

# Genome Survey Sequencing of Common Vetch (*Vicia Sativa* L.) and Genetic Diversity Analysis of Chinese Germplasm with Genomic SSR Markers

lin ma (✉ [malin@caas.cn](mailto:malin@caas.cn))

Chinese Academy of Agricultural Sciences Institute of Animal Science <https://orcid.org/0000-0003-0731-1787>

Xiao Wang

Chinese Academy of Agricultural Sciences Institute of Animal Science

Min Yan

National Animal Husbandry Station, Ministry of Agriculture Sciences

Fang Liu

National Animal Husbandry Station, Ministry of Agriculture Science

Xuemin Wang

Chinese Academy of Agricultural Sciences Institute of Animal Science

---

## Research Article

**Keywords:** Chinese common vetch, Genome survey, SSR, genetic diversity

**Posted Date:** August 26th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-832190/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Molecular Biology Reports on November 6th, 2021. See the published version at <https://doi.org/10.1007/s11033-021-06875-z>.

# Abstract

Common vetch (*Vicia sativa* L.) is an annual legume with excellent suitability in cold and dry regions. Despite its great applied potential, the genomic information regarding common vetch currently remains unavailable. In the present study, the whole genome survey of common vetch was performed using the next-generation sequencing (NGS). A total of 79.84 Gbp high quality sequence data were obtained and assembled into 3,754,145 scaffolds with an N50 length of 3,556 bp. According to the K-mer analyses, the genome size, heterozygosity rate and GC content of common vetch genome were estimated to be 1,568 Mbp, 0.4345% and 35%, respectively. In addition, a total of 76,810 putative simple sequence repeats (SSRs) were identified. Among them, dinucleotide was the most abundant SSR type (44.94%), followed by Tri- (35.82%), Tetra- (13.22%), Penta- (4.47%) and Hexanucleotide (1.54%). Furthermore, a total of 58,175 SSR primer pairs were designed and ten of them were validated in Chinese common vetch. Further analysis showed that Chinese common vetch harbored high genetic diversity and could be clustered into two main subgroups. This is the first report about the genome features of common vetch, and the information will help to design whole genome sequencing strategies. The newly identified SSRs in this study provide basic molecular markers for germplasm characterization, genetic diversity and QTL mapping studies for common vetch.

## Introduction

Common vetch (*Vicia sativa* L.) is one kind of self-pollinating annual legume with excellent suitability in cold and dry regions, such as western Asia and northern Africa [1]. As an inexpensive and rich source of protein, common vetch grows quickly and is normally used for feeding livestock on account of its high digestibility [2]. In addition, common vetch can also be used as a green manure crop on idle farmland due to its strong ability to fix nitrogen [3]. Moreover, following with the increasing global population and plant-based protein demands, common vetch could be exploited as a future potential protein source due to its widely adaption in marginal cropping zones with severe drought and cold conditions [4].

However, the genetic information of common vetch remains largely unknown, although some studies have performed transcriptome analysis, genetic diversity analysis with ESTs based on SSR markers, and EMS-induced mutation development [2, 5-7]. Nevertheless, the genomic information regarding common vetch currently remains unavailable, as well as the systematic analysis of Chinese common vetch germplasm. The lack of the reference genome sequence impedes the advances in functional genomics and molecular breeding of this species [8]. Therefore, it is necessary to conduct the genome survey sequencing which would obtain basic knowledge on the genome structure of common vetch, providing a foundation for the further research of this specie.

Recently, the next-generation sequencing (NGS) has been employed as the cost-effective approach to conduct genome survey sequencing [9-11]. Except the basic knowledge of genome structure, the genome survey sequencing will provide a large number of simple sequence repeats (SSRs), which could be developed into molecular markers [12-14]. As the versatile DNA-based markers one kind of markers, SSRs

markers showed multiple advantages including co-dominant, more informative and more economical, which were generally used in plant genetic researches, such as population diversity, genetic linkage mapping and evolutionary studies [15, 16]. A larger number of SSR markers are essential for comprehensive genome-wide association studies (GWAS), as well as the quantitative trait locus mapping (QTL) and marker-assisted selection (MAS) [14, 17].

In the present research, we perform the de novo whole genome sequencing of common vetch through NGS, and then assembly to construct a reference genome database. The results showed that the genome of common vetch was estimated to be 1,568 Mbp with a heterozygosity of 0.4345%. A total of 76,810 putative SSRs were identified and 58,175 of them were designed as potential SSR markers. In addition, 10 SSRs were validated in 68 Chinese common vetch accessions and suggested the high genetic diversity of Chinese common vetch. Taken together, this study firstly reported the de novo whole genome sequencing of common vetch and firstly analyzed the genetic diversity of Chinese germplasm resources. The genome database and potential SSR markers would provide the foundation for further genomic functional and evolutionary analyses of common vetch, as well as accumulating the development of its molecular breeding.

## Materials And Methods

### Plant materials and growth condition

The common vetch (*Vicia sativa*) cv. 'Lanjian No.1' from Lanzhou University (Lanzhou, Gansu, China) was chosen for the genome survey. The seeds were planted in 10 cm pots and grown in the greenhouse at 24/22°C (day/night) temperature with 16 h light (380-400  $\mu\text{E}/\text{m}^2/\text{s}$ ). After growing for three weeks, fresh leaves from one individual plant were collected and quickly frozen in liquid nitrogen for DNA isolation.

A total of 68 common vetch accessions originating from China were used for genetic diversity analysis (Table S1). The seeds were sterilized in 75% ethanol for 5 min and rinsed with sterile water 5 times. They were placed on filter paper in dishes and then subsequently cultured in a growth chamber at 25°C. The eight-day-old seedlings were prepared for DNA extraction.

### DNA extraction and genome sequencing

Total genomic DNA was isolated by using the CTAB method with modifications [18]. DNA concentrations were measured on a Nanodrop (Thermo Fisher Scientific, Waltham, MA). DNA quality was detected on a Qubit (Thermo Fisher Scientific, Waltham, MA). The genomic paired-end library with 300-400 bp short-inserts was constructed and sequenced on an Illumina NovaSeq 6000 (Illumina Inc. San Diego, CA, USA) with PE 150 sequencing methods.

### K-mer analyses and genome size estimation

All of the clean data were used for K-mer analysis using Jellyfish software [19]. Based on the results of K-mer frequency distributions (K-mer=17), the characteristics of the genome, including genome size, repeat content and heterozygosity rate, were estimated by using GenomeScope [20].

### **Genome assembling and guanine plus cytosine (GC) content analysis**

SOAPdenovo software was used for genome assembly [21]. In brief, a de Bruijn graph was constructed based on the overlapping relationship reads from SOAPdenovo software, and contigs were output after simplifying the de Bruijn graph. Scaffolds were constructed based on the contigs, and gaps inside the scaffolds were filled by employing GAPCloser. Here, 10-kb nonoverlapping sliding windows along the assembled sequence were used to calculate the average GC sequencing depth [22].

### **Genomic microsatellite identification**

The MicroSATellite (MISA) software (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) was employed to detect the genomic SSRs. The search parameters were respectively set for identifying various types of SSRs, including Di-, Tri-, Tetra-, Penta- and Hexa-nucleotide SSR motifs with a minimum of 6, 5, 4, 4 and 4 repeats, respectively. Primer 3 software were used for primers designing for each SSR locus, with the following parameters: 18-25 primer size, 90-250 bp product size, 70% GC content, and annealing temperature of 55-65°C.

### **SSR genotyping**

The selected 10 pairs SSR primers with additional different fluorescent probes on forward primers were used for polymorphism screening of 68 common vetch accessions in China (Table S2). The PCR was conducted in a 20 µL reaction system containing 20 ng genomic DNA, 10µL 2×Taq Master Mix (Genestar, Beijing, China) and 0.5 µL each of the forward and reverse primers. The PCR parameters were as follows: 94°C for 5min ; 35 cycles: 94°C for 30s, 58°C for 30s, 72°C for 45s; the final extension at 72°C for 7min. The PCR products were diluted five times and then resolved in an ABI3130xl Genetic Analyzer (Applied Biosystems, CA, USA). Fragments size and data analysis were determined by using mapmaker 5.0 software.

### **Genetic diversity analysis**

The allelic diversity and genetic variation parameters including the number of different alleles ( $N_a$ ) and effective alleles ( $N_e$ ), the index of observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), Shannon's Information Index ( $I$ ) and the polymorphism information content (PIC) were calculated by fragment size in GenAlEx 6.5 software [23]. The genetic diversity among the 68 accessions was determined by using Dissimilarity Analysis and Representation for windows (DARwin) software. The dendrogram was generated by using the UPGMA phylogenetic cluster analysis [24].

## **Results**

## Genome sequencing and sequence assembly

To avoid the influence of potential heterozygosity, we extracted DNA from the single plant leaves of common vetch for libraries constructing (Fig. 1). After filtering the low quality data, we obtained approximately 79.84 Gbp of high-quality data from the sequencing library, which were approximately 51 times of the estimated genome size. The Q20 and Q30 of the obtained data were greater than 97% and 92%, indicating the reliability of the genome survey sequencing. We then de novo assembled (K-mer = 75) all of the high quality data by using the de Bruijn graph-based SOAPdenovo software. A total of 4,227,942 raw contigs were obtained, and the total length of raw contigs was 1,475,990,986 bp and the contig N50 length of 1,245 bp (Table 1). Finally, the assembled common vetch genome consisted of 3,754,145 scaffolds which had a total length of 1,516,858,186 bp, and the scaffold N50 length of 3,556 bp (Table 1).

## Genomic characteristics

The peak K-mer depth and the number of K-mers were calculated as 45 and 70,575,281,718, respectively, based on the K-mer analysis (K-mer = 17). The genome size of common vetch was estimated at 1,568 Mbp, while the heterozygosity rate of this genome was 0.4345%, indicating that common vetch was a self-pollinating species (Fig. 2a).

In order to investigate the guanine plus cytosine (GC) content of the common vetch genome, we built a scatterplot graph by using scaffolds larger than 500bp, elucidating the information on sequencing data bias (Fig. 2a). The results showed that the GC content of the common vetch genome was 35%, which was consistent with the main peak in the scatterplot graph. Moreover, we also noticed that the confidence area (shown in red) was around the peak at 35, suggesting that the DNA sample for genome survey sequencing was not polluted by DNA from other species.

## Genomic SSR markers development

The assembled scaffolds were employed for genomic SSR search via the MISA software (<http://pgrc.ipk-gatersleben.de/misa/misa.html>). A total of 76,810 putative SSRs were identified from 58,373 isoforms and 12,050 isoforms contained more than one SSR. Among the identified putative SSRs, 4,932 SSRs were present in compound formation. We found that the most abundant SSR type was Dinucleotide, accounting for 44.94% of the total SSRs, followed by Tri- (35.82%), Tetra- (13.22%), Penta- (4.47%) and hexa nucleotide (1.54%) SSRs (Fig. 3). The density of SSRs identified in the assembled common vetch genome was one SSR per 20.41 kb.

The SSRs were categorized by their repeat motifs. The most abundant repeats were AG/CT (17.29%) and AC/GT (15.54%), followed by AT/AT (12.02%), AAC/GTT (10.07%), AAT/ATT (9.93%) and AAG/CTT (9.37%), and AAAT/ATTT (4.95%). The most abundant pentanucleotide repeats were AAAAT/ATTTT (1.27%) and AAACC/CGTTT (0.79%) (Fig. 4). Furthermore, we designed primers for 58,175 SSRs by using Primer 3.0 software. The detailed primers are shown in Table S3.

## Genetic diversity and cluster analysis of Chinese common vetch

Ten SSR markers with polymorphisms were selected randomly to investigate genetic diversity of 68 Chinese common vetch accessions. In total, we obtained 76 alleles from the 10 SSR loci (Table 2). For each SSR loci, the number of different alleles ( $N_a$ ) and the effective number of alleles ( $N_e$ ) were ranged from 3 (SSR-12) to 16 (SSR-13) and 1.2786 (SSR-5) to 6.1286 (SSR-13), respectively. The mean  $N_a$  and  $N_e$  were 7.6 and 3.4905. The index of observed heterozygosity ( $H_o$ ) and expected heterozygosity ( $H_e$ ) ranged from 0 (SSR-12) to 0.1765 (SSR-10) and 0.3195 (SSR-5) to 0.8430 (SSR-13), with the average of 0.0632 and 0.6438, respectively. The polymorphism information content (PIC) ranged from 0.217802 (SSR-5) to 0.836845 (SSR-13) with an average of 0.639076. Other parameter, such as Shannon's information index ( $I$ ), ranged from 0.5341(SSR-5) to 2.1759 (SSR-5) with an average of 1.3387. Together, we noticed that SSR-13 harbored the highest polymorphism, followed by SSR-14, and the polymorphism of SSR-5 was the lowest (Table 2). These results suggested that the 68 common vetch accessions from China harbored high genetic diversity.

In addition, we also constructed the hierarchical tree of the Chinese common vetch accessions based on dissimilarity data, to infer phylogenetic relationships among these 68 accessions. Unweighted neighbor-joining analysis resulted in a dendrogram with two main subgroups (A and B) with 6 and 10 clusters, respectively (Fig. 5). In detail, subgroup A consisted of 33 accessions and most of them were wild accessions or landraces; in contrast, subgroup B was composed of 35 accessions but only 17 of them were wild accessions (4) and landraces (13). In addition, we hardly connected the clusters with their original places, suggesting that more markers should be hired in further population structure analysis.

## Discussion And Conclusion

With the increasing global population, global protein demand is predicted to notably increase by 50% by 2050 [25]. The extreme climate changes resulting from global warming and the urgent demand for protein have led to the search for suitable species which provide sustainable protein resources [26]. Common vetch (*Vicia sativa L.*), with excellent drought and cold tolerance, could be exploited to cope with the requirements of humans and livestock [5, 8]. However, the high  $\gamma$ -glutamyl-b-cyano-alanine (GBCA) content in common vetch seeds has restricted its application in agriculture, and the traditional strategies have failed to breed common vetch with no GBCA [4].

Modern biotechnology-based high quality references have shown promise in accelerating crop improvement [27]. With the development of functional genomics research, we can obtain the toxin-free common vetch with the use of through modern biotechnology, including over-expression or RNA interference of the candidate genes, or editing the genome by CRISPR-Cas (clustered regularly interspaced short palindromic repeats - Cas protein). These biotechnology methods have already been successfully applied in other major crops, such as rice, wheat, maize and barley [28].

In the present study, we performed a genome survey of common vetch through NGS and obtained the genome information, including genome size, heterozygosity and GC content (Table 1, Fig. 1). This

approach has been used to analyse a number of plant genomes, such as pistachio [14], *Acer truncatum* Bunge [29], *Akebia trifoliata* [30] and yellow horn [31]. The depth distribution (K-mer=17) indicated that the genome size of common vetch was 1,568 Mbp and the heterozygosity rate was 0.4345% (Fig. 2). The GC content of the common vetch genome was calculated to be 35%. To our knowledge, this report of genomic information of common vetch is the first of its kind, and lays the foundation for future genome assembly and subsequent functional genome research.

The total number of common vetch populations is quite difficult to estimate, as this species was distributed worldwide. More than 20,000 accessions were kept in the Plant Genetic Resources (PGR) [2]. It is difficult for farmers to directly use or incorporate into breeding programmes due to the large number of germplasm resources. A core collection for common vetch, which consists of 5% total accessions and represents 95% of the genetic diversity, needed to be constructed through evaluating the genetic relationships between accessions [32]. SSRs with significant dominance were used in evaluating genetic diversity in populations, and the genome survey can also provide extremely useful sources for SSR identification [33].

In this study, we identified 76,810 putative SSRs and 58,175 of them were designed as potential SSR markers. Ten of validated SSRs were selected to investigate the genetic diversity of 68 Chinese common vetch accessions with an average PIC was 0.639076 (Table 2). The results showed that the Chinese common vetch accessions harbored high genetic diversity, as well as the efficacy of the SSR markers developed in the present study. The hierarchical tree of 68 Chinese common vetch accessions indicated that these accessions could be clustered into two main subgroups. Further analysis showed that the subgroup A represented the wild and landraces accessions, whereas subgroup B represented the cultivars and commercial variety (Fig. 5). However, 10 SSR markers were largely insufficient for common vetch molecular fingerprint construction and further population structure analysis. Combined with the further research in genome assembly in chromosomal level, the larger number of SSRs identified in this study shows high potential application in construction of the common vetch core collection. Genome-wide association studies (GWAS) in core collection is an effective way for candidate gene identification in functional genome research [2]. Moreover, the larger number of SSR markers are essential for high density linkage map construction in quantitative trait locus (QTL) mapping [34]. The further selection and verification of more SSRs and their corresponding markers should be developed for functional genome research, as well as the molecular marker assisted breeding in the future.

## Declarations

**Funding** This study was supported by the earmarked fund for China Agriculture Research System (CARS-34), National Crop Germplasm Resources Center (NICGR-78) and National Program for Forage Germplasm Conservation (2130135).

**Acknowledgements** We thank Prof. Yanrong Wang (College of Pastoral Agriculture Science and Technology, Lanzhou University) for gifting the seeds of common vetch cv. 'Lanjian No.1'.

**Author contribution** LM and XMW conceived and designed the experiments. LM and XW performed the experiment. LM, MY and FL analyzed the data; LM and XMW wrote the manuscript.

**Conflict of interests** The authors declare no conflict of interest.

**Data Availability** The raw data in this research was deposited in the short read archive (SRA) databank (<http://www.ncbi.nlm.nih.gov/sra/>) and are available under the accession number PRJNA730328.

**Ethical approval** The article does not contain any studies with human participants or animals performed by any of the authors.

**Consent for publication** All the authors have read and consented to submit the manuscript.

## References

1. Robertson LD, Singh KB, Erskine W, AbdElMoneim AM (1996) Useful genetic diversity in germplasm collections of food and forage legumes from West Asia and North Africa. *Genet Resour Crop Evol* 43(5): 447-460.
2. De la Rosa L, Lopez-Roman MI, Gonzalez JM, Zambrana E, Marcos-Prado T, Ramirez-Parra E (2021) Common Vetch, valuable germplasm for resilient agriculture: genetic characterization and Spanish core collection development. *Front Plant Sci* 12:1-6.
3. Cakmakci S, Aydinoglu B, Karaca M, Bilgen M (2006) Heritability of yield components in common vetch (*Vicia sativa* L.). *Acta Agric Scand B Soil Plant Sci* 56: 54-59.
4. Nguyen V, Riley S, Nagel S, Fisk I, Searle IR (2020) Common vetch: a drought tolerant, high protein neglected leguminous crop with potential as a sustainable food source. *Front Plant Sci* 11:818.
5. Min XY, Liu ZP, Wang YR, Liu WX (2020) Comparative transcriptomic analysis provides insights into the coordinated mechanisms of leaves and roots response to cold stress in common vetch. *Ind Crops Prod* 158:112949.
6. Shi HF, Geng BH, Zhao YL, Liu YJ, Huang RS, Zhao PY, Guo ZF (2020) EMS-induced mutations in common vetch (*Vicia sativa* L.) and two mutants without anthocyanin accumulation showing increased cold tolerance. *Grassl Sci* 67(2):148-155.
7. Dong R, Dong DK, Luo D, Zhou Q, Chai XT, Zhang JY, Xie WG, Liu WX, Dong Y, Wang YR, Liu ZP (2017) Transcriptome analyses reveal candidate pod shattering-associated genes involved in the pod ventral sutures of common vetch (*Vicia sativa* L.). *Front Plant Sci* 8:649.
8. De la Rosa L, Zambrana E, Ramirez-Parra E (2020) Molecular bases for drought tolerance in common vetch: designing new molecular breeding tools. *BMC Plant Biol* 20:71.
9. Zhou W, Hu YY, Sui ZH, Fu F, Wang JG, Chang LP, Guo WH, Li BB (2013) Genome survey sequencing and genetic background characterization of *Gracilariopsis lemaneiformis* (*Rhodophyta*) based on next-generation sequencing. *PLoS One* 8(7):e69909.

10. He Y, Xiao HT, Deng C, Xiong L, Nie H, Peng C (2016) Survey of the genome of *Pogostemon cablin* provides insights into its evolutionary history and sesquiterpenoid biosynthesis. *Sci Rep* 6:26405.
11. Wang CR, Yan HD, Li J, Zhou SF, Liu T, Zhang XQ, Huang LK (2018) Genome survey sequencing of purple elephant grass (*Pennisetum purpureum* Schum 'Zise') and identification of its SSR markers. *Mol Breed* 38:94.
12. Jiao Y, Jia HM, Li XW, Chai ML, Jia HJ, Chen Z, Wang GY, Chai CY, van de Weg E, Gao ZS (2012) Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genomics* 13:201.
13. Wei X, Wang LH, Zhang YX, Qi XQ, Wang XL, Ding X, Zhang J, Zhang XR (2014) Development of simple sequence repeat (SSR) markers of Sesame (*Sesamum indicum*) from a genome survey. *Molecules* 19(4): 5150-5162.
14. Motalebipour EZ, Kafkas S, Khodaeiaminjan M, Coban N, Gozel H (2016) Genome survey of pistachio (*Pistacia vera* L.) by next generation sequencing: development of novel SSR markers and genetic diversity in *Pistacia* species. *BMC Genomics* 17:998.
15. Varshney RK, Graner A, Sorrells ME (2005) Genetic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23(1): 48-55.
16. Zalapa JE, Cuevas H, Zhu HY, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am J Bot* 99(2):193-208.
17. Blackburn A, Sidhu G, Schillinger WF, Skinner D, Gill K (2021) QTL mapping using GBS and SSR genotyping reveals genomic regions controlling wheat coleoptile length and seedling emergence. *Euphytica* 217:45.
18. Ma L, Hao CY, Liu HX, Hou J, Li T, Zhang XY (2019) Diversity and sub-functionalization of *TaGW8* homoeologs hold potential for genetic yield improvement in wheat. *Crop J* 7(6): 830-844.
19. Marcais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764-770.
20. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33: 2202-2204.
21. Luo RB, Liu BH, Xie YL, Li ZY, Huang WH, Yuan JY, He GZ, Chen YX, Pan Q, Liu YJ, Tang JB, Wu GX, Zhang H, Shi YJ, Liu Y, Yu C, Wang B, Lu Y, Han CL, Cheung DW, Yiu SM, Peng SL, Zhu XQ, Liu GM, Liao XK, Li YR, Yang HM, Wang J, Lam TW, Wang J (2015) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1(1):18.
22. Li RQ, Yu C, Li YR, Lam TW, Yiu SM, Kristiansen K, Wang J (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25(15): 1966-1967.
23. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537-2539.
24. Perrier X, Jacquemoud-Collet JP (2006) Darwin software: available online at: <http://darwin.cirad.fr/darwin>.

25. Henchion M, Hayes M, Mullen AM, Fenelon M, Tiwari B (2017) Future protein supply and demand: strategies and factors influencing a sustainable equilibrium. *Foods* 6:53.
26. Westhoek H, Rood T, van den Berg M, Janse J, Nijdam D, Reudink M, Stehfest E (2011) The protein puzzle: the consumption and production of meat, dairy and fish in the european union. *Eur J Food Res Rev* 1(3):123-144.
27. Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol* 147(3): 969-977.
28. Liu X, Wu SR, Xu J, Sui C, Wei JH (2017) Application of CRISPR/Cas9 in plant biology. *Acta Pharm Sin B* 7(3): 292-302.
29. Wang RK, Fan JS, Chang P, Zhu L, Zhao MR, Li LL (2019) Genome survey sequencing of *Acer truncatum Bunge* to identify genomic information, simple sequence repeat (SSR) markers and complete chloroplast genome. *Forests* 10(2):87.
30. Zhang Z, Zhang JW, Yang Q, Li B, Zhou W, Wang ZZ (2021) Genome survey sequencing and genetic diversity of cultivated *Akebia trifolata* assessed via phenotypes and SSR markers. *Mol Biol Rep* 48:241-250.
31. Bi QX, Zhao Y, Cui YF, Wang LB (2019) Genome survey sequencing and genetic background characterization of yellow horn based on next-generation sequencing. *Mol Biol Rep* 46:4303-4312.
32. Ulukan H (2011) Plant genetic resources and breeding: current scenario and future prospects. *Int J Agric Biol* 13(3): 447-454.
33. Vieira MLC, Santini L, Diniz AL, Munhoz CD (2016) Microsatellite markers: what they mean and why they are so useful. *Genet Mol Biol* 39 (3): 312-328.
34. Kumar, M., Choi, J.Y., Kumari, N., et al. (2015) Molecular breeding in *Brassica* for salt tolerance: importance of microsatellite (SSR) markers for molecular breeding in *Brassica*. *Front Plant Sci* 6:688.

## Tables

Table 1  
Information of the assembled genome  
sequences of common vetch.

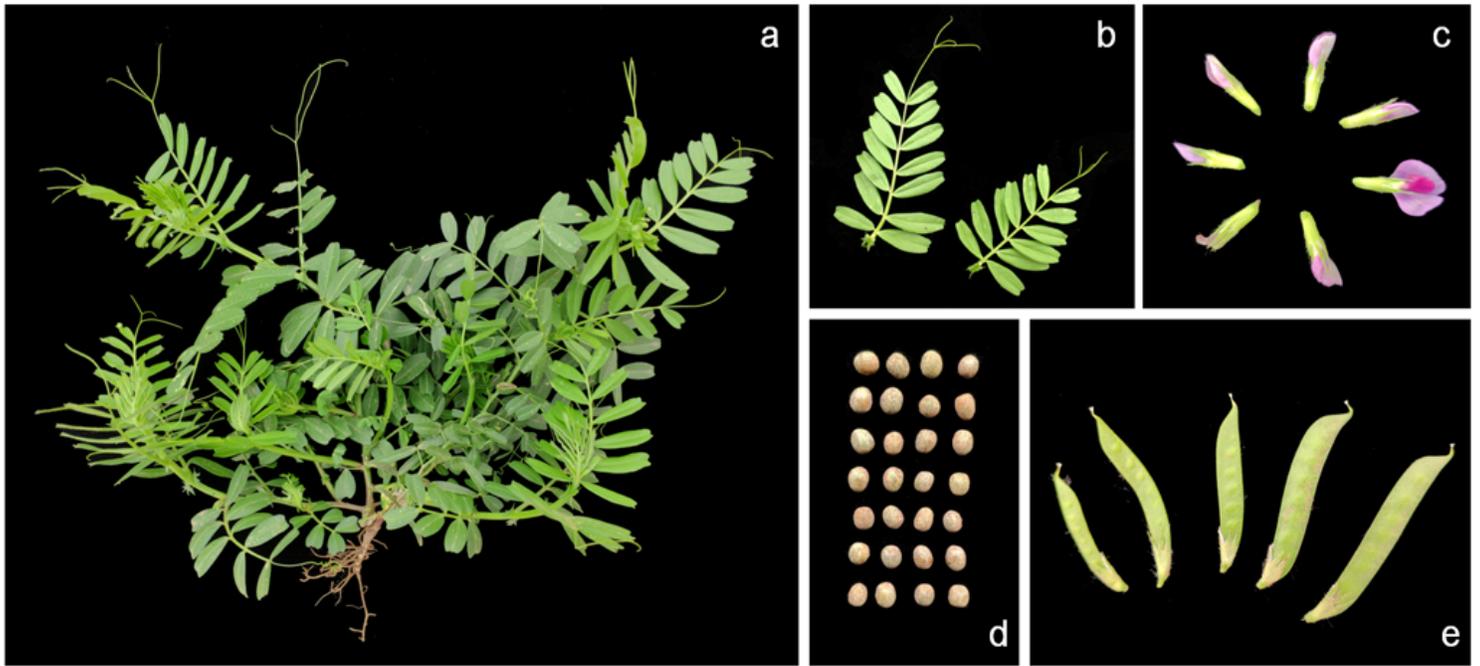
| <b>Contigs</b>      |               |
|---------------------|---------------|
| Number of sequences | 4,227,942     |
| Total length (bp)   | 1,475,990,986 |
| Max length (bp)     | 38,860        |
| N50 length (bp)     | 1,245         |
| N90 length (bp)     | 115           |
| <b>Scaffolds</b>    |               |
| Number of sequences | 3,754,145     |
| Total length (bp)   | 1,516,858,186 |
| Max length (bp)     | 89,299        |
| N50 length (bp)     | 3,556         |
| N90 length (bp)     | 116           |
| GC content          | 35.94%        |

Table 2  
Diversity statistic from 10 SSR tested in Chinese common vetch accessions  
(n = 68).

| Locus  | Na     | Ne     | I      | Ho     | He     | PIC      |
|--------|--------|--------|--------|--------|--------|----------|
| SSR-1  | 4      | 3.0552 | 1.1539 | 0.0294 | 0.6777 | 0.672682 |
| SSR-2  | 5      | 2.874  | 1.0305 | 0.0735 | 0.5670 | 0.562873 |
| SSR-3  | 6      | 2.7956 | 1.1717 | 0.0735 | 0.6471 | 0.642264 |
| SSR-5  | 6      | 1.2786 | 0.5341 | 0.0294 | 0.3195 | 0.217802 |
| SSR-9  | 10     | 3.6918 | 1.7656 | 0.0147 | 0.7345 | 0.729144 |
| SSR-10 | 5      | 2.3289 | 0.9759 | 0.1765 | 0.5748 | 0.570627 |
| SSR-11 | 11     | 5.318  | 1.9084 | 0.0441 | 0.8180 | 0.811961 |
| SSR-12 | 3      | 2.0588 | 0.7595 | 0.0000 | 0.5180 | 0.514268 |
| SSR-13 | 16     | 6.1286 | 2.1759 | 0.0735 | 0.8430 | 0.836845 |
| SSR-14 | 10     | 5.9626 | 1.9111 | 0.1176 | 0.8385 | 0.83229  |
| MEAN   | 7.6    | 3.4905 | 1.3387 | 0.0632 | 0.6438 | 0.639076 |
| SD     | 4.0332 | 1.7276 | 0.5584 | 0.0528 | 0.1897 | 0.188304 |

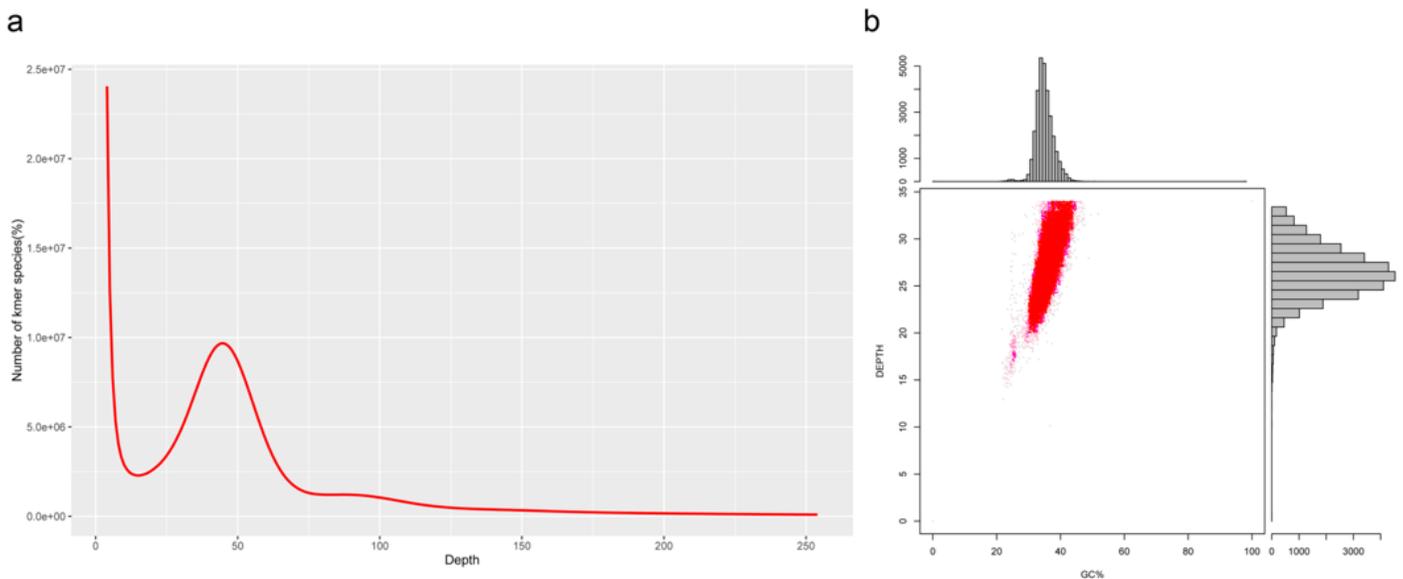
Na Observed number of alleles; Ne Effective number of alleles; I Shannon's information index, Ho observed heterozygosity; He expected heterozygosity.

## Figures



**Figure 1**

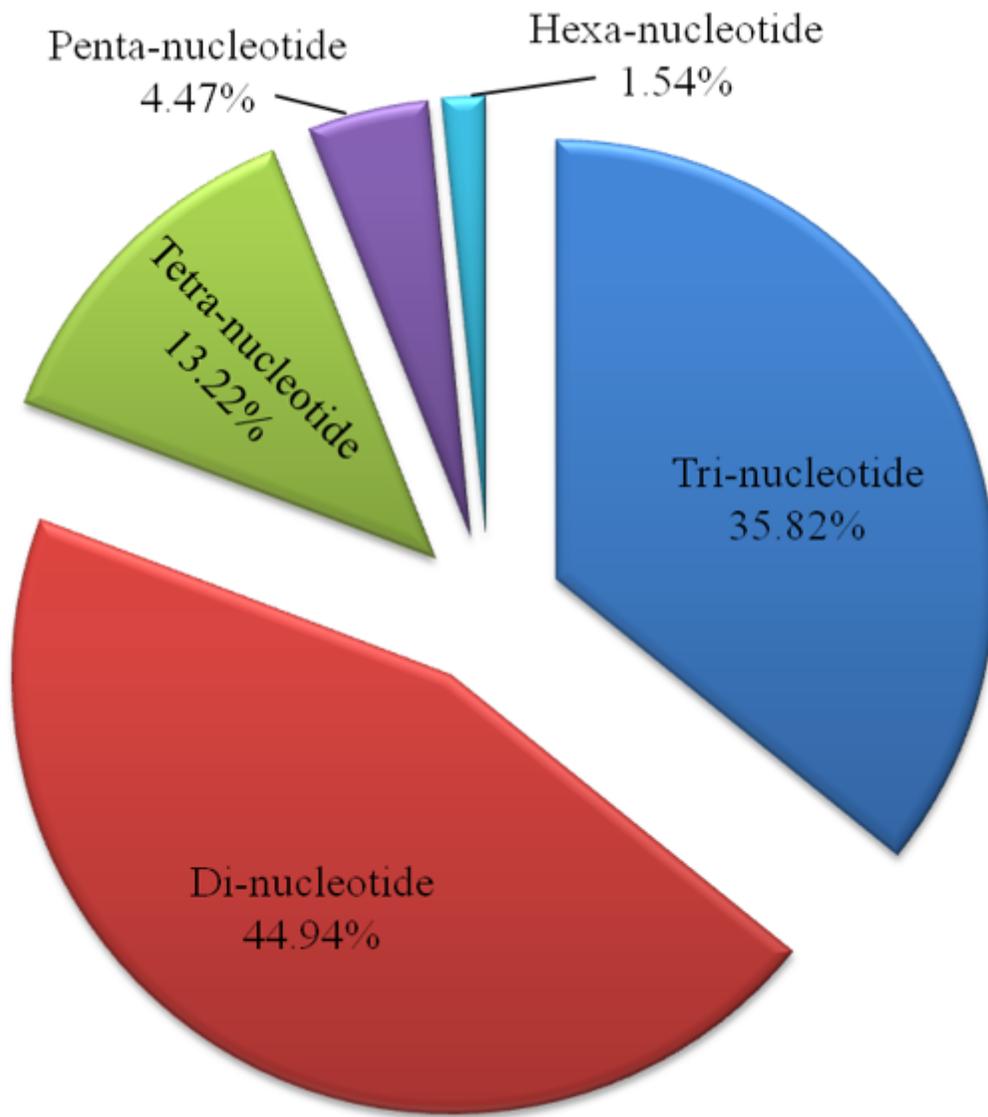
The morphological characteristics of common vetch cv "Lanjian No.1". a The plant. b The leaves. c The flowers. d The seeds. e The seedpods.



**Figure 2**

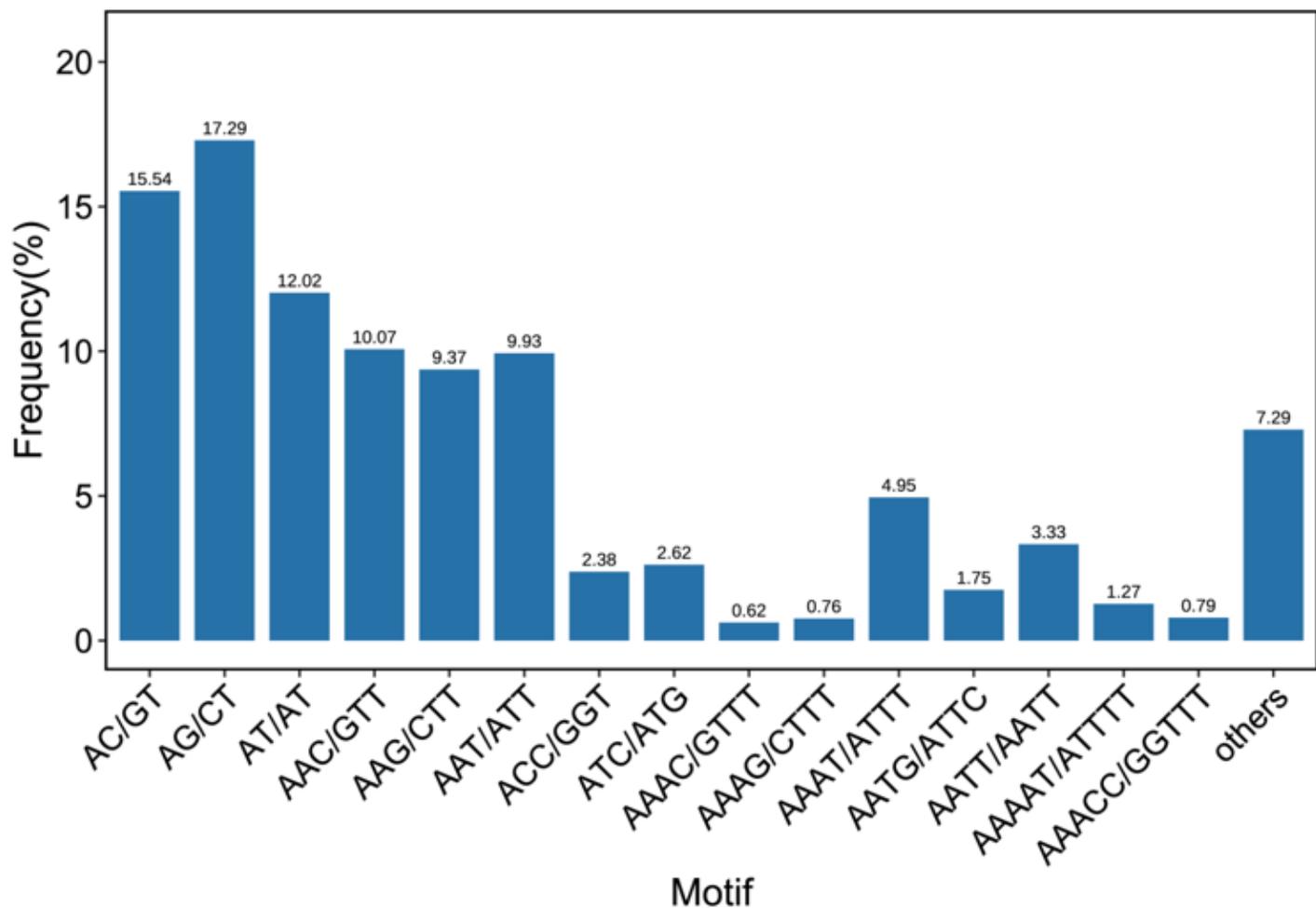
K-mer frequency distribution at K-mer = 17 depth and GC content and depth correction analysis. a The estimated genome size of common vetch was determined based on the following formula: genome size = K-mer depth. The x-axis is depth and y-axis represents the frequency at the particular depth divided by the total frequency of all depths. b The x-axis represents the GC content and the y-axis is the sequence depth.

The distribution of the sequence depth is on the right side, while the distribution of the GC content is at the top.



**Figure 3**

Distribution of 76,810 putative SSRs in common vetch genome based on repeat type.



**Figure 4**

The frequency distributions of representative SSR motifs in common vetch. The x-axis and y-axis represent motif type and the motif frequency, respectively.

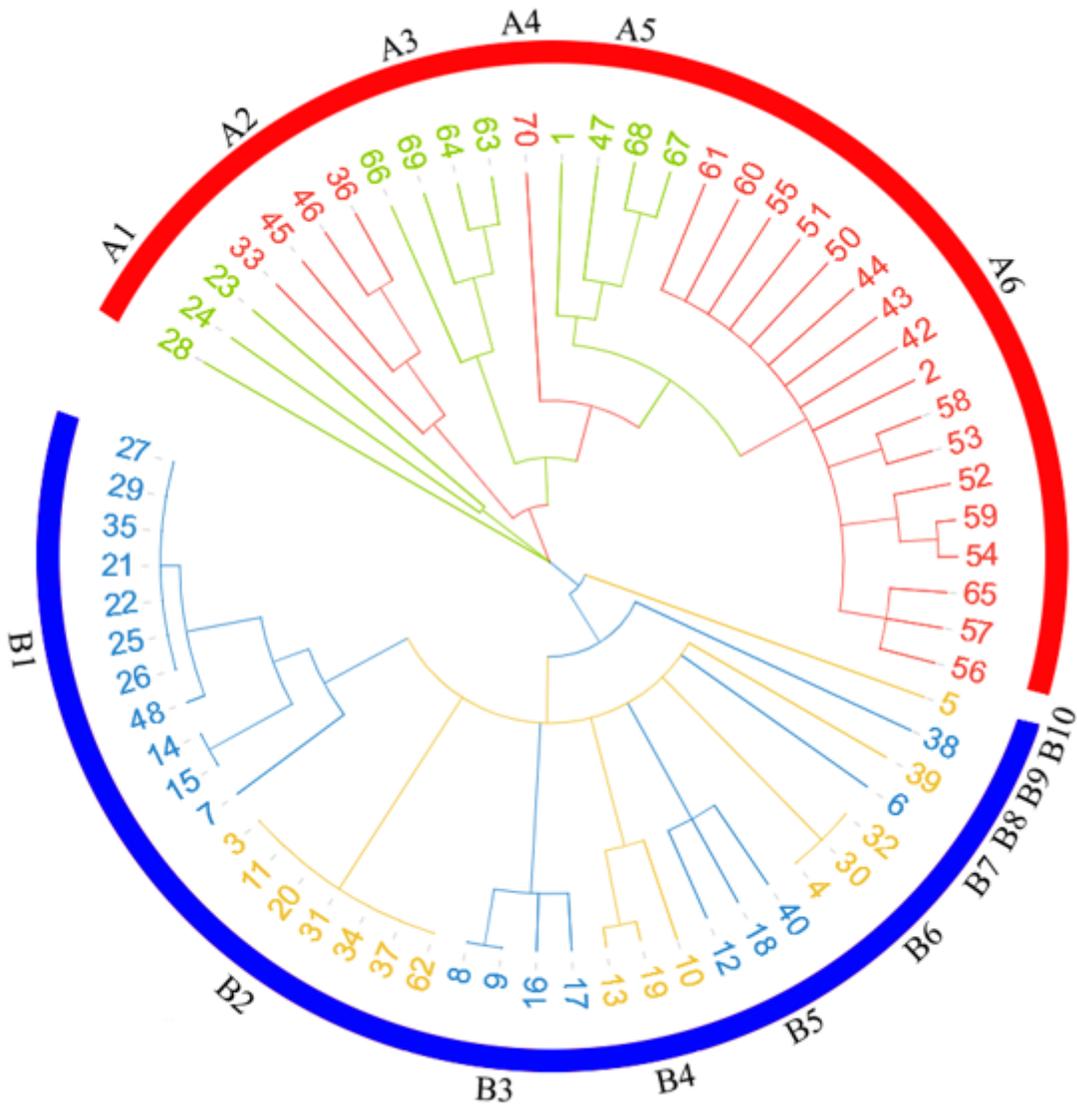


Figure 5

Cluster diagram for 68 individuals of Chinese common vetch by UPGMA method.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplymentarytables.xls](#)