshuffling in the *Arabidopsis* genome**. *Plant Cell* 2000, **12:**1093-102.
29. Paterson, AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang C, Katsar CS, Lan T, Lin Y, Ming R, *et al.*: **Comparative genomics of plant chromosomes**. *Plant Cell* 2000, **12:**1523-1540.
30. Vision TJ, Brown DG, Tanksley SD: **The origins of genomic duplications in *Arabidopsis***. *Science* 2000, **290:**2114-2117.
31. Song K, Tang K, Osborn TC: **Development of synthetic *Brassica* amphidiploids by reciprocal hybridization and comparison to natural amphidiploids.** *Theor Appl Genet* 1993, **86:**811-821.

32. Song K, Lu P, Tang K, Osborn T: **Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution.** *Proc Natl Acad Sci USA* 1995, **92:**7719-7723.
33. Schranz ME, Osborn TC: **Novel flowering time variation in resynthesized polyploid *Brassica napus*.** *J Hered* 2001, **91:**242-246.
34. Wayne RK, Ostrander EA: **Origin, genetic diversity, and genome structure of the domestic dog**. *Bioessays* 1999, **21:**247-257.