bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The majority of somatic mutations in fruit trees are layer-specific

Manish Goel[1,2], José A. Campoy[2,3], Kristin Krause[2,#], Lisa C. Baus[1,2], Anshupa Sahu[4,5], Hequan Sun[2,#], Birgit Walkemeier[2], Magdalena Marek[6], Randy Beaudry[7], David Ruiz[8], Bruno Huettel[6] and Korbinian Schneeberger[1,2,9,*]

[1] Faculty of Biology, LMU Munich, Planegg-Martinsried, Germany

[2] Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany

[3] Department of Agronomical Engineering, Institute of Plant Biotechnology, Universidad Politécnica de Cartagena, Cartagena, Spain

[4] Institute for Medical Biometry, Informatics and Epidemiology, University Hospital Bonn, Bonn, Germany

[5] Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Bonn, Germany

[6] Max Planck-Genome Centre Cologne, Cologne, Germany

[7] Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA

[8] Department of Plant Breeding, CEBAS-CSIC, P.O. Box 164, E-30100, Espinardo, Murcia, Spain

[9] CEPLAS (Cluster of Excellence on Plant Sciences), Heinrich-Heine University, Düsseldorf, Germany.

[#] present addresses: Illumina Solutions Center Berlin, Berlin, Germany (KK), Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, China (HS)

[*] Correspondence to Korbinian Schneeberger (k.schneeberger@lmu.de)

One sentence summary: Somatic mutations in fruit trees occur in specific cell layers.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

Plant meristems consist of three distinct stem cells layers. The identities of these layers are maintained in developing organs. Somatic mutations in the meristem thus propagate only into the respective layers of the differentiated organs. We analysed the genomes of individual cell layers in an apricot tree and found an unexpectedly high mutation load, where >90% of the somatic mutations were layer-specific. The layers had similar mutation spectra but layer 1 (epidermis) had substantially more mutations than layer 2 (mesocarp). Most somatic mutations were shared between fruits and neighbouring leaves corroborating their meristematic origin. Single-cell transcriptomics revealed their layer-specific effects forming the basis for layer-specific phenotypes. Our work unveils the hidden abundance of layer-specific meristematic mutations and provides insights into their identification and understanding.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

Plant meristems are specialized tissues of undifferentiated stem cells. They facilitate growth by continuously dividing and producing new cells that undergo differentiation leading to the formation of new branches and organs. Meristems are structured in different zones known as tunica and corpus (*1–3*). Typically, the tunica consists of two peripheral layers (L1 and L2) while the corpus consists of one interior layer (L3) (*4*, *5*). During the development of new tissues, the identity of the different layers is mostly conserved implying that newly formed tissue is also formed of three, mostly distinct layers (*6*, *7*). For example, in leaves, the epidermis is formed of cells that developed from L1, mesophyll cells are derived from L2, and cells of the vascular tissue from L3.

While somatic mutations in differentiated tissue will only affect small parts of a tree, somatic mutations in stem cells can be propagated to large areas of the plants leading to genomic mosaicism between entire branches and organs (*3*, *8*). Such meristematic mutations can induce bud sport mutants, where significant parts of the plant appear or behave differently as compared to the rest of the tree including traits with agronomical value. Consequently, bud sports (and the underlying somatic mutations) are frequently used to improve crop species, including fruit trees, where conventional breeding based on the introgression of additional traits is slow and tedious (*9–13*).

Multiple studies have analysed genomic mosaicism within individual plants and discussed various aspects of somatic mutations like mutation rates, genome-wide distribution, and propagation through the branches (*14–19*). However, most analyses were limited to bulked samples where cells from all layers were sequenced and analysed together. This can miss certain layer-specific mutations and limits our understanding of the prevalence and spectra of somatic mutations in specific cell layers (*13*).

Here, we describe the genomic heterogeneity between cell layers caused by somatic mutations in the meristem in an apricot tree. Using a new *de novo* genome assembly, we identified and characterised somatic mutations within fruits and leaves sampled from the tips of multiple branches. Analysing layer-enriched tissues revealed an unexpectedly high load of layer-specific mutations that were hidden in somatic mutation analyses so far. We confirmed their meristematic origin and layer-specific identity by analysing the distribution of individual mutations between the organs and using single-cell transcriptomics analysis of the leaves. Our results provide a holistic understanding of the genomic heterogeneity between the different meristematic cell lineages that remain the building blocks of plants' genomic architecture.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Results

Identifying somatic mutations in a fruit tree

Prunus armeniaca cultivar (cv.) Rojo Pasión is a registered apricot variety that originated from a cross between *P. armeniaca* cv. Orange Red and cv. Currot (*20*). For this study, we selected a Rojo Pasión tree growing in an orchard close to Murcia, Spain. This tree is of particular interest as some of its branches show low chill requirements and early-flowering phenotype (*21*). Its primary stem branched into three secondary stems, which further divided into several branches. In the spring of 2020, we collected leaf and fruit samples from the tips of seven branches (Fig. 1A) (Supplementary Methods).

We first generated a chromosome-level haplotype resolved genome assembly of the diploid genome of Rojo Pasión using DNA from leaves of several different branches. PacBio HiFi reads were separated into two sets of reads derived from either the maternal or paternal haplotypes using the trio-binning approach (*22*, *23*). The read sets were independently assembled into a haplotype-resolved, chromosome-level assembly (*k-mer*-based phasing accuracy: 99.99%, assembly completeness: >98%, assembly QV: >45) (Supplementary Methods, Supplementary Figures S1-3) (*24*, *25*). The two haplotype-specific assemblies were 233.2 and 234.8 Mbp long (estimated genome size: ~242.5 Mbp) with NG50 values of 26.9 and 28.0 Mbp for the haplotypes derived from Currot and Orange Red, respectively. We annotated 28,355 and 28,473 genes in the two genomes corresponding to a BUSCO completeness score of 96.8% for both haplotypes (Supplementary Methods) (*26*).

We dissected the peel and the flesh of individual fruits from seven different branches to generate layer-enriched samples. This enriched the peel samples for cells from L1 and the flesh samples for cells from L2 (Fig. 1A, Supplementary Methods) (*7*). In apricot fruits, cells from L3 are restricted to a thin layer of the endocarp, which was not considered in this study. We also collected seven individual leaves that were adjacent to each of the selected fruits. The DNA of these 21 samples (14 from fruits and seven from leaves) were sequenced using Illumina paired-end sequencing with very high sequence coverage ranging from 198 to 422x per sample (Supplementary Table S1, Supplementary Methods).

After aligning the sequencing reads to the new assembly, we combined multiple pipelines to identify somatic mutations in the fruits (Supplementary Methods, Supplementary Tables S2-S6, Supplementary Figure S4). Overall, we identified 215 small *de novo* mutations (point mutations and small indels), six loss of heterozygosity (LOH) mutations, where heterozygous alleles changed into homozygous alleles, and two complex mutations (Supplementary Tables S3-6).

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Despite thorough screening, no transposable element movement or meiosis-like chromosome arm exchanges could be identified (*27*). To confirm that the somatic mutations were real, we targeted the validation of 20 somatic mutations with digital PCR. Of those, we could confirm 14 mutations, while for the remaining six either the primers or probes were not conclusive (Supplementary Methods, Supplementary Tables S7-S10, Supplementary Figures S5-S7).

Most of the somatic mutations are specific to individual layers

Among the 215 small-scale mutations (point mutations and small indels), 205 were found in the layer-enriched samples of the fruits while the remaining were found in leaves only (discussed in the following sections).

The vast majority of these 205 mutations in fruits were specific to individual layers (94% (n=193)), while only 6% (n=12) were shared between L1 and L2 samples. Likewise, all six LOH mutations were exclusively found in L1 samples. The low number of shared mutations suggested limited but existing cellular exchanges between the meristematic cell layers (*3*, *28*).

Among the layer-specific mutations, there were significantly more mutations (64% (n=123)) in L1 as compared to L2 (36% (n=70)) (Fig. 1B). The higher mutation load in L1 was consistent across all branches suggesting different mutational processes in the different layers of the meristem and pointing to a relaxed control of genome integrity in L1 (Fig. 1A, C) (*29*).

The 193 layer-specific mutations in L1 or L2 included 84 point mutations and 109 small indels. The point mutations were enriched for transitions with a strong bias for C to T and G to A mutations as described for germline mutations in plants before (Fig. 1D) (*15*, *27*, *30*, *31*). Interestingly, however, this bias was not only found in the mutations of L2 (which can be propagated to the next generation) but also among the mutations in L1 (which are not propagated to the next generation) (*6*). A large fraction of these mutations occurred at CpG sites, the most common DNA methylation context in plants, which can trigger the formation of C to T mutations (Fig. 1E) (*31*). In plants, UV-B based DNA damage is associated with mutations, however, the somatic mutations, including those specific to the outer L1 layer, were not enriched for mutation types associated with UV-B (*32*). Like the point mutations, indel mutations also showed a strong bias in their spectrum. Around 70% (76 of the 109) of the indels were 2bp long and occurred in AT dinucleotide microsatellite regions potentially introduced by DNA polymerase slippage (Fig. 1F) (*33*, *34*).

Most somatic mutations were found in transposable elements and repeats (Supplementary Figure S8) (*27*). However, when normalized for genomic abundance, we found that the

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

frequencies of mutations in these regions were similar to mutations in the rest of the genome except for coding regions that featured a significantly lower mutation rate in both L1 and L2 (left tail Fisher's exact test, p-adjusted for L1: 0.005, L2: 0.007) (Fig. 1G). While this is reminiscent of a recent report suggesting lower mutation rates in evolutionarily conserved genes (*35*), this bias, at least in parts, can also be explained by an under-representation of microsatellite regions within genes (*34*). We also checked the effect of sequence diversity between the haplotypes (heterozygosity) on the frequency of somatic mutations (*36*), but for both L1 and L2, somatic mutations were distributed randomly with no observable effect correlated to large-scale or small-scale nucleotide diversity as it was suggested before (Fig. 1H, I, Supplementary Figure S9) (*37*, *38*).

The distribution of somatic mutations throughout the tree

The formation of new branches starts with the development of new axillary meristems from some founding cells of the apical meristem (*39*). If the founding cells carry somatic mutations, the newly formed axillary meristem will inherit these mutations and propagate them into the new branch (*16*).

To understand the propagation of mutations within the tree, we analysed the distribution of the somatic mutations across the different branches of the tree (*13*, *40*). Overall, we found 31% (64 out of 205) somatic mutations in multiple branches. Almost all of them (94% (n=60)) were found in neighbouring branches where their sharing patterns between the branches agreed with the topology of the tree. Four mutations however did not follow this pattern (Fig. 1J, Supplementary Figure S10). Two of them were found in B2 and B3 but not in B1, one was found in B1 and B4 but not in B2 and B3, and the remaining was found in all branches except one. The strong agreement between the distribution of the mutations across branches and the tree topology suggested that meristematic mutations are highly likely to be propagated into new branches - even though we might be blind to some cases where somatic mutations were not propagated. While the distribution of the somatic mutations strongly agreed with the topology of the tree, branch length did not correlate with the number of somatic mutations (Supplementary Figure S11).

The distribution of somatic mutations across the tree was mostly similar for L1 and L2 with one remarkable exception. We found 21 mutations in L1 that were present in all seven branches while only one such mutation was found in L2 (Fig. 1J, Supplementary Figure S12). This high number of mutations occurring in all branches (as compared to mutations in six, five, or four branches) suggested that they might be clonally inherited from the mother tree and did not

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(only) occur during the early development of the tree. Intriguingly, this scenario would also agree with a striking enrichment of mutations in L1.

Mutational load of neighbouring organs reveals meristematic origin

In addition to the 64 mutations that were shared in fruits from different branches, we found 141 somatic mutations that were specific to individual branches. The high number of branch-specific mutations was in contrast to the high probability of mutations growing into newly formed branches, questioning if these mutations occurred in the meristems or the developing fruits.

To understand the origin of these somatic mutations in more detail, we sequenced the DNA of seven leaves that grew adjacent to the sequenced fruits from the seven branches. While leaf tissue also includes all three cell layers, we did not generate layer-specific samples, instead, we deeply sequenced the DNA from bulked leaf cells (sequence coverage from 198-422x). Each of these bulked samples included cells from layers 1, 2, and 3, however, at different amounts (*41*).

We identified 70 somatic mutations within the leaves (Fig. 2A) (Supplementary Methods, Supplementary Table S3). Most of the mutations found in the leaves were also found in the closely linked fruits (85% (n=60)). The remaining mutations (15% (n=10)) were specific to individual leaves. These mutations were also absent in all other samples of the tree suggesting that they were truly specific to individual leaves (Fig. 2B).

In contrast, only 30% (60 out of 205) of the mutations in fruits were shared with the neighbouring leaves including a striking difference between the individual layers. While 70% (49 out of 70) of the L2-specific mutations in the fruits were also found in leaves, not a single L1-specific mutation in the fruits was identified in leaves (0 out of 123) (Fig. 2B).

This difference in mutation sharing in individual layers could result from an under-representation of L1 cells in the leaves as it could affect the identification of L1 mutations in the heterogenous leaf samples (Supplementary Figure S13). To test this, we performed a targeted search for the mutations that were originally found in L1 in the fruits. We re-analysed the sequencing data of the leaves by lowering the required read support and found evidence for 80% of the L1 mutations (Fig. 2C). Lowering the read support requirement during *de novo* mutation identification (to one or two reads used for genotyping) would have led to an extremely high false positive rate, which in turn would make it impossible to find the set of true mutations (*42*, *43*). On the contrary, the number of mutations that could be found in the L2 layers of the fruits and their corresponding leaves was mostly unaffected by changes in read support cut-off

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

implying that almost all of the L2 mutations that were shared between leaves and fruits were already identified (Fig. 2C).

Although we found evidence for 80% of the L1 mutations in the leaves, it is possible that some of them resulted from false positives and sequencing errors. Therefore, to understand how many mutations were truly shared between closely neighbouring organs, we zoomed in on L2 mutations as those were most reliably identified in both organs. To validate the presence/absence of the L2 mutations in both sample types, we compared the read frequencies of the mutations in the fruit and leaf samples. The frequencies of the mutations that were shared in both organs were very similar (Supplementary Figure S14, S15), however, mutations that were specific to either L2 layers or leaves had almost no read support in the respective other samples. This supported the annotation of the presence and absence of the mutations.

Overall, we found that 65% (60 out of 92) of the mutations identified in L2 in fruits or in the leaves were shared between the neighbouring organs and were of putative meristematic origin. This ratio was mostly conserved at all seven individual branches (Fig. 2D). The remaining mutations (35% (n=32)) were observed in either the leaves or the fruits only. All these 32 mutations were limited to single branches and were not found in other branches. This suggests that these mutations either happened during the development of the specific organs or that the mutations were absent in the founding cells of the meristem from which these organs developed (*40*). Taking together, two-thirds of the somatic mutations were conclusively of meristematic origin, however, the likely percentage of meristematic mutations could even be much higher.

Meristematic mutations affect the expression of genes only in the respective layer

Layer-specific, meristematic mutations can affect the function of cells that originated from respective layers. Mutations in layer 2, for example, are expected to affect all the different cell types within the mesophyll of a leaf, while all other cells should not be affected.

To test if meristematic mutations affect all cell types of only the mutation-containing layer, we sequenced the transcriptomes of the same leaves that we used for mutation identification before. Conventional bulk RNA sequencing of entire leaves, however, combines the transcripts of different cell populations including the cells from different layers. Therefore, to identify and analyse the expression of specific somatic mutations in individual cell populations, we used single-cell RNA sequencing for four different branches (B2, B4, B5, and B6) (Fig. 3A, Supplementary Methods).

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The four datasets consisted of a total of 10,293 cells corresponding to fifteen different transcription clusters (Fig. 3B). The clusters were classified as epidermis (L1 consisting of two transcription clusters), mesophyll (L2; six clusters) or vascular (L3; four clusters) based on cell-type specific marker genes, gene set enrichment analysis and comparison to other leaf expression atlases (Supplementary Methods, Supplementary Tables S11-S15) (*44–46*). Most cells corresponded to mesophyll (78.2% (n=8051)), while the epidermis (8.9% (n=917)) and vascular (7.2% (n=744)) were represented by fewer cells (Supplementary Figure S16). The low representation of L1 cells (epidermis) within the leaves' samples corroborated our earlier observation that it is difficult to find somatic mutations in the L1 layer in bulked leaf tissue. The remaining cells (5.6% (n=581)) were found in three small clusters that were specific to individual branches. One of these clusters consisted of cells expressing biotic stress response genes while the two other clusters showed gene expression patterns similar to dividing cells. These branch-specific clusters are likely a reflection of the differences in the micro-environment and developmental stages of the different leaves and outline the sampling biases in single-cell experiments from wild collections.

The scRNA-seq libraries used here typically sequence only the ends of transcripts making it inefficient to probe somatic mutations (*47*). To analyse entire transcripts, we also sequenced each of the single-cell libraries with PacBio HiFi long-read sequencing (scISO-seq) (Fig. 3A, Supplementary Methods). This approach ensured that transcripts from the same cells had the same barcodes in both scRNA-seq and scISO-seq datasets. This allowed us to use the expression-based clustering generated with scRNA-seq datasets to cluster the scISO-seq reads and search for mutations based on full-length sequenced transcripts (Supplementary Figure S17). This led to an additional sequencing of 8.2 to 10.5 Mb (12.5–20%) of genomic regions, which were not sequenced using the scRNA-seq alone. Now, these regions could also be analyzed for somatic mutations (Supplementary Figure S18).

Overall, we identified mutant alleles for six somatic point mutations in these four branches (Fig. 3C). Reassuringly, all mutant alleles were found in the branches in which they were originally found. In the genome sequencing data, five of these mutations were identified in both the leaf as well as the L2 samples of the fruits, while one was found in the leaf only. All of these were absent in the L1 samples of the fruits.

All six L2 mutations were almost exclusively found in cells assigned to L2 clusters (Fig. 3C). Across all mutations, 48 L2 cells featured the mutant alleles, while only three L1 and none of the L3 cells featured reads with mutant alleles. While we suspect some migration of cells between

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

layers 1 and 2, the low number of layer 1 cells could also result from inconsistencies in the clustering of the cells.

Two of the mutations were highly expressed (SM18 and SM165 in Fig. 3C). Cells carrying the mutant alleles of these two mutations could be found in most of the transcription clusters of L2 (SM18 was found in five and SM165 in all six transcription clusters of L2). The other four mutations were only supported by a few cells due to the overall low transcript counts and not due to genetic heterogeneity in the L2 cells (Supplementary Figure S19, S20). The number of cells with the mutant alleles of these four mutations in the L2 matched the number of cells with reads of the wild-type allele in the branch suggesting that they are also genuine L2 mutations.

In summary, with the help of single-cell expression data, we could increase the resolution at which we can analyse the distribution of somatic mutations within complex tissues. We found meristematic mutations (once they grow into developing tissue) retain their layer-specific identity and occur only in the cell lineages that develop from the respective meristematic layers.

Conclusions

We analysed the spectra, distribution, and expression of somatic mutations in the different cell layers of an apricot tree. The different layers have their origin in the architecture of the meristem. The meristem is organized in separate layers leading to distinct cell lineages in the newly formed branches and organs. In each of the layers, we found distinct sets of somatic mutations with only a small proportion of mutations shared between them. The cells of specific layers from different branches were more similar to each other as compared to cells from different layers that developed in the same organ. This suggested that the individual layers of the tree developed almost independently and that selection on somatic cells would act independently on individual layers.

Layer 1 (the outer layer often forming the epidermis) of the fruit tree had significantly more mutations compared to layer 2, similar to a recent report in potato, where regenerated plants revealed different mutational loads depending on the tissue they were regenerated from (*48*). Most of the somatic mutations were shared between adjacent organs implying that they were of meristematic origin. We also confirmed the layer-specific identity of some of the layer-specific somatic mutations using scRNA sequencing of leaves, showing that layer-specific somatic mutations alter the transcripts exclusively in the cells of the respective layer and thereby forming the basis for layer-specific phenotypes.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

We observed a remarkably high number of layer-specific somatic mutations within an individual tree. Particularly, layer 1, which is typically underrepresented in bulked samples compared to layer 2, had a strikingly higher mutational load. This finding aligns well with the frequent identification of bud sports specific to layer 1 in various fruit trees (*13*). However, as we described here, identification of genomic mutations in layer 1 is difficult, which may elucidate why, even among commercially relevant cultivars, the genetic basis of only a few sport mutants has been uncovered so far.

Our work suggests that a focused analysis of individual cell layers is necessary for obtaining a comprehensive understanding of the presence and distribution of somatic mutations. Although less straightforward than analysing bulked samples, layer-specific analyses can reveal the previously hidden spectrum of mutations, offering a holistic perspective on the accumulation and propagation of somatic mutations.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Acknowledgements

The authors would like to thank Hernán López for helpful discussions, Samija Amar for help in developing illustrations and figures, and Antonio Molina for providing access to the plant material.

Funding

This work was funded by the "Humboldt Research Fellowship for Experienced Researchers" (Alexander von Humboldt Foundation) (J.A.C.), the Marie Skłodowska-Curie Individual Fellowship PrunMut (789673) (J.A.C.), Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2048/1–390686111 (K.S.) and European Research Council (ERC) grant 'INTERACT' (802629) (K.S.).

Author contributions

Conceptualization: MG, JAC, KK, RB, KS

Methodology: MG, JAC, KK, LCB, AS, HS, BW, MM

Investigation: MG, JAC, KK, LCB, AS, HS

Funding acquisition: JAC, KS

Supervision: DR, BH, KS

Writing – original draft: MG

Writing – review & editing: MG, KS

Competing interests

The authors declare that they have no competing interests.

Data and materials availability

Sequencing data, assemblies, and annotations will be made available on acceptance of the manuscript under Bioproject PRJEB71142. All scripts used for analysis and processing are available at https://github.com/schneebergerlab/apricot_layer_mutations.

Supplementary Materials

Methods and materials
Supplementary Figures S1 to S20;
Supplementary Tables S1 to S15

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References (49-85)

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1. A. Schmidt, Histologishe Studien an phanerogamen Vegetationspunkten. *Bot. Arch.* **8**, 345–404 (1924).

2. R. A. Popham, Principal Types of Vegetative Shoot Apex Organization in Vascular Plants. (1951).

3. M. H. Frank, D. H. Chitwood, Plant chimeras: The good, the bad, and the 'Bizzaria.' *Developmental Biology* **419**, 41–53 (2016).

4. E. Huala, I. Sussex, Determination and Cell Interactions in Reproductive Meristems. *Plant Cell* **5**, 1157–1165 (1993).

5. H. Dermen, Directional Cell Division in Shoot Apices. *Cytologia* **34**, 541–558 (1969).

6. P. D. Jenik, V. F. Irish, Regulation of cell proliferation patterns by homeotic genes during Arabidopsis floral development. *Development* **127**, 1267–1276 (2000).

7. H. Dermen, HISTOGENETIC FACTORS IN COLOR AND FUZZLESS PEACH SPORTS. *Journal of Heredity* **47**, 64–76 (1956).

8. E. J. Klekowski, "Mutation, Developmental Selection, and Plant Evolution" in *Mutation, Developmental Selection, and Plant Evolution* (Columbia University Press, 2019).

9. I. Granhall, Spontaneous and Induced Bud Mutations in Fruit Trees. *Acta Agriculturae Scandinavica* **4**, 594–600 (1954).

10. A. D. SHAMEL, C. S. POMEROY, BUD MUTATIONS IN HORTICULTURAL CROPS. *Journal of Heredity* **27**, 487–494 (1936).

11. W. R. Okie, *Handbook of Peach and Nectarine Varieties: Performance in the Southeastern United States and Index of Names* (U.S. Department of Agriculture, Agricultural Research Service, 1998).

12. D. J. Schoen, S. T. Schultz, Somatic Mutation and Evolution in Plants. *Annual Review of Ecology, Evolution, and Systematics* **50**, 49–73 (2019).

13. T. M. Foster, M. J. Aranzana, Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. *Horticulture Research* **5**, 44 (2018).

14. V. C. T. Hanlon, S. P. Otto, S. N. Aitken, Somatic mutations substantially increase the per-generation mutation rate in the conifer Picea sitchensis. *Evolution Letters* **3**, 348–358 (2019).

15. E. Perez-roman, A. L. Usach, Single-nucleotide mosaicism in citrus□: Estimations of somatic. 1–13 (2021).

16. S. Tomimoto, A. Satake, Modelling somatic mutation accumulation and expansion in a long-lived tree with hierarchical modular architecture. *Journal of Theoretical Biology* **565**, 111465 (2023).

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

17. Z. Zheng, H. Hu, W. Lei, J. Zhang, M. Zhu, Y. Li, X. Zhang, J. Ma, D. Wan, T. Ma, G. Ren, D. Ru, Somatic mutations during rapid clonal domestication of Populus alba var. pyramidalis. *Evolutionary Applications* **15**, 1875–1887 (2022).

18. Y. Ji, X. Chen, S. Lin, M. B. Traw, D. Tian, S. Yang, L. Wang, J. Huang, High level of somatic mutations detected in a diploid banana wild relative Musa basjoo. *Mol Genet Genomics* **298**, 67–77 (2023).

19. L. Wang, Y. Ji, Y. Hu, H. Hu, X. Jia, M. Jiang, X. Zhang, L. Zhao, Y. Zhang, Y. Jia, C. Qin, L. Yu, J. Huang, S. Yang, L. D. Hurst, D. Tian, The architecture of intra-organism mutation rate variation in plants. *PLOS Biology* **17**, e3000191 (2019).

20. J. Egea, F. Dicenta, L. Burgos, `Rojo Pasión' Apricot. *HortScience* **39**, 1490–1491 (2004).

21. D. Ruiz, B. E. García-Gómez, J. Egea, A. Molina, P. Martínez-Gómez, J. A. Campoy, Phenotypical characterization and molecular fingerprinting of natural early-flowering mutants in apricot (Prunus armeniaca L.) and Japanese plum (P. salicina Lindl.). *Scientia Horticulturae* **254**, 187–192 (2019).

22. J. A. Campoy, H. Sun, M. Goel, W. B. Jiao, K. Folz-Donahue, N. Wang, M. Rubio, C. Liu, C. Kukat, D. Ruiz, B. Huettel, K. Schneeberger, Gamete binning: chromosome-level and haplotype-resolved genome assembly enabled by high-throughput single-cell sequencing of gamete genomes. *Genome Biology* **21**, 306 (2020).

23. S. Koren, A. Rhie, B. P. Walenz, A. T. Dilthey, D. M. Bickhart, S. B. Kingan, S. Hiendleder, J. L. Williams, T. P. L. Smith, A. M. Phillippy, De novo assembly of haplotype-resolved genomes with trio binning. *Nature Biotechnology* **36**, 1174–1182 (2018).

24. A. Rhie, B. P. Walenz, S. Koren, A. M. Phillippy, Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**, 245 (2020).

25. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).

26. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

27. B. T. Hofmeister, J. Denkena, M. Colomé-Tatché, Y. Shahryary, R. Hazarika, J. Grimwood, S. Mamidi, J. Jenkins, P. P. Grabowski, A. Sreedasyam, S. Shu, K. Barry, K. Lail, C. Adam, A. Lipzen, R. Sorek, D. Kudrna, J. Talag, R. Wing, D. W. Hall, D. Jacobsen, G. A. Tuskan, J. Schmutz, F. Johannes, R. J. Schmitz, A genome assembly and the somatic genetic and epigenetic mutation rate in a wild long-lived perennial Populus trichocarpa. *Genome Biology* **21**, 259 (2020).

28. X. Xu, C. Smaczniak, J. M. Muino, K. Kaufmann, Cell identity specification in plants: lessons from flower development. *Journal of Experimental Botany* **72**, 4202–4217 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

29. R. K. Yadav, M. Tavakkoli, M. Xie, T. Girke, G. V. Reddy, A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. *Development* **141**, 2735–2744 (2014).

30. T. H. Jukes, Transitions, transversions, and the molecular evolutionary clock. *J Mol Evol* **26**, 87–98 (1987).

31. S. Ossowski, K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, M. Lynch, The rate and molecular spectrum of spontaneous mutations in arabidopsis thaliana. *Science* **327**, 92–94 (2010).

32. E. M. Willing, T. Piofczyk, A. Albert, J. B. Winkler, K. Schneeberger, A. Pecinka, UVR2 ensures transgenerational genome stability under simulated natural UV-B in Arabidopsis thaliana. *Nature Communications* **7**, 1–9 (2016).

33. H. Ellegren, Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* **5**, 435–445 (2004).

34. T. N. Marriage, S. Hudman, M. E. Mort, M. E. Orive, R. G. Shaw, J. K. Kelly, Direct estimation of the mutation rate at dinucleotide microsatellite loci in Arabidopsis thaliana (Brassicaceae). *Heredity* **103**, 310–317 (2009).

35. J. G. Monroe, T. Srikant, P. Carbonell-Bejerano, C. Becker, M. Lensink, M. Exposito-Alonso, M. Klein, J. Hildebrandt, M. Neumann, D. Kliebenstein, M. L. Weng, E. Imbert, J. Ågren, M. T. Rutter, C. B. Fenster, D. Weigel, Mutation bias reflects natural selection in Arabidopsis thaliana. *Nature* **602**, 101–105 (2022).

36. S. Yang, L. Wang, J. Huang, X. Zhang, Y. Yuan, J.-Q. Chen, L. D. Hurst, D. Tian, Parent–progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* **523**, 463–467 (2015).

37. M. Goel, H. Sun, W. B. Jiao, K. Schneeberger, SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology* **20**, 277 (2019).

38. M. Goel, K. Schneeberger, plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**, 2922–2926 (2022).

39. A. Burian, P. Barbier de Reuille, C. Kuhlemeier, Patterns of Stem Cell Divisions Contribute to Plant Longevity. *Current Biology* **26**, 1385–1394 (2016).

40. L. Yu, C. Boström, S. Franzenburg, T. Bayer, T. Dagan, T. B. H. Reusch, Somatic genetic drift and multilevel selection in a clonal seagrass. *Nature Ecology and Evolution* **4**, 952–962 (2020).

41. D. Manuela, M. Xu, Patterning a Leaf by Establishing Polarities. *Frontiers in Plant Science* **11** (2020).

42. Z. Chen, Y. Yuan, X. Chen, J. Chen, S. Lin, X. Li, H. Du, Systematic comparison of somatic variant calling performance among different sequencing depth and mutation frequency. *Sci Rep* **10**, 3501 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

43. D. Sims, I. Sudbery, N. E. Ilott, A. Heger, C. P. Ponting, Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* **15**, 121–132 (2014).

44. D. Szklarczyk, A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L. J. Jensen, C. von Mering, The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* **49**, D605–D612 (2021).

45. J.-Y. Kim, E. Symeonidi, T. Y. Pang, T. Denyer, D. Weidauer, M. Bezrutczyk, M. Miras, N. Zöllner, T. Hartwig, M. M. Wudick, M. Lercher, L.-Q. Chen, M. C. P. Timmermans, W. B. Frommer, Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *The Plant Cell* **33**, 511–530 (2021).

46. R. Tenorio Berrío, K. Verstaen, N. Vandamme, J. Pevernagie, I. Achon, J. Van Duyse, G. Van Isterdael, Y. Saeys, L. De Veylder, D. Inzé, M. Dubois, Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells. *Plant Physiology* **188**, 898–918 (2022).

47. A. A. Petti, S. R. Williams, C. A. Miller, I. T. Fiddes, S. N. Srivatsan, D. Y. Chen, C. C. Fronick, R. S. Fulton, D. M. Church, T. J. Ley, A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. *Nature Communications* **10** (2019).

48. K. R. Amundson, M. P. A. Marimuthu, O. Nguyen, K. Sarika, I. J. DeMarco, A. Phan, I. M. Henry, L. Comai, Differential mutation accumulation in plant meristematic layers. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.09.25.559363.

49. S. Nurk, B. P. Walenz, A. Rhie, M. R. Vollger, G. A. Logsdon, R. Grothe, K. H. Miga, E. E. Eichler, A. M. Phillippy, S. Koren, HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **30**, 1291–1305 (2020).

50. M. Kokot, M. Długosz, S. Deorowicz, KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* **33**, 2759–2761 (2017).

51. H. Jiang, R. Lei, S.-W. Ding, S. Zhu, Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC bioinformatics* **15**, 182 (2014).

52. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

53. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–9 (2009).

54. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **9**, e112963 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

55. R. Vaser, I. Sovic, N. Nagarajan, M. Sikic, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.*, gr.214270.116 (2017).

56. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).

57. J. Zhang, X. Zhang, H. Tang, Q. Zhang, X. Hua, X. Ma, F. Zhu, T. Jones, X. Zhu, J. Bowers, C. M. Wai, C. Zheng, Y. Shi, S. Chen, X. Xu, J. Yue, D. R. Nelson, L. Huang, Z. Li, H. Xu, D. Zhou, Y. Wang, W. Hu, J. Lin, Y. Deng, N. Pandey, M. Mancini, D. Zerpa, J. K. Nguyen, L. Wang, L. Yu, Y. Xin, L. Ge, J. Arro, J. O. Han, S. Chakrabarty, M. Pushko, W. Zhang, Y. Ma, P. Ma, M. Lv, F. Chen, G. Zheng, J. Xu, Z. Yang, F. Deng, X. Chen, Z. Liao, X. Zhang, Z. Lin, H. Lin, H. Yan, Z. Kuang, W. Zhong, P. Liang, G. Wang, Y. Yuan, J. Shi, J. Hou, J. Lin, J. Jin, P. Cao, Q. Shen, Q. Jiang, P. Zhou, Y. Ma, X. Zhang, R. Xu, J. Liu, Y. Zhou, H. Jia, Q. Ma, R. Qi, Z. Zhang, J. Fang, H. Fang, J. Song, M. Wang, G. Dong, G. Wang, Z. Chen, T. Ma, H. Liu, S. R. Dhungana, S. E. Huss, X. Yang, A. Sharma, J. H. Trujillo, M. C. Martinez, M. Hudson, J. J. Riascos, M. Schuler, L.-Q. Chen, D. M. Braun, L. Li, Q. Yu, J. Wang, K. Wang, M. C. Schatz, D. Heckerman, M.-A. Van Sluys, G. M. Souza, P. H. Moore, D. Sankoff, R. VanBuren, A. H. Paterson, C. Nagai, R. Ming, Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. *Nat Genet* **50**, 1565–1573 (2018).

58. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

59. R. K. Dale, B. S. Pedersen, A. R. Quinlan, Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).

60. SALSA: A tool to scaffold long read assemblies with Hi-C, MarBL (2023); https://github.com/marbl/SALSA.

61. J. R. Belyeu, M. Chowdhury, J. Brown, B. S. Pedersen, M. J. Cormier, A. R. Quinlan, R. M. Layer, Samplot: a platform for structural variant visual validation and automated filtering. *Genome Biology* **22**, 161 (2021).

62. T. Alioto, K. G. Alexiou, A. Bardil, F. Barteri, R. Castanera, F. Cruz, A. Dhingra, H. Duval, Á. Fernández i Martí, L. Frias, B. Galán, J. L. García, W. Howad, J. Gómez-Garrido, M. Gut, I. Julca, J. Morata, P. Puigdomènech, P. Ribeca, M. J. Rubio Cabetas, A. Vlasova, M. Wirthensohn, J. Garcia-Mas, T. Gabaldón, J. M. Casacuberta, P. Arús, Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. *The Plant Journal* **101**, 455–472 (2020).

63. Q. Zhang, W. Chen, L. Sun, F. Zhao, B. Huang, W. Yang, Y. Tao, J. Wang, Z. Yuan, G. Fan, Z. Xing, C. Han, H. Pan, X. Zhong, W. Shi, X. Liang, D. Du, F. Sun, Z. Xu, R. Hao, T. Lv, Y. Lv, Z. Zheng, M. Sun, L. Luo, M. Cai, Y. Gao, J. Wang, Y. Yin, X. Xu, T. Cheng, J. Wang, The genome of Prunus mume. *Nat Commun* **3**, 1318 (2012).

64. I. Verde, A. G. Abbott, S. Scalabrin, S. Jung, S. Shu, F. Marroni, T. Zhebentyayeva, M. T. Dettori, J. Grimwood, F. Cattonaro, A. Zuccolo, L. Rossini, J. Jenkins, E. Vendramin, L. A. Meisel, V. Decroocq, B. Sosinski, S. Prochnik, T. Mitros, A. Policriti, G. Cipriani, L. Dondini, S. Ficklin, D. M. Goodstein, P. Xuan, C. D. Fabbro, V. Aramini, D. Copetti, S. Gonzalez, D. S. Horner, R. Falchi, S. Lucas, E. Mica, J. Maldonado, B. Lazzari, D. Bielenberg, R. Pirona, M. Miculan, A. Barakat, R. Testolin, A. Stella, S. Tartarini, P. Tonutti, P. Arús, A. Orellana,

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

C. Wells, D. Main, G. Vizzotto, H. Silva, F. Salamini, J. Schmutz, M. Morgante, D. S. Rokhsar, The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* **45**, 487–494 (2013).

65. M. Alonge, L. Lebeigle, M. Kirsche, K. Jenike, S. Ou, S. Aganezov, X. Wang, Z. B. Lippman, M. C. Schatz, S. Soyk, Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology* **23**, 258 (2022).

66. M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* **34**, W435–W439 (2006).

67. W. H. Majoros, M. Pertea, S. L. Salzberg, TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).

68. A. D. Johnson, R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O'Donnell, P. I. W. de Bakker, SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939 (2008).

69. M. Pertea, G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell, S. L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295 (2015).

70. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).

71. B. J. Haas, S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen, J. Orvis, O. White, C. R. Buell, J. R. Wortman, Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**, R7 (2008).

72. P. Smit, AFA, Hubley, R & Green, RepeatMasker Open-4.0., *http://www.repeatmasker.org*.

73. O. Keller, F. Odronitz, M. Stanke, M. Kollmar, S. Waack, Scipio: Using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC Bioinformatics* **9**, 278 (2008).

74. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr, L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**, 5654–5666 (2003).

75. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644–652 (2011).

76. J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S. K. Forslund, H. Cook, D. R. Mende, I. Letunic, T. Rattei, L. J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309–D314 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

77. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).

78. A. Khanna, D. E. Larson, S. N. Srivatsan, M. Mosior, T. E. Abbott, S. Kiwala, T. J. Ley, E. J. Duncavage, M. J. Walter, J. R. Walker, O. L. Griffith, M. Griffith, C. A. Miller, Bam-readcount - rapid generation of basepair-resolution sequence metrics. *Journal of Open Source Software* **7**, 3722 (2022).

79. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178–192 (2013).

80. K. Okonechnikov, O. Golosova, M. Fursov, the UGENE team, Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).

81. R. Team, R: A language and environment for statistical computing. *MSOR connections* (2014).

82. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

83. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

84. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157 (2015).

85. M. van Kempen, S. S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C. L. M. Gilchrist, J. Söding, M. Steinegger, Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*, 1–4 (2023).

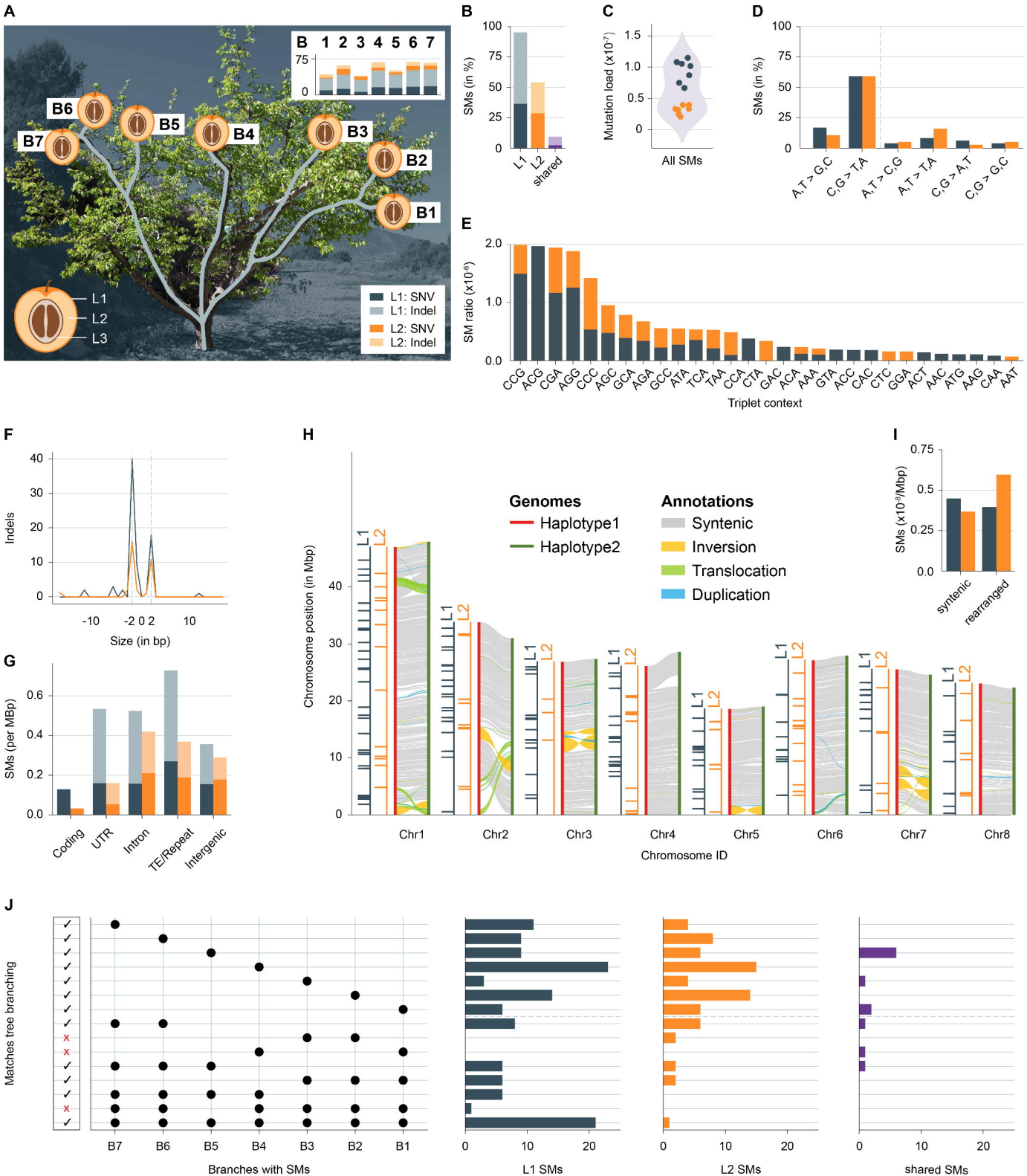
bioRxiv preprint doi: https://doi.org/10.1101/2024.01.04.573414; this version posted January 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

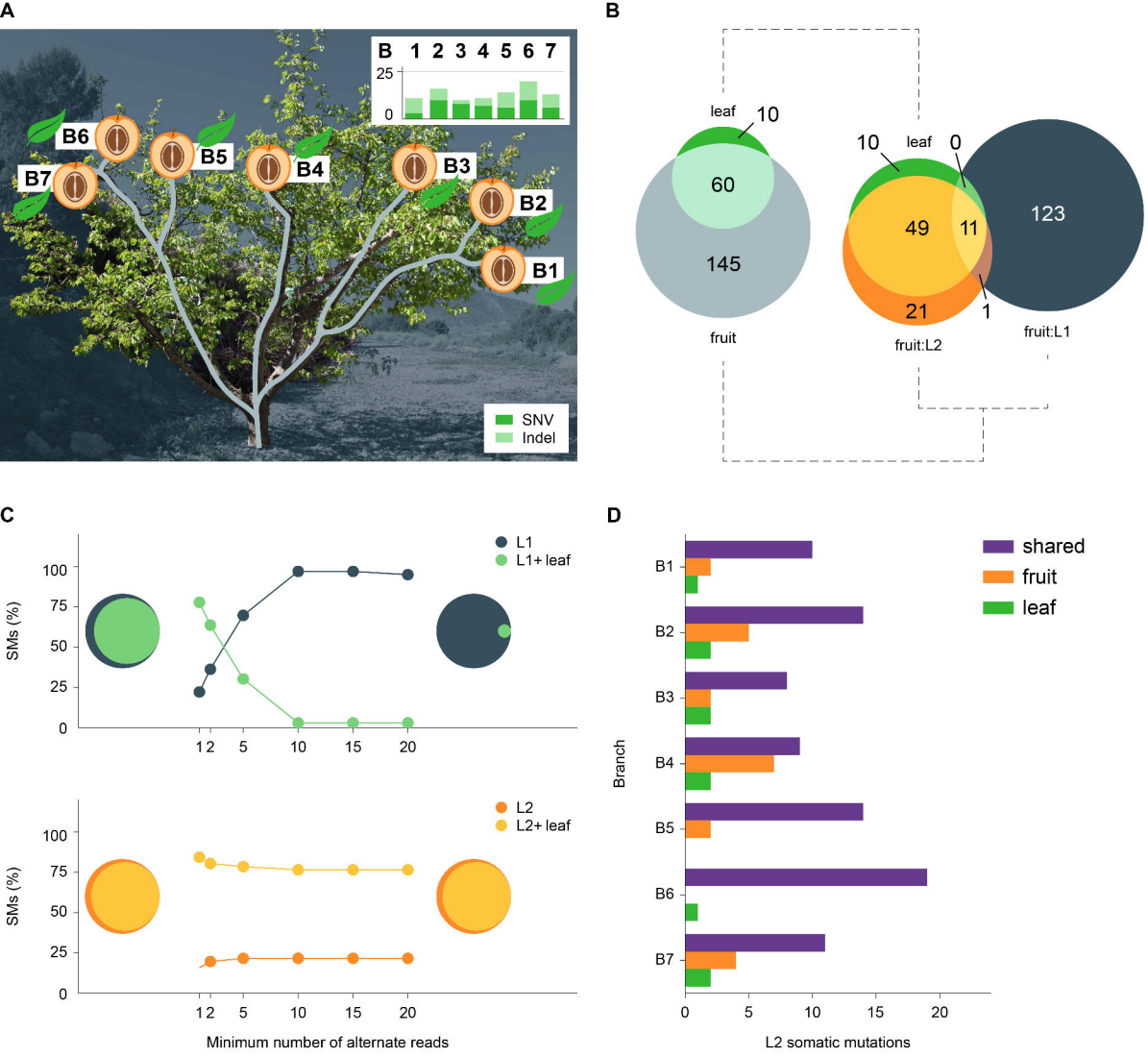
Figure Legend

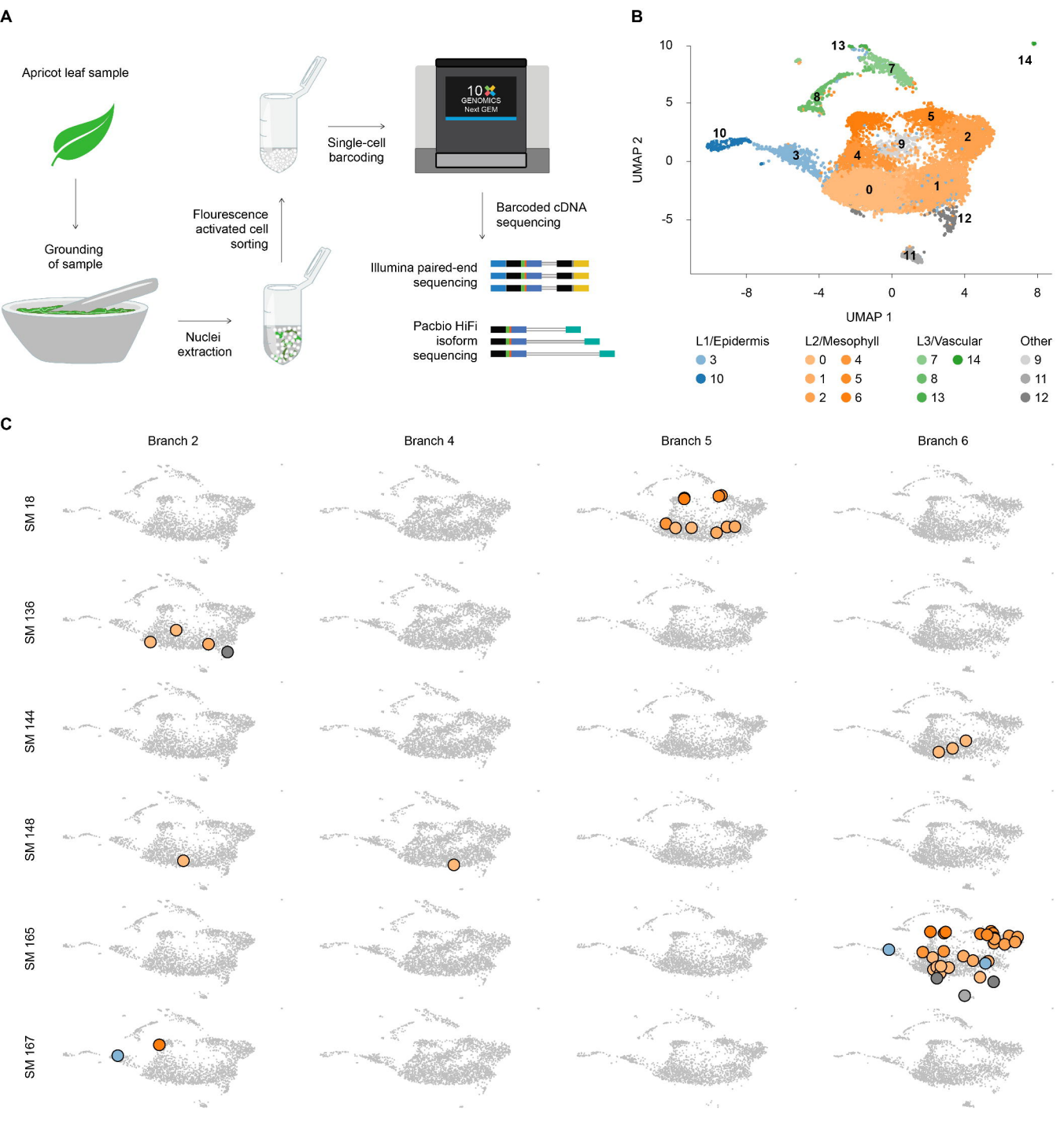
Fig. 1 . Layer-specific mutations within different branches. A) Schematic layout of the seven branches in the apricot tree. The apricot fruit peel consists of L1, mesocarp of L2, and pre-endocarp of L3 cells. Bar plots show the number of somatic mutations identified in each branch with mutations in L1 in grey and mutations in L2 in orange. SNVs have a dark hue while indels are light. **B)** Number of somatic mutations in L1, L2 or both. **C)** Mutation load in L1 and L2. **D)** Percent of transition and transversion point mutations. **E)** Triplet context of point mutations (base in the middle is mutated and canonical triplets are used). **F)** Indel size distribution. Negative values correspond to deletion whereas positive values correspond to insertions. **G)** Normalised somatic mutation frequency in different genomic regions. UTR: untranslated regions; TE: transposable element. **H)** Plotsr visualisation of structural differences between the two haplotypes and the distribution of somatic mutations across the chromosomes. **I)** Normalised somatic mutation frequency in syntenic and structurally rearranged regions. The y-values were normalised by the total number of mutations identified in the layer. **J)** Distribution of somatic mutations across branches. The five panels show (left to right) i) whether the mutation set follows the topology of the tree (black checkmark) or not (red cross), ii) the sets of branches in which mutations are present, each row corresponds to a set and the black point marks the branches present in the set, iii-v) the number of somatic mutations in only L1, only L2, or both that are present in the set.

Fig. 2. Genomic differences between neighbouring fruits and leaves. A) Leaves next to the selected fruits were used for somatic mutation identification. The bar plots show the number of somatic mutations identified in each branch with SNVs in dark green and indels in light green. **B)** Both Venn diagrams show the number of somatic mutations shared between fruits and leaves. Left: all mutations; right: somatic mutations in fruits are shown separately for L1 and L2. **C)** Genotyping of L1 and L2 somatic mutations of the fruits in the neighbouring leaves. Upper panel: genotyping L1 and leaf samples for mutations identified in L1. The X-axis shows the minimum number of reads with alternate alleles required to genotype a somatic mutation. The Y-axis shows the percent of somatic mutations that were present either only in L1 or in both. The Venn diagrams on the left and right show the sharing of somatic mutations between L1 and leaf for read cut-offs of 1 and 20, respectively. Lower panel: similar plot for genotyping L2 and leaf samples for mutations identified in L2. **D)** Number of L2 somatic mutations (X-axis) that were present either only in fruit, only in leaf, or in both.

Fig. 3. Layer-specific expression of somatic mutations. A) Workflow for generating scRNA-seq and scISO-seq sequencing data from a single leaf. **B)** Expression atlas of apricot leaves. The 15 clusters are numbered 0-14. Clusters from L1, L2, and L3 are coloured with blue, orange, and green shades, respectively. Clusters corresponding to micro-environmental and developmental variations, which were mostly specific to individual branches, are coloured in shades of grey. **C)** Cells containing reads with mutant alleles for six somatic mutations. The four columns correspond to four branches. Each row corresponds to one somatic mutation. Grey cells do not have any read with the mutant allele, whereas cells with mutant alleles are coloured based on the cluster they are from.


A
B 1 2 3 4 5 6 7
75
0
B6
B5
B4
B3
B7
B2
B1
L1
L2
L3
L1: SNV
L1: Indel
L2: SNV
L2: Indel
B
SMs (in %)
100
75
50
25
0
L1
L2
shared
C
Mutation load (x10^-7)
1.5
1.0
0.5
0
All SMs
D
SMs (in %)
A,T > G,C
C,G > T,A
A,T > C,G
A,T > T,A
C,G > A,T
C,G > G,C
E
SM ratio (x10^-6)
2.0
1.0
0.0
CCG ACG CGA AGG CCC AGC GCA AGA GCC ATA TCA TAA CCA CTA GAC ACA AAA GTA ACC CAC CTC GGA ACT AAC ATG AAG CAA AAT
Triplet context
F
Indels
40
30
20
10
0
-10
-2 0 2
10
Size (in bp)
G
SMs (per MBp)
0.6
0.4
0.2
0.0
Coding
UTR
Intron
TE/Repeat
Intergenic
H
Genomes
Haplotype1
Haplotype2
Annotations
Syntenic
Inversion
Translocation
Duplication
Chromosome position (in Mbp)
40
30
20
10
0
L1
L2
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chromosome ID
I
SMs (x10^-3/Mbp)
0.75
0.5
0.25
0.0
syntenic
rearranged
J
Matches tree branching
B7
B6
B5
B4
B3
B2
B1
Branches with SMs
L1 SMs
L2 SMs
shared SMs
0
10
20


A
B 1 2 3 4 5 6 7
25
0
B6
B7
B5
B4
B3
B2
B1
SNV
Indel
B
leaf
10
60
145
fruit
10
leaf
0
49
11
123
21
1
fruit:L2
fruit:L1
C
L1
L1+ leaf
SMs (%)
100
75
50
25
0
1 2 5 10 15 20
L2
L2+ leaf
Minimum number of alternate reads
D
shared
fruit
leaf
Branch
B1
B2
B3
B4
B5
B6
B7
0 5 10 15 20
L2 somatic mutations


A
Apricot leaf sample
Grounding of sample
Nuclei extraction
Flourescence activated cell sorting
Single-cell barcoding
10 GENOMICS Next GEM
Barcoded cDNA sequencing
Illumina paired-end sequencing
Pacbio HiFi isoform sequencing
B
UMAP 2
UMAP 1
L1/Epidermis
3
10
L2/Mesophyll
0
1
2
4
5
6
L3/Vascular
7
8
13
14
Other
9
11
12
C
Branch 2
Branch 4
Branch 5
Branch 6
SM 18
SM 136
SM 144
SM 148
SM 165
SM 167