PUBLISHED: 5 APRIL 2016 | ARTICLE NUMBER: 16042 | DOI: 10.1038/NPLANTS.2016.42

GENOMICS

Decoding the ancestors of peanut

Cultivated peanut has a large, complex genome, so obtaining its entire sequence is challenging. *De novo* assemblies of two diploid ancestor genomes provide high-quality reference sequences for decoding allotetraploid peanut genomes, and will become valuable resources for breeding and evolutionary studies.

Bin Han

Cultivated peanut (*Arachis hypogaea*) is a major oilseed crop of global importance. The origin of allotetraploid peanut is believed to be a single hybridization event between two wild species, *Arachis duranensis* ('A' subgenome) and *Arachis ipanensis* ('B' subgenome), giving rise to a wild diploid hybrid, followed by a chromosome duplication[1]. Domestication of the wild allotetraploid is thought to have resulted in the present-day cultivated peanut[2,3]. In a recent study in *Nature Genetics*, Bertioli and colleagues[4] present high-quality assemblies of the diploid genomes of *A. duranensis* and *A. ipaensis*, the suspected ancestors of allotetraploid cultivated peanut.

Polyploidization is a widespread process that has played a major role in plant speciation and adaptation[5,6]. However, polyploid formation causes reproductive isolation from the progenitor species, and many cultivated polyploid species have experienced a genetic bottleneck around the time of their domestication[5], which has greatly narrowed their genetic diversity. Alongside its importance for oilseed crop genetic study, peanut is also a good example of how such a bottleneck leads to low genetic diversity in a cultivated species. The high-quality genome sequences of both cultivated allotetraploid peanut and its closely related wild diploid ancestors would therefore provide crucial data for understanding peanut biology and evolution. However, the genome of cultivated peanut is both complex — it is an allotetraploid containing two closely related subgenomes, totalling four sets of chromosomes — and large (some 2.7 gigabases), making the assembly of chromosomal pseudomolecules far from straightforward.

Instead, Bertioli *et al.* provide high-quality reference sequences with which to decode the allotetraploid peanut genome. Of the respective 1.281 Gb and 1.581 Gb final assemblies of *A. duranensis* and *A. ipaensis* that they report, ~90% is represented in scaffolds of 10 kb or longer.

© DESIGN PICS INC/ALAMY STOCK PHOTO

The consequent comprehensive annotation of both genomes has characterized many aspects of the wild diploid genomes, such as gene models, gene duplications, transposon elements, patterns of genome-wide DNA methylation, and candidate resistance genes against pests and diseases.

In particular, comparisons between or within the wild diploid and cultivated allotetraploid species revealed some genome rearrangements to be related to the genetic attributes of polyploids. Most of the assembled chromosome sequences (pseudomolecules) showed one-to-one correspondence between the two wild species, and the divergence of the two species was estimated to have occurred about 2.16 million years ago. To identify the relationship between the wild diploid and cultivated allotetraploid species, the authors used Moleculo technology to generate long sequences from *A. hypopaea*, which they then used to construct a draft genome of cultivated peanut to map onto the combined pseudomolecules of the ancestral diploids. The one-to-one correspondences between the diploid chromosomal pseodumolecules and cultivated linkage groups showed that the similarities between the *A.hypogaea* Moleculo reads and the pseudomolecules of *A.duranensis* and *A.ipaensis* were 98.36% and 99.96%, respectively. These data suggest that *A. ipaensis* might be a direct descendant of the same population that contributed the B subgenome to cultivated peanut. Genetic recombination events between the homeologous A and B subgenomes were also identified. These events occurred more frequently in the regions where the homeologous chromosome pairs were collinear.

This study provides strong evidence that tetraploid cultivated peanut has experienced extreme genetic bottlenecks and reproductive isolation since polyploidization. The data analyses show that the genome of *A. ipaensis* has extremely high DNA similarity with the B subgenome of polyploid peanut. For most plants, following polyploidization, sequence identity between the diploid progenitors and the polyploidy subgenomes would have

been dispersed by genetic recombination in subsequent generations.

Peanut is widely cultivated in tropical, subtropical and warm temperate climates. However, the low level of genetic diversity between cultivated peanuts with restricted natural resistance to biotic and abiotic stresses has greatly limited the application of molecular breeding approaches for the genetic improvement of peanut[7,8]. This has driven researchers to investigate wild diploid genome sequences as a potential source of genetic polymorphisms for peanut breeding. Several agronomically important traits related to pests and disease resistance have been detected in the wild species[9]. The genome-wide polymorphisms can also be used in genomic-assisted breeding such as construction of introgression lines and generation of varieties possessing alleles from wild species[10].

Understanding the origins and domestication processes of cultivated crops is important for modern crop breeding[11,12]. The wild diploid genomes reported by Bertioli *et al*. have provided a domestication context for the wealth of existing data on genetic variation in cultivated peanut. Many traits will have been changed significantly between wild species and cultivated peanut by the phenotype-driven artificial selection of ancient farmers such as higher yield, easier growth, and preferable taste. Completion of a high-quality reference genome sequence for cultivated peanut will aid both the ancient practice of farmers' crop improvement and our understanding of crop evolution. By determining gene frequency data in cultivated populations of peanut and other crops, the changes in genetic diversity associated with domestication traits can be identified. ❐

Bin Han is at the National Center for Gene Research, CAS Center for Excellence of Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200233, China.
e-mail: bhan@ncgr.ac.cn

References

1. Kochert, G. *et al. Am. J. Bot.* **83,** 1282–1291 (1996).
2. Seijo, G. *et al. Am. J. Bot.* **94,** 1963–1971 (2007).
3. Robledo, G. & Seijo, G. *Theor. Appl. Genet.* **121,** 1033–1046 (2010).
4. Bertioli, D. J. *et al. Nature Genet.* http://dx.doi.org/10.1038/ng.3517 (2016).
5. Ramsey, J. & Schemske, D. W. *Annu. Rev. Ecol. Syst.* **19,** 467–501 (1998).
6. Soltis, P. S. & Soltis, D. E. *Proc. Natl Acad. Sci. USA* **97,** 7051–7057 (2000).
7. Halward, T. M. *et al. Genome* **34,** 1013–1020 (1991).
8. Herselman, L. *Euphytica* **133,** 319–327 (2003).
9. Mallikarjuna, N. *et al. Plant Breed.* **123,** 573–576 (2002).
10. Burow, M. D. *et al. Genome* **52,** 107–119 (2009).
11. Huang, X. *et al. Nature* **490,** 497–501 (2012).
12. Huang, X. & Han, B. *Annu. Rev. Plant Biol.* **65,** 531–551 (2014).