

A PacBio and Hi-C based proximity-guided assembly of *C. pallidicaule*.

Hayley Jennifer Hansen Mangelson

A prospectus submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Science

Jeff Maughan, Chair
Rick Jellen
David Jarvis
Brad Geary

Department of Plant and Wildlife Sciences
Brigham Young University

Copyright © 2018 Hayley Jennifer Hansen Mangelson

All Rights Reserved

TABLE OF CONTENTS

TITLE PAGE	1
TABLE OF CONTENTS	2
LITERATURE REVIEW	3
Introduction to <i>Chenopodium pallidicaule</i>	3
Genetic resources for <i>C. pallidicaule</i>	6
Advantages of a whole-genome assembly	8
METHODS.....	9
Proximity-Guided Assembly.....	9
Gap Closing with PacBio Long-reads and PBJelly.....	9
Proximity-Guided Assembly 2.....	10
Resequencing.....	10
SNP Discovery	11
Chloroplast Genome Reconstruction.....	11
Annotation	11
TIMELINE	11
LITERATURE CITED	13

LITERATURE REVIEW

Introduction to Chenopodium pallidicaule

Dating back to at least 5,000 BC, *C. pallidicaule* Aellen is a crop of import to ancient South American societies (Repo-Carrasco, Espinoza, and Jacobsen 2003). It is known as cañihua in Quechua and alternatively as cañigua, cañahua, cañawa, and kañiwa (Gade 1970). Following the Spanish conquest, cultivation was likely discouraged due to its association with Incan society in the minds of European colonists (Ruas et al. 1999). However, farmers in the Altiplano region have continued to grow this traditional crop due to its resistance to frost, drought, salinity, and pests in addition to its high nutritional quality (Gade 1970). It is grown alongside Andean tubers and traditional pseudocereals, such as quinoa (*C. quinoa* Wiild.) and kiwicha (*Amaranthus caudatus*, L.). In spite of the growing popularity of *C. quinoa*, cañihua remains practically unknown and wholly underutilized as a food source (Rastrelli et al. 1996).

The Andean Altiplano is the coldest and harshest portion of the 7,000 kilometer Andean mountain range (Ranilla et al. 2009). This semi-arid region (Peñarrieta et al. 2008a; Rastrelli et al. 1996) consists of stony, low pH (as low as 4.8) soils (Ruas et al. 1999), and is subject to frequent frosts and hails (Peñarrieta et al. 2008b), offering a challenging environment for plant growth. Cañihua is cultivated in two major centers proximal to Lake Titicaca [La Paz, Bolivia, and Puno, Peru (Vargas et al. 2011)] at high altitudes, ranging from 3.6 to 4.6 kilometers above sea level. High altitude increases exposure to damaging ultra-violet rays that further complicate plant growth (Peñarrieta et al. 2008b). Farmers in these regions depend on cañihua as a low-maintenance, tolerant (Ruas et al. 1999), alimentary crop that

will yield when other crops fail. It is often grown in rotation with tubers because of its ability to grow on marginal and depleted soils (Gade 1970; Rastrelli et al. 1996).

Sowing of cañihua occurs between September and November with the onset of the rainy season. The soil is typically hoed, then seed is broadcast by hand. Germination occurs around 4° C, followed by flowering at about 10°C and seed set at about 15°C. Little or no irrigation or fertilization is required for crop maintenance (Gade 1970; Peñarrieta et al. 2008b; Rastrelli et al. 1996).

A crop of cañihua will reach maturity in 95-173 days (depending on the ecotype), which is a substantially shorter lifespan than the other predominant pseudocereal, quinoa (Rodríguez et al. 2017; Vargas et al. 2011). In addition to differences in time to maturity, cañihua plants are also much shorter than quinoa, reaching merely 25-60cm. There are three main growth habits displayed across the species: erect, semiprostrate, and prostrate (Vargas et al. 2011). Because the crop is only partially domesticated, seeds do not mature at one time and seed shattering causes significant yield loss (Gade 1970; Rodríguez et al. 2017).

At the end of the growing season, mature plants are harvested by hand and piled on a blanket, where they are traditionally threshed by beating with curved sticks. They are left to sun-dry, then threshed once more. Typical yields range from 375 to 2,968 kilograms per hectare (Vargas et al. 2011). The seeds are then soaked in water to remove the pericarp and rinse off bitter saponins. They are then sun-dried and cooked or further processed (Gade 1970). Traditionally, cañihua seeds are added to soups (Rastrelli et al. 1996) or toasted and ground into a meal called kañiwako (Repo-Carrasco-Valencia et al. 2010). This can then be used to make mush, breads, cakes, and fermented drinks (Gade 1970; Rastrelli et al. 1996). Stalks and chaff are often saved and used as animal feed (Gade 1970; Ruas et al. 1999).

Like its close relative, quinoa, cañihua has a unique nutritional profile that makes it an excellent food source. The protein content of these Andean grains has been of particular interest. Cañihua contains 15-18% protein, with a complete set of essential amino acids, including 5-6% lysine, which is typically limited in grains (Peñarrieta et al. 2008b). In many native areas, quinoa and cañihua are principle protein sources because of a lack of animal protein available to the indigent people. According to Ranilla et al. (2009), the protein quality is at least equal to milk. With a poverty rate of nearly 50% in the rural highlands of the Altiplano, cañihua is an incredibly important resource in the prevention of poverty-induced malnutrition (Repo-Carrasco, Espinoza, and Jacobsen 2003).

In addition to high quality protein, cañihua also offers a wide variety of antioxidants, phenolic compounds, and flavonoids. The increase in antioxidants is likely due to the high altitude cultivation and exposure to intense ultra-violet radiation and, therefore, free radicals (Peñarrieta et al. 2008b). The appreciable concentrations of antioxidants and phenolic compounds mean cañihua may have considerable value for human nutrition. Repo-Carrasco-Valencia et al. (2010) compared flavonoid concentrations to berries, which are known to have very high flavonoid content, and found that the percentage of total dry matter was comparable. Quercetin and isorhamnetin in particular were found in exceptionally high concentrations. Traditional cereals contain no flavonoids, so cañihua may prove an important source of these health-promoting compounds. Cañihua seeds also contain a phenolic compound called vanillic acid. This is a flavor enhancer that gives cañihua a pleasant taste, particularly as a fermented drink (Peñarrieta et al. 2008b; Repo-Carrasco-Valencia et al. 2010).

Genetic resources for C. pallidicaule

D.W. Gade (1970) wisely noted nearly half a century ago that the continued presence of cañihua in the Altiplano depends on its genetic transformation into a more efficient food producer. In spite of its unique agronomic and nutritional qualities, very few genetic resources have been developed for this crop. In 1999, a phylogenetic study of 19 *Chenopodium* species used random amplified polymorphic DNAs (RAPD) to analyze variation within the genus. The results showed clustering into five major groups: Group 1, *C. nuttalliae*, Group 2, *C. quinoa*, Group 3, *C. berlandieri* and *C. album*, Group 4, *C. pallidicaule*, and Group 5, *C. ambrosiodes*. Two *C. pallidicaule* accessions were included in the study, and were found to be genetically very similar. However, there were high levels of dissimilarity between *C. pallidicaule* and *C. quinoa*. The authors conclude that it may be possible to introgress favorable traits present in other *Chenopodium* species into quinoa by crossing, then backcrossing to restore fertility. However, the study was inconclusive as to whether *C. pallidicaule* was a close enough relative to perform this kind of cross (Ruas et al. 1999).

Further genomic research was reported in 2011 by Vargas, et al. (2011), who developed the first microsatellite markers for *C. pallidicaule*. From 616 microsatellite markers screened (192 derived from *C. pallidicaule* genomic sequences and 424 previously identified in *C. quinoa*), Vargas et al. identified 34 polymorphic markers, exhibiting a total of 154 different alleles. Forty percent of the *C. quinoa*-derived markers amplified, confirming their close genetic history. Of these, only 12 were polymorphic. Vargas states that these regions may be useful in identifying syntenic relationships and may suggest potential for cross-species transferability. The microsatellite data was used to create a phylogeny, which showed clear distinctions between wild and cultivated varieties and a distinct subclade of only erect

morphotypes. Other morphotypes were not predictive of genetic distance. There was no clear geographic differentiation, presumably due to the trading culture of the native Andean people. Additionally, outcrossing was estimated to occur about 1.8% of the time. *In situ* genetic diversity for *C. pallidicaule* is a concern for the perpetuation of the crop, and Vargas suggests utilizing wild varieties to maintain diversity in cultivation.

In 2015, Kolano et al. (2015) published an analysis describing the genome size and rDNA loci of 23 *Chenopodium* diploid species ($2n = 2x = 18$), including *C. pallidicaule*. The study also identified phylogenetic relationships, grouping *C. pallidicaule* with other American diploids (separate from Old World diploids). The American diploids possess much smaller genomes than other diploids, with *C. pallidicaule* measuring at less than 1 pg/2C.

The most recent paper providing insight into the genome of *C. pallidicaule* was published in *Nature* in 2017 (Jarvis, et al.). The purpose of this publication was to present a high quality reference assembly of the *C. quinoa* genome. As a part of the genome analysis of quinoa, two diploid species (*C. pallidicaule* and *C. suecicum*) were sequenced using Illumina short-reads. Quinoa is an ancient allopolyploid ($2n = 4x = 36$), presumably resulting from a cross between North American and Eurasian diploids, representing the A- and B- subgenomes of modern quinoa, respectively (Štorchová et al. 2015). While cañihua is not believed to be the direct parent of quinoa, it is an A-genome diploid with the ability to provide insight into the evolutionary history of quinoa and potential to be improved as an independent crop. An accession of *C. pallidicaule* was selected from the USDA collection (PI 478407) for sequencing. K-mer analyses estimated a genome size of approximately 452 Mb, of which 337,010,935bp were assembled in a draft assembly of the *C. pallidicaule* genome. The scaffold N50 was 356,818bp, and the longest scaffold was 2,949,784bp. Using the reads from

the cañihua assembly, 156 of the quinoa scaffolds (totaling 202.6 Mb) were assigned to the A sub-genome.

Advantages of a whole-genome assembly

Whole-genome assemblies have proven an effective tool in rapid crop improvement and genetic diversification. With new sequencing and assembly technologies, obtaining genomic information is becoming less expensive and much faster. Bioinformatic techniques are also rapidly improving to meet the demand for insightful analyses of large datasets. Already, whole-genome assemblies have been useful in enabling trait identification and alteration, speeding up the breeding process, and optimizing crop performance (Bevan et al. 2017).

While the implications of these rapid technological changes have certainly benefited elite crops, they have also had a dramatic effect on breeding of ‘orphaned’ (minor, regionally important) crops. According to Varshney et al (2012), high-throughput genotyping and sequencing technologies can enhance breeding specifically by facilitating low-coverage sequencing for high-resolution germplasm fingerprinting, accurate mapping of loci associated with traits of interest, and enabling genome-assisted breeding. The authors are concerned, however, that advances in genomics will mean very little to the developing world and their regional crops unless they are able to collect consistent phenotypic data, as well.

Several ‘orphaned’ legume crops have been successfully improved by the application of modern genetic and phenotypic technologies. For example, whole-genome sequencing followed by re-sequencing of germplasm collection led to the discovery of several million SNPs in chickpea. Genetic and physical maps were developed and utilized in combination with yield-affecting traits such as drought resistance. QTL analysis was used to identify main-effect loci, and 45 were identified, 12 of which were located in a ‘QTL-hotspot’ region. This region was introgressed into the leading chickpea variety using marker-assisted backcrossing. The

application of powerful genomic and phenotypic tools to breeding has resulted in 29 lines of chickpea that are higher yielding than the initial elite line in both irrigated and rainfed conditions (Varshney 2016).

Cañihua is an ‘orphaned’ crop that has great, untapped potential. Genomic resources to assist in the development of this pseudocereal into a fully domesticated crop with consistently high yields, in addition to its extant high nutrient content and incredible tolerance to abiotic stresses (Ruas et al. 1999) are much needed. A high quality, whole-genome sequence is a valuable and flexible genomic tool that can be applied to the improvement of this unique crop (Bevan et al. 2017). I intend to further the research that has already been performed by using the existing Illumina assembly of *C. pallidicaule* (Jarvis, et al. 2017) and developing a chromosome-scale assembly using PacBio longreads and proximity-guided assembly technologies. Additionally, I will resequence 30 accessions of *C. pallidicaule* from both the USDA and Bolivian collections. These sequences will help quantify existing genetic diversity in the species for future use in breeding. I will also compare the final assembly to the latest assembly of *C. quinoa* in order further understand the relationship of the A genome diploid to its allotetraploid relative.

METHODS

Proximity-Guided Assembly

PI 478407 (the same accession sequenced by Jarvis, et al (2017)) will be grown from seed in a 4” pot in the growth chamber. When the plant has reached its mature size, it will be dark-treated for approximately 72 hours, then tissue will be collected, packed with dry ice, and sent to Phase Genomics (Lightfoot et al. 2017) along with the initial Illumina/ALLPATHS assembly (Jarvis, et al. 2017) for *in-vivo*, Hi-C based proximity-guided ligation.

Gap Closing with PacBio Long-reads and PBJelly

Tissue will be collected from PI 478407 that has been grown to mature height in the growth chamber and dark treated for ~72 hours. Extraction will be performed using the Qiagen Genomic-tip, then three samples will be sent to the Brigham Young University DNA Sequencing Lab. There, long-read sequence data will be collected by the Pacific Biosciences (PacBio) single-molecule, real-time sequencing technology (Quail et al. 2012).

When sequence data is available, BWA (Li 2013) will be used to trim and align PacBio reads to the first proximity-guided assembly (PGA1). The QUIVER (Chin et al. 2013) pipeline will be used to identify the consensus sequence between PacBio reads. Then, because PacBio reads have a higher error rate than Illumina, the original Illumina short-reads will be re-aligned to the consensus assembly and discrepancies will be changed in favor of the Illumina data using Pilon (Walker et al. 2014).

Proximity-Guided Assembly 2

A second proximity-guided assembly (PGA2) will be performed by Phase Genomics using the newly gap-filled and polished assembly. The assembly will be broken at each “N” (unknown base-pair), and the resulting contigs will be re-scaffolded. This will work to confirm that gap-filling has occurred correctly.

Resequencing

For resequencing, 7 USDA and 2 wild Peruvian accessions will be grown in the growth chamber, then an SDS extraction will be performed according to the protocol outlined by Sambrook et al. (1989). An additional 21 lines will be provided by collaborators at the Universidad Mayor de San Andres, La Paz, Bolivia. All extracted DNA will be sent to NovoGene for 10x, whole genome, Illumina sequencing.

SNP Discovery

INTERSNP (Herold et al. 2009) will be used to identify variants between the 30 accessions used for resequencing, then hapmap output files will be analyzed by SNPhylo (Lee et al. 2014) with the bootstrapping parameter set to 1000. The resulting phylogenetic tree will be visualized using FigTree.

Chloroplast Genome Reconstruction

The chloroplast and mitochondrial genomes soon to be published by Jarvis et al. will be used as a reference to develop chloroplast assemblies for *C. pallidicaule*. I will again use BWA (Li 2013) to align *C. pallidicaule* reads to the *C. quinoa* assembly.

Annotation

I will use the MAKER (Cantarel et al. n.d.) pipeline and RNA-seq data published by Jarvis et al. (2017) to annotate the chromosome-scale assembly of *C. pallidicaule*. I will also annotate the chloroplast genome using the existing RNA-seq data.

TIMELINE

The majority of this project has already been completed. I collected tissue and sent it to Phase Genomics in July 2017, and also extracted DNA for PacBio sequencing in the same month. I received results from both in late August and proceeded with BWA in September 2017. PBJelly, QUIVER, Pilon were also completed in September and the gap-closed, polished assembly was sent to Phase Genomics for a second proximity-guided assembly. USDA varieties were grown in the BYU growth-chamber starting in late September, then extracted in late October and early November. The extractions were sent to NovoGene shortly after the arrival of the Bolivian samples in early November. Problems with the re-scaffolding of the assembly by Phase

Genomics resulted in delayed results. This prevented me from completing genome annotation. There were also quality control problems with the NovoGene resequencing samples, so only 19 were ready by late December. I extracted more DNA in early December and sent them to NovoGene to try again. The first 19 samples were used for SNP discovery to report at the annual Plant and Animal Genome conference in January. Proximity-guided assembly 2 was completed days before the conference after a new program, Polar Star, was used to break the PacBio reads at low quality data points. I received the remainder of the NovoGene data in early February and successfully used INTERSNP to identify variants. I had trouble reformatting the .hapmap output files before inputting the data into SNPhylo, so I did not complete SNPhylo until early March. The remaining tasks are genome annotation and chloroplast genome reconstruction. I expect these tasks by the end of Spring term, 2018.

LITERATURE CITED

- Bevan, Michael W. et al. 2017. “Genomic Innovation for Crop Improvement.” *Nature* 543(7645): 346–54. <http://www.nature.com/doi/10.1038/nature22011> (March 25, 2018).
- Cantarel, BL et al. “MAKER: An Easy-to-Use Annotation Pipeline Designed for Emerging Model Organism Genomes.” *genome.cshlp.org*. <http://genome.cshlp.org/content/18/1/188.short> (March 1, 2018).
- Chin, Chen-Shan et al. 2013. “Nonhybrid, Finished Microbial Genome Assemblies from Long-Read SMRT Sequencing Data.” *Nature Methods* 10(6): 563–69. <http://www.nature.com/articles/nmeth.2474> (April 10, 2018).
- Gade, Daniel W. 1970. “Ethnobotany of Cañihua (*Chenopodium Pallidicaule*), Rustic Seed Crop of the Altiplano.” *Economic Botany* 24(1): 55–61. <http://link.springer.com/10.1007/BF02860637> (March 20, 2018).
- Herold, C. et al. 2009. “INTERSNP: Genome-Wide Interaction Analysis Guided by a Prior Information.” *Bioinformatics* 25(24): 3275–81. <http://www.ncbi.nlm.nih.gov/pubmed/19837719> (March 28, 2018).
- Jarvis, David E, Yung Shwen Ho, Damien J Lightfoot, Sandra M Schmöckel, Bo Li, Theo J A Borm, Hajime Ohyanagi, Katsuhiko Mineta, Craig T Michell, Noha Saber, Najeh M Kharbatia, Ryan R Rupper, Aaron R Sharp, Nadine Dally, Berin A Boughton, Yong H Woo, Ge Gao, Elio G W M Schijlen, Xiujie Guo, Afaque A Momin, Sónia Negrão, Salim Al-Babili, Christoph Gehring, Ute Roessner, Christian Jung, Kevin Murphy, Stefan T Arold, Takashi Gojobori, C Gerard Van Der Linden, et al. 2017. “The Genome of

Chenopodium Quinoa.” *Nature* 542: 307–510.

<https://www.nature.com/articles/nature21370.pdf> (March 20, 2018).

Jarvis, David E., Yung Shwen Ho, Damien J. Lightfoot, Sandra M. Schmöckel, Bo Li, Theo J. A.

Borm, Hajime Ohyanagi, Katsuhiko Mineta, Craig T. Michell, Noha Saber, Najeh M.

Kharbatia, Ryan R. Rupper, Aaron R. Sharp, Nadine Dally, Berin A. Boughton, Yong H.

Woo, Ge Gao, Elio G. W. M. Schijlen, Xiujie Guo, Afaque A. Momin, Sónia Negrão, Salim

Al-Babili, Christoph Gehring, Ute Roessner, Christian Jung, Kevin Murphy, Stefan T.

Arold, Takashi Gojobori, C. Gerard van der Linden, et al. 2017. “The Genome of

Chenopodium Quinoa.” *Nature* 542(7641): 307–12.

<http://www.nature.com/doi/10.1038/nature21370> (March 1, 2018).

Kolano, Bozena, Dorota Siwinska, Jamie McCann, and Hanna Weiss-Schneeweiss. 2015. “The

Evolution of Genome Size and rDNA in Diploid Species of *Chenopodium* S.l.

(Amaranthaceae).” *Botanical Journal of the Linnean Society* 179(2): 218–35.

<https://academic.oup.com/botlinnean/article-lookup/doi/10.1111/boj.12321> (March 25, 2018).

Lee, Tae-Ho et al. 2014. “SNPhylo: A Pipeline to Construct a Phylogenetic Tree from Huge SNP Data.” *BMC Genomics* 15(1): 162.

<http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-162> (March 1, 2018).

Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” <http://arxiv.org/abs/1303.3997> (March 27, 2018).

Lightfoot, D. J. et al. 2017. “Single-Molecule Sequencing and Hi-C-Based Proximity-Guided Assembly of Amaranth (*Amaranthus Hypochondriacus*) Chromosomes Provide Insights into Genome Evolution.” *BMC Biology* 15(1): 74.

- <http://bmcbiol.biomedcentral.com/articles/10.1186/s12915-017-0412-4> (March 27, 2018).
- Peñarrieta, J. Mauricio, J. Antonio Alvarado, Björn Åkesson, and Björn Bergenståhl. 2008a. “Total Antioxidant Capacity and Content of Flavonoids and Other Phenolic Compounds in Canihua (*Chenopodium Pallidicaule*): An Andean Pseudocereal.” *Molecular Nutrition & Food Research* 52(6): 708–17. <http://doi.wiley.com/10.1002/mnfr.200700189> (March 20, 2018).
- . 2008b. “Total Antioxidant Capacity and Content of Flavonoids and Other Phenolic Compounds in Canihua (*Chenopodium Pallidicaule*): An Andean Pseudocereal.” *Molecular Nutrition & Food Research* 52(6): 708–17. <http://doi.wiley.com/10.1002/mnfr.200700189> (March 20, 2018).
- Quail, Michael et al. 2012. “A Tale of Three next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers.” *BMC Genomics* 13(1): 341. <http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-341> (March 27, 2018).
- Ranilla, Lena Galvez et al. 2009. “Evaluation of Indigenous Grains from the Peruvian Andean Region for Antidiabetes and Antihypertension Potential Using In Vitro Methods.” *Journal of Medicinal Food* 12(4): 704–13. <http://www.liebertonline.com/doi/abs/10.1089/jmf.2008.0122> (March 20, 2018).
- Rastrelli, Luca, Francesco De Simone, Oreste Schettino, and Antonio Dini. 1996. “Constituents of *Chenopodium Pallidicaule* (Cañihua) Seeds: Isolation and Characterization of New Triterpene Saponins.” *Journal of Agricultural and Food Chemistry* 44: 3528–33. <https://pubs-acscs-org.erl.lib.byu.edu/doi/abs/10.1021/jf950253p> (March 20, 2018).
- Repo-Carrasco-Valencia, Ritva, Jarkko K. Hellström, Juha-Matti Pihlava, and Pirjo H. Mattila. 2010. “Flavonoids and Other Phenolic Compounds in Andean Indigenous Grains: Quinoa

- (Chenopodium Quinoa), Kañiwa (Chenopodium Pallidicaule) and Kiwicha (Amaranthus Caudatus).” *Food Chemistry* 120(1): 128–33. <https://www-sciencedirect-com.erl.lib.byu.edu/science/article/pii/S0308814609011662> (March 20, 2018).
- Repo-Carrasco, R., C. Espinoza, and S. -E. Jacobsen. 2003. “Nutritional Value and Use of the Andean Crops Quinoa (Chenopodium Quinoa) and Kañiwa (Chenopodium Pallidicaule).” *Food Reviews International* 19(1–2): 179–89. <http://www.tandfonline.com/doi/abs/10.1081/FRI-120018884> (March 20, 2018).
- Rodríguez, J. P. et al. 2017. “Seed Shattering of Cañahua (*Chenopodium Pallidicaule* Aellen).” *Journal of Agronomy and Crop Science* 203(3): 254–67. <http://doi.wiley.com/10.1111/jac.12192> (March 20, 2018).
- Ruas, Paulo M. et al. 1999. “Genetic Relationship among 19 Accessions of Six Species of *Chenopodium* L., by Random Amplified Polymorphic DNA Fragments (RAPD).” *Euphytica* 105(1): 25–32. <http://link.springer.com/10.1023/A:1003480414735> (March 20, 2018).
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual. Second Edition*. New York: Greene Publishing Associates and John Wiley & Sons.
- Štorchová, Helena et al. 2015. “The Introns in Flowering Locus T-like (FTL) Genes Are Useful Markers for Tracking Paternity in Tetraploid *Chenopodium Quinoa* Willd.” *Genetic Resources and Crop Evolution* 62(6): 913–25. <http://link.springer.com/10.1007/s10722-014-0200-8> (March 25, 2018).
- Vargas, A. et al. 2011. “Development and Use of Microsatellite Markers for Genetic Diversity Analysis of Cañahua (*Chenopodium Pallidicaule* Aellen).” *Genetic Resources and Crop Evolution* 58(5): 727–39. <http://link.springer.com/10.1007/s10722-010-9615-z> (March 20, 2018).

Varshney, Rajeev K et al. 2012. “Can Genomics Boost Productivity of Orphan Crops?” *Nature Biotechnology* 30(12): 1172–76. <https://www.nature.com/articles/nbt.2440.pdf?origin=ppub> (March 26, 2018).

———. 2016. “Exciting Journey of 10 Years from Genomes to Fields and Markets: Some Success Stories of Genomics-Assisted Breeding in Chickpea, Pigeonpea and1. Varshney, R. K. Exciting Journey of 10 Years from Genomes to Fields and Markets: Some Success Stories of Genom.” *Plant Science* 242: 98–107. <https://www.sciencedirect.com/science/article/pii/S0168945215300662?via%3Dihub> (March 25, 2018).

Walker, Bruce J et al. 2014. “Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement.” *PloS one* 9(11): e112963. <http://www.ncbi.nlm.nih.gov/pubmed/25409509> (March 27, 2018).