UNIVERSITY OF COPENHAGEN



Københavns Universitet

Occurrence of aneuploidy across the range of coast redwood (Sequoia sempervirens)

Nikolaeva, Alexandra Sasha; Santangelo, James; Smith, Lydia; Dodd, Richard; Nielsen, Rasmus

Published in:
G3: Genes, Genomes, Genetics

DOI:
10.1093/g3journal/jkaf063

Publication date:
2025

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Nikolaeva, A. S., Santangelo, J., Smith, L., Dodd, R., & Nielsen, R. (2025). Occurrence of aneuploidy across the range of coast redwood (*Sequoia sempervirens*). *G3: Genes, Genomes, Genetics*, *15*(5), Article jkaf063. https://doi.org/10.1093/g3journal/jkaf063

Download date: 18. Aug. 2026

G3, 2025, **15(5)**, jkaf063
https://doi.org/10.1093/g3journal/jkaf063
Advance Access Publication Date: 25 March 2025
Plant Genetics and Genomics

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

Occurrence of aneuploidy across the range of coast redwood (*Sequoia sempervirens*)

Alexandra Sasha Nikolaeva [1,2,*] James Santangelo,[3] Lydia Smith,[4] Richard Dodd,[1] Rasmus Nielsen [3,5,6]

[1]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA
[2]Center for Computational Biology, University of California, Berkeley, CA 94720, USA
[3]Department of Integrative Biology, University of California, Berkeley, CA 94720, USA
[4]Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA
[5]Department of Statistics, University of California, Berkeley, CA 94720, USA
[6]Center for GeoGenetics, University of Copenhagen, Oester Voldgade 5-7 Copenhagen K 1350, Denmark

*Corresponding author: Department of Integrative Biology, University of California, Berkeley, 3040 Valley Life Sciences Building #3140. Berkeley, CA 94720-3140. Email: sashanikolaeva@berkeley.edu

Aneuploidy, a condition characterized by an abnormal number of chromosomes, can have significant consequences for fitness of an organism, often manifesting in reduced fertility and other developmental challenges. In plants, aneuploidy is particularly complex to study, especially in polyploid species such as coast redwood (*Sequoia sempervirens* (D. Don) Endl.), which is a hexaploid conifer ($2n = 6x = 66$). This study leverages a novel Markov Chain Monte Carlo method based on sequence depth to investigate the occurrence of aneuploidy across the range of coast redwood. We show that aneuploidy, defined here as a whole-chromosome gain or loss, is prevalent in second-growth redwoods, predominantly as additional chromosomes, while vegetatively propagated plants frequently experience chromosome loss. Although our study does not directly assess the fitness of aneuploids, the frequency of chromosomal instability observed in vegetatively propagated plants compared to second-growth and old-growth trees raises questions about their long-term developmental viability and potential to become established trees. These findings have significant implications for redwood conservation and restoration strategies, especially as methods such as tissue culture propagation becomes the primary mode of producing nursery stock plants used in reforestation.

Keywords: polyploidy; aneuploidy; tissue culture; coast redwood; *Sequoia sempervirens*

Introduction

Aneuploidy—or a change in the base number (n) of homeologous chromosomes (Li and Zhu 2022)—is generally considered deleterious, as an unbalanced chromosome dosage can lead to the loss of fitness of an organism. In many diploid organisms, including humans, aneuploidy is linked to a shorter life span and can cause certain health symptoms. For example, trisomy 21 is a genetic condition responsible for Down syndrome in humans and is often associated with learning disabilities, congenital heart diseases, Alzheimer's diseases, leukemia, cancers, etc. (Asim *et al.* 2015).

While the causes and consequences of aneuploidy have been reasonably well-studied in humans and other animals, less research has focused on plants (Li *et al.* 2021). Some plants appear to tolerate aneuploidy better than others, especially polyploid plants that have not undergone the diploidization process (Birchler 2013). For example, in hexaploid wheat (*Triticum aestivum*), changes in the copy number of a homeologous set of chromosomes (an extra or a missing chromosome 4B and 7A) did not significantly affect the expression levels of that set as a whole (Zeng *et al.* 2020). In contrast, in a population of synthetic tetraploid *Arabidopsis thaliana*, the presence of an extra chromosome 5 led to significant detrimental effects, including altered gene expression on the trisomic chromosome, changes in gene expression on other chromosomes, and overall genome instability (Huettel *et al.* 2008). Laboratory experiments with self-pollinated allopolyploids like *Brassica napus* have shown that aneuploids are selected against, as gene expression becomes imbalanced (Xiong *et al.* 2011).

Key causes of aneuploidy in better-studied diploid genomes include incorrect attachments between chromosomes and spindle fibers, which can go undetected by the cell's checkpoint systems. Extra centrosomes can create abnormal spindle shapes, leading to mis-segregation during meiosis or mitosis (Compton 2011). Meiotic aneuploidy most often arises from nondisjunction during chromosome segregation in gametes due to errors like premature chromatid separation (Fuchs *et al.* 2018), while mitotic aneuploidy occurs in somatic cells due to spindle assembly failures, often influenced by chemicals that disrupt microtubule dynamics (Sharma 1990). Age-related defects in chromosome cohesion also contribute, as do stresses on DNA replication and repair processes. Factors like oxidative stress and mechanical stress during cell division further disrupt chromosome segregation (Li and Zhu 2022).

Environmental factors are known to influence levels of aneuploidy in natural populations. On a cellular level, for example,

Received on 08 December 2024; accepted on 10 March 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of The Genetics Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

water stress has been demonstrated to cause meiotic chromosome abnormalities in rice, barley, and other agricultural crops, often reducing fertility of male plants (Fuchs et al. 2018). Given this effect, it is therefore not surprising that at the population level aneuploidy has been also found to occur at the extremes of nonagricultural species ranges at the edges of preferred climatic conditions, where the species is subject to increased environmental pressures. For example, in diploid Scots pine (*Pinus silvestris*) at the edges of its natural range in Khakassia, Sedel-nikova et al. (2023) found a wide range of genomic and chromosomal rearrangements, including aneuploidy and mixoploidy, which they attributed to a combination of factors such as dry and nutrient-poor conditions typical of mountainous, gravelly stepped landscapes, as well as reproductive isolation of the population.

Coast redwood (*Sequoia sempervirens* (D. Don) Endl.,) is a long-lived conifer restricted to the coastal fog belt from southern Oregon to central California. It reproduces mainly asexually, but seed reproduction also has an important role in population dynamics (Jameson and Robards 2007). The species is hexaploid (Hirayoshi and Nakamura 1943; Stebbins 1948; Ahuja and Neale 2002) and a tentative autopolyploid (Scott et al. 2016), although earlier studies indicate partial allopolyploidy (Stebbins 1948). The degree of sequence differentiation among all six copies of redwood chromosomes has not been extensively studied and the number of chromosomes per gamete (gametophyte) is also not known. In a distantly related species of cypress (*Cupressus sempervirens*), megagametophytes have been found to contain "an even and odd series of DNA contents: 1C, 2C, 3C, 4C, 5C etc., where C is the amount of DNA in the haploid genome" (El Maâtaoui and Pichot 1999).

The examination of allozyme inheritance patterns conducted by Rogers (Rogers 1997) indicates that hexasomic inheritance, in which each combination of homologous chromosomes is equally likely to be formed in a gamete, is largely preserved in redwood. These results exclude strictly disomic inheritance, where homologous chromosomes always pair in a particular configuration, making some combinations impossible. However, multisomic inheritance does not rule out the possibility of bivalent pairing configurations in an autopolyploids (Qu et al. 1998; Li et al. 2021), and that might also be true for coast redwood. A later study, examining the chromosome segregation patterns in meiosis (Hizume et al. 2014) showed that many of the redwood chromosomes do pair as bivalents, although full multivalent meiotic rings and other chromosomal configurations, including monovalents, trivalents, and tetravalents were also present. It is possible that certain redwood chromosomes have started to diverge in bivalent pairs, but this process is slow and still allows for multivalent pairing.

The inconsistent meiotic behavior could help explain low (up to 15%) seed germination in coast redwood (Roy 1966). As Stebbins pointed out in a seminal paper on polyploid plants, (Stebbins 1947), irregular chromosomal configurations often lead to sterility and formation of aneuploids. Later studies (Crowley and Rees 1968) suggested that formation of univalents and trivalents is often responsible for low fertility as such gametes often result in an unbalanced number of chromosomes in a cell. More recent studies confirmed this suggestion in *Brassica napus* plants, in which seed yield and pollen viability were inversely correlated with increasing aneuploidy (Xiong et al. 2011). Interestingly, seed viability in redwood tends to increase with age of the stand where the seed is collected, with the maximum viability reached when trees are over 250 years, before tapering off and decreasing at over 1,200 years of age, with some exceptions (Metcalf 1924). The extent to which seed viability is affected by imperfect segregation in meiosis and the resulting aneuploidy remains an open question.

In this research, we evaluated cases of aneuploidy across the geographic range and temporal scale of redwoods. We utilized short-read sequencing to identify aneuploidy in both old-growth and second-growth trees and compared these patterns with those observed in plants propagated vegetatively in a greenhouse (DLT dataset for "De La Torre," De La Torre et al. 2021). The primary challenge in this research lies in accommodating technical variance in sequencing depth among individuals and chromosomes. While this is relatively straightforward and standard practice in organisms with well-described and curated genomes, such as humans and model organisms, it becomes significantly more complex in poorly understood hexaploid genomes. To address these challenges, we developed a new Markov Chain Monte Carlo (MCMC) method for inferring variation in chromosome number from population samples of nonmodel organisms. This approach allows us to accurately detect and analyze aneuploidy despite the inherent difficulties posed by the unique and understudied coast redwood genome.

Materials and methods

Sample collection

To investigate aneuploidy across the range of coast redwood, we used a paired study design, in which pairs of populations are selected in such a way that the populations are geographically close but experience substantially different selective environments (Lotterhos and Whitlock 2015).

Sampling was conducted in unmanaged old-growth and redwood second-growth forests defined using the LEMMA forest structure dataset provided by the Save the Redwoods League (Cowan et al. 2017). The history of how the sampling site was managed was verified through consultations with landowners. Although we did not have an accurate estimate of the age of the trees—it is a notoriously difficult task to estimate the age of redwoods (Waring and O'Hara 2006)—we assessed based on management history that the old-growth forests were over 300 years old (conservative estimate) and the second-growth trees were between 40–100 years old.

For each location, we collected between one and five foliage samples. The trees were located at a minimum of 60 m from each other as this was the maximum distance between the clones reported in the most recent study on redwood clonality (Narayan 2015) to minimize our chances of sampling clonal trees in one location. The foliage was collected from the lowest branches of established trees or epicormic or basal sprouts when the lowest branches were not reachable, which was often the case in old-growth stands. In some locations, the so-called "sun foliage," or the foliage that typically grows near the tree tops and has different phenotypic characteristics, was also collected from the ground. The location of each sample was recorded on Avenza maps (Avenza Maps™ v5.1.1).

Samples then were placed on ice and transported to the University of California, Berkeley campus within 2 days. Upon arrival, they were immediately stored in a −80°C freezer. In total, samples from 305 trees were collected (Fig. 1).

We also utilized a second coast redwood exome dataset described in (De La Torre et al. 2021), we refer to it here as DLT dataset, by the initials of the first author. The dataset included 82 samples, originating from vegetatively propagated plants collected as a part of the previous common garden experiment at

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

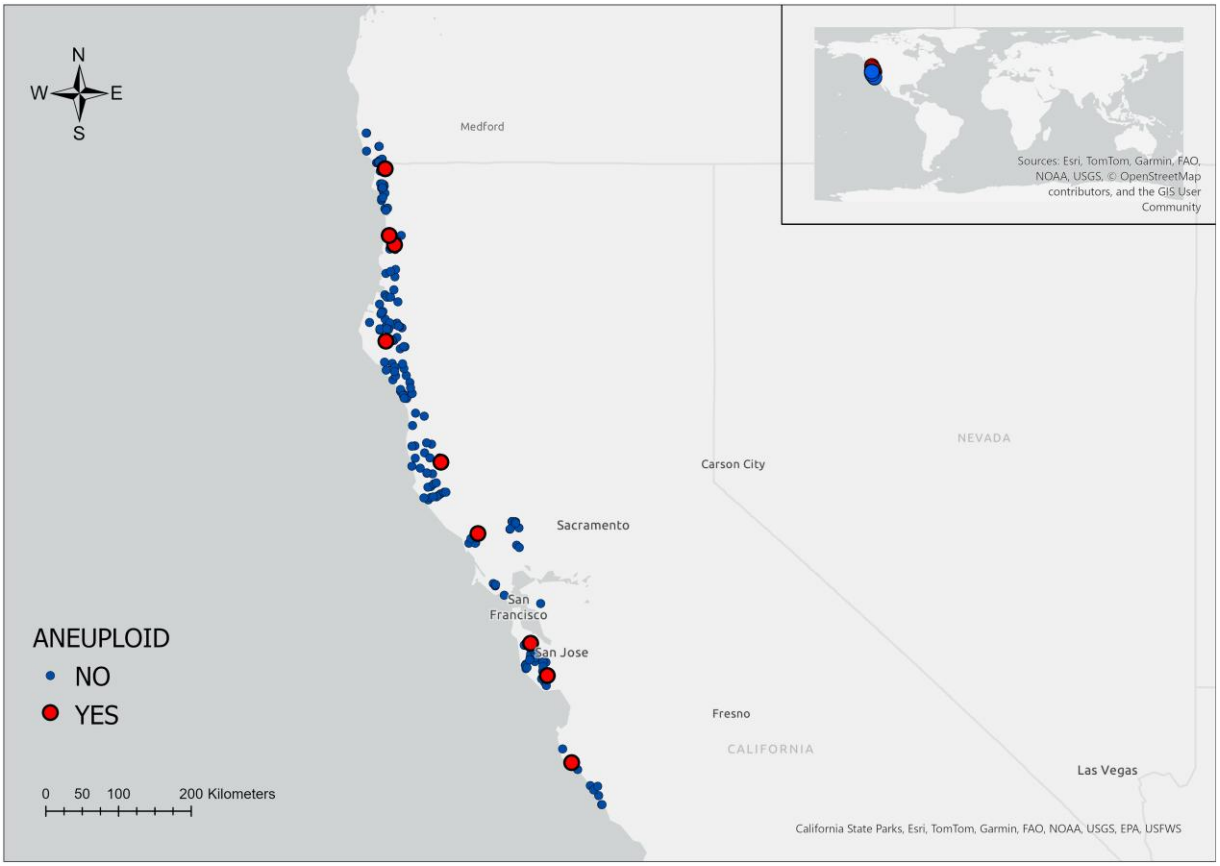


Fig. 1. Sampling distribution across coast redwood range.

the Russell Research Station (Kuser *et al.* 1995). The cuttings from the common garden were propagated in a greenhouse and the foliage material was sequenced from the cuttings at the University of California, Davis.

DNA extraction and sequencing

DNA was extracted from leaf tissue using a modified CTAB protocol (Doyle 1990) with changes made to the tissue preparation and DNA purification steps. Briefly, modifications to the protocol included a chloroform prewash applied to the homogenized tissue to remove the secondary metabolite compounds and a second ethanol wash. We also performed a magnetic bead clean-up using Solid Phase Reversible Immobilization (SPRI) Beads using a modified protocol from (Meyer and Kircher 2010). DNA was then resuspended in purified water. The concentration and purity of DNA were assessed on a SpectraMax M2 plate reader using the Biotium Accuclear High-Sensitivity Kit.

A portion of each extraction was diluted to 10 ng/μL at a volume of 110 μL using 10 mM Tris elution buffer, pH 8. The aliquot was sonicated on a qSonica Q800R sonicator at 40% amplitude and 15 s on/off pulse for 5 min of active sonication time. A double-sided SPRI bead cleaning process was used to size-select fragmented DNA to ~300–500 bp and to concentrate to a volume of 12.5 μL. The enzymatic steps of library preparation followed a modified Kapa Hyper Prep (Roche Diagnostics) protocol. After end repair and a-tailing, a universal stub adapter was ligated, and then was extended to full length during amplification with TruSeq-style unique dual-indexing oligos provided by the Functional Genomics Laboratory (University of California, Berkeley). After the final cleaning, libraries were eluted in water. Samples were assessed for sizing on an agarose gel or Bioanalyzer DNA 1,000 chip (Agilent Technologies) and quantified using the Biotium Accuclear High-Sensitivity Kit.

Libraries were then combined into 36 pools of 8 libraries per each pool. One thousand nanograms of library was used as input, such that each capture pool contained 8 μg. Capture hybridizations were completed in sets of 4–8 captures at a time using the Twist Target Enrichment Standard Hybridization v2 Protocol and Kit. Manufacturer's protocol was followed, except in addition to the standard blocking elements of the Twist kit, we also added additional adapter and indexing oligo blockers provided by Roche to compensate for the additional library material being used. (13.4 μL of Universal Blocking Oligos and 50 μL of Kapa Enhancer Reagent per capture reaction.) After captures, pools were split in half for enrichment PCR: the first half of the pool was amplified with 6–8 cycles of amplification depending on what number had worked well with previous captures. Then, it was cleaned with SPRI beads and assessed on a Qubit v.2 Fluorometer. If the result was high, the cycle number was lowered for the second amplification reaction; if low, it was raised. After the second cleaning, all pools were combined and assessed on a Qubit. Final concentrations of the enriched capture pools ranged from 4.26 to 25.2 ng/μL (median of 14.6 ng/μL; average of 14.2 ng/μL). At the Vincent J. Coates Genomics Sequencing Lab, all 36 captures were pooled together in equimolar amounts based on qPCR assessment using Kapa Biosystems Illumina standards (QB3 Genomics, UC Berkeley, Berkeley, CA, USA RRID:SCR_022170). This final pool of all captured libraries was sequenced across 4 lanes of Illumina NovaSeq X 10B to collect paired-end 150 bp data. (Sequencing was performed at the UCSF CAT, supported by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants.)

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

The total size of target space was 17.7Mb. Targets were selected using available annotations for redwood genome (Neale *et al.* 2022), and filtered to 60% identity using sequences that had an alignment to the custom conifer database, NCBI's Plant RefSeq, or UniProt database. This selection criterion was chosen to enrich for the most conserved sequences across the target exome.

Out of 305 collected samples, 285 were sequenced. One sample had insufficient sequencing depth and was excluded from analysis. Two samples were technical duplicates that were also removed. We also excluded a number of samples (8) that came from a suspected clonal group of trees, for the total of 274 samples we used in the analysis.

Reference genomes

For this analysis, we used two existing reference genomes. The first, published by The Redwood Genome Project (RGP) (Neale *et al.* 2022), is ~26.5 gigabases, reflecting the triploid size of the genome. There is little synteny of the RGP assembly to the genome assembly of sister species of giant sequoia (*Sequoiadendron giganteum*) (Fu *et al.* 2023). Biologically, this lack of synteny is highly unlikely, given that giant sequoia's genome is syntenic with both metasequoia genome (Metasequoia glyptostroboides) (Fu *et al.* 2023) and Japanese cedar (*Cryptomeria japonica*). We assumed that the lack of synteny between giant sequoia and coast redwood genomes is likely caused by an imperfect genome assembly.

The second, known as the PacBio redwood genome, was sequenced in 2020 using 33-fold long HiFi reads and is publicly available (Hatas *et al.* 2020). We utilized both the primary assembly of the PacBio genome (48.5 Gbp), and the full 51 Gbp of assembled unitigs. It is important to note that both reference genomes were assembled with tools primarily designed for phasing diploid genomes, and neither of the genomes are haplotype-resolved or chromosome level. The RGP genome was assembled using the MaSuRCA assembler (Zimin *et al.* 2013) and the HiRise scaffolder (Koch 2016), whereas the PacBio genome was assembled with the HiFiasm assembler (Cheng *et al.* 2021).

Confirming ploidy of the reference genomes

Given the hexaploid nature of the genome and the possibility of some gene copies missing in the reference genomes, leading to incorrect estimations of gene and chromosome loss, it was necessary to confirm that the annotation sequences we used in designing exome probes were found in exactly six copies in the reference. To confirm that, we used the set of exome sequences (CDS) from the annotation of the RGP triploid reference as a query input to the BLAST program (blastn v.2.9.0-2) (Altschul *et al.* 1990) against the PacBio hexaploid reference: We performed a BLASTn search using the PacBio reference genome (ssempervirens.p_utg.fa) as the database and a query dataset (55 K.fa). The search utilized 10 threads, and the output was formatted as a tab-delimited table including the following fields: query sequence ID, subject sequence ID, percent identity, alignment length, number of mismatches, number of gap openings, query start and end positions, subject start and end positions, *e*-value, bit score, and number of identical matches. The BLAST output was then filtered using custom R-scripts (GITHUB). Briefly, this code is used to calculate the frequency of hits for each query sequence. We create a data frame, `query_frequencies`, which represents the distribution of hit frequencies for each query sequence in a BLAST analysis. It is created by counting how many times each unique hit frequency (number of hits per query sequence) occurs. For example, if a frequency of 2 hits is observed for 10 query sequences, `query_frequencies` will have a row with `Var1 = 2` and `Freq = 10`. Each frequency is then shown as a bar in a histogram plot. We repeated this analysis for the RGP genome and for the giant sequoia reference genome (Scott *et al.* 2020). We also aligned giant sequoia reference genome to a related species of Japanese cedar (*Cryptomeria japonica*), using minimap2 (Li 2018) with the `-x asm10` preset to align two reference genomes, replicating the results described in (Fu *et al.* 2023).

Normalization of reads depth

Raw reads were aligned to the closely related but diploid species of giant sequoia reference genome (Scott *et al.* 2020) using BWA (v. 0.7.17-r1188). Giant sequoia's genome has a high level of synteny to cryptomeria's genome (Fig. 3a) (Fu *et al.* 2023), and this gave us higher confidence in the quality of the giant sequoia's assembly and therefore, the success of our approach. Alignments were then deduplicated, collated, and sorted with samtools (v.1.17) (Li *et al.* 2009). Reads with a minimum quality score of 60 (MQ ≥ 60) were selected, and the idxstats tool from samtools was used to calculate the count of mapped reads to each chromosome.

To visualize chromosome specific differences in sequencing dosage, we computed a read depth for each individual sample i for each chromosome j, standardized by the total read depth of the sample and the total read depth for the chromosome over all samples:

$$\alpha_{ij} = \frac{n_{ij} \times N}{\sum_k n_{ik} \times \sum_l n_{ij}},$$

where n_{ij} is the read count for sample i on chromosome j and $N = \sum_i \sum_j n_{ij}$ is the total read count across all samples and chromosomes. A value of $\alpha_{ij} > 1$ (or $\alpha_{ij} < 1$) indicates that sample i has more (or fewer) reads on chromosome j than expected given the total number of reads for sample i and the total number of reads across all samples for chromosome j.

Taking into account the possibility of varying read depths due to differences in the number of amplification cycles, we stratified the dataset into five groups according to the total number of PCR cycles (Fig. A1). Then, α was calculated separately for each group (Appendix A).

Bayesian model for determining ploidy per chromosome

We developed a Bayesian statistical framework to infer the number of chromosomes in coast redwood.

Prior distributions

The model assumes a uniform Dirichlet prior for the expected proportion of reads assigned to each chromosome: $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, which can be interpreted as the length of the mapping target on each chromosome under the assumption of constant mapping probability along the length of the genome:

$$\lambda \sim \text{Dirichlet}(\mathbf{1}),$$

where $\mathbf{1}$ denotes a vector of ones, and m is the number of chromosomes.

The prior for the ploidy level (k_{ij}) for individual i and chromosome j, is a reflected and truncated geometric prior distribution, centered around the expected ploidy of 6:

$$P(k_{ij}) = \frac{(1-p)^{|k-6|}}{\sum_{i=0}^{12}(1-p)^{|i-6|}}, \quad 0 \le k \le 12.$$

We assign a uniform[0, 1] hyperprior to p.

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

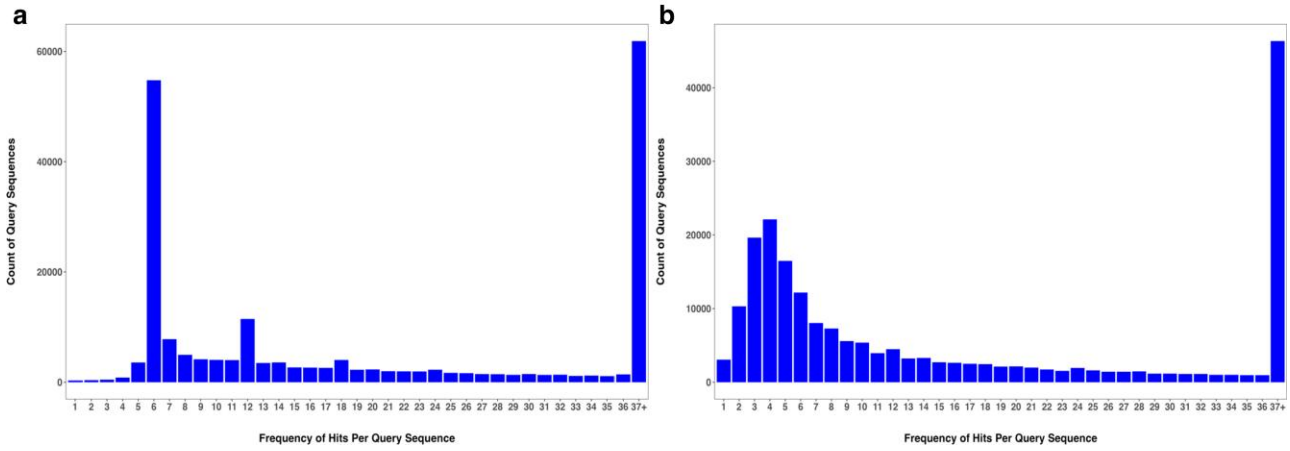


Fig. 2. BLAST query frequency distribution, a) PacBio reference genome, b) RGP reference genome.

Likelihood function

The likelihood function is defined by a multinomial distribution, which models the read counts across different chromosomes for each individual as:

$$L(\lambda, k \mid \text{data}) = \prod_{i=1}^{n}\prod_{j=1}^{m}\left(\frac{\lambda_j k_{ij}}{\sum_{\upsilon=1}^{m}\lambda_\upsilon k_{i\upsilon}}\right)^{n_{ij}},$$

where n is total number of samples.

MCMC algorithm

To estimate the posterior distributions of the parameters, we use a Metropolis–Hastings Markov Chain Monte Carlo (MCMC) approach. The MCMC algorithm iteratively updates the parameters λ, k, and p through a series of steps designed to explore the parameter space.

Updating kernels

***Updates of* λ.** New values of λ, $\lambda\prime$, are proposed using a reflected exponential distribution, ensuring the new values remain within the permissible range, [0,1]. The update rule is given by:

$$\lambda_j' = \text{Reflect}(\lambda_j + \text{Exponential}(rate)),$$

where the Reflect function iteratively sets $x \leftarrow -x$ if $x < 0$ or $x \leftarrow 2 - x$ if $x > 1$ until $0 \le x \le 1$. This ensures that the proposed value of $\lambda\prime_j$ is between 0 and 1 and maintains symmetry of the update kernel. Other values of λ are then updated as

$$\lambda_i' = \lambda_i \frac{(\lambda_j' - 1.0)}{(\lambda_j - 1.0)},\ i \neq j.$$

Because of the symmetric proposal kernel and the uniform Dirichlet prior, only the likelihood enters into the Metropolis–Hastings ratio for this update. The value of the *rate* used in our analyses is 100.0 corresponding to an exponential with mean 0.01.

***Update of* k.** The ploidy levels, $\{k_{ij}\}$, are updated independently for each chromosome and each individuals using a simple symmetric random walk, on a circle of integers {0, 1, …, 12}, where 0 and 12 are connected states such that symmetry of the updates are preserved. The Metropolis–Hastings ratio then includes the likelihoods and the prior, but not the symmetric update probabilities.

***Update of* p.** The parameter p is updated using a reflected exponential, with mean 0.01 (rate 100,0), similar to the one used for updates of λ. Because $p \sim U[0, 1]$ and the proposal kernel is symmetric, only the likelihood appears in the Metropolis–Hastings ratio.

MCMC runs. The algorithm cycles between updating all $\{k_{ij}\}$, all $\{\lambda_j\}$, and p and runs for a predefined number of iterations (100,000), using the first 10,000 as a burn-in. Convergence is assessed by running multiple chains with different starting point and evaluating the variance of the parameters across chains and within chains, and the auto-correlation of the sampled values across iterations, Fig. A2).

The validity of the implementation was tested by ensuring that the prior distribution was recovered as the posterior when running the program without data and by comparing the likelihood calculations to an independent implementation in Mathematica. A program written in C implementing the algorithm is available from Github.

Results

Ploidy of the reference genomes

Of the 207,167 total sequences blasted, 54,771 sequences were found in sets of 6 (328,626 qseqid hits) (see definition of a set in the Methods section), reflecting the hexaploid nature of the genome (Fig. 2a) in the PacBio full genome. After filtering for duplicates, the resulting number of sequences in the reference set was 39,840.

However, most CDS sequences were found in high copy numbers, reaching up to 23,645 copies in one homologous set. This observation is consistent with previous findings that the coast redwood genome is rich in repetitive elements, which cover about 70% of the genome (Neale *et al.* 2022). The average identity (pident value from the BLAST results table) between sequence copies found in sets of six was 99.36%, with a minimum value of 72.13% and a maximum value of 100%.

Another minor frequency peak was observed at 12 sequences per set, with 11,475 sets in total (Fig. 2a). From looking at the sequence identity of these sequences, it appeared that they could be split into groups of six—one group of six sequences among which the mean identity was comparable with the previous group (99.38%), and another group where the mean value of the pident

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

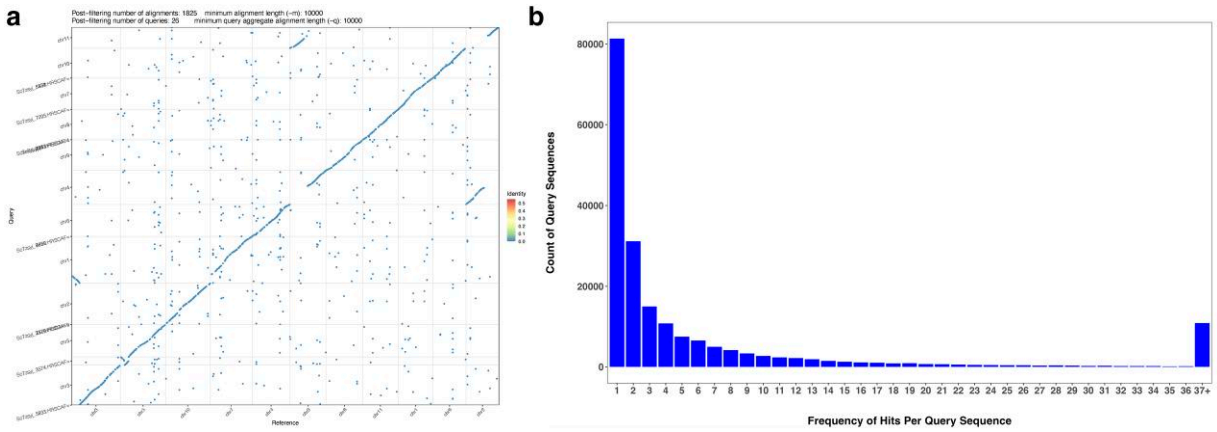


Fig. 3. Validation of giant sequoia (*Sequoiadendron giganteum*) genome assembly, a) minimap2 alignment of giant sequoia genome (query) to Japanese cedar (*Cryptomeria japonica*) (reference), b) giant sequoia BLAST query frequency distribution.

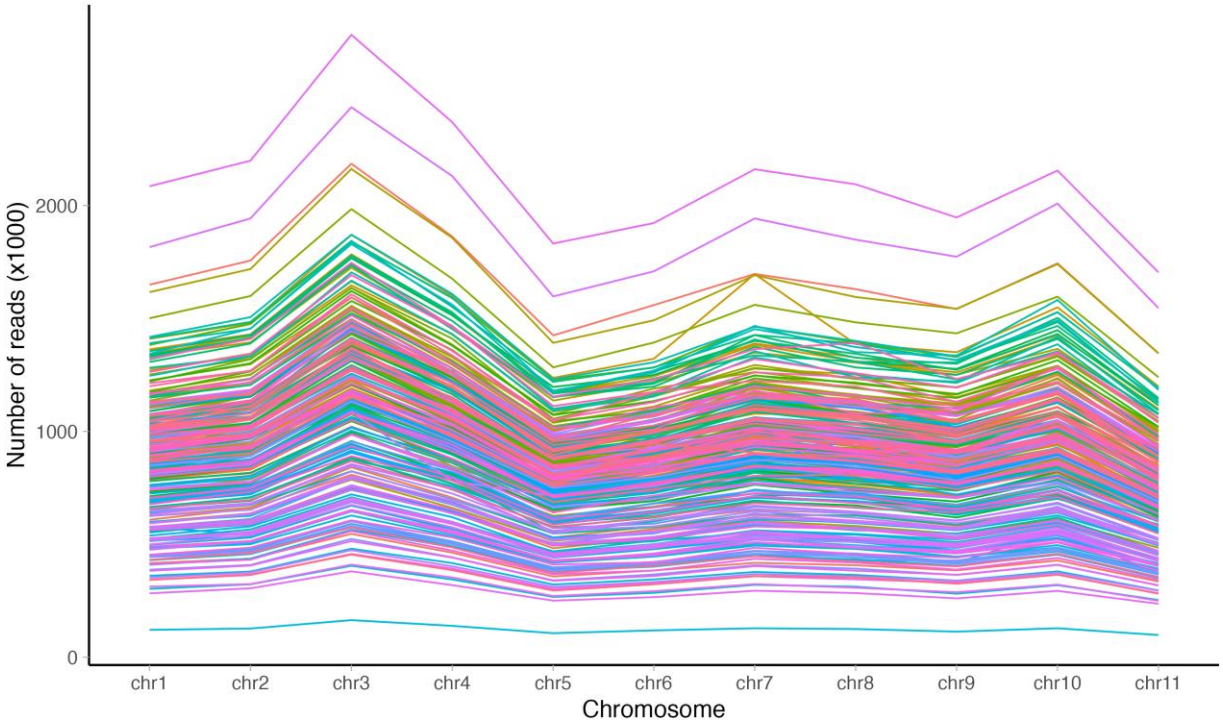


Fig. 4. Raw read depth in coast redwood (*Sequoia sempervirens*) in relation to giant sequoia (*Sequoiadendron giganteum*) chromosomes. Each line represents an individual sample.

was lower (95.63%). These sequences are possibly duplicates from an older whole genome duplication event.

Repeating this analysis for the RGP genome, we found that most often CDS sequences were found in sets of four (22,105 unique sequences), followed by sets of three (19,638 sequences), indicating that this reference genome might not be triploid for all coding sequences, as expected (Fig. 2b). This finding disagrees with the results of the k-mer analysis of the RGP genome using GenomeScope 2.0 software (Ranallo-Benavidez *et al.* 2020), which indicated triploidy of the reference.

As a validation step, we also performed this analysis for the reference genome of giant sequoia, and found a frequency peak at 1 sequence, with 81,364 total sets (Fig. 3b).

Ploidy per chromosome

The total number of reads (raw depth) aligning to giant sequoia (*Sequoiadendron giganteum*) chromosomes is shown in Fig. 4, with the average of 9.13×10^2 reads aligning per chromosome.

There were variations in α on all chromosomes except for chr2 and chr5, characterized by either an increase or a decrease in α in the expected number of reads, indicating a deviation from the hexaploid chromosome number (Fig. 5). To statistically assess these deviations and estimate the ploidy of each chromosome within each sample, we applied the MCMC algorithm described above.

Additionally, to address the observed variance in α, we also ran the MCMC algorithm separately for each subgroup analysis based

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

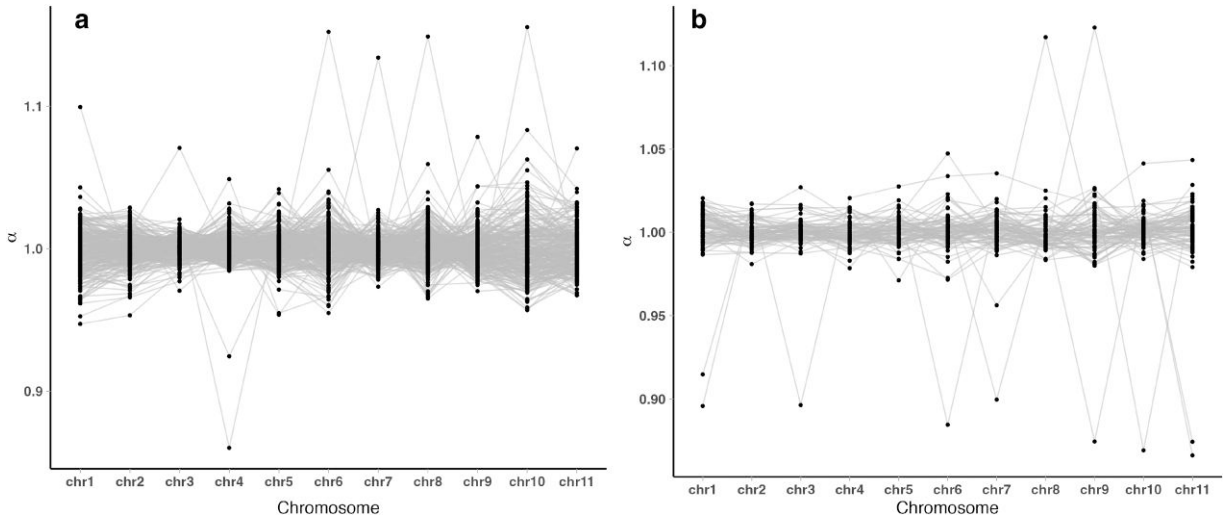


Fig. 5. a) Read depth normalization (α) in coast redwood (*Sequoia sempervirens*) in relation to giant sequoia (*Sequoiadendron giganteum*) chromosomes, this study; b) Read depth normalization (α) in coast redwood (*Sequoia sempervirens*) in relation to giant sequoia (*Sequoiadendron giganteum*) chromosomes, DLT dataset (De La Torre *et al.* 2021). Each line represents an individual sample.

Table 1. Aneuploidy counts across all samples from this study and the DLT dataset, (De La Torre *et al.* 2021).

	This study (n = 274)		DLT dataset (n = 82)	
Chromosome	**Ploidy 5**	**Ploidy 7**	**Ploidy 5**	**Ploidy 7**
Chr1	0	1	2	0
Chr2	0	0	0	0
Chr3	0	1	1	0
Chr4	1	0	0	0
Chr5	0	0	0	0
Chr6	0	1	1	0
Chr7	0	1	1	0
Chr8	0	1	0	1
Chr9	0	1	1	1
Chr10	0	2	1	0
Chr11	0	0	2	0

on the number of PCR amplification cycles for each sample. This allowed us to assess whether variations in amplification cycles might have influenced chromosome counts inferred by the MCMC. However, stratification by PCR cycle groups did not change the results.

In total, aneuploidy was detected in 9 samples (out of 274) and 8 of these samples showed an increase in the number of chromosomes per set, with a ploidy of 7, while one sample had a loss of a chromosome (Table 2).

An increase on chromosome 10 was observed most frequently, with 2 samples affected. Other chromosomes with increases in ploidy number were 1, 3, 6, 7, 8 and 9. There was also a loss of chromosome for 1 sample on chromosome 4.

In the DLT dataset, we observed 11 samples (of 82) with aneuploidy present. Two of the samples had an increase of the chromosomal number to 7 (on chromosomes 8 and 9), and the rest had a loss of chromosome (chromosomes 1, 3, 6, 7, 9, 10 and 11) (Table 1).

The number of samples with an increase in the chromosomal number per set (ploidy of 7) was comparable between the two datasets (Fisher's exact test $p \approx 0.535$). The number of samples with a decrease or loss of chromosome is significantly higher in the DLT dataset (Fisher's exact test $p \approx 1.039 \times 10^{-5}$).

Table 2. Aneuploidy counts by growth stage across all samples from this study and the DLT dataset (De La Torre *et al.* 2021).

	Old growth	Second growth	DLT dataset
Loss of chromosome	0	1	9
Gain of chromosome	0	8	2
Total aneuploids	0	9	11
No aneuploidy	56	209	71
Total trees	56	218	82
Percent aneuploids	0	4.3062	13.4146

Temporal and geographic distribution of aneuploid trees

The GPS coordinates were recorded for every tree. Instances of aneuploidy were found throughout the range of species (Fig. 1), and there was no geographic clustering of aneuploids. The only sample with a missing chromosome (ploidy of 5) was found in the southern part.

The stand age data were also collected in the field. There were 56 old-growth trees and 218 second-growth trees. Aneuploidy was present in second-growth trees (4.3%) and in the tissue-culture dataset (13.41%) (Table 2), but missing from the old-growth trees. However, there was no statistically significant difference between the number of aneuploids in old-growth and second-growth trees (one-tailed Fisher's exact test $P \approx 0.21$).

We were not able to identify the exact geographic locations of the samples in the DLT dataset, although the sampling distribution was range wide. All of the samples from the DLT were seedlings grown from vegetatively propagated cuttings (De La Torre *et al.* 2021).

Discussion

Detection of aneuploidy and study limitations

Recent studies investigating aneuploidy events in plants have found that aneuploids are typically located at the edges of a species' range, signaling a potential stress-related genome restructuring that could also be adaptive (Sedel-nikova *et al.* 2023). However, aneuploidy, its causes and especially its effects in plants

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

remains poorly studied, perhaps due to the laborious experimental and laboratory procedures necessary to identify aneuploid plants. Polyploid plants present additional challenges since detecting such chromosomal variation often involves extensive laboratory techniques such as measuring DNA content via flow cytometry (Pellicer *et al.* 2021) or QF-PCR (Henry *et al.* 2006). In autopolyploids, where chromosomes may form multivalent rings during meiosis, karyotyping alone may not be sufficient to detect aneuploidy. More labor-intensive methods, such as fluorescence in situ hybridization (FISH) with custom probes targeting individual chromosomes, may be necessary to detect aneuploidy.

In this study, we explored chromosome numbers in coast redwood populations across the range using a new computational method that utilizes sequencing depth data. However, this method might not be applicable to detecting partial aneuploids where only parts of chromosomes are added or missing. Mosaic aneuploidy, where only some of the cells might be aneuploid, would also be hard to detect using our method. Chimera trees with various levels of mosaicism have been recorded in coast redwood (Moore 2023), and more studies are needed to confirm the frequency of such mosaicism in redwood populations.

Additionally, our assumption was that all trees in the two datasets were hexaploid, which means our method cannot distinguish between a hexaploid and a tetraploid plant with aneuploidy.

Aneuploidy can arise from errors during meiosis and mitosis (Orr *et al.* 2015). When mechanisms ensuring correct segregation of chromosomes during these processes fail, the resulting cells might have an unbalanced number of chromosomes. Here, we defined aneuploidy as a whole-chromosome gain or loss. As more accurate chromosome-level reference genome for the species become available, it will open up more avenues for investigating other structural variations such as deletions, insertions, and translocations.

Reference set of CDS sequences

Our BLAST analysis is a relatively simple way to confirm the hexaploid nature of the coast redwood genome. It is notable that the most frequent category in which CDS sequences were found was 6, but we also found sequences in sets of 12. These sequences are possibly sequences that existed in two copies in the original diploid genome that then underwent two rounds of whole genome duplication, resulting in 12 copies of that gene. It is possible that these sequences date back to the genome duplication event that occurred at the base of all seed plants (Stull *et al.* 2021). The distribution of frequencies of CDS sequences in the gametophyte reference genome (RGP), however, did not show the expected pattern of 3 BLAST hits per sequence. Instead, we observed that sequences were most often present in copies of 4, followed by sets of 3. This could be explained biologically by the true presence of four copies of the genome in the gametophyte, but the more likely explanation is an imperfect genome assembly, where the assembler might have failed to collapse repetitive sequences properly, possibly due to a higher than expected differentiation between the homologous haplotypes. We also found that this disagrees with the results of the GenomeScope2.0 analysis that indicated the triploidy of the RGP reference. Similar disagreements have been previously reported for the baobab genome *Adansonia digitata* (Kitony *et al.* 2024), where GenomeScope2.0 suggested a diploid homozygous genome for a confirmed tetraploid, aligning with the caution from the GenomeScope2.0 authors (Ranallo-Benavidez *et al.* 2020) that their tool may underestimate ploidy levels beyond certain heterozygosity thresholds. One possible reason could be that, at higher levels of heterozygosity, k-mers become too divergent to be consistently recognized as matching pairs, potentially leading to an underestimation of ploidy. However, the precise factors contributing to this limitation remain unclear.

We chose to use the giant sequoia genome as a reference for our analysis because the majority of the CDS sequences were found only in one copy—as expected from a haploid representation of a diploid genome. Therefore, deviations in normalized read depth beyond what is expected for a hexaploid species would be indicative of the presence or absence of extra chromosome copies relative to the expected ploidy level.

Consequences of aneuploidy for tree fitness

Coast redwood is a long-lived species that employs a clonal mode of reproduction (Narayan 2015). It is common for polyploids to reproduce vegetatively and such a mode is often regarded as an escape strategy from the barriers to sexual reproduction (Comai 2005). This strategy might be successful for polyploids needing to occupy changing ecological niches during significant environmental perturbations (Van de Peer *et al.* 2021) and do so quickly, but it might come at a cost to individual plants that harbor substantial and detrimental structural genome changes (Charalambous *et al.* 2023).

In this complex ecological system, it is useful to distinguish between a tree's meiotic age (the time elapsed since a particular stem cell began its lineage through meiosis) and its developmental stage. The meiotic age does not necessarily correspond to the plant's developmental stage. Terms such as "old-growth," "second-growth," or "vegetative sprout" describe the developmental stage of an individual plant, regardless of whether it originated from a seed or a sprout. Thus, the developmental stage refers to the period since a plant tissue, whether a seed or a vegetative sprout, began developing into an independent organism. For example, in our dataset, an old-growth tree that originated from a seed might have matching developmental and meiotic ages, whereas one developed from a sprout would have a higher meiotic age. Likewise, although vegetative sprouts are considered younger from a developmental perspective, they might represent some of the oldest tissues in the dataset when viewed from a meiotic standpoint.

Our results suggest a pattern, in which there is more aneuploidy in the vegetatively propagated plants than in the second-growth stands. Also, we do not find any aneuploidy in the older cohorts of trees. Below we discuss several possible explanations for this observation.

First, it is possible that aneuploidy could accumulate neutrally over time in plant tissues. However, the observed absence of aneuploidy in old-growth trees cannot be explained by this process alone. If aneuploidy were merely a by-product of time, we would expect it to be prevalent or even increase in older trees, which we do not observe.

When aneuploid cells arise within a plant, competition among aneuploid and euploid cells may counter the spread of aneuploidy. A second possibility is, therefore, that trees initially have a high probability of being aneuploid, or high proportion of aneuploid cells, due to the bottleneck in cell population size that occur during the early stages of development. However, euploid cells may then in later developmental stages outcompete the aneuploid cells leading to reduced aneuploidy in older trees. Additionally, beyond cellular competition, there might be unknown cellular mechanisms that actively prevent aneuploid cells from becoming established. Similar mechanisms have been proposed recently in a study on the triploid aspen clone Pando (Pineau *et al.* 2024).

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

However, the simplest explanation for our observations is that selection against aneuploidy occurs at the whole-organism level. If trees tend to have lower fitness i.e. be less successful at competing, if they are aneuploid, it would explain why vegetatively propagated trees that face little competition have more aneuploidy than older trees that might have faced more competition and more environmental stress, and why old-growth trees similarly have less aneuploidy than second-growth trees. Aneuploidy can perhaps explain previously reported strong differences in plant performance for height and volume gains among redwood cultivars (Morrison *et al.* 2022), even though in our study we do not directly assess the fitness of plants.

Our study did not aim to determine the origins or causes of aneuploidy (meiotic vs. mitotic). Clonal plants do have an advantage as compared to the plants originating from seed because they do not need to put resources into root system development. After a tree is cut (or damaged, for example after a fire), it can actively resprout and re-establish itself. It is possible that resource availability protects clonal sprouts from the negative effects of dosage imbalance in the early stages of plant development, but as trees become mature and competition increases, the seed trees with euploid number of chromosomes might outcompete vegetative sprouts.

Management implications

A common goal in redwood restoration projects is the return of the old-growth forest structure (O'Hara *et al.* 2017; Dagley *et al.* 2018; Iberle *et al.* 2020) after the intensive harvesting of the previous century. However, to meet the old-growth structure objective, it is preferable that genetics of individual ramets are taken into account. It is possible that not all redwood clones might be able to reach the old-growth stages, due to the fitness differences between aneuploid and euploid trees.

Importantly, here we do not advocate for the systematic removal of aneuploid trees in conservation efforts to achieve the goal of old-growth structure. In certain cases, and in some redwood stands it might be best for long-term restoration goals to use multiple restoration strategies, including natural recovery (Russell *et al.* 2014). While aneuploidy in general likely has negative consequences for the fitness of an individual, it can also provide evolutionary flexibility by promoting genome and chromosome instability (CIN), facilitating cellular adaptation, and redistribution of resources within a cell (Liu *et al.* 2015; Millet *et al.* 2015; Simonetti *et al.* 2019). Additionally, while our results suggest strong fitness effects of aneuploidy, more studies are necessary to quantify those fitness effects, and specifically the phenotypic differences between euploids and aneuploids to understand the effects of aneuploidy on tree survival, seed viability etc. and evaluate the tree performance not only in the sense of the short-term growth but also its ability to survive in the changing climate.

The species is also a valuable timber resource and managing this resource might require different approaches than in restoration. Many of the redwood commercial stands are currently managed on short rotations of ~50 years (O'Hara *et al.* 2017), and in such stands persistence of the trees into the old growth states as well as seed viability might not be the main objective. After harvesting, such stands are also restored using the tissue-culture planting material. Precommercial thinning of such stands is often recommended to increase the stand volume increment. This is when avoiding planting aneuploids becomes especially important.

Caution should be factored into the decision-making process regarding which cultivars to plant in the field, given the prevalence of aneuploids with a missing chromosome among the vegetatively propagated plants. The initial performance of the seedlings should be taken into account, as the negative effects of aneuploidy are likely to manifest themselves in the very early stages of development. Another important factor to take into account is how many generations of the tissue-culture replication a particular clonal line has been through. It is not yet clear how aneuploidy propagates throughout clonal generations, but some examples from the literature indicate that aneuploidy remains and increases in plants propagated by selfing. In *Brassica napus*, for example, aneuploidy increased from 24% to 94% after only 10 generations (Xiong *et al.* 2011).

Identifying aneuploid trees either in greenhouse plants or in the second-growth wild stands is another outstanding question. There are methods to identify aneuploidy in human cells and it should not be difficult to adapt those methods to identification of aneuploids in plants. Quantitative PCR (qPCR) is a common approach that detects aneuploidy by amplifying DNA sequences from targeted chromosomes and quantifying them in real time, comparing the results against a reference chromosome to determine if extra or missing copies are present. More comprehensive methods include fluorescence in situ hybridization, where fluorescent probes specific to chromosomes are hybridized to cell nuclei, allowing for direct visualization and counting of chromosomes under a fluorescence microscope. However, many of these methods require chromosome-specific probes, which are not yet available for redwood.

Conclusions

This study aimed to evaluate instances of aneuploidy in coast redwood, assuming a baseline ploidy of six. Our analysis shows that coding sequence are found in sets of six and they are the most abundant category, reflecting the genome's hexaploid nature. At the population level, we observe structural instability due to aneuploidy, with extra chromosomes more common than chromosome loss. Unique among forest trees, coast redwood is a rare hexaploid and autopolyploid conifer (Scott *et al.* 2016) that reproduces vegetatively and can live for thousands of years, making structure genome errors costly. Aneuploidy is present in second-growth populations, where extra chromosomes are more common than a chromosome loss, whereas vegetatively propagated plants mainly exhibit missing-chromosome aneuploidy. These findings have significant implications for coast redwood restoration and management.

Data availability

The raw sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive (BioProject: PRJNA1163354). The data are publicly accessible and can be freely downloaded for academic and research purposes.

Acknowledgments

The authors thank undergraduate researchers Liam Galleher, Claire Whicker, Simone Stevens, Nic Dutch, and Jenifer Camarena for their assistance in sample collection and DNA extraction. We also thank Michelle Davila for help in library preparation. We are also grateful to participating land managers who provided access to sampling locations, including those at

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

The Napa Valley Reserve (Paul Asmuth), California State Parks, The California Department of Forestry and Fire Protection(CalFire) and especially Lynn Webb, Green Diamond Resource Company (Carlos Gantz and Scott Whittington), Mendocino Redwood Company, Humboldt Redwood Company, and The Lyme Timber Company.

Funding

This work was supported by Save-the-Redwoods League direct grant 168 given to Alexandra Nikolaeva, as well as a Continuing Fellowship Student Award, a Researcher Starter Grant, and The Hannah M. and Frank Schwabacher Memorial Scholarship from the Department of Environmental Science, Policy, and Management at the University of California, Berkeley. JSS was supported by a postdoctoral fellowship from the Miller Institute for Basic Research in Science, University of California, Berkeley.

Conflicts of interest

The author(s) declare no conflicts of interest.

Author contributions

Conceptualization: ASN, RD, and RN. Methodology: ASN, RD, RN, LS, and JS. Data collection: ASN. Data analysis: ASN, RN, and JS. Writing—original draft: ASN, RN, and LS. Writing—review & editing: ASN, RD, RN, LS, and JS. Funding acquisition: ASN, RD, and RN.

Literature cited

Ahuja MR, Neale DB. 2002. Origins of polyploidy in coast redwood (*Sequoia sempervirens* (D. DON) ENDL.) and relationship of coast redwood to other genera of Taxodiaceae. Silvae Genet. 51(2-3): 93–99.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Asim A, Kumar A, Muthuswamy S, Jain S, Agarwal S. 2015. Down syndrome: an insight of the disease. J Biomed Sci. 22(1):41. https://doi.org/10.1186/s12929-015-0138-y

Birchler JA. 2013. Aneuploidy in plants and flies: the origin of studies of genomic imbalance.

Charalambous C, Webster A, Schuh M. 2023. Aneuploidy in mammalian oocytes and the impact of maternal ageing. Nat Rev Mol Cell Biol. 24(1):27–44. https://doi.org/10.1038/s41580-022-00517-3

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 18(2):170–175. https://doi.org/10.1038/s41592-020-01056-5

Comai L. 2005. The advantages and disadvantages of being polyploid.

Compton DA. 2011. Mechanisms of aneuploidy. Curr Opin Cell Biol. 23(1):109–113. https://doi.org/10.1016/j.ceb.2010.08.007

Cowan P, Burns EE, Campbell R. 2017. A GIS approach to identifying the distribution and structure of coast redwood across its range. In: Standiford RB, Valachovic Y, editors. Coast Redwood Science Symposium—2016: Past Successes and Future Direction. Gen. Tech. Rep. PSW-GTR-258. Albany (CA): U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station. p. 357–359.

Crowley JG, Rees H. 1968. Fertility and selection in tetraploid Lolium. Chromosoma. 24(3):300–308. https://doi.org/10.1007/BF00336197

Dagley CM, Berrill J, Leonard LP, Kim YG. 2018. Restoration thinning enhances growth and diversity in mixed redwood/Douglas—fir stands in Northern California, U.S.A. Restor Ecol. 26(6): 1170–1179. https://doi.org/10.1111/rec.2018.26.issue-6

De La Torre A, Sekhwal M, Neale D. 2021. Selective sweeps and polygenic adaptation drive local adaptation along moisture and temperature gradients in natural populations of coast redwood and giant sequoia. Genes (Basel). 12(11):1826. https://doi.org/10.3390/genes12111826

Doyle J. 1990. Isolation of plant DNA from fresh tissue. Focus. 12: 13–15.

El Maâtaoui M, Pichot C. 1999. Nuclear and cell fusion cause polyploidy in the megagametophyte of common cypress, *Cupressus sempervirens* L. Planta. 208(3):345–351. https://doi.org/10.1007/s004250050568

Fu F, Song C, Wen C, Yang L, Guo Y, Yang X, Shu Z, Li X, Feng Y, Liu B, *et al.* 2023. The Metasequoia genome and evolutionary relationships among redwoods. Plant Commun. 4(6):100643. https://doi.org/10.1016/j.xplc.2023.100643

Fuchs LK, Jenkins G, Phillips DW. 2018. Anthropogenic impacts on meiosis in plants. Front Plant Sci. 9:1429. https://doi.org/10.3389/fpls.2018.01429

Hatas E, Young G, Vierra M. 2020. Go big or go home—tackling a giant genome.

Henry IM, Dilkes BP, Comai L. 2006. Molecular karyotyping and aneuploidy detection in *Arabidopsis thaliana* using quantitative fluorescent polymerase chain reaction. Plant J. 48(2):307–319. https://doi.org/10.1111/tpj.2006.48.issue-2

Hirayoshi I, Nakamura Y. 1943. Chromosome number of *Sequoia sempervirens*. Bot Zool. 2:73–75.

Hizume M, Kaneko K, Miyake T. 2014. A method for preparation of meiotic chromosomes of conifers and its applications. Chr Bot. 9(3):83–88. https://doi.org/10.3199/iscb.9.83

Huettel B, Kreil DP, Matzke M, Matzke AJM. 2008. Effects of aneuploidy on genome structure, expression, and interphase organization in *Arabidopsis thaliana*. PLoS Genet. 4(10):e1000226. https://doi.org/10.1371/journal.pgen.1000226

Iberle BG, Van Pelt R, Sillett SC. 2020. Development of mature second-growth *Sequoia sempervirens* forests. For Ecol Manage. 459:117816. https://doi.org/10.1016/j.foreco.2019.117816

Jameson MJ, Robards TA. 2007. Coast redwood regeneration survival and growth in Mendocino county, California. West J Appl For. 22(3):171–175. https://doi.org/10.1093/wjaf/22.3.171

Kitony JK, Colt K, Abramson BW, Hartwick NT, Petrus S, Konozy EHE, Karimi N, Yant L, Michael TP. 2024. Chromosome-level baobab genome illuminates its evolutionary trajectory and environmental adaptation. Nat Commun. 15(1):8833. https://doi.org/10.1038/s41467-024-53157-w

Koch L. 2016. Chicago HighRise for genome scaffolding. Nat Rev Genet. 17(4):194–194. https://doi.org/10.1038/nrg.2016.18

Kuser J, Bailley A, Franclet A, Libby W. 1995. Early results of a range-wide provenance test of *Sequoia sempervirens*. For Genet Res. 23: 21–25.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

Li Z, McKibben MT, Finch GS, Blischak PD, Sutherland BL, Barker MS. 2021. Patterns and processes of diploidization in land plants. Annu Rev Plant Biol. 72(1):387–410. https://doi.org/10.1146/arplant.2021.72.issue-1

Li R, Zhu J. 2022. Effects of aneuploidy on cell behaviour and function. Nat Rev Mol Cell Biol. 23(4):250–265. https://doi.org/10.1038/s41580-021-00436-9

Liu G, Yong MYJ, Yurieva M, Srinivasan KG, Liu J, Lim JSY, Poidinger M, Wright GD, Zolezzi F, Choi H, *et al.* 2015. Gene essentiality is a quantitative property linked to cellular evolvability. Cell. 163(6):1388–1399. https://doi.org/10.1016/j.cell.2015.10.069

Lotterhos KE, Whitlock MC. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol. 24(5):1031–1046. https://doi.org/10.1111/mec.2015.24.issue-5

Metcalf W. 1924. Artificial reproduction of redwood (*Sequoia sempervirens*). J For. 22(8):873–893.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010(6):pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Millet C, Ausiannikava D, Le Bihan T, Granneman S, Makovets S. 2015. Cell populations can use aneuploidy to survive telomerase insufficiency. Nat Commun. 6(1):8664. https://doi.org/10.1038/ncomms9664

Moore Z. 2023. Somatic mutation accumulation in coast redwood (*Sequoia sempervirens* (D. Don) Endl.). [PhD thesis]. University of California, Davis, Davis.

Morrison WM, Armstrong AD, Webb LA, Dagley CM, Cahill KG, Berrill J-P. 2022. Performance and genetic analysis of coast redwood cultivars for afforestation of converted grassland in California. New For (Dordr). 53(1):1–16. https://doi.org/10.1007/s11056-021-09841-6

Narayan L. 2015. Clonal diversity, patterns, and structure in old coast redwood forests. [PhD thesis]. [Berkeley (CA) USA]: University of California.

Neale DB, Zimin AV, Zaman S, Scott AD, Shrestha B, Workman RE, Puiu D, Allen BJ, Moore ZJ, Sekhwal MK, *et al.*. 2022. Assembled and annotated 26.5 Gbp coast redwood genome: a resource for estimating evolutionary adaptive potential and investigating hexaploid origin. G3 (Bethesda). 12(1). https://doi.org/10.1093/g3journal/jkab380.

O'Hara KL, Cox LE, Nikolaeva S, Bauer JJ, Hedges R. 2017. Regeneration dynamics of coast redwood, a sprouting conifer species: a review with implications for management and restoration. Forests. 8(5):144. https://doi.org/10.3390/f8050144

Orr B, Godek KM, Compton D. 2015. Aneuploidy. Curr Biol. 25(13):R538–R542. https://doi.org/10.1016/j.cub.2015.05.010

Pellicer J, Powell RF, Leitch IJ. 2021. The application of flow cytometry for estimating genome size, ploidy level endopolyploidy, and reproductive modes in plants. Methods Mol Biol. 2222(2):325–361. https://doi.org/10.1007/978-1-0716-0997-2

Pineau RM, Mock KE, Morris J, Kraklow V, Brunelle A, Pageot A, Ratcliff WC, Gompert Z. 2024. Mosaic of Somatic Mutations in Earth's Oldest Living Organism, Pando.

Qu L, Hancock JF, Whallon JH. 1998. Evolution in an autopolyploid group displaying predominantly bivalent pairing at meiosis: genomic similarity of diploid *Vaccinium darrowi* and Autotetraploid *V. corymbosum* (Ericaceae). Am J Bot. 85(5):698–703. https://doi.org/10.2307/2446540

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 11(1):1432. https://doi.org/10.1038/s41467-020-14998-3

Rogers DL. 1997. Inheritance of allozymes from seed tissues of the hexaploid gymnosperm, *Sequoia sempervirens* (D. Don) Endl. (Coast redwood). Heredity (Edinb). 78(2):166–175. https://doi.org/10.1038/hdy.1997.24

Roy DF. 1966. Silvical characteristics of redwood (*Sequoia sempervirens* [D. Don] Endl.) In: Res. Paper PSW-RP-28. Berkeley, CA: Pacific Southwest Forest & Range Experiment Station Forest Service, U. S. Department of Agriculture. p. 20.

Russell W, Sinclair J, Michels KH. 2014. Restoration of coast redwood (*Sequoia sempervirens*) forests through natural recovery. Open J For. 04(2):106–111. https://doi.org/10.4236/ojf.2014.42016

Scott AD, Stenz NWM, Ingvarsson PK, Baum DA. 2016. Whole genome duplication in coast redwood (*Sequoia sempervirens*) and its implications for explaining the rarity of polyploidy in conifers. New Phytol. 211(1):186–193. https://doi.org/10.1111/nph.2016.211.issue-1

Scott AD, Zimin AV, Puiu D, Workman R, Britton M, Zaman S, Caballero M, Read AC, Bogdanove AJ, Burns E, *et al.* 2020. A reference genome sequence for giant sequoia. G3 (Bethesda). 10(11):3907–3919. https://doi.org/10.1534/g3.120.401612

Sedel-nikova TS, Pimenov AV, Efremov SP. 2023. Karyological study of an isolated Scots Pine (*Pinus sylvestris*) population in shirinskaya steppe of the Republic of Khakassia. Biol Bull. 50(5):901–911. https://doi.org/10.1134/S1062359023601301

Sharma CBSR. 1990. Chemically induced aneuploidy in higher plants. Mutagenesis. 5(2):105–126. https://doi.org/10.1093/mutage/5.2.105

Simonetti G, Bruno S, Padella A, Tenti E, Martinelli G. 2019. Aneuploidy: cancer strength or vulnerability? Int J Cancer. 144(1):8–25. https://doi.org/10.1002/ijc.v144.1

Stebbins GL. 1947. Types of polyploids: their classification and significance. Adv Genet. 1(12):403–429. https://doi.org/10.1016/S0065-2660(08)60490-3

Stebbins GL. 1948. The chromosomes and relationships of *Metasequoia* and *Sequoia*. Science. 108(2796):95–98. https://doi.org/10.1126/science.108.2796.95

Stull GW, Qu X-J, Parins-Fukuchi C, Yang Y-Y, Yang J-B, Yang Z-Y, Hu Y, Ma H, Soltis PS, Soltis DE, *et al.* 2021. Gene duplications and phylogenomic conflict underlie major pulses of phenotypic evolution in gymnosperms. Nat Plants. 7(8):1015–1025. https://doi.org/10.1038/s41477-021-00964-4

Van de Peer Y, Ashman T-L, Soltis PS, Soltis DE. 2021. Polyploidy: an evolutionary and ecological force in stressful times. Plant Cell. 33(1):11–26. https://doi.org/10.1093/plcell/koaa015

Waring KM, O'Hara KL. 2006. Estimating relative error in growth ring analyses of second-growth coast redwood (*Sequoia sempervirens*). Can J For Res. 36(9):2216–2222. https://doi.org/10.1139/x06-127

Xiong Z, Gaeta RT, Pires JC. 2011. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Natl Acad Sci U S A. 108(19):7908–7913. https://doi.org/10.1073/pnas.1014138108

Zeng D, Guan J, Luo J, Zhao L, Li Y, Chen W, Zhang L, Ning S, Yuan Z, Li A, *et al.* 2020. A transcriptomic view of the ability of nascent hexaploid wheat to tolerate aneuploidy. BMC Plant Biol. 20(1):97. https://doi.org/10.1186/s12870-020-2309-6

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. Bioinformatics. 29(21):2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Editor: E. Akhunov

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

Appendix: Read depth normalization by the PCR group

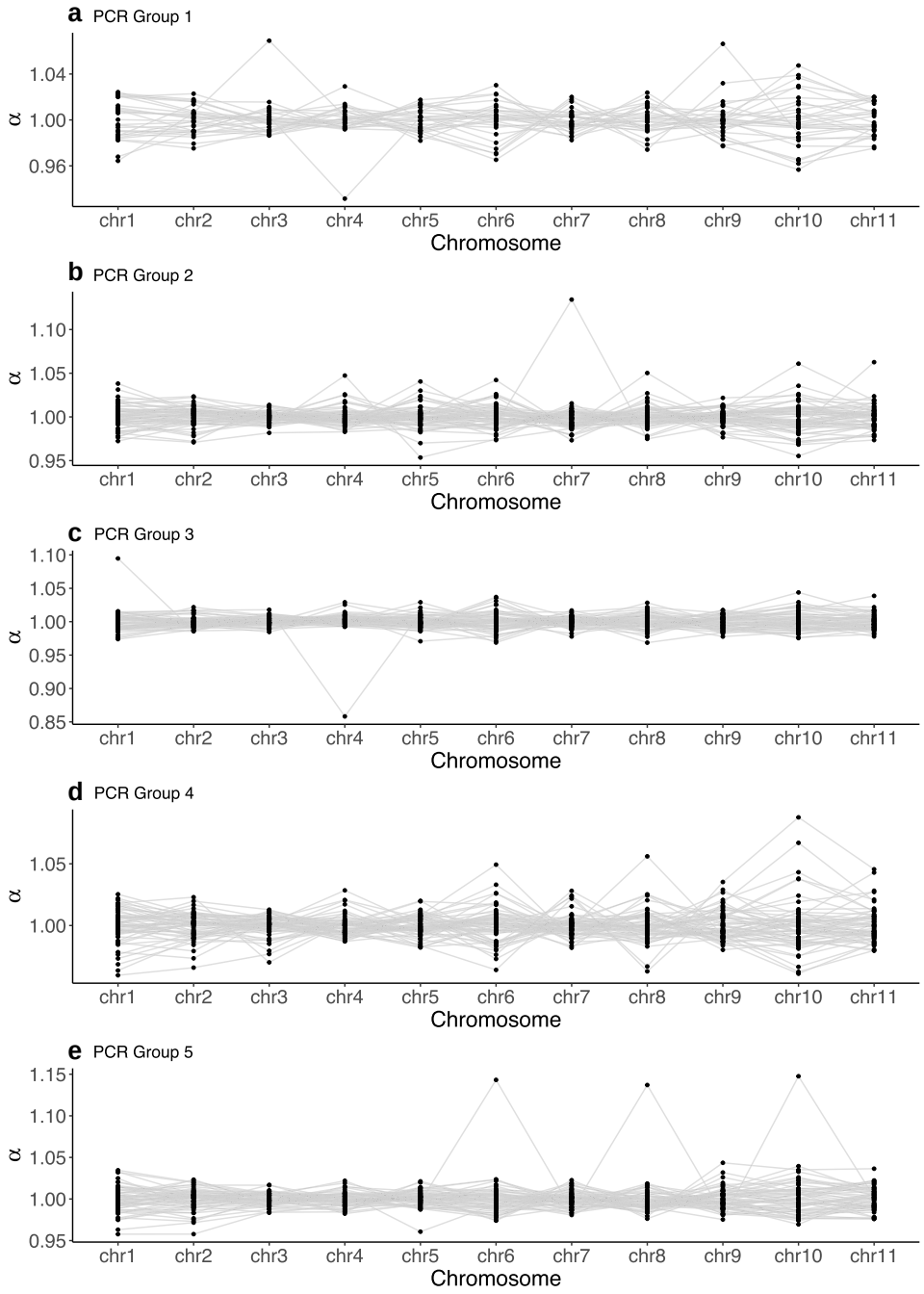


Fig. A1. Read depth normalization by PCR group, a–e) PCR groups from 1 to 5.

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025

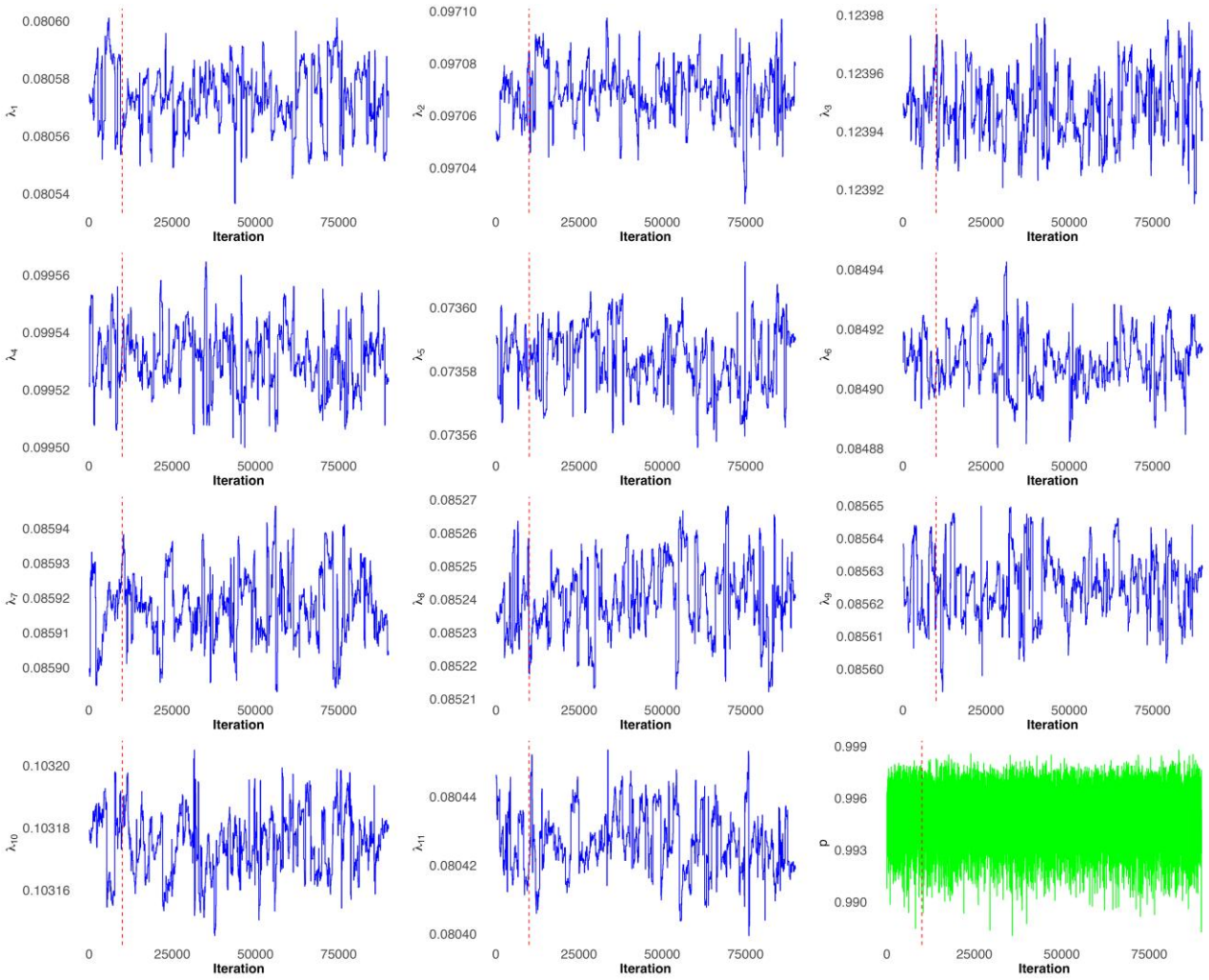


Fig. A2. MCMC trace plots.

Downloaded from https://academic.oup.com/g3journal/article/15/5/jkaf063/8092943 by The Library. Institute of Molecular Biology. University of Copenhagen. user on 23 September 2025