Monitoring rapid evolution of plant populations at scale with Pool-Sequencing

Lucas Czech[1,%], Yunru Peng[1,%], Patricia Lang[2], Tatiana Bellagio[1,2], Julia Hildebrandt[3], Katrin Fritschi[3], Rebecca Schwab[3], Beth Rowan[3], Niek Scheepens[4], Detlef Weigel[3], Francois Vasseur[3,5], GrENE-net consortium[$], Moises Exposito-Alonso[1,2,3,6*].

[1]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA
[2]Department of Biology, Stanford University, Stanford, CA 94305, USA
[3]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany.
[4]Institute of Evolution and Ecology, University of Tübingen, 72076 Tübingen, Germany.
[5]CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, F-34090 Montpellier, France
[6]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA
[$]GrENE-net consortium authors and affiliations listed in the appendix.
[%]Shared co-first authors
[*]To whom correspondence should be addressed: moisesexpositoalonso@gmail.com

Keywords: Evolve-and-Resequence in plants, Pool-Sequencing, rapid adaptation, GrENE-net.org.

Draft last updated Nov 25

Abstract

The change in allele frequencies of a population over time is the fundamental signal of evolution. By monitoring this signal, we can analyze the effects of natural selection and genetic drift on populations. To efficiently track the change, large experimental or wild populations can be sequenced as pools of individuals over time using next-generation sequencing. Here, we present a set of experiments using hundreds of genotypes of *Arabidopsis thaliana* to showcase the power of this approach to study rapid evolution at scale. First, we validate that sequencing DNA directly extracted from pools of flowers from multiple individuals produces comparable results to sequencing DNA pooled at equal quantity from multiple individuals. Further, sequencing pools of 25-50 individuals at ~40X coverage adequately recovers genome-wide frequencies in diverse populations ($r>0.95$). A large number of replication and data are required to robustly test evolutionary adaptation to different environments. Therefore, we provide open source tools that streamline sequencing data curation and calculate various population genetic statistics two orders of magnitude faster than current software. Finally, we conducted a multi-year outdoor evolution experiment to show signals of rapid evolution in multiple genomic regions. We demonstrate how these methods can be scaled to study hundreds of populations across many climates.

Introduction

Adaptation to new and local environments based on within-species genetic variation is a core question in ecology and evolutionary genetics. The study of population variation in phenotypes and genotypes across geographic landscapes that vary in local environments, and the common gardens quantify current selection in a specific environment is provided by field experiments in which multiple genotypes of a species are grown together and traits and fitness are measured, are powerful tools to measure natural selection from the environment (Clausen et al., 1941; Kingsolver et al., 2001; Savolainen et al., 2013). These have typically studied natural selection within a generation, showing strong natural selection, which paradoxically does not appear to occur across generations (Merilä et al., 2001). Multi-year experiments where phenotypes and genetics are tracked would be ideal to study these evolutionary forces and question their predictability (Grant and Grant, 2002; Nosil et al., 2018). Next-Generation Sequencing now allows scrutiny of the genome variation of a population repeatedly, often called "Evolve and Resequence" (E&R), has enabled the direct observation of evolutionary forces such as drift and natural selection in real time, especially in the bacterial and animal model systems *E. coli* and *D. melanogaster* (Bergland et al., 2014; Good et al., 2017; Schlötterer et al., 2014). In the traditional genome sequencing approach, each individual is processed independently into one DNA sequencing library. In the Pool-seq approach, multiple individuals sampled from the same population are processed into one DNA sequencing library (Futschik and Schlötterer, 2010). Therefore, while individual haplotypes are lost in the Pool-seq approach, population-level allele frequencies are obtained in a cost-effective manner (Schlötterer et al., 2014). The Pool-seq approach has been applied to study rapid selective sweeps and polygenic adaptation (Pritchard et al., 2010) and the associations of these to meta-data such as climate (Günther and Coop, 2013; Hancock et al., 2011) or quantitative traits (Endler et al., 2016).

The decreasing cost of next-generation sequencing has enabled affordable re-sequencing, but we argue two other key improvements in library preparation and computation allow E&R to be scaled up to dozens to hundreds of wild populations over time. Genomic DNA library preparation using Tn5 transposase (Baym et al., 2015) has enabled lowering preparation time (~2h/96 pooled samples) and cost (~$3/pooled sample). A new C++ implementation for fast computing of population genetic statistics for Pool-seq, `g_renedalf`, (Czech and Exposito-Alonso, 2021a), based on the original Perl-based Popoolation software (Kofler et al., 2011a, 2011b), now allows for >100 speed in computation allowing analyses of thousands of pooled libraries. Although these methods are species-free and can be applied to any organism in the tree of life, we describe several proof-of-concept experiments using *Arabidopsis thaliana.* We showcase these methods' impact for studying rapid adaptation in plant evolutionary ecology studies, which are typically using individual-based methods such as common garden experiments and within-generation natural selection to different environments (Anderson and Wadgymar, 2019; Brachi et al., 2010; Exposito-Alonso et al., 2019; Fournier-Level et al., 2011; Lovell et al., 2021; Lowry et al., 2009; Monnahan et al., 2020).

To showcase the power of Pool-seq and E&R strategies to study population genetics and rapid evolution, we conduct a set of four experiments using from a few to several hundred natural accessions of *Arabidopsis thaliana*. Our main goal here is to provide evidence that our experimental setup can be used to obtain allele frequency data of equal quality compared to established

approaches, while being comparatively simple and inexpensive. In particular, we address the following challenges: (**Experiment 1**) We sequenced a mixture of seeds pooled from several hundreds of *A. thaliana* ecotypes from the 1001 Genomes Project (1001 Genomes Consortium, 2016), which can be used as a population founder for multiple evolution experiments. (**Experiment 2**). We constructed sequencing libraries of two wild type ecotypes, as the smallest population unit ?, to assess the variation in the resulting allele frequencies in the Pool-seq approach versus the conventional approach of pooling DNA extracts equally by concentration quantification. (**Experiment 3**) We conducted various poolings of genotypes and tissue types (i.e. leaf or flower) to understand the effect of individual pooling and coverage in allele frequency inferences. (**Experiment 4**) We trialed an "Evolve & Resequence common garden" experiment, to test our methods in real outdoor settings and test whether any signal of rapid evolution can be detected in a few generations.

Computational methods for large-scale Pool-seq datasets

To tackle the large amount of sequencing data produced in this project, we developed novel software tools, which will find further use in the ongoing GrENE-net project.

First, we implemented `g_renepipe` (Czech and Exposito-Alonso, 2021b), a pipeline based on Snakemake workflow management system (Köster and Rahmann, 2012; Mölder et al., 2021) to process raw sequence data into variant calls and allele frequencies. We used `g_renepipe` to process the data of all four experiments. Unless otherwise specified, we used the following tools in the pipeline: trimmomatic (Bolger et al., 2014) for read trimming, bwa mem (Li and Durbin, 2009) for mapping against the reference genome, samtools (Li et al., 2009) for working with bam and pileup files. We furthermore employed several quality control tools that are built into `g_renepipe` for ensuring that our sequence data is of sufficient quality (Andrews and Others, 2017; Ewels et al., 2016; Li et al., 2009; Okonechnikov et al., 2016) (for config file used in each analysis, see the online code repository https://github.com/lczech/grenepilot-paper).

The `g_renepipe` automatization of SNP and frequency calling allows us to test a number of variant filters and compare in a standardized fashion. Specifically, we focused quality controls related to: (1) free discovery of genetic variation *vs* utilizing only SNPs previously discovered in individual sequencing from the 1001 Genomes project (1001 Genomes Consortium, 2016), (2) coverage filters to reduce sampling noise (Lynch et al., 2014), (3) base quality filters and (4) mapping quality filters to reduce the likelihood of false positive mutations. These follow essentially the same rubric as the heuristic PoolSNP approach used in the Drosophila Evolution over Space and Time resource (Kapun et al., 2021, 2020). Although Pool-seq dedicated SNP callers exist based on likelihood and bayesian methods (Guirao-Rico and González, 2021), the advantage of these is clear to detect *bona fide* SNPs and reduce false positive SNPs. However, if a set of SNPs is already known to vary in the species, as is the case with the 1001 Genomes Project, the allele frequencies are well estimated as long as appropriate filters are implemented (Guirao-Rico and González, 2021) and there is sufficient coverage (Tilk et al., 2019). An extension to `g_renepipe` to improve allele frequency even at low coverage, given whole-genome information and linkage disequilibrium among founder populations is available, will be developed in the future using HARP and HAFpipe softwares (Kessner et al., 2013; Tilk et al., 2019).

Second, we developed the C++ based command line tool `g_renedalf` (Czech and Exposito-Alonso, 2021a) re-implementing and extending the methods of the state-of-the-art software PoPoolation (Kofler et al., 2011a, 2011b). For details on implemented equations and corrections, see the **Supplemental Mathematical Appendix**. In brief, population genetic metrics such as Waterson's θ, π, Tajima's D, and F_{ST}, account for Pool-seq sources of noise, including number of individuals pooled, coverage per base pair, and base call PHRED qualities. `g_renedalf` furthermore provides several auxiliary methods such as conversions between file formats (vcf to frequency table, vcf to sync file, etc.).

Experiment 1: Sequencing a seed mixture of 231 ecotypes to characterize a diversity panel

Rationale: In this experiment, we established a genetically diverse panel of *A. thaliana* natural accessions. We sequenced the seed mix of this panel to assess the ability of Pool-seq to correctly recover genome-wide allele frequencies of a large genome panel.

Setup: The founder seed mix for this experiment was sourced from the seeds of 231 ecotypes, 229 of which are part of the (1001 Genomes Consortium, 2016) and available from the Arabidopsis Biological Resource Center (ABRC) under accession CS78942 (https://abrc.osu.edu/stocks/465820). The remaining 2 ecotypes are available through the Israel Plant Gene Bank (https://igb.agri.gov.il/) under accession numbers 24208 and 22863 (see **Dataset S1**). Seeds were pooled at roughly equal proportions using an allometric seed number-to-weight relationship (weight in grams = b x seed number; where the coefficient ranges $b=1.369\times10^{-5}$—2.449×10^{-5}; for accessions 9965 and 7058, which are representative accessions of extreme seed weights).

Analysis: Eight tubes containing varying weights of the founder seed mix (**Table S1**) were germinated and homogenized using a FastPrep-24 (MP Biomedicals, Irvine, CA, USA). DNA extraction was done using a Qiagen DNeasy Plant Mini kit (Hilden, Germany). Each DNA extract was made into one TruSeq library. The eight TruSeq libraries were multiplexed and sequenced together on one lane of a HiSeq 3000 sequencer (Illumina, San Diego, California, USA). The total sequencing output was 9.54×10^{10} base pairs and the average genome-wide coverage was 500X. Raw sequence data were processed with our `g_renepipe` workflow (Czech and Exposito-Alonso, 2021b) to trim and map the reads against the reference genome (Berardini et al., 2015; Lamesch et al., 2012). We then calculated the minor allele frequencies (MAF) at each biallelic position, based on bam/pileup files counting the ratio of reads containing either reference or alternative alleles, using our `g_renedalf` tool (Aka. pileup-based frequency). Although likely unnecessary, because users of Pool-seq may also want to utilize variant callers typically used in individual sequencing, we also ran `g_renepipe` with three different variant callers: BCFtools (Li, 2011), freebayes (Garrison and Marth, 2012), and GATK HaplotypeCaller (McKenna et al., 2010). These tools are not primarily designed for calling variants from Pool-seq data, but the resulting VCF file of each caller can be turned into a frequency table by extracting the Allelic Depth ("AD") format field at each genome position for each sample, a process also implemented in `g_renedalf`. We then run GATK HaplotypeCaller and freebayes using the ploidy option equaling the pool size (`-ploidy 2470` and `--ploidy 2470 --pooled-discrete`, respectively. We also tried `--pooled-continuous` from freebayes. Note that each seed sample used for sequencing is estimated to have about 2,500 seeds based on

weight). These runs resulted in prohibitively long runtimes (GATK HaplotypeCaller) and memory usage (freebayes), making it impractical to be used on large datasets and large pool sizes. Furthermore, even if those analyses had succeeded, the resulting VCF files would contain 2,470 genotype calls for each sample at each position, leading to impractically large files. We hence ran the three callers with default ploidy of 2, to study their artifacts in Pool-seq applications, as we assume other researchers may resource in default settings.

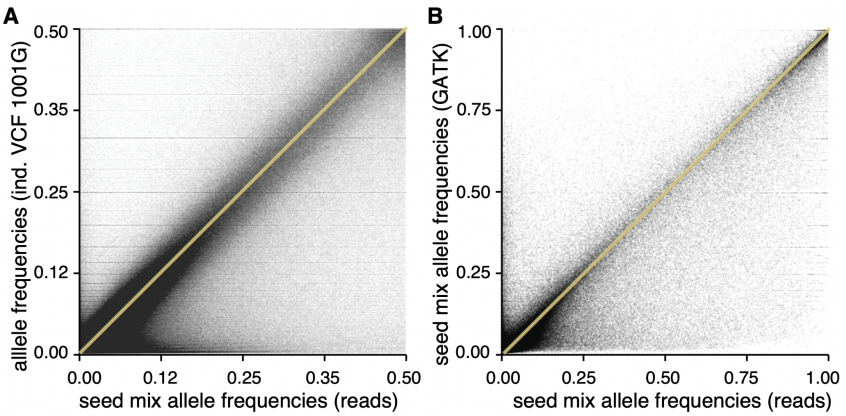

Fig. 1 | Direct sequencing of experimental founder seeds capture the 1001 Genomes variation
(A) (B)

Results We then compared this resulting allele frequencies from sequencing pools of seeds among different pipeline runs (**Fig. 1A, Fig SX**) and with the allele frequency based on the 1001 Genomes VCF file. This was conducted by subsetting to the 229 genotypes that overlap with those used for the seed mix and calculating the allele frequency *in silico* per allele across the genotypes (**Fig. 1B, Fig SX**). The frequency distribution of the sequenced seed correlates tightly with the frequencies estimated from the 1001 Genomes VCF (**Fig. S1A**), indicating that the seed mix correctly represents the genomic diversity of *A. thaliana*—the Pearson correlation coefficients between the frequencies are 0.9.x and 0.9x [Lucas to add], respectively.

The comparison of raw allele frequencies from different SNP callers revealed a trend to upwardly or downwardly bias allele frequencies, especially those found at very low frequency (<5%). We believe this effect is most dramatic in GATK HaplotypeCaller in its default diploid likelihood mode. GATK, which is tuned for human SNP calling, aims to call genetic variants that fit the reference homozygote, heterozygote, or alternative homozygote scheme, and utilizes local genome information to reject certain reads, likely leading to allele frequency noise (standard deviation from y=x line, SD=0.153) (**Fig. SX**). BCFtools, which is a relatively light-weight variant caller that is based on per-position data but also utilizes a diploid likelihood model has a higher mass around the y=x relationship (SD=0.097) but is still affected in upwardly biasing low frequency variants that may be inferred diploid heterozygous. Finally, Freebayes has the lowest noise around the y=x relationship (SD=0.625). Stringent coverage filters decrease allele frequency noise for all cases (**Fig. SX**).

Experiment 2: Two ecotypes to understand biases of DNA contribution to pooled samples and sequencing noise

Rationale One important assumption in population inferences based on Pool-seq data is that each

individual contributes an equal amount of sequencing reads. However, the deviation in DNA contribution by pooling organs from different individuals or entire individuals—as is common practice in *D. melanogaster* E&R experiments—has not been tested in plants such as *A. thaliana*. In this experiment, we sequenced two *A. thaliana* ecotypes (i.e. the smallest possible pool size of 2) to assess the variation in DNA contribution.

Setup To quantify the deviation of two flowers in their DNA content when pooled together, we sequenced three replicates of two flowers each. The first ecotype was the laboratory inbred strain Col-0, which was the type strain used to assemble the reference genome (Lamesch et al., 2012). The second, a natural accession (inbred in greenhouse propagations) from the 1001 Genomes project (1001 Genomes Consortium, 2016), was RUM-20 (#9925), which diverges from Col-0 by 1.00756 million base pairs according to the 1001 Genomes data. To compare with the ideal laboratory case where DNA is pooled at equal proportion to ensure lack of bias, we also extracted DNA from an individual leaf of Col-0 and RUM-20 and performed equal pooling by mass before library preparation (**Fig. 2A**, **Table S2X**). DNA was extracted with the CTAB method and processed into whole-genome sequencing libraries via a modified Nextera protocol (**Supplemental Appendix I: Library preparation**).

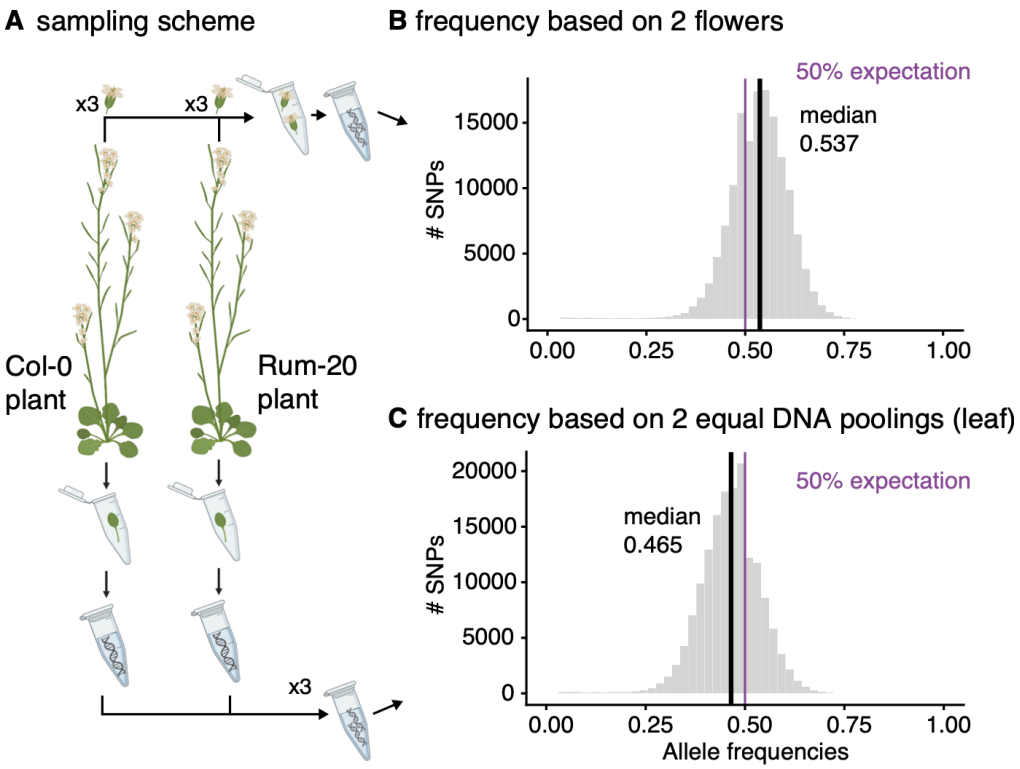

Fig. 2 | Experimental design (Exp. 2) to test the relative contribution to DNA sequencing output
(A) *Flower and leaf tissues are sampled from two ecotypes, Col-0 and RUM-20. Three replicates of two flowers were collected into the same tube (the Pool-seq method) while leaves were collected separately (the conventional method). Leaf DNA extracts were pooled equally by mass to create three replicates of DNA input for library preparation.* ***(B)*** *Distribution of allele frequencies in a replicate of 2-flower pool directly extracted and whole-genome sequenced* ***(C)*** *The equivalent in (B) for two separate DNA extracts carefully pooled at equal mass.*

Analysis: Assuming that both pooled individuals are homozygous and that both contributed equal amounts of DNA, the possible allele frequencies for two individuals (of which one is identical to the reference genome) are either 0% or 50%. Deviations from these hence indicate differences in DNA content. To test this, we again trimmed and mapped the reads with our `g_enepipe` workflow, and computed frequencies from bam files with `g_enedalf`, as described in Experiment 1.

Results: This proof-of-concept analysis provided a number of clues on the power and potential biases of Pool-seq to study population evolution in real time. Firstly, here we show that extraction and Pool-Seq of groups of flowers is not only more cost effective than separate extractions and later pooling, but it also produces the same accuracy in allele frequencies (**Fig. 2B-C**). Secondly, subsetting the dataset to SNPs passing several filters including stringent mapping quality (-q 60 option) and matching of forward/reverse read mapping (-f option), base quality (-Q 30 option), and very low alternative allele count shows the expected (MAC>2), enable us to retrieve allele frequency distributions expected under a regular binomial sampling of alleles (**Fig. S16**).

Experiment 3: Combinatorial experiments of pool sizes and tissue type sequencing to determine optimal sampling schemes

Rationale In this experiment, we evaluated the ability of Pool-seq to recover correct allele frequencies from pooled samples made up of 5 to100 flowers and leaves sampled from *A. thaliana* plants. We studied whether (A) individual leaf DNA extraction and library preparation with equal DNA input and (B) pooled flower DNA extraction and direct library preparation produce comparable population estimates. Previous experiments of direct Pool-sequencing of whole *Drosophila melanogaster* flies indicate that the frequency estimation per population requires 100 individuals at 50x coverage for perfect allele frequency retrieval (Gautier et al., 2013). Therefore, we also tested up to 100 individuals per pool as the maximum sample number.

Setup We grew 231 genotypes (**Dataset S1**) in 2,500 pots with single plants (replicating similar conditions of large evolving populations outdoors) and sampled 50 random plants for our test. We potted 2,500 plants to ensure future experiments would suffer the least due to genetic drift and bottlenecks induced by laboratory propagation. Flowers were sampled from 50 plants to produce 5 combinations of different numbers of flowers: 5, 10, 25, 50, 100 (**Fig. 2**). We subsampled the 50 pots to produce pools of 5, 10, and 25 different plants. For the 100 flower combination, two flowers of each of the 50 pots were sampled to simulate sampling of 100 individuals in the GrENE-net outdoor experiments. For the same plants for which flowers were collected, also one leaf per plant was collected and stored separately.

Grinding, extraction and library preparation steps are described in the **Supplemental Appendix I: DNA extraction** and **Library preparation**. The leaf DNA pooling was done for the same individuals for which flower subsamples of 5, 10, 25, and 50 individuals were taken before (**Table S4X**). Therefore, we expect the allele frequencies of the equimolar pool of leaf DNA and the frequencies of flower extracts to be identical, unless our method has systematic biases.

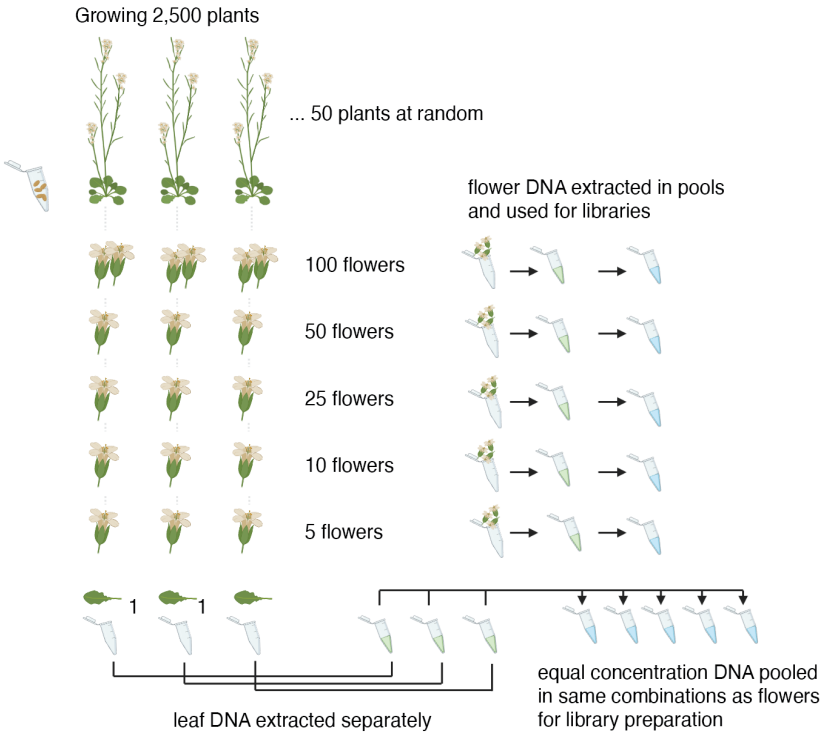

Fig. 3 | Experimental design (Exp. 3) to test Pool-Seq with plant flowers
A total of 2,500 pots were planted with diverse ecotypes of A. thaliana *(**Table S2**) in order to not create a bottleneck. 50 plants were sampled at random avoiding applying any natural selection. Different combinations of these 50 plants were used (**Table S4**). Created with BioRender.com*

Results The chosen metric to study the accuracy of frequency recovery of a large population was Pearson's correlation coefficient between genome-wide allele frequencies (alternative allele count / coverage) from the founder seed sequencing and either flower or plant sequences of the 50 or less randomly chosen individuals. The seed sequencing ideally represents the true frequency of the founder seed mix, as potential DNA content differences across seeds are averages over hundreds of thousands and over different extraction and library preparations (any seed sequencing pair showed Pearson's $r > 0.95$). Because accuracy of allele frequency retrieval may vary along the genome due to random coverage variation or differential mapping quality, we calculated correlation coefficients across coverages. The frequency retrieval is highly accurate ($r > 0.95$) as long as more than 50 individuals are sampled and the sequencing coverage is greater than 25X, regardless of the tissue type or the pooling method (**Fig. 3**). Therefore, together with Experiment 2 we have established that sequencing pools of flowers is comparably accurate as equimolar sequencing pools of leaf DNA extracts.

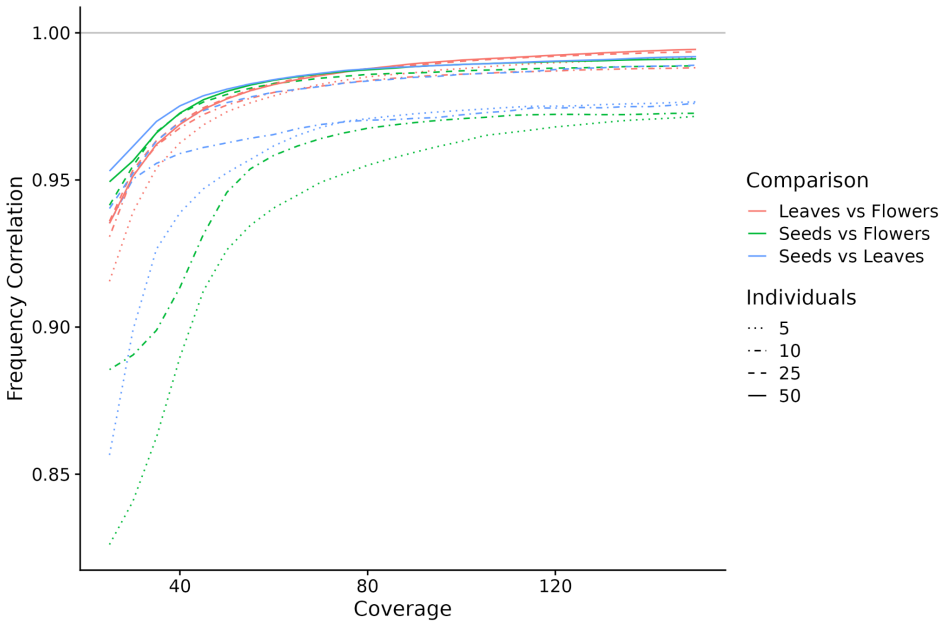

Fig. 4 | Correlation between allele frequencies estimated from direct Pool-Sequencing of flowers* vs *leaves pooled at equal DNA concentration.
Genome-Wide allele frequencies of seeds from Experiment 1 were compared to frequencies of direct pools of flowers and equimolar pools of DNA from leaves (Fig. 3). Variant positions were binned by coverage (x-axis) and Pearson correlation coefficient was computed for a given comparison (y-axis). Comparisons are also split by the number of individuals used in the pools.

Experiment 4: Multi-year outdoor field experiment to showcase the power to track rapid evolution

Rationale Ultimately, the cost-effective and scalable Pool-Seq approach is aimed at tracking evolution of populations in time. To showcase its power, we conducted an outdoor experiment over two growth seasons starting from a rich population of *A. thaliana* ecotypes.

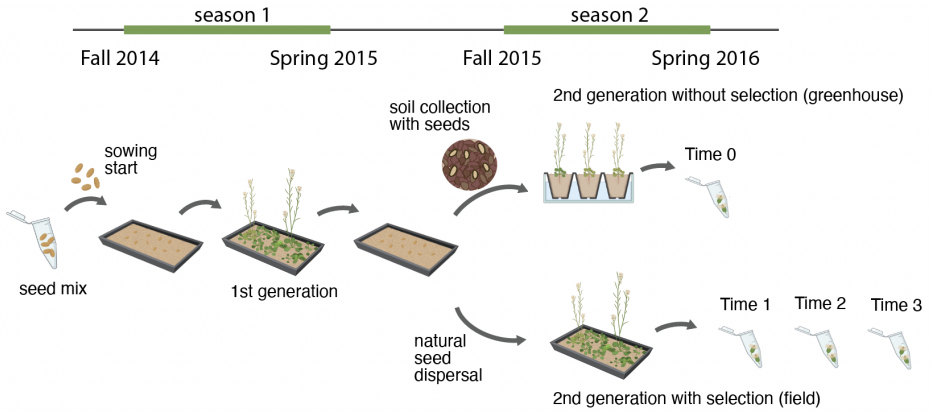

Fig. 5 | Design of the outdoor field experiment (Exp. 4)
*(**A**) Setup of 3 population replicates. Example germination of seedlings (**B**) and flowering plants (**C**). (**D**) Example sampling of flowers for pool-sequencing. This timeline was replicated 3 times in parallel.*

Setup. This experiment was performed on an outdoor experimental field at the Max Planck Institute of Developmental Biology campus (48.537723, 9.058746, Tübingen, Germany) (**Fig. 4**). A seed mix of 451 natural accessions—a larger set of the 1001 Genomes enabled by higher seed availability than in Experiment 1(**Dataset S1,** under same accession CS78942 in ABRC stocks, https://abrc.osu.edu/stocks/465820)—was sowed in three plots in Fall 2014. After one complete generation in the wild with natural fruit shattering and seed dispersal during late Spring, soil samples were collected (Francois, when was this done? before germination?) and transferred to an indoors growth chamber to avoid any natural selection involved in germination or survival. Between 50-110 adult plants were sampled as a baseline (Time 0 in **Table S5 and Fig. 5**). In Spring of the second generation, between 50-110 surviving and reproductive adults were sampled for sequencing at 3 different time points to capture the whole breadth of flowering: 80-200, 60-300, and 10-100 flowers depending on the abundance of adults (Times 1-3 in **Table S5 and Fig. 5**). The flowers collected at Time 1, 2, and 3 were sequenced separately and analyzed in different combinations. A total of 1,398 individuals were sequenced in these pooled samples. We used Pool-seq adjusted population genetic statistics from `g_enedalf` to study genome-wide patterns of F_{ST} across all combinations of replicates and timepoints accounting for pool size (**Table S5**) and genome-wide variation in coverage.

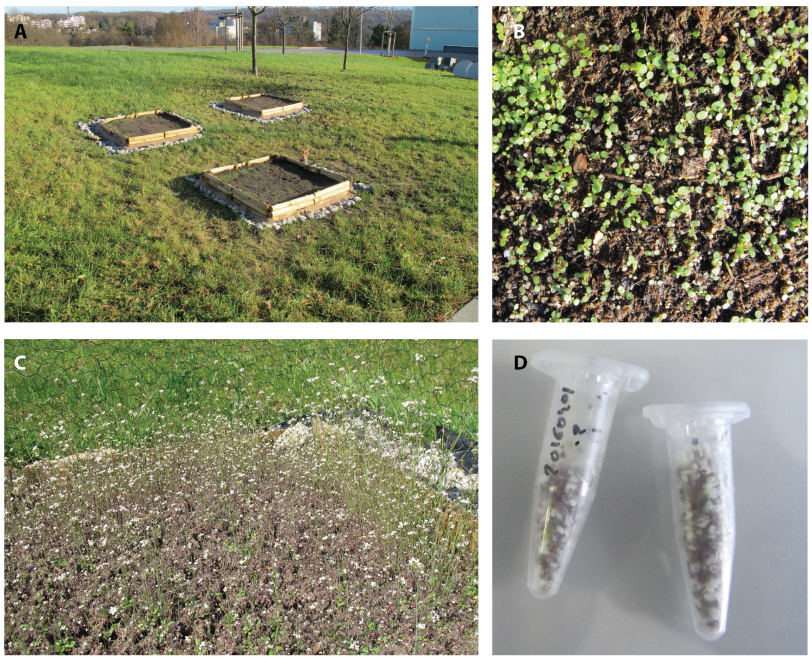

Fig. 6 | Pictures of outdoor field experiment (Exp. 4)
*(**A**) Setup of 3 population replicates. (**B**) Example germination of seedlings. (**C**) Abundant flowering patch. (**D**) Example sampling of flowers for pool-sequencing. This timeline was replicated 3 times in parallel.*

Results. Plants successfully established in dense patches in the experiment (**Fig. 6**). In the order of tens of thousands of seedlings were observed per plot replicate, theoretically enabling efficient natural selection in the order of selection coefficients of 0.1% fitness effects ($N_e s > 1$) (Charlesworth and Charlesworth, 2010). The observed genomic differentiation based on F_{ST} averaged in 10Kb windows between replicates and timepoints in the field (**Fig. S12**, Exp. 4) was much larger than that

of multiple sequencing runs of large seed replicates (Exp. 1), as expected under the absence of selection and large pool sizes of seeds. The latter can be used as a baseline for noise in this E&R experiment due to limited coverage and individual pool size, and other experimental factors. Scanning the genome for F_{ST} peaks between Time 0 (no selection) with Time 1-3 (with natural selection, sampled in three different times in Spring) revealed locations more differentiated than the background (**Fig. 7, Fig. S 2**).

One of such peaks was localized in chromosome 5 near the region of the gene *Flowering Locus C* (*FLC*, AT5G10140), MADS-box transcription factor and master regulator of flowering time. The region with elevated F_{ST} is before the transcription start site of *FLC*, leading us to speculate that variation in the promoter region could be under natural selection in these experiments. These signals were repeated in the three comparisons between time points 1 and 3 across replicates (**Fig. 5B**). This Experiment 4 was conducted in parallel to a common garden experiment 1.51 km away (48.545809, 9.042449) with similarly rich and highly-overlapping *A. thaliana* ecotype sets (Exposito-Alonso et al., 2019). The latter directly measured survival of each ecotype in pot and estimated the number of seeds produced if the plant reached adult reproductive stage. Exposito-Alonso et al. conducted Genome-Wide Associations to identify genetic variants that explained variation in fitness, which is one of the most direct ways to quantify natural selection in an environment. Ultimately, we expect that genetically-based fitness differences among ecotypes would cause genotype and allele frequencies to change over time. As expected, we found an overlap between the identified F_{ST} peak and an elevated number of SNPs associated with seed set with *P<0.0004* (**Fig. 7**) (thp condition, (Exposito-Alonso et al., 2019)). This overlap indicates flowering time may have been under selection in Exp. 4. The fact that flowering time was correlated at the individual level with relative seed production (Spearman's rho= -0.404, S = 31048965, $P < 2.2 \times 10^{-16}$) and survival (rho= -0.187, S = 26399658, $P = 2.074 \times 10^{-5}$) in the common garden experiments further evidences our finding of selection over the FLC locus in the the multi-year E&R field experiment.

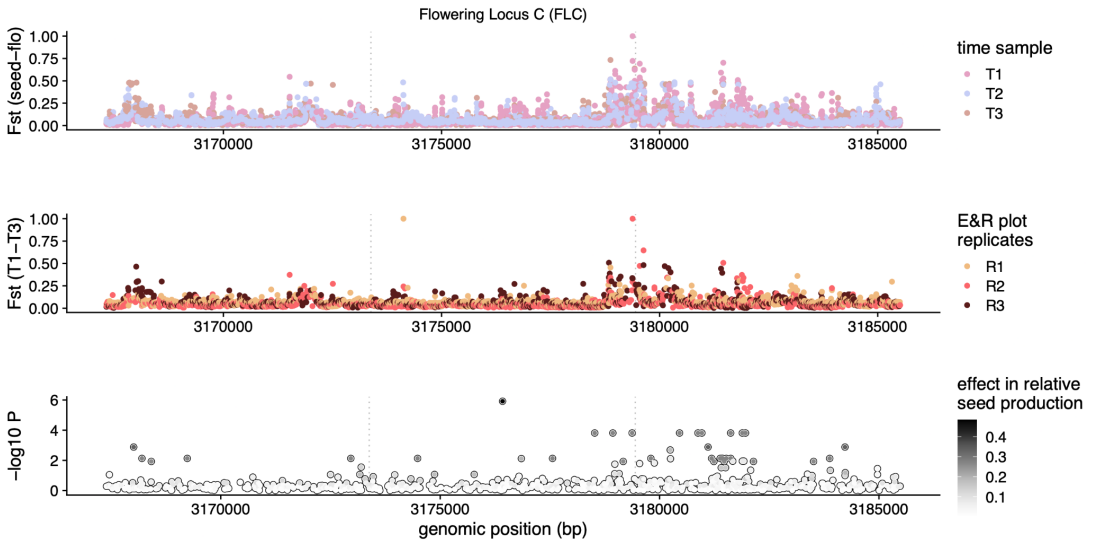

Fig . 7 | Temporal allele frequency change in a multi-year Evolve & Resequence experiment compared to fitness effects in a common garden experiment.
*(**A-B**) Temporal allele frequency differentiation (Fst) in the Flowering Locus C region on chromosome 5 showing peaks of differentiation around the first exon and the upstream promoter region of the gene (positions around 3,180,000, note the protein coding strand is the reverse). (**A**) Differentiation between the baseline "without selection" and the flower samples of surviving adults in nature. (**B**) and between the first and last sampling in the flowering season. (**C**) Genome-Wide Association between genetic variants in the 1001 Genomes and outdoor seed production in a common garden experiment (Exposito-Alonso et al., 2019) 1.51 kilometers away from the Evolve & Resequence experiment in (A-B).*

Discussion and outlook

The paradigm that evolution is a slow process is being challenged by accumulating evidence in animal and plant species that genotype frequencies within populations fluctuate or change in the order of seasons or within decades following environmental changes (Bergland et al., 2014; Franks and Weis, 2008). Scalable whole-genome approaches based on Pool-Seq (Schlötterer et al., 2014) have enabled population genomic resources across continental scales such as the "Drosophila Evolution over Space and Time" (DEST) resource (Kapun et al., 2021) (https://dest.bio) or Drosophila Evolve & Resequence approaches across generations (Rudman et al., 2021). However, projects of such scale to study plant evolution over time do not yet exist. Here we present Pool-seq laboratory protocols and new efficient software implementations scalable to thousands of experimental or natural populations of plants and envision the future of a globally distributed experiment with *Arabidopsis thaliana*. Here, we document new advances in whole-genome Pool-Seq protocols and computational methods to track genetic evolution of plants in real time at low cost.

First, we show that directly whole-genome sequence of a mixture of seeds can properly characterize the standing genetic variation of a hypothetical starting pool of founder individuals for an E&R experiment and options of quality parameters in the SNP calling as well as a streamlined frequency calling pipeline (Czech and Exposito-Alonso, 2021b). Plant evolutionary ecologists using annual plants as model systems may find the consistency of size of flowers even in small plants, may enable tracking of genotypes of successfully reproducing adults more cost-effectively and precisely than by separate leaf extracts or combination of leaf punches, used so far in a number of projects (Fracassetti et al., 2015; Roda et al., 2017), since leaves vary in many Brassicacea species or other families more than an order of magnitude in size while flower development is relatively constant among dramatically different plants. Protocols where sampling strategies are clear and objective, such as the sampling of a flower per plant, would potentially enable more citizen-science Pool-Seq projects.

Second, we have updated the now-classic 2011's PoPoolation software (Kofler et al., 2011a) with a C++ implementation from scratch. If we are, for instance, to generate in the order of ~2,000 whole-genome samples in *A. thaliana* (~5 Tbp) and conduct pairwise F_{ST} among samples, we would require ~100 days in a X computer cores of of given characteristics, Lucas fill in here whereas the new implementation `grenedalf` would run in less than a day (Czech and Exposito-Alonso, 2021a). This tool also has a number of utilities such as <Lucas fill in here> which become cumbersome with customized scripts.

Third, to showcase the approach could work in an E&R experiment using the 1001 *Arabidopsis* Genomes, we conduct outdoor field experiments. The fact that linkage decays surprisingly fast in *A. thaliana* (**Fig. S9**) (Kim et al., 2007), perhaps owed to that a ~2-16% recombination is sufficient to shuffle standing genetic variation (Bomblies et al., 2010; Platt et al., 2010), may enable mapping of certain adaptive loci at relative narrowly (**Fig. 7**), perhaps even at higher mapping than Genome-Wide Associations (Atwell et al., 2010). The success of Experiment 4 motivates the use of this approach for larger efforts, and seems to provide a sensitivity that may be even higher than highly expensive and costly experiments that can only be done in few environments (Agren and Schemske, 2012; Exposito-Alonso et al., 2019, 2018; Fournier-Level et al., 2011; Manzano-Piedras et al., 2014).

Despite the suggestive association between fitness effect sizes and allele frequency changes, and the parallel change in frequency in the three replicates, further evidence is needed to confirm natural selection is behind these genome-wide allele frequency changes. We have, however, detected consistent changes in allele frequency, and we are able to accurately retrieve such frequencies at a comparable quality compared to other state-of-the-art methods using common gardens [Cite Exposito]. Differently from common garden experiments, E&R experiments could be conducted in tens or hundreds of locations at relatively low cost. To improve this, we have began a project called "Genomics of rapid Evolution to Novel Environments" network. The experiment here can be considered as a pilot study for the internationally distributed E&R GrENE-net project which involves 45 field sites (https://grenenet.org), whose dataset creation is ongoing.

The design of the GrENE-net experiment is more powerful than Experiment 4 presented here in that we have kept track of these populations since 2017 until 2021/2022—providing a better temporal tracking—including the original founder seed mix (part of Experiment 1 here). Climate x genotype x fitness effects have been so far limited by the number of environments (2-10) [Cite, Lowry]. In total, participants from 45 locations joined GrENE-net, with successfully reproducing populations in 32 locations. This will enable us to compare magnitude of allele frequency changes with intensity of climate stressors, drivers of these frequency dynamics and locate potential causal loci involved in adaptation. Further, at each location, 12 population replicates similar to Experiment 4 were created using each 0.1 g of the founder seed mix (0.1 g ~ about 5,000 seeds) to ensure enough replication. This approach is much more powerful than the Experiment 4 presented here, where even if we identified parallel F_{ST} trends in three replicates, there was only one temporal comparison rather than yearly. Not only that, but the GrENE-net experiments have multiple generations of outdoor samples, providing a better temporal tracking of allele trajectories as well as replication across climate gradients, enabling us to compare magnitude of allele frequency changes with intensity of climate stressors. The GrENE-net experiment will not only enable researchers to make a direct link between environment and natural selection at the allele frequency level, but will certainly spark theoretical development in evolutionary genetics as well as empower molecular biologists searching mechanisms of adaptation. If biologists are to anticipate biotic changes to future extreme climate conditions, long-term highly-spatially-replicated Evolve & Resequence datasets are paramount.

Additional Information

Data and Code availability Reads were deposited at ENA with accession number: TBD.

Genomes of founder populations are available as part of the 1001 Arabidopsis Genomes project: http://1001genomes.org/data/GMI-MPI/releases/v3.1/. Scripts of the analyses in this manuscript are available at https://github.com/lczech/grenepilot-paper, which contains all settings used for the runs of our `grenepipe` (Czech and Exposito-Alonso, 2021b) workflow for variant calling, as well as all python and R scripts for the figures presented here. Genome frequency manipulations and Pool-Seq corrected population genetic statistics are implemented in `grenedalf` (Czech and Exposito-Alonso, 2021a).

Acknowledgements We thank Robert Kofler for discussing the PoPoolation implementation and the members of the Moi Lab for feedback on the manuscript and analyses.

Funding statement This work was funded by an ERC Advanced Grant IMMUNEMESIS and the Max Planck Society (D.W.) and by the Carnegie Institution for Science and National Institutes of Health's Early Investigator Award (1DP5OD029506-01) (M.E.-A.). The computing for this project was performed on the Calc and Memex clusters from the Carnegie Institution for Science.

Disclosure statement The authors declare no competing financial interests. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contribution MEA, FV, NJS, conceived the project. MEA and DW acquired financial support for the project leading to this publication, provided study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools. MEA managed and coordinated the research activity planning and execution. LC, MEA, YP, TB, conducted statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data. LC, MEA conducted programming, software development, and tested code components. MEA, RP, PL, FV, JH, KF, BR, RS, conducted research, performing experiments or data collection. LC, YP, TB, wrote the first manuscript draft and all authors edited and reviewed the latest manuscript version.

References

1001 Genomes Consortium. 2016. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell* **166**:481–491. doi: 10.1016/j.cell.2016.05.063

Agren J, Schemske DW. 2012. Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. *New Phytol* **194**:1112–1122. doi:10.1111/j.1469-8137.2012.04112.x

Anderson JT, Wadgymar SM. 2019. Climate change disrupts local adaptation and favours upslope migration. *Ecol Lett*. doi: 10.1111/ele.13427

Andrews S, Others. 2017. FastQC: a quality control tool for high throughput sequence data. 2010.

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JDG, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M. 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **465**:627–631. doi: 10.1038/nature08800

Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* **10**:e0128036. doi:10.1371/journal.pone.0128036

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E. 2015. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* **53**:474–485. doi: 10.1002/dvg.22877

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLoS Genet* **10**:e1004775. doi: 10.1371/journal.pgen.1004775

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. doi: 10.1093/bioinformatics/btu170

Bomblies K, Yant L, Laitinen R a., Kim S-T, Hollister JD, Warthmann N, Fitz J, Weigel D. 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of Arabidopsis thaliana. *PLoS Genet* **6**:e1000890–e1000890. doi:10.1371/journal.pgen.1000890

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F. 2010. Linkage and association mapping of Arabidopsis thaliana flowering time in nature. *PLoS Genet* **6**:e1000940. doi: 10.1371/journal.pgen.1000940

Charlesworth B, Charlesworth D. 2010. Elements of Evolutionary Genetics. W. H. Freeman.

Clausen J, Keck DD, Hiesey WM. 1941. Regional Differentiation in Plant Species. *Am Nat* **75**:231–250.

Czech L, Exposito-Alonso M. 2021a. grenedalf: population genetic statistics to study rapid evolution with Pool-Seq. *in prep*.

Czech L, Exposito-Alonso M. 2021b. grenepipe: A flexible, scalable, and reproducible pipeline to automate variant and frequency calling from sequence reads. *arXiv*.

Endler L, Betancourt AJ, Nolte V, Schlötterer C. 2016. Reconciling Differences in Pool-GWAS Between Populations: A Case Study of Female Abdominal Pigmentation in Drosophila melanogaster. *Genetics* **202**:843–855. doi: 10.1534/genetics.115.183376

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**:3047–3048. doi:10.1093/bioinformatics/btw354

Exposito-Alonso M, 500 Genomes Field Experiment Team, Burbano HA, Bossdorf O, Nielsen R, Weigel D. 2019. Natural selection in the *Arabidopsis thaliana* genome in present and future climates. *Nature* **573**:126–129. doi: 10.1038/s41586-019-1520-9

Exposito-Alonso M, Brennan AC, Alonso-Blanco C, Picó FX. 2018. Spatio-temporal variation in fitness responses to contrasting environments in Arabidopsis thaliana. *Evolution*. doi:10.1111/evo.13508

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in Arabidopsis thaliana. *Science* **334**:86–89. doi:10.1126/science.1209271

Fracassetti M, Griffin PC, Willi Y. 2015. Validation of Pooled Whole-Genome Re-Sequencing in Arabidopsis lyrata. *PLoS One* **10**:e0140462. doi: 10.1371/journal.pone.0140462

Franks SJ, Weis AE. 2008. A change in climate causes rapid evolution of multiple life-history traits and their interactions in an annual plant. *J Evol Biol* **21**:1321–1334. doi:10.1111/j.1420-9101.2008.01566.x

Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* **186**:207–218. doi:10.1534/genetics.110.114397

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv*

[q-bioGN].

Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P, Kerdelhué C, Estoup A. 2013. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol Ecol* **22**:3766–3779. doi: 10.1111/mec.12360

Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. *Nature*. doi: 10.1038/nature24287

Grant PR, Grant BR. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**:707–711. doi: 10.1126/science.1070315

Guirao-Rico S, González J. 2021. Benchmarking the performance of Pool-seq SNP callers using simulated and real sequencing data. *Mol Ecol Resour* **21**:1216–1229. doi:10.1111/1755-0998.13343

Günther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. *Genetics* **195**:205–220. doi: 10.1534/genetics.113.152462

Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, Utermann G, Pritchard JK, Coop G, Di Rienzo A. 2011. Adaptations to climate-mediated selective pressures in humans. *PLoS Genet* **7**:e1001375. doi: 10.1371/journal.pgen.1001375

Kapun M, Barrón MG, Staubach F, Obbard DJ, Wiberg RAW, Vieira J, Goubert C, Rota-Stabelli O, Kankare M, Bogaerts-Márquez M, Haudry A, Waidele L, Kozeretska I, Pasyukova EG, Loeschcke V, Pascual M, Vieira CP, Serga S, Montchamp-Moreau C, Abbott J, Gibert P, Porcelli D, Posnien N, Sánchez-Gracia A, Grath S, Sucena É, Bergland AO, Guerreiro MPG, Onder BS, Argyridou E, Guio L, Schou MF, Deplancke B, Vieira C, Ritchie MG, Zwaan BJ, Tauber E, Orengo DJ, Puerma E, Aguadé M, Schmidt P, Parsch J, Betancourt AJ, Flatt T, González J. 2020. Genomic Analysis of European Drosophila melanogaster Populations Reveals Longitudinal Structure, Continent-Wide Selection, and Previously Unknown DNA Viruses. *Mol Biol Evol* **37**:2661–2678. doi:10.1093/molbev/msaa120

Kapun M, Nunez JCB, Bogaerts-Márquez M, Murga-Moreno J, Paris M, Outten J, Coronado-Zamora M, Tern C, Rota-Stabelli O, García Guerreiro MP, Casillas S, Orengo DJ, Puerma E, Kankare M, Ometto L, Loeschcke V, Onder BS, Abbott JK, Schaeffer SW, Rajpurohit S, Behrman EL, Schou MF, Merritt TJS, Lazzaro BP, Glaser-Schmitt A, Argyridou E, Staubach F, Wang Y, Tauber E, Serga SV, Fabian DK, Dyer KA, Wheat CW, Parsch J, Grath S, Veselinovic MS, Stamenkovic-Radak M, Jelic M, Buendía-Ruíz AJ, Josefa Gómez-Julián M, Luisa Espinosa-Jimenez M, Gallardo-Jiménez FD, Patenkovic A, Eric K, Tanaskovic M, Ullastres A, Guio L, Merenciano M, Guirao-Rico S, Horváth V, Obbard DJ, Pasyukova E, Alatortsev VE, Vieira CP, Vieira J, Roberto Torres J, Kozeretska I, Maistrenko OM, Montchamp-Moreau C, Mukha DV, Machado HE, Barbadilla A, Petrov D, Schmidt P, Gonzalez J, Flatt T, Bergland AO. 2021. Drosophila Evolution over Space and Time (DEST) - A New Population Genomics Resource. *bioRxiv*. doi:10.1101/2021.02.01.428994

Kessner D, Turner TL, Novembre J. 2013. Maximum Likelihood Estimation of Frequencies of Known Haplotypes from Pooled Sequence Data. *Molecular Biology and Evolution*. doi:10.1093/molbev/mst016

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M. 2007. Recombination and linkage disequilibrium in Arabidopsis thaliana. *Nat Genet* **39**:1151–1155. doi: 10.1038/ng2115

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P. 2001. The strength of phenotypic selection in natural populations. *Am Nat* **157**:245–261. doi:10.1086/319193

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C.

2011a. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* **6**:e15925. doi: 10.1371/journal.pone.0015925

Kofler R, Pandey RV, Schlötterer C. 2011b. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**:3435–3436. doi:10.1093/bioinformatics/btr589

Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**:2520–2522. doi: 10.1093/bioinformatics/bts480

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* **40**:D1202–10. doi: 10.1093/nar/gkr1090

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–2993. doi:10.1093/bioinformatics/btr509

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**:1754–1760. doi: 10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi: 10.1093/bioinformatics/btp352

Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, Sreedasyam A, Healey A, Session A, Shu S, Barry K, Bonos S, Boston L, Daum C, Deshpande S, Ewing A, Grabowski PP, Haque T, Harrison M, Jiang J, Kudrna D, Lipzen A, Pendergast TH 4th, Plott C, Qi P, Saski CA, Shakirov EV, Sims D, Sharma M, Sharma R, Stewart A, Singan VR, Tang Y, Thibivillier S, Webber J, Weng X, Williams M, Wu GA, Yoshinaga Y, Zane M, Zhang L, Zhang J, Behrman KD, Boe AR, Fay PA, Fritschi FB, Jastrow JD, Lloyd-Reilley J, Martínez-Reyna JM, Matamala R, Mitchell RB, Rouquette FM Jr, Ronald P, Saha M, Tobias CM, Udvardi M, Wing RA, Wu Y, Bartley LE, Casler M, Devos KM, Lowry DB, Rokhsar DS, Grimwood J, Juenger TE, Schmutz J. 2021. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature*. doi:10.1038/s41586-020-03127-1

Lowry DB, Hall MC, Salt DE, Willis JH. 2009. Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland Mimulus guttatus. *New Phytol* **183**:776–788. doi:10.1111/j.1469-8137.2009.02901.x

Lynch M, Bost D, Wilson S, Maruki T, Harrison S. 2014. Population-genetic inference from pooled-sequencing data. *Genome Biol Evol* **6**:1210–1218. doi:10.1093/gbe/evu085

Manzano-Piedras E, Marcer A, Alonso-Blanco C, Picó FX. 2014. Deciphering the adjustment between environment and life history in annuals: lessons from a geographically-explicit approach in Arabidopsis thaliana. *PLoS One* **9**:e87836. doi: 10.1371/journal.pone.0087836

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**:1297–1303. doi:10.1101/gr.107524.110

Merilä J, Sheldon BC, Kruuk LE. 2001. Explaining stasis: microevolutionary studies in natural populations. *Genetica* **112-113**:199–222.

Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, Others. 2021. Sustainable data analysis with Snakemake. F1000Res 10: 33.

Monnahan PJ, Colicchio J, Fishman L, Macdonald SJ, Kelly JK. 2020. Predicting evolutionary change at

the DNA level in a natural Mimulus population. *bioRxiv*. doi: 10.1101/2020.06.23.166736

Nosil P, Villoutreix R, de Carvalho CF, Farkas TE, Soria-Carrasco V, Feder JL, Crespi BJ, Gompert Z. 2018. Natural selection and the predictability of evolution in Timema stick insects. *Science* **359**:765–770. doi: 10.1126/science.aap9125

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**:292–294. doi:10.1093/bioinformatics/btv566

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Agren J, Bossdorf O, Byers D, Donohue K, Dunning M, Holub EB, Hudson A, Le Corre V, Loudet O, Roux F, Warthmann N, Weigel D, Rivero L, Scholl R, Nordborg M, Bergelson J, Borevitz JO. 2010. The scale of population structure in Arabidopsis thaliana. *PLoS Genet* **6**:e1000843. doi:10.1371/journal.pgen.1000843

Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* **20**:R208–15. doi: 10.1016/j.cub.2009.11.055

Roda F, Walter GM, Nipper R, Ortiz-Barrientos D. 2017. Genomic clustering of adaptive loci during parallel evolution of an Australian wildflower. *Mol Ecol* **26**:3687–3699. doi: 10.1111/mec.14150

Rudman SM, Greenblum SI, Rajpurohit S, Betancourt NJ, Hanna J, Tilk S, Yokoyama T, Petrov DA, Schmidt P. 2021. Direct observation of adaptive tracking on ecological timescales in Drosophila. *bioRxiv*. doi: 10.1101/2021.04.27.441526

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nat Rev Genet* **14**:807–820. doi: 10.1038/nrg3522

Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals -- mining genome-wide polymorphism data without big funding. *Nat Rev Genet* **15**:749–763. doi:10.1038/nrg3803

Tilk S, Bergland A, Goodman A, Schmidt P, Petrov D, Greenblum S. 2019. Accurate Allele Frequencies from Ultra-low Coverage Pool-Seq Samples in Evolve-and-Resequence Experiments. *G3* **9**:4159–4168. doi: 10.1534/g3.119.400755

Genomics of **r**apid **E**volution in **N**ovel **E**nvironments

A Coordinated Distributed Evolution Experiment with Arabidopsis thaliana

Supplemental Information Guide

Monitoring adaptation and demography of plant experimental populations with Pool-Sequencing

Supplemental Materials & Methods: Extended DNA preparation, sequencing methods, and computational analyses.

Google drive link

Supplemental Mathematical Appendix: Population genetic equations adapted for Pool-Seq.

Google drive link

GrENE-net.org consortia authors

TBA