

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

José Gonçalves-Dias¹ and Markus G. Stetter ^{1,2,*}

¹Institute for Plant Sciences, University of Cologne, Cologne 50674, Germany

²Cluster of Excellence on Plant Sciences, University of Cologne, Cologne 50674, Germany

*Corresponding author: Department for Plant Sciences, University of Cologne, Cologne, Germany. m.stetter@uni-koeln.de

Abstract

The combination of genomic, physiological, and population genetic research has accelerated the understanding and improvement of numerous crops. For nonmodel 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 (*Amaranthus hypochondriacus*, *Amaranthus cruentus*, and *Amaranthus caudatus*) and two wild relatives (*Amaranthus hybridus* and *Amaranthus quitensis*) 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 (Lemmon et al. 2018; Wallace 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 (Huang et al. 2012; Hufford 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; Jin et al. 2013; Alonso-Blanco et al. 2016). Online genome browser platforms such as Ensemble (Bolser et al. 2016) and Phytosome (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; Krishnakumar et al. 2015; Dash et al. 2016; Kudo et al. 2017; Mansueto et al. 2017), but do not allow a

direct inference of a population scale genome-wide diversity across related species. Population genetic genome browsers that provide population-scale summary statistics are available for only a few non-plant model species (Casillas et al. 2018). For plant and crop species, in particular minor crops, such resources are currently unavailable.

Novel and under-utilized crops have a high potential to contribute to sustainable food production, as many such crops are tolerant to abiotic and biotic factors and are of high nutritional value (Mayes et al. 2012). Amaranth is an under-utilized crop that has been cultivated for its grains as pseudocereal and its edible leaves as a vegetable (Sauer 1967; Joshi et al. 2018). Three-grain amaranth species, *Amaranthus caudatus* L., *Amaranthus cruentus* L., and *Amaranthus hypochondriacus* L., have been domesticated for their grain from a common wild ancestor, *A. hybridus* L. (Stetter et al. 2020). Another wild relative, *A. quitensis* Kunth, is suspected to be involved in the domestication of the South American *A. caudatus*, although its role and contribution to the crop remain unclear (Stetter et al. 2017, 2020). The repeated domestication of amaranth presents an interesting model to study genetic parallelisms along selection gradients, and the combination of genomics, quantitative genetics, and molecular dissection of gene function has a high potential to improve grain amaranth.

First resources that allow the functional study of traits have been developed for amaranth. On the one hand, numerous genomic resources, including a high-quality reference genome

Received: December 17, 2020. Accepted: March 07, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-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

(Lightfoot *et al.* 2017) and transcriptome (Clouse *et al.* 2016), genome-wide marker data (Mallory *et al.* 2008; Stetter *et al.* 2017, 2020), and QTL regions for different traits (Lightfoot *et al.* 2017; Stetter *et al.* 2020) have been identified. On the other hand, a number of molecular methods have been adapted for the crop, 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 ortholog 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 accessions 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, Supplementary 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 PCAnsd (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 (Supplementary 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) (Korneliusson *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 (π) and Wu and Watterson estimator (θ) in nonoverlapping 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 `bedgraph2bigwig` (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 and light gray 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 (Korneliusson *et al.* 2014). We used these values as input to calculate pairwise F_{st} in nonoverlapping windows of 5000 bp along the genome.

We employed ANGSD (Korneliusson *et al.* 2014) with the parameters described above for π and θ to calculate Tajima's D in nonoverlapping 5 kb windows. Using the nucleotide diversity estimated for each of the species, we calculated relative nucleotide diversity. We divided π for each of the domesticated species (*A. caudatus*, *A. cruentus*, and *A. hypochondriacus*) by π 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 based on the subsampled VCF data from Stetter *et al.* (2020). We used the `scikit-allel` python library (<https://doi.org/10.5281/zenodo.597309>, last accessed 04/07/2021) to calculate per site heterozygosity statistics (H_{exp} , H_{obs} , 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 considered windows on the top 1% μ 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 species was added. Further, we added the “Variant” category, providing variant data for biallelic SNPs within each species from Stetter *et al.* (2020) (not including variants fixed between populations).

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).

Data 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.

Results

Sample filtering

Amaranthus species are difficult to taxonomically classify because of their high morphological similarity (Sauer 1967). Therefore, we subsampled 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 (Supplementary 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 Supplementary Table S1).

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 a 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 Supplementary Table 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 F_{st} comparisons in 5 kb windows. The genome-wide pairwise F_{st} ranged from 0.17 between *A. caudatus* and *A. quitensis* to 0.68 between *A. caudatus* and *A. cruentus*. As observed before, F_{st} 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 a higher F_{st} 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 is 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 nonoverlapping 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.

Table 1 Tracks available in PopAmaranth

Track	Description
Annotation	
Reference Genome v2.0	<i>Amaranthus hypochondriacus</i> reference genome v2.0 (Lightfoot et al. 2017)
Gene Annotation v2.1	<i>Amaranthus hypochondriacus</i> gene annotation with subfeatures, including CDS, mRNA, and UTRs
Differentiation	
F_{st}	Fixation Index, average pairwise differences (Weir and Cockerham 1984)
Diversity	
Wu and 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 (<i>A. hybridus</i>)
Selective Sweep (RAiSD (μ))	μ statistic for selective sweep detection
Variant Call	
VCF	Called SNPs with a given species and their genotype frequency

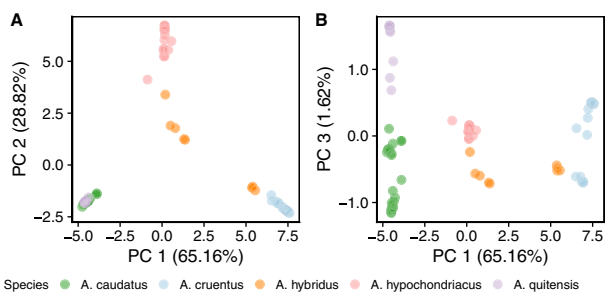


Figure 1 Principal component analysis with filtered samples. Each dot represents each of the 88 samples. *A. caudatus* (green), *A. cruentus* (blue), *A. hybridus* (orange), *A. hypochondriacus* (rose), and *A. quitensis* (purple). Axis shows the percentage of variance explained by each principal component.

After merging adjacent outliers, we found 973 nonoverlapping windows with positive selection signals in *A. caudatus*, 1096 in *A. cruentus*, 1121 *A. hypochondriacus*, 2452 *A. hybridus*, and 1275 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 - TSAR1 (AH-019562), a key regulator for seed saponin content in *C. quinoa* (Jarvis et al. 2017). We found signals of selective sweeps in the three-grain amaranth species. Furthermore, the relative diversity compared to the wild ancestor was below the genome-wide mean, suggesting selection during amaranth domestication (Supplementary 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 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 (Kozioł et al. 2012). Our browser allows the selection of genotypes

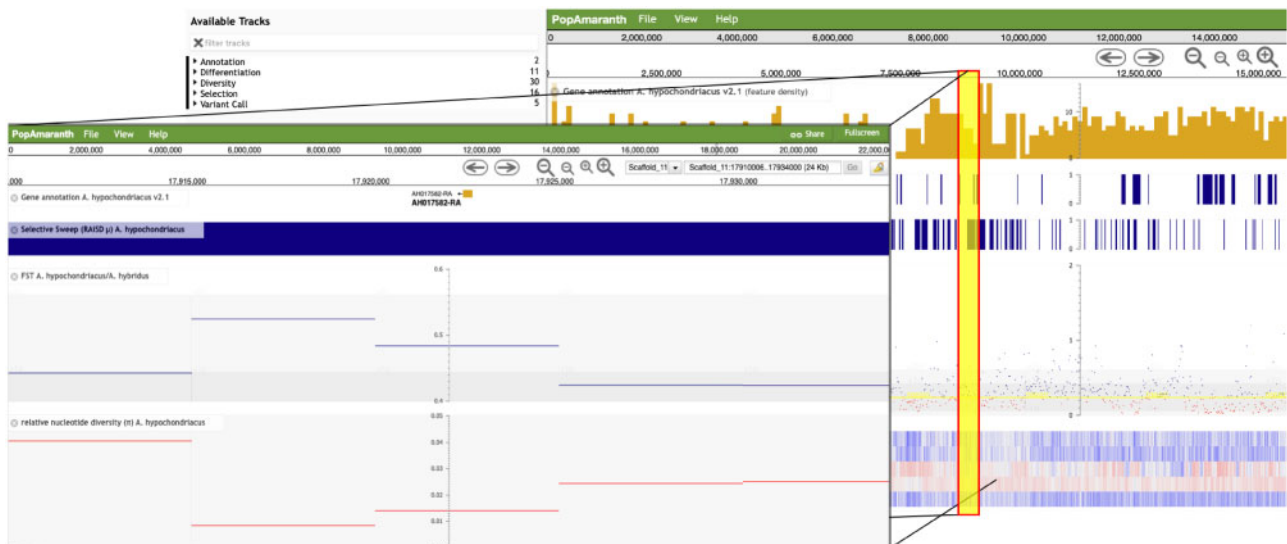


Figure 2 PopAmaranth screen view. Background panel: Zoomed out user view along a chromosome. Search field provides access to genome positions or gene names. Front panel: example is illustrated with a zoom-in region for the water-stress related *MIF1* gene (AH-017582). The blue bar shows a selective sweep in *A. hypochondriacus*, which is supported by high F_{st} between wild ancestor and crop and low π in the crop.

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 2003). Our subsampling approach is conservative regarding genetic diversity, as it excludes more differentiated individuals from the analysis. Reported values of genetic differentiation (F_{st}) between species could be inflated due to the lack of intermediate individuals. The increased differentiation by subsampling potentially led to the higher F_{st} 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.

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.

Acknowledgments

The authors thank the RRZ at the 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

This work was supported by funding from the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy—EXC-2048/1—Project ID 390686111, and DFG grant STE 2654/5-1.

Conflicts of interest

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

Literature cited

- Alachiotis N, Pavlidis P. 2018. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun Biol.* 1:1–11.
- Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, *et al.* 2016. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell.* 166:481–491.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Bolser D, Staines DM, Pritchard E, Kersey P. 2016. Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. *Methods Mol Biol.* 1374:115–140.
- Casillas S, Mulet R, Villegas-Mirón P, Hervas S, Sanz E, *et al.* 2018. PopHuman: the human population genomics browser. *Nucleic Acids Res.* 46:D1003–D1010.
- Castellanos-Arévalo AP, Estrada-Luna AA, Cabrera-Ponce JL, Valencia-Lozano E, Herrera-Ubaldo H, *et al.* 2020. *Agrobacterium* rhizogenes-mediated transformation of grain (*Amaranthus hypochondriacus*) and leafy (*A. hybridus*) amaranths. *Plant Cell Rep.* 39:1143–1160.
- Clouse J, Adhikary D, Page J, Ramaraj T, Deyholos M, *et al.* 2016. The amaranth genome: genome, transcriptome, and physical map assembly. *Plant Genome.* 9:1–14.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics.* 27:2156–2158.
- Dash S, Campbell JD, Cannon EK, Cleary AM, Huang W, *et al.* 2016. Legume information system (LegumeInfo.org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Res.* 44:D1181–D1188.
- Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, *et al.* 2015. The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Res.* 43:D1036–D1041.
- Fernie AR, Yan J. 2019. *De novo* domestication: an alternative route toward new crops for the future. *Mol Plant.* 12:615–631.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, *et al.* 2012. Phytzome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40:D1178–D1186.
- Huang X, Kurata N, Wang Z-X, Wang A, Zhao Q, *et al.* 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature.* 490:497–501.
- Huerta-Ocampo J, León-Galván M, Ortega-Cruz L, Barrera-Pacheco A, De León-Rodríguez A, *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 Biol (Stuttg).* 13:472–482.
- Hufford MB, Xu X, Van Heerwaarden J, Pyhäjärvi T, Chia J-M, *et al.* 2012. Comparative population genomics of maize domestication and improvement. *Nat Genet.* 44:808–811.
- Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, *et al.* 2017. The genome of *Chenopodium quinoa*. *Nature.* 542:307–312.
- Jin J, Liu J, Wang H, Wong L, Chua N-H. 2013. PLncDB: plant long non-coding RNA database. *Bioinformatics.* 29:1068–1071.
- Joshi DC, Sood S, Hosahatti R, Kant L, Pattanayak A, *et al.* 2018. From zero to hero: the past, present and future of grain amaranth breeding. *Theor Appl Genet.* 131:1807–1823.

- Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*. 26:2204–2207.
- Korneliusson TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*. 15:356.
- Kozlowski L, Rieseberg LH, Kane N, Bever JD. 2012. Reduced drought tolerance during domestication and the evolution of weediness results from tolerance–growth trade-offs. *Evolution*. 66:3803–3814.
- Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, et al. 2015. Araport: the Arabidopsis information portal. *Nucleic Acids Res*. 43:D1003–D1009.
- Kudo T, Kobayashi M, Terashima S, Katayama M, Ozaki S, et al. 2017. TOMATOMICS: a web database for integrated omics information in tomato. *Plant Cell Physiol*. 58:e8.
- Lawrence CJ, Dong Q, Polacco ML, Seigfried TE, Brendel V. 2004. MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res*. 32:D393–D397.
- Lemmon ZH, Reem NT, Dalrymple J, Soyk S, Swartwood KE, et al. 2018. Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat Plants*. 4:766–770.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25: 2078–2079.
- Lightfoot D, Jarvis DE, Ramaraj T, Lee R, Jellen E, et al. 2017. Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biol*. 15:74.
- Mallory MA, Hall RV, McNabb AR, Pratt DB, Jellen EN, et al. 2008. Development and characterization of microsatellite markers for the grain amaranths. *Crop Sci*. 48:1098–1106.
- Mansueti L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, et al. 2017. Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Res*. 45:D1075–D1081.
- Massange-Sanchez JA, Palmeros-Suarez PA, Espitia-Rangel E, Rodriguez-Arevalo I, Sanchez-Segura L, 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 JA, Palmeros-Suarez PA, Martinez-Gallardo NA, Castrillon-Arbelaez PA, Aviles-Arnaut H, et al. 2015. The novel and taxonomically restricted Ah24 gene from grain amaranth (*Amaranthus hypochondriacus*) has a dual role in development and defense. *Front Plant Sci*. 6:602.
- Mayes S, Massawe F, Alderson P, Roberts J, Azam-Ali S, et al. 2012. The potential for underutilized crops to improve security of food production. *J Exp Bot*. 63:1075–1079.
- Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*. 210: 719–731.
- Montgomery JS, Giacomini D, Waithaka B, Lanz C, Murphy BP, et al. 2020. Draft Genomes of *Amaranthus tuberculatus*, *Amaranthus hybridus*, and *Amaranthus palmeri*. *Genome Biol Evol*. 12: 1988–1993.
- Mroczek A. 2015. Phytochemistry and bioactivity of triterpene saponins from *Amaranthaceae* family. *Phytochem Rev*. 14:577–605.
- Nei M, Li W-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA*. 76:5269–5273.
- Oleszek W, Junkuszew M, Stochmal A. 1999. Determination and toxicity of saponins from *Amaranthus cruentus* seeds. *J Agric Food Chem*. 47:3685–3687.
- Parra-Cota FI, Peña-Cabriales JJ, de los Santos-Villalobos S, Martínez-Gallardo NA, Délano-Frier JP. 2014. *Burkholderia 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.
- Rieseberg LH, Wendel J. 2003. Plant speciation: rise of the poor cousins. *New Phytol*. 161:3–8.
- Ross-Ibarra J, Morrell PL, Gaut BS. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc Natl Acad Sci USA*. 104:8641–8648.
- Sauer JD. 1967. The grain amaranths and their relatives: a revised taxonomic and geographic survey. *Ann Missouri Bot Garden*. 54: 103–137.
- Sedeek KE, Mahas A, Mahfouz M. 2019. Plant genome engineering for targeted improvement of crop traits. *Front Plant Sci*. 10:114.
- Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. 2009. JBrowse: a next-generation genome browser. *Genome Res*. 19: 1630–1638.
- Stetter MG, Müller T, Schmid KJ. 2017. Genomic and phenotypic evidence for an incomplete domestication of South American grain amaranth (*Amaranthus caudatus*). *Mol Ecol*. 26:871–886.
- Stetter MG, Vidal-Villarejo M, Schmid KJ. 2020. Parallel seed color adaptation during multiple domestication attempts of an ancient new world grain. *Mol Biol Evol*. 37:1407–1419.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123:585–595.
- Wallace JG, Rodgers-Melnick E, Buckler ES. 2018. On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annu Rev Genet*. 52:421–444.
- Wang B, Lin Z, Li X, Zhao Y, Zhao B, et al. 2020. Genome-wide selection and genetic improvement during modern maize breeding. *Nat Genet*. 52:565–567.
- Watterson G. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 7:256–276.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*. 38:1358–1370.
- Wright S. 1950. Genetical structure of populations. *Nature*. 166: 247–249.

Communicating editor: A. Kern