

Arabidopsis natural variants and the Indian scenario

Sribash Roy

Arabidopsis thaliana is the model species of choice in plant science. It is the first plant species whose genome was sequenced in 2001. One of the important factors that has largely contributed in the growth of *Arabidopsis* as a model plant is existence of its natural variants across the globe and its availability from public sources. These natural variants have helped in discovering a large number of quantitative trait loci associated with specific traits and other functional alleles. The 1001 genome consortium was launched in 2009 to unearth the genetic and epigenetic variations in natural accessions spread across the globe. However, there was no report of detailed work on Indian populations of *Arabidopsis* before 2015. The Indian populations of *Arabidopsis thaliana* are unique and may provide valuable information on its evolution and adaptation under different climatic conditions. Since major conclusions on the origin and evolution of *Arabidopsis thaliana* from different studies were drawn without including the Indian populations, inclusion of these populations in global data analysis may help unearth new findings.

Historical perspective of *Arabidopsis* research

Arabidopsis thaliana, a tiny weed is widely used as the model plant by researchers. The earliest mention of the species in the literature was in a paper by Alexander Braun in 1873, where he described a mutant plant found in a field near Berlin, Germany¹. However, it was Friedrich Laibach who described the chromosome numbers of *Arabidopsis*¹. He is viewed, without question, as the father of experimental *Arabidopsis* research. In a seminal paper, Laibach for the first time proposed *Arabidopsis thaliana* as a model plant for genetic studies¹. In 1951, he and his students emphasized the importance of natural variation in the study of physiology such as flowering time and seed dormancy. Since then, it has been almost seven decades that researchers have used *A. thaliana* as a model to understand every aspect of plant life. There was increased interest in *Arabidopsis* research among researchers in the sixties. However, it gained momentum during late eighties and early nineties with researchers exploiting the genetic potential of *Arabidopsis* to characterize plant-specific processes. Thereafter, several excellent reviews summarizing the progress on plant research using *Arabidopsis* as a model have been published from time to time¹⁻⁴. In a most recent review, celebrating 50 years of the first *Arabidopsis* conference in 2015, Provart *et al.*⁴ meticulously summarized almost every aspect of plant life that has been described using *Arabidopsis*. They included 54,033

Arabidopsis papers published till 2015 to summarize how it has contributed in our understanding in the fields of cell biology, biochemistry, genetics, epigenetics, physiology, development biology, genomics, systems biology, etc. Thus as of today (7 November 2017), going by the number of publications available in PubMed, *Arabidopsis* remains as the plant with the highest number of publications (30,051; retrieved using 'Arabidopsis' as the search term in the title of the research article in PubMed advance search option). Even the important crop plants such as rice, wheat, etc. could not keep pace with *Arabidopsis* research in terms of research publications. For example, following the above search criterion, rice and wheat retrieved 23,946 and 21,500 publications respectively. It indicates that *Arabidopsis* still remains the first choice for plant biologists.

One of the important factors that contributed immensely to the growth of *Arabidopsis* as model plant is its natural variation among so-called eco-types. The analysis of genetic variation of populations found significant structure on a global scale, as well as long-range isolation by distance among different regions⁵⁻⁹. Similarly, at the regional scale, several laboratories have initiated the development of new *A. thaliana* collections for genetic variation studies such as northern Europe^{10,11}, France¹², Central Asia⁸ and China¹³. The first report on polymorphism patterns on a large-scale study by Nordborg *et al.*⁶ revealed a species-wide pattern of isolation by distance, with linkage disequilibrium decaying similarly to that in humans. This

work was followed by generation of large-scale single nucleotide polymorphism (SNP) data by several authors¹⁴⁻¹⁷. Some of these studies have shown patterns of climate adaptation using SNPs^{15,18}. Developments in quantitative genetics enabled identification of genomic regions associated with specific traits. So far, large natural variations have been reported for every phenotypic trait investigated. Several genes underlying QTLs have been identified using biparental inbred populations of *Arabidopsis*¹⁹⁻²¹. The existence of natural variations has also led to the discovery of functional allele of genes. One example is *FRI*, a major gene in the control of flowering time, which is mutated both in Col and Ler, but was identified in late-flowering natural accessions²². With the advancement of sequencing technology, genome-wide association (GWA) is rapidly becoming a preferred choice for studying the genetics of natural variations. While extensive linkage disequilibrium in *Arabidopsis* is a major disadvantage in identifying the causal site(s) among many highly associated ones, availability of naturally occurring inbred ('pure') lines makes it possible to capture local genotypes. In one of the largest samplings from a single geographic region, 180 inbred lines were sequenced to find evidence of selection and genome-wide association studies (GWAS) in local populations²³. Similarly, several hundred lines have been sequenced in other studies to compare with the existing reference genome²⁴⁻²⁶. GWAS have been extensively used as a tool for dissecting natural variations²⁷⁻²⁹. More recently, The 1001 Genomes

Consortium (2016) described global polymorphisms in *A. thaliana* using more than 1135 re-sequenced natural accessions representing the native Eurasian and North African range and recently colonized North America³⁰. This study provided a broader aspect of the species history that it is a complex mixture of survivors from multiple glacial refugia. It suggested that those which survived expansion/invasion during glacial period might have favoured population expansion. The study also provided a powerful GWAS platform to decipher how genetic variation translates into phenotypic variation in response to environmental challenges. Interestingly, this and other global studies did not include Indian populations (except two accessions – Kas-1 and Kas 2). This may be due to non-availability of accessions in stock centres from this region. It will be interesting to see if inclusion of accessions of Indian origin has any impact on the observations of the 1001 genomes study, considering our earlier observation that Indian populations might have evolved in isolation³¹.

Indian scenario

Though *A. thaliana* had gained the status of a model plant species during the nineties, publications from India were a mere 13 till 2000. This is based on the search term ‘Arabidopsis’ in the title/abstract, ‘India’ as affiliation and ‘publication date’, 1980 to 2000 in PubMed advance search option. The number of publications from 2000 to 2017 with the same criterion was 823 (as of 7 November 2019). However, this number may be more if other sources of publication could be considered. In contrast, while China had only 29 publications during 1980 to 2000, this increased to 7501 during 2000–2017 when searched with the same criterion as above, except with affiliation term as China (as on 7 November 2019).

It is interesting to note that there was no detailed report on *A. thaliana* populations from India till 2015 (ref. 7). Though there was mention of herbarium specimen and occurrence of the species in Western Himalaya, no detailed work was carried out before the report by Singh *et al.*³². There are only two Indian accessions available in the *Arabidopsis* stock centre (Kas-1 and Kas-2). How-

ever, the Indian origin of Kas-1 has been questioned by Vander Zwan and Campanella³³ who reported an European origin for the same. This was latter supported by Zhang *et al.*³⁴ on the basis of retroposons analysis. Interestingly, analysis of six Western Himalayan populations using chloroplast markers by the present author’s group also showed that the Kas-1 accession clustered together with the European accessions, but not with other Indian accessions³¹.

We initially collected six populations from different altitudes of Western Himalaya and reported the morphological trait variations³². Subsequently, we genetically characterized these populations using 19 simple sequence repeats (SSRs) and 12 chloroplast markers, and showed how these populations might be genetically distinct from the other world populations³¹. Further, using population-level transcriptome data, we observed how light intensity plays a critical role in the emergence of population-level variations in the coding regions of these

genes. We also identified high light intensity-tolerant putative genes from the highest altitude populations³⁵. It is to be noted that one of the populations studied by us is in the highest elevation habitat reported so far for the species. Availability of populations across wide climatic range due to elevation makes them ideal for studying plant adaptation. Encouraged by the initial observations, we further explored new populations and successfully collected 12 of them from this region. We characterized these populations using 43 SSR markers spanning all five chromosomes. Data suggest that most of these 12 populations from distinct clades when analysed using clustering algorithms (Figure 1). Using whole-genome sequencing approach, these populations also showed similar observations.

Conclusion

A. thaliana is the model species for modern plant biology research. From

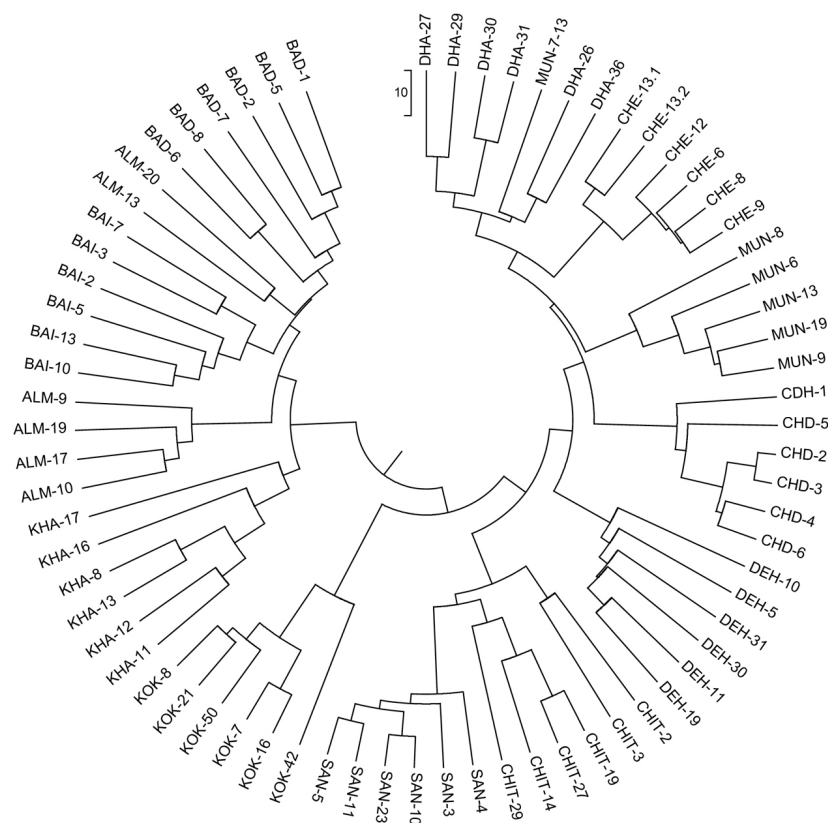


Figure 1. Unrooted neighbour joining trees of West Himalayan accessions. Genetic distance matrix of 43 MS markers was utilized for construction of tree using MEGA 5.0 (ref. 36). The three-letter initials correspond to name of the collection site followed by accession number.

understanding the different plant-specific processes such as physiological, biochemical, molecular, etc. to population-level studies, thousands of accessions and hundreds of populations have been studied worldwide. Recently, emphasis has been given on local-level population studies to explore local-level adaptation mechanisms. The Himalayan populations provide excellent opportunity in this direction. The natural variations in the coding region of the genes might provide a clue to their local adaptation and environmental factors. A greater collaboration among ecologists, molecular biologists, computational biologists, etc. may provide further insight into the evolution and adaptation mechanisms of the Himalayan populations of *A. thaliana*.

1. Koornneef, M. and Meinke, D., *Plant J.*, 2010, **61**(6), 909–921.
2. Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D. and Koornneef, M., *Science*, 1998, **282**(662), 679–682.
3. Meyerowitz, E. M., *Plant Physiol.*, 2001, **125**(1), 15–19.
4. Provart, N. J. *et al.*, *New Phytol.*, 2016, **209**(3), 921–944.
5. Beck, J. B., Schmuths, H. and Schaal, B. A., *Mol. Ecol.*, 2008, **17**(3), 902–915.
6. Nordborg, M. *et al.*, *PLoS Biol.*, 2005, **3**, e196.
7. Ostrowski, M.-F. *et al.*, *Mol. Ecol.*, 2006, **15**, 1507–1517.
8. Schmid, K. J., Törjék, O., Meyer, R., Schmuths, H., Hoffmann, M. H. and Altmann, T., *Theor. Appl. Genet.*, 2006, **112**, 1104–1114.
9. Sharbel, T. F., Haubold, B. and Mitchell-Olds, T., *Mol. Ecol.*, 2000, **9**(12), 2109–2118.
10. Bakker, E. G., Stahl, E. A., Toomajian, C., Nordborg, M., Kreitman, M. and Bergelson, J., *Mol. Ecol.*, 2006, **15**(5), 1405–1418.
11. Stenoien, H. K., Fenster, C. B., Tonteri, A. and Savolainen, O., *Mol. Ecol.*, 2005, **14**(1), 137–148.
12. Le Corre, V., *Mol. Ecol.*, 2005, **14**(13), 4181–4192.
13. He, F., Kang, D., Ren, Y., Qu, L.-J., Zhen, Y. and Gu, H., *Heredity*, 2007, **99**, 423–431.
14. Clark, R. M. *et al.*, *Science*, 2007, **317**(5836), 338–342.
15. Hancock, A. M. *et al.*, *Science*, 2011, **334**(6052), 83–86.
16. Horton, M. W. *et al.*, *Nature Genet.*, 2012, **44**(2), 212–216.
17. Kim, S. *et al.*, *Nature Genet.*, 2007, **39**(9), 1151–1155.
18. Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J. and Wilczek, A. M., *Science*, 2011, **334**(6052), 86–89.
19. Alonso-Blanco, C., Aarts, M. G. M., Bentsink, L., Keurentjes, J. J. B., Reymond, M., Vreugdenhil, D. and Koornneef, M., *Plant Cell*, 2009, **21**, 1877–1896.
20. Keurentjes, J. J. B. W. G., van Eeuwijk, F., Nordborg, M. and Koornneef, M., *Plant Genet. Resour.*, 2011, **9**, 185–188.
21. Weinig, C. *et al.*, *Genetics*, 2002, **162**(4), 1875–1884.
22. Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R. and Dean, C., *Science*, 2000, **290**(5490), 344–347.
23. Long, Q. *et al.*, *Nature Genet.*, 2013, **45**(8), 884–890.
24. Cao, J. *et al.*, *Nature Genet.*, 2011, **43**(10), 956–963.
25. Gan, X. *et al.*, *Nature*, 2011, **477**(7365), 419–423.
26. Schmitz, R. J. *et al.*, *Nature*, 2013, **495**(7440), 193–198.
27. Chao, D.-Y. *et al.*, *PLoS Genet.*, 2012, **8**, e1002923.
28. Karasov, T. L. *et al.*, *Nature*, 2014, **512**, 436–440.
29. Todesco, M. *et al.*, *Nature*, 2010, **465**(7298), 632–636.
30. Genomes Consortium, *Cell*, 2016, **166**(2), 481–491.
31. Tyagi, A. *et al.*, *AoB Plants*, 2015, **8** (plv145; doi:10.1093/aobpla/plv145).
32. Singh, A., Tyagi, A., Tripathi, A. M., Gokhale, S. M., Singh, N. and Roy, S., *Curr. Sci.*, 2015, **108**(12), 2213–2222.
33. Vander Zwan, C. B. S. and Campanella, J. J., *Syst. Bot.*, 2000, **25**, 47–59.
34. Zhang, Y., Wu, Y., Liu, Y. and Han, B., *Plant Physiol.*, 2005, **138**, 935–948.
35. Tyagi, A., Yadav, A., Tripathi, A. M. and Roy, S., *Sci Rep.*, 2016, **6**, 26160.
36. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S., *Mol. Biol. Evol.*, 2011, **28**(10), 2731–2739.

ACKNOWLEDGEMENT. I thank Council of Scientific and Industrial Research and Department of Biotechnology, New Delhi for providing funds.

Sribash Roy is in the Genetics and Molecular Biology Division, CSIR-National Botanical Research Institute, Lucknow 226 010, India and Academy of Scientific and Innovative Research, Anusandhan Bhawan, 2 Rafi Marg, New Delhi 110 001, India.
e-mail: sribashroy@nbri.res.in