PopAmaranth: A population genetic genome browser for grain amaranths and their wild relatives

José Gonçalves-Dias* and Markus G. Stetter†,‡,1
*Dept. of Plant Sciences, University of Cologne, Cologne, Germany, †Cluster of Excellence on Plant Sciences, University of Cologne, Cologne, Germany

The combination of genomic, physiological, and population genetic research has accelerated the understanding and improvement of numerous crops. For non-model crops the lack of interdisciplinary research hinders their improvement. Grain amaranth is an ancient nutritious pseudocereal that has been domesticated three times in different regions of the Americas. We present and employ PopAmaranth, a population genetic genome browser, which provides an accessible representation of the genetic variation of the three grain amaranth species (A. hypochondriacus, A. cruentus, and A. caudatus) along the A. hypochondriacus reference sequence. We performed population-scale diversity and selection analysis from whole-genome sequencing data of 88 curated genetically and taxonomically unambiguously classified accessions. We employ the platform to show that genetic diversity in the water stress-related MIF1 gene declined during amaranth domestication and provide evidence for convergent saponin reduction between amaranth and quinoa. PopAmaranth is available through amaranthGDB at amaranthgdb.org/popamaranth.html

KEYWORDS Amaranthus, Orphan Crop, Genetic Diversity, Genome Browser, Population genetics

Introduction

Genome sequencing, genome-assisted breeding, and molecular breeding techniques have accelerated the improvement of numerous major crops (Wallace et al. 2018; Lemmon et al. 2018). The availability of genome-wide diversity data of crops and their wild relatives has allowed to identify and study candidate genes of agronomic significance (Hufford et al. 2012; Huang et al. 2012; Wang et al. 2020). These candidates can then be validated through molecular genetics (Ross-Ibarra et al. 2007; Fernie and Yan 2019; Sedeek et al. 2019; Wang et al. 2020). To facilitate the interdisciplinary use of population genetic results, it is essential to provide summary statistics in an intuitive and user-friendly way.

Different platforms have been developed to make genomic resources available across disciplines and have enabled the integration of complementary research areas (Lawrence et al. 2004; Alonso-Blanco et al. 2016; Jin et al. 2013). Online genome browser platforms such as Ensemble (Bolser et al. 2016) and Phytozome (Goodstein et al. 2012) have become a standard interface to interact with genome sequences and annotations and are used across research fields. Genome browsers provide access to reference genome sequences and gene annotations for numerous plant species but most browsers only provide data for a single reference individual per species. Species-specific browsers include sequence data and variant calls for a large number of individuals (e.g., Lawrence et al. 2004; Dash et al. 2016; Krishnakumar et al. 2015; Mansueto et al. 2017; Kudo et al. 2017), but do not allow

© The Author(s) (2021). Published by Oxford University Press on behalf of the Genetics Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited.
For commercial re-use, please contact journals.permissions@oup.com
including molecular gene function identification (Massange-Sanchez et al. 2016), state-of-the-art transient ‘hairy’ roots expression systems (Castellanos-Arévalo et al. 2020), and stress physiology assays (Parra-Cota et al. 2014; Massange-Sanchez et al. 2015). Combined, these resources can elevate amaranth research and improvement if results and data are available and accessible for researchers across disciplines.

Here, we present PopAmaranth, an interactive genome-wide population genetic browser for amaranth. PopAmaranth facilitates browsing a number of population genetic summary statistics and selection signals, gene annotation, and variant calls of the three grain amaranths and two wild relatives along the amaranth genome. We defined a curated set of 88 morphologically and genetically identified samples with whole-genome sequencing data to represent the five populations. Currently, PopAmaranth provides three categories of summary statistics, namely genetic diversity, population differentiation, and selection signals, plus variant calls and annotation tracks, in a total of more than 40 tracks. We show how the tool allows a user-friendly way to screen evolutionary signals for candidate genes and compare them between populations by identifying selection signals in a stress gene previously identified in one of the grain amaranths and in an orthologue quinoa domestication gene that shows convergent signals of selection in amaranth. PopAmaranth is embedded in amaranthGDB and is accessible from amaranthgdb.org/popamaranth.html.

**Methods**

**Data and filtering**

We used whole-genome sequencing data of 116 accession from five amaranth species, including the three grain amaranths (24 A. hypochondriacus, 24 A. cruentus, and 34 A. caudatus samples) and their two wild relatives, 9 A. hybridus and 25 A. quitensis (Stetter et al. 2020, Table S1). The sequencing reads were aligned to the A. hypochondriacus reference sequence V 2.0 (Lightfoot et al. 2017).

We performed principal component analysis (PCA) on the full set of accessions to remove individuals with ambiguous species clustering using PCAngsd (version 0.982) (Meisner and Albrechtsen 2018) followed by prcomp to calculate principal components. We manually excluded samples that did not genetically cluster with the morphologically designated species information in their passport data after visual evaluation of the first three PCs (Figure S1). We only used bam files of remaining individuals for summary statistic estimation and subset the VCF file from Stetter et al. (2020) using VCFtools 0.1.16 (Danecek et al. 2011) to only include sites that segregate in this set.

We calculated the site allele frequency likelihood based on individual genotype likelihoods for each of the five species using the -doSaf 1 function on ANGSD (version 0.930) (Korneliussen et al. 2014). We removed sites with a minimum map quality below 30, minimum base qscore below 20, and a flagstat (Li et al. 2009) above 255, keeping only primary reads (-doSaf 1, -GL 2, -remove_bads 1, -minMapQ 30, -minQ 20). In addition, we removed all sites with more than 66% missing values (-minInd=1/3/n).

**Population genetic browser tracks**

Using realSFS saf2theta functions on ANGSD, we calculated the folded site frequency spectrum and estimated per site thetas (population scaled mutation rate). Consequently, we calculated nucleotide diversity ($\pi$) and Wu and Watterson estimator ($\theta$) in non-overlapping windows of 5000 bp using the do_stats function of ANGSD. We only kept windows with more than 30% of the sites called in a given window.

To data as browser tracks we converted the files to bigWig format using UCSC bedgraphToBigWig (Kent et al. 2010). Within the genome browser, a yellow horizontal line denotes the genome-wide mean for each of the summary statistics, values below the mean are shown in red and above the mean in blue and we indicated the strength of deviation by adding dark gray and light grey shadings for one and two standard deviations from the mean, respectively.

We calculated pairwise Weir-Cockerham $F_{st}$ (Wright 1950) as a measure for genetic differentiation for each pair of populations using ANGSD (Korneliussen et al. 2014). We used these values as input to calculate pairwise $F_{st}$ in non-overlapping windows of 5000bp along the genome.

We employed ANGSD (Korneliussen et al. 2014) with the parameters described above for $\pi$ and $\theta$ to calculate Tajima’s D in non-overlapping 5 kb windows. Using the nucleotide diversity estimated for each of the species, we calculated relative nucleotide diversity. We divided $\pi$ for each of the domesticated species (A. caudatus, A. cruentus, and A. hypochondriacus) by $\pi$ of their wild ancestor, A. hybridus. We only used windows where both species had data after filtering for the number of genotyped sites.

Variant based statistics were calculated base on the subsampled VCF data from Stetter et al. (2020). We used the scikit-allel python library (https://doi.org/10.5281/zenodo.597309) to calculate per site heterozygosity statistics ($H_{xy}$ $H_{bsc}$, and $F$) for each of the five populations. We applied Raised Accuracy in Sweep Detection (RAiSD) (Alachiotis and Pavlidis 2018) with default setting (20 SNP windows) on the subset VCF data from Stetter et al. (2020) to detect signals of selective sweeps within each population. We consider windows on the top 1% $\mu$ values as outliers and under positive selection (A. caudatus: 17650 windows; A. cruentus 16546; A. hypochondriacus: 17932 A. hybridus: 43415; and A. quitensis: 15854). We merged all overlapping windows to create stretches of selective sweeps.

**Browser implementation and annotation**

We provided access to the summary statistics described above as an interactive tool through JBrowse 1.16.9 ( Skinner et al. 2009). We added the reference sequence and gene annotation, including exons, introns, CDS, mRNA, and UTRs from Lightfoot et al. (2017) available through Phytozome (Goodstein et al. 2012). For each summary statistic a color gradient summary plot combining all the sites called in a given window.

We employed ANGSD (Korneliussen et al. 2014) with the parameters described above for $\pi$ and $\theta$ to calculate Tajima’s $D$ in non-overlapping 5 kb windows. Using the nucleotide diversity estimated for each of the species, we calculated relative nucleotide diversity. We divided $\pi$ for each of the domesticated species (A. caudatus, A. cruentus, and A. hypochondriacus) by $\pi$ of their wild ancestor, A. hybridus. We only used windows where both species had data after filtering for the number of genotyped sites.

Variant based statistics were calculated base on the subsampled VCF data from Stetter et al. (2020). We used the scikit-allel python library (https://doi.org/10.5281/zenodo.597309) to calculate per site heterozygosity statistics ($H_{xy}$ $H_{bsc}$, and $F$) for each of the five populations. We applied Raised Accuracy in Sweep Detection (RAiSD) (Alachiotis and Pavlidis 2018) with default setting (20 SNP windows) on the subset VCF data from Stetter et al. (2020) to detect signals of selective sweeps within each population. We consider windows on the top 1% $\mu$ values as outliers and under positive selection (A. caudatus: 17650 windows; A. cruentus 16546; A. hypochondriacus: 17932 A. hybridus: 43415; and A. quitensis: 15854). We merged all overlapping windows to create stretches of selective sweeps.

**PopAmaranth application to candidate genes**

We downloaded the sequence of the water stress-related MIF1 gene reported in Huerta-Ocampo et al. (2011) from the NCBI database and used BLASTn (Altschul et al. 1990) to identify the gene ID in the A. hypochondriacus V2 reference sequence on Phytozome. Using the same procedure, we studied the triterpene saponin biosynthesis activating regulator-1 (TSAR-1) gene from Chenopodium quinoa (Jarvis et al. 2017).
Results

Sample filtering
Amaranthus species are difficult to taxonomically classify because of their high morphological similarity (Sauer 1967). Therefore, we sub-sampled the original dataset from Stetter et al. (2020) based on the genetic clustering in the PCA and species delimitation in Germplasm Resources Information Network (GRIN). We selected each species according to their clustering in the first three principal components (Figure S1). After filtering, our sample consisted of 88 genetically and morphologically defined samples representing the five species, with 28 individuals classified as A. caudatus L., 21 A. cruentus L., 18 A. hypochondriacus L., 12 Amaranth quitensis Kunth, and 9 A. hybridus L. (Figure 1 and table S1).

![Figure 1 Principal Component Analysis with filtered samples](https://academic.oup.com/g3journal/advance-article/doi/10.1093/g3journal/jkab103/6208888)

Categories and Tracks
We created PopAmaranth relative to the high-quality A. hypochondriacus reference genome (Lightfoot et al. 2017) and added the gene annotation as functional guide. We calculated nine summary statistics from whole-genome sequencing data for each of the five species. The tracks are grouped into five categories, namely annotation, differentiation, diversity, selection, and variant calls (Table 1 and S2). Each category includes tracks one color gradient summary track combining data of a summary statistic for all species.

Differentiation Tracks in the differentiation category represent all pairwise Fst comparisons in 5 kb windows. The genome-wide pairwise Fst ranged from 0.17 between A. caudatus and A. quitensis to 0.69 between A. caudatus and A. cruentus. As observed before, Fst between crop species was higher than between the crops and their wild ancestor for A. caudatus and A. hypochondriacus (Stetter et al. 2020). Although, we found higher Fst between A. cruentus and A. hybridus (0.69) than between A. cruentus and A. hypochondriacus (0.57).

Diversity Genetic diversity patterns along the genome can give insights into the evolutionary history of a population. Hence, we calculated several diversity statistics along the genome. Inbreeding coefficients and expected and observed heterozygosity are reported on a per-site basis for each SNP that segregated within a population. In addition to SNP-based statistics, we provide windowed diversity measures, including Wu & Watterson θ and nucleotide diversity π in 5 kb non-overlapping windows. Consistent with previous findings, the three grain amaranths had a lower mean π (0.005-0.010) compared to their wild ancestor A. hybridus (0.019) (Stetter et al. 2020). Wu & Watterson θ was also lower for domesticated amaranth species (0.004-0.007) compared to A. hybridus (0.023).

Selection We calculated three different summary statistics to detect signals of selection along the genome. Tracks displaying Tajima’s D were calculated in 5 kb windows for each species. Tajima’s D was higher for domesticated species (1.443 in A. caudatus, 1.773 in A. cruentus, and -0.105 in A. hypochondriacus) than for their wild ancestor A. hybridus (-0.597), indicating a domestication bottleneck. A. quitensis had a mean Tajima’s D of 2.037 also suggesting a recent population contraction.

We employed RAiSD to detect signals of selective sweeps in 20 SNP windows within each species. The top 1% of all windows were considered outliers and suggest regions of positive selection. After merging adjacent outliers, we found 973 non-overlapping windows with positive selection signals in A. caudatus, 1,096 in A. cruentus, 1,121 A. hypochondriacus, 2,452 A. hybridus, and 1,275 windows in A. quitensis. To investigate the signal of domestication-related selection, we added the relative nucleotide diversity between each crop and their wild ancestor A. hybridus in 5 kb windows. While the genome-wide π was lower for all three crops (see "Diversity"), relative π allows to visualize deviations from this genome-wide mean and detect outlier signals in individual regions.

Variant Calls Individual variants give access to an individuals’ genotype. Molecular biologists might be interested in evaluating natural alleles of a gene of interest, and plant breeders could use individuals with specific variants to enrich their gene pools. We provide variant data for all five species representing their genotype frequency within the population. Each variant track only displays variants within the given population (not including fixed variants between populations). A total of 4,961,210 variants for A. caudatus, 4,075,368 for A. cruentus, 4,551,278 for A. hypochondriacus, 12,238,589 for A. hybridus, and 2,342,505 for A. quitensis along the genome are available.

PopAmaranth case study
To show the utility of PopAmaranth, we evaluated the evolutionary signals for a gene that was molecularly shown to be involved in the response of A. hypochondriacus to water stress (Huerta-Ocampo et al. 2011). We found that MIF1 (AH-017582) showed lower nucleotide diversity, decreased expected heterozygosity, and a relative nucleotide diversity below the genome-wide average in all three grain amaranth species. Also, we identified a selective sweep in A. hypochondriacus around this gene (Figure 2). Our findings combined with the “Variant call” tracks in the browser allow to select accessions with contrasting genotypes to identify the causal allele for the expression difference in A. hypochondriacus and compare wild and domesticated amaranth for their drought response.

In addition to the amaranth specific use, PopAmaranth facilitates the evaluation of hypothesis beyond the species. To show its utility to study convergent selection signals across distant families, we evaluated population genetic signals around the amaranth ortholog to the triterpene saponin biosynthesis activating regulator 1 - TSARI (AH-019562), a key regulator for seed saponin content in Chenopodium quinoa (Jarvis et al. 2017). We found signals of selective sweeps in the three grain amaranth
We identify previously unknown signals of selection in stress-tolerance of genotypes with different alleles within grain amaranths and the higher-level framework of weedy amaranth, enabling the identification of stress-tolerance alleles and potentially the reintroduction of such alleles into the crop domestication. In most crops, domestication led to a reduction in stress resilience compared to their wild ancestor was below the genome-wide mean, suggesting selection during amaranth domestication (Figure S2).

Discussion

Over the last decades, large-scale population genomic data revealed insights into the evolution and adaptation of crops. Providing access to results in a user-friendly and interactive way opens paths to better integrate data from different research areas. Our population genomic genome browser, PopAmaranth, aims to provide such an intuitive tool for amaranth population genetic results. The inclusion of five different species involved in the crop domestication history of amaranth facilitates hypothesis testing along this evolutionary gradient.

For other plant species, i.e., maize (Lawrence et al. 2004), tomato (Fernandez-Pozo et al. 2015), and arabidopsis (Alonso-Blanco et al. 2016) accessible platforms of genomic and evolutionary data are integral parts of the research communities. We hope that PopAmaranth and the higher-level framework amaranthGDB will help establish an amaranth community that benefits from the interdisciplinary exchange.

Our results show how PopAmaranth can be employed to add an evolutionary perspective to different molecular questions. We identify previously unknown signals of selection in stress-related MIF1 gene, which might have been under selection during amaranth domestication. In most crops, domestication led to a reduction in stress resilience compared to their wild ancestors. Hence, the reduction in diversity might represent selection against the tolerant allele to free resources for increased crop productivity (Koziol et al. 2012). Our browser allows the selection of genotypes with different alleles within grain amaranths and in wild amaranth, enabling the identification of stress-tolerance alleles and potentially the reintroduction of such alleles into breeding programs.

On a broader scale, PopAmaranth also facilitates the comparison of convergent adaptation signals between more distant taxa. For instance, our finding of convergent selection between quinoa and amaranth in a saponin-related gene suggests that in both quinoa and amaranth the saponin content was reduced to improve the palatability of the grains (Jarvis et al. 2017). Saponins confer toxicity to protect wild plants against birds but reduce the nutritional quality of seeds for human consumption and animal feed (Oleszek et al. 1999; Mroczek 2015). Hence, our platform allowed to identify the convergent selection between the two pseudocereals, demonstrating its utility to evaluate selection signals across taxa. This is of particular use for close relatives of weedy Amaranthus species that are of evolutionary and agronomic interest and have been aligned to the same reference genome used in our browser (Montgomery et al. 2020).

We aimed for a generalized usage of diversity and differentiation estimates. Therefore, we only selected unambiguous samples of each species, based on morphological and genetic classifications. A clear grouping is crucial for a reference tool, as misclassified samples would confound population-wide signals (Rieseberg and Wendel 2004). Our sub-sampling approach is conservative regarding genetic diversity, as it excludes more differentiated individuals from the analysis. Reported values of genetic differentiation (Fst) between species could be inflated due to the lack of intermediate individuals. The increased differentiation by sub-sampling potentially led to the higher Fst value between A. cruentus and A. hybridus compared to previous results (Stetter et al. 2020). While there is a trade-off between including additional individuals and the potential for undiscovered diversity, our goal was a defined and distinguished set of samples representing each species. The inclusion of only core individuals of each species further allows the comparison and classification of less distinct individuals using our set.

| Table 1 Tracks available in PopAmaranth |
|-----------------------------------------|
| **Track**                              | **Description**                             |
| Reference Genome v2.0                  | Amaranthus hypochondriacus reference genome v2.0 (Lightfoot et al. 2017) |
| Gene Annotation v2.1                   | Amaranthus hypochondriacus gene annotation with subfeatures, including CDS, mRNA and UTRs |
| **Differentiation**                    |                                             |
| Fst                                     | Fixation Index, average pairwise differences (Weir and Cockerham 1984) |
| Diversity                               |                                             |
| Wu & Watterson θ                       | Estimator of genetic diversity in a population (Watterson 1975) |
| Expected heterozygosity                | Expected heterozygosity for a SNP under Hardy-Weinberg equilibrium |
| Observed heterozygosity                | Observed heterozygosity for a SNP genotype. |
| Inbreeding coefficient                 | Inbreeding coefficient (F) for each variant |
| Nucleotide diversity (π)               | Nei’s nucleotide diversity (Nei and Li 1979) |
| **Selection**                          |                                             |
| Tajima’s D                             | Scaled difference between the mean number of pairwise differences and the number of segregating sites (Tajima 1989) |
| Relative nucleotide diversity          | Ratio of nucleotide diversity between a domesticated species and their wild ancestor (A. hybridus) |
| Selective Sweep (RAiSD (µ))            | µ statistic for selective sweep detection |
| VCF Call                               | Called SNPs with a given species and their genotype frequency |

species. Furthermore, the relative diversity compared to the wild ancestor was below the genome-wide mean, suggesting selection during amaranth domestication (Figure S2).
Altogether, we incorporated a well-defined set of individuals with congruent data filtering to estimate population-wide diversity statistics for the three grain amaranth species and two wild relatives. The identification of selection signals in candidate genes within amaranth and beyond shows the utility of the browser for a range of researchers. PopAmaranth and the amaranthGDB platform will help build and grow the amaranth research community and facilitate interdisciplinary research to ultimately improve the crop.

Availability

PopAmaranth is available at https://amaranthgdb.org/popamaranth.html. A static version of the browser and data in table format can be found at: https://doi.org/10.6084/m9.figshare.13340798 Code is available at https://github.com/cropevolution/PopAmaranth. Tracks data for the region in observation can be downloaded directly from the tracks options.

Acknowledgments

We thank the RRZ at University of Cologne, for hosting PopAmaranth, Benedict Wieters and the de Meaux lab for testing and feedback on the browser, and all members of the Stetter lab for discussion and suggestions.

Funding and Competing Interests

We acknowledge the support of the Deutsche Forschungsgemeinschaft under Germany’s Excellence Strategy – EXC-2048/1 – Project ID 390686111 to MGS. The authors declare no conflict of interest.

References

Alachiotis, N. and P. Pavlidis, 2018 RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Communications biology 1: 1–11.
Alonso-Blanco, C., J. Andrade, C. Becker, F. Bemm, J. Bergelson, et al., 2016 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. Cell 166: 481–491.
Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. Journal of molecular biology 215: 403–410.
Bolser, D., D. M. Staines, E. Pritchard, and P. Kersey, 2016 Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. In Plant bioinformatics, pp. 115–140, Springer.
Casillas, S., R. Mulet, P. Villegas-Mirón, S. Hervas, E. Sanz, et al., 2018 PopHuman: the human population genomics browser. Nucleic acids research 46: D1003–D1010.
Castellanos-Arévalo, A. P., A. A. Estrada-Luna, J. L. Cabrera-Ponce, E. Valencia-Lozano, H. Herrera-Ubaldo, et al., 2020 Agrobacterium rhizogenes-mediated transformation of grain (Amaranthus hypochondriacus) and leafy (A. hybridus) amaranths. Plant Cell Reports.
Clouse, J., D. Adhikary, J. Page, T. Ramaraj, M. Deyholos, et al., 2016 The amaranth genome: genome, transcriptome, and physical map assembly. The Plant Genome 9: 1–14.
Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, et al., 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158.
Dash, S., J. D. Campbell, E. K. Cannon, A. M. Cleary, W. Huang, et al., 2016 Legume information system (LegumeInfo.org): a key component of a set of federated data resources for the legume family. Nucleic acids research 44: D1181–D1188.
Fernández-Pozo, N., N. Menda, J. D. Edwards, S. Saha, I. Y. Tecle, et al., 2015 The Sol Genomics Network (SGN)—from...
genotype to phenotype to breeding. Nucleic acids research 43: D1036–D1041.
Fernie, A. R. and J. Yan, 2019 De novo domestication: an alternative route toward new crops for the future. Molecular plant 12: 615–631.
Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, et al., 2012 Phytozome: a comparative platform for green plant genomics. Nucleic acids research 40: D1178–D1186.
Huang, X., N. Kurata, Z.-X. Wang, A. Wang, Q. Zhao, et al., 2012 A map of rice genome variation reveals the origin of cultivated rice. Nature 490: 497–501.
Huerta-Ocampo, J., M. León-Galván, L. Ortega-Cruz, A. Barrera-Pacheco, A. De León-Rodriguez, et al., 2011 Water stress induces up-regulation of DOF1 and MIF1 transcription factors and down-regulation of proteins involved in secondary metabolism in amaranth roots (Amaranthus hypochondriacus L.). Plant Biology 13: 472–482.
Hufford, M. B., X. Xu, J. Van Heerwaarden, T. Pyhäjärvi, J.-M. Chia, et al., 2012 Comparative population genomics of maize domestication and improvement. Nature genetics 44: 808.
Jarvis, D. E., Y. S. Ho, D. J. Lightfoot, S. M. Schmölkel, B. Li, et al., 2017 The genome of Chenopodium quinoa. Nature 542: 307.
Jin, J., J. Liu, H. Wang, L. Wang, and N.-H. Chua, 2013 PLncDB: plant long non-coding RNA database. Bioinformatics 29: 1068–1071.
Joshi, D. C., S. Sood, R. Hosahatti, L. Kant, A. Pattanayak, et al., 2018 From zero to hero: the past, present and future of grain amaranth breeding. Theoretical and Applied Genetics 131: 1807–1823.
Kent, W. J., A. S. Zweig, G. Barber, A. S. Hinrichs, and D. Karolchik, 2010 BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26: 2204–2207.
Korneliussen, T. S., A. Albrechtsen, and R. Nielsen, 2014 ANGSD: analysis of next generation sequencing data. BMC bioinformatics 15: 356.
Koziol, L., L. H. Riesenberg, N. Kane, and J. D. Bever, 2012 Reduced drought tolerance during domestication and the evolution of weediness results from tolerance–growth trade-offs. Evolution: International Journal of Organic Evolution 66: 3803–3814.
Krishnakumar, V., M. R. Hanlon, S. Contrino, E. S. Ferlanti, S. Karamycheva, et al., 2015 Araport: the Arabidopsis information portal. Nucleic acids research 43: D1003–D1009.
Kudo, T., M. Kobayashi, S. Terashima, M. Katayama, S. Ozaki, et al., 2017 TOMATOMICS: a web database for integrated omics information in tomato. Plant and Cell Physiology 58: e8–e8.
Lawrence, C. J., Q. Dong, M. L. Polacco, T. E. Seigfried, and V. Brendel, 2004 MaizeGDB, the community database for maize genetics and genomics. Nucleic acids research 32: D393–D397.
Leanon, Z. H., N. T. Reem, J. Dalrymple, S. Soyk, K. E. Swartwood, et al., 2018 Rapid improvement of domestication traits in an orphan crop by genome editing. Nature plants 4: 766–770.
Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, et al., 2009 The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079.
Lightfoot, D., D. E. Jarvis, T. Ramaraj, R. Lee, E. Jellen, 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: 74.
Mallory, M. A., R. V. Hall, A. R. McNabb, D. B. Pratt, E. N. Jellen, et al., 2008 Development and characterization of microsatellite markers for the grain amaranths. Crop science 48: 1098–1106.
Mansuetu, L., R. R. Fuentes, F. N. Borja, J. Detras, J. M. Abriol-Santos, et al., 2017 Rice SNP-Seek database update: new SNPs, indels, and queries. Nucleic acids research 45: D1075–D1081.
Massange-Sanchez, J. A., P. A. Palmeros-Suarez, E. Espitia-Rangel, I. Rodriguez-Arevalo, L. Sanchez-Segura, et al., 2016 Overexpression of grain amaranth (Amaranthus hypochondriacus) AhERF or AhDOF transcription factors in Arabidopsis thaliana increases water deficit-and salt-stress tolerance, respectively, via contrasting stress-amelioration mechanisms. PloS one 11: e0164280.
Massange-Sanchez, J. A., P. A. Palmeros-Suarez, N. A. Martinez-Gallardo, P. A. Castrillon-Arbelaez, H. Aviles-Arnaut, et al., 2015 The novel and taxonomically restricted Ah24 gene from grain amaranth (Amaranthus hypochondriacus) has a dual role in development and defense. Frontiers in plant science 6: 602.
Mayes, S., F. Massawe, P. Alderson, J. Roberts, S. Azam-Ali, et al., 2012 The potential for underutilized crops to improve security of food production. Journal of experimental botany 63: 1075–1079.
Meinszer, N. and A. Albrechtsen, 2018 Inferring population structure and admixture proportions in low-depth NGS data. Genetics 210: 719–731.
Montgomery, J. S., D. Giacomin, B. Waithaka, C. Lanz, B. P. Murphy, et al., 2020 Draft Genomes of Amaranthus tuberculatus, Amaranthus hybridus, and Amaranthus palmeri. Genome biology and evolution 12: 1988–1993.
Mróczek, A., 2015 Phytochemistry and bioactivity of triterpene saponins from Amaranthaceae family. Phytochemistry Reviews 14: 577–605.
Nei, M. and W.-H. Li, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences 76: 5269–5273.
Oleszek, W., M. Junkuszew, and A. Stochmal, 1999 Determination and toxicity of saponins from Amaranthus cruentus seeds. Journal of Agricultural and Food Chemistry 47: 3685–3687.
Parra-Cota, F. L., J. J. Peña-Cabriles, S. de los Santos-Villalobos, N. A. Martínez-Gallardo, and J. P. Delano-Frier, 2014 Burkholdeia ambifaria and B. caribensis promote growth and increase yield in grain amaranth (Amaranthus cruentus and A. hypochondriacus) by improving plant nitrogen uptake. PloS one 9: e88094.
Riesenberg, L. H. and J. Wendel, 2004 Plant speciation: Rise of the poor cousins. The New Phytologist 161: 3–8.
Ross-Ibarra, J., P. L. Morrell, and B. S. Gaut, 2007 Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proceedings of the National Academy of Sciences 104: 8641–8648.
Sauer, J. D., 1967 The grain amaranths and their relatives: a revised taxonomic and geographic survey. Annals of the Missouri Botanical Garden 54: 103–137.
SedeeK, K. E., A. Mahas, and M. Mahfouz, 2019 Plant genome engineering for targeted improvement of crop traits. Frontiers in plant science 10: 114.
Skinner, M. E., A. V. Uzilov, L. D. Stein, C. J. Mungall, and I. H. Holmes, 2009 JBrowse: a next-generation genome browser. Genome research 19: 1630–1638.
American grain amaranth (*Amaranthus caudatus*). Molecular ecology **26**: 871–886.

Stetter, M. G., M. Vidal-Villarejo, and K. J. Schmid, 2020 Parallel seed color adaptation during multiple domestication attempts of an ancient new world grain. Molecular Biology and Evolution **37**: 1407–1419.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.

Wallace, J. G., E. Rodgers-Melnick, and E. S. Buckler, 2018 On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. Annual review of genetics .

Wang, B., Z. Lin, X. Li, Y. Zhao, B. Zhao, *et al.*, 2020 Genome-wide selection and genetic improvement during modern maize breeding. Nature Genetics pp. 1–7.

Watterson, G., 1975 On the number of segregating sites in genetical models without recombination. Theoretical population biology **7**: 256–276.

Weir, B. S. and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. evolution pp. 1358–1370.

Wright, S., 1950 Genetical structure of populations. Nature **166**: 247–249.