Ciencia e Investigación AGRARIA

Cien. Inv. Agr. 46(2):187-196. 2019
www.rcia.uc.cl

PLANT BREEDING, GENETICS AND GENETIC RESOURCES

DOI 10.7764/rcia.v46i2.2150

RESEARCH PAPER

New seed collections of North American pitseed goosefoot (*Chenopodium berlandieri*) and efforts to identify its diploid ancestors through whole-genome sequencing

Eric N. Jellen, David E. Jarvis, Spencer P. Hunt, Hayley H. Mangelsen, and Peter J. Maughan

Brigham Young University, Plant & Wildlife Sciences Department. 4105 LSB, Provo, Utah, 84602, USA.

Abstract

E.N. Jellen, D.E. Jarvis, S.P. Hunt, H.H. Mangelsen, and P.J. Maughan. 2019. New seed collections of North American pitseed goosefoot (*Chenopodium berlandieri*) and efforts to identify its diploid ancestors through whole-genome sequencing. Cien. Inv. Agr. 46(2): 187-196. Pitseed goosefoot (*Chenopodium berlandieri*) is an ecologically diverse wild/weedy North American species within the primary gene pool for improving South American quinoa (*Chenopodium quinoa*). Both taxa are 36-chromosome allotetraploids with subgenomes AA and BB. The A genome is found in a large number of diploids in the Americas, along with one Northeast Asian taxon, and was recently shown to be the maternal ancestor, while the paternal B genome is closely related to several extant Eurasian diploids. Two of our primary objectives were 1) to determine the extent of genetic diversity in the allotetraploid *C. berlandieri-quinoa-hircinum* complex and 2) to characterize the evolutionary path from polyploidization to domestication in these taxa. In an effort to survey genetic diversity, in 2018, we made seed collections of southern Texas, southern Great Plains, and New England coastal ecotypes of *C. berlandieri* as well as sympatric diploids. With respect to the second goal, we performed whole-genome sequencing of two Sonoran Desert *Chenopodium* A-genome diploids in subsection *Cellulata* and Andean cultivated *C. pallidicaule* in subsection *Leiosperma*. When paired reads were aligned to the whole-genome reference of *C. quinoa* strain 'QQ74', the match percentages were 99.31, 99.23, and 98.53 for *C. watsonii*, *C. sonorense*, and *C. pallidicaule*, respectively. These data strongly support *C. watsonii* as being the most closely related of these three species to the A-genome ancestor of quinoa. Ongoing sequencing efforts with a larger panel of diploids are aimed at identifying the maternal ancestor of *C. quinoa* and *C. berlandieri*, if extant.

Keywords: *Chenopodium*, DNA sequencing, genetic resources, quinoa.

Introduction

Quinoa (*Chenopodium quinoa* Willd., $2n = 4x = 36$) cultivation is spreading rapidly throughout the world in response to increasing consumer demand for this highly nutritious South American pseudocereal. In addition to their nutritional benefits, most quinoa strains tend to be highly tolerant of salt- and drought-affected production environments. Quinoa's principal weakness is that it is highly susceptible to a wide range of pests and diseases and heat during its reproductive cycle. This susceptibility occurs because

Received Mar 06, 2019. Accepted Jun 04, 2019.
Corresponding author: rick_jellen@byu.edu

the crop evolved for thousands of years under domestication in highland Andean and Chilean coastal environments, which have mild daytime high temperatures and are relatively geographically isolated from Eurasian and North American pest and pathogen populations. One manifestation of quinoa's heat susceptibility is that it is spontaneously cross-pollinated by weedy pitseed goosefoot (*C. berlandieri* Moq.) where the two co-inhabit quinoa production fields in warm North American environments (Wilson and Manhart, 1993). A series of recent cytogenetic and DNA sequencing studies have confirmed that these two taxa have a close relationship, constituting a single biological species with a genome composition of AABB and with *C. berlandieri* at the root of the allotetraploid clade that includes domesticated quinoa (Jarvis *et al.* 2017; Kolano *et al.* 2016; Walsh *et al.* 2015; Brown *et al.* 2015; Storchova *et al.* 2015). This clade also includes forms that were anciently domesticated as seed and vegetable crops at various times in Mesoamerica (Wilson and Heiser, 1979) and eastern North America (Kistler and Shapiro, 2011; Smith and Yarnell, 2009; Gremillion, 1993; Smith, 1984). In addition, sequencing of cpDNA and mtDNA genomes recently confirmed that of subgenomes A and B, the former is most likely the cytoplasmic donor and was, consequently, the maternal ancestor of the allotetraploid clade and that the Western Hemisphere was most likely where the hybridization event took place (Maughan *et al.*, 2019).

These results point to two critical lines of research, one basic and the other applied. Fundamental research should focus on the identification, collection, and preservation of the AA and BB diploid species that are most closely related to the *C. berlandieri*-*C. quinoa* complex. Previously cited studies (Maughan *et al.*, 2019; Jarvis *et al.*, 2017; Kolano *et al.*, 2016; Walsh *et al.*, 2015; Brown *et al.*, 2015; Storchova *et al.*, 2015) indicated that both *C. suecicum* Murr. and *C. ficifolium* Sm. carry the BB genome, and these two species have recently been shown to be cross-fertile (Hodkova and Mandak, 2018); however, identification of the A-genome ancestor may prove more difficult, with *C. watsonii* A. Nels. emerging as the leading candidate. *Chenopodium watsonii*, also known as Watson's stinking goosefoot, thrives on disturbed high-nitrogen soils in the upper Sonoran Desert and adjacent intermediate-elevation woodlands of Arizona, New Mexico, and Utah, although its range would certainly have moved during the 3.3-6.3 million years since the estimated polyploidization time (Jarvis *et al.*, 2017). Wilson (1981) hypothesized that southwestern North America was the best candidate for the center of origin of the allotetraploid species complex, being the only region in the Americas where *C. berlandieri zschackei* is known to be sympatric with *Cellulata* diploids. However, this hypothesis certainly demands further testing, especially given that comprehensive taxonomic descriptions, including seed and plant specimen collections, of the North and South American A-genome diploids are ongoing, as exemplified by a very recent report refining an important group of southwestern North American taxa in the subsection *Cellulata* complex that includes *C. watsonii* (Benet-Pierce and Simpson, 2017).

From an applied research perspective, efforts to collect and characterize samples representative of the broad geographic and morphological range of *C. berlandieri* and its wild/weedy South American cousin, *C. hircinum* Schrad., are essential. Pitseed goosefoot is currently subdivided taxonomically into subspecies *nuttaliae* and *berlandieri*. The former includes cultivated and associated weedy forms from South-Central Mexico, while the latter includes the wild and weedy ecotypes *berlandieri* from interior southern Texas, *boscianum* from the Gulf of Mexico coast, *bushianum* from interior eastern North America, *macrocalycium* from the Atlantic Coast, and *sinuatum* from southwestern North America. While all of these botanical varieties of *C. berlandieri* ssp. *berlandieri* have somewhat restricted geographic distributions, the additional botanical variety *zschackei* is a catch-all group that encompasses populations distributed throughout temperate North America

and possibly southward into the northern Andes (Wilson, 1981; Wilson and Heiser, 1979). A third subspecies, *jonesianum*, was created to encompass extinct cultigens from eastern North America (Smith and Yarnell, 2009). Botanical varieties of *C. berlandieri* are differentiated mostly by subtle morphological characters such as seed size, the presence of bracteate leaves, and pericarp coloration, along with habitat preference.

In this report, we outline recent pitseed goosefoot collection efforts in North America; recent observations regarding hybridization between quinoa and pitseed goosefoot and selection of individuals with desirable production characteristics from selfed progeny of these crosses; and sequencing efforts designed to narrow the list of potential A-genome ancestors of the *C. berlandieri/hircinum/quinoa* allotetraploid species complex. The principal objectives of the collection efforts in 2018 were as follows: 1) to determine if unique spring-fruiting *C. berlandieri* var. *berlandieri* is present in the mild-winter environment of subtropical southern Texas and to conduct the first systematic seed collection for this ecotype; 2) to expand the very limited existing collections and determine the degree of morphological variation present in New England coastal var. *macrocalycium*; 3) to determine the presence and variety of likely heat-tolerant strains of *C. berlandieri* within the southern Great Plains; and 4) to identify *Chenopodium* diploids (putatively, with an A-genome) sympatric with pitseed goosefoot in these regions.

Materials And Methods

Pitseed goosefoot seed collection expeditions in 2018

Personnel from Brigham Young University (BYU) made seed collection trips in 2018 to southern Texas (April), the New England coast (September), and Oklahoma (October). Seeds were sampled from populations encountered on publicly accessible roadside rights-of-way and public beaches, generally guided by previous collection sites noted on herbarium specimens. Maps of the three collections are included in Fig. 1. Collection site passport data and putative identification of *C. berlandieri* and accessions of other selected species were noted and are presented in Table 1.

Fruits from *Chenopodium* population samples were photographed at BYU using a BK PLUS Lab System (Dun Inc., Palmyra, Virginia, USA) with a 65-mm Canon (Melville, New York, USA) Macro lens at f4-5.6 with 1.5x-4x magnification. Images were captured using Capture One Pro 64 bit v. 10.2.1 (Phase One, New York, USA) and superimposed using Stacker v. 1.04 (Zerene Systems, Richland, Washington, USA), with contrast adjustment and sharpening performed using the Smart Sharpen tool in Photoshop x64 v. 12.0 (Adobe Inc., San Jose, California, USA).

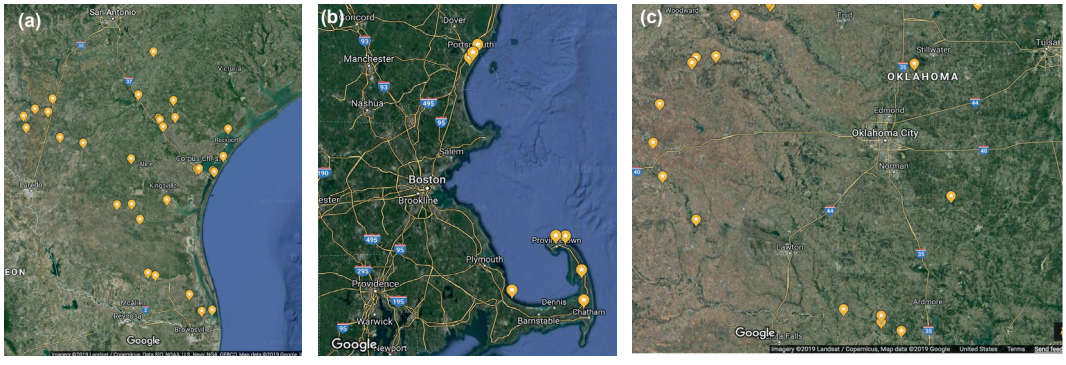

Figure 1. Maps of BYU *Chenopodium* collection trips in 2018. (a) Southern Texas; (b) New England; (c) Oklahoma.

Table 1. *Chenopodium* collections by BYU personnel in 2018. *Chenopodium berlandieri* taxa: BERB=var. *berlandieri*; BERC=var. *boscianum*; BERL=unclassified variety; BERM=var. *macrocalycium*; BERS=var. *sinuatum*; and BERZ=var. *zschackei*. ALBE=*C. albescens*; UND=undetermined species.

No.	Taxa	Date	County	State	Latitude	Longitude	Elevation
1801	BERB	4-19	La Salle	Texas	28.4284	-99.2514	118
1802	BERB	4-19	Cotulla	Texas	28.2828	-99.3045	147
1803	ALBE	4-19	Dimmit	Texas	28.3135	-99.4723	159
1804	BERB	4-19	Dimmit	Texas	28.2342	-99.6142	176
1805	BERB	4-19	Webb	Texas	28.1037	-99.5791	201
1806	BERB	4-19	Webb	Texas	27.9998	-99.1494	117
1807	BERB	4-19	Webb	Texas	27.9407	-98.8594	120
1808	ALBE	4-19	Webb	Texas	27.9407	-98.8594	120
1809	ALBE, BERB	4-19	Duval	Texas	27.5821	-98.4355	
1810	ALBE	4-20	Duval	Texas	27.5821	-98.4355	126
1811	ALBE	4-20	Jim Hogg	Texas	27.2954	-98.6317	154
1812	BERB	4-20	Brooks	Texas	27.2568	-98.4297	79
1813	ALBE	4-20	Brooks	Texas	27.2568	-98.4297	79
1814	BERB	4-20	Brooks	Texas	27.2625	-98.2495	58
1815	ALBE	4-20	Brooks	Texas	27.2625	-98.2495	58
1816	ALBE, UND	4-20	Brooks	Texas	27.1569	-98.0714	19
1817	BERB	4-20	Brooks	Texas	27.0962	-98.1463	32
1818	BERB	4-20	Hidalgo	Texas	26.4989	-98.0451	22
1819	BERB	4-20	Willacy	Texas	26.4686	-97.9517	16
1820	BERB	4-20	Cameron	Texas	26.0711	-97.3765	3
1821	BERC	4-20	Cameron	Texas	26.0803	-97.2489	3
1822	BERB	4-20	Cameron	Texas	26.2553	-97.5286	15
1823	BERB	4-21	Kleberg	Texas	27.3099	-97.8101	4
1824	BERB	4-21	Nueces	Texas	27.6545	-97.4298	6
1825	BERC	4-21	Nueces	Texas	27.6565	-97.4023	10
1826	BERC	4-21	Nueces	Texas	27.63	-97.2295	0
1827	BERC	4-21	Nueces	Texas	27.6194	-97.2119	0
1828	BERC	4-21	Nueces	Texas	27.7903	-97.0958	6
1829	BERC	4-21	Aransas	Texas	28.0951	-97.0322	0
1830	BERB	4-21	Bee	Texas	28.4169	-97.7219	60
1831	BERB	4-21	Bee	Texas	28.2265	-97.7027	49
1832	BERB	4-21	Live Oak	Texas	28.1984	-97.8939	42
1833	BERB	4-21	Live Oak	Texas	28.2176	-97.9066	41
1834	BERB	4-21	Live Oak	Texas	28.4776	-98.1658	60
1835	BERB	4-21	Karnes	Texas	28.9567	-97.9739	86
1836	BERB	4-21	San Patricio	Texas	28.1229	-97.8521	40
1856	BERM	9-21	Barnstable	Massachusetts	41.7663	-70.4816	0
1857	BERM	9-21	Barnstable	Massachusetts	41.8717	-70.0088	0
1858	BERM	9-21	Barnstable	Massachusetts	42.0496	-70.1183	0
1859	BERM	9-21	Barnstable	Massachusetts	42.0514	-70.1849	2
1860	BERM	9-21	Barnstable	Massachusetts	41.7121	-69.9933	1
1861	BERM	9-21	Barnstable	Massachusetts	41.7121	-69.9933	2
1862	BERM	9-22	Rockingham	New Hampshire	43.0417	-70.7153	2
1863	BERM	9-22	Rockingham	New Hampshire	43.0025	-70.7478	2
1864	BERM	9-22	York	Maine	43.3439	-70.4987	1
1878	BERS	10-03	Payne	Oklahoma	35.9766	-97.2384	276
1879	BERS	10-03	Osage	Oklahoma	36.435	-96.6313	262
1880	BERS	10-03	Osage	Oklahoma	36.4781	-96.949	284
1883	BERZ	10-03	Alfalfa	Oklahoma	36.7823	-98.1695	356
1886	BERL	10-03	Alfalfa	Oklahoma	36.7632	-98.1282	379
1887	BERL	10-03	Alfalfa	Oklahoma	36.6664	-98.1984	373
1891	BERL	10-04	Woods	Oklahoma	36.4361	-98.5873	444
1893	BERL	10-04	Woods	Oklahoma	36.4833	-98.6757	444
1894	BERL	10-04	Woods	Oklahoma	36.5212	-98.7466	473
1896	BERL	10-04	Major	Oklahoma	36.3621	-98.8958	535
1897	BERL	10-04	Dewey	Oklahoma	36.0434	-99.0741	553
1899	BERZ	10-04	Dewey	Oklahoma	36.0298	-99.257	628
18101	BERZ	10-04	Roger Mills	Oklahoma	35.6715	-99.5876	575
18102	BERZ	10-04	Beckham	Oklahoma	35.3742	-99.6467	584
18103	BERZ	10-04	Greer	Oklahoma	35.115	-99.554	512
18105	BERS	10-05	Jefferson	Oklahoma	34.0856	-97.9115	282
18106	BERS	10-05	Jefferson	Oklahoma	33.9843	-97.5701	257
18107	BERS	10-05	Jefferson	Oklahoma	34.0335	-97.5698	240
18109	BERS	10-05	Love	Oklahoma	33.9112	-97.3902	286
18110	BERS	10-05	Pottawatomie	Oklahoma	34.9493	-96.9221	291

DNA sequence analysis of C. pallidicaule*,* C. watsonii*, and* C. sonorense

Raw DNA sequence reads for *C. pallidicaule* USDA accession PI 478407 (BYU 1652; Jarvis *et al.,* 2017), *C. watsonii* accession BYU 873 (Yavapai Co., Arizona), and *C. sonorense* Benet-Pierce & M.G. Simpson accession BYU 17220 (Santa Cruz Co., Arizona) were generated by the Novogene Corporation (Chula Vista, California, USA) using an Illumina (San Diego, California, USA) high-fidelity, short-read platform. Raw reads were then trimmed using Trimmomatic (Bolger et al., 2014), removing low-quality reads and Illumina adapters. Reads from each accession were aligned to the quinoa QQ74 published reference genome (Jarvis *et al.*, 2017) using BWA-MEM (Li, 2013), and SAM files were then converted to BAM format, sorted and indexed using SAMtools v1.9 (Li *et al*., 2009). The GATK CollectAlignmentSummaryMetrics subprogram was used to generate mapping statistics (McKenna *et al.*, 2010).

Results

Collection trips to Texas, New England, and Oklahoma in 2018

The trip to southern Texas in April resulted in the collection of seeds from 23 populations of *C. berlandieri* var. *berlandieri*; eight sympatric populations of putative diploid, broad-leaved *C. albescens*, which at one location (BYU 1816) was intermixed with a third narrow-leaved species that might have been *C. pratericola*; and six populations of the Gulf Coastal ecotype *C. berlandieri* var. *boscianum* (Table 1, Fig. 1a). The var. *berlandieri* populations all emitted a fetid trimethylamine odor and were mostly found on highly disturbed roadsides, including areas where seasonally dry riverbeds intersected the highway. The var. *boscianum* populations also smelled strongly of trimethylamine, were restricted to sandy barrier islands and tidal estuary margins, tended towards later maturity, and were more glabrous than var. *berlandieri*. Both ecotypes included populations of plants having semi-compact to compact inflorescences, with and without extensive stem branching. As shown in Fig. 2, var. *boscianum* included populations with significantly larger fruits, which approached 1.8 mm in diameter.

The collection trip to the New England coast in September yielded seeds from nine populations of *C. berlandieri* var. *macrocalycium* (Table 1 and Figs. 1b and 3). This was a fairly common species in disturbed beach sand, especially among decaying seaweed at the high-tide line. Plants displayed a consistent phenotype with excessive branching, a yellow-green and glabrous appearance, and fruits to 2 mm in diameter in

Figure 2. (a) Comparison of fruits of *C. berlandieri* vars. *berlandieri* and *boscianum* from southern Texas. In the columns, from left to right: var. *berlandieri* populations BYU 1814, BYU 1818, BYU 1822, and BYU 1833; var. *boscianum* populations BYU 1821, BYU 1825, BYU 1828, and BYU 1829. (b) Typical habitat of *C. berlandieri* var. *berlandieri* (BYU 1806) on cracked alkaline clay with *Helianthus* spp. (c) Typical habitat of *C. berlandieri* var. *boscianum* (BYU 1827) in sandy substrate on Mustang Island.

small, widely spaced glomerules and fairly uniform in maturity. At the Chatham Estuary of Cape Cod, populations extended a few meters into adjacent deciduous forest, where they were noticeably shorter in the shade. Notably, at least one population on the New Hampshire coast was being attacked by a stem-weakening, boring maggot (Fig. 3c).

The collection trip to Oklahoma in October (Table 1 and Fig. 1c) resulted in eight collections of *C. berlandieri* var. *sinuatum*, four of var. *zschackei*, and eight of a hitherto unidentified variety with characteristics similar to those of vars. *zschackei* (yellow area at the stylar base) and *berlandieri* (strong trimethylamine odor and more conical fruit side profile). Figure 4 compares fruits from three populations of each of these ecotypes. Variety *sinuatum* was more common in northeastern Oklahoma, with var. *zschackei* predominant in the southeast and the unnamed variety more common in the western part of the state. At all sites, the species was strongly associated with disturbed, sandy soils and was especially common along roadsides that intersected with arroyos and seasonally flooded riverbanks.

In addition to the *C. berlandieri* accessions, we collected seeds from 12 populations of unidentified narrow-leaved species bearing utriculate fruits. We are currently working to identify these putative diploids via fruit morphology and DNA analyses.

Recovery of significant breeding traits from C. berlandieri x C. quinoa *hybrids*

We herein provide some examples that demonstrate the value of ecologically variable pitseed goosefoot as a quinoa improvement resource. These examples are from intertaxa populations derived via either spontaneous field crosses or intentional greenhouse crosses (Fig. 5).

Figure 5a shows vigorous and relatively uniform $F_{2:5}$ plants (O108-1-11-17-1 parent) derived from a cross between 'Ollague' and BYU 14108, a var. *sinuatum* accession collected in October 2014 in a torrid oak-grassland steppe in Cochise Co., Arizona. A single black-seeded hybrid plant was identified in the greenhouse after growing seeds harvested from a heat-stressed 'Ollague' plant exposed to BYU 14108 pollen. Among the 178 F_2 plants showing widely variable morphology in terms of plant architecture, seed size, and seed coloration, 169 (94.9%) produced at least one seed, and all but a few produced many more than five seeds. Only two of the plants, however, produced seeds that were not black, one of which (O108-1-11) bore white seeds that were similar in size to those of 'Ollague'.

Figure 5b shows representative fruit size and color variation in the parent, F_1, and F_2 plants (population R1R-2) derived from a cross between quinoa cv. 'Real-1' and BYU 937, a var. *boscianum* accession

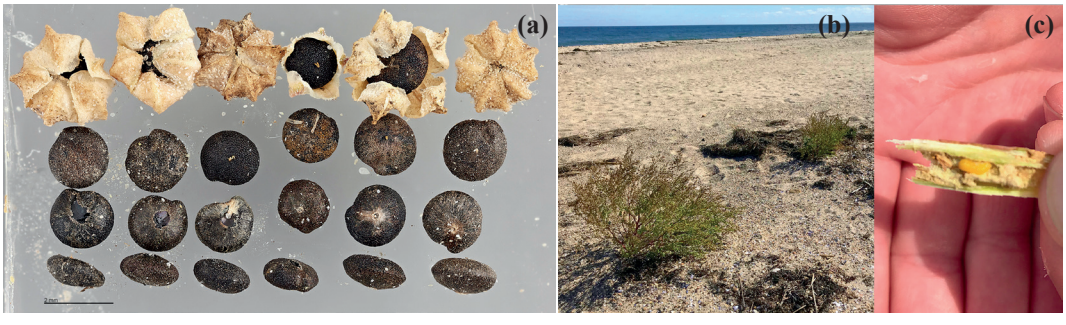

Figure 3. (a) Fruits of six *C. berlandieri* var. *macrocalycium* populations from New England. In the columns, from left to right: BYU 1856, BYU 1857, BYU 1859, BYU 1860, BYU 1863, and BYU 1864. (b) Var. *macrocalycium* growing typically along the high-tide line among seaweed detritus at Sandwich Beach, Cape Cod, Massachusetts (BYU 1856). (c) Unidentified stem-boring maggot infecting var. *macrocalycium* plants at Odiorne Point State Park, New Hampshire.

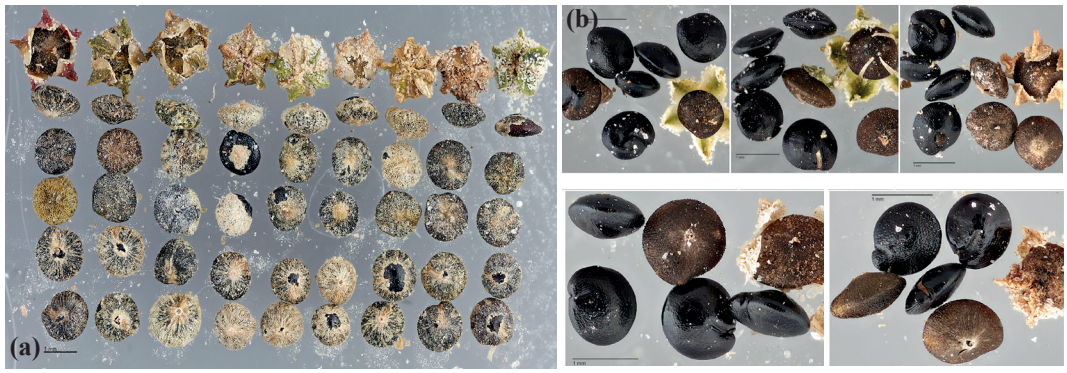

Figure 4. (a) Comparison of fruits of *C. berlandieri* from Oklahoma. In the columns, from left to right: var. *sinuatum* collections BYU 1878, BYU 1879, and BYU 1880; var. unknown collections BYU 1883, BYU 1886, and BYU 1887; and var. *zschackei* collections BYU 1899, BYU 18101, and BYU 18102. (b) Fruits of *Chenopodium* subsect. *Leiosperma* diploids collected in Oklahoma. Clockwise from middle left: BYU 1881 (Alfalfa Co.); BYU 1888 (Woods Co.); BYU 1892 (Woods Co.); BYU 18100 (Dewey Co.); and BYU 18108 (Love Co.).

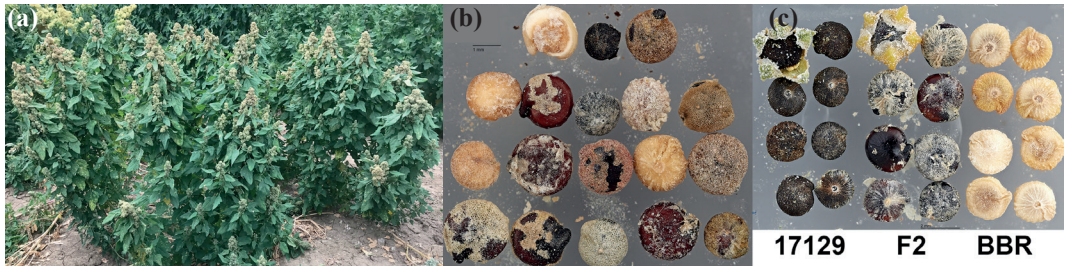

Figure 5. Characteristics of plants derived from hybridization between *C. berlandieri* and *C. quinoa*. (a) Plot of selected $F_{2:5}$ plants (O108-1-11-17-1 parent) derived from a cross between Salares quinoa cv. 'Ollague' and BYU 14108 growing at the University of California Hansen Agricultural Experiment Station near Los Angeles, California, in June 2018. BYU 14108 (var. *sinuatum*) was collected in October 2014 in the torrid upper Sonoran oak-grassland savanna zone near the Chiricahua Mountains in southern Arizona. (b) Variation in fruit size and color in the 'Real-1' (top row, left) and BYU 937 (top row, middle) parents, F_1 plants (top row, right) and fifteen F_2 plants of population R1R-2. BYU 937 (var. *boscianum*) was a single plant collected along the shore of the Galveston Bay estuary, Texas, in August 2009. (c) Fruits from one early-maturing F_2 derived from a quinoa cv. 'Brightest Brilliant Rainbow' x var. *macrocalycium* hybrid showing transgressive segregation for large seed size. The F_1 parent was identified in a row of Brightest Brilliant Rainbow growing at the Woodman Research Farm at the University of New Hampshire Agricultural Experiment Station in September 2018, and seeds from that plant were kindly provided by UNH Professor Tom Davis; 17129 = var. *macrocalycium* parent; BBR = Brightest Brilliant Rainbow.

collected in August 2009 at the high-tide line along the sandy shore of Galveston Bay, Texas. Multiple black-seeded, foul-smelling hybrids were recovered from seeds taken off a heat-stressed Real-1 plant that had been exposed to BYU 937 pollen in the greenhouse, one of which produced population R1R-2. The hybrid bore seeds that were similar in size to those of 'Real-1', while several F_2 plants produced seeds that exceeded the size of those of the quinoa parent, despite the fact that the var. *boscianum* parent produces seeds that are only slightly larger than 1 mm in diameter (Fig. 5b). This transgressive segregation pattern suggests that BYU 937 harbors at least one complementary allele for large seed size.

During the New England collection trip in September 2018, inspection of a row of 'Brightest Brilliant Rainbow' quinoa at the University of New Hampshire Experimental Farm in Durham revealed the presence of a putative F_1 plant that was very large and bushy with some phenotypic characteristics reminiscent of var. *macrocalycium*. Approximately 60 large, black seeds from this plant were kindly provided by Tom Davis and subsequently planted in the greenhouse at BYU,

resulting in plants with a wide array of whole-plant and panicle morphologies, from compact to exceptionally lax. Figure 5c presents seeds from one early-maturing, putative F_2 showing transgressive segregation for large seed size. We previously noted in an 'Ingapirca' x var. *macrocalycium* cross that segregants in the F_2 and subsequent generations were universally heat susceptible, produced less than 50% fertile progeny, and produced no individuals with large seeds that did not also have highly lax panicles (personal observations).

Comparison of genomic DNA sequences of C. watsonii*,* C. sonorense*, and* C. pallidicaule *with the quinoa QQ74 reference genome*

Table 2 presents a summary of trimmed short-read sequence mapping of three A-genome diploids – wild Watson's stinking goosefoot, wild Sonoran goosefoot and cultivated cañahua – to the lowland QQ74 quinoa reference genome reported in Jarvis *et al.* (2017). Although all three species are closely related to quinoa and belong to the A-genome diploid group, sequencing reads of Watson's stinking goosefoot showed the highest mapping percentage and the lowest mismatch rate, whereas cañahua showed the lowest mapping percentage and the highest mismatch rate.

Discussion

The diverse seed collections described herein represent a very valuable resource for improving cultivated quinoa's resistance to pests and diseases. Taking an alternative perspective, the relatively limited pool of highly diverse quinoa strains legally available to breeders outside the Andean region can be viewed as a source of domestication traits for *redomesticating* pitseed goosefoot. This is a potential staple food crop that is already biologically adapted to diverse warm-season temperate and subtropical production environments. Analyses of seed components have revealed that pitseed goosefoot and its hybrid derivatives have nutritional value comparable to that of quinoa (unpublished). This result is consistent with pitseed goosefoot, known as *kipahki* in the Skiri Pawnee language, having been a foundational food source in indigenous eastern North American agriculture five centuries prior to the sixteenth-century arrival of Europeans.

Mounting evidence from DNA sequencing and cytogenetic studies indicates that *C. watsonii* or a very close relative that is unrecognized, unanalyzed, or extinct is the donor of the A genome's chromosome sets and cytoplasmic DNA found in pitseed goosefoot and quinoa (Jarvis *et al.*, 2017; Kolano *et al.*, 2016; Walsh *et al.*, 2015; Storchova

Table 2. Summary metrics of trimmed (40-Gb pretrimmed) Illumina short-sequence paired reads of genomic DNA from the A-genome diploids *C. watsonii*, *C. sonorense*, and *C. pallidicaule* mapped to the *C. quinoa* accession QQ74 whole-genome reference (Jarvis *et al.*, 2017).

	Reads Aligned (%)	Mismatch Rate (%)	High-Quality Error Rate (%)	Indel Rate (%)	Reads Aligned in Pairs (%)
C. watsonii					
First of pair	98.37	3.18	3.02	0.28	99.30
Second of pair	98.36	3.20	3.04	0.27	99.32
Pair	98.36	3.19	3.03	0.28	99.31
C. sonorense					
First of pair	98.36	3.39	3.24	0.28	99.21
Second of pair	98.33	3.43	3.27	0.28	99.24
Pair	98.34	3.41	3.26	0.28	99.23
C. pallidicaule					
First of pair	95.36	3.45	3.54	0.27	98.52
Second of pair	95.34	3.46	3.55	0.27	98.54
Pair	95.35	3.46	3.55	0.27	98.53

et al., 2015). These same studies have pointed to Eurasian *C. ficifolium* and *C. suecicum* as the B-genome donors. Ongoing work will involve long-read DNA sequencing and chromatin proximity mapping to assemble these diploid genomes into pseudochromosome-scale assemblies. Additional efforts will focus on the collection and sequencing of other North American *Chenopodium* diploids and, in conjunction with Latin American colleagues, the collection and sequencing of South American species. Although Andean cañahua (*C. pallidicaule*) does not appear to be the direct ancestor of *C. quinoa*, the scientific community should accelerate efforts to study, improve, conserve, and utilize this unique high-altitude food crop.

Acknowledgments

The authors express gratitude to Tom Davis (UNH) for sharing seeds of the Brightest Brilliant Rainbow x *C. berlandieri* var. *macrocalycium* F_1 plant and to Tom Davis, David Brenner and BYU student Zach Jaramillo for helping with seed collections.

Resumen

E.N. Jellen, D.E. Jarvis, S.P. Hunt, H.H. Mangelsen, y P.J. Maughan. 2019. Nuevas colecciones de semillas de pata de gallo norteamericano (*Chenopodium berlandieri*) y esfuerzos para identificar a sus antepasados diploides a través de la secuenciación de todo el genoma. Cien. Inv. Agr. 46(2): 187-196. Pata de gallo (*Chenopodium berlandieri*) es una especie norteamericana, silvestre y muy diversa en su ecología, tal que representa un recurso genético para mejorar la quinua sudamericana (*Chenopodium quinoa*). Ambas entidades taxonómicas son alotetraploides con 36 cromosomas compuestas en subgenomas AA y BB. El genoma A se encuentra en muchos diploides de las Américas, junto con una especie del noreste de Asia, y recientemente fue identificado como pariente maternal, mientras el genoma paternal B se relaciona a un grupo de diploides de Eurasia. Dos de nuestros objetivos principales eran 1) determinar la diversidad genética que hay en el complejo alotetraploide de *C. berlandieri-quinoa-hircinum*; y 2) caracterizar el sendero evolucionario de poliploidización hasta domesticación en este grupo taxonómico. Para investigar la diversidad genética, en el año 2018 hicimos colecciones de semillas de poblaciones de *C. berlandieri* y diploides simpátricos en el sur de Texas, el sur de los Llanos Grandes, y en el litoral de Nueva Inglaterra. Referente al segundo objetivo, secuenciamos los genomas de dos diploides AA de la subsección *Cellulata* del Desierto de Sonora y el cultivar andino cañahua *C. pallidicaule* de subsección *Leiosperma*. Al alinear leídas pareadas de estos diploides con la secuencia referencia de *C. quinoa* cultivar 'QQ74', los porcentajes que coincidieron eran 99.31, 99.23, and 98.53 para *C. watsonii*, *C. sonorense*, y *C. pallidicaule*, respectivamente. Estos datos aportan la hipótesis que, de entre estas tres *C. watsonii* es la especie más cercana al ancestro AA de la quinua. Continuamos nuestros esfuerzos en secuenciar un panel más amplio de diploides con el fin de identificar con más seguridad el ancestro maternal de *C. quinoa* y *C. berlandieri*, sea que todavía existe.

Palabras clave: *Chenopodium*, quinua, recursos genéticos, secuenciación de ADN.

References

Benet-Pierce, N., and M.G. Simpson. 2017. Taxonomic recovery of the species in the *Chenopodium neomexicanum* (Chenopodiaceae) complex and description of *Chenopodium sonorense* sp. nov. Journal of the Torrey Botanical Society 144(3):339–356.

Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Brown, D.C., V. Cepeda-Cornejo, P.J. Maughan, and E.N. Jellen. 2015. Characterization of the *granule-bound starch synthase I* gene in *Chenopodium*. The Plant Genome 8(1):1–12.

Gremillion, K.J. 1993. Crop and weed in prehistoric eastern North America: the *Chenopodium* example. American Antiquity 58(3):496–509.

Hodkova, E., and B. Mandak. 2018. An overlooked hybrid between the two diploid *Chenopodium* species in Central Europe determined by microsatellite and morphological analysis. Plant Systematics and Evolution 304:295–312.

Jarvis, D.E., Y.S. Ho, D.J. Lightfoot, S.M. Schmockel, B. Li, T.J.A. Borm, H. Ohyanagi, K. Mineta, C.T. Michell, N. Saber, N.M. Kharbatia, R.R. Rupper, A.R. Sharp, N. Dally, B.A. Boughton, Y.H. Woo, G. Gao, E.G.W.M. Schijlen, X. Guo, A.A. Momin, S. Negrao, S. Al-Babili, C. Gehring, U. Roessner, C. Jung, K. Murphy, S.T. Arold, T. Gojobori, C.G. van der Linden, E.N. van Loo, E.N. Jellen, P.J. Maughan, and M. Tester. 2017. The genome of *Chenopodium quinoa*. Nature 542:307–312.

Kistler, L., and B. Shapiro. 2011. Ancient DNA confirms a local origin of domesticated chenopod in eastern North America. Journal of Archaeological Science 38:3549–3554.

Kolano, B., J. McCann, M. Orzechowska, D. Siwinska, E. Temsch, and H. Weiss-Schneeweiss. 2016. Molecular and cytogenetic evidence for an allotetraploid origin of *Chenopodium quinoa* and *C. berlandieri* (Amaranthaceae). Molecular Phylogenetics and Evolution 100:109–123.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Maughan, P.J., L. Chaney, D.J. Lightfoot, B.J. Cox, M. Tester, E.N. Jellen, and D.E. Jarvis. 2019. Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (*Chenopodium quinoa* Willd.). Scientific Reports 9:185.

McKenna A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297–1303.

Smith, B.D. 1984. *Chenopodium* as a prehistoric domesticate in eastern North America: evidence from Russell Cave, Alabama. Science 226:165–167.

Smith, B.D., and R.A. Yarnell. 2009. Initial formation of an indigenous crop complex in eastern North America at 3800 B.P. Proceedings of the National Academy of Sciences USA 106(16):6561–6566.

Storchova, H., J. Drabesova, D. Chab, J. Kolar, and E.N. Jellen. 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):913–925.

Walsh, B.M., D. Adhikary, P.J. Maughan, E. Emshwiller, and E.N. Jellen. 2015. *Chenopodium* polyploidy inferences from Salt Overly Sensitive 1 (*SOS1*) data. American Journal of Botany 102(4):533–543.

Wilson, H.D. 1981. Genetic variation among South American populations of tetraploid *Chenopodium* sect. *Chenopodium* subsect. *Cellulata*. Systematic Botany 6(4):380–398.

Wilson, H.D., and C.B. Heiser, Jr. 1979. The origin and evolutionary relationships of 'Huauzontle' (*Chenopodium nuttaliae* Safford), domesticated chenopod of Mexico. American Journal of Botany 66(2):198–206.

Wilson, H., and J. Manhart. 1993. Crop/weed gene flow: *Chenopodium quinoa* Willd. and *C. berlandieri* Moq. Theoretical and Applied Genetics 86:642–648.