

Article

Studying the Genetic Diversity of Yam Bean Using a New Draft Genome Assembly

Cassandra G. Tay Fernandez ¹, Kalidas Pati ^{1,2}, Anita A. Severn-Ellis ¹ , Jacqueline Batley ¹ 
and David Edwards ^{1,*} 

¹ The School of Biological Sciences, Institute of Agriculture, University of Western Australia, Perth, WA 6009, Australia; cassandria.tayfernandez@research.uwa.edu.au (C.G.T.F.); kalidas9555@gmail.com (K.P.); anita.severn-ellis@uwa.edu.au (A.A.S.-E.); jacqueline.batley@uwa.edu.au (J.B.)

² Regional Centre, ICAR-Central Tuber Crops Research Institute, Bhubaneswar 751019, Odisha, India

* Correspondence: dave.edwards@uwa.edu.au

Abstract: Yam bean (*Pachyrhizus erosus* Rich. Ex DC.) is an underutilized leguminous crop which has been used as a food source across central America and Asia. It is adapted to a range of environments and is closely related to major leguminous food crops, offering the potential to understand the genetic basis of environmental adaptation, and it may be used as a source of novel genes and alleles for the improvement of other legumes. Here, we assembled a draft genome of *P. erosus* of 460 Mbp in size containing 37,886 gene models. We used this assembly to compare three cultivars each of *P. erosus* and the closely related *P. tuberosus* and identified 10,187,899 candidate single nucleotide polymorphisms (SNPs). The SNP distribution reflects the geographic origin and morphology of the individuals.

Keywords: *Pachyrhizus erosus*; draft genome assembly; genetic diversity; SNPs; phylogenetic trees



Citation: Tay Fernandez, C.G.; Pati, K.; Severn-Ellis, A.A.; Batley, J.; Edwards, D. Studying the Genetic Diversity of Yam Bean Using a New Draft Genome Assembly. *Agronomy* **2021**, *11*, 953. <https://doi.org/10.3390/agronomy11050953>

Academic Editor: Tristan Edward Coram

Received: 14 April 2021

Accepted: 10 May 2021

Published: 11 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Yam bean (*Pachyrhizus erosus* Rich. Ex DC.) is a diploid leguminous crop grown for its starchy tuberous root, which is high in vitamins and minerals, particularly vitamin C, iron, zinc and potassium [1]. The genus *Pachyrhizus* belongs to the subtribe Glycininae, which includes the major legume crop, soybean (*Glycine max*) [2]. Yam bean is a regionally important crop in Mexico and Southeast Asia, where it is popular in traditional dishes or eaten raw. It produces a high yield in a small area of farmland (Mexico produces ~160 tons per hectare) and flourishes in humid conditions [3].

Yam bean has not been extensively studied, however, it has significant potential for improvement, both as a durable leguminous food crop for major agro-ecologies [4], as well as a potential donor of traits to related major legumes, such as insect resistance and abiotic stress tolerance [5]. Three species of yam bean have been cultivated for human consumption; Mexican yam bean (*P. erosus*), Andean yam bean (*P. ahipa*) and Amazonian yam bean (*P. tuberosus*). These three species can intercross to produce fertile, interspecific hybrids [6].

P. tuberosus and *P. erosus* are the most widely consumed species of yam bean and each of these has diverse morphologies within the species and the genus. *P. tuberosus* plants are generally larger than other *Pachyrhizus* species, with plump and kidney-shaped seeds, and wing and keel petals covered in minute hairs [7]. The skin colour of the root of *P. tuberosus* varies from whitish cream to brownish grey depending on the cultivar, and can have elongated buddings or more round offshoots. In contrast, *P. erosus* has petals that are smooth and free from hair, with flat and square seeds and pods that are either covered in short, stiff hairs or smooth at maturity [8]. *P. erosus* is primarily beige in colour and different cultivars have varying shapes, ranging from smooth and circular to more irregular, flower-shaped base (Supplementary Figure S1).

An early study on the relationship between cultivated *Pachyrhizus* taxa used restriction fragment analysis on chloroplast DNA (cpDNA) and Random Amplified Polymorphic DNA (RAPD) molecular markers [9]. This established the separation of the genus into two evolutionary branches, Mesoamerican and South American, reflecting the pattern of the species' distribution. Phylogenies constructed from sequence variants of the internal transcribed spacer (ITS) region of ribosomal DNA supported the cpDNA phylogeny [9]. These phylogenies also suggest that *P. panamensis*, *P. tuberosus* and *P. erosus* may have originated from rapid radiation of a continuously distributed ancestor, diverging due to varying climates, domestication and human selection [9]. *P. tuberosus* was found to be likely ancestral to *P. erosus* as a separate lineage, although this was not conclusive [9].

More recently, Delêtre et al. [10] developed simple sequence repeat (SSR) loci to investigate the intraspecific diversity and interspecific relationships between members of the *Pachyrhizus* genus and found loci across all yam bean species that showed high levels of polymorphism. The markers were able to distinguish varietal groups within each species and showed *P. ahipa* possessed a relatively low level of genetic variability. Santayana et al. [11] examined correlations between environmental factors and phylogeographic diversity patterns. The study also identified two distinct lineages between the Andean and Amazonian landraces using cpDNA sequences and SSR molecular markers, and the authors came to similar conclusions as Englemann et al. [9] that in *Pachyrhizus*, climate and genotypes are closely correlated.

In addition to the molecular analyses, Zanklan et al. [12] performed a genetic diversity study in cultivated yam bean using quantitative and qualitative physical traits, including size, weight and yield of various parts of the plant. The study documented the percentage of variation in each of the plants' features, such as storage root, seeds, pods, flowers, stems and leaves, linking some of the observed diversity to geographical origin. Overall, the study identified a high level of diversity on the intraspecific scale but overall lower diversity on the interspecific level. Wide differentiation was found among accessions and lines between the three tested yam bean species. The close relationship among *P. erosus*, *P. tuberosus* and *P. ahipa* indicated that only eight highly heritable characters are needed to describe the diversity within yam bean.

Here, we describe the first published draft genome assembly of *Pachyrhizus erosus*, one of the Mesoamerican yam bean species and show how it can be used to infer phylogenetic relationships. We re-sequenced two species of yam bean representing each one of the two evolutionary branches [9], *P. erosus* and *P. tuberosus* (representing the South American branch) and compared three cultivars of each. We also compared the gene content with related legumes. The assembly produced can be used as a foundation for future work and our results can be used as a stepping stone for describing the genetic content of yam bean.

2. Materials and Methods

2.1. Plant Material

P. erosus for the reference assembly was grown from seeds obtained from Eden Seeds (Lower Beechmont, QLD, Australia). Seeds for 3 additional *P. erosus* cultivars (Pe-CIP-209016, Pe-CIP-209046 and Pe-CIP-209051) and 3 *P. tuberosus* cultivars (Pe-CIP-209016, Pe-CIP-209046 and Pe-CIP-209051) were obtained from the International Potato Centre in Peru (Centro Internacional de la Papa, La Molina, Lima, Peru) under the conditions of the Standard Material Transfer Agreement of the International Treaty on Plant Genetic Resources for Food and Agriculture. Five plants for each of the accessions obtained were grown until large apple- to melon-sized tubers were formed to verify morphological uniformity.

2.2. DNA Extraction and Sequencing

Leaf tissue was collected and flash frozen for DNA extraction. DNA was extracted from only one plant for the reference as well one plant for each of the 3 cultivars from the CIP *P. erosus* and *P. tuberosus* cultivars using the Qiagen DNeasy Plant kit (Qiagen, Hilden, Germany). The DNA concentration was determined with the Qubit fluorometer

(Invitrogen, Carlsbad, CA, USA). DNA quality was confirmed using the LabChip GX Touch (PerkinElmer, Waltham, MA, USA) using the HT DNA gDNA reagents and the LabChip GX Touch (PerkinElmer, Waltham, MA, USA) while DNA purity was confirmed using the NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA 02451, USA).

Library preparation and sequencing of the yam bean reference was carried out by the Kinghorn Centre for Clinical Genomics Sequencing Facility (Darlinghurst, NSW, Australia). Sequencing libraries were prepared using the TruSeq Nano DNA HT Library Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol, and sequenced on the HiSeq X (Illumina, San Diego, CA, USA), generating 150 bp paired end sequencing data.

Sequencing libraries for the remaining *P. erosus* and *P. tuberosus* cultivars were prepared in house using the Illumina Nextera™ DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Library integrity was confirmed using the LabChip GX Touch 24 (PerkinElmer, Waltham, MA, USA) using the HT DNA HiSens Dual Protocol Reagents run on the 24 DNA Extended Range Chip (PerkinElmer, Waltham, MA, USA). The libraries were paired end (PE) sequenced on the HiSeq X (Illumina, San Diego, CA 92122, USA) at the Kinghorn Centre for Clinical Genomics' Sequencing Facility (Darlinghurst, NSW, Australia).

2.3. Data Generation and Validation

The yam bean reference genome was assembled using MaSuRCA v3.2.2 (Maryland Super-Read Celera Assembler) (University of Maryland, College Park, MD, USA) [13] and using a Jellyfish hash size of 50,000,000 with a mean insert size of 400 bp and a standard deviation of 100 bp [14]. The data were assembled using a k-mer size of 31 [15]. A custom script was then used to remove contigs that were shorter than 1 kbp. GenomeScope V.2 (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA) was used to predict the size of the assembly with a k-mer size of 31 bp and an expected coverage of 230× [16].

BUSCO v4 (Benchmarking Universal Single-Copy Orthologs) (SIB, Lausanne, Switzerland) was used to benchmark the quality of the assemblies by calculating the number of orthologues present [17]. The provided script, run_BUSCO.py, was used to compare the genome assembly with the BUSCO plant reference, *embryophyta_odb9* using the -m geno flags. Results were visualised using the generate_plot.py script with the -wd flags.

2.4. Annotation

AUGUSTUS [18] was used to annotate the genome using the *Arabidopsis* extrinsic detection files from the program Exonerate [19]. A custom script, peptideripper.pl, made a peptide file from this annotation, which was processed through OrthoFinder [20]. The custom script, orthogroup_filter.py was used on the produced file, Orthogroups.tsv to obtain the designations of the high-quality genes. The high-quality genes were extracted from the orthogroup sequences using awk and compared with the UniRef database [21] using diamond BLASTp with an E-value of 0.05 [22].

2.5. Phylogenetic Gene Trees

Assemblies and annotations for *Cajanus cajan*, *Cicer arietinum*, *Glycine max*, *Lotus japonicus*, *Lupinus angustifolius*, *Medicago truncatula*, *Phaseolus vulgaris*, *Trifolium pratense*, *Vigna angularis*, *Vigna radiata* and *Vigna unguiculata* were obtained from <https://legumeinfo.org/> (first accessed on 20 June 2018) [23] and Ensembl genomes (release 45), corresponding to the genome assemblies used in the analysis.

OrthoFinder v2.2.7 was used to compare the yam bean annotation with other legume annotations [20]. The orthogroup file was filtered to remove the orthogroups that only contained genes from yam bean, and UpSetR [24] was used to visualise the orthogroup size with respect to species. A randomly chosen orthogroup containing only single copy orthologs was aligned with MUSCLE (multiple sequence comparison by log-expectation) (Mill Valley, CA 94941, USA) [25] and clustered using the pal2nal.pl script. The file was

converted into a Phylogenetic Analysis by Maximum Likelihood (PAML) file and displayed using the interactive Tree Of Life (iTOL) program [26] (Figure 1).

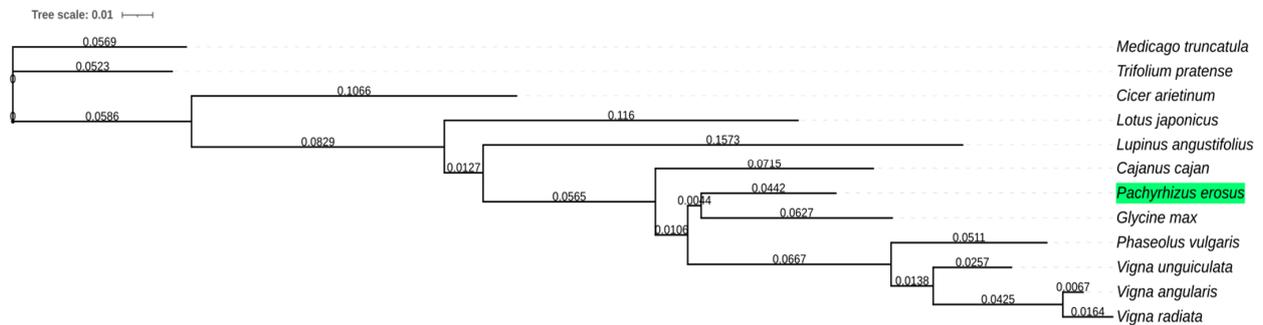


Figure 1. A phylogeny tree based on orthologous protein groups identified in OrthoFinder. Yam bean (*Pachyrhizus erosus*) is highlighted in green. Out of the legume species selected, yam bean is predicted to be most closely related to *Glycine max*.

Publicly available sequencing data of *Glycine tomentella* (SRR1176933), *Glycine dolichocarpa* (SRR1174380), *Glycine syndetika* (SRR1176843), *Glycine max* (SRR5479841), *Glycine soja* (SRR1030612), *Glycine gracilis* (SRR1028190), *Phaseolus vulgaris* (SRR8191436), *Vigna unguiculata* (SRR3716700), *Lupinus angustifolius* (SRR2869724) and *Sphenostylis stenocarpa* (SRR5647410 and SRR5647389) were obtained from the Sequence Read Archive (SRA) [27]. A Mash sketch [28] was calculated for each file using a minimum k-mer copy number of 3 (-r -m 3). The Mash distances were then calculated for each pair of sequencing data and plotted into a phylogeny tree using the Phangorn package in R [29]. The MaSuRCA assembly was indexed using the Burrows–Wheeler aligner (BWA) [30].

The tools in the package Bcftools were used to perform variant calling (standard settings). Mpileup was used to produce a BCF file with all locations in the genome to call genotypes and produce a list of variant sites. The produced statistics were visualised using the provided script plot-vcfstats. A phylogeny tree was produced from the bedtools vcf file using the R programs gdsfmt and SNPRelate [31] (Figure 2).

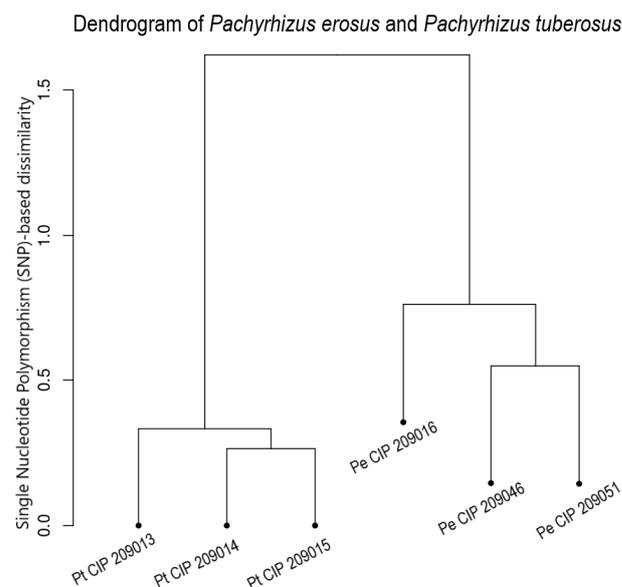


Figure 2. A single nucleotide polymorphism (SNP) phylogenetic cluster tree of six yam bean cultivars. Pt: *Pachyrhizus tuberosus*, Pe: *Pachyrhizus erosus*. The *Pachyrhizus tuberosus* cultivars were all sourced from Peru, while Pe-CIP-209016, Pe-CIP-209046 and Pe-CIP-209051 were sourced from Guatemala, Cartago Costa Rica, and Mexico, respectively.

3. Results and Discussion

A total of 145 Gbp of Illumina HiSeqX sequencing data were generated for the reference. We estimated the *Pachyrhizus erosus* genome size to be 560 Mbp based on a k-mer size of 31. This is similar to the 550 Mbp predicted based on flow cytometry measurements [32]. Assembly of the Illumina sequence data resulted in contigs totalling 460 Mbp, with an N50 of 18,412, representing 83.6% the predicted genome size (Supplementary Table S1). BUSCO analysis identified 1329 (94.9%) of the 1440 *embryophyta_obdb9* plant orthologues in the assembly (Supplementary Table S2), suggesting that while the assembly is fragmented, it contains a similar level of completeness as comparable legume assemblies such as pea [33], *Glycine max* [34], *Medicago sativa* [35] and *Vigna subterranea* [36]. Annotation of the assembly predicted 37,886 genes, similar to other related diploid legume genomes, and a total of 27,575 (72.78%) of the encoded proteins share sequence identity with predicted proteins from other legumes.

To assess diversity within and between *Pachyrhizus* species, we generated a total of 17 Gbp of sequence data each for three *P. erosus* individuals, Pe-CIP-209016, Pe-CIP-209046 and Pe-CIP-209051 originating from Guatemala, Costa Rica and Mexico respectively, as well as three Peruvian *P. tuberosus* individuals, Pt-CIP-209013, Pt-CIP-209014 and Pt-CIP-209015 (Table 1). Mapping this data to our reference assembly identified 10,187,899 single nucleotide polymorphisms (SNPs) across the genome.

Table 1. Accession number, origin and species identity of the sampled yam bean cultivars. ‘Accession number’ is the tag used to designate each species, ‘Origin’ documents where the original plant seeds come from, and ‘*Pachyrhizus* species’ describes which species the cultivars are. All the species were obtained from the International Potato Center (CIP) in Peru.

No.	Accession	Origin	<i>Pachyrhizus</i> Species	Total Gbp	Sequence Coverage
13	Pt-CIP-209013	Peru	<i>P. tuberosus</i>	16.51	35.89×
14	Pt-CIP-209014	Peru	<i>P. tuberosus</i>	12.19	26.49×
15	Pt-CIP-209015	Peru	<i>P. tuberosus</i>	16.79	36.50×
16	Pe-CIP-209016	Guatemala	<i>P. erosus</i>	21.06	45.78×
46	Pe-CIP-209046	Cartago Costa Rica	<i>P. erosus</i>	16.05	34.89×
51	Pe-CIP-209051	Mexico	<i>P. erosus</i>	19.18	41.69×

Pachyrhizus is within the Millettoid/Phaseoloid clade from the Faboideae subfamily of the Phaseoleae tribe, which also encompasses members of the genera *Glycine*, *Cicer*, *Cajanus*, *Phaseolus* and *Vigna* [37]. We generated a phylogenetic tree comparing the yam bean peptide sequences to other Phaseoleae species (*Glycine max*, *Glycine tomentella*, *Glycine dolichocarpa*, *Glycine sydetika*, *Glycine soja*, *Glycine gracilis*, *Phaseolus vulgaris*, *Vigna unguiculata*, *Lupinus angustifolius* and *Sphenostylis stenocarpa*). The tree supports *Glycine max* (and, by extension, the *Glycine* family) as being the closest major crop relative to yam bean, which had also been shown using chloroplast sequences [38] (Figure 1). A second tree was constructed based on k-mer sketches comparing yam bean to related leguminous species (Figure 3) and further supported this relationship. This corresponds with previous analyses based on plastid genome and DNA sequences, which supports the assertion that *Pachyrhizus* is closely related to *Glycine* [39]. A strict consensus cladogram generated from 7006 equally parsimonious Wagner cpDNA trees placed *Pachyrhizus*, then in the subtribe Diocleinae, as closely related to the subtribe Glycininae [40] and suggested that it should be placed into the Glycininae subtribe, which is supported by Polhill’s [41] cpDNA restriction study. Maximum likelihood phylogeny trees of legume species generated from 71 protein-coding genes also inferred that *Pachyrhizus* was closely related to *Glycine*, despite not being monophyletic with it [42].

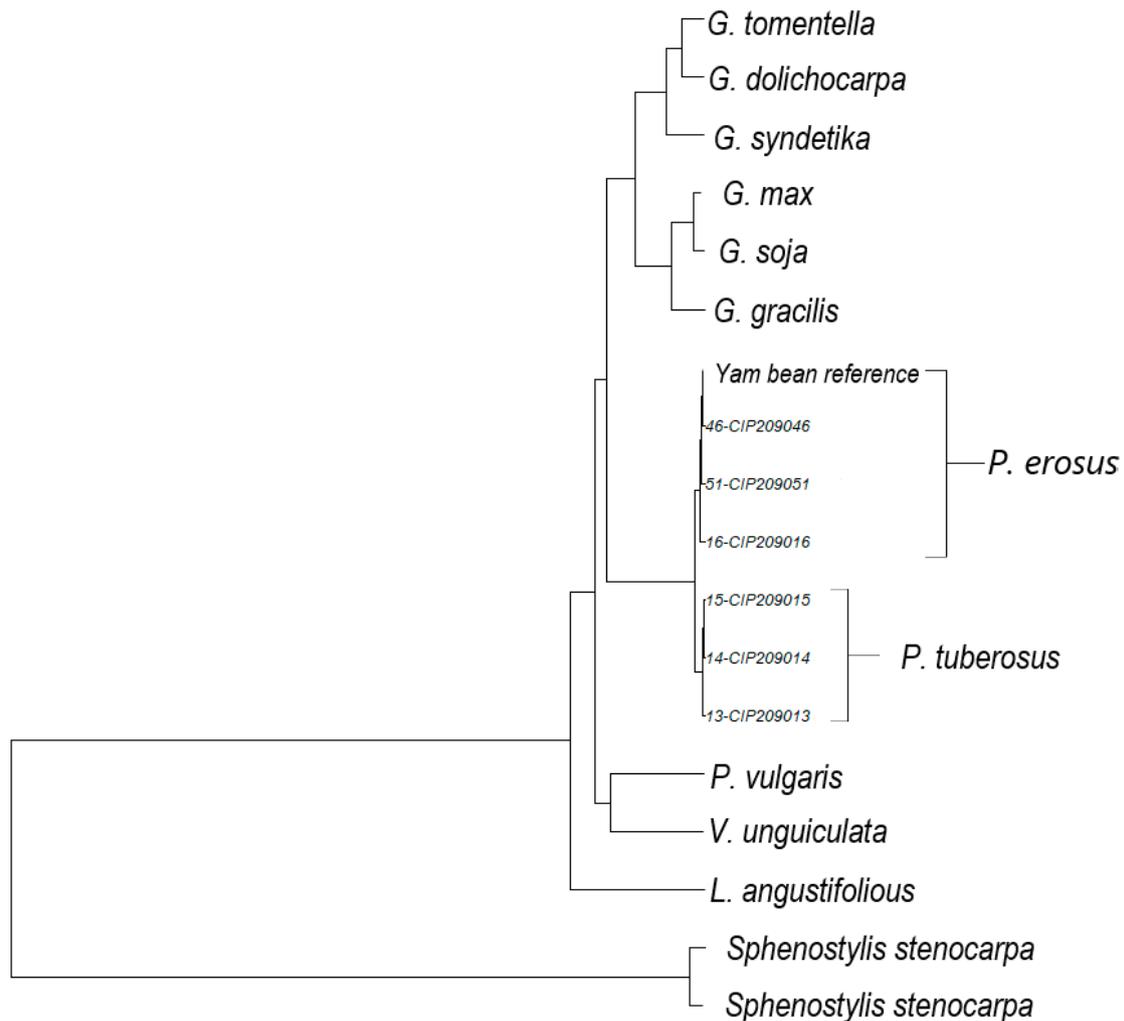


Figure 3. A Mash rooted phylogeny tree displaying assorted members from the Glycine family and other related legumes. The closest relative to either species of yam bean (*Pachyrhizus erosus* or *Pachyrhizus tuberosus*) is *Glycine max*, *Glycine soja* and *Glycine gracilis*.

The SNP cluster tree distinctly grouped the six individuals by the species of yam bean (Figure 2). SNP-based dissimilarity suggests that the two yam bean species have small but distinct differences. *P. tuberosus* and its relative *P. ahipa* are the only two South American-originating species in the *Pachyrhizus* genus [43]. The clustering performed is similar to the results of Tapia and Sørensen [43], showing that *P. tuberosus* has been cultivated for thousands of years in South America, with accessions demonstrating little morphological variation.

There is still uncertainty regarding the origin of current landraces and cultivars of *P. erosus* in Mexico and Guatemala, as *P. erosus* has no clear subgroups related to its geographical origin [9,44]. Central American and Mexican landraces of *P. erosus* were cultivated by the Mayans and Aztecs but appear to have different origins based on preliminary molecular analyses [5]. Our analysis suggests that within *P. erosus*, Pe-CIP-209016 could have diverged from Pe-CIP-209046 and Pe-CIP-209051 (Figure 2).

Englemann [9] examined cpDNA genetic marker variation within and between *Pachyrhizus* taxa. The consensus trees suggested that *P. tuberosus* was involved in the early parentage of *P. erosus*, likely stemming from an early lineage of *P. tuberosus*. However, the results are inconsistent with an analysis of the cpDNA and nuclear internal transcribed spacer (ITS) sequence variation [9]. These consensus trees were not supported by our analysis, which does not suggest a direct origin of *P. erosus* from *P. tuberosus* (Figure 2).

Wild *P. erosus* species are found across a wide range of geographical habitats, including wet forests (Mexico), deciduous forests (Guatemala) and dry savannahs (Costa Rica) [9]. The geographic distribution of the *P. erosus* individuals was reflected in the SNP distribution. Pe-CIP-209046 (Costa Rica) was found to be the most evolutionarily distant of the six yam bean individuals examined, and it was the only species to come from a dry, arid area. However, this was not reflected in the phylogenetic tree (Figure 3). A potential reason for this is that Pe-CIP-209046 is from Engelmann's [9] proposed Central American groups, where accessions are exclusively from Central America, and has SNPs shared with individuals also stemming from the group. Pe-CIP-209016 is likely from an accession that has both Central American and Mexican origins, making it distant from Pe-CIP-209046 and Pe-CIP-209051. Additional data may help to estimate the extent of the divergence between *P. erosus* from Costa Rica and *P. erosus* originating elsewhere.

Domestication of *P. tuberosus* first took place in the Peruvian Andes, and has a long history of being cultivated in various locations in South America, including Paraguay and Bolivia [5,7]. The *Pachyrhizus* genus demonstrates high levels of diversity within the species [5,9,12]. The *P. tuberosus* individuals in this study have been determined to be a more genetically similar group than the *P. erosus* individuals based on their SNP distribution.

The genetic and phenotypic diversity within *P. erosus* could be due to geographic isolation and domestication, which is in line with previous studies showing that yam bean genotypes are associated with their respective environment [5,6,10]. *P. tuberosus* and *P. erosus* also have genotype-by-environment interactions associated with fresh storage root yield [45] and crop yield [46]. Engelmann's work [9], shows that *P. erosus* accessions could be divided into four groups based on RAPD analysis: a group of *P. erosus* that originated from Central America only; two groups that are from Central America and Mexico; and a group that was exclusively made up of accessions from Mexico.

In conclusion, the *P. erosus* assembly has allowed us to infer phylogenetic links within the *Pachyrhizus* species and other legumes. The assembly has shown that *P. erosus* is more genetically distinct than *P. tuberosus*, and adds support to the *Glycine* family being the closest major crop relative to the *Pachyrhizus* species. Our analysis suggests that *P. erosus* from Costa Rica is not as evolutionarily distinct from other *P. erosus* species as expected, although more data are needed to support this conclusion.

Pachyrhizus species are mostly self-pollinating, but some crossbreeding does occur (2–4%), depending on the availability of pollinators [5]. While some research has been carried out on its genetic traits, more data are needed to further investigate the evolutionary origins of the plant and to identify genetic loci associated with adaptation and agronomic traits. Here, we have assembled the first draft genome of *P. erosus* and used it to study the phylogeny of geographically diverse *Pachyrhizus* lines. We determined the location of the *Pachyrhizus* genus in relation to other legumes and how geography affects the SNP distribution. We also annotated 37,886 genes and over 10 million SNPs. This draft genome assembly can be used as a foundation for future yam bean genomic analyses of this underutilised crop.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/agronomy11050953/s1>, Figure S1: Yam bean root images including the origin and accession numbers of each cultivar, which were grown and investigated at the University of Western Australia. (S1) *P. tuberosus* (Pt-CIP-209013) from Peru. (S2) *P. tuberosus* (Pt-CIP-209014) from Peru. (S3) *P. tuberosus* (Pt-CIP-209015) from Peru. (S4) *P. erosus* (Pe-CIP-209016) from Guatemala. (S5) *P. erosus* (Pe-CIP-209046) from Cartago Costa Rica. (S6) *P. erosus* (Pe-CIP-209051) from Mexico, Table S1: Assembly statistics of the yam bean genome assembled by MaSuRCA 3.2.2. N50 of the assembly was 6687 across 18,412 sequences, Table S2: The BUSCO (Benchmarking Universal Single-Copy Orthologs) analysis results for the yam bean MaSuRCA assembly, Table S3: The singleton/non-duplicated statistics of the yam bean genome assembled by MaSuRCA 3.2.2. compared to other six other *Pachyrhizus* cultivars using bcftools.

Author Contributions: D.E. and J.B. designed the experiments and coordinated the project. K.P. sourced and prepared the plant material. A.A.S.-E. prepared the plant material and generated the DNA sequence data. C.G.T.F. performed the bioinformatics analyses. D.E. and C.G.T.F. wrote the manuscript with contributions from the other authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was undertaken with the assistance of resources provided at the Pawsey Supercomputing Centre and was partially funded by the Australia Research Council (Projects DP200100762 and LP160100030). C.T.F. was supported by an IPRS awarded by the Australian government. K.P. was supported by an Endeavour Research Fellowship, Australia Awards, Australian Government Department of Education and Training.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw sequencing reads are available from BIOPROJ. The assembly and annotation is available from http://www.appliedbioinformatics.com.au/index.php/Yam_bean_assembly (accessed on 1 May 2021).

Acknowledgments: We would like to acknowledge the International Potato Centre which provided the seeds.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Gosálvez, M. Carcinogenesis with the insecticide rotenone. *Life Sci.* **1983**, *32*, 809–816. [[CrossRef](#)]
- Lee, J.; Hymowitz, T. A Molecular Phylogenetic Study of the Subtribe Glycininae (Leguminosae) Derived from the Chlo-roplast DNA rps16 Intron Sequences. *Am. J. Bot.* **2001**, *88*, 2064–2073. [[CrossRef](#)] [[PubMed](#)]
- National Research Council. *Lost Crops of Africa; Volume II: Vegetables*; National Academies Press: Washington, DC, USA, 2006.
- Agaba, R.; Tukamuhabwa, P.; Rubaihayo, P.; Mwangi, R.; Ssenyonjo, A.; Ndirigwe, J.; Tumwegamire, S.; Grüneberg, W. Heritability, combining ability and inheritance of storage root dry matter in yam beans. *Afr. Crop. Sci. J.* **2017**, *25*, 83. [[CrossRef](#)]
- Sørensen, M. *Yam Bean (Pachyrhizus DC.)*. *Promoting the Conservation and Use of Underutilized and Neglected Crops*; Institute of Plant Genetics and Crop Plant Research: Gatersleben, Germany; International Plant Genetic Resources Institute: Rome, Italy, 1996.
- Grüneberg, W.; Freynhagen-Leopold, P.; Delgado-Vázquez, O. A new yam bean (*Pachyrhizus* spp.) interspecific hybrid. *Genet. Resour. Crop. Evol.* **2003**, *50*, 757–766. [[CrossRef](#)]
- Sørensen, M.; Døyggaard, S.; Estrella, J.E.; Kvist, L.P.; Nielsen, P.E. Status of the South American tuberous legume *Pachyrhizus tuberosus* (Lam.) Spreng: Field observations, taxonomic analysis, linguistic studies and agronomic data on the diversity of the South American *Pachyrhizus tuberosus* (Lam.) Spreng. complex with special reference to the identification of two new cultivar groups from Ecuador and Peru. *Biodivers. Conserv.* **1997**, *6*, 1581–1625. [[CrossRef](#)]
- Sørensen, M. A taxonomic revision of the genus *Pachyrhizus* (Fabaceae-Phaseoleae). *Nord. J. Bot.* **1988**, *8*, 167–192. [[CrossRef](#)]
- Engelmann, J. Molecular Systematics of the Neotropical Tuberous Legume ‘*Pachyrhizus*’ Rich. ex DC, the Yam Bean. In *Environmental and Evolutionary Biology*; University of St Andrews: St Andrews, UK; ProQuest Dissertations Publishing: Morrisville, NC, USA, 1998.
- Delêtre, M.; Soengas, B.; Utge, J.; Lambourdière, J.; Sørensen, M. Microsatellite Markers for the Yam Bean *Pachyrhizus* (Fabaceae). *Appl. Plant Sci.* **2013**, *1*, 1200551. [[CrossRef](#)] [[PubMed](#)]
- Santayana, M.; Rossel, G.; Núñez, J.; Sørensen, M.; Delêtre, M.; Robles, R.; Fernández, V.; Grüneberg, W.J.; Heider, B. Molecular characterization of cultivated species of the genus *Pachyrhizus* Rich. ex DC. by AFLP markers: Calling for more data. *Trop. Plant Biol.* **2014**, *7*, 121–132. [[CrossRef](#)]
- Zanklan, A.S.; Becker, H.C.; Sørensen, M.; Pawelzik, E.; Grüneberg, W.J. Genetic diversity in cultivated yam bean (*Pachyrhizus* spp.) evaluated through multivariate analysis of morphological and agronomic traits. *Genet. Resour. Crop. Evol.* **2017**, *65*, 811–843. [[CrossRef](#)]
- Zimin, A.V.; Marçais, G.; Puiu, D.; Roberts, M.; Salzberg, S.L.; Yorke, J.A. The MaSuRCA genome assembler. *Bioinformatics* **2013**, *29*, 2669–2677. [[CrossRef](#)]
- Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, *27*, 764–770. [[CrossRef](#)]
- Rahman, A.; Hallgrímsson, I.; Eisen, M.; Pachter, L. Association mapping from sequencing reads using k-mers. *eLife* **2018**, *7*, e32920. [[CrossRef](#)]
- Vurture, G.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202–2204. [[CrossRef](#)] [[PubMed](#)]
- Waterhouse, R.M.; Seppey, M.; Simão, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **2018**, *35*, 543–548. [[CrossRef](#)]

18. Hoff, K.J.; Stanke, M. Predicting Genes in Single Genomes with AUGUSTUS. *Curr. Protoc. Bioinform.* **2018**, *65*, e57. [[CrossRef](#)]
19. Slater, G.S.C.; Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **2005**, *6*, 31. [[CrossRef](#)]
20. Emms, D.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157. [[CrossRef](#)]
21. Suzek, B.E.; Wang, Y.; Huang, H.; McGarvey, P.B.; Wu, C.H. The UniProt Consortium UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **2015**, *31*, 926–932. [[CrossRef](#)]
22. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [[CrossRef](#)]
23. Dash, S.; Campbell, J.D.; Cannon, E.K.S.; Cleary, A.M.; Huang, W.; Kalberer, S.R.; Karingula, V.; Rice, A.G.; Singh, J.; Umale, P.E.; et al. Legume information system (LegumeInfo.org): A key component of a set of federated data resources for the legume family. *Nucleic Acids Res.* **2016**, *44*, D1181–D1188. [[CrossRef](#)]
24. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**, *33*, 2938–2940. [[CrossRef](#)]
25. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 113. [[CrossRef](#)]
26. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [[CrossRef](#)]
27. Leinonen, R.; Sugawara, H.; Shumway, M.; on behalf of the International Nucleotide Sequence Database Collaboration. The Sequence Read Archive. *Nucleic Acids Res.* **2010**, *39*, D19–D21. [[CrossRef](#)] [[PubMed](#)]
28. Ondov, B.D.; Treangen, T.J.; Melsted, P.; Mallonee, A.B.; Bergman, N.H.; Koren, S.; Phillippy, A.M. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **2016**, *17*, 132. [[CrossRef](#)]
29. Schliep, K.P. phangorn: Phylogenetic analysis in R. *Bioinformatics* **2011**, *27*, 592–593. [[CrossRef](#)]
30. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)] [[PubMed](#)]
31. Zheng, X.; Levine, D.; Shen, J.; Gogarten, S.M.; Laurie, C.; Weir, B.S. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **2012**, *28*, 3326–3328. [[CrossRef](#)]
32. Pati, K.; Zhang, F.; Batley, J. First report of genome size and ploidy of the underutilized leguminous tuber crop Yam Bean (*Pachyrhizus erosus* and *P. tuberosus*) by flow cytometry. *Plant Genet. Resour.* **2019**, *17*, 456–459. [[CrossRef](#)]
33. Kreplak, J.; Madoui, M.-A.; Cápál, P.; Novák, P.; Labadie, K.; Aubert, G.; Bayer, P.E.; Gali, K.K.; Syme, R.A.; Main, D.; et al. A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* **2019**, *51*, 1411–1422. [[CrossRef](#)]
34. Xie, M.; Chung, C.Y.-L.; Li, M.-W.; Wong, F.-L.; Wang, X.; Liu, A.; Wang, Z.; Leung, A.K.-Y.; Wong, T.-H.; Tong, S.-W.; et al. A reference-grade wild soybean genome. *Nat. Commun.* **2019**, *10*, 1216. [[CrossRef](#)]
35. Chen, H.; Zeng, Y.; Yang, Y.; Huang, L.; Tang, B.; Zhang, H.; Hao, F.; Liu, W.; Li, Y.; Liu, Y.; et al. Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the auto-tetraploid cultivated alfalfa. *Nat. Commun.* **2020**, *11*, 2494. [[CrossRef](#)]
36. Chang, Y.; Liu, H.; Liu, M.; Liao, X.; Sahu, S.K.; Fu, Y.; Song, B.; Cheng, S.; Kariba, R.; Muthemba, S.; et al. The draft genomes of five agriculturally important African orphan crops. *GigaScience* **2019**, *8*, 152. [[CrossRef](#)]
37. Oyebanji, O.; Zhang, R.; Chen, S.-Y.; Yi, T.-S. New Insights into the Plastome Evolution of the Millettoid/Phaseolid Clade (Papilionoideae, Leguminosae). *Front. Plant Sci.* **2020**, *11*, 151. [[CrossRef](#)]
38. Doyle, J.; Chappill, J.; Bailey, C.; Kajita, T.; Herendeen, P.; Bruneau, A. *Advances in Legume Systematics, Part 9*; Royal Botanic Gardens Kew: Richmond, UK, 2000; pp. 1–20.
39. Lackey, J.A. A revised classification of the tribe Phaseoleae (Leguminosae: Papilionoideae), and its relation to canavanine distribution. *Bot. J. Linn. Soc.* **1977**, *74*, 163–178. [[CrossRef](#)]
40. Doyle, J.J.; Doyle, J.L. Chloroplast DNA Phylogeny of the Papilionoid Legume Tribe Phaseoleae. *Syst. Bot.* **1993**, *18*, 309–327. [[CrossRef](#)]
41. Polhill, R. Classification of the Leguminosae. In *Phytochemical Dictionary of the Leguminosae*; Harborne, J.B., Ed.; Chapman & Hall: Cambridge, UK, 1994; Volume 1, pp. XXXV–XLVIII.
42. Choi, I.S.; Choi, B.H. The distinct plastid genome structure of *Maackia fauriei* (Fabaceae: Papilionoideae) and its systematic implications for genistoids and tribe Sophoreae. *PLoS ONE* **2017**, *12*, e0173766. [[CrossRef](#)]
43. Tapia, C.; Sørensen, M. Morphological characterization of the genetic variation existing in a Neotropical collection of yam bean, *Pachyrhizus tuberosus* (Lam.) Spreng. *Genet. Resour. Crop. Evol.* **2003**, *50*, 681–692. [[CrossRef](#)]
44. Estrella, E.; Phillips, S.; Abbott, R.; Gillies, A.; Sørensen, M. Genetic Variation and Relationships in Agronomically Important Species of Yam Bean (*Pachyrhizus* DC.) Based on RAPD Markers. In *Acta Agriveterinaria, Proceedings of the 2. International Symposium on Tuberous Legumes, Celaya, Guanajuato, Mexico, 5–8 August 1996*; MacKeenzie: Dublin, UK, 1998; pp. 43–59.
45. Andiku, C.; Tukamuhabwa, P.; Ssebuliba, J.M.; Talwana, H.; Tumwegamire, S.; Grüneberg, W.J. Evaluation of the American Yam Bean (*Pachyrhizus* spp.) for Storage Root Yield Across Varying Eco-geographic Conditions in Uganda. *J. Agric. Sci.* **2019**, *11*, 100. [[CrossRef](#)]
46. Jean, N.; Patrick, R.; Phenihias, T.; Rolland, A.; Placide, R.; Robert, M.O.M.; Silver, T.; Vestine, K.; Evrard, K.; Grüneberg, W.J. Evaluation of Performance of Introduced Yam Bean (*Pachyrhizus* spp.) in Three Agro-Ecological Zones of Rwanda. *Trop. Plant Biol.* **2017**, *10*, 97–109. [[CrossRef](#)]