www.nature.com/ng/ March 2024 Vol. 56 No. 3
nature genetics

A genomic variation map provides insights into peanut diversity in China and associations with 28 agronomic traits

Qing Lu# *, Lu Huang#, Hao Liu#, Vanika Garg, Sunil S. Gangurde, Haifen Li, Annapurna Chitikineni, Dandan Guo, Manish K. Pandey, Shaoxiong Li, Haiyan Liu, Runfeng Wang, Quanqing Deng, Puxuan Du, Rajeev K. Varshney*, Xuanqiang Liang*, Yanbin Hong*, Xiaoping Chen*

1.Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangdong Provincial Key Laboratory of Crop Genetic Improvement, South China Peanut Sub-Centre of National Centre of Oilseed Crops Improvement, Guangzhou, China.

2.WA State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Australia.

3.International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India.

These authors contributed equally: Qing Lu, Lu Huang, Hao Liu.

luqing@gdaas.cn

Abstract

Peanut (*Arachis hypogaea* L.) is an important allotetraploid oil and food legume crop. China is one of the world's largest peanut producers and consumers. However, genomic variations underlying the migration and divergence of peanut in China are largely unknown. Here, we reported a genome-wide variation map based on resequencing of 390 peanut accessions, suggesting that peanut might have been introduced into southern and northern China separately, forming two cultivation centers. Selective sweep analysis highlights asymmetric selection between the two sub-genomes during peanut improvement. A classical pedigree from South China offers a context for examination of the impact of artificial selection on peanut genome. Genome-wide association studies identified 22,309 significant associations with 28 agronomic traits. Our findings shed light on peanut migration and diversity in China and provide valuable genomic resources for peanut improvement.

Introduction

Peanut (*Arachis hypogaea* L.) is one of the most important oil and food legumes worldwide, offering nutritional elements and economic value to address malnutrition and poverty. In China, peanut is an important source of vegetable oil for its residents and a major cash crop for increasing farmer income and lifting them out of poverty. The usage of introduced peanuts, including landraces and breeding lines, has brought great diversity to improved varieties in both northern and southern regions of China. However, the genomic variations underlying the phenotypic diversity owning to natural and artificial selections have not yet been investigated, and only few germplasms have been used for improving agronomic traits.

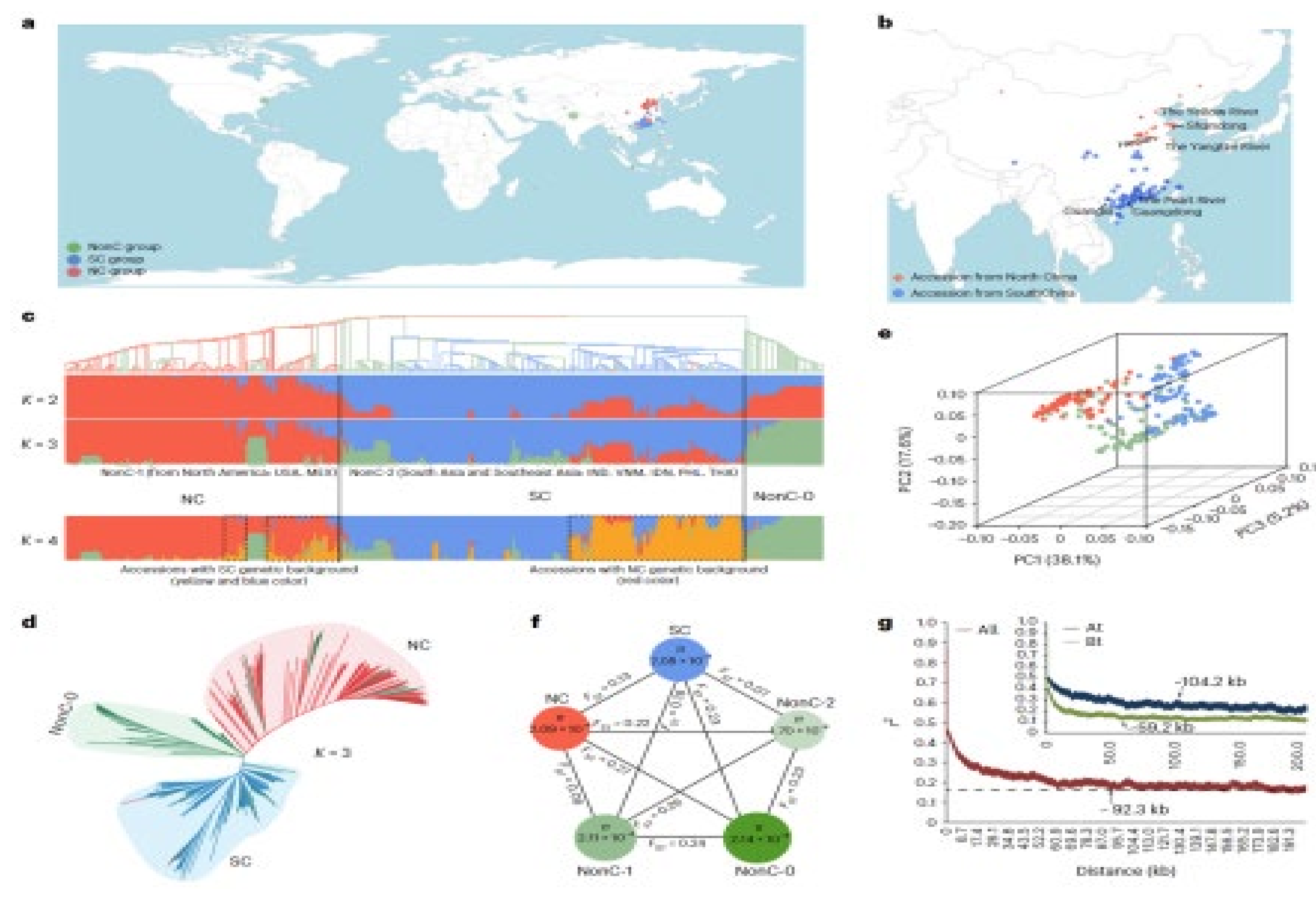
In this study, we resequenced a total of 390 peanut accessions to evaluate the genome-wide diversity of landraces and breeding lines. To further explore the genomic signatures of breeder-driven selection, we resequenced 11 elite varieties from a classical pedigree in the breeding programs in southern China. Our results provide important insights into the transmission of peanut after introduction into China, the genetic diversity and genomic variation underlying peanut agronomic traits, and valuable genomic resources and candidate genes applicable for peanut improvement.

Materials and Methods

A total of 390 peanut accessions from major peanut-growing countries were collected and conserved. Whole-genome sequencing libraries were constructed and sequenced using the Illumina HiSeq X-ten platform. Phenotyping was performed for 28 agronomic traits in four natural environments. Variant detection was done using the BWA and the Genome Analysis Toolkit. Functional annotation of SNPs was performed using ANNOVAR. Population structure and gene flow analyses were conducted. Genome-wide selective signals were identified using XPCLR. A genome-wide association study was performed using EMMAX, and candidate genes were further analyzed and characterized.

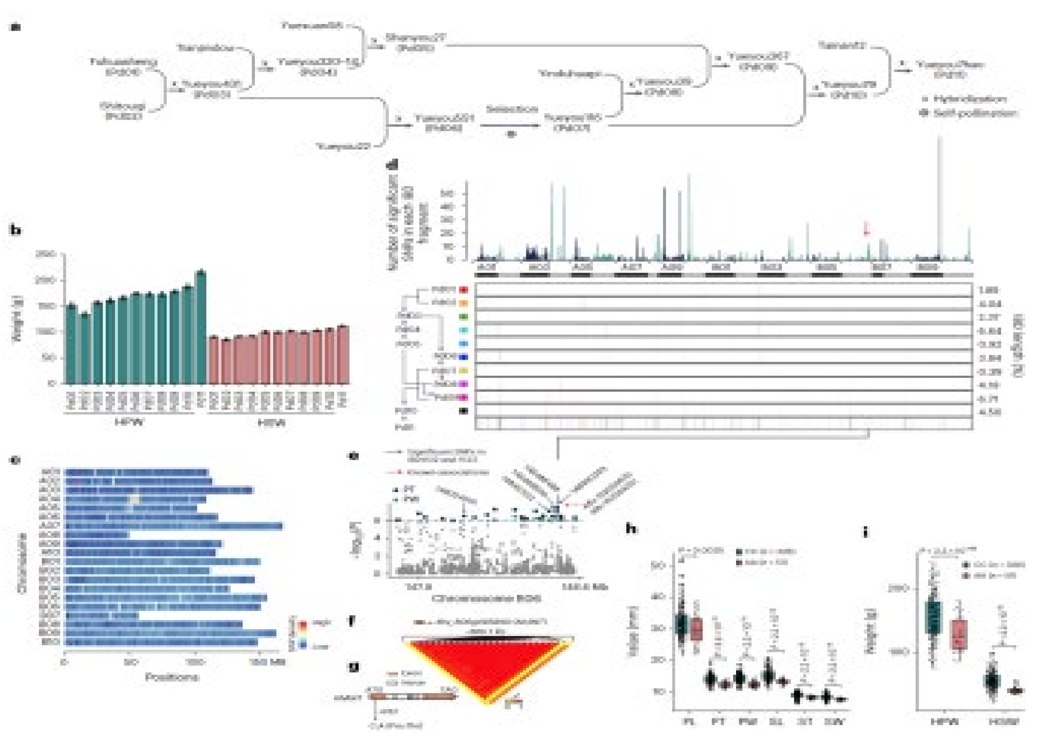
Results

1. Genome-wide variation and phylogeographic analysis



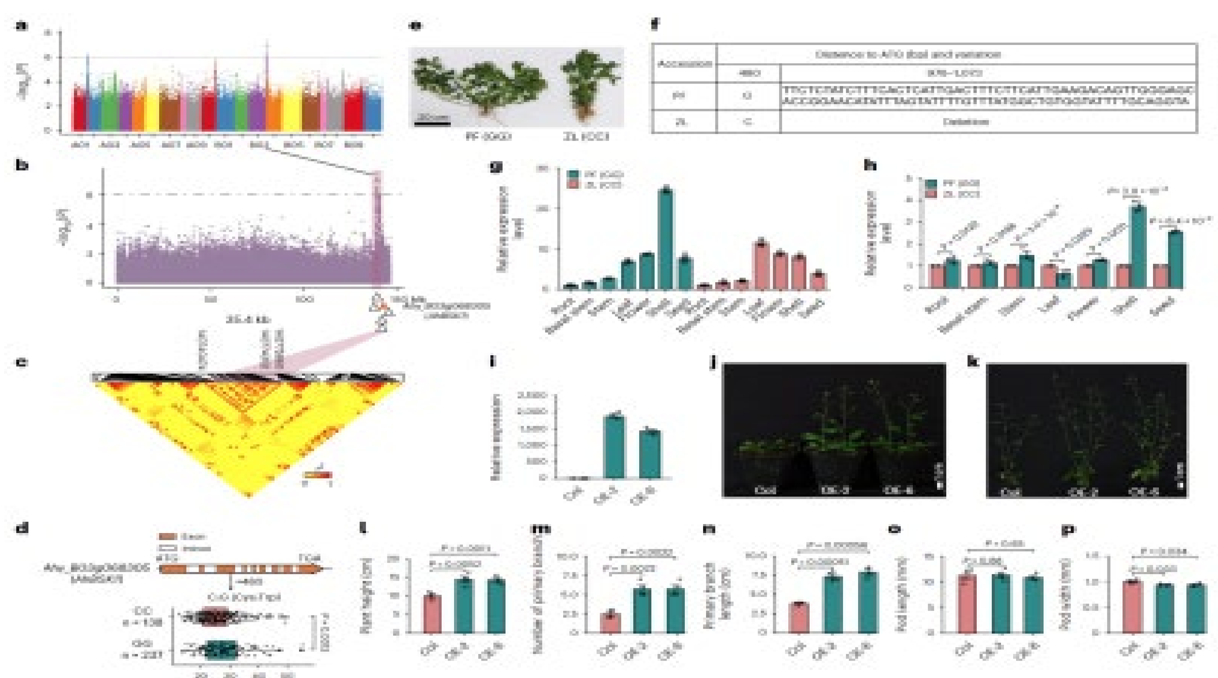
(Continued.) Geographic distribution, population structure and LD decay of 390 peanut accessions
a, Worldwide distribution of 390 peanut accessions. The size of the plot represents the sample size. b, Geographic distribution of accessions from China. c, Population structure and phylogenetic analysis. d, Phylogenetic tree of all accessions. e, Three-dimensional PCA plot of the first three principal components. f, Genetic diversity (π) and population differentiation (FST) across the three subpopulations. g, LD decay estimation of all access

2. Pedigree-based genomic signatures of artificial selection



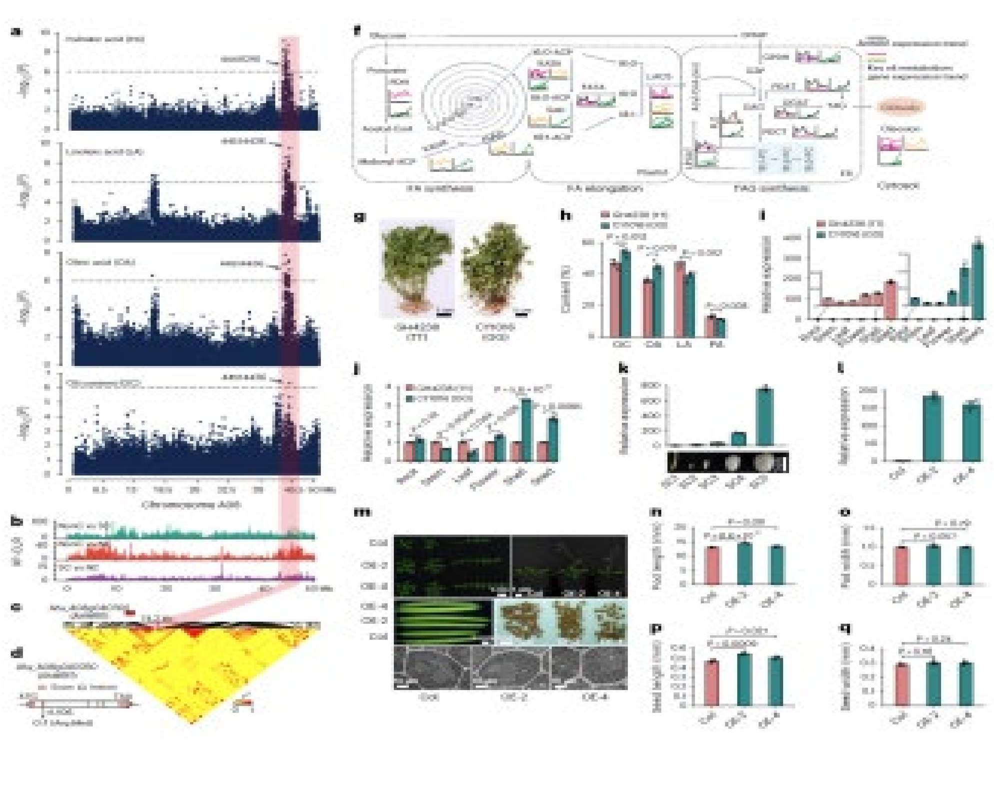
Pedigree of Yueyou7hao and GWAS for yield-related traits. a, Pedigree of Yueyou7hao. b, HPW and HSW of each variety. c, SNP density of 20 chromosomes. d, Genome flow of Yueyou7hao. e, Manhattan plot of the GWAS signals for PT and PW, overlapping with the core IBD fragments on chromosome B06. f, LDheatmap and the candidate gene *Ahy_B06g085890* (*AhANT*) in the target region. g, Gene structure and a nonsynonymous SNP of *AhANT*. h,i, Box plots for pod and seed size (h), and pod and seed weight (i) according to the genotype of the nonsynonymous SNP (C/A) in *AhANT*.

3. Candidate gene for branching habit



GWAS for branching habit and *AhBSK1* identification. a,b, Manhattan plots for PLD-associated SNPs. c, LD heatmap and the candidate gene *AhBSK1*. d, Gene structure and a nonsynonymous SNP of AhBSK1. e, Phenotypes of PF and ZL accessions. f, SNP and InDel in *AhBSK1* CDS region. g,h, Relative expression (g) and its comparison (h) of *AhBSK1* in PF and ZL accessions. i, Relative expression of *AhBSK1* in OE and wildtype (Col) Arabidopsis. j,k, The OE of *AhBSK1* resulted in earlier flowering (j) and increased biomass (k). l–p, Comparison of the plant height, number of branches, branch length, PL and PW between the OE and Col plants.

4. Candidate gene for oil biosynthesis



GWAS for oil traits and *AhWRI1* identification. a, Manhattan plots for SNPs associated with oil traits on chromosome A08. b, Selective sweep on chromosome A08. c, LD heatmap and the candidate gene *AhWRI1*.d, Gene structure and a nonsynonymous SNP of *AhWRI1*. e, Box plot of oil traits between GG and TT genotypes. f, Expression tendency of AhWRI1 and key oil metabolism genes. g,h, Phenotypes of GH4238 and CY1016 accessions and their oil traits. i,j, Relative expression(i) and its comparison (j) of *AhWRI1* in GH4238 and CY1016 accessions. k, Relative expression of *AhWRI1* in developing seeds. l, Relative expression of *AhWRI1* in OE and wild-type (Col) Arabidopsis. m, Plant (top), pod and seed (middle) and oil body (bottom) sizes of the OE and Col.

Conclusions

Peanut is a crucial crop in developing countries. Our study analyzed 390 peanut germplasm, conducting GWAS and identifying selective signals, significant associations, and candidate genes for agronomic traits. Genes like *AhANT*, *AhBSK1*, and *AhWRI1* were linked to important traits. Further research, including functional genomics methods, is necessary to validate these genes' effects in peanuts. This study provides valuable genomic data for peanut breeding and crop improvement.

Nature Genetics | *Feb. 20, 2024 | 56 | 530-540*